

Evolving origin-of-transfer sequences on staphylococcal conjugative and mobilizable plasmids—who's mimicking whom?

Karina Yui Eto^{1,2,3}, Stephen M. Kwong⁴, Patrick T. LaBreck⁵, Jade E. Crow^{2,3},
Daouda A. K. Traore^{6,7,8,9}, Nipuna Parahitiyawa², Heather M. Fairhurst², D. Scott Merrell⁵,
Neville Firth⁴, Charles S. Bond¹ and Joshua P. Ramsay^{3,2,1,*}

¹School of Molecular Sciences, University of Western Australia, Perth, WA 6009, Australia, ²Curtin Medical School, Curtin University, Perth, WA 6102, Australia, ³Curtin Health Innovation Research Institute, Curtin University, Perth, WA 6102, Australia, ⁴School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia, ⁵Department of Microbiology and Immunology, Uniformed Services University of the Health Sciences, Bethesda, Maryland, MD 20814, USA, ⁶Department of Biochemistry and Molecular Biology, Infection and Immunity Program, Monash Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia, ⁷Faculty of Natural Sciences, School of Life Sciences, Keele University, Staffordshire ST5 5BG, UK, ⁸Life Sciences Group, Institut Laue Langevin, Grenoble 38000, France and ⁹Faculté des Sciences et Techniques, Université des Sciences, des Techniques et des Technologies de Bamako (USTTB), Bamako BP E423, Mali

Received March 23, 2021; Revised April 09, 2021; Editorial Decision April 13, 2021; Accepted April 14, 2021

ABSTRACT

In *Staphylococcus aureus*, most multiresistance plasmids lack conjugation or mobilization genes for horizontal transfer. However, most are mobilizable due to carriage of origin-of-transfer (*oriT*) sequences mimicking those of conjugative plasmids related to pWBG749. pWBG749-family plasmids have diverged to carry five distinct *oriT* subtypes and non-conjugative plasmids have been identified that contain mimics of each. The relaxosome accessory factor SmpO, encoded by each conjugative plasmid, determines specificity for its cognate *oriT*. Here we characterized the binding of SmpO proteins to each *oriT*. SmpO proteins predominantly formed tetramers in solution and bound 5'-GNNNC-3' sites within each *oriT*. Four of the five SmpO proteins specifically bound their cognate *oriT*. An F7K substitution in pWBG749 SmpO switched *oriT*-binding specificity *in vitro*. *In vivo*, the F7K substitution reduced but did not abolish self-transfer of pWBG749. Notably, the substitution broadened the *oriT* subtypes that were mobilized. Thus, this substitution represents a potential evolutionary intermediate with promiscuous DNA-binding specificity that could facilitate a switch between *oriT* specificities. Phylogenetic analysis suggests pWBG749-family plasmids have switched *oriT*

specificity more than once during evolution. We hypothesize the convergent evolution of *oriT* specificity in distinct branches of the pWBG749-family phylogeny reflects indirect selection pressure to mobilize plasmids carrying non-cognate *oriT*-mimics.

INTRODUCTION

The majority of horizontally acquired antimicrobial-resistance genes in *Staphylococcus aureus* are located on circular DNA plasmids. An estimated ~90% of *S. aureus* isolates possess at least one plasmid and ~79% carry large (>20 kb) plasmids frequently harbouring collections of resistance and virulence genes (1). Evidence for the movement of plasmids between distinct lineages of *S. aureus* and their introduction from other species and genera is abundant (2). However, the majority of *S. aureus* plasmids lack conjugation genes and therefore must depend on self-transmissible mobile elements such as bacteriophage or conjugative plasmids for their horizontal transfer (1,3).

The pWBG749 family of plasmids are a recently characterized family of conjugative staphylococcal plasmids that are distinct from the well characterized pSK41/pGO1 family. pWBG749-family plasmids have been identified carrying genes for resistance to methicillin, vancomycin, penicillin, gentamicin, trimethoprim, mupirocin, cadmium and chlorhexidine (4–11). As well as directly disseminating antimicrobial resistance, pWBG749-family plasmids can mobilize numerous non-conjugative multiresistance plasmids

*To whom correspondence should be addressed. Email: josh.ramsay@curtin.edu.au

(7,8). In most well-understood mechanisms of conjugative mobilization, a mobilizable plasmid carries both a DNA relaxase gene and an origin-of-transfer (*oriT*) sequence. The expressed relaxase protein recognizes and nicks the mobilizable plasmid *oriT* and recruits ssDNA to the conjugative-plasmid-encoded type IV secretion system for transfer. In contrast, documented plasmids mobilized by pWBG749, lack their own relaxase gene and instead carry sequence mimics of the pWBG749 origin-of-transfer (*oriT*). The pWBG749-encoded relaxase therefore acts *in trans* on the mobilizable-plasmid's *oriT* mimic. An estimated 50% of sequenced non-conjugative *S. aureus* plasmids carry one or more pWBG749-like *oriT* sequences (7). *oriT* mimic-carrying plasmids have been subsequently identified in *Escherichia coli* and *Acinetobacter baumannii* (12–14), with ~20% sequenced *A. baumannii* plasmids carrying *oriT* sequences resembling those on conjugative plasmids (15). A search for *oriT* sequences in 4602 plasmids spanning 22 distinct phyla revealed that while only 29% carried recognizable relaxase genes, 69% carried at least one *oriT* sequence (16). These observations suggest the largely unexplored 'relaxase *in trans*' mechanism of plasmid mobilization is an underappreciated route for horizontal gene transfer in both gram-positive and gram-negative bacteria (1,7).

The *oriT* sequences of pWBG749-family plasmids have diverged into at least five distinct subtypes based on nucleotide identity, previously named OT49, OT45, OTUNa, OT408 and OTSep (7) (Figure 1). Each *oriT* subtype contains an identical core sequence targeted by the putative relaxase SmpP and a conserved arrangement of DNA-repeat sequences. The repeat sequences IR1 and IR3 are largely conserved in sequence between subtypes, but the IR2 repeat sequences have diverged. pWBG749 carries an OT49 *oriT* and can mobilize plasmids carrying OT49 sequences, but not those carrying OT45 or OTUNa. Conversely, conjugative plasmids pC02 and pWBG731 carry OTUNa and OT45 *oriT* subtypes, respectively, and both can mobilize plasmids carrying either OTUNa or OT45 *oriT*, but not OT49. While the pC02 and pWBG731 plasmids share the same *oriT* specificities, their conjugation-gene sequences are more divergent in comparison to pWBG731 and pWBG749 (17), suggesting that *oriT* specificity of each plasmid has evolved somewhat independently of rest of the conjugation-gene cluster. These same five *oriT* subtypes are also present as *oriT* mimics on non-conjugative plasmids and enable mobilization by pWBG749 plasmids with a matching *oriT* sequence subtype specificity (4,7,17). Some plasmids have captured 2 or 3 *oriT* mimic sequences and most often they are of different subtypes. The observation that non-conjugative plasmids have captured and retained these distinct *oriT* variants suggests these plasmids have benefited from the horizontal mobilization facilitated by various pWBG749-family members carrying these distinct *oriT*s during their evolutionary history.

A conjugative plasmid's ability to discriminate between its cognate *oriT* and other DNA sequences is presumably advantageous for both a conjugative plasmid and the bacterial host (18). Relaxase-*oriT* recognition often involves additional accessory DNA-binding proteins or 'relaxosome accessory factors' (RAFTs), which can be essential for efficient *oriT* recognition and nicking by the relax-

ase. Characterized RAFTs frequently contain a ribbon-helix-helix (RHH) DNA-binding domain within their N-termini, including TraY_F, TraM_F, TrwA_{R388}, NikA_{RP64}, TraJ_{RP64}, MobC_{RSF1010} and MbeC_{CoIE1} in gram-negative bacteria, and PcfF_{pCF10}, AuxI_{pLS20}, AusX2_{pLS20} and SmpO_{pWBG749} in gram-positive bacteria (7,19–27). RAFTs generally bind sequence(s) within or 5' of the *oriT* *nic* site. RHH domains contain an antiparallel β -sheet formed from the N-termini of the two RHH monomers, which recognizes specific DNA base-pairs in the major groove. RAFTs may also interact with the relaxase and/or the type-IV secretion system coupling protein (28).

The pWBG749-family SmpO proteins are RHH-containing RAFTs encoded between the *oriT* site and downstream *smpP* relaxase gene on each pWBG749-family plasmid. While pWBG749 does not normally mobilize plasmids carrying the OT45 *oriT* it can efficiently mobilize an OT45 *oriT*-carrying plasmid if the plasmid additionally carries the *smpO*₄₅ gene (7). This indicates variations in the *smpO* genes are likely responsible for differences in mobilization specificity between various pWBG749-family plasmids. In this work, we identified the DNA-binding sites for five distinct SmpO proteins using surface plasmon resonance (SPR) and found SmpO proteins specifically bind two sites within each *oriT*. Amino acid substitutions in SmpO₄₉ and SmpO₄₅ revealed a single change switched *oriT* specificity *in vitro* and *in vivo* and phylogenetic comparisons of conjugation-gene sequences suggested similar changes have occurred in divergent members of the pWBG749 family of conjugative plasmids.

MATERIALS AND METHODS

Conjugation and mobilization

Conjugation experiments were carried out as previously described (7).

Strain and plasmid construction

Construction and sequencing of pLI50 plasmid constructs was carried out as previously described (7,17). Details of strains, plasmids and their construction can be found in Supplementary Table S1 and oligonucleotides used for cloning can be found in Supplementary Table S2. For construction of pWBG749e-F7K, the *smpO*_{49-F7K} allele was introduced by allelic exchange using pIMAY-Z (29). Oligonucleotides #68–71 (Supplementary Table S1) were used to amplify two overlapping fragments (1–2 and 3–4) from pWBG749e. The two fragments and linearized pIMAY-Z were then assembled using Gibson Assembly (NEB). The desired pIMAY-Z construct was validated by DNA sequencing, then introduced into *S. aureus* WBG4515 harbouring pWBG749e. Blue colonies on NYE agar plates containing erythromycin (5 μ g/ml), chloramphenicol (10 μ g/ml) and X-gal (40 μ g/ml) were switched between permissive and non-permissive temperatures for pIMAY-Z replication to induce integration, excision and then plasmid loss as described previously (29). The pWBG749e-F7K plasmid was screened by PCR and confirmed by Sanger sequencing.



Figure 1. pWBG749-family *oriT* sequences examined in this work. Alignments of pWBG749-family *oriT* sequences examined in this work are shown; inverted repeats are shaded. The pWBG749-family *oriT* sequences include 1–3 copies of the AR repeat sequence (AR1–AR3); however, only the AR3 copy is essential for conjugative mobilization and AR1–AR2 are not shown here. The OTSep and OT408 sequences are aligned separately as they contain distinct AR sequences. Asterisks under alignments indicate positions with 100% nucleotide conservation. The IR2 sequences vary between *oriT* subtypes. Each of the *oriT* sequences shown are located on conjugative plasmids, except for those of pWBG762; pWBG762 carries mimics of OT49, OTUNA and OT45 (4,17). The 5'-GNNNC-3' motifs central to each of the *oriT* specificity sequences *ossA* and *ossB* discovered in this work are underlined. The *ossA* sites are shown in italics if they are on the complementary strand relative to the *ossB* site.

Protein purification

SmpO₄₉ and SmpO₄₅ coding sequences were amplified by PCR respectively from pWBG749 and pWBG745. SmpO_{49-F7K} and SmpO_{45-K7F} coding sequences were amplified from pLIOT5S9M-F7K and pLIOT9S5M-K7F. Coding sequences for other SmpO variants SmpO_{UNa}, SmpO_{Sep} and SmpO₄₀₈ were codon optimized using JCaT (<http://www.jcat.de>) (30) and synthesized (IDT). PCR products and synthesized DNA fragments were digested and cloned into NcoI/BamHI or XbaI/BamHI sites of pETM-11. All constructs were introduced to *E. coli* BL21(DE3)pLysS by electroporation. Transformed cells were grown in 1 L of LB supplemented with chloramphenicol (Cm) at 100 µg/ml and kanamycin (Km) at 50 µg/ml and incubated at 25°C and 200 rpm shaking. At an optical density at 600 nm of approximately 0.6–0.8, cells were induced with isopropyl-β-D-1-thiogalactopyranoside (IPTG) at 0.5 mM final concentration for 16 h at 18°C and 180 rpm shaking. The cells were harvested by centrifugation at 8000 g for 20 min at 4°C; 1 L culture cell pellets were gently washed in nickel binding buffer [50 mM NaH₂PO₄, pH 7.5, 1 M NaCl, 10% (v/v) glycerol, 25 mM imidazole] and centrifuged again at 8000 g for 20 min at 4°C; cell pellets were immediately used or stored at -80°C. The cell pellets were resuspended in 50 ml of nickel-affinity binding buffer supplemented with 2 µg/ml of DNase I nuclease and 300 µg/ml of lysozyme. Lysis was carried out with Emulsiflex C5 high-pressure homogeniser (Avestin), and the lysate was clarified by centrifugation (24 000 g for 45 min at 4°C). The clarified lysate was filtered (0.22 µm) prior to application onto 5 ml NiCl₂-charged HisTrap HP column (GE Healthcare). The cleavable hexahistidine-tagged SmpO protein (6H-SmpO) was eluted using an imidazole gradient (25–500 mM) over eight column-volumes. Eluted 6H-SmpO protein was diluted to 2 mg/ml with TEV protease digestion buffer [50 mM Tris-HCl pH 7.5, 250 mM NaCl, 1 mM EDTA, 5% (v/v) glycerol, 1 mM DTT], uncleavable hexahistidine-tagged TEV protease (produced in-house) was added at protease:SmpO ratio of 1:10 (w/w), then dialysed into TEV protease digestion buffer for 16 h at ambient temperature with gentle mixing. Post-digestion, dialysed SmpO TEV protease digestion reaction was centrifuged (24 000 g for 10 min at 4°C), 0.22 µm filtered and re-applied to the HisTrap col-

umn. The flowthrough containing the tag-cleaved SmpO protein was further purified using a HiLoad 16/60 Superdex 75 column (GE Healthcare) preequilibrated in SEC buffer [50 mM Tris-HCl, pH 7.5, 1 M NaCl, 1 mM EDTA, 5% v/v glycerol]. The concentration was determined from absorption at 280 nm using absorption coefficients and theoretical molecular mass values calculated from sequence using *ExPasy ProtParam* (31). Purified SmpO proteins were used in following experiments or flash-frozen with liquid nitrogen for long-term storage at -80°C. Chromatography purification steps were performed using AKTA Start and/or AKTA purifier FPLC system (GE Healthcare) at 4°C and absorbance traces at 280 nm only (AKTA Start), or 280, 260 and 230 nm (AKTA Purifier) were constantly monitored.

SEC-MALS

All SEC-MALS experiments were carried out using Superdex Increase 10/300 GL column (GE Healthcare) attached to Viskotek GPCmax VE 2001 solvent/sample module (Malvern) coupled to Viskotek 305 TDA detector array (Malvern) at room temperature. In summary, 200 µl of purified SmpO protein samples at 2–3 mg/ml in MALS buffer [10 mM HEPES, pH 7.5, 250 mM NaCl, 3 mM EDTA, 2.5% (v/v) glycerol, 0.05% (v/v) Tween 20] and bovine serum albumin (BSA) (Sigma) samples of approximately 1 mg/ml in MALS buffer were applied to the size-exclusion column pre-equilibrated with MALS buffer at flow rate of 0.3 ml/min, monitoring the refractive index, UV absorbance and left and right-angle light scattering. OmniSEC 5.10 Bio software (Malvern) was used to analyse SEC profile and to calculate molecular weight averages and dispersity using calibration settings derived from five BSA samples, then results were averaged.

Electrophoretic mobility shift assays (EMSAs)

The *oriT*₄₅ and *oriT*₄₉ regions were PCR amplified from pWBG745 and pWBG749e using primers carrying 3' adapter sequences (Supplementary Table S3). The PCR products were used as competitive unlabelled DNA in the EMSAs and amplified using IRDye800-labelled primers targeting the tag-sequences (Supplementary Table S3).

Binding reactions were set up as per Li-Cor[®] EMSA using IRDye[®] oligonucleotides protocol, except that reaction mixtures contained 0.5 $\mu\text{g}/\mu\text{l}$ sheared Herring sperm DNA, IRDye800-labelled DNA at 5 nM final concentration. SmpO₄₅, SmpO₄₉ and SmpO_{49-F7K} were diluted into 1 \times binding buffer containing 2.5% glycerol and 0.05% Tween 20, then added to the binding reactions at various concentrations. Native acrylamide/agarose hybrid gel electrophoresis was carried out in pre-run gels containing 3% 19:1 acrylamide:bis-acrylamide, 0.5% agarose, 0.5 \times TBE buffer (1 \times TBE = 89 mM Tris, 89 mM boric acid, 2 mM EDTA), 2.5% glycerol, run in 0.5 \times TBE buffer.

SPR-based DNA footprinting assays

All SPR experiments were carried out using a Biacore T200 (GE Healthcare), Biacore SA sensor chip (GE Healthcare) and applying the Re-usable DNA Capture Technique (ReDCaT) (32). The oligonucleotide arrays for the *oriT* sequences from pWBG749e, pWBG745, pC02 and *S. aureus* strain M0408 (GI:477787193) were designed using the Perl script *poop.pl* (32), and *ossA*_{sep} and *ossB*_{sep} sequences were originated from *S. aureus* strain VCU120 (GI:41866860). The oligonucleotides used on SPR assays are listed in Supplementary Table S3. A DNA-binding affinity protocol was derived from the method of Stevenson *et al.* (32). In each cycle of screening assays, unless stated, cycle steps were run at 30 $\mu\text{l}/\text{min}$; test DNA was capture on Fc2, and control ReDCaT complementary linker was capture on Fc1 both at 1 μM and 10 $\mu\text{l}/\text{min}$; subsequently SmpO protein at 1 μM was injected onto Fc1 and Fc2 for 60 s contact time and 60 s dissociation time, followed by a wash with regeneration buffer [50 mM NaOH, 1 M NaCl] and 90 s stabilization with SPR buffer [10 mM HEPES, 150–300 mM NaCl, 3 mM EDTA, 0.05% (v/v) Tween 20]. To maintain SmpO protein stability in solution and optimise DNA-binding in SPR, SmpO₄₅, SmpO₄₀₈ and SmpO_{UNa} SPR screening assays were carried out in SPR buffer containing 300 mM NaCl, while SmpO₄₉ SPR screening assays were carried out in SPR buffer with 150 mM NaCl. Assays presented in Supplementary Table S4 where all five SmpO proteins were tested together were carried out in SPR buffer containing 250 mM NaCl. Complete SPR data are also presented in Supporting Dataset S1.

RESULTS

SmpO proteins bind two *oriT*-specificity-sites *ossA* and *ossB*

Five distinct pWBG749-family SmpO proteins SmpO₄₉ (Genbank accession ACZ58780 from plasmid pWBG749), SmpO₄₅ (ACZ58661 from pWBG745, identical to SmpO of pWBG731; AXQ85809), SmpO_{UNa} (AKR47467 from plasmid pC02), SmpO₄₀₈ (ENK02073 from *S. aureus* MO408) and SmpO_{sep} (EHR84364 from *S. epidermidis* VCU120) were purified using nickel-affinity chromatography and size-exclusion chromatography (SEC) (see ‘Materials and Methods’ section). SEC coupled with multi-angle light scattering (SEC-MALS) analysis revealed SmpO₄₉, SmpO₄₅, SmpO₄₀₈ and SmpO_{sep} each existed predominantly as tetramers (84–96.6%) in solution, with smaller

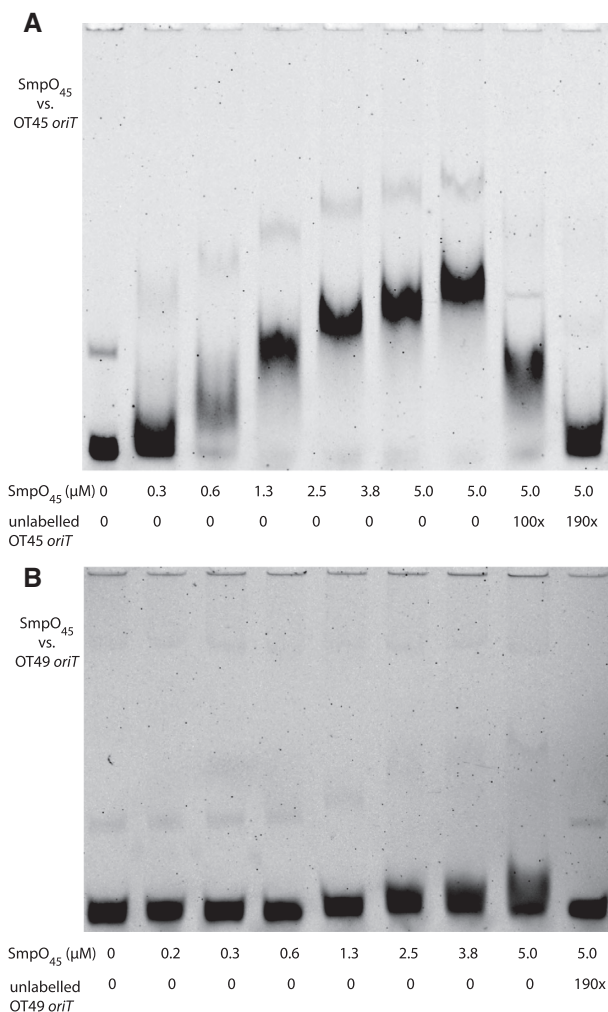


Figure 2. EMSA of SmpO₄₅ with *oriT* regions amplified from OT45 and OT49. Purified SmpO₄₅ was used in EMSA assays with a 5 nM of a 134-bp dual-IRDye800-labelled DNA region encompassing AR3-IR3 for *oriT* sequences OT45 and OT49 (Figure 1). (A) EMSA with SmpO₄₅ and labelled OT45. The final two lanes include unlabelled competitor OT45 DNA. (B) EMSA with SmpO₄₅ and labelled OT49. The final lane includes unlabelled competitor OT49 DNA. Secondary labelled DNA fragments with an apparent higher molecular weight were visible for both labelled OT45 and OT49 and were shifted in a similar manner to the main PCR products.

proportions as hexamers (3.4–16%) (Supplementary Table S5). We were unable to purify enough SmpO_{UNa} for SEC-MALS.

We first examined binding of SmpO₄₅ to the pWBG745 OT45 *oriT* (*oriT*₄₅) using EMSAs. IRDye800-labelled oligonucleotides were used to PCR-amplify a 134-bp DNA fragment from the *oriT*₄₅ sequence encompassing repeats AR3-IR3 (Figure 1). Purified SmpO₄₅ induced a series of DNA-migration shifts in the presence of increasing concentrations of SmpO₄₅ (Figure 2A). Addition of excess unlabelled *oriT*₄₅ DNA outcompeted labelled DNA and reversed the shifts. EMSAs using SmpO₄₅ with IRDye800-labelled *oriT*₄₉ DNA produced only a minor shift at high SmpO₄₅ concentrations, confirming SmpO₄₅ exhibited higher specificity for its cognate *oriT*₄₅ sequence than the non-cognate *oriT*₄₉ (Figure 2B). The PCR products for

both *oriT*₄₅ and *oriT*₄₉ contained a minor secondary species exhibiting slower gel migration that also shifted in EMSAs with increasing SmpO₄₅ concentrations. Gel purification did not remove this DNA species nor did denaturing and reannealing the DNA. This secondary product also increased with PCR-product age, suggesting some alternate secondary DNA structure formed during storage. The IR1–IR3 repeats are structurally conserved in *oriT* regions of diverse gram-positive conjugative plasmids and potentially form a ssDNA branched hairpin structure, so it is possible that these secondary structures could be the cause of this secondary band in EMSAs (7). In summary, the complex series of DNA-shifts for the *oriT*₄₅ region with SmpO₄₅ suggested there may be multiple binding sites for SmpO₄₅ in the region, multiple DNA isoforms within complexes and/or that SmpO₄₅ formed higher-order oligomers or soluble aggregates with DNA at higher concentrations.

To identify sequences recognized by each SmpO protein, DNA-binding assays were carried out by SPR using ‘Reusable DNA Capture Technology (ReDCaT)’ (32,33). Briefly, a biotinylated ssDNA oligonucleotide was permanently immobilized to a streptavidin-coated SPR chip (the ReDCaT chip), and dsDNA oligonucleotides with a 3′ overhang complementary to the biotinylated oligonucleotide were successively bound to and released from the ReDCaT chip. The SPR response for each oligonucleotide, corresponding to a change in mass on the chip’s surface, was used as a baseline from which SPR responses were measured following addition of purified SmpO protein (see ‘Materials and Methods’ section and Figure 3). Tiled arrays of 30-bp dsDNA oligonucleotides spanning the OT45, OT49, OTUNa and OT408 *oriT* sequences were synthesized (Figure 3A). For SmpO₄₅ (Figure 3), SPR responses corresponding to ~30% of the theoretical maximum response (R_{max}) were detected with oligonucleotides spanning two distinct regions within the OT45 *oriT*, named here ‘*oriT* specificity sequence A’ (*ossA*₄₅) and *ossB*₄₅. *ossB*₄₅ was located within the right-hand arm of the IR2 repeat sequence proximal to the core region, confirming our previous speculations that the more divergent IR2 repeats within each *oriT* were involved in mobilization specificity (7,17). The *ossA*₄₅-binding site was positioned ~50-bp upstream of IR2, between AR3 and IR3. Both *ossA*₄₅ and *ossB*₄₅ contained the sequence 5′-TGGGGTCA-3′. Reinspection of each of the other four *oriT* subtypes revealed they too carried sequences in a similar position that resembled the arm sequences of the IR2 region. The putative *ossA* sites for SmpO_{Sep} and SmpO_{UNa} were present on the opposite DNA strand relative to their respective *ossB* site sequences (Figures 1 and 3). SPR footprinting with purified SmpO₄₉, SmpO_{UNa} and SmpO₄₀₈ proteins and their corresponding oligonucleotide arrays confirmed they all also bound their respective *ossA* regions; however, SmpO₄₉, SmpO_{UNa} and SmpO₄₀₈ each exhibited weaker responses with their respective *ossB* sites.

To refine the DNA sequences critical for binding by each SmpO protein, we first aligned all the *ossA* and *ossB* sites to identify any common features of the *oss* sites. This revealed a conserved 5′-GNNNNC-3′ motif present within all *ossA* and *ossB* sites, which was often centred within or around a distinct region of dyad symmetry in each sequence (Figure 4A). We then designed 16-bp dsDNA oligonucleotides

containing each GNNNNC motif for *ossA*₄₉, *ossA*₄₅ and *ossA*₄₀₈. For each of these 16-bp *ossA* regions, we designed a further 16 dsDNA oligonucleotides each containing a single-base-pair substitution to the complementary base-pair for each of the 16 nucleotide positions (Figure 4B). SPR assays using these oligonucleotides with their corresponding SmpO proteins demonstrated that mutations within each GNNNNC motif were most detrimental to SmpO binding. Finally, to examine the specificity of SmpO binding for cognate *ossA* and *ossB* sites, 16-bp dsDNA oligonucleotides for each of *ossA* and *ossB* site, including predicted GNNNNC sites for the OTSep *ossA* and *ossB* sites (Figure 4A), were used in SPR experiments with each of the five purified SmpO proteins. To enable an unbiased comparison between all the proteins in a single SPR run, SPR buffer containing an intermediate but suboptimal NaCl concentration (250 mM) was used, which unfortunately substantially reduced several of the SPR responses. However, despite these limitations, each SmpO protein exhibited the strongest SPR responses for its cognate *ossA* and *ossB* sequences (Supplementary Table S4). SmpO_{UNa} also exhibited a binding response with *ossA*₄₅, consistent with the sequence similarity between the *ossA*_{UNa} and *ossA*₄₅ sites and the ability of pC02 to mobilize plasmids carrying OTUNa or OT45 *oriT* subtypes (4,17).

Incongruence between conjugative-plasmid phylogeny and *oriT* specificity

As noted here and previously (4,17), there are clear incongruences between the nucleotide sequence similarities of pWBG749-family conjugation-gene clusters and the *oriT* subtypes they carry. Alignment of the conjugation-gene cluster sequences carrying each of the five *oriT* subtypes (Figure 5A) (34–37) clustered the pWBG731/pWBG745 conjugation genes with those of pWBG749, despite these plasmids carrying distinct *oriT* recognized by SmpO proteins with distinct specificities. Conversely, the pC02 conjugation-gene cluster has diverged from pWBG731/pWBG745 both in nucleotide sequence similarity and gene order, yet these conjugative plasmids mobilize the same *oriT* sequence subtypes (OT45 and OTUNa). To search for other such discrepancies, we carried out BLASTN searches against whole-genome shotgun sequences using the *smpN-oriT-smpO* region from pC02 and looked for sequence divergence in the *oriT* region. A 39-kb contig from *S. aureus* M17027 (JGJU01000023.1) carrying a complete *smpA-X* cluster was identified. Nucleotide alignments of pC02 and pM17027 revealed the *smpA-smpN* and *smpP-smpU* regions share 96–99% nucleotide identity, while the *oriT* sequence (aligned from AR3-core, Figure 1) and *smpO* genes share only 67% and 68% nucleotide identity, respectively. The pM17027 *oriT* sequence contained the *ossA*₄₉ and *ossB*₄₉ sequences 5′-GATATCA-3′ and 5′-GATAGCA-3′, identical to those on pWBG749, not the *ossA*_{UNa} and *ossB*_{UNa} sequences present on pC02. Alignment of the pM17027 and pWBG749 *oriTs* revealed they share 78% nucleotide identity, but the *smpO* sequences from these same plasmids share only 59% identity.

Next, we investigated whether the *smpO* alleles carried by pWBG749 and pM17027 could enable mobilization via

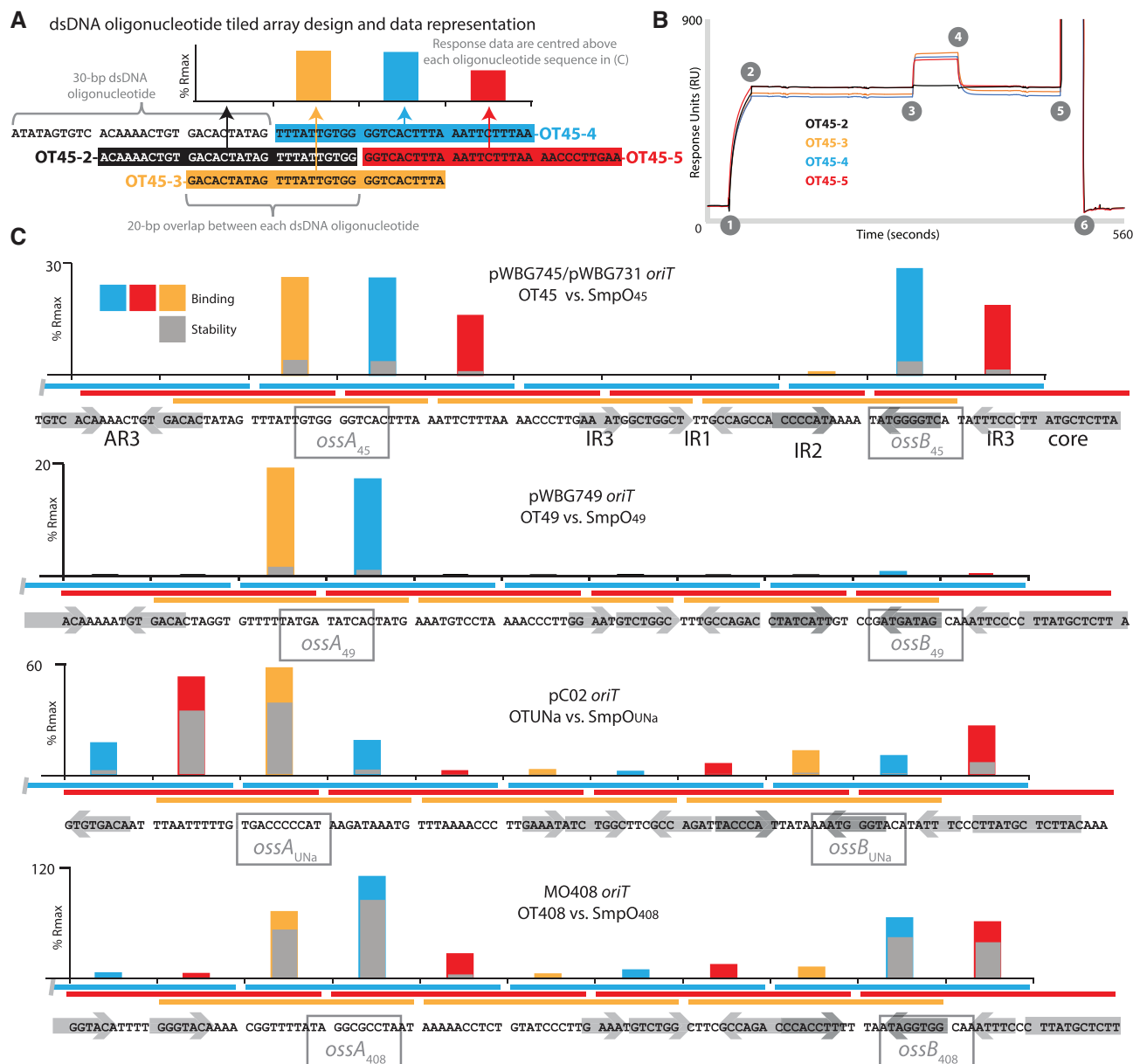


Figure 3. Surface plasmon resonance DNA-footprinting of SmpO and *oriT* regions. (A) Diagram outlining the design of dsDNA oligonucleotide arrays for each of the *oriT* regions using dsDNA oligonucleotides OT45–2, OT45–3, OT45–4 and OT45–5 as examples. SPR R_{\max} values are indicated in the graph above, with each bar corresponding to the oligonucleotide centred at that position and highlighted in the same colour. (B) Example SPR response data corresponding to oligonucleotides as labelled in (A). Numbers on the graph indicate different steps in the SPR assay process. (1) DNA capture, (2) end of DNA capture flowcell wash, (3) addition of 1 μ M purified SmpO protein, (4) protein addition ceases, (5) flowcell regeneration step starts, removing dsDNA oligonucleotide and (6) regeneration step ceases and flowcell is washed with SPR buffer prior to the next cycle with the next oligonucleotide. Responses are measured between steps 3 and 4 from the baseline established between steps 2 and 3. (C) SPR binding and stability responses for 1 μ M of each purified SmpO protein to *oriT* oligonucleotide arrays for OT45, OT49, OTUNa and OT408 (from top to bottom). The sequence of each *oriT* array examined is shown at the bottom of each graph and AR3, IR1–IR3 and the core regions are shaded. Between each graph and *oriT* sequence are coloured bars representing the dsDNA oligonucleotides tested in each SPR array, corresponding to the sequence directly below and the similarly coloured bars in the graph above each set of tiles. SPR responses (coloured bars) are expressed as a percentage of R_{\max} , the theoretical maximum binding response expected assuming a tetrameric SmpO protein binds to a single dsDNA oligonucleotide, and are calculated from the molecular weights of the oligonucleotides and SmpO tetramers. Stability measurements (grey bars) are the measured R_{\max} response 10 s after step 4 indicated in part (B). Complete SPR data are presented in Supporting Dataset S1.

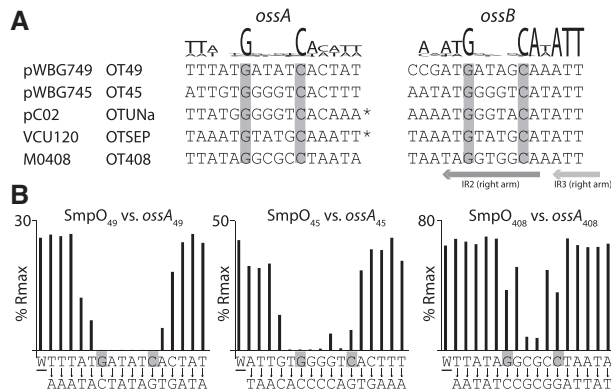


Figure 4. Single-nucleotide mutagenesis of OT49, OT45 and OT408 *oriT* sites. (A) An alignment of *ossA* and *ossB* sequences for each of the five *oriT* sequences examined in this work. The conserved G and C nucleotides centred about each *oss* site are highlighted in grey and a sequence logo indicating nucleotide conservation is shown above each alignment. The positions of the right-arms of the IR2 and IR3 repeats are indicated below as arrows. (B) Single-nucleotide mutagenesis of *ossA* sites and SPR responses with SmpO proteins. Sixteen-base-pair sequences containing *ossA*₄₉, *ossA*₄₅ and *ossA*₄₀₈ were synthesized and assayed by SPR with 1 μ M of each cognate purified SmpO protein as indicated. W indicates the SPR response for the wild-type sequence, while each of the other sequence letters represent a dsDNA oligonucleotide containing a single base-pair substitution to the complementary base-pair at that position as indicated below each base. Complete SPR data are presented in Supporting Dataset S1.

the OT49 *oriT* by pWBG731 and pC02 (Table 1). Vectors containing the pWBG749 *oriT* positioned upstream of cloned *smoO*₄₅, *smoO*₄₉ or synthesized *smoO*_{M17027}, were introduced into *S. aureus* RN4220, after which pWBG731 or pC02 were introduced into each strain to produce conjugative donors. As expected, pWBG731 and pC02 did not mobilize a plasmid carrying the OT49 *oriT* alone or a plasmid carrying the OT49 *oriT* with the *smoO*₄₅ or *smoO*_{M17027}. These results confirmed that the newly identified *smoO*_{M17027} enabled recognition of the OT49 *oriT* and that the two distantly-related plasmids pWBG731 and pC02 both mobilized plasmids carrying the non-cognate OT49 *oriT* if provided a compatible *smoO* gene.

The amino acid sequence similarities between the SmpO proteins also appeared incongruent with the *oriT* sequences they mobilized (Figure 5B). SmpO₄₉ and SmpO₄₅ share 82% amino acid identity, yet they recognize distinct *oriT* sequences (OT49 and OT45/OTUNa, respectively). Conversely, SmpO₄₉ shares only 46% amino acid identity with SmpO_{M17027} and yet they both enabled plasmid mobilization via the OT49 *oriT* (Table 1). SmpO_{M17027} instead appears more closely related in sequence to SmpO_{UNa} (64% identity Figure 5B), in line with relatedness of the other pM17027 and pC02 conjugation-gene sequences. These observations, taken together, suggested that members of the pWBG749-family of conjugative plasmids have switched *oriT* specificity between OT49 and OT45/OTUNa at least once in the pWBG749/pWBG745/pWBG731 and/or the pC02/pM17027 branches of the pWBG749 family (Figure 5A).

A single amino acid substitution in SmpO switches mobilization specificity

Given that SmpO₄₉ and SmpO₄₅ share 86% amino-acid identity, we reasoned that *oriT* specificity differences between the two proteins involved a small number of amino acid residues. SmpO₄₉ and SmpO₄₅ differ by 8 of the first 17 amino acid residues, corresponding with the predicted β -sheet and first α -helix of the RHH domain, but only by 4 of the following 67 residues (Figure 5B). To determine if the first 17 residues of SmpO governed *oriT* specificity, we replaced the coding sequence for the first 17 amino acids of SmpO₄₉ with that of SmpO₄₅ and placed this gene downstream of the OT45 *oriT* (Figures 5B and 6). This plasmid (pLIOT5S9M-N-WT45) was mobilized by pWBG749e (8) with similar efficiency to a plasmid carrying the wild-type *smoO*₄₅ gene (pKY5TO) (Figure 6 and Supplementary Table S6). Likewise, a construct carrying the OT49 *oriT* and a gene fusion encoding the first 17 amino acids of SmpO₄₉ fused the last 67 residues of SmpO₄₅ (pLIOT9S5-N-WT49) was mobilized by both pWBG731 and pC02 with similar efficiency to a plasmid carrying OT49 and the complete *smoO*₄₉ gene (pLIT9O99) (Table 1). These experiments demonstrated the SmpO residues discriminating between the OT49 and OT45 *oriT* sequences are located within the first 17 amino acids of each protein.

We next constructed a range of mutant *smoO*₄₉ alleles, each encoding 2–4 amino acid substitutions to match residues present in SmpO₄₅ (Figure 6). Eleven out of twenty constructs were mobilized by pWBG749e and each of the mobilized plasmids carried a F7K substitution (Figure 6 and Supplementary Table S6). A construct encoding SmpO₄₉ with an F7K substitution alone (SmpO_{49-F7K}) was also mobilized as efficiently as a vector carrying wild-type *smoO*₄₅. Similarly, we demonstrated that *smoO*₄₅ containing an encoded K7F substitution (SmpO_{45-K7F}) enabled mobilization of a plasmid carrying the OT49 *oriT* by pWBG731 and pC02 (Table 1). In summary, this single amino acid position in both SmpO₄₉ and SmpO₄₅ appeared responsible for the differences in specificity between OT49 and OT45 *oriT* subtypes.

To see if the *in vitro* DNA-binding specificity of the substituted SmpO proteins mirrored the observed changes in mobilization specificity, SmpO_{49-F7K} and SmpO_{45-K7F} were purified and tested for their ability to bind each of the *ossA* and *ossB* sequences. Both proteins formed tetramers and hexamers in similar proportions to their wild-type counterparts in SEC-MALS experiments (Supplementary Table S5), confirming changes did not alter the oligomerization state of the proteins. While we were unable to detect any DNA binding for SmpO_{45-K7F} in SPR experiments or EMSAs (not shown), the SmpO_{49-F7K} protein bound the non-cognate *ossA*₄₅ and *ossB*₄₅ sequences as strongly as SmpO₄₅ (Supplementary Table S4). Moreover, while SmpO₄₉ was unable to shift OT45 DNA in EMSA assays (Supplementary Figure S1A), SmpO_{49-F7K} produced a similar series of DNA shifts (Supplementary Figure S1B) to that of SmpO₄₅ (Figure 2A). We did not detect any SmpO_{49-F7K} binding to the *ossA*₄₉ and *ossB*₄₉ sequences, confirming the gain in affinity of SmpO_{49-F7K} for the OT45 *ossA* and *ossB* sites was

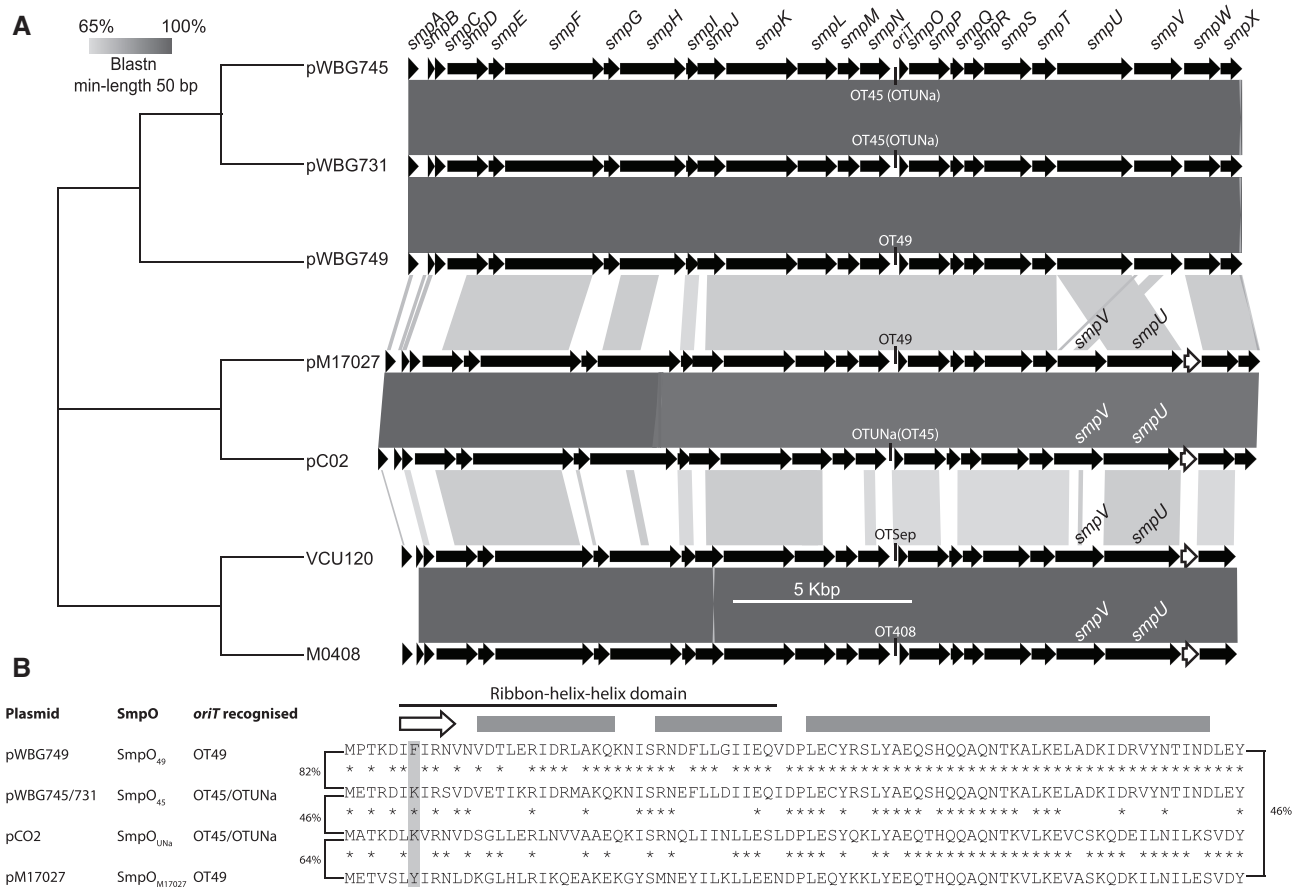


Figure 5. Incongruence between conjugation-gene clusters and their *oriT* specificity. (A) The pWBG749-family conjugation-gene clusters (*smpA-smpX*) of pWBG749 (GQ900391), pWBG745 (GQ900389), pWBG731 (MH587574), pM17027 (JGJU01000023.1), pC02 (CP012121), *S. epidermidis* VCU120 (AHL01000011) and *S. aureus* M0408 (AIW001000029) were aligned for tree construction using MEGA X (34). Black arrows indicate conserved conjugation genes *smpA-X* and the white arrows indicate a unique gene present in the pM17027, pC02, VCU120 and M0408 clusters, typified by the pC02 locus tag ACO02_2800. A Maximum Likelihood tree based on the Tamura-Nei model (35) was constructed with 1000 replicates and 100% of trees supported all nodes (36). The conjugation-gene cluster alignment was constructed using Easyfig and BLASTN (37). (B) Sequence alignments of the SmpO proteins from pWBG749 versus pWBG745/pWBG731, pWBG745/pWBG731 versus pC02 and pC02 versus pM17027. Asterisks indicate identical residues conserved between each two aligned sequences. PSI-PRED secondary-structure prediction is shown above (the arrow indicates the β -sheet region and bars represent helices) and residue 7 of the SmpO proteins is shaded. Brackets either side of the sequences indicate % amino acid identity between sequences.

concomitant with a loss in affinity for the OT49 *ossA* and *ossB* sites. Therefore, these experiments confirmed (at least for SmpO₄₉) that the changes in mobilization specificity of the F7K substitution were also reflected by DNA-binding specificity changes *in vitro*.

To investigate if the F7K change affected the ability of pWBG749 to recognize its own *oriT* in mobilization experiments we introduced the F7K change into *smpO*₄₉ on pWBG749e, producing pWBG749e-F7K. Surprisingly, conjugation was only decreased 17-fold by this mutation (Table 2). Furthermore, when pWBG749e-F7K was tested in mobilization experiments with each of the cloned *oriT* mimics from pWBG762 (8), it mobilized all three plasmids (Table 2). The mobilization rate was highest for OT45, then OTUNa, and lowest for OT49, confirming that the *oriT* preference had indeed been reversed by the F7K change but that *in vivo* recognition of the OT49 *oriT* was not completely abolished.

DISCUSSION

In this work, we located SmpO-binding sites within the *oriT* subtypes present on five pWBG749-family plasmids. Each *oriT* contained two SmpO-binding sites named here *ossA* and *ossB*. The *ossB* sites were located within the right arm of the IR2 repeat sequence in each *oriT*, consistent with our previous proposal that the differences in mobilization specificities between pWBG749-family members were related to differences in IR2 repeat sequences (7). Alignments of pWBG749-family conjugation genes suggested some family members had switched their specificity between OT45 and OT49 *oriT* types during their evolutionary history. We introduced N-terminal substitutions in the SmpO₄₉ and SmpO₄₅ genes in an attempt to change their respective mobilization specificities. A single amino-acid change in either protein switched their specificities between OT49 and OT45 *oriT*s. A substituted SmpO_{49-F7K} variant bound the OT45 *oriT* in SPR and EMSA assays as efficiently as the SmpO₄₅ pro-

Table 1. Mobilization of OT49-*oriT*-carrying plasmids by pWBG731 and pC02

Construct name	<i>oriT</i>	SmpO	pWBG731 conjugation ^a	pWBG731 mobilization	pC02 conjugation	pC02 mobilization
pLI50	None	None	2.8×10^{-4} ($\pm 4.2 \times 10^{-4}$)	$<1.0 \times 10^{-9}$	1.5×10^{-4} ($\pm 5.5 \times 10^{-5}$)	$<1.0 \times 10^{-9}$
pKY5T	OT45	None	1.2×10^{-4} ($\pm 1.7 \times 10^{-4}$)	1.4×10^{-4} ($\pm 1.5 \times 10^{-4}$)	5.9×10^{-4} ($\pm 4.8 \times 10^{-4}$)	1.6×10^{-5} ($\pm 2.5 \times 10^{-5}$)
pKY5TO	OT45	SmpO ₄₅	4.6×10^{-5} ($\pm 1.0 \times 10^{-5}$)	2.1×10^{-5} ($\pm 1.4 \times 10^{-5}$)	3.1×10^{-4} ($\pm 2.6 \times 10^{-4}$)	5.5×10^{-6} ($\pm 5.3 \times 10^{-6}$)
pLI749a	OT49	None	8.4×10^{-5} ($\pm 7.1 \times 10^{-5}$)	$<1.0 \times 10^{-8c}$	2.5×10^{-4} ($\pm 3.2 \times 10^{-4}$)	$<1.0 \times 10^{-9}$
pLIOT9S5-N-WT45	OT49	SmpO ₄₅	7.5×10^{-5} ($\pm 7.8 \times 10^{-5}$)	$<1.0 \times 10^{-8c}$	1.6×10^{-3} ($\pm 2.3 \times 10^{-3}$)	$<1.0 \times 10^{-9}$
pLIT9O99	OT49	SmpO ₄₉	2.0×10^{-4} ($\pm 7.3 \times 10^{-5}$)	1.9×10^{-4} ($\pm 1.2 \times 10^{-4}$)	2.0×10^{-4} ($\pm 2.4 \times 10^{-4}$)	2.2×10^{-6} ($\pm 3.0 \times 10^{-6}$)
pLIOT9S5-N-WT49	OT49	^b SmpO ₄₉ (1–17)::SmpO ₄₅ (18–84)	1.1×10^{-4} ($\pm 1.0 \times 10^{-4}$)	5.7×10^{-5} ($\pm 2.8 \times 10^{-5}$)	3.1×10^{-4} ($\pm 2.3 \times 10^{-4}$)	1.6×10^{-6} ($\pm 1.2 \times 10^{-6}$)
pLIOT9S5M-K7F	OT49	SmpO _{45-K7F}	6.1×10^{-4} ($\pm 6.1 \times 10^{-4}$)	4.2×10^{-5} ($\pm 3.0 \times 10^{-5}$)	3.4×10^{-4} ($\pm 3.1 \times 10^{-4}$)	5.8×10^{-7} ($\pm 4.3 \times 10^{-7}$)
pLIOT9S17027	OT49	SmpO _{M17027}	3.2×10^{-3} ($\pm 2.2 \times 10^{-4}$)	2.4×10^{-5} ($\pm 8.3 \times 10^{-6}$)	6.9×10^{-3} ($\pm 2.6 \times 10^{-3}$)	4.3×10^{-6} ($\pm 1.3 \times 10^{-6}$)

^aPer-donor transfer frequencies are the average of three independent experiments (\pm standard deviation).

^bNumbers in brackets next to SmpO correspond to regions derived from SmpO₄₉ or SmpO₄₅ as indicated.

^cSporadic presence of 1 to 3 colonies from 100 μ l of undiluted culture.

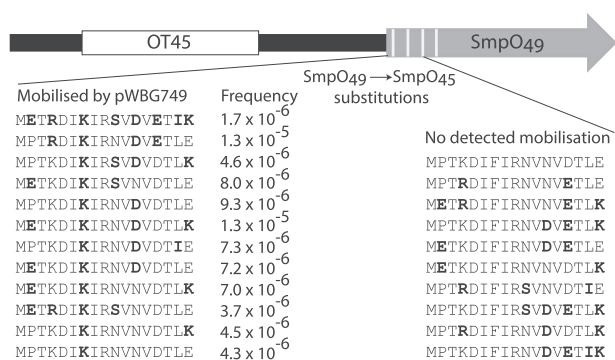


Figure 6. SmpO₄₉ mutagenesis and pWBG749e-mediated mobilization via the OT45 *oriT*. The OT45 *oriT* sequence from pWBG745 was cloned upstream of the SmpO₄₉ coding sequence. The first 17 amino acids of the SmpO₄₉ coding sequence were semi-randomly mutated to match those present in SmpO₄₅ as indicated in bold. Transfer frequency rates are given as exconjugants per donor and represent the average of three experiments. The complete dataset and construct names are presented in Supplementary Table S6.

tein. When this F7K substitution was introduced into the *smpO*₄₉ gene on pWBG749e, mobilization preference was switched to the OT45 *oriT*, but the conjugation frequency of pWBG749e was only modestly reduced. Therefore, it seems the F7K substitution may be a viable evolutionary intermediate during the evolutionary switch in DNA-binding specificity between the OT49 and OT45 *oriTs*.

RHH DNA-binding proteins decode DNA sequences using an anti-parallel β -sheet encoded by the N-termini of each protein chain in the RHH domain (38). Several RHH proteins form tetramers and bind cooperatively to two or more binding sites (23,38–42). The TraM protein for instance forms tetramers that bind cooperatively to four 5'-GANTC-3' motifs (40). In this work, EMSAs with the OT45 *oriT* revealed several complexes formed with in-

creasing concentrations of SmpO₄₅. SmpO proteins formed tetramers and hexamers in solution, so it is possible multiple SmpO proteins together bind *ossA* and *ossB* sites and form higher-order complexes. *In vivo* SmpO proteins may cooperatively bind *ossA* and *ossB*, perhaps together with the relaxase SmpP. This model could explain the relatively weak binding of *ossB* sites observed in some of the SPR assays, as *ossA* and *ossB* binding events were assayed in isolation from each other. Single nucleotide mutagenesis across three distinct *ossA* sites revealed SmpO proteins bound a ~6–10 bp region centred around a conserved 5'-GNNNNC-3' motif. The importance of the central bases in this motif, which vary between *oriT* subtypes, implicates them as determinants of SmpO binding specificity. The distance between the centres of the *ossA* and *ossB* 5'-GNNNNC-3' motifs ranged between 65 and 67 bp, suggesting spacing was also important, perhaps for higher order SmpO-*oriT* complex formation. The OTUNA and OTSep *ossA* sequences (and the OT45 *ossA* site on pWBG762) are present on the opposite DNA strand relative to the *ossB* site. Plasmids carrying SmpO_{UNa} or SmpO₄₅ can each mobilize plasmids carrying either OTUNA or OT45 *oriT*, so the orientation of the *ossA* site relative to the *ossB* is clearly not critical. This is consistent with the dyad symmetry of the 5'-GNNNNC-3' motif and the two-fold rotational symmetry of RHH domains.

In our earlier comparisons of the pWBG749-family *oriT* sequences we separated *oriT* sequences into subtypes based on global sequence alignments of the *oriT* sequences, which separated OT45 and OTUNA sites into distinct but closely related clusters. It is clear from subsequent work (4,17) and experiments here that the OT45 and OTUNA *oriT* are functionally interchangeable with respect to mobilization, despite being present on relatively distantly related members of the pWBG749 family. SPR assays suggest that despite this overlap in mobilization specificity, each of the SmpOs encoded by pWBG745 and pC02 still have a higher affinity for their cognate *oriTs*, as only SmpO_{UNa} produced weak

Table 2. Mobilization of *oriT*-mimic-carrying plasmids by pWBG749e-F7K

Construct name	Conjugative plasmid	Clone <i>oriT</i> mimic subtype	Transfer frequency*	
			Conjugative plasmid	pLI50 clone
pLI50	pWBG749e	none	4.4×10^{-3} (4.4×10^{-3})	$<1.0 \times 10^{-9}$
pLI762-49	pWBG749e	OT49	4.3×10^{-3} (9.8×10^{-4})	2.1×10^{-5} (9.6×10^{-6})
pLI762-45	pWBG749e	OT45	6.2×10^{-3} (1.3×10^{-3})	$<1.0 \times 10^{-9}$
pLI762-Una	pWBG749e	OTUNa	5.7×10^{-3} (4.2×10^{-3})	$<1.0 \times 10^{-9}$
pLI50	pWBG749e-F7K	none	2.6×10^{-4} (2.9×10^{-4})	$<1.0 \times 10^{-9}$
pLI762-49	pWBG749e-F7K	OT49	1.2×10^{-4} (5.9×10^{-5})	2.6×10^{-6} (2.6×10^{-6})
pLI762-45	pWBG749e-F7K	OT45	1.4×10^{-4} (9.1×10^{-5})	1.3×10^{-4} (1.3×10^{-4})
pLI762-Una	pWBG749e-F7K	OTUNa	1.5×10^{-4} (1.5×10^{-4})	3.5×10^{-5} (4.4×10^{-5})

*Per-donor transfer frequencies are the average of three independent experiments (\pm standard deviation).

SPR responses with the non-cognate OT45 *ossA* and *ossB* sites.

There were several discrepancies between our *in vitro* EMSA and SPR DNA-binding data and *in vivo* mobilization data. For instance, the SmpO_{45-K7F} variant was unable to bind OT49 *oriT* in SPR experiments (unpublished data) but the *smpO*_{45-K7F} allele enabled mobilization via the OT49 *oriT*. The *smpO*₄₅ allele enabled mobilizations of plasmids carrying the OTUNa *oriT* yet we did not observe binding to the OTUNa *oriT* using SPR. Furthermore, the SmpO_{49-F7K} protein was able to bind the OT45 *oriT* but not the *ossA*₄₉ or *ossB*₄₉ in SPR experiments (Supplementary Table S4), yet the pWBG749e-F7K was able to conjugatively transfer via the OT49 *oriT* and mobilize a plasmid carrying the OT49 *oriT*, albeit at reduced rates. Some of the negative results for SPR experiments may result from incomplete optimization of conditions, relative positioning of the binding site on the oligonucleotide or perhaps the absence of required secondary DNA structures normally present *in vivo*. An alternative explanation is that the levels of *smpO* gene expression *in vivo* may be high enough to compensate for low-affinity binding by some of the SmpO proteins. Several RAFs including TraM negatively regulate their own expression by binding their own promoter. A negative autoregulation model for *smpO* gene expression would fit with some of the discrepancies observed. For instance, if SmpO₄₉ represses transcription of *smpO*₄₉ by binding the pWBG749 *oriT* region, then the weaker binding of SmpO_{49-F7K} to the *oriT* would result in SmpO_{49-F7K} overexpression, tentatively explaining how pWBG749e-F7K retains the ability to transfer from and mobilize plasmids carrying an OT49 *oriT*.

Conjugative plasmids with mobilization specificity for OT49 and OT45/OTUNa *oriT* both appear in two distinct branches of the pWBG749-family tree (Figure 5A), suggesting convergent evolution has occurred in at least one branch. It is possible that homologous recombination or relaxase-mediated recombination between two conjugative plasmids, or a conjugative plasmid and a mobilizable plasmid with an *oriT* mimic, may have led to the replacement of *oriT* sequences on conjugative-plasmids. Indeed, pWBG731 carries remnants of an ancestral recombination with a pC02-like plasmid (although outside the *oriT*-*smpO* region), illustrating that recombination between divergent pWBG749-family plasmids occurs. Moreover, the conjuga-

tive plasmid pC02 carries an additional OT49 *oriT* mimic outside its conjugation-gene cluster, so intramolecular recombination between the OT49 and OTUNa *oriTs* on pC02 could conceivably produce a plasmid resembling pM17027 (4,17). There is no compelling evidence however that the *smpO* sequences have been replaced through recombination on conjugative plasmids. While the *smpO* gene sequences are less conserved than surrounding conjugation genes on closely related plasmids with distinct *oriT* specificities, they are even less similar to *smpO* genes present on distantly related plasmids with the same *oriT* specificity (Figure 5B). Overall, these observations seem to best support a model in which *oriT* sequences have been replaced on conjugative plasmids through recombination while SmpO specificity has evolved through gain-of-function mutations.

We suspect the evolutionary partitioning of conjugation genes and beneficial gene cargo onto separate plasmids in *S. aureus* reflects a molecular symbiosis between conjugative plasmids and plasmids they mobilize. Staphylococcal plasmids are smaller than those found in many other bacteria. Roughly half of documented plasmids are <10 kb and the other half are mostly <40 kb (6). This strongly suggests some selective pressure restricts plasmid size in *S. aureus*. Conjugation-gene clusters, together with plasmid housekeeping loci such as replication and partitioning genes, typically span 27 kb alone, so it is not surprising that most identified pWBG749-family plasmids lack additional genetic cargo such as antimicrobial-resistance genes (6,7,10,43). Antimicrobial-resistance genes instead mostly appear as single resistance genes on small rolling-circle-replication plasmids or clustered together on 20–40 kb non-conjugative plasmids (1,3). For these larger plasmids (>10 kb), 85% carry a pWBG749-family *oriT* mimic, of which OT49 and OT45 are the dominant subtypes (44% and 47%, respectively) (7). Antimicrobial-resistance plasmids, without needing to dedicate large regions of DNA coding potential for conjugation functions, have space to collect suites of virulence and resistance genes while maintaining an ability to be mobilized horizontally by a variety of conjugative plasmids. Importantly, conjugative plasmids likely also benefit from this arrangement by sharing in the phenotypic advantages the mobilized plasmids endow when they transfer together to a new host. We suspect the apparent switching of *oriT* sequences on pWBG749-family plasmids may reflect instances where selection pressure exists for conjugative mo-

bilization but the antimicrobial-resistance plasmid carries an incompatible *oriT* subtype. We demonstrated here that a single-codon change enables a conjugative plasmid to switch its *oriT* specificity to another *oriT* subtype while retaining an ability to self-transfer, thus providing a simple evolutionary pathway to enable mobilization of a plasmid carrying an otherwise incompatible *oriT* sequence.

The evolutionary pathways leading to changes in protein–DNA (and protein–protein) binding specificity are hotly debated (12,44–50). Evolution of a new DNA-binding specificity can be theoretically problematic depending on the evolutionary steps envisaged, which may in some scenarios involve the evolution of non-functional or redundant intermediary alleles. An alternative model that avoids non-functional intermediates is the evolution of a ‘promiscuous intermediate’ with broadened specificity for both the old DNA site and the new. pWBG749e-F7K mobilized plasmids carrying OT45/OTUNa and OT49 sites and while its mobilization preference favoured OT45, it remained capable of self-transfer from its own OT49 *oriT* albeit with reduced efficiency. While pWBG731 and pC02 plasmids also carry *smfO* alleles with a lysine at position 7, they do not mobilize plasmids carrying OT49 *oriTs*, suggesting SmpO₄₅ and SmpO_{UNa} carry additional amino acid residues that prevent OT49 *oriT* binding. This raises the question if such promiscuous mobilizing plasmids persist in nature. Clearly, we have not tested all members of the pWBG749-family for their capacity to mobilize distinct *oriT* subtypes, so promiscuous variants may indeed exist.

pWBG749-family plasmids have clearly played an important role in the movement of antimicrobial-resistance plasmids in staphylococcal species, as evidenced by the widespread prevalence of *oriT* mimics carried by them. The pWBG749 family have diverged in *oriT* specificity during evolution but non-conjugative plasmids have kept pace by acquiring mimics corresponding to each subtype. In this work, we present evidence that pWBG749-family plasmids have converged on the same *oriT* specificities in distinct branches of the pWBG749-family tree. We present a feasible evolutionary pathway enabling such *oriT* specificity changes, in which a single amino-acid change can enable recognition of a previously incompatible site. Broadly, these observations suggest dynamic interplay between conjugative and mobilizable plasmids and further highlights the important role of horizontal transfer via conjugative mobilization.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors wish to thank the 2015 undergraduate student cohort from the MICB3001 Advanced Molecular Microbiology unit (Curtin University) for assistance in the design and construction of SmpO amino acid substitution plasmids. Bacterial expression vector pETM11 was a gift from Dr Gunter Stier (EMBL, Heidelberg, Germany; <https://www.embl.de>). The contents of this publication are the sole responsibility of the author(s) and do not necessarily

reflect the views, opinions, or policies of Uniformed Services University, the Henry M. Jackson Foundation for the Advancement of Military Medicine, nor the U.S. Government. DSM is an employee of the U.S. Government. Title 17 U.S.C. 105 provides that ‘Copyright protection under this title is not available for any work of the United States Government’.

FUNDING

National Health and Medical Research Council (Australia) [APP1145697 to N.F., J.P.R.]; Australian Research Council [FT170100235 to J.P.R.]; University of Western Australia (to K.Y.E.); U.S. Department of Defense Program project grant [HT9404-12-1-0019]; Military Infectious Diseases Research Program award [HU0001-15-2-0031 to D.S.M.]. Funding for open access charge: National Health and Medical Research Council (Australia) [APP1145697].

Conflict of interest statement. None declared.

REFERENCES

- Shearer, J.E., Wireman, J., Hostetler, J., Forberger, H., Borman, J., Gill, J., Sanchez, S., Mankin, A., Lamarre, J., Lindsay, J.A. *et al.* (2011) Major families of multiresistant plasmids from geographically and epidemiologically diverse staphylococci. *G3 (Bethesda)*, **1**, 581–591.
- Richardson, E.J., Bacigalupe, R., Harrison, E.M., Weinert, L.A., Lycett, S., Vrieling, M., Robb, K., Hoskisson, P.A., Holden, M.T.G., Feil, E.J. *et al.* (2018) Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nat. Ecol. Evol.*, **2**, 1468–1478.
- Firth, N., Jensen, S.O., Kwong, S.M., Skurray, R.A. and Ramsay, J.P. (2018) Staphylococcal plasmids, transposable and integrative elements. *Microbiol Spectr.*, **6**, doi:10.1128/microbiolspec.GPP3-0030-2018.
- LaBreck, P.T., Li, Z., Gibbons, K.P. and Merrell, D.S. (2019) Conjugative and replicative biology of the *Staphylococcus aureus* antimicrobial resistance plasmid, pC02. *Plasmid*, **102**, 71–82.
- LaBreck, P.T., Rice, G.K., Paskey, A.C., Ellassal, E.M., Cer, R.Z., Law, N.N., Schlett, C.D., Bennett, J.W., Millar, E.V., Ellis, M.W. *et al.* (2018) Conjugative transfer of a novel staphylococcal plasmid encoding the biocide resistance gene, *qacA*. *Front. Microbiol.*, **9**, 2664.
- Ramsay, J.P. and Firth, N. (2017) Diverse mobilization strategies facilitate transfer of non-conjugative mobile genetic elements. *Curr. Opin. Microbiol.*, **38**, 1–9.
- O’Brien, F.G., Yui Eto, K., Murphy, R.J., Fairhurst, H.M., Coombs, G.W., Grubb, W.B. and Ramsay, J.P. (2015) Origin-of-transfer sequences facilitate mobilisation of non-conjugative antimicrobial-resistance plasmids in *Staphylococcus aureus*. *Nucleic Acids Res.*, **43**, 7971–7983.
- O’Brien, F.G., Ramsay, J.P., Monecke, S., Coombs, G.W., Robinson, O.J., Htet, Z., Alshaiikh, F.A. and Grubb, W.B. (2015) *Staphylococcus aureus* plasmids without mobilization genes are mobilized by a novel conjugative plasmid from community isolates. *J. Antimicrob. Chemother.*, **70**, 649–652.
- Rossi, F., Diaz, L., Wollam, A., Panesso, D., Zhou, Y., Rincon, S., Narechania, A., Xing, G., Di Gioia, T.S., Doi, A. *et al.* (2014) Transferable vancomycin resistance in a community-associated MRSA lineage. *N. Engl. J. Med.*, **370**, 1524–1531.
- Ramsay, J.P., Kwong, S.M., Murphy, R.J., Yui Eto, K., Price, K.J., Nguyen, Q.T., O’Brien, F.G., Grubb, W.B., Coombs, G.W. and Firth, N. (2016) An updated view of plasmid conjugation and mobilization in *Staphylococcus*. *Mob Genet Elements*, **6**, e1208317.
- Becker, K., van Alen, S., Idelevich, E.A., Schleimer, N., Seggewiss, J., Mellmann, A., Kaspar, U. and Peters, G. (2018) Plasmid-Encoded Transferable *mecB*-Mediated Methicillin Resistance in *Staphylococcus aureus*. *Emerg. Infect. Dis.*, **24**, 242–248.
- Blackwell, G.A. and Hall, R.M. (2019) Mobilisation of a small *Acinetobacter* plasmid carrying an *oriT* transfer origin by conjugative RepAci6 plasmids. *Plasmid*, **103**, 36–44.

13. Moran, R.A. and Hall, R.M. (2017) Analysis of pCERC7, a small antibiotic resistance plasmid from a commensal ST131 *Escherichia coli*, defines a diverse group of plasmids that include various segments adjacent to a multimer resolution site and encode the same NikA relaxase accessory protein enabling mobilisation. *Plasmid*, **89**, 42–48.
14. Moran, R.A. and Hall, R.M. (2019) pBuzz: a cryptic rolling-circle plasmid from a commensal *Escherichia coli* has two inversely oriented *oriT*s and is mobilised by a B/O plasmid. *Plasmid*, **101**, 10–19.
15. Salgado-Camargo, A.D., Castro-Jaimes, S., Gutierrez-Rios, R.M., Lozano, L.F., Altamirano-Pacheco, L., Silva-Sanchez, J., Perez-Oseguera, A., Volkow, P., Castillo-Ramirez, S. and Cevallos, M.A. (2020) Structure and evolution of *Acinetobacter baumannii* plasmids. *Front. Microbiol.*, **11**, 1283.
16. Zrimec, J. (2020) Multiple plasmid origin-of-transfer regions might aid the spread of antimicrobial resistance to human pathogens. *Microbiologyopen*, **9**, e1129.
17. Yui Eto, K., Firth, N., Davis, A.M., Kwong, S.M., Krysiak, M., Lee, Y.T., O'Brien, F.G., Grubb, W.B., Coombs, G.W., Bond, C.S. *et al.* (2019) Evolution of a 72-kilobase cointegrand, conjugative multiresistance plasmid in community-associated methicillin-resistant *Staphylococcus aureus* isolates from the early 1990s. *Antimicrob. Agents Chemother.*, **63**, e01560-19.
18. Garcillan-Barcia, M.P., Cuartas-Lanza, R., Cuevas, A. and de la Cruz, F. (2019) Cis-acting relaxases guarantee independent mobilization of MOBQ 4 plasmids. *Front. Microbiol.*, **10**, 2557.
19. Bowie, J.U. and Sauer, R.T. (1990) TraY proteins of F and related episomes are members of the Arc and Mnt repressor family. *J. Mol. Biol.*, **211**, 5–6.
20. Chen, Y., Staddon, J.H. and Dunny, G.M. (2007) Specificity determinants of conjugative DNA processing in the *Enterococcus faecalis* plasmid pCF10 and the *Lactococcus lactis* plasmid pRS01. *Mol. Microbiol.*, **63**, 1549–1564.
21. Moncalian, G. and de la Cruz, F. (2004) DNA binding properties of protein TrwA, a possible structural variant of the Arc repressor superfamily. *Biochim. Biophys. Acta*, **1701**, 15–23.
22. Varsaki, A., Moncalian, G., Garcillan-Barcia Mdel, P., Drainas, C. and de la Cruz, F. (2009) Analysis of ColE1 MbeC unveils an extended ribbon-helix-helix family of nicking accessory proteins. *J. Bacteriol.*, **191**, 1446–1455.
23. Yoshida, H., Furuya, N., Lin, Y.J., Guntert, P., Komano, T. and Kainosho, M. (2008) Structural basis of the role of the NikA ribbon-helix-helix domain in initiating bacterial conjugation. *J. Mol. Biol.*, **384**, 690–701.
24. Zhang, S. and Meyer, R. (1997) The relaxosome protein MobC promotes conjugal plasmid mobilization by extending DNA strand separation to the nick site at the origin of transfer. *Mol. Microbiol.*, **25**, 509–516.
25. Ziegelin, G., Furste, J.P. and Lanka, E. (1989) TraJ protein of plasmid RP4 binds to a 19-base pair invert sequence repetition within the transfer origin. *J. Biol. Chem.*, **264**, 11989–11994.
26. Ramachandran, G., Miguel-Arribas, A., Abia, D., Singh, P.K., Crespo, I., Gago-Cordoba, C., Hao, J.A., Luque-Ortega, J.R., Alfonso, C., Wu, L.J. *et al.* (2017) Discovery of a new family of relaxases in Firmicutes bacteria. *PLoS Genet.*, **13**, e1006586.
27. Alvarez-Martinez, C.E. and Christie, P.J. (2009) Biological diversity of prokaryotic type IV secretion systems. *Microbiol. Mol. Biol. Rev.*, **73**, 775–808.
28. Lu, J., Wong, J.J., Edwards, R.A., Manchak, J., Frost, L.S. and Glover, J.N. (2008) Structural basis of specific TraD-TraM recognition during F plasmid-mediated bacterial conjugation. *Mol. Microbiol.*, **70**, 89–99.
29. Monk, I.R., Tree, J.J., Howden, B.P., Stinear, T.P. and Foster, T.J. (2015) Complete bypass of restriction systems for major *Staphylococcus aureus* lineages. *mBio*, **6**, e00308-15.
30. Grote, A., Hiller, K., Scheer, M., Munch, R., Nortemann, B., Hempel, D.C. and Jahn, D. (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.*, **33**, W526–W531.
31. Gasteiger, E., H.C., G.A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005) In: Walker, J.M. (ed). *The Proteomics Protocols Handbook*. Humana Press, Totowa, NJ, pp.571–607.
32. Stevenson, C.E., Assaad, A., Chandra, G., Le, T.B., Greive, S.J., Bibb, M.J. and Lawson, D.M. (2013) Investigation of DNA sequence recognition by a streptomycete MarR family transcriptional regulator through surface plasmon resonance and X-ray crystallography. *Nucleic Acids Res.*, **41**, 7009–7022.
33. Traore, D.A.K., Wisniewski, J.A., Flanigan, S.F., Conroy, P.J., Panjikar, S., Mok, Y.F., Lao, C., Griffin, M.D.W., Adams, V., Rood, J.I. *et al.* (2018) Crystal structure of TcpK in complex with *oriT* DNA of the antibiotic resistance plasmid pCW3. *Nat. Commun.*, **9**, 3732.
34. Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, **35**, 1547–1549.
35. Tamura, K. and Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.*, **10**, 512–526.
36. Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.
37. Sullivan, M.J., Petty, N.K. and Beatson, S.A. (2011) Easyfig: a genome comparison visualizer. *Bioinformatics*, **27**, 1009–1010.
38. Schreiter, E.R. and Drennan, C.L. (2007) Ribbon-helix-helix transcription factors: variations on a theme. *Nat. Rev. Microbiol.*, **5**, 710–720.
39. Zampini, M. and Hayes, F. (2012) Combinatorial targeting of ribbon-helix-helix artificial transcription factors to chimeric recognition sites. *Nucleic Acids Res.*, **40**, 6673–6682.
40. Wong, J.J., Lu, J., Edwards, R.A., Frost, L.S. and Glover, J.N. (2011) Structural basis of cooperative DNA recognition by the plasmid conjugation factor, TraM. *Nucleic Acids Res.*, **39**, 6775–6788.
41. Miguel-Arribas, A., Hao, J.A., Luque-Ortega, J.R., Ramachandran, G., Val-Calvo, J., Gago-Cordoba, C., Gonzalez-Alvarez, D., Abia, D., Alfonso, C., Wu, L.J. *et al.* (2017) The *Bacillus subtilis* conjugative plasmid pLS20 encodes two ribbon-helix-helix type auxiliary relaxosome proteins that are essential for conjugation. *Front. Microbiol.*, **8**, 2138.
42. Moncalian, G., Grandoso, G., Llosa, M. and de la Cruz, F. (1997) *oriT*-processing and regulatory roles of TrwA protein in plasmid R388 conjugation. *J. Mol. Biol.*, **270**, 188–200.
43. Udo, E.E. and Grubb, W.B. (1990) A new class of conjugative plasmid in *Staphylococcus aureus*. *J. Med. Microbiol.*, **31**, 207–212.
44. Nakagawa, S., Gisselbrecht, S.S., Rogers, J.M., Hartl, D.L. and Bulyk, M.L. (2013) DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proc. Natl. Acad. Sci. USA*, **110**, 12349–12354.
45. Brockington, S.F., Moyroud, E., Sayou, C., Monniaux, M., Nanao, M.H., Thevenon, E., Chahtane, H., Warthmann, N., Melkonian, M., Zhang, Y. *et al.* (2015) Evolution. Response to comment on “a promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity”. *Science*, **347**, 621.
46. Sayou, C., Monniaux, M., Nanao, M.H., Moyroud, E., Brockington, S.F., Thevenon, E., Chahtane, H., Warthmann, N., Melkonian, M., Zhang, Y. *et al.* (2014) A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science*, **343**, 645–648.
47. McKeown, A.N., Bridgham, J.T., Anderson, D.W., Murphy, M.N., Ortlund, E.A. and Thornton, J.W. (2014) Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell*, **159**, 58–68.
48. Jalal, A.S.B., Tran, N.T., Stevenson, C.E., Chan, E.W., Lo, R., Tan, X., Noy, A., Lawson, D.M. and Le, T.B.K. (2020) Diversification of DNA-binding specificity by permissive and specificity-switching mutations in the ParB/Noc protein family. *Cell Rep.*, **32**, 107928.
49. Aakre, C.D., Herrou, J., Phung, T.N., Perchuk, B.S., Crosson, S. and Laub, M.T. (2015) Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell*, **163**, 594–606.
50. Resch, M., Striegl, H., Henssler, E.M., Sevvana, M., Egerer-Sieber, C., Schiltz, E., Hillen, W. and Muller, Y.A. (2008) A protein functional leap: how a single mutation reverses the function of the transcription regulator TetR. *Nucleic Acids Res.*, **36**, 4390–4401.