An improved method for oriT-directed cloning and functionalization of large bacterial genomic regions

Brian H. Kvitko\textsuperscript{a,}\textsuperscript{1}, Ian A. McMillan\textsuperscript{a,}\textsuperscript{2}, Herbert P. Schweizer\textsuperscript{a}\textsuperscript{*}

Running title: Improved DNA cloning via oriT recombination

\textsuperscript{a}Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO, 80523, USA

\textsuperscript{*} Herbert.Schweizer@colostate.edu

\textsuperscript{1}Present Address: MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI, 48824, USA

\textsuperscript{2}Present Address: Department of Molecular Bioscience and Bioengineering, University of Hawaii at Manoa, Honolulu, HI, 96822, USA
We have made significant improvements to a broad-host-range system for the cloning and manipulation of large bacterial genomic regions based on site-specific recombination between directly repeated oriT sites during conjugation. Using two suicide capture vectors carrying flanking homology regions oriT sites are recombined on either side of the target region. Using a broad-host conjugation helper plasmid the region between the oriT sites is conjugated into an Escherichia coli recipient strain where it is circularized and maintained as a chimeric mini-F vector. The cloned target region is functionalized in multiple ways to accommodate downstream manipulation. The target region is flanked with Gateway attB sites for recombination into other vectors and by rare 18-bp I-SceI restriction sites for sub-cloning. The Tn7- functionalized target can also be inserted at naturally occurring chromosomal attTn7 site(s) or maintained as a broad-host plasmid for complementation or heterologous expression studies. We have used the oriTn7 capture technique to clone and complement Burkholderia pseudomallei genomic regions up to 140-kb in size and have created isogenic Burkholderia strains with various combinations of genomic islands. We believe this system will greatly aid the cloning and genetic analysis of genomic islands, biosynthetic gene clusters and large open reading frames.
Introduction

To aid the genetic characterization of strain-to-strain genomic variation and the heterologous production of natural products, the genome scale manipulations required by synthetic biology necessitate the specific cloning and manipulation of DNA regions greater than 10 kb in size (1-4). However, PCR-amplification and restriction fragment-based cloning techniques are inefficient, inaccurate and not always possible for regions >10 kb. Several techniques have been developed to bridge this technical gap. The oriT-directed capture system developed by Chain et al. has advantages for the specific cloning of large genomic regions compared with site-specific recombinase-catalyzed ‘pop-out’ and short-flank recombineering-based strategies (2, 5-10). The system is based on the principle that recombination will occur between two directly oriented RP4 origin of transfer (oriT) sites during conjugative transfer (11). The oriT-directed capture system utilizes two suicide vectors, one of which is mini-F replicon based, each with different selectable markers. Segments flanking the target region are cloned into the suicide vectors which are then homologously recombined into the donor strain genome, resulting in flanking the target region with directly oriented oriT sites. After transformation of this target donor strain with a conjugation helper plasmid, the conjugation machinery encoded by this plasmid recognizes one of the oriT sites integrated into the genome and begins transfer in a directional manner of downstream genomic DNA including the target region and the mini-F replicon into an Escherichia coli recipient. Transfer is terminated upon reaching the second oriT site, which is then recombined with the initiating oriT in the E. coli recipient, thus generating a circular mini-F-based plasmid carrying the target genome region. The system developed by Chain et al. 2000, was used to clone specific regions in excess of 200-kb and does not rely on the location of flanking restriction sites required by RecE-based recombineering strategies (2, 5). The oriT-directed capture system does not generate deletions in the target chromosome during capture because only one strand of the genome is mobilized. This prevents any fitness effects that might be caused by large deletions seen in site-specific recombinase-based ‘pop-
out' systems (10). The circularized capture vector is simultaneously generated and recovered as an E. coli transconjugant. This eliminates the need for organism-specific conditional replicons, additional plasmid rescue steps, and the potential difficulties associated with the recovery of large plasmids. The captured DNA is recovered directly from the donor genome which eliminates the introduction of point mutations. Finally, all steps involving the handling of large DNA are conducted in vivo which reduces issues of large fragment DNA instability and large vector transformation inefficiency. This system should be functional in any organism that can be established as a conjugation donor.

We have engineered significant improvements into the oriT-directed capture system for increased functionality and the downstream manipulation of captured target regions. Most notably we created new broad-host conjugation helpers and have incorporated mini-Tn7 functionalization for single-copy genomic integration to create the oriTn7 capture system. The transposon Tn7 is capable of integrating site-specifically into a naturally evolved attTn7 genomic site located downstream of highly conserved glmS genes (12-14). The final pFTarget vector incorporates portions from each capture vector so that the large target region is carried by a complete mini-Tn7 transposon. We have built the oriTn7 capture system for use in Burkholderia pseudomallei the causative agent of melioidosis and a category B Select Agent. There is extensive strain-to-strain variation within the B. pseudomallei pan-genome, much of this in genomic regions of difference greater than 10 kb in size and many of these regions are genomic islands (15-21). The characterization of natural gene loss, acquisition and strain-to-strain variability are likely to be important factors in understanding B. pseudomallei pathogenicity and phenotypic diversity. This characterization requires the creation of isogenic strains carrying specific deletions or insertions of the genomic regions of difference. We have successfully used the oriTn7 capture system to specifically clone genomic regions from 12 kb to 140 kb corresponding to a region absent from the human clinical B. pseudomallei isolate 708a and
have used Tn7 integration to complement the matching 140 kb deletion of this region. In addition, we have used oriTn7 capture to create a set of isogenic *Burkholderia* strains carrying various combinations of two *B. pseudomallei* gene clusters whose presence is geographically biased in strains from different endemic regions. Although in its current form, the oriTn7 capture system was built specifically for use in *B. pseudomallei* and related species it was designed with portability in mind and could be readily adapted for use in other bacteria.

**Materials and Methods**

**Media and culture conditions.** Strains used in this study are listed in Supplementary Table S1. Bacteria were cultured in liquid or on agar solidified Lennox LB (MO BIO Laboratories, Carlsbad, CA) at 37°C with aeration. *Burkholderia* strains carrying temperature-sensitive plasmids were maintained at 30°C and plasmid curing was achieved by incubation at 42°C in the absence of selection. All procedures involving *B. pseudomallei* were performed in Select Agent approved Biosafety Level 3 (BSL3) facilities in the Rocky Mountain Regional Biosafety Laboratory (CSU) using approved Select Agent compliant procedures and protocols. Media were supplemented with antibiotics at the following final concentrations. For *E. coli* cultures; ampicillin (Amp), 100 μg/mL; kanamycin (Km), 40 μg/mL; zeocin (Zeo), 25 μg/mL; trimethoprim (Tmp), 100 μg/mL; chloramphenicol (Cm), 25 μg/mL; streptomycin (Sm), 50 μg/mL; gentamicin (Gm) 10 μg/mL; rifampicin (Rf), 50 μg/mL, polymyxin B (PmB), 15 μg/mL; 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), 5-bromo-4-chloro-3-indolyl glucuronide (X-gluc) and diaminopimelic acid (DAP) were used at 40 μg/mL (X-gal and X-gluc) and 200-400 μg/mL, respectively. For *Pseudomonas syringae*; Rf 50 μg/mL. For *B. thailandensis*; Km 40, μg/mL; Zeo, 25 μg/mL; Gm, 10 μg/mL; trimethoprim (Tmp), 100 μg/mL. For *B. thailandensis* AmrAB-OprA efflux pump expressing strains antibiotic concentrations were adjusted to 500-1000 μg/ml Km and 1000-2000 μg/ml Zeo. For *B. pseudomallei*; Km, 40 μg/mL; Zeo, 25 μg/mL, Gm, 10
μg/mL; PmB, 15 μg/mL. For *B. pseudomallei* AmrAB-OprA efflux pump expressing strains, antibiotic concentrations were adjusted to 500-1000 μg/ml Km and 1000-2000 μg/ml Zeo.

Counter-selection of *B. pseudomallei* was done with Tmp, 100 μg/mL and Rf, 50 μg/mL.

Arabinose (ARA) or rhamnose (RHA) were used to induce gene expression from plasmids contained in *Burkholderia* spp. at final concentrations of 0.2%.

**DNA manipulation.** Plasmid DNA was purified using the GeneJET plasmid miniprep kit (Fermentas, Glen Burnie, MD). Cleanup of DNA enzymatic reactions and gel extractions were conducted using the GenElute gel extraction kit (Sigma-Aldrich, St. Louis, MO). Bacterial genomic DNA was isolated using the Puregene genomic DNA purification kit (Gentra Systems, Qiagen, Valencia, CA). DNA boiling preparations for PCR analysis were obtained by combining 1 μL of overnight culture with 30 μL dH2O or picking from a bacterial colony or patch with a sterile toothpick followed by suspension in 30 μL dH2O and incubation at 100°C for 10 min. Site-directed mutagenesis was conducted with the Quickchange site-directed mutagenesis kit according to the manufacturer’s recommendations (Stratagene, Agilent Technologies, Santa Clara, CA). DNA restriction endonucleases, T4 ligase, calf alkaline intestinal phosphatase and Taq polymerase were obtained from New England Biolabs (Ipswich, MA) and used according to the manufacturer’s recommendations. Hi-Lo and λ DNA-mono cut mix DNA ladders were obtained from Minnesota Molecular (Minneapolis, MN) and New England Biolabs, respectively. Oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA) and are listed in Supplementary Table S2. DNA sequencing was conducted using an ABI 3130xL Genetic Analyzer (Applied Biosystems, Carlsbad, CA) at the Colorado State University Proteomics and Metabolomics Facility. For multiplex PCR, oligonucleotide primers for each probe were designed to have annealing temperature differences of 1°C and amplicon size differences of 100-bp. Multiplex oligos were combined into a master mix and used at 4-12 μM final concentration.
concentration. Southern blot analysis was performed using the NEBlot Phototope and Phototope-Star chemiluminescent labeling and detection kits from New England Biolabs following the manufacturer recommendations and using standard capillary transfer and blotting procedures (22).

Plasmid transformation of E. coli was done either by using standard electroporation or chemical transformation procedures (22). Electro-transformation of B. pseudomallei was conducted as described previously (23). Briefly, overnight cultures were harvested by centrifugation, washed three times with 300 mM sterile sucrose, and concentrated 10-fold prior to electroporation with 0.2-1.0 µg plasmid DNA.

Bacterial conjugations were conducted as either bi-parental or multi-parental matings with RHO3 (or derivative) conjugation donors or pRK2013 (or derivative) conjugation helper plasmids. One to three mL overnight cultures were harvested by centrifugation, washed twice with fresh LB, and concentrated, typically 5-fold, to 10⁸-10¹⁰ CFU/mL. Equal parts of each parent strain were combined and applied to sterile cellulose acetate or nitrocellulose filters placed on LB plates (augmented with DAP for conjugations including RHO3 donor strains) with each parent strain spotted individually as controls. Spots were allowed to dry and plates were incubated overnight at 37°C unless otherwise noted. Cells were recovered from the filters by centrifugation in 1.5 mL microcentrifuge tubes with 1 mL of LB, washed with 1 mL LB, and spread onto LB plates with appropriate selection and counter-selection to recover the desired exconjugants.

**Plasmid and strain construction.** Plasmids used in this study are listed in Supplementary Table S3. Plasmid and strain construction details are provided in Supplementary Methods.
OriTn7 capture

Capture vector LR recombination. Left and right flanking region of homology to target regions were amplified by PCR and cloned into directional pENTR vectors. (See Supplementary Methods). LR recombination between right flanking fragment containing pENTR vectors and the pUCTCAPR-GW right capture vectors was facilitated by the compatible GmR and ZeoR markers carried by pUCTCAPR. The pENTR vectors carrying various right flanking fragments were LR recombined with pUCTCAPR vectors using either LR clonase or LR clonase II from Invitrogen Life Technologies according to the manufacturer’s recommendations. LR recombinants were recovered on LB+ Gm plates, screened for ZeoR, and confirmed by NotI and/or BsrGI restriction enzyme digests.

The design of the left capture vector pFTCAPL-GW required the incorporation of a KmR resistance marker to comply with Select Agent regulations for *B. pseudomallei*. Unfortunately this conflicts with the KmR marker carried by directional pENTR/D-TOPO vectors. Two strategies were used to overcome the KmR marker conflict between the pENTR vectors and pFTCAPL vector. In the first strategy, pENTR vectors were digested with NsiI and their KmR markers were replaced with a TmpR marker by ligation with the 0.9 kb PstI TmpR marker fragment from pFTP2 to create pENTRTmp vectors. This allowed for the recovery *E. coli* EPI300 clones on LB+Km plates with correct LR recombinants without recovering a KmR pENTR background. In the second strategy, pENTR inserts were PCR amplified with primers M13F+M13R which appends the cloned inserts with attL1 and attL2 sites. The linear PCR products were then gel-purified and used in LR reactions with pFTCAPL-GW before transformation into *E. coli* EPI3000 and recovery on LB+Km plates. This PCR-based strategy is the one that we currently favor for the LR recombination of both left and right capture vectors as it is both faster and less cumbersome. EPI300 pFTCAPL LR recombinant clones were induced
to high copy with 0.2 % arabinose prior to plasmid purification and confirmed by BsrGI and other
restriction enzyme digests.

Creation of double merodiploid donor strains. Right and left capture vectors were
transformed into the conjugation strain RHO3. Using RHO3 donors, capture vectors were
conjugally transferred and homologously recombined into the B. pseudomallei strains at the
right and left flanks by bi-parental mating on LB+DAP plates followed by recovery of merodiploid
exconjugants with zeocin, gentamicin or kanamycin selection as appropriate. Due to the
occasional difficulty in recovering ZeoR exconjugants with AmrAB-OprA efflux pump proficient
strains, the ZeoR conferring right capture vectors were always recombined into the B.
pseudomallei strains first to simplify screening. Isolated exconjugant right capture vector
merodiploid colonies were screened by PCR with primer pair 536+537 for the presence of an
oriT. Confirmed TCAPR B. pseudomallei merodiploids were then used as recipients in a second
bi-parental mating to homologously recombine left capture vectors. Recovered KmR exconjugant
TCAPR TCAPL double merodiploids were then pooled and saved as oriTn7 capture donors.

OriTn7 capture matings. One mL each of LB+Km or LB+Km+Zeo grown cultures of double
merodiploid B. pseudomallei donor strains, LB+Rf or LB+Rf+Tmp grown cultures of E. coli
EPI300R1, and LB+Km or LB+Km+Gm grown cultures of DH5α/pUCPRK2013 or
DH5α/pBBRK2013 helper strains were harvested by centrifugation and washed twice with 1 mL
fresh LB medium prior to suspension in 100-200 µL LB. 20 µL each of donor, recipient, and
helper strains were combined in a multi-parental mating on a filter placed on an LB plate along
with individual parental controls and incubated overnight at 37°C. E. coli exconjugants were
recovered by selection on LB+Rf+Tmp+Km+Zeo plates. Candidate pFTarget-Tn7 clones were
struck to isolation on LB+Zeo+X-gluc plates and blue colonies were screened for KmR, GmS and
often for PmB5.
**pFTarget-Tn7 vector confirmation.** pFTarget-Tn7 vectors were confirmed by multiplex PCR, Eckhardt Gel electrophoresis, and I-SceI restriction digest. *E. coli* pFTarget-Tn7 candidate clones were inoculated into LB+Km or LB+Zeo medium with and without arabinose and grown overnight. Plasmid DNA was isolated and used as a template for multiplex PCR to confirm pFTarget-Tn7 insert identity. The Multiplex primer set consisting of primers 2506 through 2519 and 2505 were used to screen for the presence of the AMR, MBA and 140 regions. Multiplex primers 2502 pFTarget-Tn7 plasmid DNA was analyzed by I-SceI restriction digest to release the captured Tn7-Target. The Griffitts Lab modified Eckhardt gel electrophoresis protocol (25) was used to resolve and visualize large pFTarget-Tn7 vectors. *Pseudomonas syringae* pv. *phaseolicola* 1448A (Pph), 131.9 kb and 51.7 kb plasmids were used for size comparison (26). Briefly, overnight 30°C LB cultures of Pph were sub-inoculated 2:7 in fresh LB and incubated for 4 h at 30°C with aeration to an OD$_{600}$ between 0.5 and 0.7. Overnight 37°C LB Km *E. coli* pFTarget-Tn7 cultures were sub-inoculated 1:10 in fresh LB+Km medium and incubated for 5 h at 37°C with aeration to an OD$_{600}$ between 0.5 and 0.7. Cell samples were chilled on ice, washed with 0.3% sarkosyl and lysed in-well. The Eckhardt gel (1X SBE [10 mM NaOH, 1 mM EDTA, 29 mM boric acid, pH 8.0], 0.9% agarose, 0.5% SDS) was run in 1X SBE buffer at 100 V for 2.5 h total in 30 min current on and 15 min current off rest periods to prevent melting/warping of the gel. The gel was stained with SYBR Safe DNA gel stain from Invitrogen Life Technologies according to the manufacturer’s recommendations.

**Target-Tn7 transposition.** To perform target-Tn7 transposition, pFTarget-Tn7 vectors were transferred into derivatives of the RHO3 conjugation strain which can be metabolically counter-selected by plating on LB media lacking DAP. For pFTarget-Tn7 vectors with inserts less than 15 kb, copy control induced plasmid DNA was electroporated into RHO5 and transformants were recovered on LB+Km+DAP. To avoid deletion of insert sequences that routinely occurred...
when re-transforming pFTarget-Tn7 vectors with the 140 kb insert into RHO3, these plasmids were transferred into RHO3cm via conjugation. EPI300R1 pFTarget-Tn7 was used as donor, DH5α/pBBRK2013 as the conjugation helper and RHO3cm/pBBRSac3 as the recipient. The pBBRSac3 vector acts as an exclusion plasmid via plasmid incompatibility to prevent RHO3cm exconjugants from receiving the pBBRK2013 conjugation helper plasmid. Exconjugants were recovered on LB + DAP + Km + Zeo + Cm + Amp plates and then struck to isolation on LB + DAP + Km + Zeo + Cm + 10% sucrose to cure pBBRSac3. Recovered clones were screened for KmR, AmpS and GmS, and DAP auxotrophy. Plasmid DNA was isolated from RHO3cm/pFTarget-Tn7 exconjugants and re-confirmed by multiplex PCR. RHO5 or RHO3cm/pFTarget-Tn7 donors and RHO3/pTNS3 transposition helpers were conjugated with B. pseudomallei recipients and recovered on LB + Zeo or LB + Gm plates as appropriate. Target-Tn7 transposition candidates were screened for KmS to eliminate single cross-over events carrying the pFTarget-Tn7 backbone. Candidate Δ(amrRAB-oprA) complemented strains were additionally screened for GmR. B. pseudomallei Target-Tn7 integrant candidates were checked for Tn7 insertion at attTn7-1, -2 or -3 using primer 479 in combination with primers 1509, 1510, or 1511, respectively (23). B. thailandensis candidates were checked for Tn7 insertion at glmS1 or glmS2 with primer 479 and primers 618, or 619, respectively (27). The appropriate multiplex PCR described above was used to confirm the Tn7-target insert. Siderophore production phenotypes of Tn7 transposition candidates and control strains were determined by quantitative chrome azurol S assay as described previously (28) from overnight cultures grown at 37°C in CAA medium (29).

Maintenance of pFTarget-Tn7 as an RK2-based replicating plasmid. B. thailandensis strain BT36 was electroporated with pARAtrfA and transformants were recovered on LB + Zeo plates at 30°C. BT36 pARAtrfA and BT36 plasmid minus control recipients were combined with RHO5/pFTarget2-Tn7-YLF donor in biparental matings on an LB plate augmented with DAP.
and ARA, incubated overnight at 30°C and KmR exconjugants were recovered. BT36 pARAtfA
pFTarget2-Tn7-YLF maintenance was done at 30°C with ARA augmentation. To cure
pARAtfA, BT36 pARAtfA pFTarget2-Tn7-YLF exconjugants were struck to isolation on LB in
the absence of selection at 42°C. Plasmid DNA was isolated from BT36/pARAtfA/pFTarget2-
Tn7-YLF overnight cultures by boiling for PCR analysis and plasmid kit isolation for re-
transformation into *E. coli* EPI300R1.
RESULTS

To improve upon the oriT-directed capture system described by Chain et al. we constructed new left and right capture vectors and broad-host conjugation helpers which facilitate the cloning, manipulation and complementation of large specific genomic regions. The capture vectors pFTCAPL-GW and pUCTCAPR-GW were built to comply with *B. pseudomallei* Select Agent regulations, which restrict antibiotic marker use to ble (Zeo<sup>R</sup>), accCI (Gm<sup>R</sup>), nptI and nptII (Km<sup>R</sup>), However, they were also designed to allow facile swapping of the Km<sup>R</sup> uidA and FRT Zeo<sup>R</sup> markers in the pFTCAPL-GW or pUCTCAPR-GW vectors by EcoRI and XbaI digestion, respectively. Rather than use the ColE1-based pRK2013 conjugation helper plasmid to establish non-*E. coli* strains as conjugation donors, we created two new broad-host-range conjugation helper plasmids, pUCPRK2013 and pBBRK2013, by replacing the ColEI replicon of pRK2013 with either the pRO1600 replicon-containing vector pUCP24 or *Bordetella* BBR replicon-containing vector pBRR1MCS-5. These vectors both used the Select Agent compliant accCI and nptII resistance markers.

The steps and features of the oriTn7 system are summarized in Fig. 1. In brief, segments of DNA flanking the target region to the left and right are cloned into pFTCAPL-GW and pUCTCAPR-GW via Gateway LR recombination. The capture vectors are then homologously recombined with their matching genomic regions creating a merodiploid strain in which the target region is flanked by oriT sites (Fig. 1A). Using either the pUCPRK2013 or pBBRK2013 conjugation helper plasmids the merodiploid strain is then used as a conjugation donor so that the region between the oriT sites will be conjugated into an *E. coli* recipient strain (Fig. 1B). This results in the creation of a chimeric vector in which the target region is carried as cargo in a mini-Tn7 transposon and maintained by a mini-F replicon (Fig. 1C). The target region may then be inserted in single copy into the attTn7 site via Tn7 site-specific transposition for complementation or heterologous expression (Fig. 1D).
OriTn7 capture system efficacy testing. To test the efficacy of the oriTn7 capture system, two 
B. pseudomallei genomic regions of difference that cover a range of target sizes were selected 
for cloning.

The first region is an ~140 kb region encompassing the genes encoding the aminoglycoside and 
macrolide antibiotic AmrAB-OprA efflux pump and the malleobactin siderophore synthesis and 
uptake cluster which is absent from the clinical isolate 708a. Absence of the ~140 kb region has 
minimal effect on virulence as strains lacking this region remain fully lethal in a murine 
melioidosis model (30). For analysis of the ~140 kb deletion region three targets corresponding 
to different functional portions of the region were captured from B. pseudomallei 1710b (Fig. 2).

Although DNA sequence analysis using next generation sequencing would have been an option 
for assessing the integrity of the large cloned DNA fragments, we opted for more traditional 
technologies such as PCR and phenotypic analyses to assess integrity and functionality of 
cloned genes and operons.

A set of seven multiplex PCR probes, one probe approximately every 20 kb along the 140 kb 
region length, was used to confirm the presence of the captured targets (Fig. 2A). Eckhardt gel 
electrophoresis was used to resolve the large circular pFTarget-Tn7 plasmids to confirm their 
sizes and that a single plasmid was present in the capture recipients (Fig. 2B). Multiplex PCR 
using plasmid DNA templates confirmed the presence of the appropriate targets (Fig. 2C-E).

The 140 target is 142.8 kb in size and captures the entire region absent from 708a. All seven 
multiplex probe sites are present in this target. The MBA region is 64.9 kb in size and captures 
the amrAB-oprA efflux pump operon and its associated regulatory gene amrR, as well as the 
malleobactin siderophore synthesis and uptake gene clusters. Multiplex probes 1 through 4 are 
located within this region. The AMR target is 12.1 kb in size and captures only the amrAB-oprA 
efflux pump operon and its amrR regulatory gene. Only multiplex probe site 1 is carried on this 
target. For each target region the right capture flank was varied while the left capture flank was 
maintained. As it is occasionally difficult to cleanly recover Zeo<sup>R</sup> exconjugants in AmrAB-oprA 
clones.
efflux pump expressing strains, the Zeo<sup>R</sup> conferring right capture vector was always recombined into the chromosome first so that the merodiploid could be confirmed by checking for the presence of the ori<sup>T</sup> site by PCR. The number of recovered exconjugants from capture matings generally decreased with ~10<sup>8</sup> recipient cells the size of the target from 10<sup>2</sup>-10<sup>3</sup> for the AMR target to 10<sup>1</sup> for the 140 target.

The capacity of the system for complementation was tested in Bp338, a <i>B. pseudomallei</i> 1710b strain with a laboratory generated deletion that mimics the ~140 kb natural deletion of <i>B. pseudomallei</i> 708a. We were able to successfully recover Tn<sup>7</sup> integrants of all three Target-Tn<sup>7</sup> regions into Bp338 with frequencies ranging from 10<sup>2</sup> to 10<sup>3</sup> exconjugants per ~10<sup>8</sup> recipient cells. Km<sup>S</sup> Tn<sup>7</sup> integrants were identifiable from the background of Km<sup>R</sup> single crossover exconjugants at approximately 1:10 for the MBA and 140 target region and 1:3 for the AMR target region. The integrants were checked for attTn<sup>7</sup> insertions by PCR and target presence was confirmed by multiplex PCR (Fig. 2C-E). Similar attempts were made to recover Tn<sup>7</sup> integrants in <i>B. pseudomallei</i> 708a but Tn<sup>7</sup> integrants were only successfully recovered with the AMR target region. For the MBA and 140 target regions only Km<sup>R</sup> capture flank region crossover exconjugants of Bp708a were recovered (Fig. 2A). Other than acknowledging that strain 1710b is easier to genetically manipulate than other <i>B. pseudomallei</i> strains we do not understand the reason(s) for our inability to recover site-specific mini-Tn<sup>7</sup> integrants carrying the MBA and 140 regions. Target-Tn<sup>7</sup> integrants were also phenotypically confirmed. Resistance to Gm was restored in Tn<sup>7</sup> integrants with any of the three target regions but siderophore production was only restored in Tn<sup>7</sup> integrants carrying either the MBA or 140 targets (Fig. 2F).

These experiments confirmed the malleobactin siderophore phenotypes previously associated with the 140 kb region deletion either naturally present in strain 708a or genetically engineered in the strain 1710b background (31). More importantly, the newly developed large fragment capture technology allowed us to confirm what we could not do experimentally achieve
previously, that is to successfully complement the naturally present or engineered large genetic
lesions causing the observed malleobactin siderophore deficiency and AmrAB-OprA efflux
pump susceptibility phenotypes (31).

The second test region of 11 kb is the *Yersinia*-like fimbrial (YLF) gene cluster which is one of a
pair of *B. pseudomallei* gene clusters whose presence is geographically biased (24). The YLF
and *B. thailandensis*-like flagellum and chemotaxis (BTFC) gene clusters are found at the same
chromosomal location in different strains and appear to be mutually exclusive in nature (24).
The YLF cluster is more common in Thai *B. pseudomallei* isolates while Australian *B.
pseudomallei* strains more commonly carry the BTFC cluster (24). *B. thailandensis* strains carry
the BTFC cluster and provided the cluster’s namesake. Phenotypes associated with the YLF
cluster are unknown, but the BTFC cluster carries the *motA2* gene which has been shown to be
essential for intracellular motility and intercellular spread by *B. thailandensis* in the absence of
the *BimA* actin-recruitment-based motility system (32). Both YLF and BTFC carrying strains
cause human disease and are virulent in animal models. Thus neither of these regions carry
genes that would be expected to enhance virulence of a strain from which either the YLF or
BTFC cluster is absent (24, 33).

As we had already created *B. thailandensis* and Australian *B. pseudomallei* strain derivatives
deleted in their native BTFC gene clusters we sought to capture the 11.4 kb YLF gene cluster
from the Thai *B. pseudomallei* isolate 1026b and use Tn7 integration to create sets of strains
carrying various combinations of the two gene clusters. The YLF target insert was sufficiently
small that the I-SceI digest fragments of pFTarget2-Tn7-YLF were resolved and confirmed by
standard gel electrophoresis (Fig. 3A). Target-Tn7 integration was conducted and confirmed by
multiplex PCR in the Australian *B. pseudomallei* isolate MSHR305 and its ΔBTFC derivative as
well as by PCR in *B. thailandensis* E264 and its ΔBTFC derivative (Fig. 3B & 3C). (32). Thus
we have generated sets of isogenic *B. pseudomallei* and *B. thailandensis* strains carrying neither gene cluster, their native BTFC cluster, their non-native YLF cluster or both gene clusters in combination. The ability to clone the YLF and BTFC gene clusters into defined *B. pseudomallei* genetic backgrounds will facilitate future characterization of the roles that the respective clusters play in this bacterium’s biology.

**Plasmid-based maintenance of pFTarget-Tn7 vectors.** As an alternative to Tn7 integration, we sought to determine if the pFTarget-Tn7 vectors could be maintained as broad-host replicating plasmids by utilizing the RK2 replication origin on the ccFOS vector backbone. Plasmid replication from the RK2 origin requires expression of the TrfA protein which can be provided *in trans*. To this end we created pARATrfA, a temperature sensitive *Burkholderia* PBAD- trfA expression vector as a RK2 replication helper. We chose this strategy to aid in our analyses as it would result in temperature sensitive replication of pFTarget-Tn7 (Fig. 4A). For stable pFTarget-Tn7 plasmid maintenance the use of a non-temperature sensitive constitutive TrfA expression replication helper would be preferable. *B. thailandensis* BT36 was transformed with pARATrfA and used as a conjugation recipient with a RHO5/pFTarget2-Tn7-YLF donor. Exconjugants were recovered at 30°C on LB+Km+ARA plates to maintain the plasmids. Candidates were purified at 42°C to cure both plasmids. 12/12 of single colonies tested were sensitive to both Zeo and Km. Control conjugations with BT36 lacking pARATrfA resulted in the recovery of a small number of presumed crossover exconjugants whose Zeo and Km resistance could not be cured at 42°C. PCR confirmed both the presence of the YLF region in BT36/pFTarget2-Tn7-YLF plasmid replicating exconjugants and the loss of YLF in temperature cured strains (Fig. 4B). Plasmid DNA purified from pFTarget2-Tn7-YLF replicating BT36 was transformed into *E. coli* EPI300R1. Comparison between I-SceI digests of plasmid DNA from the original pFTarget2-Tn7-YLF and recovered *E. coli* clones indicated that pFTarget-Tn7-YLF was recovered intact from BT36 (Fig. 4C).
Based on our initial success we sought to determine if pARAtrfA would allow large insert pFTarget-Tn7 vectors to be maintained as replicating plasmids in B. pseudomallei 708a or if pARAtrfA might enhance Tn7 integration of large target inserts into 708a by allowing even transient pFTarget-Tn7 replication. Unfortunately, while pARAtrfA enhanced the recovery of KmR 708a flank region crossover exconjugants, it did not allow for the temperature sensitive replication of large target MBA or 140 kb insert plasmids or the recovery of large insert Tn7 integrants.

Discussion

The oriTn7 capture system offers some key advantages for the specific cloning and downstream manipulation of genomic fragments over other large fragment cloning systems. The oriTn7 capture system can easily accommodate fragment sizes from 10 to over 100 kb and is mediated by reliable bacterial conjugation. All manipulation of large DNA is conducted in vivo which eliminates the issue of DNA fragmentation. Finally, the captured target region is functionalized with Gateway attB sites and I-SceI recognition sites for transfer into other vectors and is also functionalized as a mini-Tn7 for single copy chromosomal integration or maintenance as a multi-copy plasmid by using the RK2 origin of replication.

Details of specific methodology considerations for the steps of the oriTn7 capture system are discussed below. First, to create target region specific capture vectors, left and right capture flanking regions to the desired target region are PCR amplified and TOPO cloned into directional pENTR vectors. Left and right flanking region primers are designed to amplify flanking regions of sufficient size to facilitate single homologous crossover events. In addition, flanking regions corresponding to central fragments of genes may be undesirable as they will create corresponding gene interruptions after homologous recombination. The forward primers
for each flanking region are appended with the 5' CACC directional TOPO sequence such that
the correct flanking region orientations are maintained with respect to the targets genomic
orientation throughout the procedure. The left and right flanking regions carried in the pENTR
vectors are then LR recombined with the left (pFTCAPL-GW) and right (pUCTCAPR2 or 3-GW)
capture vectors (Fig. 1A).

The left and right capture vectors are sequentially recombined into the target genome via
homologous recombination resulting in a double merodiploid donor strain with two directly
repeated oriT sites flanking the target region (Fig. 1B). The mini-F and pUC plasmid replicons in
the left and right capture vectors are incapable of replication in Burkholderia spp. and many
other bacteria. This allows antibiotic selection to identify RecA-mediated homologous
recombination events. Exconjugant double merodiploids recombinants with both capture vectors
recombined into the chromosome are pooled for use in subsequent capture matings.

Using one of the newly constructed broad-host-range conjugation helpers, pUCPRK2013 or
pBBRK2013, conjugation is initiated from one integrated oriT site of the double merodiploid
donor strain and a single strand is transferred into the E. coli recipient strain. Conjugation
continues until the second integrated oriT is reached. This halts the conjugation process
resulting in the creation of a chimeric circularized, KmR ZeoR, and GmS pFTarget-Tn7 vector
carrying the desired target region (Fig. 1C). Although both broad-host-range conjugation
helpers have proved effective for capture matings, in general, the use of pBBRK2013 has
resulted in 2-fold greater numbers of recovered exconjugants from B. pseudomallei.

Conjugation initiated from the second oriT is non-productive as it would require conjugation of
the entire chromosome.

Recovered pFTarget-Tn7 candidate exconjugants are purified and screened for GmS. This
screen checks against donor breakthrough, recovery of both the evicted capture vectors in
combination or the recovery of the pFTarget-Tn7 vector in combination with a conjugation helper plasmid. This last option was the most common cause of GmR exconjugants. In general, GmS exconjugant clones appear to be recovered at a higher percentage as the size of the target insert increases. For example, for the 140 kb fragment ~75% of the exconjugants were GmS whereas for the 12.1 kb AMR fragment ~25% were GmS. However, if no GmS clones are recovered from a capture mating, small insert pFTarget-Tn7 vectors (15 kb or less) can be isolated and re-transformed by electroporation. Additionally, for large insert pFTarget-Tn7 vectors, conjugation from an EPI300R1/pFTarget-Tn7 pBBRK2013 donor into a RHO3cm/pBBRsac3 BBR-exclusion recipient can be used to recover clones lacking the pBBRK2013 helper plasmid.

In the final step of constructing Tn7 integrant expression or complementation strain, the pFTarget-Tn7 vector is mobilized by conjugation into the desired recipient strain in combination with a Tn7 transposase-encoding helper plasmid. The expression of the site-specific Tn7 transposition pathway proteins is necessary to integrate the Target-Tn7 at glmS-associated attTn7 sites (Fig. 1D) (27). Putative Tn7 integrants are screened for single crossover events by checking for KmS. For inserts under 15 kb Tn7 integrants are easily recovered, presumably because the rate of Tn7 transposition for inserts of this size is higher than the rate of RecA-mediated recombination with the genome. To successfully recover attTn7-located integrants for inserts of large size, RecA-mediated crossover between the target insert and the genome must be restricted. This can be accomplished by limiting the homology between the target insert and genome by using a strain carrying a matching deletion of the target insert, by using a strain with limited genome homology to the target, or by integrating into a recA-deficient strain.

Although the left and right capture vectors by necessity share homology at their oriT sites, efforts were made to keep inter-vector homology to a minimum to prevent RecA-mediated homologous recombination between the two capture vectors during the creation of double

20
merodiploid donor strains. Cloning target flanking regions of larger sizes will help bias against
inter-oriT homologous recombination and it would also typically be favorable to homologously
recombine the capture vector carrying the smaller of the two flanking regions first rather than
second. However, a capture mating from a donor with inter-oriT recombined capture vectors
would only result in the recreation of the original miss-recombined capture vector post-
conjugation into the E. coli recipient. As the original capture vectors do not individually carry the
required combination of antibiotic markers to be recovered, pooling the double merodiploid
exconjugants is a desirable strategy for this step of the system.

The final captured target regions carried in pFTarget-Tn7 vectors are usefully functionalized in
various ways. The pFTarget-Tn7 vectors carry both a mini-F-based replicon, which can easily
accommodate inserts of over 100 kb, and an RK2 oriV which allows for broad-host replication of
pFTarget-Tn7 vectors if TrfA is provided in trans or induction to high copy number in
CopyControl strains that express a mutant, copy-up variant of the TrfA protein (34). The single
oriT site on the pFTarget-Tn7 vectors are both intrinsic to the system and facilitates conjugation-
based mobilization of large pFTarget-Tn7 vectors between strains. The incorporation of
Gateway cassettes into the left and right capture vectors in combination with directional
pENTR/D-TOPO vectors simplifies work-flow and primer design, but also results in the captured
target region being flanked with Gateway attB sites. The attB sites can be used to recombine
the target region into attP carrying Gateway pDONR vectors assuming that the particular
pDONR vector is capable of supporting replication of the target insert. The Target-Tn7 is flanked
by 18-bp rare cutting I-SceI endonuclease recognition sites which aid in both restriction-based
screening of candidates and allow for I-SceI-based restriction sub-cloning of targets inserts.
Lastly, the target mini-Tn7 element carries a Flp recombinase excisable FRT-ZeoR-FRT
cassette. This allows the target insert to be site-specifically transposed in single copy into the
chromosome of most Gram-negative bacteria, followed by Flp-mediated excision of the ZeoR
selection marker. Some bacteria, for instance \textit{B. thailandensis} and \textit{B. pseudomallei}, contain more than one Tn7 insertion site but simultaneous insertion into multiple sites occurs less frequently than single site insertions, and single and multiple insertions can easily be differentiated by PCR.

In addition, if both capture flanking regions have homology to the borders of a matching chromosomal deletion, the pFTARGET-Tn7 vector may also be used as an allelic exchange vector for deletion restoration. The I-SceI recognition sites in the vector backbone can serve as counter-selection markers in combination with an I-SceI endonuclease expression plasmid (35). Colorimetric screening for the \textit{uidA} gene would aid the identification of double crossover events resolving the merodiploid state. Unfortunately, this strategy cannot be used in \textit{B. pseudomallei} strains expressing AmrAB-OprA as all three currently permissible antibiotic resistance markers - Gm\textsuperscript{R}, Km\textsuperscript{R} and Zeo\textsuperscript{R} - are already exhausted leaving none to select for introduction of an I-SceI expression plasmid. However, one could envision use of other permissible non-antibiotic resistance markers such as glyphosate resistance for maintenance of the I-SceI expression vector (36).

Currently, the \textit{oriTn7} capture system is limited to use in genetically-tractable Gram-negative bacteria that can be established, at least transiently, as conjugation donors. The F-replicon and the pUC replicon must also be non-replicative in the target organism. Even within these limitations, the genera \textit{Burkholderia}, \textit{Pseudomonas}, \textit{Ralstonia} and \textit{Sinorhizobium} should all be compatible and likely many other genera as well.

\textbf{Supplementary Material}

Supplementary Methods, Supplementary Tables S1-S3 and Supplementary References.

\textbf{Acknowledgements}
Funding was provided by indirect cost recovery returns made possible by several extramural National Institutes of Health, National Institute of Allergy and Infectious Diseases research grants and the Ruth L. Kirschstein National Research Service Award F32 [AI088884] to [BK]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We would like to thank Bryan Swingle for providing pBBR1-MCS5, Alan Collmer for providing pCPP5250, Carolina Lopez for providing pEXpheS*Gen and Lily Trunck for providing B. thailandensis BT36.

References


FIGURE LEGENDS

Figure 1. oriTn7 capture system overview. A. Separate mini-Tn7 capture suicide vectors based on pFTCAPL and pUCTCAPR are homologously recombined into the *B. pseudomallei* genome using cloned left and right regions that flank the target region. The GenBank accession numbers for pFTCAPL-GW, pUCTCAPR2-GW and pUCTCAPR3-GW are JX573201, JX573199 and JX573200, respectively. B. Transfer functions encoded by the broad-host conjugation helper mobilizes DNA downstream of the oriT into the recipient *E. coli* strain. C. Transfer is halted at the second oriT site and the plasmid is circularized in the recipient *E. coli* generating a chimeric pFTarget-Tn7 vector carrying the target region. D. Co-delivery by conjugation of the pFTarget-Tn7 and the pTNS helper plasmid allows for site-specific integration of the Tn7-functionalized target region into the attTn7 site(s) of the chosen expression strain.

Abbreviations: nptII, kanamycin resistance gene; aacC1, gentamicin resistance gene; ble, zeocin resistance gene; uidA, GUS colorimetric marker; glmS, glucosamine-6-phosphate synthase gene; oriT, RP4 origin of transfer; I-SceI, I-SceI recognition site; L and R, Tn7 left and right Tn7 ends; 1 and 2, Gateway attB1 and attB2 sites; attTn7, Tn7 insertion site; FRT, Flp recombinase target site; ccFOS, CopyControl RK2 inducible replication origin and mini-F replication origin; pUC, pUC replication origin; left and right; Left and Right target flanking regions of homology.

Figure 2. Use of oriTn7 capture to clone and complement regions associated with a large *B. pseudomallei* natural deletion. A. Map of the approximately 140 kb region absent from the *B. pseudomallei* clinical isolate 708a. The relative locations of the amrAB-oprA aminoglycoside efflux pump encoding genes and its regulatory gene amrR (amr) and the malleobactin siderophore synthesis and uptake gene cluster (mba) are shown. The extents of left and right capture flanking regions for various targets are marked in red. Multiplex probe regions 1-7 are
marked in white and numbered. The three captured target regions with their sizes in kb are shown below the region map. **B.** Eckhardt gel electrophoresis of large circular pFTarget-Tn7 vectors carrying 140, MBA, and AMR targets. Expected sizes for target + vector: 140, 155.1 kb; MBA, 77.2 kb; AMR, 24.4 kb. *Pseudomonas syringae pv. phaseolicola* 1448a (Pph) plasmids are used for size comparison. **C and D.** Multiplex PCR confirmation of 140 and MBA target capture plasmids from *B. pseudomallei* 1710b and Target-Tn7 integration into Bp338, a 1710b 140 kb deletion strain. Multiplex probe amplicons are numbered as in panel A. HL= HiLo DNA ladder, the 500 bp fragment is marked with an asterisk. PCR products seen in panel C, lane Δ140 are non-specific amplicons. **E.** Multiplex PCR confirmation of AMR target capture plasmid from *B. pseudomallei* 1710b and Target-Tn7 integration into Bp338 708a a clinical isolate carrying a natural 140 kb deletion. The multiplex probe amplicon is numbered as in panel A. The upper PCR product is a non-specific amplicon as in panel C, lane Δ140 seen only when using *B. pseudomallei* genomic DNA as the multiplex PCR template. **F.** Siderophore production and gentamicin resistance (10 µg/ml) phenotypes of duplicate complemented strains. Siderophore production complementation was determined by quantitative chrome azurol S assays of culture supernatants from cells grown in iron-limited media. Mean and standard deviation of triplicate measurements are shown

**Figure 3. Creation of Burkholderia strains carrying geographically diverse gene clusters**

**A.** I-Sce I digest of pFTarget-Tn7 vectors carrying the AMR and YLF targets. Expected sizes: AMR, 10.4 kb and 14 kb; YLF 10.4 and 13.2 kb. HL and λ refer to HiLo and lambda monocut mix DNA ladders, respectively. **B and C.** Multiplex PCR and PCR confirmation of YLF gene cluster capture and the creation of *B. pseudomallei* MSHR305 and *B. thailandensis* E264 strains carrying the YLF gene cluster. YLF probe size, 350 bp. BTFC probe size 115 bp. Note that the BTFC probe will not amplify from the native *B. thailandensis* BTFC cluster (24). The 500 bp fragment is marked with an asterisk.
Figure 4. Maintenance of pFTarget-Tn7 as a plasmid in B. thailandensis. A. Expression of TrfA from the *Burkholderia* temperature sensitive $P_{BAD}$ expression plasmid allows replication from the copy-control RK2 origin on the ccFOS backbone of pFTarget-Tn7. Temperature-based curing of one pARA$trfA$ results in curing of the pFTarget-Tn7 vector. B. PCR confirmation of YLF cluster presence in pFTarget-Tn7 plasmid-maintaining B. thailandensis BT36 grown at 30°C. C. I-SceI digests of pFTarget2-Tn7-YLF alongside transformant clones recovered from a BT36/pARA$trfA$ and BT/pFTarget2-Tn7-YLF purified plasmids. HL and λ refer to HiLo and lambda monocut mix DNA ladders, respectively.
A. Capture Vector Integration

B. Donor Merodiploid Strain

C. Recipient E. coli Strain

D. Complementation-Expression Strain

pTNS (tnsABCD) transposase helper