Supplementary Information

Supplementary Figure Legends

Figure S1. Integration of plasmid pKR04 into the *M. smegmatis* and *M. tuberculosis* genomes. (A) PCR analysis of the *attL* and *attR* attachment junctions (lanes 1-3, *attL*; 4-6, *attR*) showing integration into the *attB* site in Msmeg_5156. Each lane represents PCR of individual transformants. (B) PCR analysis of the *attL* and *attR* attachment junctions (lanes 1-8, *attL*; 9-16, *attR*) showing integration into the *attB* site in *M. tuberculosis* Rv1156. Lanes 2-8 and 10-16 represent PCR from individual transformants. Lanes 1 and 9 are controls using wild-type *M. tuberculosis* DNA.

Figure S2. Alignment of Peaches and Bxz2 Integrases. The location of four domains is indicated as follows: N-terminal catalytic domain, red box; Pfam07508 domain, blue box; zinc finger domain, green box; coiled-coil domain purple box. The C-terminal end of the catalytic domain indicated aligns with positions of proteolytic cleavage in Bxb1 Int. Vertical arrow shows position of the catalytic serine residue. Predicted secondary structure elements derived from PSIPRED and HHPred analysis and alignment to the LI Int structure for residues 1-310 are shown; the C-terminal ~200 residues are sufficiently divergent that the secondary structure features cannot be confidently predicted. Putative inter-domain linker regions are indicated.

Figure S3. **Mutations at the central dinucleotides of Bxz2 attachment sites**. *attP* and *attB* share a common core of 4 bp, 5'GCTG at their centers, and strand exchange is likely to occur about two of these, with T-1 and G+1 being the best candidates as they lie at the center of sequence symmetries within each site. (A) Recombination of linear *attB* substrates with 60 bp long *attP* mutants (as indicated) in which each position from G-4 to G+2 is substituted. Only the

mutations at the T-1 and G+1 positions inhibit recombination consistent with these representing the central dinucleotide about which strand exchange occurs. (B) Recombination of linear *attP* with *attB* mutant substrate (G+1C).

Figure S4. **Predicted *attP* sites of Mycobacteriophages Timshel and Benedict.** In both Timshel and Benedict *int* is transcribed leftwards and there is a short intergenic space between the 5' end of the *int* and the adjacent gene to its right that is predicted to contain the *attP* site. The sequences shown represent plausible *attP* sites identified by the presence of imperfect inverted repeats with symmetrically conserved bases shown with a line above them. The predicted central dinucleotide is indicated (⌐┘). The translation initiation codon of the *int* gene is shown in bold.

Table S1. Sequences of oligonucleotides used in this study.

***att* site**	**Sequence**
attB	5′ CGGCGAACGCGGTCTCCATCGGGATCTGCTGATCGAGCAGCATGCCGACCAGAAGCGCGA 3′
	5′ TCGCGCTTCTGGTCGGCATGCTGCTCGATCAGCAGATCCCGATGGAGACCGCGTTCGCCG 3′
B half site	5′ GCGAGCTTCTTGGGCCCGGCGAACGCGGTCTCCATCGGGATCTGCTGATC 3′
	5′ GATCAGCAGATCCCGATGGAGACCGCGTTCGCCGGGCCCAAGAAGCTCGC 3′
B′ half site	5′ TGCTGATCGAGCAGCATGCCGACCAGAAGCGCGAACGGGTTGGACTCCAA 3′
	5′ TTGGAGTCCAACCCGTTCGCGCTTCTGGTCGGCATGCTGCTCGATCAGCA 3′
Bxz2 *attP*	5′ GCCATAACCGCAAGTGTACATCCCTCGGCTGGCCGAGACAAGTACAGTTGCGACAGACTG 3′
	5′ CAGTCTGTCGCAACTGTACTTGTCTCGGCCAGCCGAGGGATGTACACTTGCGGTTATGGC 3′
Bxz2 P half-site	5′ GGCTCTTAGCGGCTGTGCCATAACCGCAAGTGTACATCCCTCGGCTGGCC 3′
	5′ GGCCAGCCGAGGGATGTACACTTGCGGTTATGGCACAGCCGCTAAGAGCC 3′
Bxz2 P′ half-site	5′ GGCTGGCCGAGACAAGTACAGTTGCGACAGACTGTCTTCGCAGCTCAGAC 3′
	5′ GTCTGAGCTGCGAAGACAGTCTGTCGCAACTGTACTTGTCTCGGCCAGCC 3′
Bxz2 *attL*	5′ CGGCGAACGCGGTCTCCATCGGGATCTGCTGGCCGAGACAAGTACAGTTGCGACAGACTG 3′
	5′ CAGTCTGTCGCAACTGTACTTGTCTCGGCCAGCAGATCCCGATGGAGACCGCGTTCGCCG 3′
Bxz2 *attR*	5′ GCCATAACCGCAAGTGTACATCCCTCGGCTGATCGAGCAGCATGCCGACCAGAAGCGCGA 3′
	5′ TCGCGCTTCTGGTCGGCATGCTGCTCGATCAGCCGAGGGATGTACACTTGCGGTTATGGC 3′
Peaches *attP*	5′ GGCATAGTTTCCAATGTTACAGGAACTGCTGGCAGAATCCAACACATTGGAAGTCGATGA 3′
	5′ TCATCGACTTCCAATGTGTTGGATTCTGCCAGCAGTTCCTGTAACATTGGAAACTATGCC 3′
Peaches *attL*	5′ CGGCGAACGCGGTCTCCATCGGGATCTGCTGGCAGAATCCAACACATTGGAAGTCGATGA 3′
	5′ TCATCGACTTCCAATGTGTTGGATTCTGCCAGCAGATCCCGATGGAGACCGCGTTCGCCG 3′
Peaches *attR*	5′ GGCATAGTTTCCAATGTTACAGGAACTGCTGATCGAGCAGCATGCCGACCAGAAGCGCGA 3′
	5′ TCGCGCTTCTGGTCGGCATGCTGCTCGATCAGCAGTTCCTGTAACATTGGAAACTATGCC 3′
Peaches P half-site	5′ GGCTCTTAGCGGTTGTGGCATAGTTTCCAATGTTACAGGAACTGCTGGCA 3′
	5′ TGCCAGCAGTTCCTGTAACATTGGAAACTATGCCACAACCGCTAAGAGCC 3′
Peaches P′ half-site	5′ CTGCTGGCAGAATCCAACACATTGGAAGTCGATGAGAAAGGACCCCTCAC 3′
	5′ GTGAGGGGTCCTTTCTCATCGACTTCCAATGTGTTGGATTCTGCCAGCAG 3′
Hybrid-1	5′ GCCATAACCGCAAGTGTACATCCCTCGGCTGGCAGAATCCAACACATTGGAAGTCGATGA 3′
	5′ TCATCGACTTCCAATGTGTTGGATTCTGCCAGCCGAGGGATGTACACTTGCGGTTATGGC 3′
Hybrid-2	5′ GGCATAGTTTCCAATGTTACAGGAACTGCTGGCCGAGACAAGTACAGTTGCGACAGACTG 3′
	5′ CAGTCTGTCGCAACTGTACTTGTCTCGGCCAGCAGTTCCTGTAACATTGGAAACTATGCC 3′
pKR04 For primer	5′ AAAAAACATATGGGGTATGGCACCGGATTTGACAT
pKRO4 rev primer	5′ TTTTTTCATATGTGACTGGGGATAGCATCTGCAC

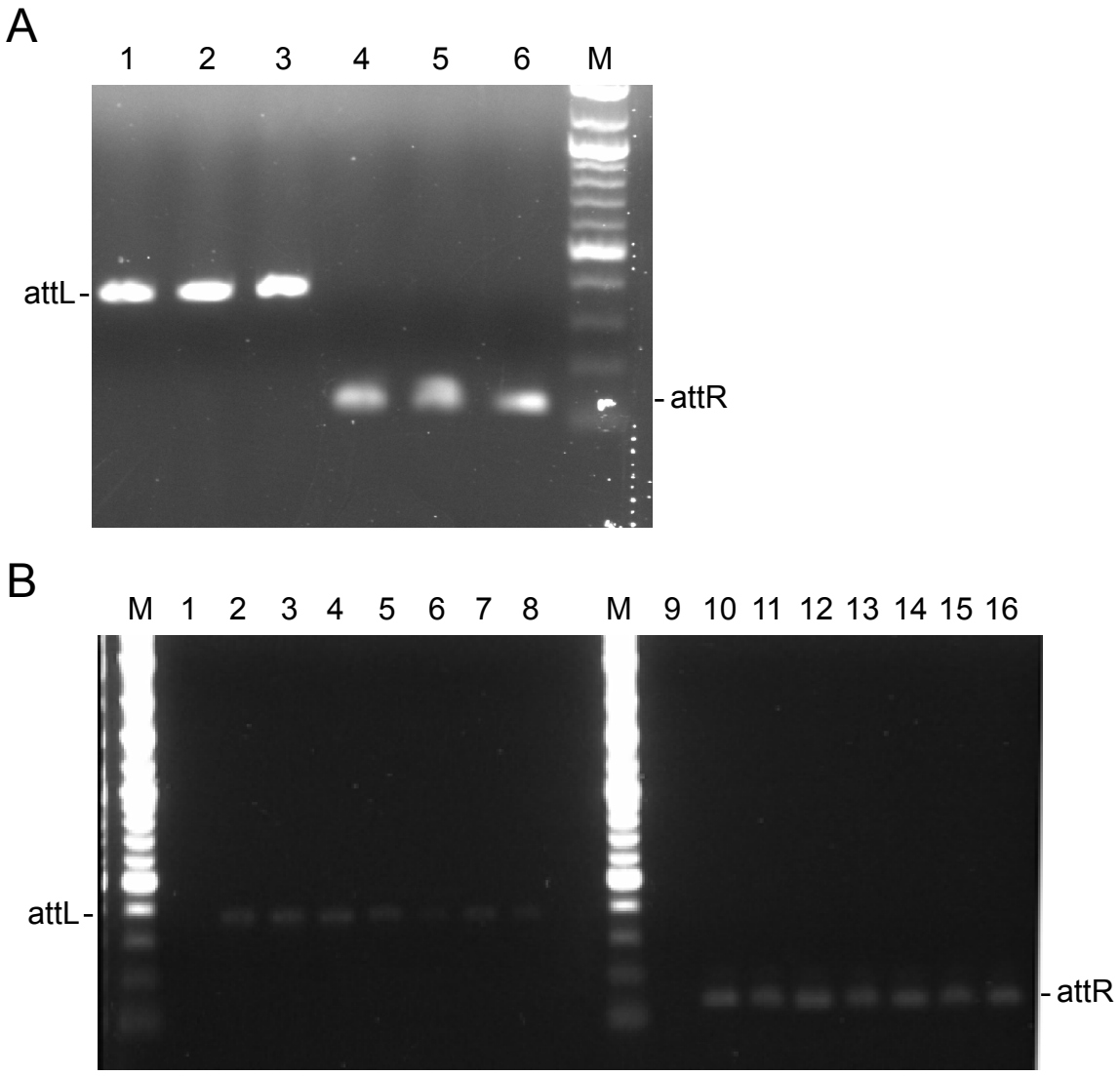
A
1 2 3 4 5 6 M
attL
attR
B
M 1 2 3 4 5 6 7 8 M 9 10 11 12 13 14 15 16
attL
attR

Figure S1

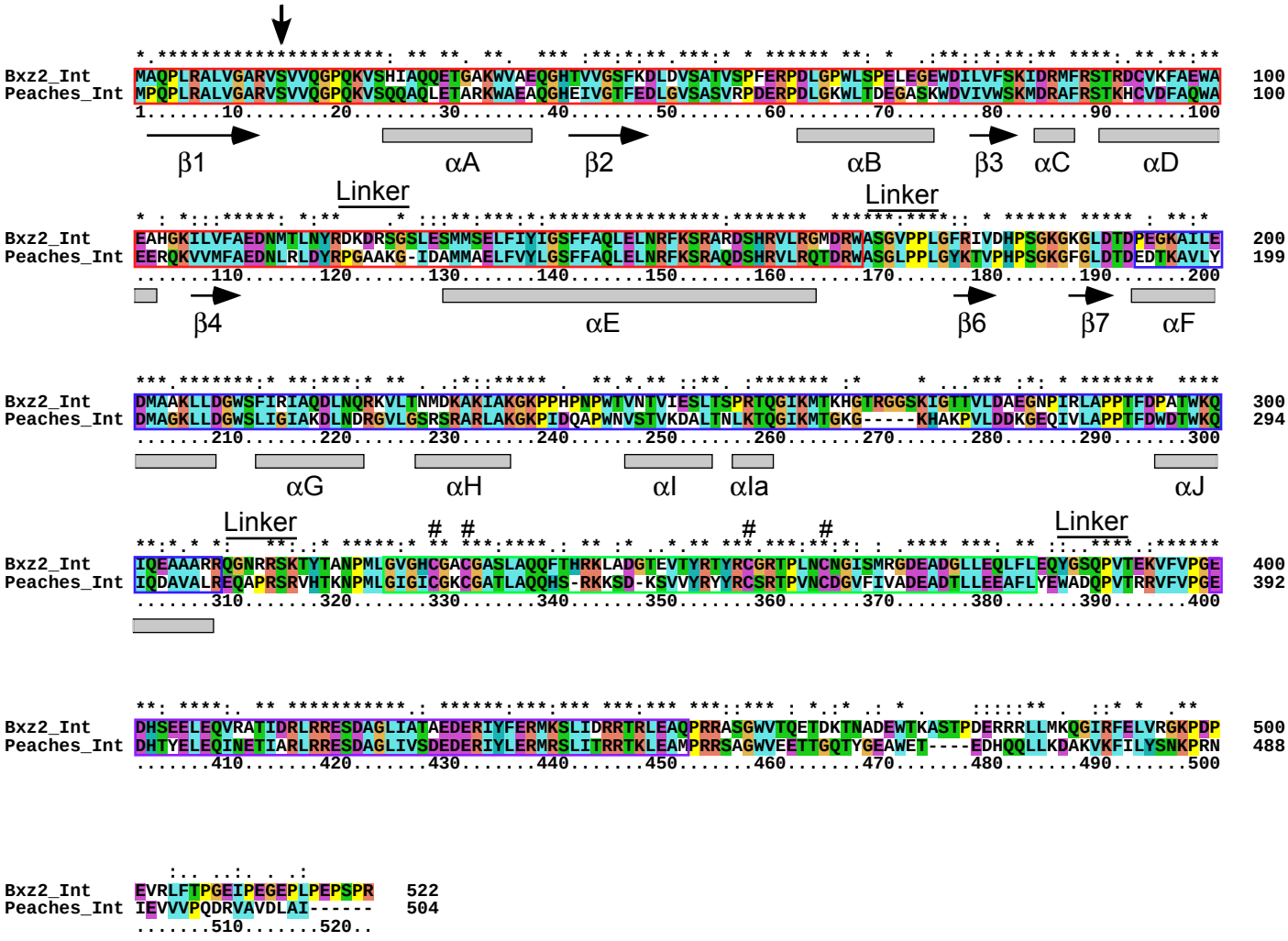
Bxz2_Int
Peaches_Int
β1
αA
β2
αB
β3
αC
αD
Linker
Linker
β4
αE
β6
β7
αF
αG
αH
αI
αIa
αJ
Linker
Linker

Figure S2

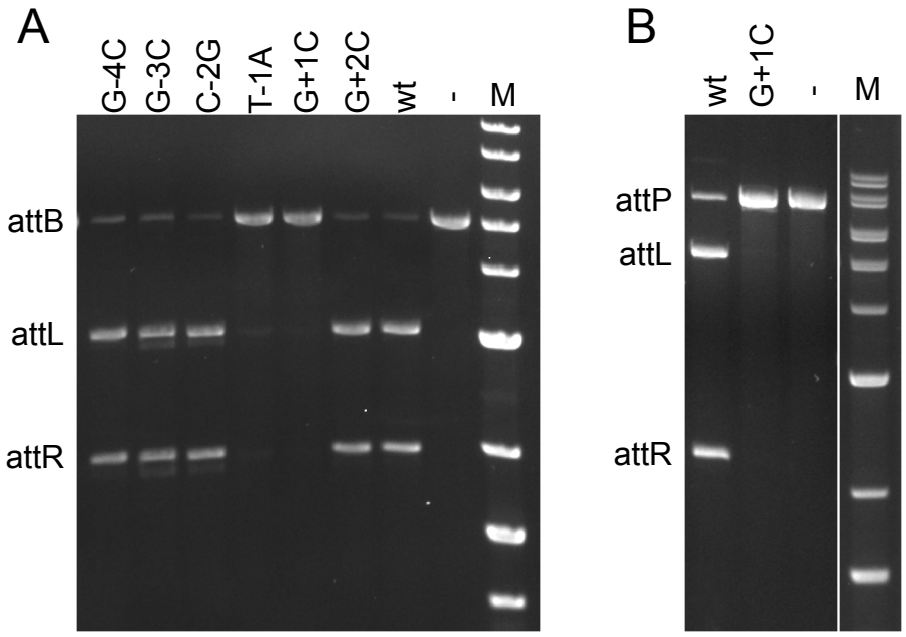
A
G-4C
G-3C
C-2G
T-1A
G+1C
G+2C
wt
-
M
attB
attL
attR
B
wt
G+1C
-
M
attP
attL
attR

Figure S3

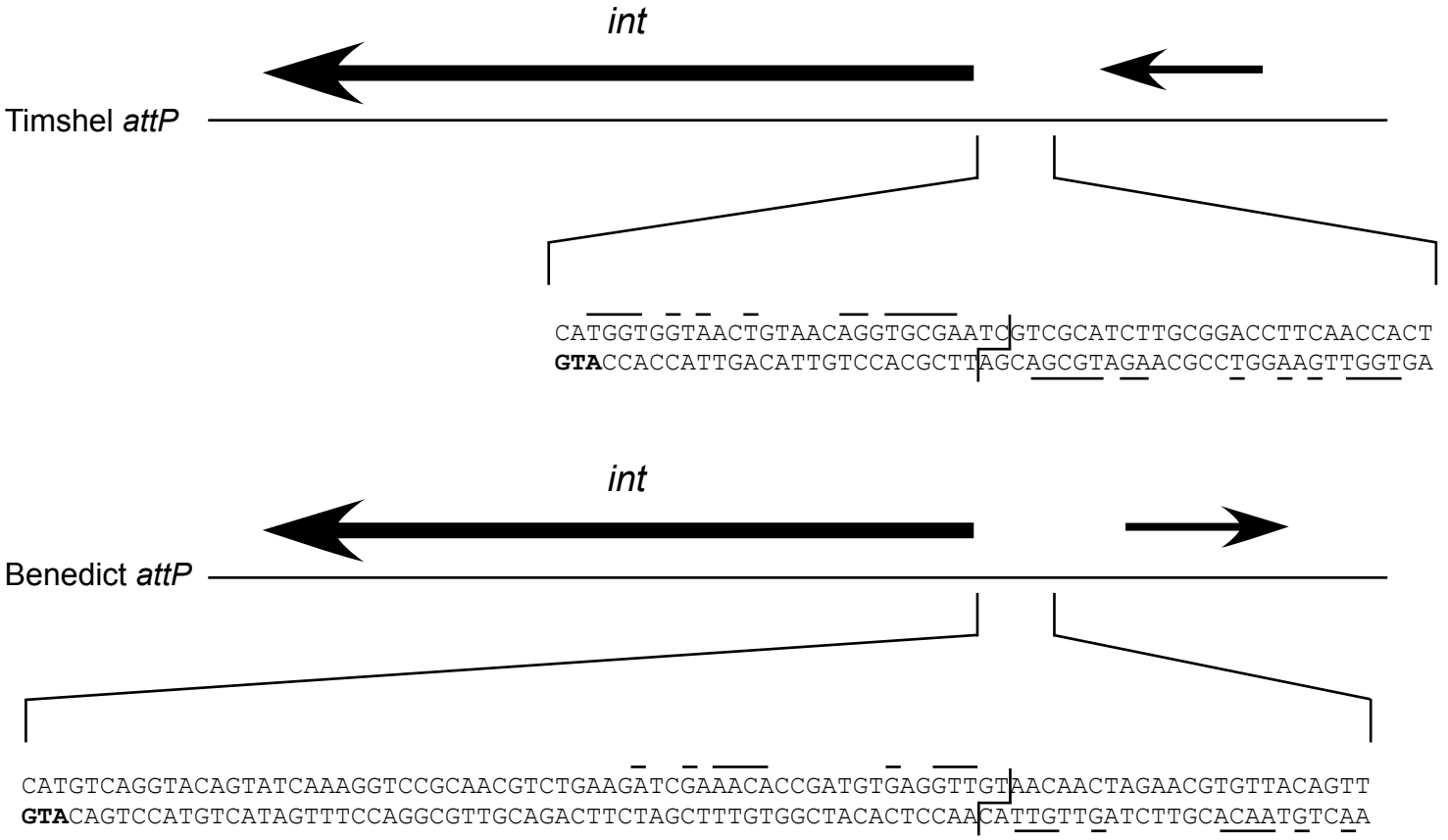
int
Timshel attP
CATGGTGGTAACTGTAACAGGTGCGAATCGTCGCATCTTGCGGACCTTCAACCACT
GTACCACCATTGACATTGTCCACGCTTAGCAGCGTAGAACGCCTGGAAGTTGGTGA
int
Benedict attP
CATGTCAGGTACAGTATCAAAGGTCCGCAACGTCTGAAGATCGAAACACCGATGTGAGGTTGTAACAACTAGAACGTGTTACAGTT
GTACAGTCCATGTCATAGTTTCCAGGCGTTGCAGACTTCTAGCTTTGTGGCTACACTCCAACATTGTTGATCTTGCACAATGTCAA

Figure S4