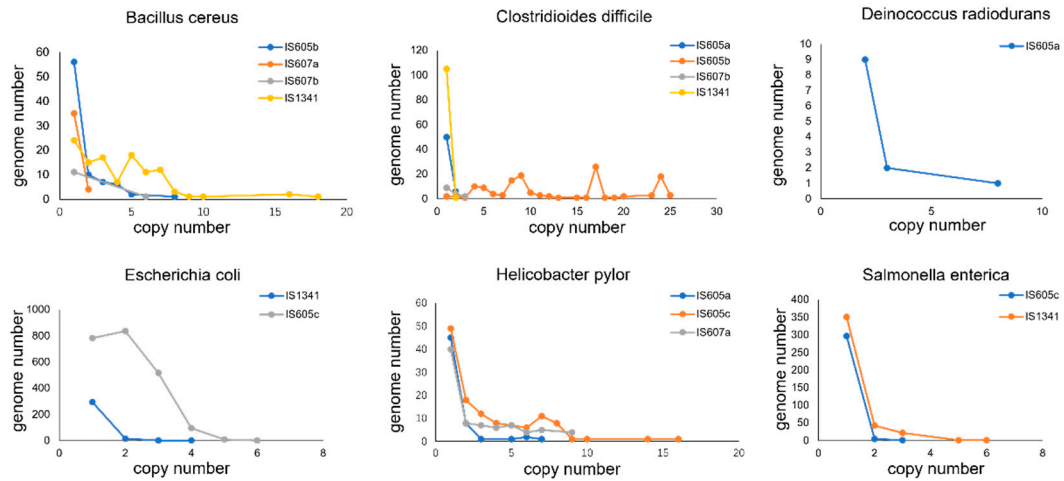


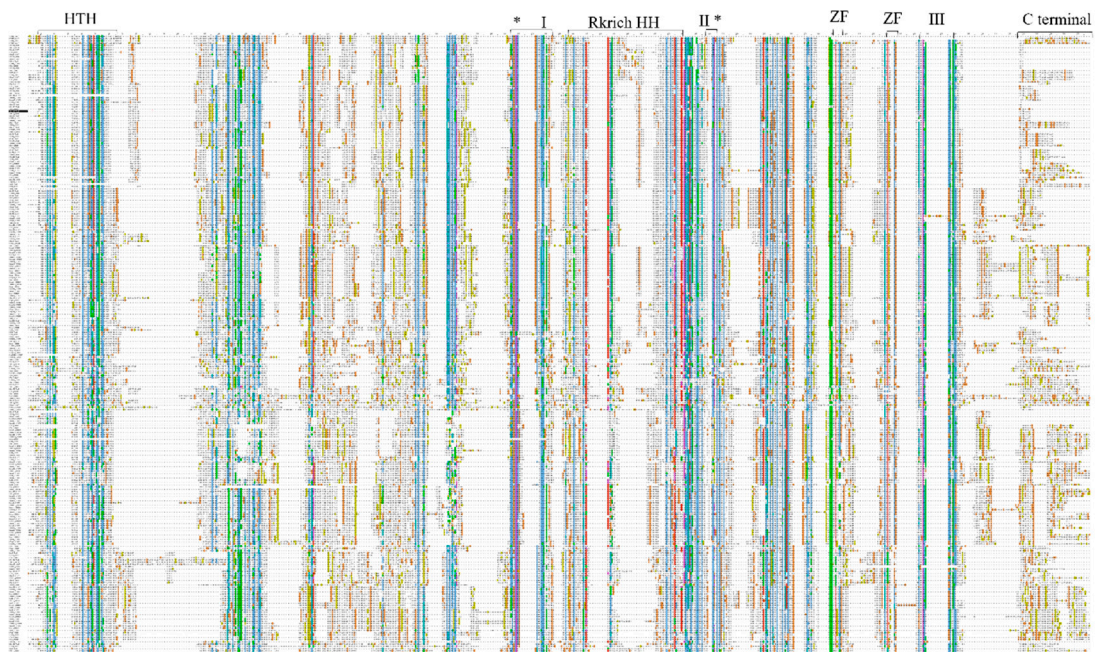
Supplementary Figure. S1. The number of TnpB hits in 19 species. The figure listed the top 19 species containing TnpB hits.

ISBce25 LE gagaaggag **gAAgtAA** CAGTTGAAATAAA **TATAAAA** CT **TATAAAA** GACCCCTCAATTA **TATAAAA** GTTTAA **TAATAAT** AATTAGA
 ISBce25 RE cataccatt **CTGTGCCAAA** ACCT **TATAAAA** GT **TATAAAG** TTGTGTTATTTCCCTGTTCTACGAGTCCGGTT **TGCCAAA** GCTACTACCGTTTGAT
 IS607 LE ttaataaag **gTtAcAA** CCTATAA CCTATAA CCTATAA CCTATAA CCTATAA TTTTGTATAGGTTATT
 IS607 RE gcatatctag **CCGT** TACTACAAA **CCTATAAA** AT **CCTATAA** AG **GCTATAA** AATT **CTCTTAA** TTT
 ISHp2 LE tgtttgataa **TA** T **TATTA** AACTTTCA **ATTTAAT** TTT **GTTTAAA** **TATATTA** T **TGTATAA** TC **TAAGAAA** TTCACCTCAAAGGTA
 ISHp2 RE aatgactta **gAT** TAAA TATTAAA TATTAAA TATTAAA TATTAAA TATTAAA TATTAAA TATTAAA
 ISBce17 LE aaacataat **GG** AG **AAAGAAA** ACAGT **AAACAAAACAT** AGACA **CAAGTA** GGCATATATGGTGA **AAAGTA** ATTATG **AAACAA** TATAAAC
 ISBce17 RE tatgagatat **CT** GCTACTA **AAAATA** **AAACAA** ATGTATATTCTTAGAGTATTCCTGCTTCTACGTCGCCCTGCCTCGCTAATCA
 ISCd3 LE taaacattca **GG** AGCTTATCTTTAGTTAATA **AAATAT** **AAATAT** AGCTGTCTATAGTCAAGAGAT **AAATAT** ATGAAAAACAATAA
 ISCd3 RE gacatagtat **TC** TGCTACTCA **AAATAT** GGTCA **AAATAT** AT **CCATAT** CGAGATTTTCTTCTCAITGAACCTTCGCTTGCCTAAGCTACTA
 ISHp3 LE ttggagtaat **TA** **AACTAAT** AAAAAA **AACTATTG** ACAAAACAT **AACTAAT** CTATG **ATAAAAT** AGCT **ATATAAT** AAATTAA
 ISHp3 RE gagttagatg **CT** GTTCCAT **AAACAT** AGACA **AAATGCC** TACACATACAGT **AAATCGC** ACTTCGTTGAT **AAATGAA** ATTCCTCATTT

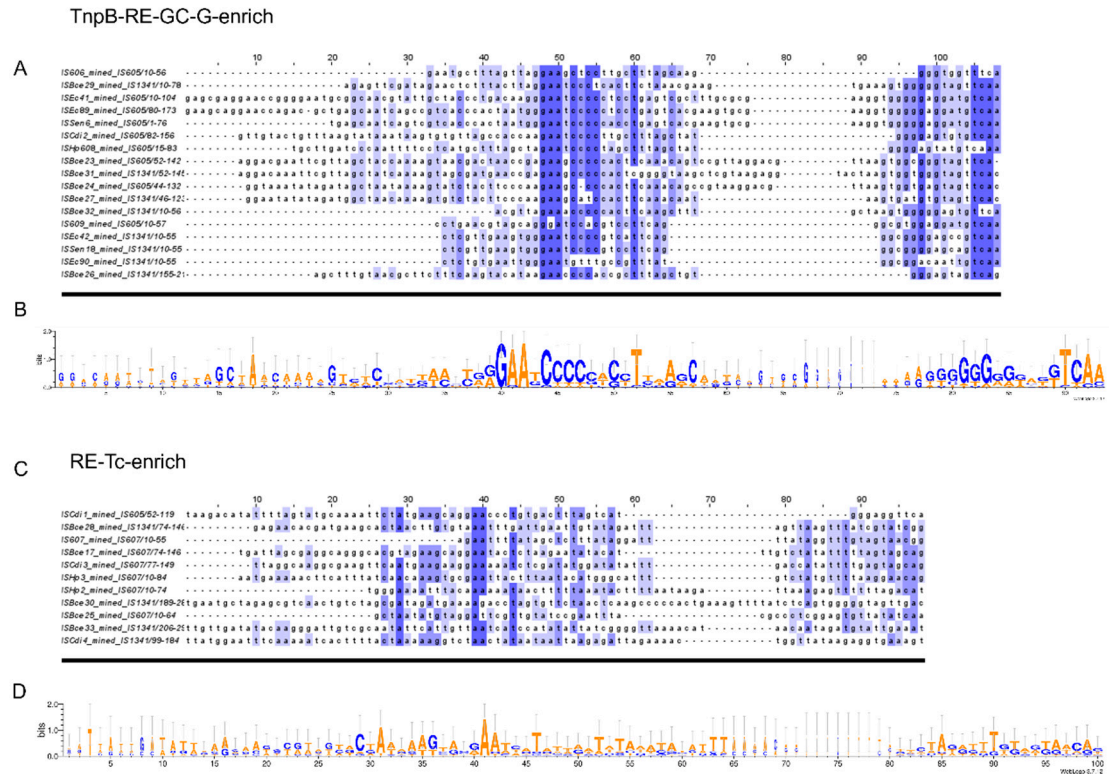
Supplementary Figure. S2. The LE and RE structures of the *IS607* group. The cleavage sites are shown in blue. The direct repeat sequences are shown in green. The top nucleotide sequences show variations from the direct repeat sequences.



Supplementary Figure. S3. The various copy number of IS elements in different genomes. The most of genomes only contains one copy, and a small fraction contains multiple copies..



Supplementary Figure. S4. Alignment of 274 TnpB sequences includes represented sequences from ISfinder database and 39-mined sequence from six species.



Supplementary Figure. S5. The character of RE sequences. (A) The alignment of part of TnpB-mined RE sequences. Two motifs were found which showed GC-rich position. (B) The logo of this type RE showed C-rich and G-rich conserved motifs. (C) The alignment of AT-TG-rich type RE sequences which showed two conserved motifs. (D) The logo of AT-TG -rich RE. This type RE showed one TA-rich motif and one TG-rich motif.