SUPPLEMENTARY INFORMATION for

Chromosomal replication dynamics and interaction with the β sliding clamp determine orientation of bacterial mobile elements

by Manuel J. Gómez, Héctor Díaz-Maldonado, Enrique González-Tortuero
and Francisco J. López de Saro

- **Supplementary Methods**
- **Table S1.** List of Pfam domains used to identify bacterial transposases analyzed in this study.
- **Table S2.** Pfam-based architecture description of IS structures and classification into IS families.
- **Table S3 (see attached PDF file).** List of chromosomes analyzed, IS orientation data, and statistical significance for the non-random orientation of IS elements.
- **Table S4 (see attached PDF file)**. List of chromosomes with 6 or more copies of a given IS family.
- **Table S5.** Alignments of regions containing β motifs in transposases.

SUPLEMMENTARY METHODS

Purification of *E. coli* β. The DnaN gene of *E. coli* MG1655 was amplified by PCR from genomic DNA with oligonucleotides #1 (5'-CGGCTAGCCATGGAATTTACCGTAGAACGTGAGC-3') and #2 (5'-GGCCTAGGATCCTTACAGTCTCATTGGCATGACAACATAAGCCGC-3'), cloned (NcoI-BamHI) in vector pET16b (Novagen Inc.), and sequenced. This plasmid was used to overexpress β in *E. coli* BL21(DE3). 15 g of cells (dry weight) were resuspended in Buffer A (100 mM TrisCl, 100 mM NaCl, 2 mM EDTA, 5% glycerine, 1 mM β-mercaptoethanol, 1 mM PMSF, pH 8) and processed four times with a French press. All purification procedures took place at 4ºC. The lysed cells were diluted to 150 mL and centrifuged (17,200 g, 30 m, 4ºC). The supernatant was treated with 0,2% polyethyleneimine and the precipitate removed by centrifugation. The supernatant was then brought up to 50% saturation with $[NH_4]_2SO_4$. After centrifugation, the supernatant was further treated with $[NH_4]_2SO_4$ to 70% saturation and centrifuged. The precipitate, containing β, was dissolved in 27 mL of Buffer B (100 mM TrisCl, 100 mM NaCl, 2 mM EDTA, 5% glycerine, 1 mM β-mercaptoethanol, pH 7.2) and dialyzed against 2 L of this buffer for 12 h. This fraction (27 mL) was then applied on a 30 mL Q sepharose FF (GE Healthcare) ion-exchange chromatography column equilibrated in Buffer B. Protein was eluted with Buffer B and a NaCl gradient (320 mL, 0.1-1.0 M) and fractions (3 mL) collected. Fractions containing β (30 mL) were pooled and dialyzed against 1 L of Buffer C (50 mM TrisCl pH 7.2) + 1 M $[NH_4]_2SO_4$. This fraction (250 mg) was then applied on a 20 mL Phenyl Sepharose 6 FF (GE

Healthcare) chromatography column equilibrated in Buffer C. Hydrophobic interaction chromatography was performed in Buffer C with a double gradient (100 mL) of $[NH_4]_2SO_4$ (1.0-0 M) and ethylene glycol (0-60%). Fractions containing β were pooled and dialyzed against 2 L of Buffer D (50 mM TrisCl, 50 mM NaCl, 1 mM EDTA, 10% glycerine, 1 mM DTT, pH 8). This fraction (2.7 mg mL^{-1}) was aliquoted and stored at -80ºC.

Purification of GST-Pol IV^{LF}. The C-terminal domain of *E. coli* DNA polymerase IV (residues Val243-Leu351) was amplified by PCR using genomic DNA from strain *E. coli* MG1655 and oligonucleotides #3 (5’-GAATGACGGATCCGTCGGCGTGGAACGCACGATGGCGGAAG-3’) and #4 (5’-GAATGACCTCGAGTCATAATCCCAGCACCAGTTGTCTTTCCATTTGCGGG-3’), and cloned (BamHI-XhoI) in pGEX-4T-3 (GE Healthcare). The resulting soluble GST fusion protein (335 amino acids, 39 kDa) was purified by affinity chromatography with Glutathione Sepharose 4B following the standard procedures recommended by the manufacturer (Amersham).

Purification of *Methanosarcina barkeri* PCNA. Genomic DNA from *Methanosarcina barkeri* Fusaro (DSM 804) was obtained from DSMZ (Braunschweig, Germany) and used to amplify the gene encoding PCNA by PCR using the oligonucleotides #5 (5’-GGATGACCATGGTTAAGGCAGCAATTAATGCAGAGCTTC-3’) and #6 (5’-GGATGAGGATCCTCAGTCCGACTCAATTCTTGGAGCCAGG-3’). The PCNA gene was cloned (NcoI-BamHI) into vector pET16b and the protein overexpressed in *E. coli* BL21(DE3). Purification of PCNA followed the same protocol as for *E. coli* β with the following differences: a) $[NH_4]_2SO_4$ was raised to 40% saturation to remove contaminants and then to 60% to precipitate PCNA; b) the NaCl gradient on the Q sepharose FF chromatography column was from 0.2 to 1.0 M.

TABLE S1. List of Pfam domains used to identify the bacterial transposases analyzed in this study.

DDE_2
DDE_3
DDE_4
DDE_4_2
DDE_5
DDE_Tnp_1
DDE_Tnp_1_2
DDE_Tnp_1_3
DDE_Tnp_1_4
DDE_Tnp_1_5
DDE_Tnp_1_6
DDE_Tnp_1_assoc
DDE_Tnp_2
DDE_Tnp_IS1
DDE_Tnp_IS1595
DDE_Tnp_IS240
DDE_Tnp_IS66
DDE_Tnp_IS66_C
DDE_Tnp_ISAZ013
DDE_Tnp_ISL3
DDE_Tnp_Tn3
DEDD_Tnp_IS110
Dimer_Tnp_Tn5
DUF2080
DUF4096
DUF4158
DUF4277
DUF4338
DUF4351
DUF4372
DUF772
HTH_17
HTH_21
HTH_23
HTH_24
HTH_28
HTH_29
HTH_32
HTH_33
HTH_38
HTH_7
HTH_OrfB_IS605
HTH_Tnp_1
HTH_Tnp_IS1
HTH_Tnp_IS630
HTH_Tnp_IS66
HTH_Tnp_ISL3
HTH_Tnp_Mu_1
HTH_Tnp_Mu_2
IstB_IS21
IstB_IS21_ATP
LZ_Tnp_IS481
LZ_Tnp_IS66
MULE
Mu-transpos_C
Nterm_IS4
OrfB_IS605
OrfB_Zn_ribbon
Phage-MuB_C
Resolvase
rve
rve_2
rve_3
Tn7_Tnp_TnsA_C
Tn7_Tnp_TnsA_N
Tn7_TnsC_Int
TnpB_IS66
Transposase_20
Transposase_31
Transposase_mut
Y1_Tnp
Y2_Tnp
Zn_Tnp_IS1
Zn_Tnp_IS1595
Zn_Tnp_IS91

TABLE S2. Pfam-based architecture description of IS structures and classification into IS families. The table lists 209 gene clusters corresponding to IS elements. Gene symbols are enclosed by characters that represent an arrow (--gene->), to describe their relative orientation, and are connected by double colons (::). The gene symbols represent the domain organization of encoded proteins. Domain names follow Pfam nomenclature and are connected by equal signs (=). IS family nomenclature follows that of IS finder (www-is.biotoul.fr//is.html), except for IS4, IS5, IS66, and ISNCY, as described in accompanying Notes. Importantly, our classification aims to group the domain architectures and proteins of each IS family according to the transposase domains detected and does not take into account other aspects of IS structure, such as the flanking inverted repeats present in many ISs.

IS FAMILY	***IS STRUCTURE***	***NOTES***
IS1	--DDE_Tnp_IS1-> --Zn_Tnp_IS1=HTH_Tnp_IS1->::--DDE_Tnp_IS1-> --Zn_Tnp_IS1=HTH_Tnp_IS1-> --Zn_Tnp_IS1=DDE_Tnp_IS1-> --HTH_23=DDE_Tnp_IS1-> --Zn_Tnp_IS1=HTH_Tnp_IS1=DDE_Tnp_IS1-> --HTH_Tnp_IS1->::--DDE_Tnp_IS1-> --Zn_Tnp_IS1->::--DDE_Tnp_IS1->	
IS3	--HTH_Tnp_1->::--HTH_21=rve-> --HTH_21=rve-> --HTH_Tnp_1->::--HTH_21=rve_3-> --HTH_29=rve-> --rve_3-> --HTH_28=HTH_28->::--HTH_21=rve-> --rve_2-> --HTH_Tnp_1=HTH_21=rve-> --HTH_28->::--HTH_21=rve-> --HTH_Tnp_1->::--rve_3-> --HTH_21=rve_3-> --HTH_Tnp_1->::--HTH_21-> --HTH_28=HTH_21=rve-> --HTH_28=rve-> --HTH_28->::--HTH_21=rve_2-> --HTH_21=rve_2-> --HTH_32=rve-> --HTH_28=rve_3-> --HTH_Tnp_1=HTH_28->::--HTH_21=rve-> --HTH_Tnp_1=HTH_21=rve_3-> --HTH_Tnp_1->::--HTH_29=rve-> --HTH_29=rve_3-> --HTH_28=HTH_Tnp_1=HTH_28->::--HTH_21=rve-> --HTH_21->::--rve_2-> --HTH_29=rve=rve_2-> --HTH_Tnp_1->::--HTH_21=rve_2-> --HTH_Tnp_1=rve-> --HTH_32=rve_3-> --HTH_Tnp_1->::--HTH_21->::--rve_3-> --HTH_28=HTH_21=rve_2-> --HTH_38=HTH_Tnp_1->::--HTH_21=rve_2-> --HTH_28->::--rve_2-> --HTH_28=HTH_28=HTH_28->::--HTH_21=rve-> --HTH_21->::--rve_3->	

	--HTH_Tnp_1->::--HTH_32=rve-> <-HTH_Tnp_1--::--rve_3-> --HTH_28=HTH_28=HTH_21=rve-> --HTH_28->::--HTH_21->::--rve_2-> --HTH_28->::--rve_3-> --HTH_Tnp_1->::--HTH_38=rve-> --HTH_Tnp_1->::<-HTH_38=rve-- --HTH_29=rve=HTH_29-> --HTH_Tnp_1=rve_3-> --HTH_23=HTH_38=rve-> --HTH_28=HTH_28->::--HTH_21->::--rve_2-> --HTH_Tnp_1->::--rve_2-> --HTH_23=HTH_21=rve->	
IS4a	--Nterm_IS4=DDE_Tnp_1-> --Nterm_IS4-> --Nterm_IS4->::--DDE_Tnp_1->	Includes members of subgroup IS4 of family IS4
IS4b	--DUF4372=DDE_Tnp_1-> --DUF4372-> --DUF4372->::--DDE_Tnp_1->	Includes members of subgroup IS4Sa of family IS4
IS4c	--Dimer_Tnp_Tn5-> --DDE_Tnp_1=Dimer_Tnp_Tn5->	Includes members of the IS50 subgroup of family IS4
IS5a	--DUF772=DDE_Tnp_1-> --DUF772-> --DUF772=DDE_Tnp_1_2-> --DUF772->::--DDE_Tnp_1-> --DUF772->::--DDE_Tnp_1_2->	Includes members of the IS5 subgroup of family IS5
IS5b	--DUF4096->::--DDE_Tnp_1_2-> --DUF4096-> --DUF4096=DDE_Tnp_1-> --DUF4096=DDE_Tnp_1_2-> --DUF4096->::--DDE_Tnp_1->	Includes members of subgroups IS427 and IS1031 of family IS5
IS5c	--DDE_Tnp_1_5->	Includes members of subgroup ISL2 of family IS5
IS5d	--DDE_4-> --DDE_4_2-> --DDE_4_2->::--DDE_4-> --DDE_4_2=DDE_4->	Includes members of subgroup IS903 of family IS5
IS6	--DDE_Tnp_IS240-> --Zn_Tnp_IS1=DDE_Tnp_IS240->	
IS21	--IstB_IS21-> --HTH_38=rve->::--IstB_IS21-> --IstB_IS21_ATP=IstB_IS21-> --HTH_23=rve->::--IstB_IS21-> --HTH_7=rve->::--IstB_IS21-> --HTH_23->::--IstB_IS21-> --HTH_23=rve->::--IstB_IS21_ATP=IstB_IS21-> --HTH_29=rve->::--IstB_IS21-> --IstB_IS21->::--HTH_21=rve->	
IS30	--HTH_38=rve-> --HTH_38->	
IS66a	--HTH_Tnp_1->::--TnpB_IS66->::--LZ_Tnp_IS66=HTH_Tnp_IS66=DDE_Tnp_IS66=DDE_Tnp_IS66_C-> --TnpB_IS66->::--LZ_Tnp_IS66=HTH_Tnp_IS66=DDE_Tnp_IS66=DDE_Tnp_IS66_C->	The IS66 family was divided in two groups according to Gourbeyre *et al.* (2010) Route 66: investigations into the organisation and

	--TnpB_IS66->::--LZ_Tnp_IS66=HTH_Tnp_IS66=DDE_Tnp_IS66-> --LZ_Tnp_IS66=HTH_Tnp_IS66=DDE_Tnp_IS66=DDE_Tnp_IS66_C-> --LZ_Tnp_IS66=HTH_Tnp_IS66=DDE_Tnp_IS66-> --TnpB_IS66->::--LZ_Tnp_IS66=DDE_Tnp_IS66=DDE_Tnp_IS66_C-> --HTH_Tnp_1->::--TnpB_IS66->::--LZ_Tnp_IS66=HTH_Tnp_IS66=DDE_Tnp_IS66-> --HTH_Tnp_1->::--LZ_Tnp_IS66=HTH_Tnp_IS66=DDE_Tnp_IS66=DDE_Tnp_IS66_C-> --TnpB_IS66->::--LZ_Tnp_IS66->::--HTH_Tnp_IS66=DDE_Tnp_IS66=DDE_Tnp_IS66_C-> --TnpB_IS66->::--LZ_Tnp_IS66->::--HTH_Tnp_IS66=DDE_Tnp_IS66-> --TnpB_IS66->::--HTH_Tnp_IS66=DDE_Tnp_IS66-> --TnpB_IS66->::--LZ_Tnp_IS66=DDE_Tnp_IS66-> --HTH_Tnp_1->::--TnpB_IS66->::--LZ_Tnp_IS66-> --HTH_Tnp_IS66=DDE_Tnp_IS66=DDE_Tnp_IS66_C-> --TnpB_IS66->::--HTH_Tnp_IS66=DDE_Tnp_IS66=DDE_Tnp_IS66_C-> --HTH_23=HTH_Tnp_1->::--TnpB_IS66->::--LZ_Tnp_IS66=HTH_Tnp_IS66=DDE_Tnp_IS66=DDE_Tnp_IS66_C-> --TnpB_IS66->::--LZ_Tnp_IS66=HTH_Tnp_IS66->	distribution of the IS66 family of prokaryotic insertion sequences. Research in Microbiology **161**: 136-143. IS66a includes the ‘classical’ IS66 architecture.
IS66b	--HTH_Tnp_IS66=DDE_Tnp_IS66->	Includes the subgroup ISBst12 of IS66 (Gourbeyre *et al*., 2010). See IS66a.
IS91	--Zn_Tnp_IS91=Y2_Tnp-> --Y2_Tnp-> --Zn_Tnp_IS91->::--Y2_Tnp->	
IS110	--DEDD_Tnp_IS110=Transposase_20-> --Transposase_20-> --DEDD_Tnp_IS110-> --DEDD_Tnp_IS110->::--Transposase_20->	
IS200	--Y1_Tnp-> --Y1_Tnp=Y1_Tnp->	
IS200/IS605	<-HTH_OrfB_IS605=OrfB_IS605=OrfB_Zn_ribbon--::--Y1_Tnp-> --Y1_Tnp->::--HTH_OrfB_IS605=OrfB_IS605=OrfB_Zn_ribbon-> --Y1_Tnp->::--OrfB_IS605=OrfB_Zn_ribbon-> <-Y1_Tnp--::--OrfB_IS605=OrfB_Zn_ribbon-> <-OrfB_IS605--::--Y1_Tnp-> <-HTH_OrfB_IS605=OrfB_IS605--::--Y1_Tnp-> <-Y1_Tnp--::--HTH_OrfB_IS605->::--OrfB_IS605=OrfB_Zn_ribbon->	We have chosen TnpA (domain Y1_Tnp) as a proxy to score the orientation of the insertion sequences in this family.
IS256	--Transposase_mut->	
IS481	--LZ_Tnp_IS481=rve-> --LZ_Tnp_IS481->	
IS607	--Resolvase->::--HTH_OrfB_IS605=OrfB_IS605=OrfB_Zn_ribbon-> --HTH_17=Resolvase->::--OrfB_IS605=OrfB_Zn_ribbon-> --MerR=Resolvase-> --Resolvase->::--HTH_OrfB_IS605=OrfB_IS605-> --HTH_17=Resolvase-> --MerR=Resolvase->::--HTH_OrfB_IS605=OrfB_IS605=OrfB_Zn_ribbon-> --MerR_1=Resolvase->::--OrfB_IS605=OrfB_Zn_ribbon-> --HTH_17=Resolvase->::--OrfB_Zn_ribbon-> --MerR=Resolvase->::--OrfB_Zn_ribbon-> --MerR_1=Resolvase->::--OrfB_IS605-> --Resolvase->::--OrfB_IS605=OrfB_Zn_ribbon-> --MerR_1=Resolvase->::--HTH_OrfB_IS605=OrfB_IS605=OrfB_Zn_ribbon-> --MerR=Resolvase->::--OrfB_IS605=OrfB_Zn_ribbon-> --Resolvase->::--OrfB_IS605-> --HTH_17=Resolvase->::--OrfB_IS605->	

IS630	--HTH_29=DDE_3-> --DDE_3-> --HTH_Tnp_IS630=DDE_3-> --HTH_Tnp_IS630->::--DDE_3-> --HTH_29->::--DDE_3-> --HTH_23=HTH_33=DDE_3-> --HTH_Tnp_IS630-> --HTH_23=HTH_33->::--DDE_3-> --HTH_33=DDE_3-> --HTH_28->::--DDE_3-> --HTH_28=HTH_33->::--DDE_3-> --HTH_33->::--DDE_3-> --HTH_23->::--DDE_3-> --HTH_28=DDE_3-> --HTH_23=DDE_3-> --HTH_23->::--HTH_33->::--DDE_3->	
IS701	--DDE_5-> --DDE_5=DDE_Tnp_1-> --DDE_Tnp_1->::--DDE_5-> --DDE_5->::--DDE_Tnp_1-> --DDE_Tnp_1->::<-DDE_5-- <-DDE_Tnp_1--::--DDE_5->	
IS982	--DDE_Tnp_1_3-> --DDE_Tnp_1_3=DDE_Tnp_1_3->	
IS1182	--DUF772=DDE_Tnp_1_6-> --DDE_Tnp_1_6-> --DUF772->::--DDE_Tnp_1_6->	
IS1380	--DDE_Tnp_1_4->	
IS1595	--DDE_Tnp_IS1595-> --Zn_Tnp_IS1595=DDE_Tnp_IS1595-> --Zn_Tnp_IS1595->::--DDE_Tnp_IS1595->	
IS1634	--DUF4277-> --DUF4277=DDE_Tnp_1->	
ISAs1	--DDE_Tnp_1_assoc=DDE_Tnp_1-> --DDE_Tnp_1_assoc-> --DDE_Tnp_1_assoc->::--DDE_Tnp_1-> --DUF4338=DDE_Tnp_1_assoc=DDE_Tnp_1-> --DUF4338->::--DDE_Tnp_1_assoc->	
ISAzo13	--DDE_Tnp_ISAZ013->	
ISL3	--HTH_Tnp_ISL3=DDE_Tnp_ISL3-> --DDE_Tnp_ISL3-> --HTH_Tnp_ISL3->::--DDE_Tnp_ISL3-> --HTH_Tnp_ISL3=DDE_Tnp_ISL3=DDE_Tnp_ISL3-> --DDE_Tnp_ISL3=DDE_Tnp_ISL3-> --HTH_Tnp_ISL3->::--DDE_Tnp_ISL3=DDE_Tnp_ISL3->	
ISNCYa	--Transposase_31-> --DUF4351-> --Transposase_31=DUF4351->	IS*Plu*15
ISNCYb	--MULE->	ISM1
ISNCYc	--DDE_Tnp_2->	ISC1217
ISNCYd	--DUF2080->	ISA1214

Tn3	--DUF4158=DDE_Tnp_Tn3-> --DDE_Tnp_Tn3-> --Resolvase=HTH_7->::--DUF4158=DDE_Tnp_Tn3-> <-Resolvase=HTH_7--::--DUF4158=DDE_Tnp_Tn3-> --Resolvase->::--DUF4158=DDE_Tnp_Tn3-> <-DUF4158=DDE_Tnp_Tn3--::--Resolvase-> --DUF4158->::--DDE_Tnp_Tn3-> --Resolvase=HTH_7->::--DUF4158->	
Tn7	--Tn7_Tnp_TnsA_N=Tn7_Tnp_TnsA_C-> --Tn7_Tnp_TnsA_N->::--rve=Mu-transpos_C-> --Tn7_Tnp_TnsA_N=Tn7_Tnp_TnsA_C->::--rve=Mu-transpos_C-> --Tn7_Tnp_TnsA_N=Tn7_Tnp_TnsA_C->::--HTH_28=rve=Mu-transpos_C-> --Tn7_Tnp_TnsA_N=Tn7_Tnp_TnsA_C=rve=Mu-transpos_C-> --Tn7_Tnp_TnsA_N=rve=Mu-transpos_C-> --Tn7_Tnp_TnsA_N=Tn7_Tnp_TnsA_C=rve->	

TABLE S3 (In annexed PDF file). Statistical significance for the non-random distribution of IS elements at the chromosomal level. Each chromosome is listed with the number of ISs found for each IS family. The number of ISs found in and each of the two possible orientations relative to the local GC skew sign (same/anti) are indicated (See Materials and Methods for details). *p*-values were calculated for each chromosome and IS family, and coloured according to its significance (red for $p<10^{-3}$, orange for $10^{-3}<p<10^{-2}$ and yellow for $10^{-2}<p<0.05$).

Chromosomes used for calculation of statistical significance at the Phylum level (Table 1) are marked red in the first column.

TABLE S4 (in annexed PDF file). Statistical significance for the non-random distribution of IS elements at the chromosomal level. As in Table S3, but including only chromosomes with 6 or more copies of a give IS family, and only one chromosome for each species. The table has been ordered by IS family and *p*-value. Significant p-values are labelled (red for $p<10^{-3}$, orange for $10^{-3}<p<10^{-2}$ and yellow for $10^{-2}<p<0.05$).

TABLE S5. Sequences of transposase regions containing the β motif. *Escherichia coli* peptides used for biochemical analysis (Fig. 2a) are boxed, and residues putatively corresponding to the β binding motif are in bold. Phylum code: Actinobacteria (A), Bacteroidetes/Chlorobi (B), Chlamydiae/Verrucomicrobia (V), Cyanobacteria (C), Deinococcus-Thermus (D), Firmicutes (F), Fusobacteria (U), Nitrospirae (N), Planctomycetes (M), Proteobacteria (P), Spirochaetes (S), Tenericutes (T), Thermotogae (H) and Archaea (Arch.).

IS5a

P	*Escherichia coli* (AAB53644)	305-RESQIQGVAENDN**QL**AM**LF**TLANLFRADQMIRQWERSH*
P	*Yersinia enterocolitica* (CAE46781)	231-TKVRFRGLVRNTA**QL**V**TLF**ALSNLWMARRHLLASAGEVRL*
P	*Vibrio cholerae* (EKG68054)	289-IKARYKGLMKNDN**QL**AM**LF**TLANLVKVDQLIRRQARSA*
P	*Pseudomonas putida* (CAB99195)	286-VKTRFRGLAKNTA**QL**V**TLF**ALSNLWMARRHLLTNAGEVRL*
P	*Legionella pneumophila* (CBX00305)	308-IVICVMVVNQALV**Q**SMD**L**TAYRY*
P	*Neisseria meningitides* (CBA09446)	333-ARAAYFGLSKVSA**Q**SH**L**KAMCLNLLKAANRLSAPAAA*
P	*Ralstonia solanacearum* (AAR23794)	289-RKVRYKGLAKNTA**QL**F**SLF**GLANLVLARRQLLASPGSIAS*
F	*Bacillus* sp. (ZP_01722218)	415-RWTTLRGLKKLSM**Q**AM**L**T**F**AAMNLKKLATWTWQVA*
F	*Streptococcus thermophilus* (CAE52417)	311-TLTNLLYNIFRFE**QI**KR**L**GLKSWA*
F	*Lactobacillus acidophilus* (YP_004292069)	549-LCFYIRAKNRVSS**Q**T**LF**KRKIKLKSTSKLNP*
F	*Thermoanaerobacter ethanolicus* (ZP_08211387)	478-KQGLKFYFVNKLV**Q**IK**LF**AFLYKYYR*
C	*Crocosphaera watsonii* (EAM49691)	446-ITFLVVNLSKLLR**QL**L**SLF**LSLFTNNRTGNSSNRLSLILIIL*

IS30

P	*Escherichia coli* (NP_415922)	329-TNGLIRQYFPKKTCLAQYT**Q**HE**L**DLVAAQLNNRPRKTLKFKT-370
P	*Yersinia kristensenii* (ZP_04623201)	329-TNSLIRQYLPKKTCLAQHS**Q**EV**L**NQIADELNDRPRKTLKFKT-370
P	*Pseudoalteromonas citrea* (ZP_10273944)	331-TNRLLRQYFPKKTSLHGFD**Q**SY**L**DKIANKLNNRPRRILNYLT-372
P	*Burkholderia xenovorans* (YP_555985)	331-TNRLLRQYLPHGTQLDHYS**Q**AD**L**NKIAARLNERPRKTLGFRS-372
P	*Pseudomonas fluorescens* (YP_348496)	332-TNGLLRQYFLKGTDLAEHS**Q**AT**L**NEVARQLNSRPRKTLDYET-373
P	*Ralstonia eutropha* (YP_728612)	316-TNGLLRQYLPKGTDLSVYS**Q**AK**L**NAIARRLNERPRKTLNFDT-357
P	*Rickettsia massiliae* (YP_001499437)	274-MNSMIHRILPKNTDITTVT**Q**RG**L**DNVAEILNNMPRKIFGYKT-315
F	*Streptococcus mitis* (YP_003445190)	312-NHTVLRAILPKGTSFDQLT**Q**KD**V**NLVISHVNSLKREEFQGKS-353
F	*Streptococcus salivarius* (CAA78948)	289-FNGLLREFIPKGCSLKELN**Q**N**LL**EDYTKAINERPRRIHGYQS-330
F	*Streptococcus pneumonia* (CCG13893)	313-NHTLIRDILPKGTSFDNLT**Q**ED**I**NLVCSHVNSVKRAALNGKS-354
F	*Staphylococcus aureus* (YP_039551)	261-TNGLLREFFPKKTDLAKVN**Q**EQ**L**NYALDSINYRPRKCLNWKF-302
S	*Leptospira biflexa* (YP_001840433)	286-FFPKGTDFSKLKKSQIKKV**Q**T**LL**NQRPRKTLNWNTPEEEIRA-318
S	*Treponema succinifaciens* (YP_004365962)	319-NGLPRKRLGYKTPEELFNE**QL**D**LI**YRL*
A	*Frankia* sp. (YP_001508751)	434-NDRPRKTLDWKTPTEAMNN**QL**L**SL**QQPGVARTG*

IS66a (TnpB)

P	*Escherichia coli* (YP_424826)	66-LFTKRLERGRFVWPVTRDG-KVHLTPA**QLS**M**LL**EGINWKHPKRTERAGIRI*
P	*E. coli* (YP_003235003)	64-LFTKRLEEGQFIWPAVRDG-KVSITRS**QL**AM**LL**DKLDWRQPKTSSRNSLTML*
P	*Shigella sonnei* (YP_313273)	66-LFTRRLERGRFVWPVTRDG-KVHLTPA**QLS**M**LL**EGIDWKHPKRTERAGIRI*
P	*S. boydii* (YP_407150)	66-LFTRRLEEGQFIWPAVRDG-KVSITRS**QL**AM**LL**DKLDWRQPKTSRLNALTML*
P	*Polymorphum gilvum* (YP_004305836)	66-LVHKRLECGKFVWPQAQDG-VMRISSA**Q**MAA**LF**EGLDWRLVRPERARRPLVAG*
P	*Ochrobactrum intermedium* (ZP_04679718)	72-LFTKKLERGRFIWPSAADG-TVVITPA**QL**G**YLL**EGIDWRMPQKTWRPTSAG*
P	*Methylocystis* sp. (ZP_08074780)	66-LFAKRLEDGEFRWPKIEDG-TMRLSAT**Q**F**S**A**LL**EGLDWKRVHTKETPAPALPG*
P	*Agrobacterium vitis* (YP_002540116)	72-LFYKVLERGYFPWPRAKEG-VAPLTQA**QLS**M**LV**EGIDWRRPAWTSAPARTG*
P	*Rhizobium leguminosarum* (YP_002973065)	66-LAYKRLEEHTFTWPGIKDG-LMTLTHA**QF**EA**LF**AFDLVVSGHA*
P	*Pelagibaca bermudensis* (ZP_01442077)	66-MAYKRLEESTFTWPAIRDG-AMTLNRA**QF**EA**LF**AGLDWRRVRSLEVRRPAVAE*
P	*Desulfovibrio magneticus* (YP_002952289)	66-LWHKRLERHVFRWPTR-EAEVLAIDSR**QL**AW**LL**DGLDPLAVTGHSRLEYSTLF*
P	*Magnetococcus marinus* (YP_864253)	67-LWQKRLEKDRFHWLRQGGAAEIQITGR**QL**NW**LL**DGYNLAAMKGHNKLHFSSIV*
P	*A. ferrooxidans* (YP_002218731)	65-LVYRRLDQGRLHWPRADAG-ALELSAA**Q**WAM**LV**EGRPWTPLPTLEKCTPKLL*
P	*Burkholderia glumae* (YP_002907560)	66-LMLKRLEADHFVWPHR-EQAVIELTTE**QL**HW**LL**DGIDIDAMQRHPARRYRHAS*
P	*Cupriavidus taiwanensis* (YP_001795972)	66-LFLKRLEADRFAWPR--GAAVATLSVE**QL**HW**LL**DGIDISAVQRHPPRHYQRAV*
F	*Streptococcus pneumoniae* (YP_002742797)	67-LLYKRFENGRLTWPST-EKDVKALTPE**QV**DW**LM**KGFSITPKINPSESRDFY*
F	*Bacillus pseudofirmus* (YP_003426249)	66-LYYRRLEKGTFPWPEDSSSSPQMISHR**QF**RW**LL**DGLSIDQKSAHPKVTAQRVI*
F	*B. selenitireducens* (YP_003698933)	66-LYYRRLEKGRFPWPTSGSDEPMIITER**QL**RW**LL**DGLPLDQKGAHRKMNPEKVV*

IS66a (TnpC)

```
P Escherichia coli (YP_003235004)             MSQKYLIRIAELERLLSEQAEALRQKDQQLSLVEETEAFLRSALTRAEEKIEEDEREIEHLRAQIEKLRRMLFGTRSEKLRREVELAEALL-----KQREQDSDRY
P Shigella dysenteriae (EFW49566)             MNQKYLIRIAELECQL-------RQKDQQLSLVEETEAFLRSALARAEEKIEEDEREIEHLRAQIEKLRRMLFGTRSEKLRREVEQAEALL-----KQREQDSDRY
P Yersinia pestis (YP_003566509)            MVMSQDYLARIAALEDAL-------RQKDNQLSLVAETESFLRSALARAEEKIENEEREIEHLRAQIEKLRRMLFGTRSEKLRRQVEEAEALL-----KQQEQQSDRY
P Klebsiella pneumoniae (YP_001687996)        MNHDYLARIAALEDAL-------RQKDSQLSLVAETESFLRSALARAEEKIENEEREIEHLRAQIEKLRRMLFGTRSEKLRRQVEEAEALL-----KQQEQQSDRY
P Pantoea vagans (YP_003933479)                MKRSLSAENDRLRALL--------DTQQRSLEQMAEYNRLLS-----RRVAAYASEINRLKALVAKLQRMQFGKSSEKLREKTARQVREAEE---RISGLQEEMA

P Escherichia coli (ZP_07592975)              MDISLLSTTR-------DPEQLRALAIAMVQKAMAESQNLANVVQEKDRNIAELQNRIRILEEQMKLARQQRFGKKCESLAG---MQRSLFEE-DVDADIAEISAH
P Shewanella benthica (ZP_02157382)                                       MASQYSEIAELKQS---------------VQRLLEPFRLAQQQRFGASSESHN----YQGELFNE---AEVTLD-EPE
P Fulvimarina pelagi (ZP_01440753)            MATALEALPD-------------DPGTLKAMLIA--ERVRSE-----------------RLEQIIKELQRHRFGRRAETLPE---DQLLLGLE-DVEQGVAVEEAE
P Magnetococcus marinus (YP_866522)           MKIKPQTLPD-------------DPAELKALVQSLQEEMKL-------------------LREQLHILISKRFGRSSEKYDP---NQLGLFDEAELIGATAADDVE
P Aliivibrio salmonicida (YP_002262549)          MIDKIKP--------------LPDTIDELKALVLQLENK---------------YNRLLEQFRLAQHQRFGKSSESDS----TQFDLFNE---TEEEIIIEND
P Brucella ceti (ZP_03787083)                  MLNRGQHLPR-------------DPDILVGMILERDAEIER------------------LKVLLKAANAKPFGQRSEQLAHMVERQIRLDLG-DVVHEPEVASAE
P B. cellulosilyticus (YP_004096872)           MRPMEKIKNDKDAQF--------YKERAEKLELEKEELEAK---------------LKWYEEQFRLSQKRQFGASSEQTT----GQLSLFNE---IEDTSNKDVE
P Methylocystis sp. (ZP_08074907)              MSRAAADLPE-------------DPAELRRFAEALAAEVHAK------------TLLIEKLKMQLAVLRRARFGRSSEKLDRDIEQLELLIG-DMEESDAERQA-
P Polymorphum gilvum (YP_004305837)            MDAAVLAREN---------------ALLKARLIEVEAALAES------------QEANRRLEDILRTSQREKFGKRSEKLSP--DQFNLPLE-DAELAQGVLEA-
M Planctomyces maris (ZP_01851762)                                        MIHQLG---ETVGE-----------QQREVEQLKHFIDRLLRQRFGARSEKIA---PNQMSLFDE--PEAAEEATDPE
B Parabacteroides merdae (ZP_02032857)         MIHTDTMELIIKNQQEQIKGLLETNRTLVESNQKLMEQTGE-----------LQQKVQELLSQVAWLNRQLFGRKSEKLASLDPNQLALFDT--LANPRQEE---
B Bacteroides salanitronis (YP_004258102)      MIRQDTMEQIIRSQQEQIAGLLETNRSLVESNGKLLEQTDA-----------LQRKIQELLSQIAWLNRQLFGRRSEKLAALDPNQLSLFDS--VPATGQDEDIR
V Chthoniobacter flavus (ZP_03131569)           MPDSTFPN------------------AAELLARIAELEKE---------------NALLRQKIDALARKIYGVSSEKLDP---AQLHLLLQ---GLDEPGKAPE
F Lactobacillus casei (YP_005858453)                                   MSAEAVVTTEQFEYLKQE---------------NALLREQVEFLMRRLYGTKRESLTD--GQVDLFDQ--TKTFVAPTVP-

F Streptococcus pneumoniae (NP_358890)                                               MKIIQQQSAIIDSLTNELSLLCEQVAYLTQKLSGKSSEKSVCPFGQLNLFEEESPSEKDGDVPS*
F Lactobacillus parafarraginis (ZP_09394150)                      MILSLKVKQRTVTQLEIDELRKENAELRALVAKQAKQIELLQEQVNYLMSKLYGKSSEQTPEDGQTSLFEDDENGVFEQPESTGE*
F Enterococcus faecium (EJX82788)                                     METTDTLLQLLQEAHKTNQAQQQTIQNLTTEIQLLNEKVNYLTNKLFGRSKETLFEETNGQLNLFSDEEISVSVPEA-77
```

IS91

```
P Escherichia coli (ACO24927)           463-IEDPKVIEQILKHLKQKTAKANAAKQRELPPERAPPLTPSLFDPSQSRLFD*
P Gamma proteobacterium (ZP_01617481)   453-IEDPSVIKKILEHLDAKSMAL--TSANQLPEPRAPP-QAELFD*
P Gamma proteobacterium (CBL46704)      464-IEDPAVIEKILQHLAMKESLP----LPRVHEARAPPDQAALFQL*
P Nitrococcus mobilis (ZP_01126307)     435-IEDPEIIEKILAHLDAHVTEP---EATRRPPCRAPP-QRGLFDETG*
P Aliivibrio salmonicida (CAQ77725)     357-GYVKVDPYECILCESRLVFTNFRVGNSVNDLVTHAIVQSELRAA*
F Lactobacillus casei (YP_806573)       394-RHYMLEVNQNIAKEAYQTKYKQAEAAAYDRCRFSWERQRRIYLSEMPQA*
F Clostridium hiranonis (ZP_03294328)   154-IWHYKYGLIYNVLDKSNYKRIIYEEIIEKEISLNTTTQKELF*
M Kuenenia stuttgartiensis (Q1Q731)     249-IYNEIEEIMRGKYEPKEEKVIKPEGDGGTIRPTPRRVQIPLFSM*
```

IS200

P *Escherichia coli* (ZP_03029803)	125-WSWGYFVDTVGV--NEEIIRRYARYQEKMEQTHEQ**Q**ME**LL**E*
P *Vibrio cholera* (AAC01554)	106-WARGYFVDTVGV--NEEIIRRYVRHQDKKELEPEQ**QL**E**LL**RD*
P *Shewanella baltica* (YP_001364938)	106-WARGYFVDTVGV--NEEIIRRYVRHQDKQDQEHEA**QLSL**QMM*
P *Shewanella woodyi* (YP_001761123)	107-WQRGYFVDTVGI--NEEVIRRYVKHQEKVEKQEQP**QL**D**L**K*
P *Serratia odorifera* (ZP_06639291)	138-WARGYCVDTVGI--NEEMIRKYVKYQEKHEVE-ES**QL**P**L**KEV*
P *Nitrosomonas* sp. (YP_004295562)	107-WTDGYYVATVGERADWGEVERYVKNQGKPKEE-LR**QL**E**FF***
P *Escherichia coli* (NP_752024)	123-WCRGYYVDTVGK--NTAKIQDYIKHQLEEDKM-GE**QLSI**PYPGSPFTGRK*
P *Yersinia pestis* (YP_005509676)	123-WCRGYYVDTVGK--NTARIQEYIKHQLEEDKM-GE**QLSI**PYPVSPFTGRK*
F *Streptococcus pneumoniae* (ZP_01829269)	107-WCRGYYVDTVGR--NQKVIAEYIQNQLQEDRV-AD**QLTLF**ESVDPFT-150
F *Clostridium botulinum* (YP_002650746)	105-WCRGYYVDTVGR--NKKAITEYIKNQQKEDMI-SD**QISL**KEYMDPFK-148
F *Bacillus halodurans* (BAD18222)	106-WCRGFYVDTVGR--NKKQIQEYIRNQLREDYM-GD**QLTLF**EEYDPFT-149
F *Enterococcus faecium* (ZP_05832064)	105-WCRGYYVDTVGK--NAKKIEEYIANQLQEDLE-YD**Q**M**TL**KEYIDPFT-148
F *Ruminococcus albus* (YP_004103952)	108-WCRGYYVDTVGK--NKKKIAEYIRNQLQEDIV-CD**QISLF**ETVDPFT-151
F *Staphylococcus epidermidis* (YP_188749)	107-WCKGFYVDTVGR--NKKVIENYIRNQLQEDIV-AD**QIS**MEEYLDPFT-150
F *Filifactor alocis* (YP_005053889)	110-WCKGYYVDTVGK--NTKAIQEYISNQLKVDRE-SD**QLSIF**DPRDPFTGSK*
U *Fusobacterium nucleatum* (ZP_00143858)	105-WCRGYYVDTVGR--NKERIAQYIKNQIEEDKI-MD**Q**M**TL**KEYFDPFN-148
P *Xenorhabdus nematophila* (YP_003713015)	106-WTDGYFASTVGKHGDEQMIGRYVQNQGKKYHK-LHSDH**QL**A**LF***
P *Psychromonas ingrahamii* (YP_944951)	107-WTQSVFVETIGNATEE-VIRKYVQNQLIELDRKEINSD**QL**D**LF***
C *Microcystis aeruginosa* (CAO87910)	106-WSDGYYASTVGKHGDEGMIARYVKEQDKEYLQ-LHQNL**QLSLF***
C *Acaryochloris marina* (YP_001519691)	104-WSRGYFVSTVGR--DEEVIRRYIRHQEQEEQK-LE---**QL**N**LF**RA*
B *Pelodictyon phaeoclathratiforme* (YP_002018173)	106-WTDGYYAGTVGKHGNEDMIGKYVKGQGGTYQK-RYSDY**QLSLF***
(Arch.) *Methanosarcina barkeri* (YP_307176)	106-WSDGGYIGTVGDGTTSDVIKSYIENQGNQEEKEAYK**Q**MK**I**ID**F**Q*
(Arch.) *Methanosarcina mazei* (NP_632811)	106-WSSGKFYRSVGNVTAD-TIKHYIKESQKKPKTEVKSSKSAKPD**Q**RI**I**DD**F***
(Arch.) *Methanosaeta concilii* (YP_004384718)	106-WSSGKFFRSVGNVTADTIQ-HYIKESQGKPKAESKVCRSRESG**Q**RR**L**DD**F***
(Arch.) *Sulfolobus islandicus* (YP_002831273)	108-WSPSYFLATSG**Q**VT**L**EV**L**K-KYVESQGKE*
(Arch.) *Halobacterium salinarum* (YP_001688288)	106-WQPGYFLATTG**Q**VS**I**DT**L**M-DYVDDQ*
(Arch.) *Haloarcula marismortui* (YP_134246)	108-WNDSYCLISTG**Q**VS**L**DV**L**K-QYVEDQRE*

IS200/IS605 (OrfB) Cyanobacteria

Cyanothece sp. (YP_002371738)	367-DLNHGLLAGTAPNFMNTQKERIGEYI**QLSLF**DPTLFGG*
Microcystis aeruginosa (YP_001656728)	371-IENRGKNAVGLTVLENACGGDLTGVV**QL**N**LF**DLVKSLRTKN-411
Arthrospira platensis (BAI93705)	374-NGCGERVRLSVKKAHLNEASTRPAFQ**QLSIF**DLLK*
Synechococcus sp. (ZP_05040300)	360-ASVMGVVASADGVFDNPLNSMNQSAQ**QL**T**LF**PMSA*
Synechocystis sp. (NP_441190)	368-KNGRGGKRQTTSVAASGEASTHRKAI**QL**T**LF**AS*
Trichodesmium erythraeum (YP_721194)	404-ALSKGKTCRKKPANCDEMLTRFESFK**QL**N**LF**D*
Microcoleus chthonoplastes (ZP_05026728)	410-ALVGLSINQPGGTGLSCKLSRTIKYV**QLSLF**DDFRATKNPDLSS*
Crocosphaera watsonii (ZP_00513803)	435-TPKPELTGSSHRETSVSLEIEPGNPQ**QLSLF**EWVNGEVIPC*
Acaryochloris marina (YP_001521276)	266-NGCRRECKSEVNSAVLSDASTRLVDK**QL**E**LF**AS*

IS1380

P	*Escherichia coli* (YP_003829282)	271-WEKDRRFVVSRVLKPEKE---RA**QLSLL**EGS----EYDYFFFVTN-TTLLSEKV-316
P	Gamma proteobacterium (ZP_05061507)	293-WSRSRRFIAVRRLAKVKK-EGPQ**Q**S**LII**EPV-----YDYFCYVTT-ERLTPWQA-339
P	*Azoarcus* sp (YP_195484)	286-WKKAKRTLRLVVRVTERTIDKKG**Q**H**LL**APEIEIEGWWTSLD-------VPMADV-332
F	*Bacillus coagulans* (YP_004859304)	280-WEKPRRVAVIRKADKYEE----D**QL**Q**LF**DF-----LWDYEAIVTT-MDWEPMDI-323
F	*Enterococcus casseliflavus* (AAX38177)	270-WEKDRRFVVSRVLKPEKE---RA**QISLL**EG----SEYEYFFFVTN-TTLLSEKV-315
F	*Desulfotomaculum* sp. (YP_001111807)	316-WSKARRFVFIRETQEPKV--SGE**QL**N**F**-DL----KTFDYQVIITSSDEYNPEEV-362
F	*Geobacillus* sp. (YP_002949920)	289-IDGNTYTYVQVTQVTERTMERNG**QL**M**LV**PDYEVES-YWVRLKGY--EHVRMSDV-339
B	*Tannerella forsythia* (YP_005013031)	278-WQKPRRIVIVRQKIEKRPQAGGK**QLSLF**PEDEIHRNYRYSAYFTN-QTCSMVDV-330
B	*Prevotella buccae* (ZP_07884135)	291-PRRIVMVRQEVEKRPKAAGKQVR**QL**E**LF**EDEQDFGKYRYSCFVTN-LALPAKIV-343
A	*Rhodococcus jostii* (YP_707285)	314-HQIPGR-LVVRRIPDLRPPKDQG**Q**GT**LF**DI----WRFHAFFTTTDPDDLDTVDA-362
A	*Mycobacterium gilvum* (YP_001133130)	278-RGWPAGMRVIARR--ERPHPGA-**QL**R**L**TDDNG--WRITCFATNTP--GWSIADL-323

Tn7 (TnsC)

P	*Escherichia coli* (EIL57895)	491-VKRTEWGCLSEEDLRRRYADGPESEAYDRFK**Q**AG**LIL**DLRPLLRAS*
P	*Escherichia coli* (YP_006903369)	438-LQTADDLFERIGNPDLMQHAAAEEKPLELLR**Q**EG**VF**CEHPLKVFA*
P	*Acinetobacter haemolyticus* (ZP_06726085)	508-KQWNELADTDLRHIFSMNPTSSESIYTQLDR**Q**G**LIL**QMSDLLS*
P	*Acidithiobacillus caldus* (YP_004750330)	505-KKEKWHALDSHDLRFLFSQASSEEDVYKSLG**Q**KG**LFF**DTESWVDSYG*
P	*Vibrio cholerae* (EGS60813)	509-PNSWDELPSDDLRFIKSQCKNSTDVHTALNK**Q**GM**LL**DLSSVLNLAG*
P	*Desulfovibrio aespoeensis* (YP_004122276)	500-IPRKDWHSLGSEDLRFVFSQTDSKDMYEHLK**Q**N**S**M**IF**DMEDWLRKPG*
F	*Clostridium thermocellum* (YP_001037540)	511-MDENDIRYIVEEGRKNKKSAYESLNEKGLIK**QV**EKD**F**FKAV*
F	*Bacillus cellulosilyticus* (YP_004093238)	487-LNNNDKSTEKVISNDLRTIIKKGKEQQVTAY**Q**AL**L**DAGVIKQDYPERVVS*
C	*Crinalium epipsammum* (YP_007140767)	482-RKITNKPQLITRETANNATTLIEIVTIGTAH**QIS**G**Y**EALKLAGLIQPTTEY*

ISL3

P	*Escherichia coli* (ZP_07122173)	351-IIRGEENYAFRFISLMCEKEPELKIAQ**QLVL**EFYRILKT*
P	*Methylobacterium extorquens* (YP_003065656)	478-GPVEGQVNRLKLIKRSMYGRAGFDLLR**Q**R**VL**HAA*
P	*Shewanella* sp. (YP_733579)	364-NGLAEAINSQVRQLRVKAMGYRNSERF**Q**RA**IL**FHFGKLDMGFHQ*
P	*Syntrophus aciditrophicus* (YP_461433)	276-LKPDHKERLDALMQANQPLFVMHSMKE**QL**R**LF***
A	*Mycobacterium gilvum* (YP_001131597)	398-RLINARGYGHHTAETLTSMIYLCLGGL**Q**VK**L**PTRT*
F	*Lactobacillus brevis* (ZP_03940084)	386-EGVNRKIKAIGRTAYGYRNFDHYRTRI**QM**E**LF**A*
F	*Streptococcus mitis* (YP_003445409)	408-FRNYYHFRNRIILITKMFGPKQKGIKQ**QLVV***
F	*Staphylococcus haemolyticus* (YP_252293)	407-YRNYDNLRNRIIITSRLFASTTKKEIK**Q**PK**V**A*
H	*Petrotoga mobilis* (YP_001567687)	342-GLAEGSVNKLKVVKRIMYGRNKFEMLR**Q**K**VLF**LENNG*

ISNCYa

P	*Escherichia coli* (EIL58166)	261-RQEERQEFALRLLSKGMSREDVAEMANLPLAEID**K**VIN**L**I*
P	*Photorhabdus luminescens* (Q7N2E9)	284-EEGRAEGKLETARALLRHGVSLDIIVTSTGLSRD**K**IEM**L**KH*
P	*Proteus mirabilis* (YP_002152329)	267-EKGIGLGVEKVARSLLQQGVDLNIIMQCTGLTRE**K**IES**L**KH*
F	*Selenomonas sputigena* (ZP_05899258)	253-VTEIVLNMLKKKKWSLQDISEVSQWPLDKIESLG**K**MHQ**L**L*
F	*Thermoanaerobacter italicus* (YP_003476143)	251-GKKEGKIAVAKKLLEKGMDIDEIAEITELSKEEI**K**KL**L**N*
F	*Clostridium botulinum* (ZP_04824118)	294-KKAIEDAIGFLRLGVSEEIVSKGTGLPIEKVREL**K**DK**I**NN*
F	*Caldicellulosiruptor owensensis* (YP_004003521)	256-ATIEIAKRLIQKGFSDEEVAELTELQIEKVKELR**K**SM**L**N*
S	*Leptospira interrogans* (Q8F556)	265-GKVETKIDIARNMLLDGASLEYVLKITRLTEQEL**K**DHG**L**L*

TABLE S6. Statistical significance for the non-random orientation of IS elements in the Phyla Bacterioidetes, Cyanobacteria and Spirochaeta. See legend to Table 1 for details.

	Bacteroidetes			**Cyanobacteria**			**Spirochaeta**		
	Chr.	***IS***	***Orient.***	***Chr.***	***IS***	***Orient.***	***Chr.***	***IS***	***Orient.***
IS1	7	29	0.426	7	49	0.283	-	-	-
IS3	28	171	0.333	-	-	-	13	50	0.894
IS4b	11	48	0.206	-	-	-	-	-	-
IS4c	-	-	-	4	26	0.254	-	-	-
IS5a	16	112	0.775	-	-	-	5	28	0.656
IS5b	10	68	0.630	13	142	0.470	-	-	-
IS5c	-	-	-	2	22	0.312	-	-	-
IS5d	5	62	0.199	11	154	0.556	-	-	-
IS21	14	62	0.393	-	-	-	-	-	-
IS30	6	23	0.812	-	-	-	6	24	0.869
IS66a	9	30	0.513	-	-	-	-	-	-
IS66b	2	21	0.263	2	22	0.663	-	-	-
IS110	38	151	0.119	6	26	0.398	8	116	0.577
IS200	36	132	0.546	14	59	0.825	-	-	-
IS200/IS605	-	-	-	11	45	0.459	-	-	-
IS256	17	68	0.265	-	-	-	7	66	0.0603
IS607	-	-	-	8	39	0.833	-	-	-
IS630	7	71	0.533	16	392	0.884	-	-	-
IS701	3	11	0.551	11	85	0.855	-	-	-
IS982	17	104	0.676	-	-	-	-	-	-
IS1182	20	78	0.926	3	23	0.583	-	-	-
IS1380	-	-	-	4	22	0.833	-	-	-
IS1595	16	88	0.263	-	-	-	6	24	0.124
IS1634	-	-	-	5	46	0.518	-	-	-
ISAs1	9	28	0.147	6	34	0.366	-	-	-
ISAzo13	-	-		5	29	0.359	-	-	-
ISL3	23	81	0.764	8	73	0.775	-	-	-
ISNCYa	-	-	-	19	102	0.247	8	21	0.572