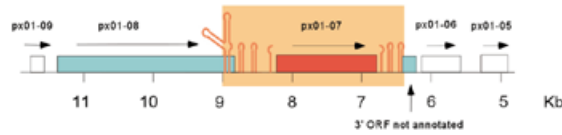


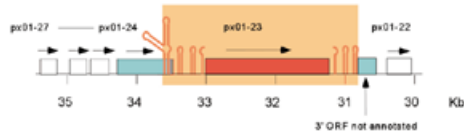
**SUPPLEMENTARY MATERIAL**

**I. Insertion pattern 1: Introns that interrupt an ORF**

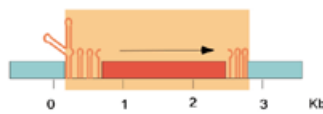
*Bacillus anthracis* (B.a.11)  
Accession: AF065404  
(6445-8975)  
NC\_001496  
(6445-8975)



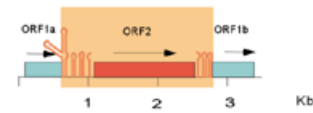
*Bacillus anthracis* (B.a.12)  
Accession: AF065404  
(30896-33785)  
NC\_001496  
(30896-33785)



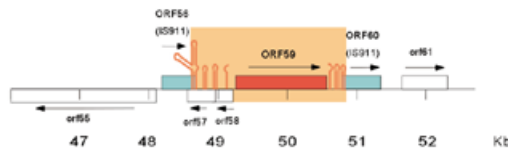
*Clostridium difficile* (C.d.11)  
Accession: X98906  
(13-2958)



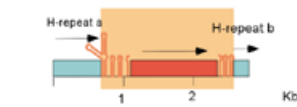
*Calothrix* sp. (C.sp.11)  
Accession: X71404  
(446-2898)



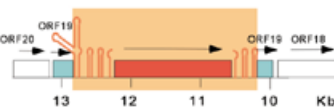
*Escherichia coli* (E.c.14)  
Accession: AB024946  
(48555-50824)



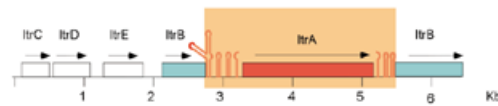
*Escherichia coli* (E.c.12)  
Accession: X77508  
(518-2408)



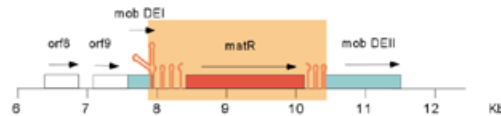
*Enterococcus faecalis* (E.f.11)  
Accession: AC091242  
(10115-12672)



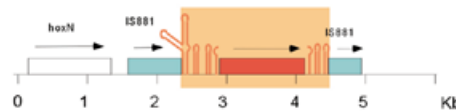
*Lactococcus lactis* (L.l.11)  
Accession: U50902 (2854-5345)  
AF209910 (7350-9841)  
X89922 (985-3478)



*Lactococcus lactis* (L.l.11)  
Accession: AF243383  
(7831-10322)



*Ralstonia eutropha* (R.e.11)  
Accession: AF261712  
(2355-4192)



**Notes**

Genbank annotation marks only ORFs and not the intron boundaries

Frame shift in domain X. There is a second conserved reading frame in the 3' exon that would be out of frame with pX01-24 ORF in the spliced exons. In *Aeropyrum pernix* the homolog to this alternative 3' ORF is directly upstream and overlapping the homolog of pX01-24, but with a frame shift between the two ORFs

Interrupted ORF is homologous to orf4 in Tn916

Interrupts a H-repeat contained in an Rhs element

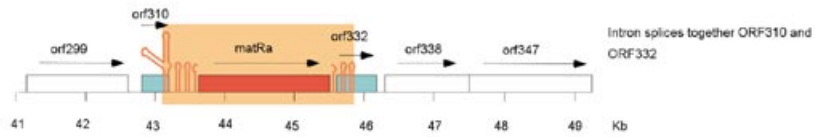
Flanking sequence without intron found in Genbank Acc# U09422

Intron is inserted into ItrB (relaxase). 99% identical to intron in AF243383

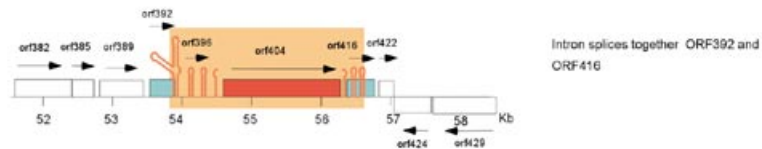
Intron is inserted into mobilisation protein A. *Lactococcus lactis* self-transfer sex factor. 99% identical to intron in U50902, AF209910 and X89922

Frame shift in RT domain 6

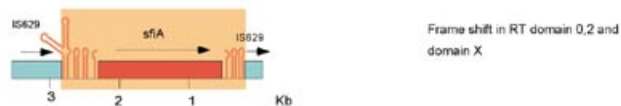
*Novosphingobium aromaticivorans* (N.a.I1)  
 Accession: AF079317  
 (43084-45662)  
 NC\_002033  
 (43084-45662)



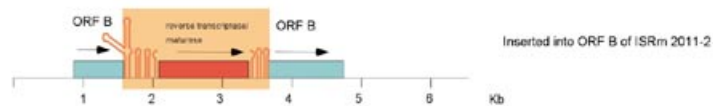
*Novosphingobium aromaticivorans* (N.a.I2)  
 Accession: AF079317  
 (53812-56360)  
 NC\_002033  
 (53812-56360)



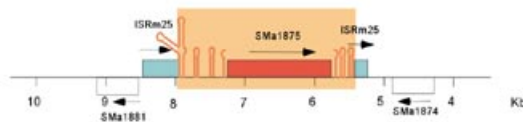
*Shigella flexneri* (S.f.I1)  
 Accession: U97469  
 (516-2787)  
 AF200662  
 (21025-23296)



*Sinorhizobium meliloti* (S.me.I1)  
 Accession: Y11597 (1-1864)  
 AL603647 (80617-82500)  
 AL603644 (105027-106910)  
 AE007285 (6583-8466)

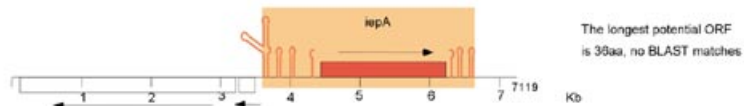


*Sinorhizobium meliloti* (S.me.I2)  
 Accession: AE007289  
 (5487-7696)

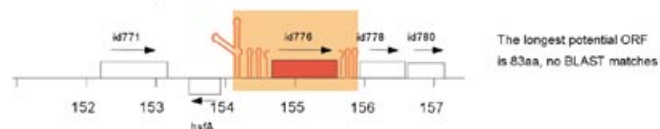


## II. Insertion pattern 2: Introns not inserted into an ORF

*Bacillus megaterium* (B.me.I1)  
 Accession: AB022306  
 (3853-6572)



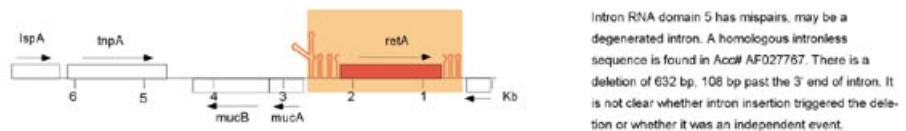
*Bradyrhizobium japonicum* (B.j.I1)  
 Accession: AF322013  
 (154084-155888)



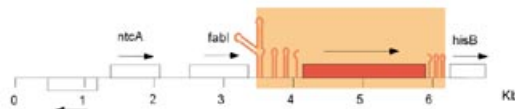
*Pseudomonas putida* (P.p.I2)  
 Accession: Y18996  
 (37-2242)



*Serratia marcescens* (S.me.I1)  
 Accession: AF027768  
 (658-2565)



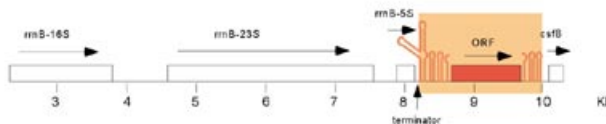
*Trichodesmium sp. IMS101 (T.sp.11)*  
 Accession: AF302392  
 (3712-6232)



The the longest potential ORF is 51aa, no BLAST matches. Frame shift in RT domain 4 and Zn domain.

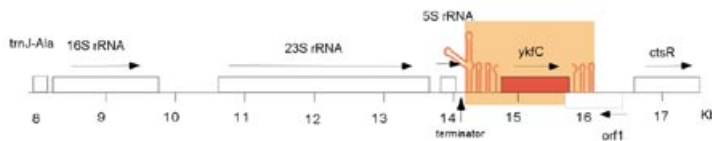
**III. Insertion pattern 3: Introns inserted directly after a terminator**

*Bacillus halodurans (B.h.11)*  
 Accession: AB031210  
 (8118-10000)  
 AP001507 (56387-58209)  
 NC\_002570 (56387-56289)



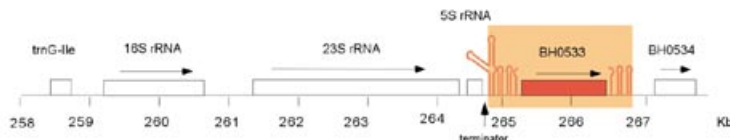
rrn operon B

*Bacillus halodurans (B.h.11)*  
 Accession: AB031211  
 (14079-15061)  
 AP001507 (130149-132031)  
 NC\_002570 (130149-132031)



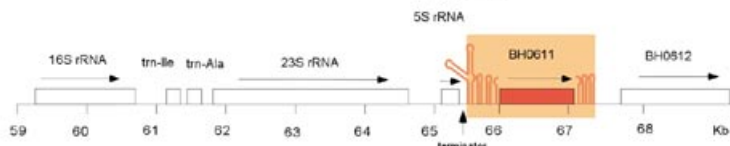
rrn operon D

*Bacillus halodurans (B.h.11)*  
 Accession: AP001508  
 (264921-266803)  
 AB031213 (8838-10716)  
 NC\_002570 (565021-569903)



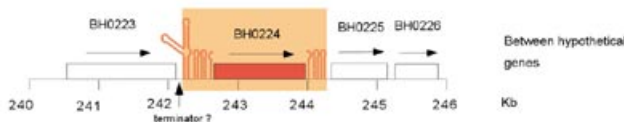
rrn operon F

*Bacillus halodurans (B.h.11)*  
 Accession: AP001509  
 (65481-67363)  
 AB031214 (8480-10362)  
 NC\_002570 (862481-684363)



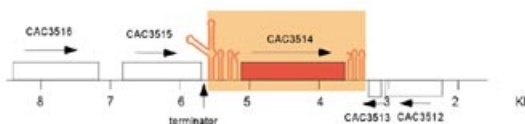
rrn operon G

*Bacillus halodurans (B.h.11)*  
 Accession: AP001507  
 (242165-244047)  
 NC\_002570 (242165-244047)

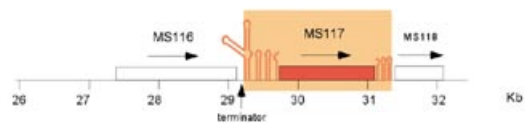


Between hypothetical genes

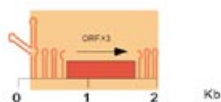
*Clostridium acetobutylicum (C.a.11)*  
 Accession: AE007848  
 (3618-5538)



*Microscilla sp. PRE1 (M.sp.11)*  
 Accession: AF339846  
 (26388-31287)  
 NC\_002808  
 (26338-31287)



*Pseudomonas alcaligenes (P.a.12)*  
 Accession: AF323437  
 (1-1924)



Frame shift in domain X

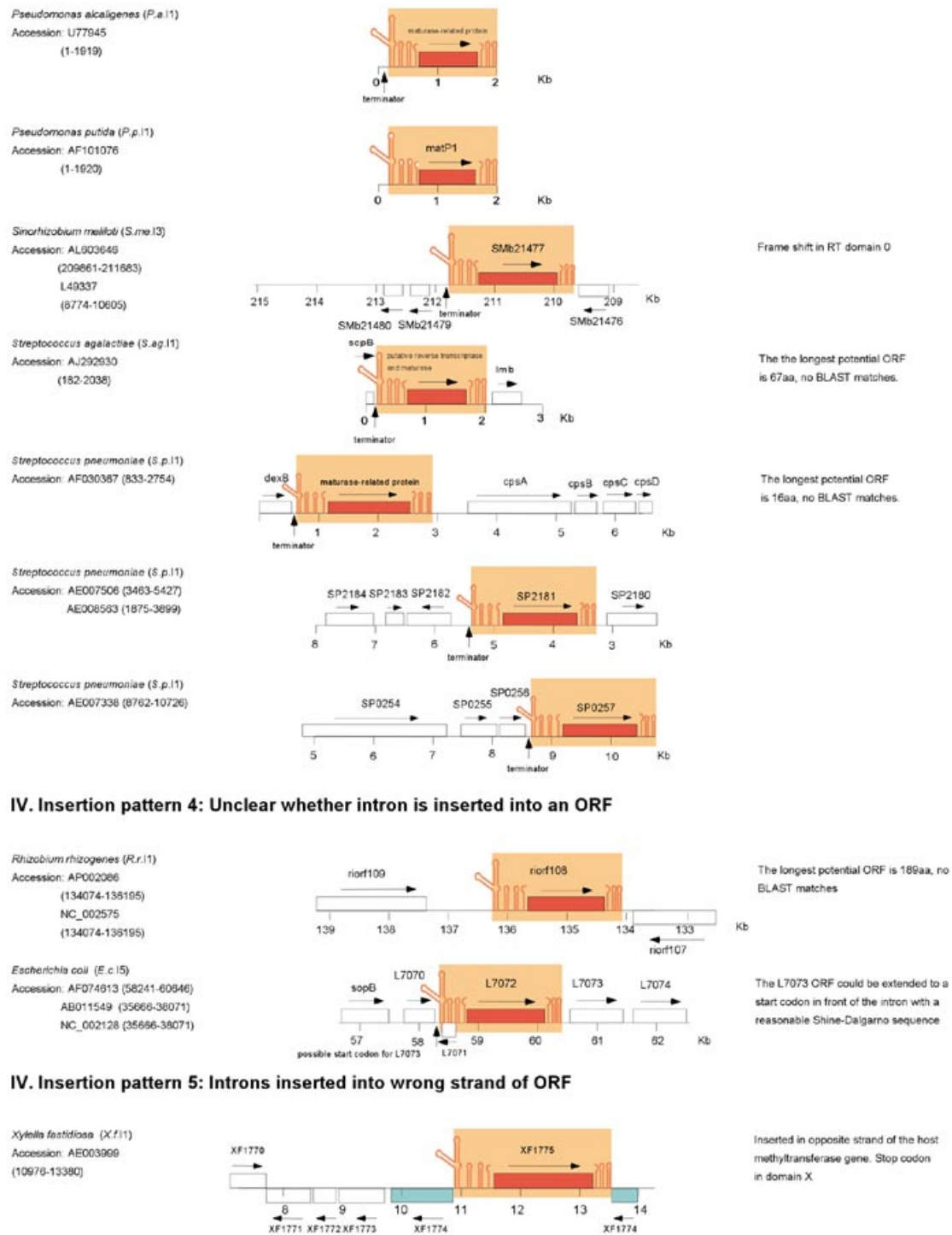


Figure S1. Group II introns that interrupt an ORF.