Supplementary table 1: Putative iron uptake systems

Genes	Characteristics
DIP0108-DIP0111	Putative DtxR-regulated operon siderophore-dependent iron uptake system. Irp6A.
DIP0282 -DIP0284	ABC transporter
DIP0582-DIP0594	ABC transporter
DIP0613-DIP0619	ABC transporter; DIP0619 is 35% identical to DtxR.
DIP0625-DIP0629	Hmu (hemin) gene cluster; ABC transporter. Present in *C. glutamicum*
DIP1059-DIP1063	Iron-siderophore uptake system. ABC transporter. Present in *C. glutamicum*
DIP1082-DIP1087	ABC transporter

Supplementary table 2: Putative sortases

CDS	Length (aa)	Database similarities
DIP0233*	329	*Actinomyces naeslundii* putative fimbrial associated protein (EMBL:U85709) *Staphylococcus aureus* sortase StrA (EMBL:AF162687)
DIP0236*	312	*A. naeslundii* putative fimbria associated protein (EMBL:AF019629) *S. aureus* sortase StrA (EMBL:AF162687)
DIP2012*	289	*Actinomyces viscosus* sortase-like protein (EMBL:AF106034) *Streptococcus pneumoniae* putative sortase (EMBL:AE007358)
DIP2224*	348	*A. naeslundii* putative fimbria-associated protein (EMBL: AF019629) *S. aureus* sortase StrA (EMBL:AF162687)
DIP2225*	318	*A. naeslundii* putative fimbria-associated protein (EMBL: AF019629) *S. aureus* sortase StrA (EMBL:AF162687)
DIP2272	273	*Clostridium acetobutylicum* sortase CAC0204 (EMBL: AE007533) *A. naeslundii* putative fimbria-associated protein (EMBL: AF019629) *S. pneumoniae* sortase, putative SP0466 (EMBL:AE007358)

* Situated within anomalous G+C content/GC skew regions

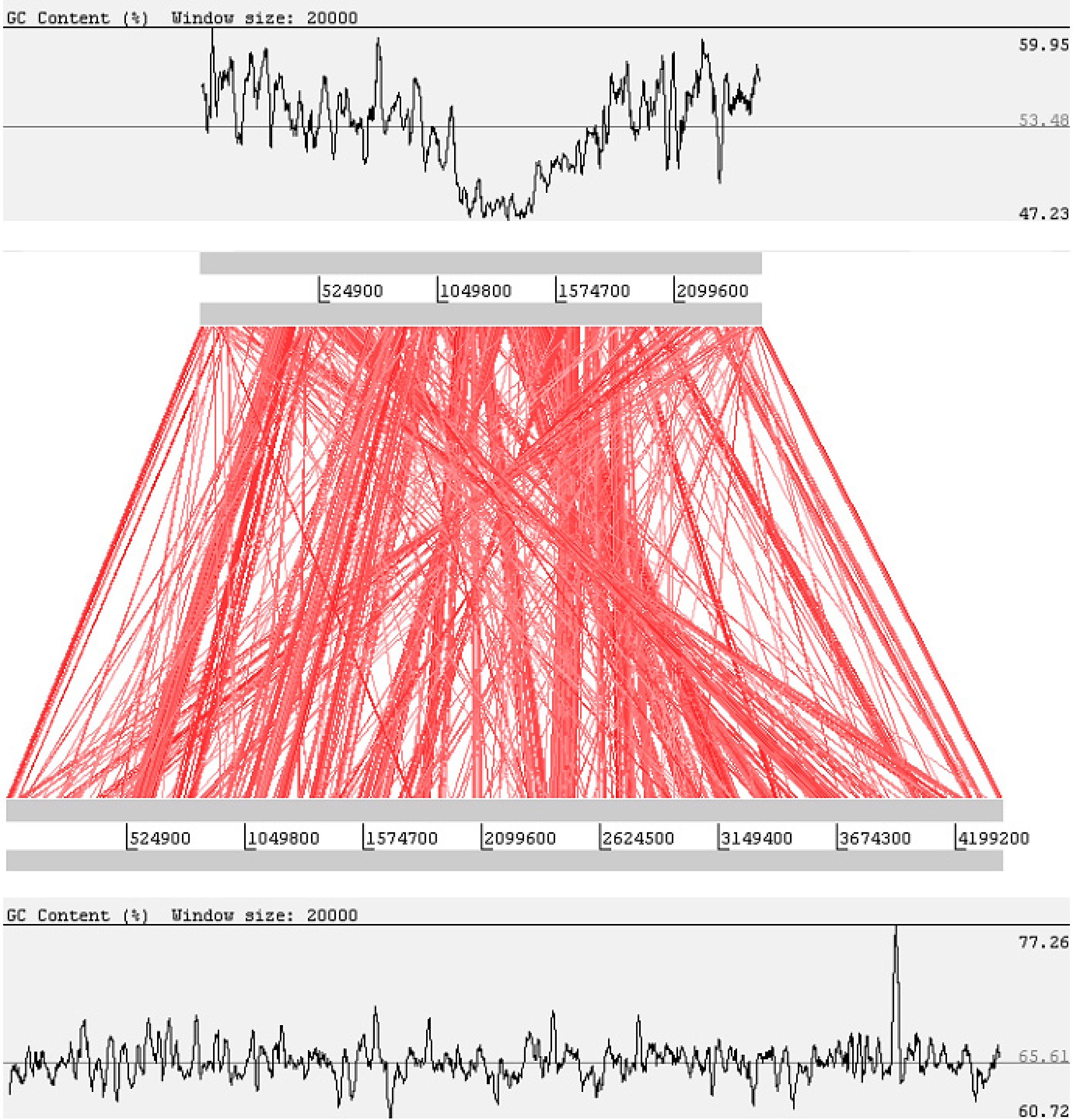
Supplementary table 3: CDSs containing potential sortase recognition motifs

CDS	CDS length (aa)	Potential sortase recognition motif	C-terminal Signal peptide length (aa)	Database similarities
DIP0235*	490	LTMPG	23	*Actinomyces viscosus* type-1 fimbrial major subunit precursor FimP (EMBL:AF10634)
DIP0237*	263	LALTG	23	No significant database matches
DIP0238*	1039	LPKTG	23	Very weak similarity to Actinomyces fimbrial proteins
DIP0239*	149	LPSTG	ND	Pseudogene
DIP0278	1080	LARTG	18	Low similarity to *S. aureus* Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding protein SdrE (EMBL:AP003359)
DIP0443*	284	LPDTG	23	No significant database matches
DIP1724	1075	LANTG	22	No significant database matches. Contains also a putative TAT recognition motif (RRVFAS)
DIP2010*	1872	LPLTG	22	*A. viscosus* putative usher-like protein precursor (EMBL:AF106034) *A. naeslundii* putative fimbrial associated protein (EMBL:U85708)
DIP2011*	205	LAFTG	22	No significant database matches
DIP2013*	525	LPLTG	19	*A. naeslundii* fimbrial subunit type 2 precursor (SW:FM2_ACTNA) and type-1 fimbrial major subunit precursor FimP (EMBL:AF107019)
DIP2062	623	LAATG	22	*Streptococcus agalactiae* BPS protein precursor Sar5 (EMBL:AJ133114) *Staphylococcus epidermis* putative cell-surface adhesin SdrF (EMBL:AF245041)

DIP2066*	1080	LPKTG	23	Low similarity to *A. naeslundii* putative fimbrial associated protein (EMBL:U85708)
DIP2093*	951	LAATG	23	Low similarity to *S. aureus* Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding protein SdrC (EMBL:AP003131)
DIP2223*	269	LGNTG	22	No significant database matches
DIP2226*	555	LPLTG	23	*A. naeslundii* fimbrial subunit type 2 precursor (SW:FM2_ACTNA) *A. viscosus* fimbrial subunit type 1 precursor (SW:FM1_ACTVI)
DIP2227*	1375	LPLTG	23	*A. naeslundii* hypothetical protein (EMBL:AJ401093)
DIP2325	879	LAETG	22	No significant database matches
DIP2370	904	LLSTG	23	Low similarity to *M. tuberculosis* hypothetical Rv3909 (EMBL:Z94121)

* Situated within anomalous G+C content/GC skew regions

Supplementary Figure 1. ACT comparison showing location of orthologous genes between *Mycobacterium tuberculosis* and *Corynebacterium diphtheriae*. The top panel shows the G+C content of *C. diphtheriae*, plotted over a 20kb window with the base-pair numbers below. The central panel shows the positions of pairs of orthologous genes between the genomes (as calculated by reciprocal best FASTA matches). The bottom panel shows the G+C content of *M. tuberculosis*, plotted over a 20kb window, with the base-pair numbers above.


GC Content (%) Window size: 20000
59.95
53.48
47.23
524900
1049800
1574700
2099600
524900
1049800
1574700
2099600
2624500
3149400
3674300
4199200
GC Content (%) Window size: 20000
77.26
65.61
60.72