

Supplementary Materials

Figures:

Figure S1. Mycobacterial ESX loci.

Figure S2. Network of ESX loci in mycobacteria, *Nocardia*, and mycobacterial plasmids with tip labels.

Figure S3. Maximum likelihood phylogeny of ESX loci in mycobacteria, *Nocardia*, and mycobacterial plasmids.

Figure S4. ESX loci in *Nocardia*.

Figure S5. Phylogeny of ESX-5.

Figure S6. Episodic directional selection during the evolution of ESX loci with tip labels.

Figure S7. Episodic directional selection during the evolution of ESX loci (Gblocks trimmed).

Tables:

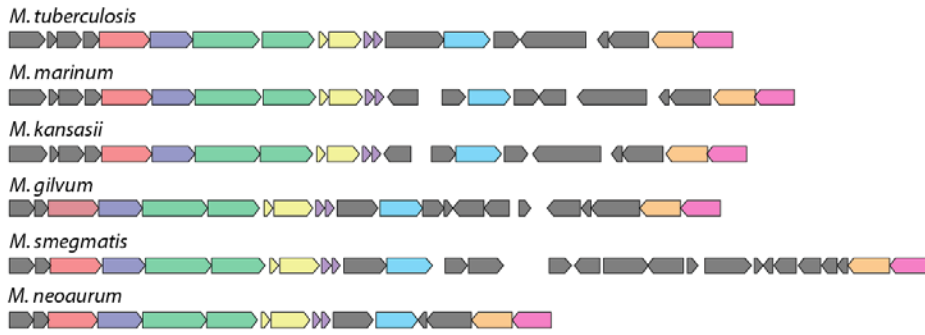
Table S1. Accession numbers for finished genomes

Table S2. Accession numbers for sequence read data used in plasmid assemblies.

Table S3. Accession numbers for finished plasmids.

Table S4. Shared plasmid gene content.

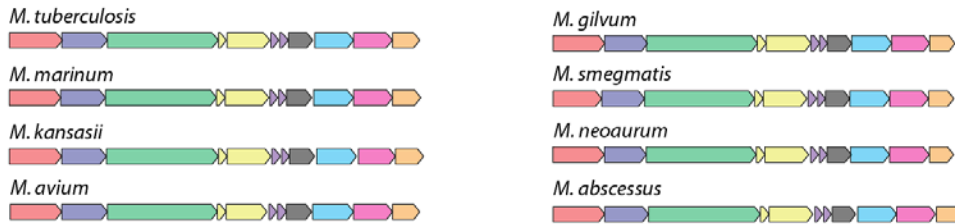
ESX-1



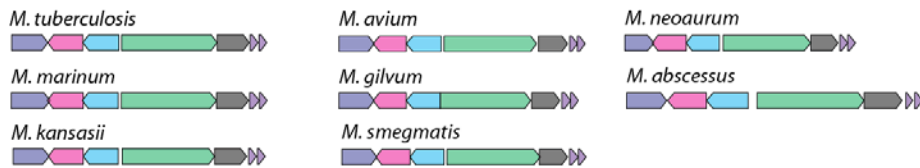
ESX-2



ESX-3



ESX-4



ESX-5

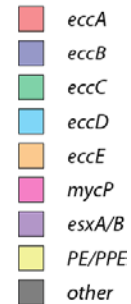
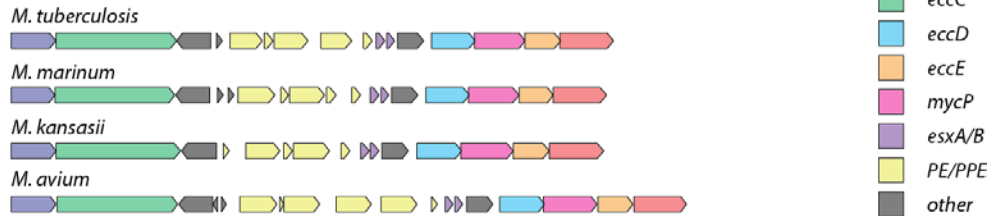


Figure S1. Mycobacterial ESX loci. Core gene content in the ESX loci are colored as follows: *eccA*- red, *eccB*- dark blue, *eccC*- green, *eccD*- light blue, *eccE*- orange, *mycP*- pink, *esxA/B*- purple, *PE/PPE*- yellow. Other variable genes in the loci are black. Orthologs and paralogs are based on OrthoMCL (Li et al. 2003) output. Locus diagrams were made using GenomeTools (Gremme et al. 2013).

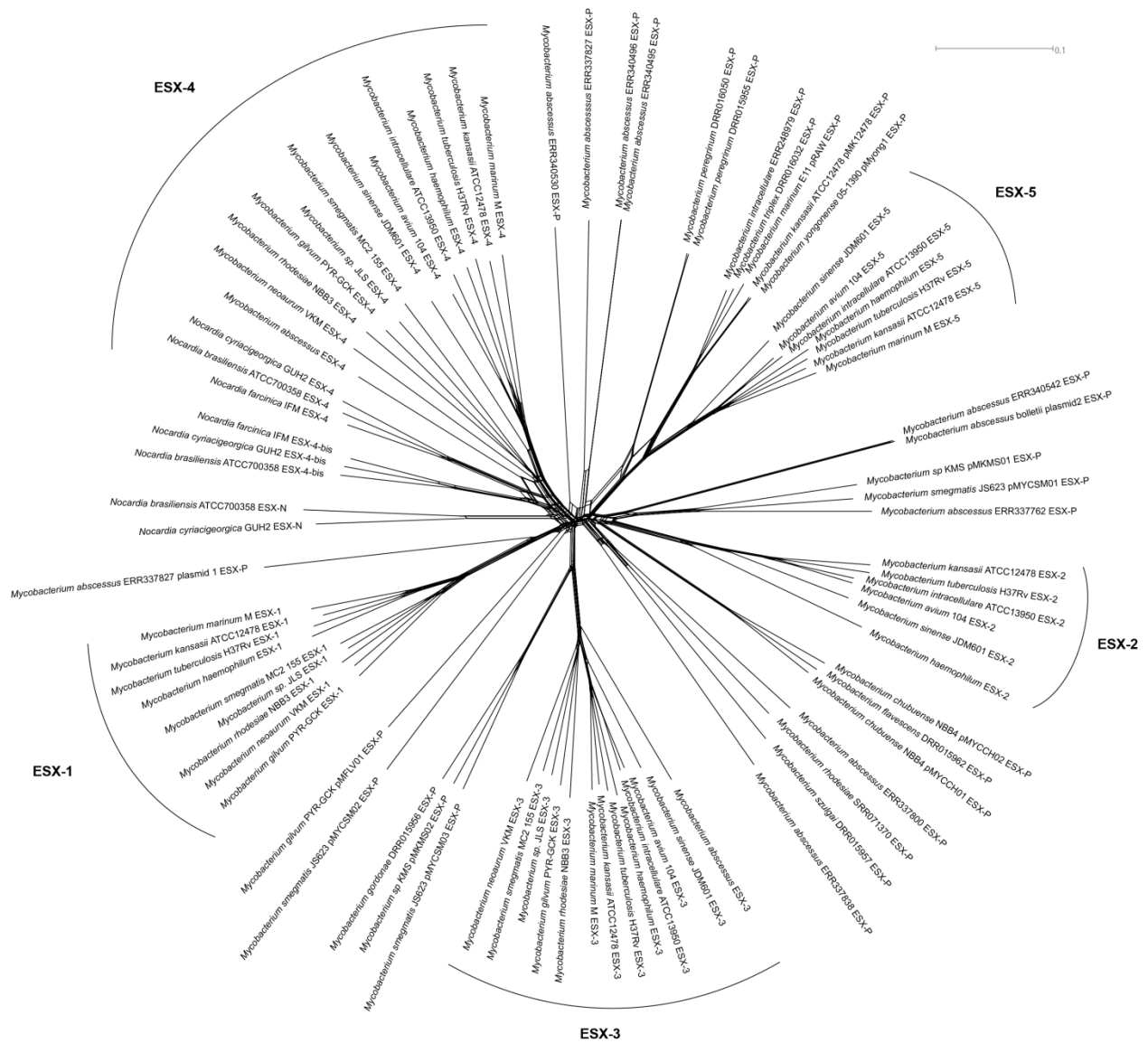


Figure S2. Network of ESX loci in mycobacteria, *Nocardia*, and mycobacterial plasmids. The network was created in SplitsTree4 (Huson and Bryant 2006) from a concatenated alignment of *eccA*, *eccB*, *eccC*, *eccD*, *eccE*, and *mycP*. The PHI test (Bruen et al. 2006) was insignificant for this alignment.

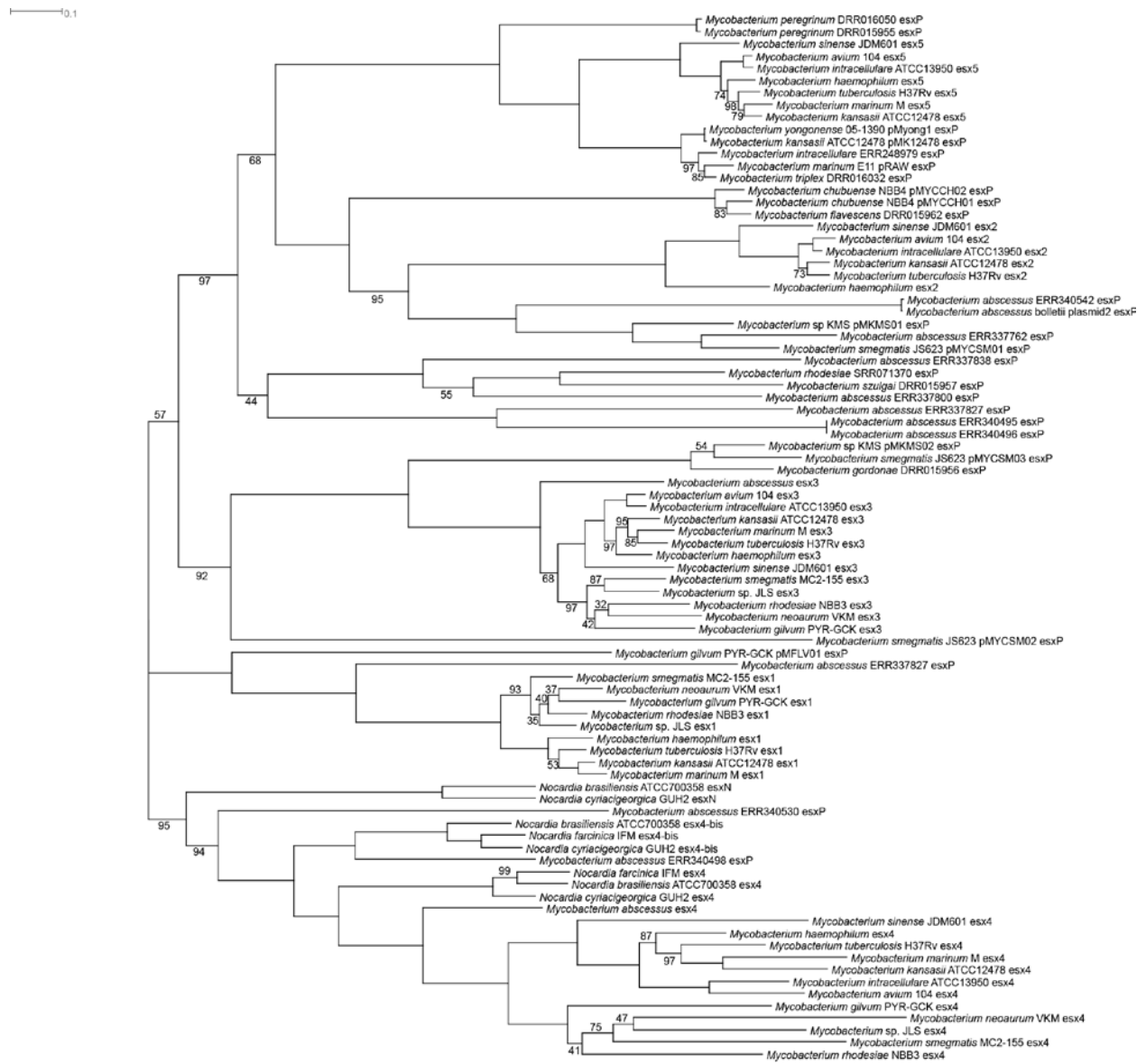


Figure S3. Maximum likelihood phylogeny of ESX loci in mycobacteria, *Nocardia*, and mycobacterial plasmids. The maximum likelihood phylogeny was created with RAxML (Stamatakis 2014) from a concatenated alignment of *eccA*, *eccB*, *eccC*, *eccD*, *eccE*, and *mycP* with poorly aligned columns removed with GUIDANCE (Sela et al. 2015). Branches without labels have a bootstrap value of 100.

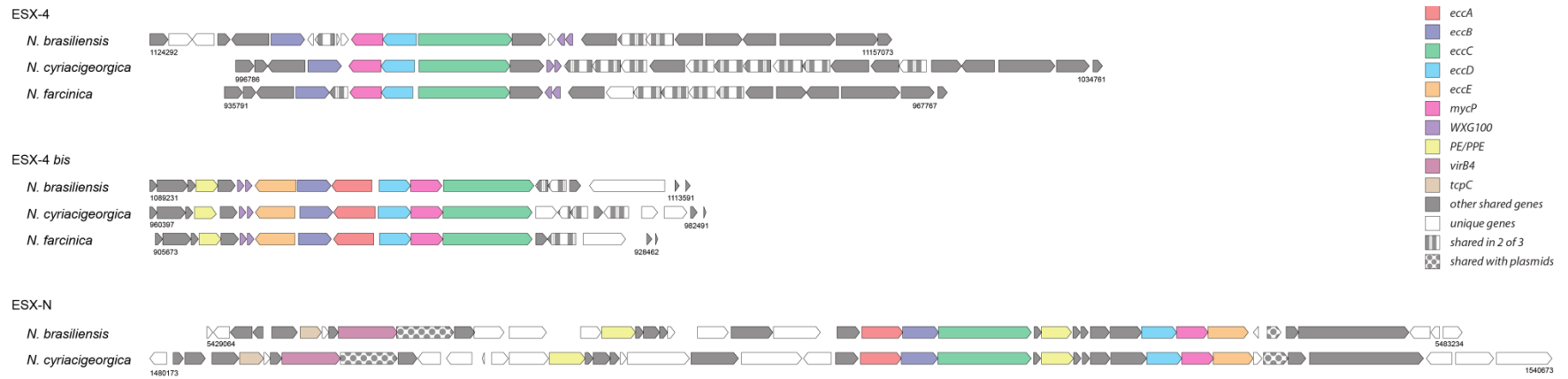


Figure S4. ESX loci in *Nocardia*. ESX genes are colored as follows: *eccA*- red, *eccB*- dark blue, *eccC*- green, *eccD*- light blue, *eccE*- orange, *mycP*- pink, *WXG100*- dark purple, *PE/PPE*- yellow. Type IV secretion system genes are colored as follows: *virB4*- light purple, *tcpC*- tan. Flanking genes shared between all strains are black, genes unique to the locus are white, genes found in two of three *Nocardia* species are striped, and genes found only in ESX-N and mycobacteria plasmids are dotted. Orthologs and paralogs are based on OrthoMCL (Li et al. 2003) output. Locus diagrams were made using GenomeTools (Gremme et al. 2013).

—|0.01

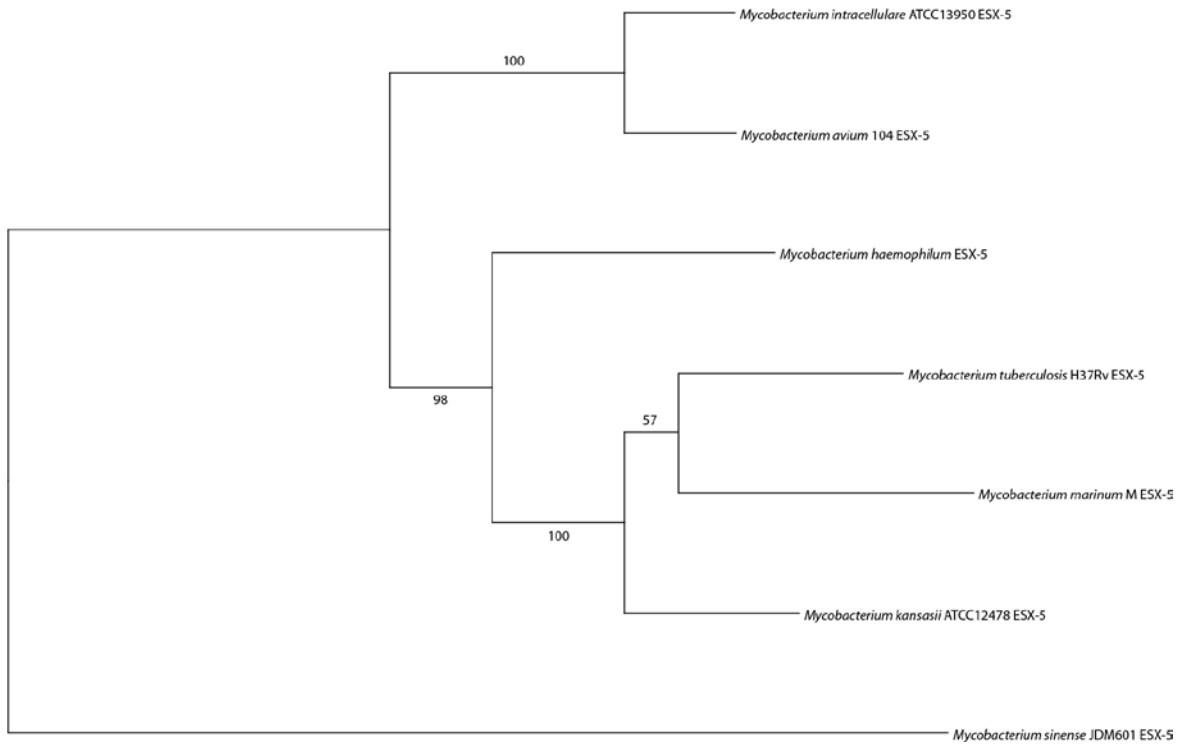


Figure S5. Phylogeny of ESX-5. The maximum likelihood phylogeny was created with RAXML (Stamatakis 2014) from a concatenated alignment of *eccA*, *eccB*, *eccC*, *eccD*, *eccE*, *mycP*, *cyp143*, and *Rv1794* with poorly aligned columns removed with GUIDANCE (Sela et al. 2015).

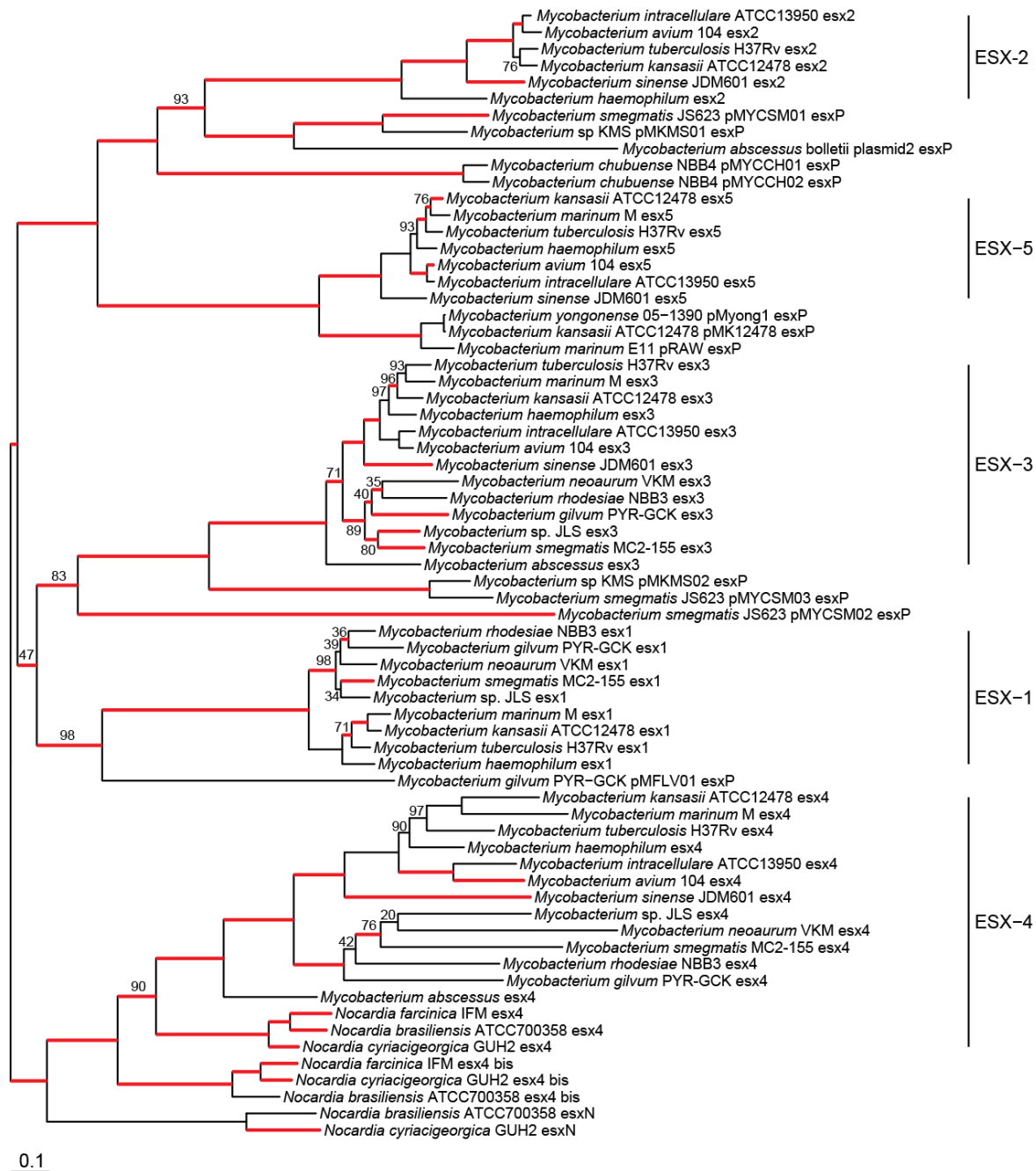


Figure S6. Episodic directional selection during the evolution of ESX loci. The maximum likelihood phylogeny was created with RAXML (Stamatakis 2014) from a concatenated alignment of *eccA*, *eccB*, *eccC*, *eccD*, *eccE*, and *mycP* with poorly aligned columns removed with GUIDANCE (Sela et al. 2015). Branches without labels have a bootstrap value of 100. We used the aBSREL test implemented in HyPhy (Smith et al. 2015) to identify branches with significant evidence of episodic directional selection. These branches are highlighted in red.

0.1

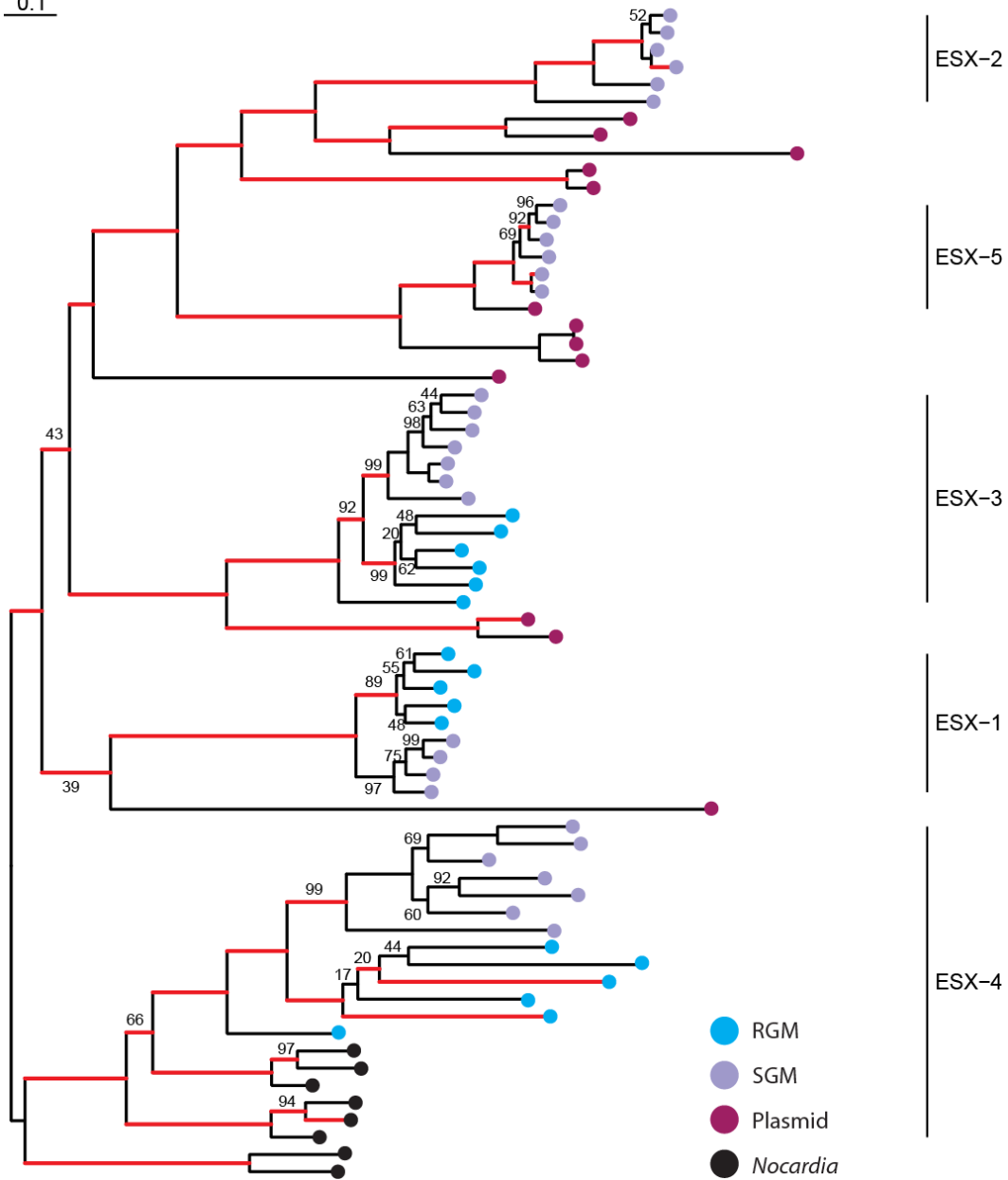


Figure S7. Episodic directional selection during the evolution of ESX loci (Gblocks trimmed). Light blue dots correspond to ESX loci from rapid growing mycobacterial chromosomes, light purple dots correspond to ESX loci from slow growing mycobacterial chromosomes, magenta dots correspond to ESX loci from mycobacterial plasmids, and black dots correspond to ESX loci from *Nocardia* chromosomes. All chromosomal ESX duplications (ESX-1-5) have basal plasmid loci, suggesting that the ancestral ESX locus was likely plasmid borne. The maximum likelihood phylogeny was created with RAxML (Stamatakis 2014) from a concatenated alignment of *eccA*, *eccB*, *eccC*, *eccD*, *eccE*, and *mycP* trimmed with Gblocks (Castresana 2000). Branches without labels have a bootstrap value of 100. We used the aBSREL test implemented in HyPhy (Smith et al. 2015) to identify branches with significant evidence of episodic directional selection. These branches are highlighted in red.

Table S1. Accession numbers for finished genomes.

Strain name	Accession
<i>Acidothermus cellulolyticus</i> 11B	NC_008578.1
<i>Corynebacterium diphtheriae</i> NCTC13129	NC_008578.2
<i>Corynebacterium efficiens</i> YS314	NC_004369.1
<i>Corynebacterium glutamicum</i> ATCC13032	NC_003450.3
<i>Corynebacterium jeikeium</i> K411	NC_007164.1
<i>Corynebacterium pseudotuberculosis</i> 106A	NC_017308.1
<i>Mycobacterium abscessus</i>	NC_010397.1
<i>Mycobacterium africanum</i> GM041182	NC_015758.1
<i>Mycobacterium avium</i> 104	NC_008595.1
<i>Mycobacterium avium paratuberculosis</i> MAP4	NC_021200.1
<i>Mycobacterium bovis</i> AF212297	NC_002945.3
<i>Mycobacterium canettii</i> CIPT 140010059	NC_015848.1
<i>Mycobacterium chelonae</i>	NZ_CP010946.1
<i>Mycobacterium chubuense</i> NBB4	NC_018027.1
<i>Mycobacterium gilvum</i> PYRGCK	NC_009338.1
<i>Mycobacterium goodii</i> X7B	NZ_CP012150.1
<i>Mycobacterium haemophilum</i> DSM 44634	NZ_CP011883.2
<i>Mycobacterium indicus pranii</i> MTCC 9506	NC_018612.1
<i>Mycobacterium intracellulare</i> ATCC13950	NC_016946.1
<i>Mycobacterium</i> JDM601	NC_015576.1
<i>Mycobacterium</i> JLS	NC_009077.1
<i>Mycobacterium kansasii</i> ATCC12478	NC_022663.1
<i>Mycobacterium</i> KMS	NC_008705.1
<i>Mycobacterium leprae</i> Br4923	NC_011896.1
<i>Mycobacterium liflandii</i> 128FXT	NC_020133.1
<i>Mycobacterium marinum</i> M	NC_010612.1
<i>Mycobacterium</i> MCS	NC_008146.1
<i>Mycobacterium</i> MOTT36Y	NC_017904.1
<i>Mycobacterium neoaurum</i> VKM Ac1815D	NC_023036.2
<i>Mycobacterium rhodesiae</i> NBB3	NC_016604.1
<i>Mycobacterium smegmatis</i> JS623	NC_019966.1
<i>Mycobacterium smegmatis</i> MC2 155	NC_018289.1
<i>Mycobacterium tuberculosis</i> H37Rv	NC_000962.3
<i>Mycobacterium ulcerans</i> Agy99	NC_008611.1
<i>Mycobacterium vanbaalenii</i> PYR1	NC_008726.1
<i>Mycobacterium yongonense</i> 051390	NC_021715.1
<i>Nocardia brasiliensis</i> ATCC700358	NC_018681.1
<i>Nocardia cyriacigeorgica</i> GUH2	NC_016887.1
<i>Nocardia farcinica</i> IFM10152	NC_006361.1
<i>Rhodococcus equi</i> 103S	NC_014659.1
<i>Rhodococcus erythropolis</i> PR4	NC_012490.1

<i>Rhodococcus jostii</i> RHA1	NC_008268.1
<i>Rhodococcus opacus</i> B4	NC_012522.1
<i>Rhodococcus pyridinivorans</i> SB3094	NC_023150.1
<i>Streptomyces albus</i> J1074	NC_020990.1
<i>Ilumatobacter coccineus</i> YM16-304	NC_020520.1
<i>Nocardia brasiliensis</i> ATCC 700358	NC_018681.1
<i>Intrasporangium calvum</i> DSM 43043	NC_014830.1
<i>Modestobacter marinus</i> BC501	NC_017955.1
<i>Verrucosipora maris</i> AB-18-032	NC_015434.1
<i>Rhodococcus pyridinivorans</i> SB3094	NC_023150.1
<i>Bifidobacterium thermophilum</i> RBL67	NC_020546.1
<i>Olsenella uli</i> DSM 7084	NC_014363.1
<i>Gardnerella vaginalis</i> 409-05	NC_013721.1
<i>Saccharomonospora viridis</i> DSM 43017	NC_013159.1
<i>Conexibacter woesei</i> DSM 14684	NC_013739.1

Table S2. Accession numbers for sequence read data used in plasmid assemblies.

Accession	Genus	Species	Strain
SRR071370	<i>Mycobacterium</i>	<i>rhodesiae</i>	2498
ERR248979	<i>Mycobacterium</i>	<i>intracellulare</i>	MIMRI002-sc-2012-12-17T14:56:10Z-1528566
ERR337762	<i>Mycobacterium</i>	<i>abscessus</i>	JRH116-sc-2013-05-15T12:33:12Z-1626758
ERR337800	<i>Mycobacterium</i>	<i>abscessus</i>	SMRL154-sc-2013-05-15T12:33:54Z-1626796
ERR337827	<i>Mycobacterium</i>	<i>abscessus</i>	SMRL266-sc-2013-05-15T12:35:56Z-1626909
ERR337838	<i>Mycobacterium</i>	<i>abscessus</i>	SMRL277-sc-2013-05-15T12:36:08Z-1626920
ERR340495	<i>Mycobacterium</i>	<i>abscessus</i>	SMRL429-sc-2013-05-15T12:38:52Z-1627074
ERR340496	<i>Mycobacterium</i>	<i>abscessus</i>	SMRL430-sc-2013-05-15T12:38:53Z-1627075
ERR340530	<i>Mycobacterium</i>	<i>abscessus</i>	SMRL464-sc-2013-05-15T12:39:30Z-1627109
ERR340542	<i>Mycobacterium</i>	<i>abscessus</i>	RBL476-sc-2013-05-15T12:39:44Z-1627122
DRR015955	<i>Mycobacterium</i>	<i>peregrinum</i>	DRS014063
DRR015956	<i>Mycobacterium</i>	<i>gordonae</i>	DRS014064
DRR015957	<i>Mycobacterium</i>	<i>szulgai</i>	DRS014065
DRR015962	<i>Mycobacterium</i>	<i>flavescens</i>	DRS014070
DRR016032	<i>Mycobacterium</i>	<i>triplex</i>	DRS014140
DRR016050	<i>Mycobacterium</i>	<i>peregrinum</i>	DRS014158

Table S3. Accession numbers for finished plasmids.

Accession	Plasmid name
NC_021279.1	<i>Mycobacterium abscessus</i> bolletii plasmid2
NC_018022.1	<i>Mycobacterium chubuense</i> NBB4 pMYCCH01
NC_018023.1	<i>Mycobacterium chubuense</i> NBB4 pMYCCH02
CP000657.1	<i>Mycobacterium gilvum</i> PYR-GCK pMFLV01
CP006836.1	<i>Mycobacterium kansasii</i> ATCC12478 pMK12478
HG917973.1	<i>Mycobacterium marinum</i> E11 pRAW
CP003079.1	<i>Mycobacterium smegmatis</i> JS623 pMYCSM01
CP003080.1	<i>Mycobacterium smegmatis</i> JS623 pMYCSM02
CP003081.1	<i>Mycobacterium smegmatis</i> JS623 pMYCSM03
CP000519.1	<i>Mycobacterium sp</i> KMS pMKMS01
CP000520.1	<i>Mycobacterium sp</i> KMS pMKMS02
NC_020275.1	<i>Mycobacterium yongonense</i> 05-1390 pMyong1

Table S4. Shared plasmid gene content.

Plasmid Group (# of plasmids in group)	Shared genes found only in group	Additional shared genes allowing one missing plasmid and duplicate genes
ESX-1P (2)	3 (hypothetical)	-
ESX-2P (8)	0	-
ESX-3P (4)	0	6 (DNA polymerase III subunit epsilon, 5 hypothetical)
ESX-5P (7)	0	-
ESX-2/5P (15)	0	-

References:

- Bruen TC, Philippe H, Bryant D. 2006. A Simple and Robust Statistical Test for Detecting the Presence of Recombination. *Genetics* 172:2665–2681.
- Castresana J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* 17:540–552.
- Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 10:645–656.
- Huson DH, Bryant D. 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.* 23:254–267.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 13:2178–2189.
- Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* 43:W7–W14.
- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Pond SLK. 2015. Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Mol. Biol. Evol.* 32:1342–1353.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.