

Genomic reconnaissance of clinical isolates of emerging human pathogen *Mycobacterium abscessus* reveals high evolutionary potential

Siew Woh Choo^{1,4} ϕ , Wei Yee Wee^{1,4}, Yun Fong Ngeow⁵, Wayne Mitchell^{2,3}, Joon Liang Tan^{1,5}, Guat Jah Wong^{1,4}, Yongbing Zhao^{6,7}, Jingfa Xiao⁶

*Genome Informatics Research Laboratory, High Impact Research (HIR) Building, University of Malaya, 50603 Kuala Lumpur, Malaysia.*¹

*Division of Information Sciences, School of Computer Engineering, Nanyang Technological University, Singapore.*²

*Synthetic Biology Group, Lanzatech, Auckland, New Zealand.*³

*Dental Research and Training Unit, Faculty of Dentistry, University of Malaya, 50603 Kuala Lumpur, Malaysia.*⁴

*Department of Medical Microbiology, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.*⁵

*CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China.*⁶

*University of Chinese Academy of Sciences, Beijing 100049, China.*⁷

ϕ =Authors for correspondence:

Dr. Siew Woh Choo

Genome Informatics Research Laboratory,
High Impact Research Building,
University of Malaya, 50603 Kuala Lumpur,
Malaysia

Tel: +60 3 79676463

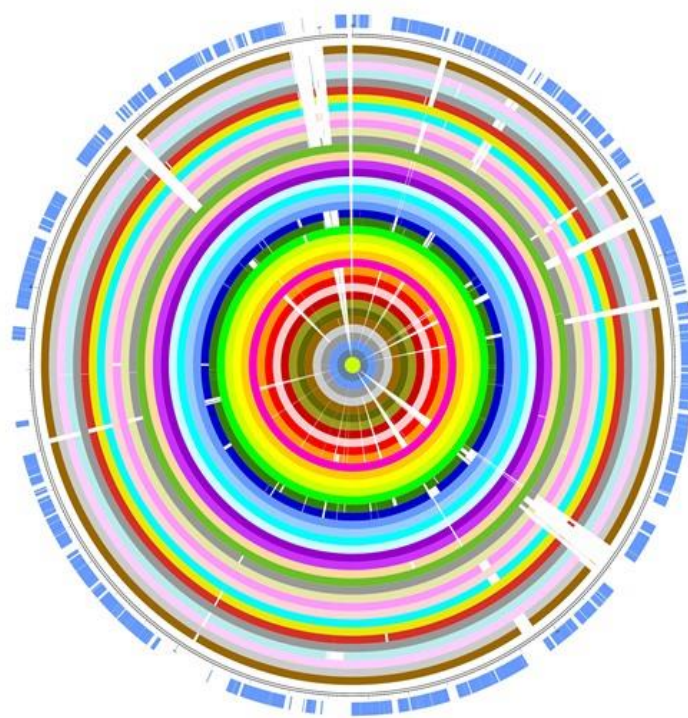
Email address: lchoo@um.edu.my

Supplementary Table S1. Summary statistics of 40 Ma genomes.

Strain	M18	M24	M93	M115	M139	M148
Status	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs
Genome Size (Mbp)	4.8	5.5	5.0	4.9	5.0	5.1
GC Content (%)	64.2	63.8	64.1	64.1	64.2	64
Number of CDS	4,850	5,605	4,970	4,947	4,983	5,363
Number of tRNAs	45	86	47	47	46	46
Number of rRNAs	3	3	3	3	3	3
Genome Identity (%)	97.3	97.4	99.4	97.6	97.8	97.1
Genome Coverage (%)	86.6	87.4	89.8	86.5	89	85.9
Strain	M154	M156	M159	M172	M94	M152
Status	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs
Genome Size (Mbp)	4.9	5	4.9	5.2	5	4.8
GC Content (%)	64.2	64.1	64.3	64	64	64.2
Number of CDS	4,709	5,039	4,864	5,244	5,146	4,833
Number of tRNAs	45	45	48	47	62	46
Number of rRNAs	3	3	3	3	3	3
Genome Identity (%)	97.3	97.3	97.3	97.3	99.3	99.4
Genome Coverage (%)	86.6	87.5	86.2	88.1	91.3	91.6
Strain	47J26	3A0119R	3A0810R	3A0122R	3A0122S	4S0116R
Status	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs
Genome Size (Mbp)	4.8	5.2	5.2	5.2	5.2	4.8
GC Content (%)	64.2	64.1	64.1	64.1	64.1	64.2
Number of CDS	4,840	5,343	5,344	5,248	5,255	4,776
Number of tRNAs	47	46	46	46	46	46

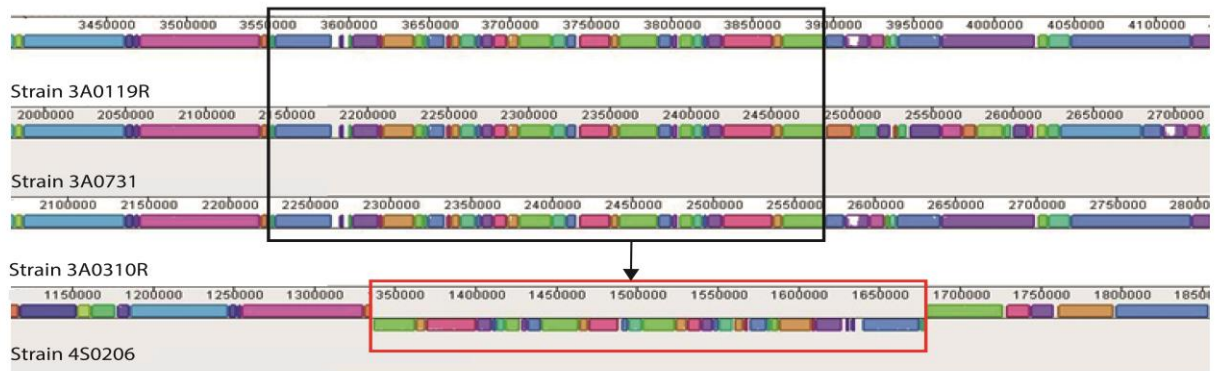
Number of rRNAs	3	3	3	3	3	3
Genome Identity (%)	97.3	99.9	99.9	99.9	99.9	99.5
Genome Coverage (%)	87.6	98.9	98.9	98.1	95.7	84.9
Strain	6G1108	3A0731	3A0930R	3A0930S	4S0206	4S0303
Status	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs
Genome Size (Mbp)	5.3	5.3	5.2	5.2	4.8	4.8
GC Content (%)	64.1	64.1	64.1	64.1	64.2	64.2
Number of CDS	5,353	5,412	5,335	5,310	4,786	4,788
Number of tRNAs	69	46	46	46	46	46
Number of rRNAs	3	3	3	3	3	3
Genome Identity (%)	99.9	99.9	99.9	99.9	99.5	99.5
Genome Coverage (%)	99.4	97.9	98.8	98.7	90.1	90.5
Strain	4S0726RA	4S0726RB	5S0304	5S1212	5S0921	5S1215
Status	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs
Genome Size (Mbp)	4.8	4.8	5.2	5.2	5.2	5.2
GC Content (%)	64.2	64.2	64	64	64	64
Number of CDS	4,777	4,788	5,227	5,226	5,236	5,196
Number of tRNAs	46	46	82	82	82	81
Number of rRNAs	3	3	3	3	3	3
Genome Identity (%)	99.5	99.5	97.4	97.4	97.4	97.4
Genome Coverage (%)	84.3	90.4	85.9	86.1	86.3	85.2
Strain	5S0817	5S0708	5S0422	5S0421	6G0125R	6G0212
Status	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs
Genome Size (Mbp)	5.2	5.2	5.3	5.2	5.1	5.1
GC Content (%)	64	64	64	64	64.1	64.1
Number of CDS	5,233	5,221	5,303	5,217	5,105	5,102
Number of tRNAs	82	82	82	82	46	46
Number of rRNAs	3	3	3	3	3	3
Genome Identity (%)	97.4	97.4	97.4	97.4	99.9	99.9

Genome Coverage (%)	84.2	84.7	86.3	86.1	99.8	99.4
Strain	6G0728R	6G0125S	6G0728S	ATCC19977		
Status	Scaffolds/Contigs	Scaffolds/Contigs	Scaffolds/Contigs	Complete genome		
Genome Size (Mbp)	5.3	5.3	5.3	5		
GC Content (%)	64.1	64.1	64.1	64.1		
Number of CDS	5,343	5,332	5,350	5,004		
Number of tRNAs	69	69	69	46		
Number of rRNAs	3	3	3	3		
Genome Identity (%)	99.9	99.9	99.9	100		
Genome Coverage (%)	99.9	99.7	99.5	100		

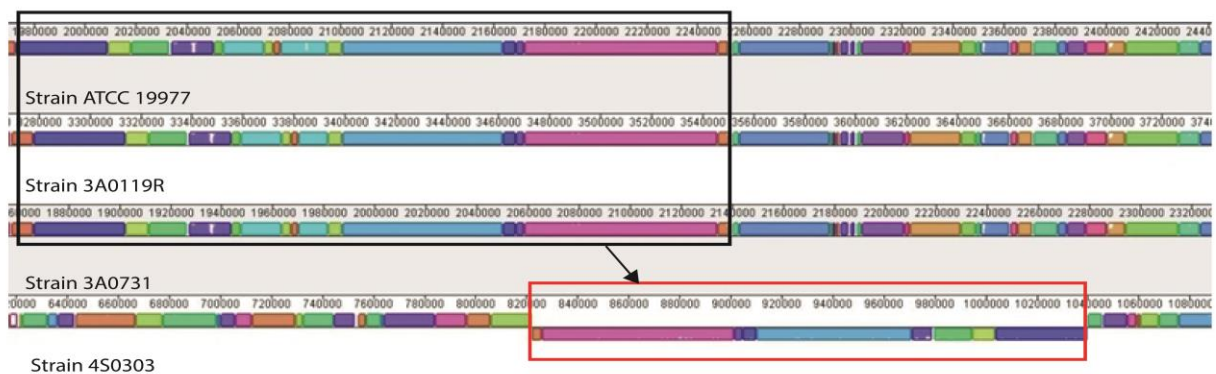


Supplementary Fig. S2: Mapping of Ma genome sequences onto the reference *M. abscessus* ATCC 19977. Each track represents one strain. In general, all genome sequences have almost covered most of the reference ma genome. From the inner track to outer track: 4S0206, 4S0303, 4S0726-RA, 4S0726-RB, 5S0304, 5S0421, 5S0422, 5S0708, 5S0817, 5S0921, 5S1212, 5S1215, 6G0125R, 6G0125S, 6G0212, 6G0728R, 6G0728S, 47J26, 4S0116S, 3A0119R, 3A0122R, 3A0122S, 3A0731, 3A0810, 3A0930R, 3A0930S, 6G1108, M172, M24, M93, M94, M115, M139, M148, M152, M154, M156, M159, M18.

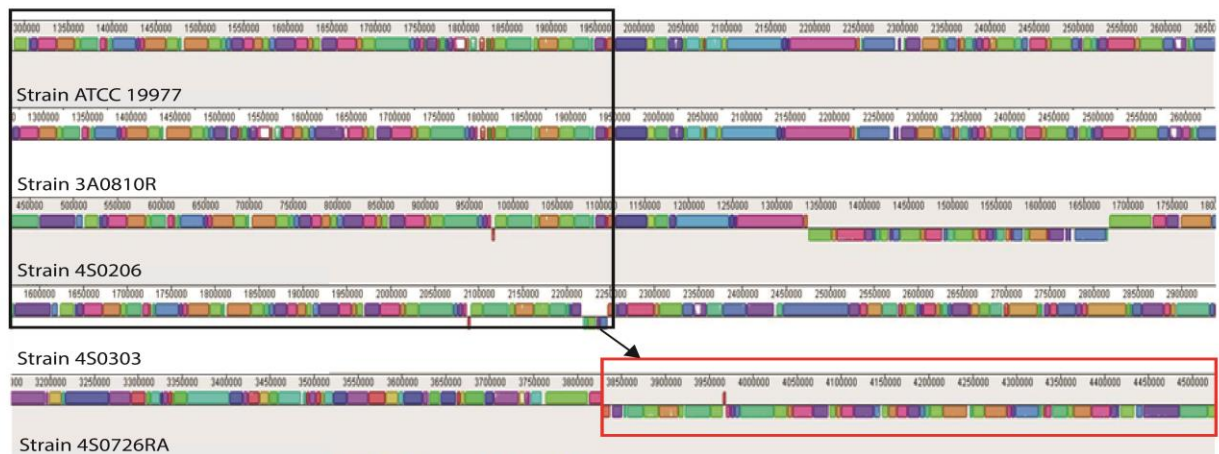
(A)



(B)

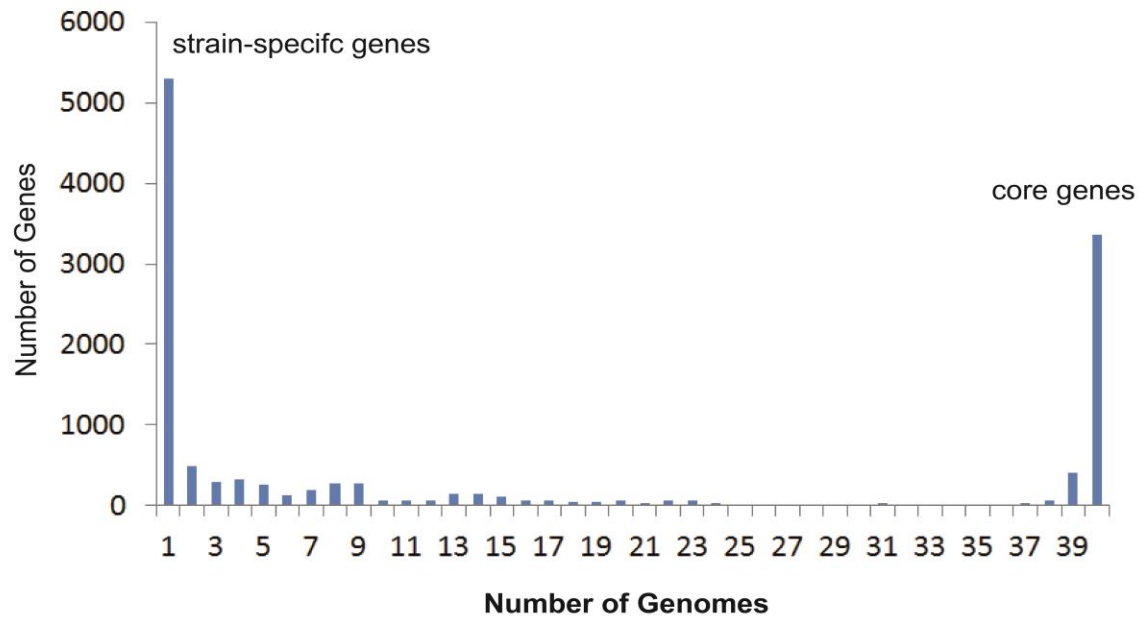


(C)

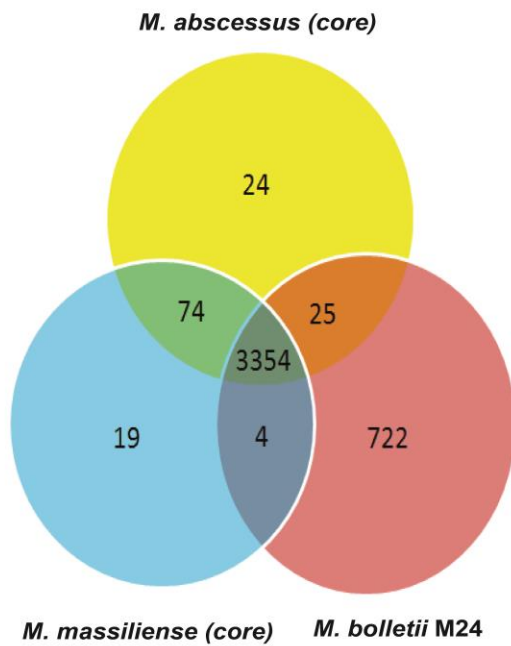


Supplementary Fig. S3: Examples of large inversions found in the *Ma* genomes. The inversions were boxed in red. Due to large number of strains that we used in this study, selected strains were shown in this figure. (A) 350kb inversion. (B) 220kb inversion. (C) 700kb inversion. The smaller inversion observed in the strain 4S0206 is not boxed because it is the same inversion in (A).

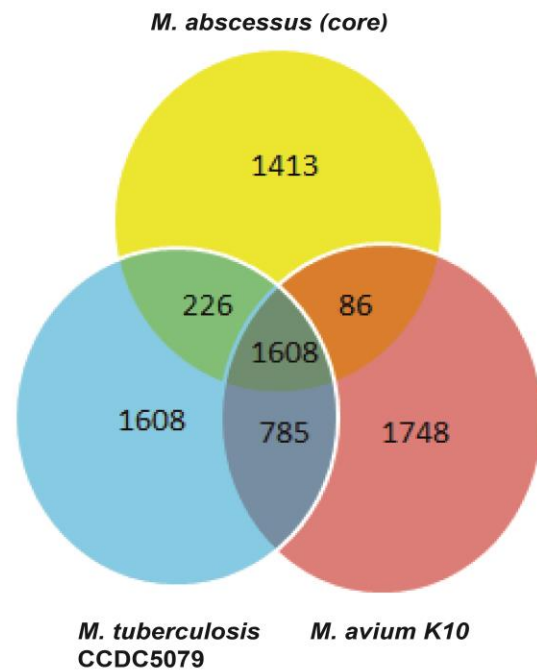
(A)



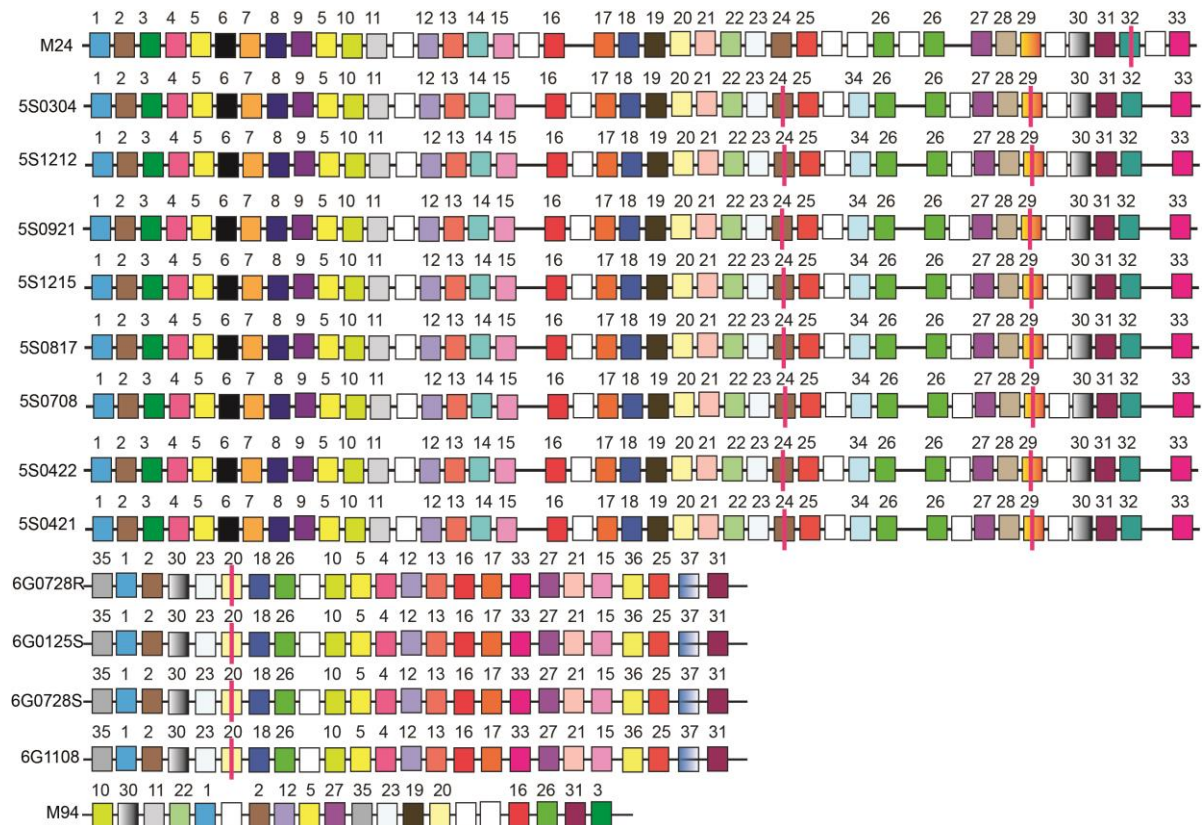
(B)



(C)

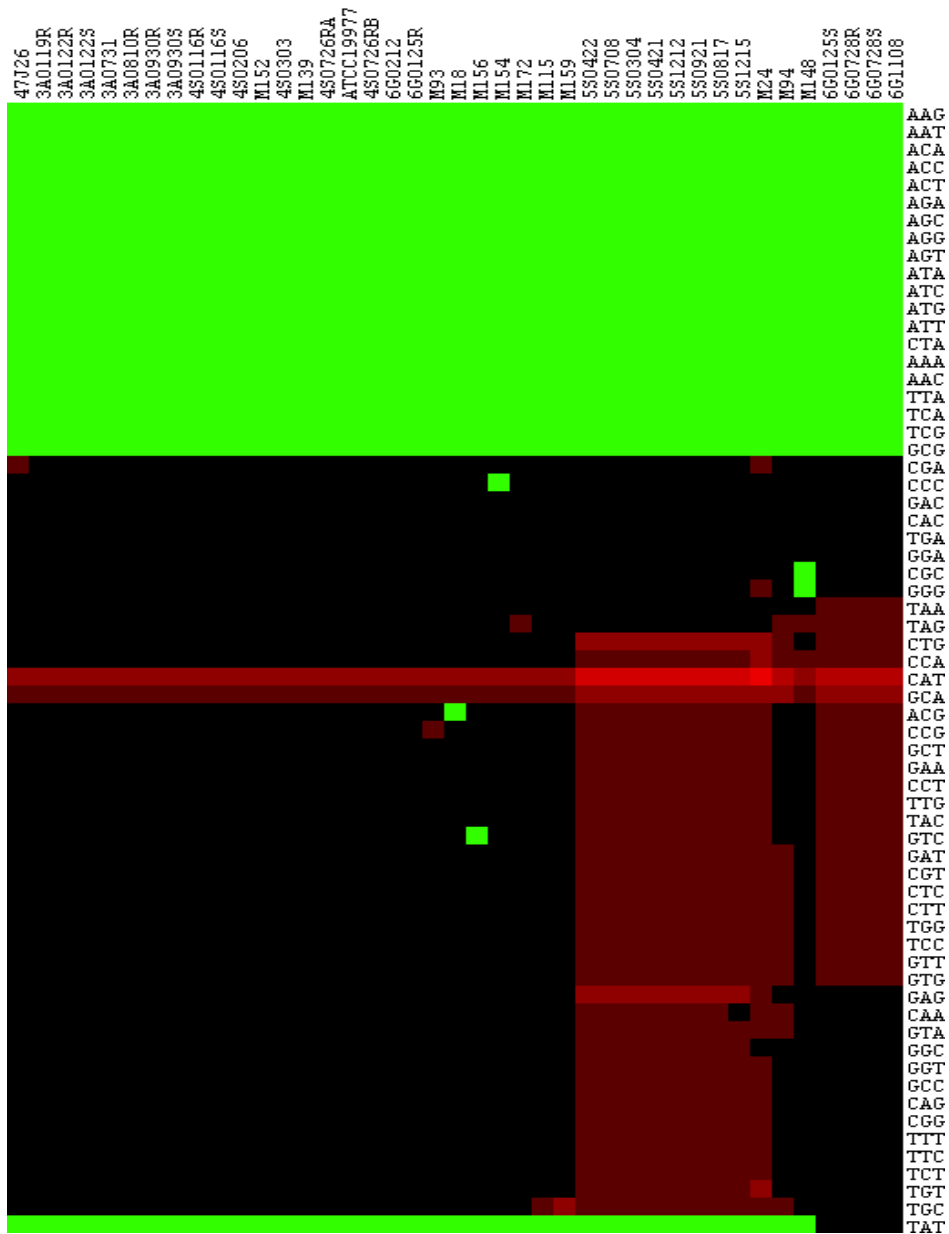


Supplementary Fig. S4: Comparative analysis of Ma genomes. (A) Genome Set Size versus Frequency of Common Genes in Each Size Set. The most left bar represents the strain-specific genes in Ma, whereas the most right bar represents the core genes found in the complete set of genomes. Genes list comparisons: (B) Comparison between Ma subspecies genes. (C) Comparison between fast-grower (Ma) and slow-growers (*M. tuberculosis* and *M. avium*).

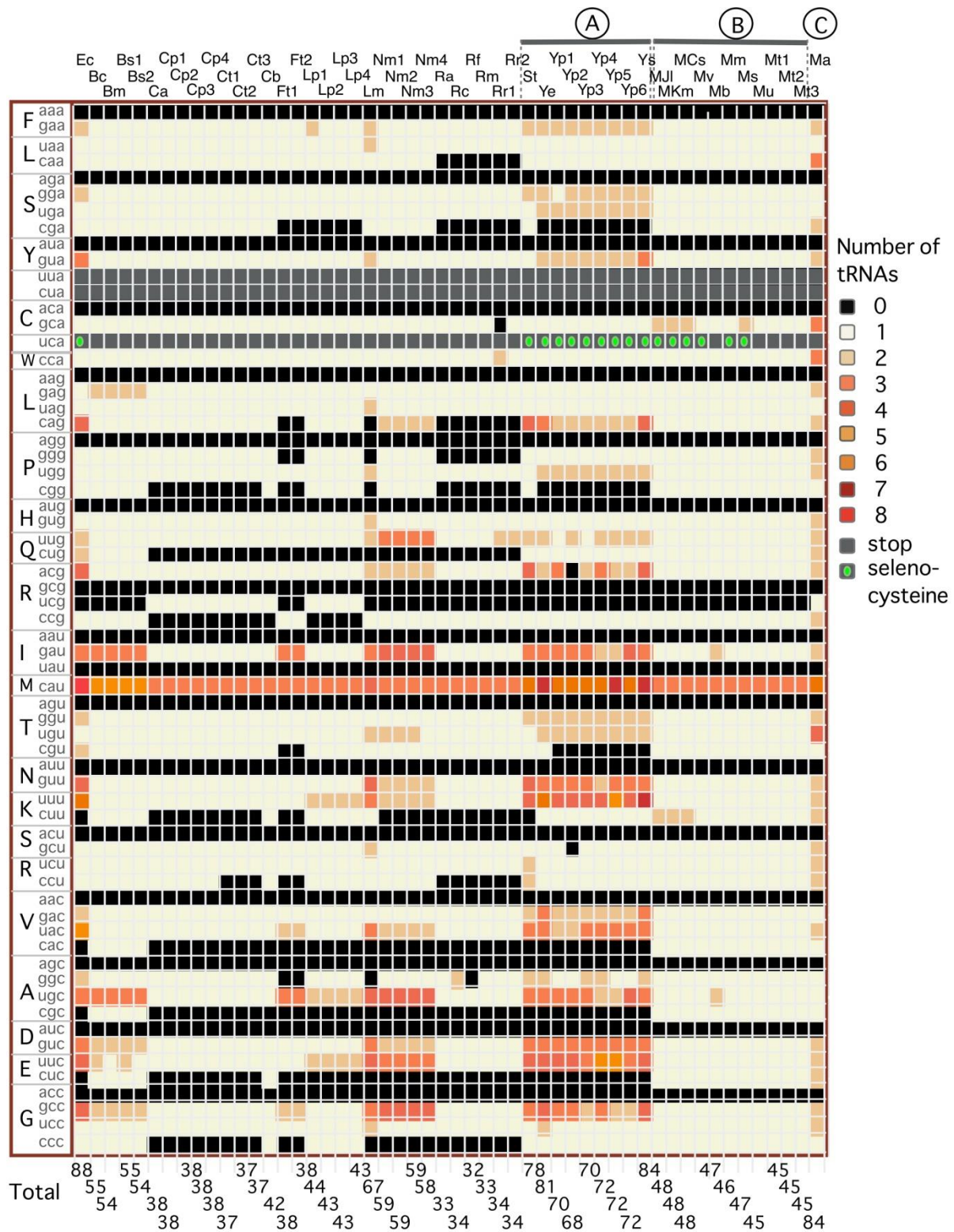


Supplementary Fig. S5: The order of the tRNA gene sequences in the genomic islands. In general, there are 3 distinct islands. The order of the tRNAs in the M24 tRNA island is very similar to the order in the islands found in 5S0304, 5S1212, 5S0921, 5S1215, 5S0817, 5S0708, 5S0422 and 5S0421 from the US group although we observed some pseudo-tRNAs such as tRNA-Glu-TTC and tRNA-Pro-GGG, probably accumulated over evolutionary time. The observed tRNA islands in the strains 6G0728R, 6G0125S, 6G0728S and 6G1108 are identical, whereas M94 has a very different order of tRNAs in its island. A white box indicates a coding sequence, whereas the red line down a box represents a pseudo-tRNA. Manual inspection of these islands in genome browser showed that, based on the similarity of the order of the tRNA genes in the islands, these islands could be classified into three distinct groups: G1(M24, 5S0304, 5S1212, 5S0921, 5S1215, 5S0817, 5S0708, 5S0422 and 5S0421); G2 (6G0728R, 6G0125S, 6G0728S and 6G1108); and G3 (M94) (Fig.4).

Legend: 1=tRNA-Lys-CTT; 2= tRNA-Gly-TCC; 3=tRNA-Leu-CAA; 4=tRNA-Gln-TTG; 5=tRNA-Gln-CTG; 6=tRNA-Thr-TGT; 7=tRNA-Thr-GGT; 8=tRNA-Gly-GCC; 9=tRNA-Leu-CAG; 10=tRNA-Asn-GTT; 11=tRNA-Ala-TGC; 12=tRNA-Glu-CTC; 13=tRNA-Phe-GAA; 14=tRNA-Leu-GAG; 15=tRNA-Arg-ACG; 16=tRNA-Ile-GAT; 17=tRNA-Asp-GTC; 18=tRNA-Val-TAC; 19=tRNA-Tyr-GTA; 20=tRNA-Pro-TGG; 21=tRNA-Arg-CCG; 22=tRNA-Lys-TTT; 23=tRNA-Trp-CCA; 24=tRNA-Glu-TTC; 25=tRNA-Arg-CCT; 26=tRNA-Met-CAT; 27=tRNA-His-GTG; 28=tRNA-Arg-TCT; 29=tRNA-Pro-GGG; 30=tRNA-Cys-GCA; 31=tRNA-Thr-CGT; 32=tRNA-Pro-CGG; 33=tRNA-Ser-GCT; 34=tRNA-Ala-GGC; 35=tRNA-Leu-TAG; 36=tRNA-Leu-TAA; and 37=tRNA-Ile-TAT.



Supplementary Fig. S6: The distribution of anticodons in *Ma* strains. Red=2-5; green=0;black=1. Of our 40 *Ma* strains, 14 (35%) showed distinct tRNA anticodon distributions. They are strains 5S0304, 5S1212, 5S0921, 5S1215, 5S0817, 5S0708, 5S0422, 5S0421, M24, M94, 6G0728R, 6G0125S, 6G0728S and 6G1108. M24 shows highest number of tRNA-CAT copies (6 copies) compared to 3 copies which is found in most bacteria. Interestingly, the TAT-anticodon tRNAs were only found in 6G0728R, 6G0125S, 6G0728S and 6G1108, but not in other *Ma* strains .



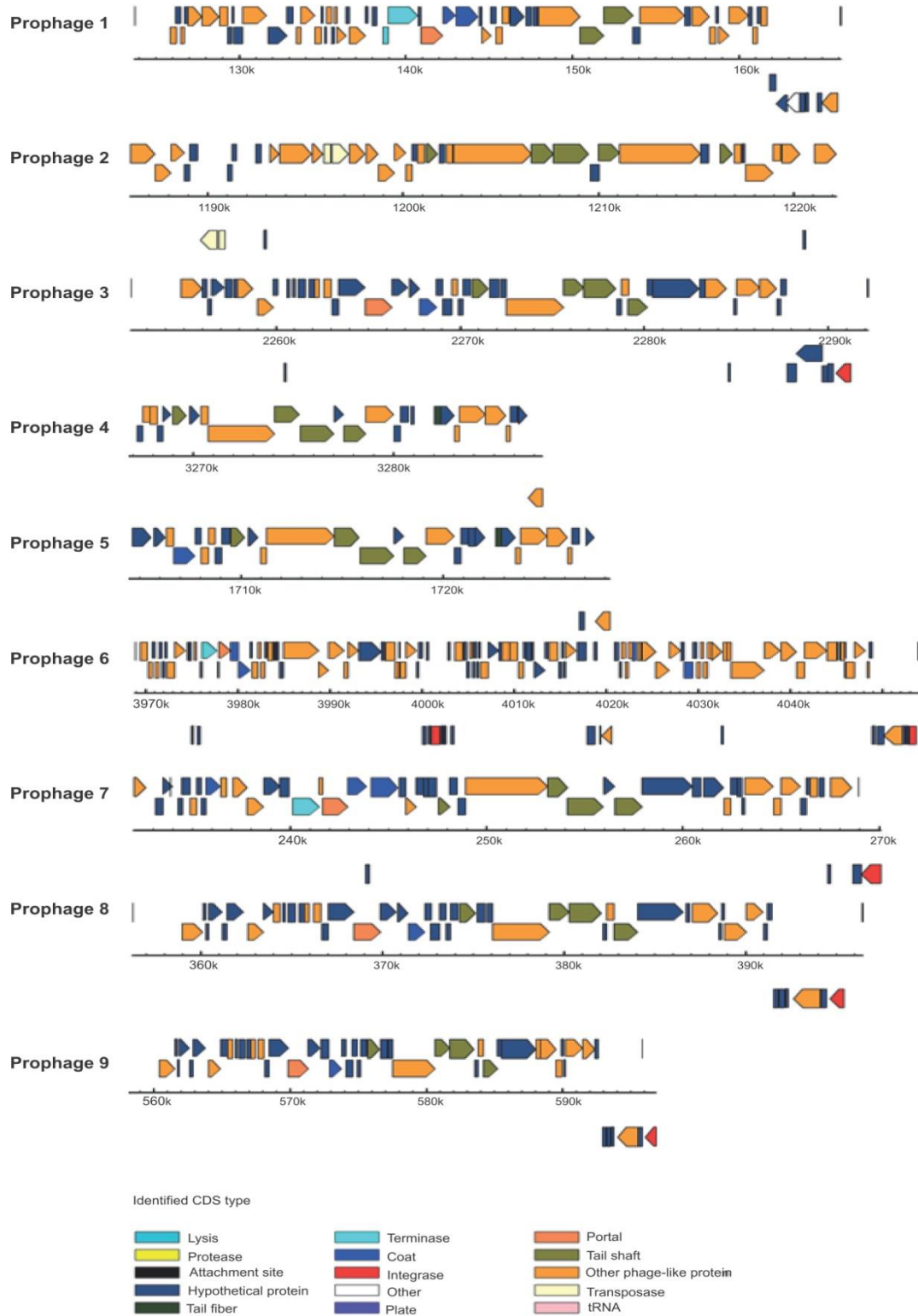
Supplementary Fig. S7: tRNA Frequency by anti-codon in obligate and facultative intracellular pathogens.

Legend. tRNA frequencies for 51 intracellular bacterial pathogens and *Escherichia coli* were extracted from the tRNADB (Nucleic Acids Res. 2009 Jan;37(Database issue):D93-7. doi: 10.1093/nar/gkn787. Epub 2008 Nov 4. GtRNadb: a database of transfer RNA genes detected in genomic sequence. Chan

PP, Lowe TM.) or calculated de novo with ARAGORN (ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Laslett D, Canback B. *Nucleic Acids Res.* 2004 Jan 2;32(1):11-6. Print 2004.).

The frequencies of each tRNA, classified by anticodon, in each species, is represented in the figure by color (see figure key). The cluster of *Yersinia* species. is marked A. The cluster of *Mycobacterium* species is marked B, except for *Mycobacterium abscessus* (marked C). Note that the *M. Abscessus*' tRNA compliment resembles *Yersinia* but not other mycobacteria. The species and the abbreviations used in the figure follow. Obligately intracellular bacteria are denoted by (*) after the full name. Unmarked species are facultative intracellular pathogens.

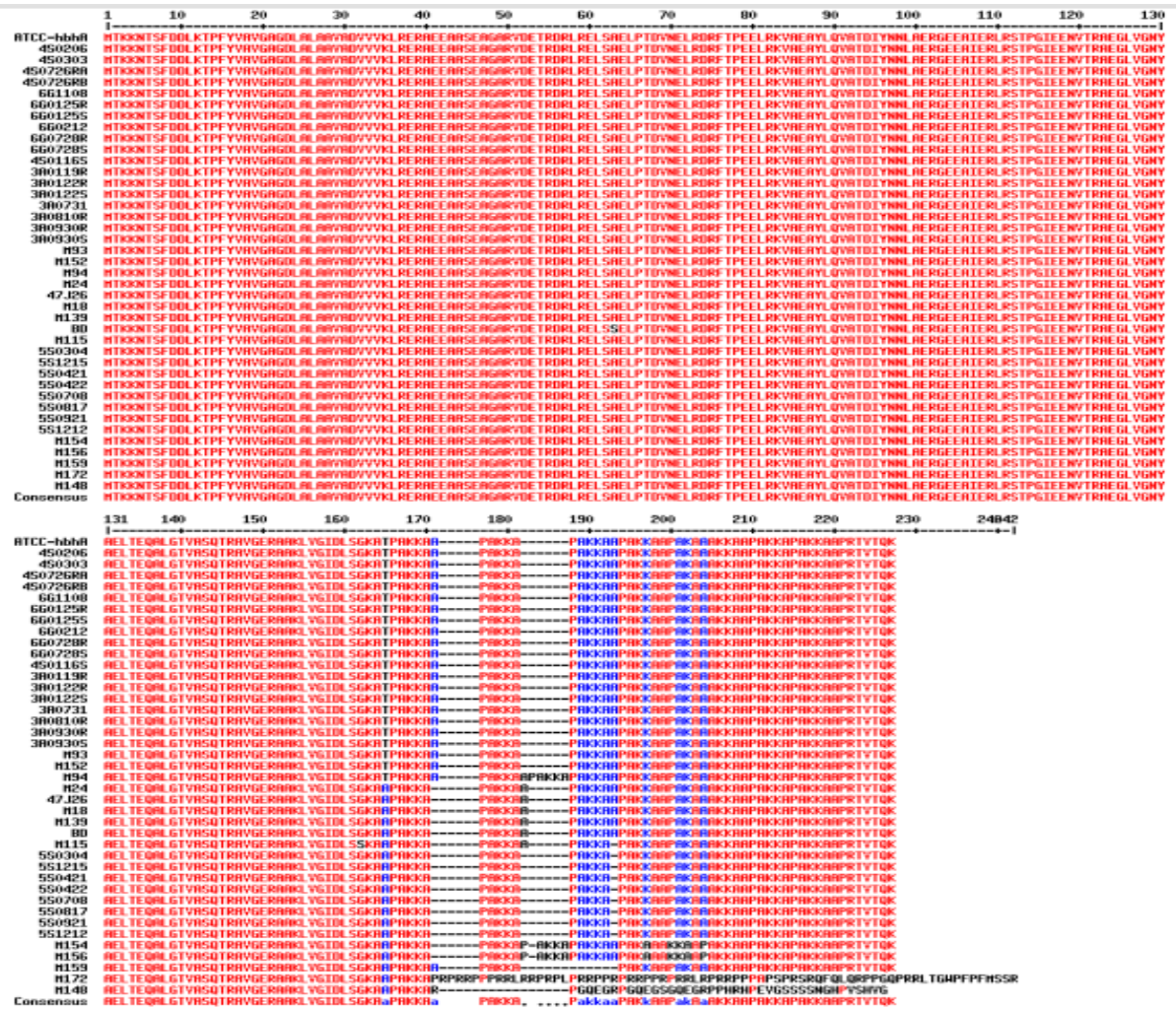
*Escherichia coli*_K12, Ec; *Brucella canis*, Bc; *Brucella melitensis*_chrI, Bm; *Brucella suis*_1330, Bs1; *Brucella suis*_ATCC_23445_chrI, Bs2; *Chlamydomphila abortus*_S26_3 (*), Ca; *Chlamydomphila pneumoniae*_AR39(*), Cp1; *Chlamydomphila pneumoniae*_CWL029(*), Cp2; *Chlamydomphila pneumoniae*_J138,(*), Cp3; *Chlamydomphila pneumoniae*_TW_183(*), Cp4; *Chlamydia trachomatis*, (*) Ct1; *Chlamydia trachomatis*_434_Bu(*), Ct2; *Chlamydia trachomatis*_L2b_UCH_1_proctitis(*), Ct3; *Coxiella burnetii*_RSA_331(*), Cb; *Francisella tularensis*_holarctica_FTA, Ft1; *Francisella tularensis*_mediasiatica_FSC147, Ft2; *Legionella pneumophila*_Corby, Lp1; *Legionella pneumophila*_Lens, Lp2; *Legionella pneumophila*_Paris, Lp3; *Legionella pneumophila*_Philadelphia_1, Lp4; *Listeria monocytogenes*_chr, Lm; *Neisseria meningitidis*_053442; Nm2; *Neisseria meningitidis*_FAM18; Nm3; *Neisseria meningitidis*_MC58; Nm4; *Neisseria meningitidis*_Z2491, Nm5; *Rickettsia akari*_Hartford, Ra(*); *Rickettsia canadensis*_URRWXCal2 *Rickettsia felis*_URRWXCal, (*) Rc; *Rickettsia felis*_URRWXCal2 *Rickettsia felis*_URRWXCal2(*), Rf; *Rickettsia massiliae*_MTU5(*), Rm; *Rickettsia rickettsii*_Iowa(*), Rr1; *Rickettsia rickettsii*_Sheila_Smith(*), Rr2; *Salmonella typhi*_Ty2 St; *Yersinia enterocolitica*_8081 Ye; *Yersinia pestis*_Angola, Yp1; *Yersinia pestis*_Antiqua *Yersinia pestis*_Antiqua., Yp2; *Yersinia pestis*_CO92_chr.trna50-ValTAC Yp3; *Yersinia pestis*_Nepal516 *Yersinia pestis*_Nepal516, Yp4; *Yersinia pestis*_Pestoides_F, Yp5; *Yersinia pestis*_biovar_Mediaevails_str_91001, Yp6; *Yersinia pseudotuberculosis*_YPIII, Ys2; *Mycobacterium*_JLS, MJl; *Mycobacterium*_KMS, MKm; *Mycobacterium*_MCS *Mycobacterium*_sp_MCS, MCs; *Mycobacterium avium*_104, My; *Mycobacterium bovis*_BCG_Pasteur_1173P2, Mb; *Mycobacterium marinum*_M, Mm; *Mycobacterium smegmatis*_MC2_155, Ms; *Mycobacterium ulcerans*_Agy99, Mu; *Mycobacterium tuberculosis*_CDC1551, Mt1; *Mycobacterium tuberculosis*_F11, Mt2; *Mycobacterium tuberculosis*_H37Ra, Mt4; *Mycobacterium abscessus*_M24, Ma



Supplementary Fig. S8: The structure of the putative intact prophages found in the *Ma* genomes. These prophages consist of phage-associated proteins like phage tail, coat and enzymes like integrase and terminase which assist in the integration of phage into bacteria. Black thin bars in prophages 1,3,6,7,8 and 9 represent the Att sites for site-specified recombination.

Prophages/ strains	3A0810R	3A0119R	3A0930R	3A0930S	3A0731	3A0122R	3A0122S	4S0206	4S0303	4S0726RA	4S0726RB	5S0421	5S0708	5S1212	5S1215	M93	M156	M172
1	Yellow	Yellow	Yellow	Yellow	Yellow													
2	Light Blue	Light Blue	Light Blue	Light Blue	Light Blue	Light Blue	Light Blue											
3	Light Blue	Light Blue	Light Blue	Light Blue	Light Blue	Light Blue	Light Blue											
4								Blue	Blue	Blue	Blue							
5												Red	Red	Red	Red			
6						Green	Green											
7																Orange		
8																	Purple	
9																		Teal

Supplementary Fig. S9: Distribution of the putative intact prophages in the Ma genomes. The Ma genomes that share a same prophage are highlighted in the same color. There are 6 prophages that are shared by more than one genome and 3 prophages that are each uniquely inserted into a single genome



Supplementary Fig. S10: Sequence alignment of *hbhA* genes in *Ma* genomes. Strains ATCC-*hbhA* to M94 (*M. abscessus* sensu stricto) show amino acid T in the 165th position, while the remaining strains (*M. bolletii* and *M. massiliense*) show amino acid A instead of T.

Supplementary Table S11: Summary statistics of tRNA Islands

Strain	M24	M94	6G1108	6G0728S	6G0125S	6G0728R	5S0421
Contig number	Contig4	Contig33	Contig1	Contig0	Contig0	Contig0	Contig4
Start Position	11,282	51,078	74,604	115,920	75,137	116,345	345,049
End Position	20,728	57,467	77,594	118,910	78,127	119,335	355,287
Number of CDS	7	3	1	1	1	1	4
Number of tRNAs	35	17	23	23	23	23	36
tRNA density	31.8%	19.1%	55.9%	55.9%	55.9%	55.9%	25.5%
Avg distance between tRNAs (bp)	167.8	164.2	59.0	59.0	59.0	59.0	207.7
Island size (bp)	9,445	6,388	2,989	2,989	2,989	2,989	10,237
Strain	5S0422	5S0708	5S0817	5S1215	5S0921	5S1212	5S0304
Contig number	Contig10	Contig5	Contig6	Contig3	Contig7	Contig0	Contig6
Start Position	345,080	2590,958	2984,997	3726,488	603,919	2566,534	345,138
End Position	355,318	2601,196	2995,235	3736,726	614,157	2576,699	355,376
Number of CDS	4	3	4	4	4	4	4
Number of tRNAs	36	36	36	36	36	36	36
tRNA density	25.5%	25.5%	24.8%	24.7%	25.5%	25.5%	25.5%
Avg distance between tRNAs (bp)	215.9	216.9	216.9	219.7	216.9	216.9	216.9
Island size (bp)	10,237	10,237	10,237	10,237	10,237	10,237	10,237

Supplementary Table S12: Summary of 9 prophage regions predicted in Ma genomes by PHAST.

Prophages	Length(Kb)	CDS	ATT-site show up	GC content
1	41.1	62	Yes	64.1
2	43.6	56	Yes	63.7
3	36.2	42	No	60.3
4	32.2	43	No	63.71
5	32.7	45	No	63.5
6	85	120	Yes	64.2
7	38	51	Yes	63.8
8	40.1	49	Yes	63.1
9	50.7	50	Yes	63.5