

Simple sequence repeats in mycobacterial genomes

VATTIPALLY B SREENU, PANKAJ KUMAR, JAVAREGOWDA NAGARAJU and HAMPAPATHALU A NAGARAJARAM

J. Biosci. 32(1), January 2007, 3–15, © Indian Academy of Sciences

Supplementary Material 1

The list of microsatellite rich as well as poor regions in the five mycobacterial genomes.

	Local GC%	Repeat rich(+)/ Repeat poor(-)	Total ORFs	Number of hypothetical proteins	Non hypothetical proteins
<i>M. avium</i>					
19-20	0.72	+	11	8	SodA, GlpQ1, prephenatedehydratase
34-35	0.70	+	8	7	DNA polymerase III subunits
39-40	0.71	+	7	5	SelD, PfpI
44-45	0.73	+	9	7	TrbB, CspA_1
47-48	0.70	+	8	4	EphA, FtsH, GTP cyclohydrolase I, FolP
63-64	0.71	+	8	5	phosphoribosylamine--glycine ligase, GgtA, adenylosuccinatelyase
92-93	0.69	+	7	4	succinyl-CoA synthetase subunit beta, succinyl-CoA synthetase alpha subunit, acetyl-CoA acetyltransferase
118-119	0.73	+	8	4	LipI, primosome assembly protein PriA, methionyl-tRNA formyltransferase, Fmu
128-129	0.71	+	5	2	MutA, methylmalonyl-CoA mutase, arginine/ornithine transport system ATPase
145-146	0.69	+	8	3	lysyl-tRNA synthetase, translation initiation factor IF-3, 50S ribosomal protein L35, 50S ribosomal protein L20, TsnR
146-147	0.72	+	7	0	phenylalanyl-tRNA synthetase beta subunit, N-acetyl-gamma-glutamyl-phosphate reductase, bifunctional ornithine acetyltransferase/N-acetylglutamate synthase, acetylglutamate kinase, acetylornithine aminotransferase, ornithine carbamoyltransferase, arginine repressor
152-153	0.70	+	10	7	tyrosine recombinase, cytidylate kinase, GTP-binding protein EngA
166-167	0.69	+	7	6	PE_5
171-172	0.69	+	9	5	6-phosphogluconate dehydrogenase, Ndh, short chain dehydrogenase, ModA
176-177	0.68	+	11	6	LppE, short chain dehydrogenase, chorismate mutase, FbpB, AdhA_2
196-197	0.71	+	0	0	
210-211	0.72	+	7	4	UDP-N-acetylmuramoylalanine-D-glutamate--2,6-diaminopimelate ligase, PbpB, S-adenosyl-methyltransferase
257-258	0.69	+	5	4	glycerol-3-phosphate acyltransferase
274-275	0.66	+	6	4	AlkA, Ogt
285-286	0.69	+	6	3	alpha-ketoglutarate decarboxylase, LpqZ, malate dehydrogenase
289-290	0.72	+	7	4	acyl-CoA synthase, succinyl-diaminopimelate desuccinylase, PE_6
292-293	0.70	+	7	5	acyl-CoA synthase, sulfate adenylyltransferase
328-329	0.69	+	11	7	ribosome releasing factor, uridylate kinase, amidase, elongation factor Ts
342-343	0.69	+	7	5	isopentenyl pyrophosphate isomerase, Phr
359-360	0.70	+	8	7	CatB

361-362	0.72	+	7	5	pyruvatedecarboxylase, PknJ
367-368	0.73	+	7	5	LipV, molybdopterinbiosynthesisproteinMoeB
389-390	0.72	+	6	4	CtpI, AdhD
404-405	0.71	+	5	3	MmpL3, tRNA(guanine-N(7)-)-methyltransferase
423-424	0.72	+	9	8	AtsG
448-449	0.73	+	10	7	glutamate-1-semialdehydeaminotransferase, CcsA, CcsB
449-450	0.74	+	9	7	1,4-dihydroxy-2-naphthoateoctaprenyltransferase,, 5'-methylthioadenosinephosphorylase
88-89	0.61	-	10	10	
283-284	0.67	-	8	6	DeaD, LprE
309-310	0.64	-	12	12	
415-416	0.63	-	8	8	
418-419	0.62	-	7	4	IS1110transposase, MmpS1, MmpL4_5
<i>M. leprae</i>					
37-38	0.54	-	5	3	thiaminebiosynthesisproteinThiC, phosphomethylpyr imidinekinase
157-158	0.57	-	4	2	proteasome[beta]-typesubunit2, proteasome[alpha]- typesubunit1
164-165	0.56	-	1	1	
198-199	0.55	-	1	1	
243-244	0.56	-	6	3	probableLysR-familytranscriptionalregulator, alkylhy droperoxidoreductase, L-lactatedehydrogenase
263-264	0.56	-	3	0	putativeaminopeptidase2, phosphoribosylformylglyci namidinesynthasesubunitI, phosphoribosylformylglyc inamidinesynthase
326-327	0.52	-	8	2	putativecelldivisionprotein, putativecelldivisi onprotein, glucose-inhibiteddivisionproteinB, putativeinnermembraneprotein, ribonucleaseP, 50SribosomalproteinL34
<i>M. bovis</i>					
33-34	0.75	+	4	2	PE-PGRSFAMILYPROTEIN[FIRST, PE- PGRSFAMILYPROTEIN
35-36	0.67	+	10	3	PEFAMILYPROTEIN, PPEFAMILYPROTEIN, LOWMOLECULARWEIGHTPROTEIN, PROBABLECONSERVEDTRANSMEMBRANEPROTEIN, PROBABLEPROTEASEPRECURSOR, PROBABLECONSERVEDTRANSMEMBRANEPROTEIN, PROBABLETRANS-ACONITATEMETHYLTRAN SFERASETAM
36-37	0.66	+	8	4	PROBABLESULFATASE, PE- PGRSFAMILYPROTEIN, PROBABLETRANSCRIPTIONALREGULATORYPROTEIN, PROBABLEDEHYDROGENASE/REDUCTASE
37-38	0.62	+	5	1	PPEFAMILYPROTEIN, PUTATIVEOXIDOREDUCTASE, PROBABLECONSERVEDINTEGRALMEMBRANE, POSSIBLECONSERVEDEXPORTEDPROTEIN

42-43	0.63	+	4	0	molecular chaperone DnaK, PROBABLE GRP E PROTEIN (HSP-70), PROBABLE CHAPERONE PROTEIN D NA J1, PROBABLE HEAT SHOCK PROTEIN
43-44	0.64	+	5	3	adenylosuccinatesynthetase, PROBABLE CONSERVED INTEGRAL MEMBRANE
67-68	0.71	+	7	5	PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN, PE-PGRS FAMILY PROTEIN
83-84	0.70	+	10	5	POSSIBLE TRANSCRIPTIONAL REGULATORY PROTEIN, PUTATIVE TRANSPOSASE (FRAGMENT), PE-PGRS FAMILY PROTEIN, POSSIBLE TRANSCRIPTIONAL REGULATORY PROTEIN, PE-PGRS FAMILY PROTEIN
84-85	0.70	+	8	4	PE-PGRS FAMILY PROTEIN, PROBABLE 3-HYDROXYISOBUTYRATE DEHYDROGENASE MMSB, PROBABLE ACYL-CO A DEHYDROGENASE FADE9, PROBABLE METHYL MALONATE-SEMIALDEHYDE DEHYDROGENASE MMSA
85-86	0.63	+	10	4	PPE FAMILY PROTEIN, PUTATIVE TRANSPOSASE (FRAGMENT), POSSIBLE TWO COMPONENTS SYSTEM, POSSIBLE TWO COMPONENTS SYSTEM, POSSIBLE ZINC-CONTAINING ALCOHOL DEHYDROGENASE, POSSIBLE FERREDOXIN
92-93	0.70	+	8	3	PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN, POSSIBLE DEAMINASE, PUTATIVE TRANSPOSASE (FRAGMENT), PE-PGRS FAMILY PROTEIN, PE-PGRS FAMILY PROTEIN
97-98	0.66	+	8	2	PROBABLE ACYL-CO A DEHYDROGENASE FADE10, POSSIBLE CONSERVED EXPORTED PROTEIN, POSSIBLE CONSERVED TRANS MEMBRANE PROTEIN, PPE FAMILY PROTEIN, POSSIBLE CONSERVED TRANS MEMBRANE PROTEIN, POSSIBLE TRANSCRIPTIONAL REGULATORY PROTEIN
109-110	0.69	+	7	1	PE-PGRS FAMILY PROTEIN, PE-PGRS FAMILY PROTEIN, 50S ribosomal protein L32, PE-PGRS FAMILY PROTEIN, MYCOBACTERIAL PERSISTENCE REGULATORY MRPA, PROBABLE TWO COMPONENTS SENSOR
121-122	0.71	+	8	1	POSSIBLE HEMOLYSIN-LIKE PROTEIN, SHORT (C15) CHAIN Z-ISOPRENYL, PE-PGRS FAMILY PROTEIN, PPE FAMILY PROTEIN, PROBABLE CELLULOSE CELA2A (ENDO-1,4-BETA-GLUCANASE), PROBABLE CELLULOSE CELA2B (ENDO-1,4-BETA-GLUCANASE)
126-127	0.66	+	9	3	5-methyltetrahydropteroyl tri glutamate--homocysteinemethyltransferase, PPE FAMILY PROTEIN, POSSIBLE ACETYL-CO A ACETYL TRANSFERASE (ACETOACETYL-CO A), POSSIBLE ENOYL-CO A HYDRATASE, PUTATIVE OXIDOREDUCTASE, PROBABLE INTEGRAL MEMBRANE PROTEIN

162-163	0.70	+	5	0	6-phosphogluconolactonase, PUTATIVEOXPPCYCLEPROTEIN, glucose-6-phosphate 1-dehydrogenase, transaldolase, transketolase
163-164	0.70	+	6	2	PE-PGRSFAMILYPROTEIN, POSSIBLETRANSCRIPTIONALACTIVATORPROTEIN, PROBABLEQUINONEREDUCTASEQOR, PROBABLEUNIDENTIFIEDANTIBIOTIC-TRANSPORTINTEGRAL
196-197	0.64	+	5	3	POSSIBLEINTEGRALMEMBRANEPROTEIN, acyl-CoAsynthase
215-216	0.61	+	3	1	isocitratelase, PPEFAMILYPROTEIN
228-229	0.68	+	0	0	
244-245	0.65	+	9	2	Probableconservedtransmembraneprotein, PROBABLECONSERVEDMEMBRANEPROTEIN, Possibleconservedintegralmembrane, PROBABLETRANSMEMBRANECYTOCHROME C, Probableasparagine synthetaseAsnB, ProbablecarbohydratekinaseCbhK, POSSIBLECONSERVEDMEMBRANEPROTEIN
276-277	0.71	+	4	0	enoyl-CoAhydratase, PE-PGRSFAMILYPROTEIN[FIRST, PROBABLETRANSCRIPTIONALREGULATORYPROTEIN, HYPOTHETICALALANINERICHPROTEIN
277-278	0.67	+	9	4	PE-PGRSFAMILYPROTEIN[FIRST, dihydrolipoamideacetyltransferase, PROBABLEPYRUVATEDEHYDROGENASEE1, PROBABLEPYRUVATEDEHYDROGENASEE1, PROBABLECITRATE(PRO-3S)-LYASE(BETA
290-291	0.66	+	8	1	pyridoxinebiosynthesisprotein, pyridoxamine5'-phosphateoxidase, PPEFAMILYPROTEIN, PROBABLECONSERVEDMEMBRANEPROTEIN, ALPHA-MANNOSYLTRANSFERASEPIMA, lipidAbiosynthesislauroyl, PROBABLEPISYNTHASEPGSA1
312-313	0.65	+	7	1	dihydrolipoamidedehydrogenase, POSSIBLENICKEL-TRANSPORTINTEGRALMEMBRANE, shortchaindehydrogenase, PROBABLEALDEHYDEDEHYDROGENASEALDC, POSSIBLEAMIDOTRANSFERASE, PROBABLEGLUTAMINESYNTHETASEGLNA4
333-334	0.66	+	10	1	PROBABLELIPOPROTEINLPGA, PUTATIVEESAT-6LIKEPROTEIN, PPEFAMILYPROTEIN, PEfamilyprotein, PUTATIVESECRETEDESAT-6LIKE, PEFAMILYPROTEIN, PPEFAMILYPROTEIN, PEFAMILYPROTEIN, PROBABLETRANSPPOSASE
348-349	0.63	+	6	0	NADHdehydrogenasesubunitN, PPEFAMILYPROTEIN, PPEFAMILYPROTEIN, POSSIBLETRANSCRIPTIONALREGULATORYPROTEIN, POSSIBLEDIOXYGENASE, POSSIBLEINTEGRALMEMBRANEPROTEIN
368-369	0.64	+	5	1	isocitratelase, O-acetylhomoserinesulfhydrylase, homoserineO-acetyltransferase, POSSIBLEMETHYLTRANSFERASE(METHYLASE)

369-370	0.72	+	2	1	PE-PGRSFAMILYPROTEIN[FIRST
370-371	0.64	+	3	0	PPEFAMILYPROTEIN[FIRST, PROBABLETRANSPOSASE, PROBABLETRANSPOSASE
371-372	0.64	+	0	0	
375-376	0.68	+	8	2	POSSIBLETRANSPOSASE, POSSIBLETRANSPOSASE, PE-PGRSFAMILYPROTEIN, POSSIBLEDEHYDROGENASE, PROBA BLECONSERVEDLIPOPROTEINLPQD, shortchaindehydrogenase
386-387	0.64	+	9	0	MCE-FAMILYPROTEINMCE4D, MCE- FAMILYPROTEINMCE4C, MCE- FAMILYPROTEINMCE4B, MCE- FAMILYPROTEINMCE4A, CONSERVEDH YPOTHETICALINTEGRALMEMBRANE, CONSERVEDHYPOTHETICALINTEGRALMEM BRANE, 3-ketoacyl-(acyl-carrier-protein)reductase, PROBABLEFERREDOXINFDXD, PROBABLEACYL- COADEHYDROGENASEFADE26
387-388	0.76	+	2	0	acyl-CoAsynthase, PE-PGRSFAMILYPROTEIN
388-389	0.75	+	3	2	PE-PGRSFAMILYPROTEIN
389-390	0.70	+	7	2	PE-PGRSFAMILYPROTEIN, acyl-CoAsynthase, enoyl-CoAhydratase, PROBABLECYTOCHROME P450MONOOXYGENASE, PROBABLECYTOCH ROME P450MONOOXYGENASE
394-395	0.68	+	8	0	PROBABLETRANSCRIPTIONALREGUL ATORYPROTEIN, PPEFAMILYPROTEIN, shortchaindehydrogenase, PROBABLEACYL- COADEHYDROGENASEFADE30, acyl-CoAsynthase, PROBABLEACYL- COADEHYDROGENASEFADE31, PROBABLEACYL- COADEHYDROGENASEFADE32, PROBABLEACYL- COADEHYDROGENASEFADE33
398-399	0.66	+	7	1	PROBABLEATP-DEPENDENTCLPPROTEASE, PROBABLELSR2PROTEINPRECURSOR, lysyl- tRNA synthetase, aspartate 1-decarboxylaseprecursor, pantoate--beta-alanine ligase, CONSERVEDHYPO THETICALALANINEAND
426-427	0.64	+	9	2	POSSIBLEMEMBRANEPROTEIN, POSSIBLEHISTONE-LIKEPROTEINHNS, ribonucleaseactivityregulatorprotein, MONOOXYGENASEETHA, TRANSCRIP TIONALREGULATORYREPRESSORPRO TEIN, POSSIBLEMEMBRANEPROTEIN, PUTATIVENADH-DEPENDENTGLUTAMATESY NTHASE

343-344	0.62	-	10	2	PROBABLENADPH:ADRENODOXINOXIDOREDUCTASEFPRA, POSSIBLEALKYLDIHYDROXYACETONEPHOSPHATESYNTHASEAGPS, PROBABLEMOLYBDENUMCOFACTORBIOSYNTHESIS, PROBABLEPTERIN-4-ALPHA-CARBINOLAMINEDEHYDRATASEMOAB1, molybdenumcofactorbiosynthesisprotein, PROBABLEMOLYBDENUMCOFACTORBIOSYNTHESIS, POSSIBLEPHOSPHATASE, PROBABLETRANSPOSASE
<i>M. tuberculosis</i> CDC1551					
33-34	0.72	+	11	9	PE_PGRS (15839660), PPE (15839665)
35-36	0.67	+	9	5	PEfamilyprotein, secretedantigen,putative, subtilasefamilyprotein, trans-aconitatemethyltransferase
36-37	0.66	+	6	2	PE_PGRSfamilyprotein, DNA-bindingprotein, CopGfamily, transcriptionalregulator,TetRfamily, oxidoreductase,short-chaindehydrogenase/reductasefamily
37-38	0.62	+			
42-43	0.63	+	5	0	heat shock protein (grpE) (15839737), heat shock protein (dnaJ) (15839738), transcriptional regulator HspR(15839739), PPE (15839740), PPE (15839741)
43-44	0.64	+	7	5	adenylosuccinate synthetase (15839743), divalent cation transporter (15839748)
67-68	0.71	+	5	2	nitroreductase,cobalaminbiosynthesisprotein, PAP2superfamilyprotein, baiEprotein
83-84	0.69	+	15	12	transcriptionalregulator,MarRfamily, IS1557transposase, PE_PGRSfamilyprotein
84-85	0.70	+	9	5	PE_PGRSfamilyprotein, 3-hydroxyisobutyratedehydrogenase, acyl-CoAdehydrogenase, methylmalonicacidsemialdehydedehydrogenase
85-86	0.63	+	10	3	PPEfamilyprotein, DNA-bindingresponseregulator, sensorhistidinekinase, HITfamilyprotein, steroid isomerase,putative, zinc-bindingdehydrogenase, ferredoxin-relatedprotein
92-93	0.71	+	7	3	transcriptionalregulator,ArsRfamily, truncatedIS1605transposase, PE_PGRSfamilyprotein, PE_PGRSfamilyprotein
96-97	0.69	+	10	5	molybdenumcofactorbiosynthesisprotein, molybdopterinbiosynthesisMogprotein, molybdopterincofactorbiosynthesisprotein, molybdenumcofactorbiosynthesisprotein, cold-shockdomainfamilyprotein
97-98	0.66	+	9	7	acyl-CoAdehydrogenase,putative, PPEfamilyprotein
109-110	0.69	+	5	0	PE_PGRSfamilyprotein, 50SribosomalproteinL32, PE_PGRSfamilyprotein, DNA-bindingresponseregulator, sensorhistidinekinase
121-122	0.69	+	9	3	undecaprenyldiphosphatesynthase, PEfamilyprotein, PEfamilyprotein, cellulase-relatedprotein, PE_PGRSfamilyprotein, pantothenatekinase

126-127	0.66	+	12	7	PPEfamilyprotein, ketoacyl-CoAthiolase-relatedprotein, enoyl-CoAhydratase/isomerasefamilyprotein, enoyl-CoAhydratase, enoyl-CoAhydratase
163-164	0.73	+	6	2	PE_PGRS (15840909), cytochrome c oxidase folding protein (15840911), PE_PGRS (15840912), quinone oxidoreductase (15840914)
197-198	0.63	+	6	4	PPEfamilyprotein, phospholipaseC
198-199	0.67	+	5	1	molybdopterinoxidoreductase, membraneprotein,M mpLfamily, IS6110,transposase, serineesterase,cutinasefamily
204-205	0.66	+	7	2	PPEfamilyprotein, PE_PGRSfamilyprotein, PEfamilyprotein, PPEfamilyprotein, PPEfamilyprotein
216-217	0.60	+	5	1	PPE (15841389), PPE (15841389) acyltransferase family protein, lipoprotein (15841392)
230-231	0.68	+	1	1	
241-242	0.71	+	6	0	N-acetylglucosaminytransferase, celldivisionproteinFtsW, UDP-N-acetylmuramoyl-L-alanyl-D-glutamatesynthetase, phospho-N-acetylmuramoyl-pentapeptide-transferase, UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-alanylligase, UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelateligase
246-247	0.66	+	9	4	cytochromecoxidase,subunit, asparaginesynthetase,putative, carbohydratekinase,PfkBfamily, HesB/YadR/YfhFfamilyprotein, nicotinate-nucleotide--dimethylb enzimidazolephosphoribosyltransferase
279-280	0.73	+	3	1	PE_PGRSfamilyprotein, transcriptionalregulator,LuxRfamily
293-294	0.66	+	8	0	pyridoxamine5'-phosphateoxidase, PPEfamilyprotein, MutT/nudixfamilyprotein, glycosyltransferase, lipidAbiosynthesislauroyl, CDP-diacylglycerol--glycerol-3-phosphate3-phosphatidyltransferase,putative, HITfamilyprotein, threonyl-tRNAsynthetase
315-316	0.68	+	7	1	cobyrinicacida,c-diamidesynthase, cob(I)yrinicacida,c-diamideadenosyltransferase, magnesiumchelataase,putative, ElaAfamilyprotein, malate:quinoneoxidoreductase, PE_PGRSfamilyprotein
351-352	0.66	+	8	0	ATPsynthasesubunitE, NADHdehydrogenaseI,F, NADHdehydrogenasegammassubunit, NADHdehydrogenasesubunitH, NADHdehydrogenasesubunitI, NADHdehydrogenasesubunitJ, NADHdehydrogenasesubunitL, NADHdehydrogenasesubunitL
352-353	0.63	+	6	1	NADHdehydrogenasesubunitN, PPEfamilyprotein, PPEfamilyprotein, transcriptionalregulator,TetRfamily, Rieske2Fe-2Sfamilyprotein
373-374	0.71	+	3	1	IS1608', transposase (15842945), IS1561', transposase (15842946)
374-375	0.64	+	1	1	

375-376	0.64	+	0	0	
379-380	0.69	+	9	2	IS1560transposase, IS1560transposase, PE_PGRSfamilyprotein, MaoCfamilyprotein, lipoprotein,putative, shortchaindehydrogenase, cyclopropane-fatty-acyl-phospholipidsynthase1
387-388	0.66	+	17	15	integrase,putative, bacteriophageprotein
392-393	0.78	+	2	0	PE_PGRS (15843119), PE_PGRS (15843120)
393-394	0.74	+	2	1	fatty-acid-CoAligase-relatedprotein
394-395	0.71	+	7	3	acyl-CoAsynthase, enoyl-CoAhydratase, P450heme-thiolateprotein, N5,N10-methylenetetrahydromethanopterinreductase-relatedprotein
399-400	0.68	+	7	0	PPEfamilyprotein, shortchaindehydrogenase, acyl-CoAdehydrogenase,putative, acyl-CoAsynthase, acyl-CoAdehydrogenase,putative, acyl-CoAdehydrogenase,putative, acyl-CoAdehydrogenase,putative
402-403	0.70	+	9	5	carbonicanhydrase, A/G-specificadenineglycosylase,putative, PE_PGRSfamilyprotein, hydrolase,alpha/betahydrolasefold
403-404	0.65	+	8	1	ATP-dependentClpprotease,ATP-binding, lsr2protein, lysyl-tRNAsynthetase, transcriptionalactivator,putative,Baf, aspartate1-decarboxylaseprecursor, pantoate-beta-alanineligase, chalcone/stilbenesynthasefamilyprotein
207-208	0.65	-	6	3	glycinedehydrogenase, haloalkanedehalogenase, PEG1/MESTprotein

M. tuberculosis H37Rv

33-34	0.72	+	7	3	acyl-CoA synthase (fadD27) (15607416), PE_PGRS (15607419), PE_PGRS (15607420), PPE (15607421)
35-36	0.67	+	9	3	ESAT-6LIKEPROTEINESXG, LOWMOLECULARWEIGHTEIGHTPROTEIN, PROBABLECONSERVEDTRANSMEMBRANEPROTEIN, PROBABLEMEMBRANE-ANCHOREDMYCOSINMYCP3, PROBABLECONSERVEDTRANSMEMBRANEPROTEIN, PROBABLETRANS-ACONITATEMETHYLTRANSFERASE
36-37	0.66	+	7	4	PE-PGRSFAMILYPROTEIN, PROBABLETRANSCRIPTIONALREGULATORYPROTEIN, PROBABLEDEHYDROGENASE/REDUCTASE
37-38	0.62	+	6	2	PPEFAMILYPROTEIN, PUTATIVEOXIDOREDUCTASE, PROBABLECONSERVEDINTEGRALMEMBRANE, POSSIBLECONSERVEDEXPORTEDPROTEIN
42-43	0.63	+	5	0	heat shock protein (grpE) (15607492), heat shock protein (dnaJ) (15607493), heat shock protein (hspR) (15607494), PPE (15607495), PPE (15607496)
43-44	0.64	+	7	5	adenylosuccinate synthase (purA) (15607498), magnesium ion transporter (mgtE) (15607503)
67-68	0.70	+	7	5	PE-PGRSFAMILYPROTEIN, PROBABLECONSERVEDLIPOPROTEINLPQN
83-84	0.71	+	10	7	PE_PGRS (15607882), PE_PGRS (15607886), PE_PGRS (15607887)

84-85	0.67	+	8	4	PROBABLE3-HYDROXYISOBUTYRATEDEHYDROGENASEMMSB, PROBABLEACYL-CoADEHYDROGENASEFADE9, PROBABLEMETHYLMALONATE-SEMIALDEHYDEDEHYDROGENASEMMSA, PE-PGRSFAMILYPROTEIN
92-93	0.71	+	7	2	PROBABLETRANSCRIPTIONALREGULATORYPROTEIN, POSSIBLEDEAMINASE, POSSIBLETRANSPOSASE(FRAGMENT), PE-PGRSFAMILYPROTEIN, PE-PGRSFAMILYPROTEIN
96-97	0.69	+	10	2	molybdenumcofactorbiosynthesisprotein, PROBABLEMOLYBDOPTERINBIOSYNTHESISMOG, PROBABLEMOLYBDENUMCOFACTORBIOSYNTHESIS, POSSIBLERESUSCITATION-PROMOTINGFACTORRPFA, PROBABLEMOLYBDENUMCOFACTORBIOSYNTHESIS, molybdenumcofactorbiosynthesisprotein, POSSIBLECONSERVEDINTEGRALMEMBRANE, PROBABLECOLDSHOCK-LIKEPROTEIN
97-98	0.66	+	8	2	PROBABLEACYL-CoADEHYDROGENASEFADE10, POSSIBLECONSERVEDEXPORTEDPROTEIN, POSSIBLECONSERVEDTRANSMEMBRANEPROTEIN, PPEFAMILYPROTEIN, POSSIBLECONSERVEDTRANSMEMBRANEPROTEIN, POSSIBLETRANSCRIPTIONALREGULATORYPROTEIN
109-110	0.69	+	7	1	PE-PGRSFAMILYPROTEIN, PE-PGRSFAMILYPROTEIN, 50SribosomalproteinL32, PE-PGRSFAMILYPROTEIN, MYCOBACTERIALPERSISTENCEREGULATORMRPA, PROBABLETWO COMPONENTSENSOR
121-122	0.71	+	8	1	SHORT(C15)CHAINZ-ISOPRENYL, PE-PGRSFAMILYPROTEIN, PEFAMILYPROTEIN, PEFAMILYPROTEIN, PROBABLECELLULASECELA2A(ENDO-1,4-BETA-GLUCANASE), PROBABLECELLULASECELA2B(ENDO-1,4-BETA-GLUCANASE), PE-PGRSFAMILYPROTEIN
126-127	0.66	+	10	3	PPEFAMILYPROTEIN, POSSIBLEACETYL-CoAACETYLTRANSFERASE(ACETOACETYL-CoA, POSSIBLEENOYL-CoAHYDRATASE, POSSIBLEOXIDOREDUCTASE, PROBABLEINTEGRALMEMBRANEPROTEIN, enoyl-CoAhydratase, enoyl-CoAhydratase
163-164	0.73	+	5	1	PE_PGRS (15608588), cytochrome c oxidase assembly factor (ctaB) (15608589), PE_PGRS (15608590), quinone oxidoreductase (qor) (15608592)
198-199	0.63	+	7	2	PPEFAMILYPROTEIN, PROBABLEPHOSPHOLIPASEC4, PUTATIVETRANSPOSASE, PUTATIVETRANSPOSASE, PROBABLECUTINASECUT1
216-217	0.61	+	3	0	PROBABLEISOCITRATATELYASEeaceAa, isocitratatelyase, PPEFAMILYPROTEIN
229-230	0.67	+	2	1	ProbablelipoproteinlppI

263-264	0.63	+	6	0	PROBABLEMEMBRANE-ASSOCIATEDPHOSPHOLIPASEC, PPEFAMILYPROTEIN, PPEFAMILYPROTEIN, PROBABLETRANSPOSASE, PROBABLETRANSPOSASE, PPEFAMILYPROTEIN
280-281	0.70	+	7	4	HYPOTHETICALALANINERICHPROTEIN, PE-PGRSFAMILYPROTEIN, dihydrolipoamideacetyltransferase
284-285	0.69	+	1	0	PROBABLEFATTYACIDSYNTHASE
293-294	0.66	+	11	4	PROBABLEACYL-CoATHIOESTERASEII, pyridoxinebiosynthesisprotein, pyridoxamine5'-phosphateoxidase, PPEFAMILYPROTEIN, PROBABLECONSERVEDMEMBRANEPROTEIN, ALPHAMANNOSYLTRANSFERASEPIMA, lipidAbiosynthesislauroyl
316-317	0.66	+	7	2	malate:quinoneoxidoreductase, PE-PGRSFAMILYPROTEIN, dihydrolipoamidedehydrogenase, POSSIBLENICKEL-TRANSPORTINTEGRALMEMBRANE, shortchaindehydrogenase
320-321	0.68	+	9	4	PPEFAMILYPROTEIN, POSSIBLEOXIDOREDUCTASE, tyrosinerecombinase, POSSIBLEMYCOBACTINUTILIZATIONPROTEIN, formatedehydrogenaseaccessoryprotein
337-338	0.67	+	10	2	aspartyl/glutamyl-tRNAamidotransferasesubunitC, DNALigase, PROBABLELIPOPROTEINLPQA, ESAT-6LIKEPROTEINESXQ, PPEFAMILYPROTEIN, PEFAMILYPROTEIN, SECRETEDESAT-6LIKEPROTEIN, ESAT-6LIKEPROTEINESXS
373-374	0.67	+	2	0	PE_PGRS (15610480), PE_PGRS (15610481)
374-375	0.67	+	2	1	PPE (15610483)
375-376	0.65	+	3	2	PPE (15610486)
376-377	0.65	+	6	5	methylenetetrahydrofolate dehydrogenase (fold) (15610492)
380-381	0.69	+	8	0	POSSIBLETRANSPOSASE, POSSIBLETRANSPOSASE, PE-PGRSFAMILYPROTEIN, POSSIBLEDEHYDROGENASE, PROBABLECONSERVEDLIPOPROTEINLPQD, shortchaindehydrogenase, CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPIDSYNTHASE1CMAA1, PROBABLENUCLEOSIDEHYDROLASEIUNH
392-393	0.70	+	6	0	CONSERVEDHYPOTHETICALINTEGRALMEMBRANE, 3-ketoacyl-(acyl-carrier-protein)reductase, PROBABLEFERREDOXINFDXD, PROBABLEACYL-CoADEHYDROGENASEFADE26, PROBABLEACYL-CoADEHYDROGENASEFADE27, acyl-CoAsynthase
393-394	0.75	+	4	1	PE_PGRS (15610644), probable acetohydroxyacid synthase I large subunit (ilvX) (15610645), PE_PGRS(15610647)

394-395	0.79	+	3	0	PE_PGRS (15610648), acyl-CoA synthase (fadD18) (15610649), PE_PGRS (15610650)
397-398	0.66	+	7	0	PPEFAMILYPROTEIN, 4-hydroxy-2-ketovaleratealdolase, acetaldehydedehydrogenase, PROBABLEHYDRATASE, 3-ketosteroid-delta-1-dehydrogenase, PROBABLEDEHYDROGENASE, PPEFAMILYPROTEIN
403-404	0.69	+	7	2	PROBABLEADENINEGLYCOSYLASEMUTY, PE-PGRSFAMILYPROTEIN, POSSIBLEHYDROLASE, PROBABLECONSERVEDLIPOPROTEINLPQF, PE-PGRSFAMILYPROTEIN

Supplementary Material 2

The distribution of microsatellite motifs of sizes di-hexa according to their GC content.

GC% of the repeat motif	Mycobacterium avium	Mycobacterium bovis	Mycobacterium tuberculosis H37Rv	Mycobacterium tuberculosis CDC 1551	Mycobacterium leprae
			Di		
0.00	1637	4164	4227	4229	7707
0.50	60293	62571	63590	63442	59404
1.00	101590	71466	72409	72362	32346
			Tri		
0.00	124	216	218	217	1095
0.17	0	0	0	1	0
0.33	12451	12359	12566	12537	13386
0.67	54113	41371	41822	41900	26829
1.00	47159	34623	35062	34911	10845
			Tetra		
0.00	12	21	21	21	142
0.25	267	495	493	495	1128
0.50	1805	2516	2533	2539	3062
0.75	5526	5721	5810	5794	4002
1.00	8387	4481	4524	4516	1119
			Penta		
0.00	1	0	0	0	42
0.20	16	22	23	24	36
0.40	186	274	295	281	548
0.60	840	1081	1137	1117	1139
0.80	2157	1809	1828	1832	902
1.00	2920	1170	1183	1180	271
			Hexa		
0.00	0	0	0	0	13
0.17	1	5	5	5	62
0.33	43	53	52	52	151
0.50	453	413	414	416	339
0.67	1295	862	877	873	427
0.83	1941	1051	1045	1058	288
1.00	729	364	375	367	48