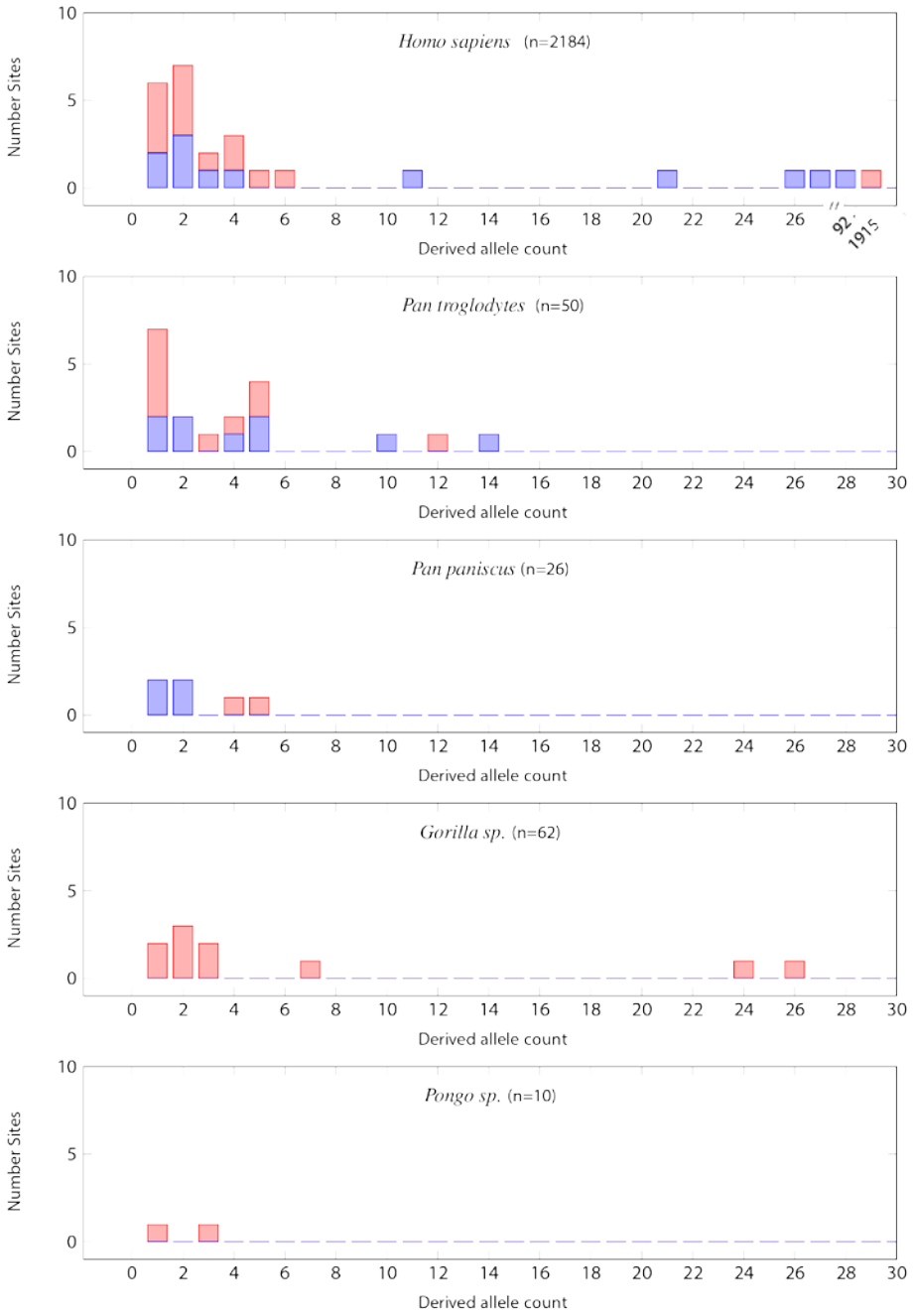
Supplementary Material

Supplementary Table 1: The rs6761637 SNP is in LD with other polymorphisms associated with infection

SNP	R^2	χ^2	*p*-value	Source	Location	Phenotype
rs2278589	0.2914	60.0213	**<0.0001**	CHB	Intron 1	*Mtb*+
rs6751745	0.3729	76.9289	**<0.0001**	CHB	Intron 13	*Mtb*+
rs17009726	0.932	191.9832	**<0.0001**	CHB	Intron 1	*Mtb*+
rs12998782	0.6245	128.6414	**<0.0001**	CHB	Intron 1	*Mtb*+
rs13389814	0.7092	160.2789	**<0.0001**	GWD	Intron 16	*Mtb*+
rs12998782	0.0141	3.1829	0.0744	GWD	Intron 1	*Mtb*+
rs4491733	0.0051	1.144	0.2848	GWD	Intron 1	*Mtb*+
rs7559955	0.0141	3.1829	0.0744	GWD	Intron 1	*Mtb*-
rs41279766	0.0001	0.0744	0.7851	EUR	Exon3 (Spacer)	Sepsis

Abbreviations: CHB = Han Chinese in Beijing (n=120), GWD = Gambian in Western Africa (n=179), EUR = European (n=535), *Mtb*+ = Pulmonary Tuberculosis Susceptibility, *Mtb*- = Pulmonary Tuberculosis Resistance.

1. Footnotes: rs2278589, and rs6751745 were linked to *Mtb* susceptibility by Thuong *et al* (2016). rs17009726 was linked to *Mtb* susceptibility by Ma *et al* (2011). rs12998782 was linked to *Mtb* susceptibility by Lao *et al* (2017). rs13389814, rs12998782, rs4491733, and rs7559955 were linked to *Mtb* susceptibility and resistance by Bowdish *et al* (2013). rs41279766 was linked to sepsis in COPD patients by Thomsen *et al* (2012).



Supplementary Figure 1: A site frequency spectrum of MARCO SNPs across primate species. The site frequency spectra for MARCO are shown for each of the primate species. Red squares indicate non-synonymous changes while the blue squares indicate synonymous changes. The largest number of derived alleles in humans is at residue 282, with 1915 derived alleles. Note that the species have not been randomly sampled and hence these cannot be considered proper spectra. The y-axes each indicate the count of the number of observed sites with the observed number of variant alleles. The sample sizes are shown in parentheses. The human spectrum has a discontinuous x-axis.

Supplementary Table 2: Human variation within MARCO.

'Population'	11969906-Syn	11969932-Non	11969959-Non	11972670-Syn	11972706-Syn	11972771-Non	11972793-Non	11972815-Non	11972909-Non	11973204-Syn	11973235-Non	11973239-Non	11973963-Non	11973977-Non	11973978-Non	11973954-Syn	11973984-Syn	11973999-Syn	11973917-Syn	11974803-Syn	11975015-syn	11975197-syn	11975206-Syn	11975229-Non	11975290-Syn	11975291-Non	Total	θ_H	H	θ_W	θ_π	D
ASW	0	0	1	0	0	1	0	0	0	0	0	0	87	0	0	4	0	0	17	0	0	0	1	0	0	0	122	1.07	-0.30	1.12	0.77	-0.68
CEU	1	0	0	0	0	0	1	0	0	0	0	0	161	0	0	0	0	0	0	0	0	6	6	0	0	0	170	1.81	-1.55	0.88	0.26	-1.38
CHB	0	0	0	0	0	0	0	1	1	0	0	0	160	0	5	0	0	0	0	0	0	0	0	1	0	0	194	1.37	-1.00	0.86	0.37	-1.09
CHS	0	0	0	0	2	0	0	1	0	0	0	0	170	0	1	0	0	0	0	1	1	0	0	0	0	0	200	1.45	-1.14	1.02	0.32	-1.40
CLM	1	0	0	0	0	0	0	0	0	0	0	0	115	0	0	1	0	0	2	0	0	8	1	0	0	0	120	1.86	-1.57	1.12	0.29	-1.62*
FIN	0	0	0	0	0	0	0	0	0	0	0	1	183	0	0	0	0	0	0	0	0	3	2	0	1	0	186	1.95	-1.84	0.86	0.11	-1.70*
GBR	0	0	0	0	0	0	2	0	0	0	0	0	170	0	0	0	0	0	0	0	0	0	5	0	0	1	178	1.84	-1.66	0.69	0.17	-1.35
IBS	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	1	1	0	0	0	28	1.79	-1.51	0.77	0.28	-1.53
JPT	0	0	0	0	0	0	0	0	0	0	0	0	170	0	0	0	1	0	0	0	0	0	0	0	0	1	178	1.83	-1.73*	0.52	0.11	-1.29
LWK	0	5	3	0	1	1	0	0	0	0	0	0	145	3	0	5	0	5	44	0	0	0	4	0	0	0	194	1.23	-0.23	1.71	1.01	-0.97
MXL	0	0	0	4	0	0	1	0	0	0	2	0	123	0	0	0	0	0	2	0	0	4	0	0	0	0	132	1.75	-1.43	1.10	0.32	-1.52
PUR	0	0	0	0	0	0	0	0	0	2	0	0	104	0	0	0	0	0	1	0	0	1	3	0	1	0	110	1.81	-1.56	1.14	0.25	-1.73*
TSI	0	0	0	0	0	0	0	0	0	0	0	0	186	0	0	0	0	0	2	0	0	4	3	0	0	0	196	1.81	-1.58	1.03	0.23	-1.58*
YRI	0	0	0	0	0	0	0	0	0	0	0	0	115	0	0	11	0	6	24	0	0	0	0	0	0	0	176	0.91	-0.03	0.70	0.88	0.47
AFR	0	5	4	0	1	2	0	0	0	0	0	0	347	3	0	20	0	11	85	0	0	0	5	0	0	0	492	1.06	-0.16	1.48	0.91	-0.82
AMR	1	0	0	4	0	0	1	0	0	2	2	0	342	0	0	1	0	0	5	0	0	13	4	0	1	0	362	1.79	-1.50	1.70	0.29	-1.86*
EAS	0	0	0	0	2	0	0	2	1	0	0	0	500	0	6	0	1	0	0	1	1	0	0	1	0	1	572	1.53	-1.25	1.44	0.28	-1.69*
EUR	1	0	0	0	0	0	3	0	0	0	0	1	726	0	0	0	0	0	2	0	0	14	17	0	1	1	758	1.84	-1.65	1.25	0.18	-1.68*
ALL	2	5	4	4	3	2	4	2	1	2	2	1	1915	3	6	21	1	11	92	1	1	27	26	1	2	2	2184	1.54	-1.12	3.15	0.42	-2.01*

The variation within MARCO was analyzed using the human genomes from PhaseI of the 1000 Human genome project (v3.20101123). The SNPs for each population along with estimates for Fay & Wu's estimate of θ_H, Fay & Wu's H statistic, Watterson's estimate of θ_W, Tajima's estimate of θ_π and Tajima's statistic D are shown. Simulations were carried out as described in the Methods section. Estimates of D that are significant ($\Pr < 0.025$; two-tailed) are shown by asterisks. One of the estimates of H is significant. The simulations are based on a simple basic neutral model, with constant population size, no recombination, no selection, random mating, and an infinite-sites model. However the actual history of the human population is complex, with a large number of demographic parameters that have not been explored. As such none of these 'populations' meet the assumptions of this test. Therefore we consider these tests to be only suggestive and justification to gather further experimental evidence. In the majority of the 'populations' the high frequency of the derived allele at position 282 (119739063bp in chromosome 2) stands out. Hence we further examined this SNP.

(The 'populations' listed are ASW - Americans of African Ancestry in SW USA, CEU - Utah Residents (CEPH) with Northern and Western European Ancestry, CHB - Han Chinese in Bejing, China, CHS - Southern Han Chinese, CLM - Colombians from Medellin, Colombia, FIN - Finnish in Finland, GBR - British in England and Scotland, IBS - Iberian Population in Spain, JPT - Japanese in Tokyo, Japan, LWK - Luhya in Webuye, Kenya, MXL - Mexican Ancestry from Los Angeles USA, PUR - Puerto Ricans from Puerto Rico, TSI - Toscani in Italia, YRI - Yoruba in Ibadan, Nigeria. The 'super populations' listed are AFR African, AMR Ad Mixed American, EAS East Asian, EUR European)