

Supplementary Materials FOR PUBLICATION

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

Table S1: A list of the soil and site characteristics for each of the 143 soil samples. This table also contains 16s rRNA and hsp65 amplicon read counts per sample, and the percent relative abundance of reads assigned to the *Mycobacterium* genus as determined by the 16S rRNA gene analyses. Abbreviations are as follows: *Clay* = *Clay+Silt (%)*; *OC* = *Organic carbon (%)*; *TN* = *Total nitrogen (%)*; *TPmgKg*=*Total phosphorus (mg P Kg)*; *CN* = *Carbon:Nitrogen ratio*; *pHwater* = *pH*; *AridityIndex* = *Aridity Index*; *MDR* = *Mean diurnal temperature range (°C)*; *MAXT* = *Maximum temperature (°C)*; *MINT* = *Minimum temperature (°C)*; *PSEA* = *Precipitation seasonality*; *UV=UV light*; *NPP20032015* = *Net primary productivity (NDVI index)*; *Ecosystem* = *Ecosystem type*; *ElectricalConductivityuScm* = *Electrical conductivity (μS cm)*; *Myco_16S_relabun* = *mycobacteria relative abundance*; *16S_reads* = *16S reads (pre-ordination)*; *hsp65_reads* = *hsp65 reads (pre presence/absence category)*.

(separate attachment)

23 **Table S2:** A table of Pearson correlations between pairs of each environmental variable
 24 measured. For abbreviation definitions, see Table S1.

	Var1	Var2	Correlation		Var1	Var2	Correlation
1	TN	OC	0.8131	52	UV	Altitude	0.1781
2	MINT	UV	0.8062	53	MAXT	PSEA	-0.1640
3	NPP20032015	AridityIndex	0.7842	54	Clay	AridityIndex	0.1626
4	NPP20032015	pHwater	-0.7249	55	PSEA	MINT	-0.1615
5	pHwater	AridityIndex	-0.7066	56	MINT	Altitude	-0.1606
6	TN	ECuScm	0.6770	57	PSEA	ECuScm	-0.1566
7	OC	ECuScm	0.6355	58	TN	Altitude	0.1556
8	Latitude	UV	-0.5940	59	OC	Altitude	0.1556
9	MAXT	pHwater	0.5767	60	UV	AridityIndex	0.1394
10	Latitude	MINT	-0.5623	61	MINT	AridityIndex	0.1380
11	NPP20032015	TN	0.5297	62	Longitude	CN	-0.1379
12	NPP20032015	OC	0.4927	63	CN	MINT	-0.1373
13	TN	AridityIndex	0.4701	64	Latitude	TN	0.1352
14	NPP20032015	ECuScm	0.4625	65	Longitude	PSEA	0.1252
15	pHwater	OC	-0.4571	66	PSEA	OC	-0.1230
16	ECuScm	AridityIndex	0.4536	67	MINT	pHwater	0.1190
17	OC	AridityIndex	0.4447	68	Longitude	Clay	-0.1188
18	MAXT	AridityIndex	-0.4263	69	Longitude	MINT	0.1179
19	pHwater	TN	-0.4148	70	pHwater	Clay	-0.1168
20	CN	pHwater	-0.4033	71	Latitude	Clay	0.1135
21	CN	OC	0.4024	72	OC	Clay	0.1075
22	Longitude	Latitude	-0.3856	73	CN	UV	-0.1074
23	MAXT	TN	-0.3833	74	TN	Clay	0.1074
24	MAXT	MINT	0.3712	75	PSEA	Altitude	-0.1047
25	MAXT	OC	-0.3711	76	NPP20032015	Altitude	-0.1021
26	CN	NPP20032015	0.3401	77	Latitude	pHwater	0.0988
27	Longitude	AridityIndex	-0.3340	78	pHwater	Altitude	0.0980
28	PSEA	NPP20032015	-0.3283	79	PSEA	TN	-0.0973
29	Longitude	NPP20032015	-0.3268	80	Latitude	NPP20032015	-0.0957
30	MAXT	NPP20032015	-0.3247	81	Latitude	AridityIndex	0.0874
31	Longitude	ECuScm	-0.3228	82	Latitude	OC	0.0842
32	Longitude	Altitude	-0.3226	83	ECuScm	Altitude	0.0788
33	pHwater	ECuScm	-0.3141	84	CN	Clay	0.0745
34	Longitude	MAXT	0.2877	85	MINT	OC	-0.0742
35	MAXT	Altitude	-0.2739	86	Altitude	AridityIndex	-0.0699
36	Longitude	TN	-0.2620	87	PSEA	pHwater	0.0585
37	MAXT	ECuScm	-0.2552	88	CN	TN	0.0576
38	CN	AridityIndex	0.2521	89	MINT	ECuScm	-0.0527
39	Longitude	OC	-0.2431	90	MINT	TN	-0.0516
40	NPP20032015	Clay	0.2399	91	Latitude	PSEA	-0.0514
41	Longitude	pHwater	0.2380	92	Latitude	MAXT	-0.0505
42	UV	NPP20032015	0.2246	93	Latitude	CN	0.0501
43	PSEA	AridityIndex	-0.2222	94	UV	Clay	-0.0444
44	MAXT	UV	0.2108	95	CN	Altitude	0.0388
45	CN	PSEA	-0.2100	96	PSEA	UV	0.0370
46	Latitude	Altitude	0.2042	97	Longitude	UV	-0.0354
47	ECuScm	Clay	0.2010	98	UV	ECuScm	-0.0291
48	MAXT	CN	-0.1953	99	MINT	Clay	0.0189
49	Altitude	Clay	-0.1925	100	UV	TN	-0.0183
50	Latitude	ECuScm	0.1887	101	UV	OC	-0.0155
51	MINT	NPP20032015	0.1884	102	PSEA	Clay	0.0139
52	CN	ECuScm	0.1798	103	UV	pHwater	-0.0090
				104	MAXT	Clay	-0.0076

26 **Table S3:** Output of Random Forest Models identifying environmental variables important for
27 predicting distributions of the *Mycobacterium* genus as well as the top 20 most ubiquitous
28 mycobacterial lineages. The table includes the percent variation described by the model overall
29 as well as the importance of the top 4 environmental predictors. For abbreviation definitions,
30 see Table S1. Four of these mycobacterial clades included described members: **Clade 10** (*M.*
31 *stomatepiae* DSM 45059 T, *M. genavense* DSM 44424 T, *M. florentinum* DSM 44852 T, *M.*
32 *lentiflavum* CIP 105465 T, *M. montefiorensis* DSM 44602 ATC, *M. triplex* ATCC 700071 CIP 106) ;
33 **Clade 12** (*M. novocastrensis* CIP 105546 T, *M. rutilum* czh-117 T); **Clade 31** (*M. arosiense* DSM
34 45069 T1921 A, *M. intracellulare* ATCC 13950 T, *M. yongonense* 05-1390 T, *M. avium* subsp.
35 *paratuberculosis*, *M. bouchedurhonense* CIP 109827, *M. avium* subsp. *silvaticum* ATC, *M.*
36 *colombiense* CIP 108962 T, *M. chimaera* CIP 107892, *M. timonense* CIP 109830 T, *M. vulneris*
37 *NLA000700772* T, *M. marseillense* 5356591 CIP 10, *M. avium* subsp. *avium* ATCC 252), and
38 **Clade 44** (*M. doricum* DSM 44339 T, *M. monacense* DSM 44395 T).

39

Group Modeled	% Variation Explained	Predictor 1	Percent Explained	p value	Predictor 2	Percent Explained	p value	Predictor 3	Percent Explained	p value	Predictor 4	Percent Explained	p value
16s Mycobacteria	25	MINT	31	****	MAXT	18	*	pHwater	15	*	AridityIndex	16	NS
clade 5	61	AridityIndex	35	****	MAXT	25	**	MINT	24	**	pHwater	20	*
clade 69	55	pHwater	33	****	AridityIndex	25	**	MINT	18	**	CN	12	NS
clade 29	50	AridityIndex	31	****	MINT	24	**	ElectricalCond	22	**	OC	16	NS
clade 10	43	AridityIndex	34	****	MINT	27	**	pHwater	23	**	MAXT	15	NS
clade 9	39	pHwater	15	NS	CN	11	NS	ElectricalCond	10	*	MINT	8	NS
clade 15	38	MINT	25	**	AridityIndex	23	**	pHwater	21	*	OC	14	NS
clade 24	33	MAXT	31	***	AridityIndex	25	**	MINT	21	**	pHwater	17	*
clade 1	30	MAXT	27	***	ElectricalCond	16	*	pHwater	12	*	AridityIndex	12	NS
clade 12	29	pHwater	40	****	MAXT	22	**	AridityIndex	21	****	MINT	21	**
clade 23	23	pHwater	24	**	AridityIndex	19	*	OC	18	*	MINT	17	*
clade 38	22	AridityIndex	19	*	MAXT	18	*	OC	18	NS	pHwater	14	NS
clade 21	22	AridityIndex	19	*	MAXT	17	*	MINT	14	*	pHwater	13	NS
clade 46	29	AridityIndex	19	*	pHwater	17	*	MINT	14	*	OC	12	NS
clade 13	14	pHwater	20	*	OC	19	*	AridityIndex	9	NS	MINT	7	NS
clade 44	3	ElectricalCond	14	NS	pHwater	12	NS	AridityIndex	11	NS	MAXT	6	NS
clade 3	2	pHwater	14	NS	AridityIndex	11	NS	OC	6	NS	MAXT	5	NS
clade 31	-4	pHwater	10	NS	AridityIndex	10	NS	CN	5	NS	MINT	1	NS
clade 54	-17	MAXT	14	NS	OC	13	NS	AridityIndex	12	NS	MINT	8	NS
clade 74	-13	MAXT	12	NS	AridityIndex	12	NS	OC	12	NS	pHwater	11	NS
clade 45	-8	pHwater	17	*	AridityIndex	15	NS	MAXT	11	NS	Clay	8	NS
p < 0.0001	****												
p < 0.001	***												
p < 0.01	**												
p < 0.05	*												
p > 0.05	NS												

40

41

42

43

44

45

46

47

48

49

50

51

52

53 **Table S4:** Summary of the statistical tests performed on mycobacterial clade presence and
 54 absence in different environmental conditions. Results of Spearman correlations and Wilcoxon
 55 rank-sum tests are included for the twenty most ubiquitous clades. Four of these mycobacterial
 56 clades included described members: Clade 10, Clade 12, Clade 31, Clade 44 (for detailed list of
 57 named species see Table S3).

Clade	soil pH				Aridity Index				Minimum Annual Temperature			
	pH Wilcoxon		pH Spearman Correlation		AI Wilcoxon		AI Spearman Correlation		MINT Wilcoxon		MINT Spearman Correlation	
	W	pval	spm R	spm pval	W	pval	spm R	spm pval	W	pval	spm R	spm pval
Clade 1	1352	0.037348	-0.17	0.037	620.5	0.010398	0.21	0.010	1229.5	0.192412	-0.11	0.192
Clade 3	839.5	0.418773	-0.07	0.419	461	0.042245	0.17	0.041	1174.5	0.000872	-0.28	0.001
Clade 5	2780	5.45E-09	-0.49	0.000	527.5	3.14E-08	0.46	0.000	2522	5.86E-06	-0.38	0.000
Clade 9	1140	0.000224	-0.31	0.000	221	0.000423	0.30	0.000	1066	0.001878	-0.26	0.002
Clade 10	2471	3.31E-08	-0.46	0.000	518	7.78E-07	0.41	0.000	969	0.011643	0.21	0.011
Clade 12	729.5	8.85E-13	0.60	0.000	3687	8.15E-07	-0.41	0.000	2259	0.368674	0.08	0.369
Clade 13	212	2.80E-05	0.35	0.000	1044	0.069028	-0.15	0.068	574	0.11583	0.13	0.115
Clade 15	2043.5	0.017671	-0.20	0.017	820	0.000102	0.33	0.000	2573	3.70E-07	-0.43	0.000
Clade 21	1457.5	2.43E-05	-0.35	0.000	205	6.66E-06	0.38	0.000	565	0.046128	0.17	0.045
Clade 23	2213	0.04331	-0.17	0.043	1166.5	0.002665	0.25	0.002	2778	2.17E-06	-0.40	0.000
Clade 24	2710.5	2.97E-05	-0.35	0.000	1031.5	0.000144	0.32	0.000	2416	0.005494	-0.23	0.005
Clade 29	2590	0.000713	-0.28	0.001	765.5	2.06E-07	0.43	0.000	3115.5	4.69E-09	-0.49	0.000
Clade 31	6900	2.26E-12	-0.41	0.000	1176	4.38E-10	0.37	0.000	2824	0.015427	0.14	0.087
Clade 38	751.5	0.524397	-0.05	0.524	398.5	0.03316	0.18	0.032	1071.5	0.001621	-0.26	0.001
Clade 44	559.5	3.46E-05	0.35	0.000	1994.5	6.97E-05	-0.33	0.000	1052	0.176003	0.11	0.176
Clade 45	984	0.001914	-0.26	0.002	276	0.006289	0.23	0.006	560.5	0.701078	0.03	0.699
Clade 46	1166.5	0.001083	-0.27	0.001	257	0.000363	0.30	0.000	1005.5	0.039634	-0.17	0.039
Clade 54	640	0.069159	0.15	0.068	941	0.837018	-0.02	0.835	1145.5	0.112952	-0.13	0.112
Clade 69	1835.5	2.53E-07	-0.43	0.000	384.5	4.85E-05	0.34	0.000	1475	0.004177	-0.24	0.004
Clade 74	880	0.265635	-0.09	0.265	406.5	0.014643	0.20	0.014	1105	0.005015	-0.23	0.005

58

59

60

61

62

63

64

65

66

67

68 **Table S5:** Summary of the Spearman partial correlations tests performed to determine the
69 relationship between mycobacterial clade presence/absence and latitude and longitude while
70 controlling for the effect of predictive environmental variables. Also included is the same test to
71 determine the relationship between mycobacterial clade presence/absence and environmental
72 variables while controlling for the effect of latitude and longitude. Matrix results of Spearman
73 partial correlations are included for the clades represented in Figure 4. Four of these
74 mycobacterial clades included described members: Clade 10, Clade 12, Clade 31, Clade 44 (for
75 detailed list of named species see Table S3).

76

Clade 12					pH					Clade 69							
Rs Estimate					Rs Estimate					Rs Estimate							
clade_12	clade_12	pH	Longitude(sin)	Longitude(cos)	Latitude	clade_5	clade_5	pH	Longitude(sin)	Longitude(cos)	Latitude	clade_69	clade_69	pH	Longitude(sin)	Longitude(cos)	Latitude
clade_12	1.00	0.59	-0.02	-0.02	0.17	clade_5	1.00	-0.44	0.14	-0.04	0.36	clade_69	1.00	-0.37	0.13	0.06	0.18
pH	0.59	1.00	0.39	-0.12	-0.22	pH	-0.44	1.00	-0.16	-0.17	0.03	pH	-0.37	1.00	-0.19	-0.13	-0.07
Longitude(sin)	-0.02	-0.19	1.00	-0.09	0.13	Longitude(sin)	0.14	-0.16	1.00	-0.09	0.07	Longitude(sin)	0.13	-0.19	1.00	-0.10	0.10
Longitude(cos)	-0.02	-0.12	-0.09	1.00	-0.21	Longitude(cos)	-0.04	-0.17	-0.09	1.00	-0.19	Longitude(cos)	0.06	-0.13	-0.10	1.00	-0.23
Latitude	0.17	-0.22	0.13	-0.21	1.00	Latitude	0.36	0.03	0.07	-0.19	1.00	Latitude	0.18	-0.07	0.10	-0.23	1.00
p value					p value					p value							
clade_12	0.000	0.00	0.824	0.782	0.045	clade_5	0.00	0.00	0.09	0.62	0.00	clade_69	0.00	0.00	0.12	0.48	0.04
pH	0.000	1.00	0.021	0.163	0.010	pH	0.00	0.00	0.05	0.05	0.69	pH	0.00	0.00	0.03	0.12	0.42
Longitude(sin)	0.824	0.02	0.000	0.276	0.125	Longitude(sin)	0.09	0.05	0.00	0.32	0.43	Longitude(sin)	0.12	0.03	0.00	0.24	0.23
Longitude(cos)	0.782	0.16	0.276	0.000	0.011	Longitude(cos)	0.62	0.05	0.32	0.00	0.02	Longitude(cos)	0.48	0.12	0.24	0.00	0.01
Latitude	0.045	0.01	0.125	0.011	0.000	Latitude	0.00	0.69	0.43	0.02	0.00	Latitude	0.04	0.42	0.23	0.01	0.00

Clade 10					Aridity Index					Clade 5							
Rs Estimate					Rs Estimate					Rs Estimate							
clade_10	clade_10	pH	Longitude(sin)	Longitude(cos)	Latitude	clade_29	clade_29	AridityIndex	Longitude(sin)	Longitude(cos)	Latitude	clade_5	clade_5	AridityIndex	Longitude(sin)	Longitude(cos)	Latitude
clade_10	1.00	-0.47	0.16	-0.10	-0.33	clade_29	1.00	0.30	0.02	0.06	0.37	clade_5	1.00	0.34	0.19	0.01	0.27
pH	-0.47	1.00	-0.34	-0.10	-0.28	AridityIndex	0.30	1.00	0.17	0.07	0.20	AridityIndex	0.34	1.00	0.10	0.08	0.23
Longitude(sin)	0.16	-0.34	1.00	-0.11	0.17	Longitude(sin)	0.02	0.17	1.00	-0.07	0.08	Longitude(sin)	0.19	0.10	1.00	-0.07	0.04
Longitude(cos)	0.10	-0.10	-0.11	1.00	-0.17	Longitude(cos)	0.06	0.07	-0.07	1.00	-0.22	Longitude(cos)	0.01	0.08	-0.07	1.00	-0.21
Latitude	-0.33	-0.28	0.17	-0.17	1.00	Latitude	0.37	0.20	0.08	-0.22	1.00	Latitude	0.27	0.23	0.04	-0.21	1.00
p value					p value					p value							
clade_10	0.00	0.00	0.05	0.25	0.00	clade_29	0.00	0.00	0.81	0.51	0.00	clade_5	0.00	0.00	0.02	0.94	0.00
pH	0.00	0.00	0.09	0.25	0.00	AridityIndex	0.00	0.00	0.05	0.44	0.02	AridityIndex	0.00	0.00	0.22	0.35	0.01
Longitude(sin)	0.05	0.09	0.00	0.21	0.04	Longitude(sin)	0.81	0.05	0.00	0.42	0.35	Longitude(sin)	0.02	0.22	0.00	0.43	0.67
Longitude(cos)	0.25	0.25	0.21	0.00	0.04	Longitude(cos)	0.51	0.44	0.42	0.00	0.01	Longitude(cos)	0.94	0.35	0.43	0.00	0.01
Latitude	0.00	0.00	0.04	0.04	0.00	Latitude	0.00	0.02	0.35	0.01	0.00	Latitude	0.00	0.01	0.67	0.01	0.00

Clade 69					Clade 10					Clade 31							
Rs Estimate					Rs Estimate					Rs Estimate							
clade_69	clade_69	AridityIndex	Longitude(sin)	Longitude(cos)	Latitude	clade_10	clade_10	AridityIndex	Longitude(sin)	Longitude(cos)	Latitude	clade_31	clade_31	AridityIndex	Longitude(sin)	Longitude(cos)	Latitude
clade_69	1.00	0.24	0.18	0.10	0.13	clade_10	1.00	0.52	0.20	0.14	-0.44	clade_31	1.00	0.44	0.06	0.04	-0.32
AridityIndex	0.24	1.00	0.13	0.06	0.31	AridityIndex	0.52	1.00	0.05	0.00	0.50	AridityIndex	0.44	1.00	0.14	0.06	0.44
Longitude(sin)	0.18	0.13	1.00	-0.08	0.07	Longitude(sin)	0.20	0.05	1.00	-0.09	-0.17	Longitude(sin)	0.06	0.14	1.00	-0.07	0.11
Longitude(cos)	0.10	0.06	-0.08	1.00	-0.23	Longitude(cos)	0.14	0.00	-0.09	1.00	-0.13	Longitude(cos)	0.04	0.06	-0.07	1.00	-0.19
Latitude	0.13	0.31	0.07	-0.23	1.00	Latitude	-0.44	0.50	0.17	-0.13	1.00	Latitude	-0.32	0.44	0.11	-0.19	1.00
p value					p value					p value							
clade_69	0.00	0.00	0.03	0.25	0.14	clade_10	0.00	0.00	0.02	0.10	0.00	clade_31	0.00	0.00	0.49	0.60	0.00
AridityIndex	0.00	0.00	0.12	0.48	0.00	AridityIndex	0.00	0.00	0.58	1.00	0.00	AridityIndex	0.00	0.00	0.10	0.50	0.00
Longitude(sin)	0.03	0.12	0.00	0.32	0.41	Longitude(sin)	0.02	0.58	0.00	0.27	0.04	Longitude(sin)	0.49	0.10	0.00	0.41	0.20
Longitude(cos)	0.25	0.48	0.32	0.00	0.01	Longitude(cos)	0.10	1.00	0.27	0.00	0.12	Longitude(cos)	0.60	0.50	0.41	0.00	0.02
Latitude	0.14	0.00	0.41	0.01	0.00	Latitude	0.00	0.00	0.04	0.12	0.00	Latitude	0.00	0.00	0.20	0.02	0.00

78

79

80

81

82

83

84

85

86

87

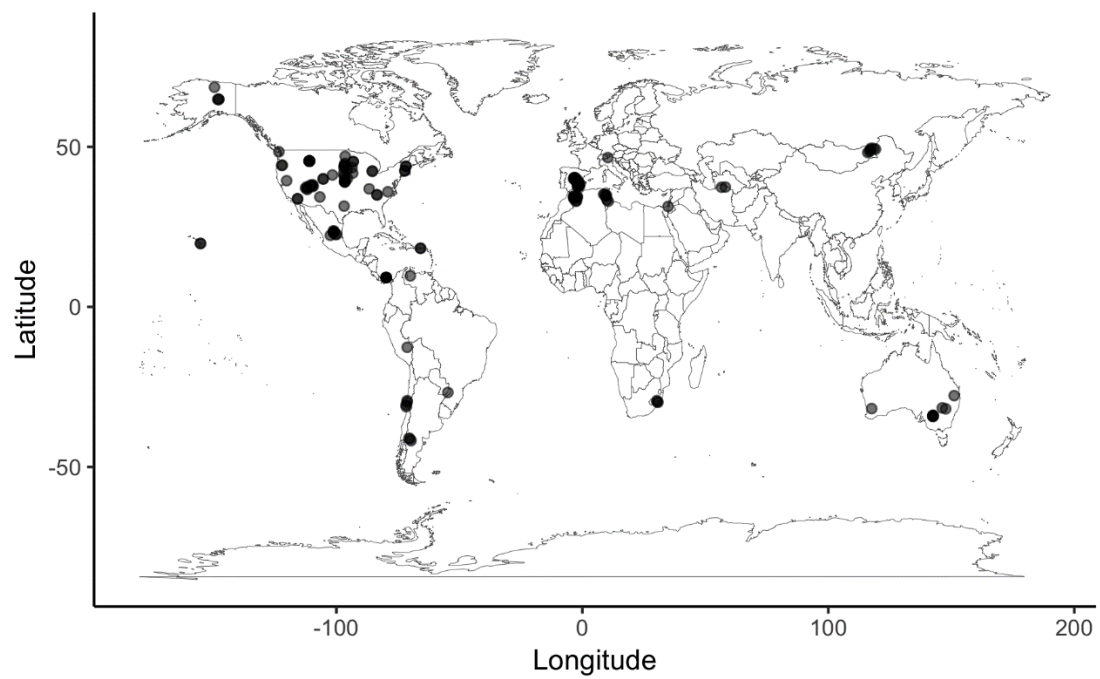
88

89

90

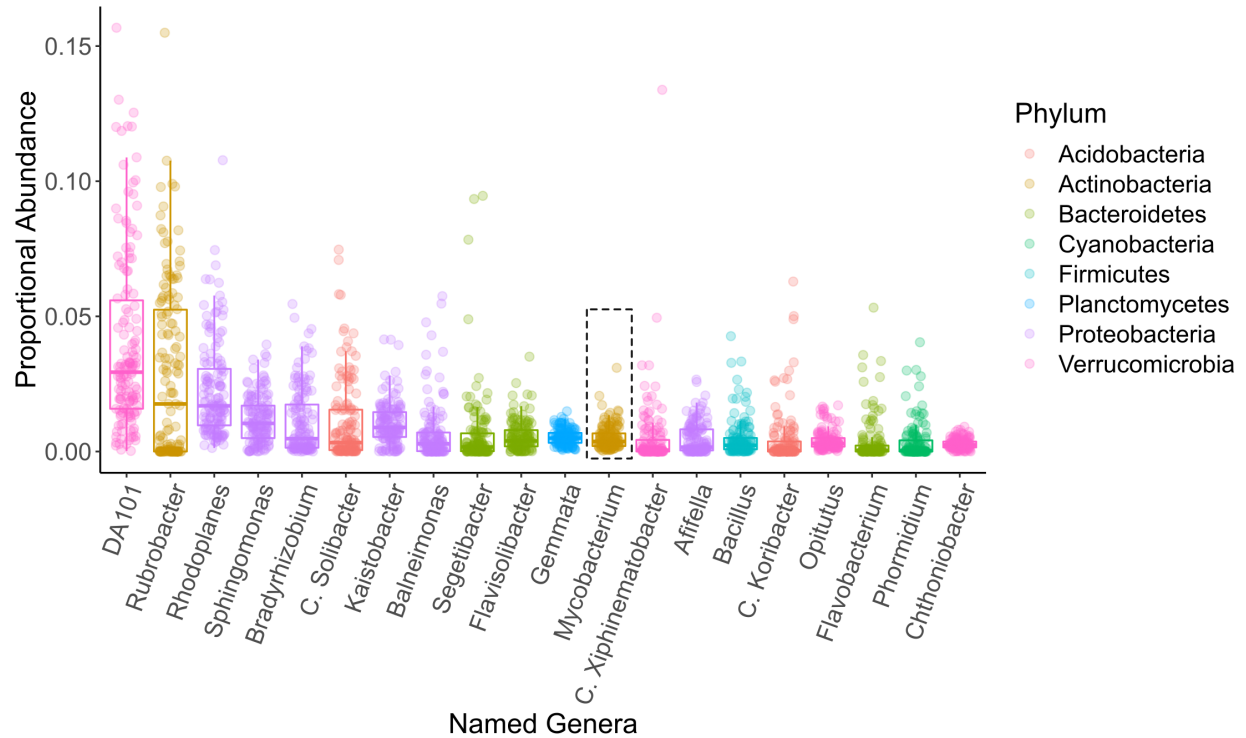
91

92



93

94 **FIGURE S1:** Map of soil sampling locations across the globe (n=143). Darker dots represent
95 overlap of nearby sampling sites.



96

97

98 **Figure S2: The top twenty most abundant named genera across 143 soil samples, ordered by**

99 **average relative abundance.** Dots represent the relative abundance of each genus per

100 individual sample, as identified by 16srRNA gene sequencing. The dot color represents the

101 phylum-level classification of each genus. The relative abundance of mycobacteria ranges

102 from 0.03% to 2.9% with an average of 0.52%, making it the 12th most abundant genus on

103 average.

104

105

106