

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Truman RW, Singh P, Sharma R, et al. Probable zoonotic leprosy in the southern United States. N Engl J Med 2011;364:1626-33.

Supplementary Data

Table S1: Demographics and genotype of *M. leprae* from humans and armadillos in US (Excel sheet). The reads from NHDP-55, NHDP-98 and I-30 resequencing analysis may be found in the trace depository under SRP001064.

Table S2. Primers used for PCR and genotyping

Locus	Sequence (5'-3')
SNP-7614-F	CGC TGA GAC GAT GGG CAA TTA CC
SNP-7614-R	TTG CCC ATC AAC CAA GGG ATA CCG
SNP-10505-F	CCA ATT TGC TTG CAT TCC A
SNP-10505-R	AAC ACT GCA CCG ACC AAC TC
InDel_17915-F	GGG GAG GCG AGT ATA GAT CC
InDel_17915-R	AGG CGC TGG TCT GCT ATC TA
SNP-22488-F	ATG CTC TGC GGT ACC AGT CT
SNP-22488-R	AGC CCT ACC TGA TCG ACA AC
SNP-29563-F	AGC GCC AAA CAG CAC TAA AG
SNP-29563-R	TCT GAC GAT ATG GCG ACA AC
SNP-365411-F	CGA GGT ATT TAC CGC AAT GG
SNP-365411-R	CGT GGT TAG GGT GAC AAC AC
SNP-38202-F	GGG TGC ACG TAG AGT TGG AT
SNP-38202-R	CTC CGC CGA CGT ACA AAT
SNP-45791-F	TGG GAA TTC TTA TAA ATC CAT TCT G
SNP-45791-R	AGA GGG TTG AAA TTG GCA CA
SNP-116225-F	GCT TTC CAT TCA CGA CCA G
SNP-116225-R	AAG TTC GAT CGG TTC ACC AC
INDEL-408352-F	TGT TGG GCA GTA GCA GGT TA
INDEL-408352-R	ATC CAG ACG CTA TTG GCA AC
SNP-574640-F	CGT TCG ACA AGC TCA AGG TC
SNP-574640-R	GCC GAT CAA TCC CAT GAG TA
SNP-604326-F	GCC CAG TTA AGC GGT CTT TA
SNP-604326-R	AGG TCC CGT CAA GAA CCA G
SNP-623366-F	ATG GCT ACC GTG TGC AAT G
SNP-623366-R	TCT CCA AGG TGA GCT CGA TT
SNP-1017410-F	ATC CTG GGA TTG GAG GAT TC
SNP-1017410-R	TGT TCC TCA CTG CGA TGT CT
SNP-1113926-F	CCT CAT TTG CTG TGG ACC TCA GTG
SNP-1113926-R	ATG TCA TCT TCC AGG GCG AAA AGC
SNP-1164294-F	CGA CGT GGG TAC AGT CGA T
SNP-1164294-R	GGT AGT CAC CCG GAG CAC
SNP-1291021-F	GTA GTT CGT TAC CGC CAG GA
SNP-1291021-R	TGA GTT TTG GTC GTG ACC TG
SNP-1450891-F	CAA TAC TGC GGT CGT CTC AA
SNP-1450891-R	AAA TGA CCA TCA CGA CGA CA
SNP-1527056-F	TTA AGC CAA TAT CGC CGA AC
SNP-1527056-R	GTT GAC AAC GGC TGA CAG AT
SNP-156948-F	ACA CCA CCA AGC CGT ATT GT
SNP-156948-R	CTT TTC CTG TTG GGT TGG TG
SNP-1588116-F	GCA GGT CCT TAC CGA TGA GA
SNP-1588116-R	GCT AGC GAC TCA ACT GAG CA
SNP-1734701-F	AGA GCG GAG TGT AGT CAC CAG
SNP-1734701-R	CTT CTT TGT GCC CGC ATA TC
SNP-1787203-F	GCG AGT TGA ACA CCA CGA C
SNP-1787203-R	CGC AGT TGA TCG GTA TTG G

SNP-1928726-F	AAC AGA TCG GCC ATG AAC TC
SNP-1928726-R	GAT CGA AGA GAC TCG GAT GC
SNP-193848 -R	CGC CAC CGT AGA ATT TGT TT
SNP-193848-F	CCC AGT CCT GTT CGT AAT CC
SNP-2200539-F	ATC ACA TTC GCT CCC CTT TT
SNP-2200539-R	AAA CGC CTG TTA AGC GAT TG
SNP-2999984-F	AGC TGG CTG CGA CTA CAA CT
SNP-2999984-R	CGA TCT TCG AGG ACT TGT CA
SNP-3110867-F	CCA GGG ATT ATT GTG CTG CT
SNP-3110867-R	GGG GAA TTG CTT TGT TAT CG
SNP-3112678-F	GGC ACA GGA GAC TCT GGT TG
SNP-3112678-R	CCA TCC ATT TAT TCC AGA GCA
INDEL-3224930-F	ATC ACC GGT AAC ATC CAC CA
INDEL-3224930-R	TAA CGA AGT GGG GAA ACA GG

Fragment spanning the reported SNP was amplified with set of primers and sequenced with reverse and forward primers. Sequence was aligned with TN reference genome to identify corresponding SNP variation in particular strain.

Table S3. Primers used for VNTR analysis			
VNTR locus		Sequence	Product size (bp) with TN strain
AC8a	F	GTG TTA CGC GGA ACC AGG CA	124
	R	CCA TCT GTT GGT ACT ACT GA	
AC8b	F	GAT GCG ACT ATC ACT CGC ACG CAG TT	140
	R	GCT GGT TTC CTT CTA GTC CC	
AC9	F	GCC TGG TGC CCG GAC AAT GC	146
	R	ACA TCA CAC TGA TCT CGC CGG CGC T	
AT17	F	GAC ACA CTC GAT CTC AGT AG	180
	R	TTA GCA GGA CGA TTG TAC AG	
GGT5	F	TCA CCA TCG ACG CTC CGG GT	161
	R	TCG GCC TGG TTG TCT GCC TT	
GTA9	F	CGC AGA TGCAACGAT CAC	122
	R	AAT ATG CAT GCC GGT GGT	
6-7	F	CTA CTT GCG CGC CAC CGC CA	191
	R	GCC GTC GCC AGG TTT TGC AG	
12-5	F	CTG GTC CAC TTG CGG TAC GAC	289
	R	GGA GAA GGA GGC CGA ATA CA	
21-3	F	TGT TGA AAT TTG GCG GCC AT	180
	R	TGC AAG GAG TGC TCA GCT AT	
23-3	F	CAG TCG CCC GGA TAC TGT TA	190
	R	TAA ATC CGC TCC CAA ATC TT	
27-5	F	GTG CTG TGC CTG CAG CCG TT	270
	R	TCC CCA AAG CCG CCG AAT CC	

Primers were used to amplify the VNTR loci and the resultant fragments analysed by gel electrophoresis. After confirmation of a single band, fragment analysis was performed in ABI 310 genetic analyzer. Each electrophoretogram was compared with known molecular weight markers to determine number of repeats at every VNTR locus.

Table S4: SNP and InDel marker's restricted to 3I strains (Excel sheet). Details of the 52 SNPs where all four resequenced genomes (NHDP-55,63,98 & I-30) conflicted with the standard strains (TN, Br 4923 and Thai 53), and 21 locations where the 3I strains further differed among themselves.

Table S4: SNP and InDel marker's restricted to 3I strains

		Type	TN genome position	TN, Br4923, Thai53	NHDP-55 (57X)	NHDP-63 (46X)	NHDP-98 (78X)	I-30 (22X)	Strains conflicting from TN genome	
1	52 SNPs WHERE ALL FOUR RESEQUENCED GENOMES (NHDP-55,63,98 & I-30) WERE CONFLICTING FROM COMPLETE SEQUENCES GENOMES (TN, Br 4923 AND Thai 53)	1	SNP	883	G	A	A	A	NHDP-55,63,98 & I-30	
2		2	SNP	1,011	G	A	A	A	NHDP-55,63,98 & I-30	
3		3	SNP	7,614	C	T	T	T	NHDP-55,63,98 & I-30	
4		4	SNP	10,505	C	T	T	T	NHDP-55,63,98 & I-30	
5		5	SNP	15,985	G	A	A	A	NHDP-55,63,98 & I-30	
6		6	InDel	17,915	TTGGTG GTGTA	-	-	-	-	NHDP-55,63,98 & I-30
7		7	SNP	22,488	G	A	A	A	NHDP-55,63,98 & I-30	
8		8	SNP	29,563	C	T	T	T	NHDP-55,63,98 & I-30	
9		9	SNP	38,202	C	T	T	T	NHDP-55,63,98 & I-30	
10		10	SNP	45,791	T	C	C	C	NHDP-55,63,98 & I-30	
11		11	SNP	48,196	G	A	A	A	NHDP-55,63,98 & I-30	
12		12	SNP	116,225	C	T	T	T	NHDP-55,63,98 & I-30	
13		13	SNP	156,948	C	T	T	T	NHDP-55,63,98 & I-30	
14		14	SNP	193,848	C	T	T	T	NHDP-55,63,98 & I-30	
15		15	SNP	305310	G	A	A	A	NHDP-55,63,98 & I-30	
16		16	SNP	365,411	A	G	G	G	NHDP-55,63,98 & I-30	
17		17	InDel	408,352	-	G	G	G	G	NHDP-55,63,98 & I-30
18		18	SNP	574,640	G	A	A	A	A	NHDP-55,63,98 & I-30
19		19	SNP	604,326	T	C	C	C	C	NHDP-55,63,98 & I-30
20		20	SNP	623,366	C	T	T	T	T	NHDP-55,63,98 & I-30
21		21	SNP	774,705	G	A	A	A	A	NHDP-55,63,98 & I-30
22		22	SNP	954,665	T	C	C	C	C	NHDP-55,63,98 & I-30
23		23	SNP	1,017,410	G	A	A	A	A	NHDP-55,63,98 & I-30
24		24	SNP	1,038,672	C	T	T	T	T	NHDP-55,63,98 & I-30
25		25	SNP	1,113,926	A	G	G	G	G	NHDP-55,63,98 & I-30
26		26	SNP	1,160,429	C	G	G	G	G	NHDP-55,63,98 & I-30
27		27	SNP	1,164,294	C	T	T	T	T	NHDP-55,63,98 & I-30
28		28	SNP	1,217,339	G	A	A	A	A	NHDP-55,63,98 & I-30
29		29	SNP	1,291,021	G	A	A	A	A	NHDP-55,63,98 & I-30
30		30	SNP	1,329,462	C	T	T	T	T	NHDP-55,63,98 & I-30
31		31	SNP	1,450,891	G	A	A	A	A	NHDP-55,63,98 & I-30
32		32	SNP	1,466,044	C	A	A	A	A	NHDP-55,63,98 & I-30
33		33	SNP	1,527,056	G	C	C	C	C	NHDP-55,63,98 & I-30
34		34	SNP	1,585,804	A	G	G	G	G	NHDP-55,63,98 & I-30
35		35	SNP	1,588,116	G	C	C	C	C	NHDP-55,63,98 & I-30
36		36	SNP	1,657,377	G	T	T	T	T	NHDP-55,63,98 & I-30
37		37	SNP	1,734,701	G	A	A	A	A	NHDP-55,63,98 & I-30
38		38	SNP	1,759,655	G	A	A	A	A	NHDP-55,63,98 & I-30
39		39	SNP	1,787,403	C	T	T	T	T	NHDP-55,63,98 & I-30
40		40	SNP	1,897,193	C	A	A	A	A	NHDP-55,63,98 & I-30
41		41	SNP	1,900,012	G	A	A	A	A	NHDP-55,63,98 & I-30
42		42	SNP	1,928,726	A	C	C	C	C	NHDP-55,63,98 & I-30
43		43	SNP	1,962,043	C	T	T	T	T	NHDP-55,63,98 & I-30
44		44	SNP	2,200,739	C	T	T	T	T	NHDP-55,63,98 & I-30
45		45	SNP	2,542,320	G	A	A	A	A	NHDP-55,63,98 & I-30
46		46	SNP	2,714,886	T	A	A	A	A	NHDP-55,63,98 & I-30
47		47	SNP	2,784,331	A	G	G	G	G	NHDP-55,63,98 & I-30
48		48	SNP	2,978,382	C	A	A	A	A	NHDP-55,63,98 & I-30
49		49	SNP	2,999,984	C	T	T	T	T	NHDP-55,63,98 & I-30
50		50	SNP	3,110,867	T	C	C	C	C	NHDP-55,63,98 & I-30
51		51	SNP	3,112,678	C	T	T	T	T	NHDP-55,63,98 & I-30
52		52	InDel	3,224,930	-	G	G	G	G	NHDP-55,63,98 & I-30
53	21 SNPs IDENTIFIED WHERE AT LEAST ONE STRAIN DIFFERED FROM THE REST OF THREE	1	SNP	604178	G	A	G	G	NHDP-55	
54		2	SNP	1514420	T	C	T	T	NHDP-55	
55		3	InDel	1978176	C	-	C	C	NHDP-55	
56		4	SNP	2116512	A	G	A	A	NHDP-55	
57		5	InDel	2322401	G	-	G	G	NHDP-55	
58		6	SNP	21139	G	G	T	G	NHDP-63	
59		7	SNP	509579	G	G	A	G	NHDP-63	
60		8	SNP	530686	C	C	T	C	NHDP-63	
61		9	SNP	1070054	C	C	T	C	NHDP-63	
62		10	SNP	1206344	C	C	T	C	NHDP-63	
63		11	SNP	1996640	C	C	T	C	NHDP-63	
64		12	SNP	2119399	A	A	G	A	NHDP-63	
65		13	SNP	2184972	T	C	T	C	NHDP-63	
66		14	InDel	3106517	-	A	-	A	NHDP-63	
67		15	InDel	667972	-	G	G	-	NHDP-98	
68		16	SNP	2116786	C	T	T	T	I-30 (W-09)	
69		17	SNP	509024	T	C	C	T	NHDP-55 & 63, I-30	
70		18	SNP	1223297	C	T	T	C	NHDP-55 & 63, I-30	
71		19	SNP	1600438	T	G	G	T	NHDP-55 & 63, I-30	
72		20	SNP	1797215	C	T	T	C	NHDP-55 & 63, I-30	
73		21	SNP	1983613	C	T	T	C	NHDP-55 & 63, I-30	

52 SNPs WHERE ALL FOUR RE-SEQUENCED GENOMES (NHDP-55,63,98 & I-30) CONFLICTED WITH STANDARD STRAINS (TN, Br 4923 AND Thai 53)

21 SNPs IDENTIFIED WHERE AT LEAST ONE STRAIN DIFFERED FROM THE OTHER THREE

Table S5. Genotypes of *M. leprae* infecting humans and armadillos in the US

Source	Report State	EE	Residence history*	Genotype	n	VNTR profile [#]									
						AC8a	AC8b	AC9	GGT5	GTA9	6-7	12-5	21-3	23-3	27-5
HUMAN PATIENTS REPORTED IN US	LA	Y	NFR	1A-v1	1	11	7	9	5	9	7	5	3	2	5
	CO	N	Philippines	1A-v2	1	8	8	8	5	9	7	5	3	2	5
	FL	N	Paraguay	1D-v1	1	7	7	9	4	13	5	4	2	2	5
	LA	Y	NFR	3I-1-v1	1	9	7	8	4	10	5	5	2	2	4
	LA	Y	Puerto Rico	3I-1-v2	1	9	7	8	4	9	5	4	2	2	5
	NY	N	Dominican Republic	3I-1-v3	1	10	7	8	4	10	6	5	2	2	4
	MA	N	Brazil	3I-1-v4	1	10	7	8	4	9	6	5	2	2	4
	LA, TX, AR, GA	Y	Various	3I-2-v1	25	10	7	8	4	10	7	5	2	2	4
	NM	Y	Mexico	3I-2-v2	1	11	7	8	4	10	6	5	2	2	4
	CA	Y	Mexico	3I-2-v3	1	10	8	8	4	10	7	5	2	2	4
	AR	Y	NFR	3I-2-v4	1	10	7	9	4	10	6	5	2	4	4
	TX	Y	Mexico	3I-2-v5	1	10	7	8	4	10	7	5	3	2	4
	LA	Y	NFR	3I-2-v6	1	10	7	8	4	10	7	4	2	2	4
	LA	Y	NFR	3I-2-v7	1	10	7	7	4	10	7	5	2	2	4
	LA	Y	NFR	3I-2-v8	1	10	7	7	4	10	6	5	2	2	4
	TX	Y	Mexico	3I-2-v9	1	8	7	8	4	9	7	5	2	2	4
	NJ	N	Puerto Rico	3I-2-v2	1	11	7	8	4	10	6	5	2	2	4
	PR	N	Puerto Rico	3I-2-v11	1	10	8	9	5	10	7	5	2	2	4
	CA	N	Samoa	3I-2-v7	1	10	7	7	4	10	7	5	2	2	4
	TX	Y	Guam	3K-v1	1	8	8	8	4	9	7	4	2	2	5
	TX	Y	Philippines	3K-v2	1	8	8	8	4	8	7	4	2	2	5
OR	N	Micronesia	3K-v3	1	9	7	9	4	8	9	4	1	2	5	
SC	N	Philippines	3K-v4	1	8	8	9	5	11	6	4	3	2	4	
LA	Y	NFR	3L-v1	1	10	7	8	4	8	7	5	2	2	5	
OH	N	Tanzania	3L-v2	1	10	7	8	4	13	6	3	2	2	6	
MI	N	India	3L-v3	1	8	7	8	4	8	6	4	2	2	5	
WARM	AL, AR, MS, TX	Y	NFR	3I-2-v1	28	10	7	8	4	10	7	5	2	2	4
	LA, AL	Y	NFR	3I-2-v13	3	10	7	8	4	10	7	6	2	2	4
	LA	Y	NFR	3I-2-v14	2	10	7	8	4	10	7	5	2	1	4

M. leprae genotypes were determined as shown in Figure 2. Complete data of all 114 SNPs and 10 VNTR Loci is available as supplementary Table S1. #, VNTR profile represents the number of repeat copies found at a particular locus. *, Residence history is as reported by patient; NFR: No Foreign Residence; EE: Endemic Exposure, living in areas with potential exposure to *M. leprae* through armadillos; n: number of strains with same genotype; WARM: Wild Armadillo *M. leprae*.

Table S6. Case control studies and contingency analysis:

A. 3I-1 v 3I-2 Strain Distribution in North and South America

Fisher's Exact Test, Two-sided $P < 0.001$

Odds Ratio=27.000, 95% Confidence Interval 8.312 to 87.7

Data analyzed:

	3I-1	3I-2	Total
South	48 (46%)	16 (15%)	64 (62%)
North	4 (4%)	36 (35%)	40 (38%)
Total	52 (50%)	52 (50%)	104 (100%)

B. 3I-2-v1 Association with Armadillo Exposure (EE/NE)

Fisher's Exact Test, Two-sided $P < 0.001$

Odds Ratio=54.517, 95% Confidence Interval 3.031 to 980.67

Data analyzed: (0.5 added to each value for analysis) Includes Reference Strains

	Endemic Exposure (EE)	Non-Endemic Exposure (NE)	Total
3I-2-v1	25 (46%)	0 (0%)	64 (46%)
Other	14 (26%)	15 (28%)	29 (54%)
Total	39 (72%)	15 (28%)	54 (100%)

C. 3I-2-v1 Association with Residence History (NFR/FR)

Fisher's Exact Test, Two-sided P<0.0001

Odds Ratio=16.5, 95% Confidence Interval 4.207 to 64.713

Data analyzed: Includes Reference Strains

	No Foreign Residence (NFR)	Foreign Residence (FR)	Total
3I-2-v1	21 (39%)	4 (7%)	25 (46%)
Other	7 (13%)	22 (41%)	29 (54%)
Total	28 (52%)	26 (48%)	54 (100%)

D. 3I-2-v1 Associations with Patient Reported Armadillo Contact History

Fisher's Exact Test, Two-sided P=0.3147

Odds Ratio=4.0, 95% Confidence Interval 0.4468 to 35.807

Data analyzed:

	3I-2-v1	Other	Total
Yes	6 (40%)	2 (13%)	8 (53%)
No	3 (20%)	4 (27%)	7 (47%)
Total	9 (60%)	6 (40%)	15 (100%)