Lepr Rev (2009) **80,** 302–315

Genetic diversity of *Mycobacterium leprae* isolates from Brazilian leprosy patients

AMANDA NOGUEIRA BRUM FONTES*,
RAMA MURTHY SAKAMURI**,
IDA MARIA FOSCHIANI DIAS BAPTISTA***,
SOMEI URA***, MILTON OZÓRIO MORAES****,
ALEJANDRA NÓBREGA MARTÍNEZ****,
EUZENIR NUNES SARNO****,
PATRICK J. BRENNAN**,
VARALAKSHMI D. VISSA** & PHILIP NOEL SUFFYS*
**Laboratório de Biologia Molecular Aplicada a Micobactérias, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil*
***Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado, USA*
****Instituto Lauro de Souza Lima, Bauru, São Paulo, Brazil*
*****Laboratório de Hanseníase, Instituto Oswaldo Cruz, Brazil*

Accepted for publication 10 September 2009

Summary

Introduction Leprosy is a chronic disease caused by infection with *Mycobacterium leprae*, an obligate intracellular parasite. A problem in studying the transmission of leprosy is the small amount of variation in bacterial genomic DNA. The discovery of variable number of tandem repeats (VNTRs) allowed the detection of strain variation in areas with a high prevalence of leprosy. Four genotypes of *M. leprae* based on three single-nucleotide polymorphism (SNPs) were also discovered to be useful for analysis of the global spread of leprosy.

Methods In this present study, we examined the allelic diversity of *M. leprae* at 16 select VNTR and three SNP loci using 89 clinical isolates obtained from patients mainly from the neighbouring states of São Paulo and Rio de Janeiro Brazil.

Results and conclusion By use of a PCR-RFLP-based procedure that allows the recognition of SNP types 3 and 4 without the need for the more expensive DNA sequencing steps, characterisation of the main *M. leprae* genotypes was easy. When applied on the study population, it was found that the SNP type 3 is most frequent in these two states of Brazil, and that VNTRs provided further discrimination of the isolates. Two Short Tandem Repeats (STRs) were monomorphic, with the remaining 14 STRs represented by two to 18 alleles. Epidemiological associations with,

Correspondence to: Philip Noel Suffys, Laboratório de Biologia Molecular Aplicada a Micobactérias, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil (e-mail: psuffys@gene.dbbm.fiocruz.br)

 0305-7518/09/064053+14 $1.00

township or state were not evident in this random collection and require further investigations. In phylogenetic trees, branches formed by all 16 STRs clearly separated SNP type 3 organisms from the other types while the allelic patterns of two minisatellite loci 27-5 and 12-5 were highly correlated with SNP type 3. This strain typing study provide the basis for comparison of *M. leprae* strain types within Brazil and with those from other countries, and informed selection of genomic markers and methods for future studies.

Introduction

Leprosy is a slow, chronic infection caused by *Mycobacterium leprae* and still a major health problem in developing countries; notably Brazil. Genotyping of causative agents of infectious diseases is a powerful tool for epidemiological analysis of transmission. Successive genetic studies performed since 2000 revealed polymorphic regions in the genome of *M. leprae*[1] which might be useful for genotyping. Variable numbers of tandem repeats (VNTR), such as runs of di- or trinucleotides, have been examined as potential typing markers and found to vary in copy number between strains of *M. leprae*.[2–10] Another elegant study investigated single nucleotide polymorphisms (SNPs) and found that following the reductive evolution and genome decay events undergone by the leprosy bacillus, the genomes of extant isolates are highly conserved and SNPs are rare.[1,11] Three SNPs were found to be informative and have been used to understand the evolution and global spread of *M. leprae*.[11] Here we present findings of a study comparing the performance of three SNPs and multiple-locus VNTR analysis (MLVA). Biopsy samples from Brazilian leprosy patients were analysed in order to begin to understand the ecology of *M. leprae* in this region.

Material and Methods

PATIENTS AND CLINICAL SAMPLES

This study was performed on skin biopsies derived from 89 leprosy patients who had been submitted to standardised diagnostic procedures based on presence of skin lesions, nerve damage and microscopy-based detection of acid-fast bacilli. Among these, 70 cases (79%) had conclusive data on the clinical form of the disease and included 53 cases of lepromatous leprosy (LL), 14 patients with borderline lepromatous leprosy (BL), two individuals with borderline-borderline leprosy (BB) and a single case of the borderline tuberculoid disease form. Most patients were residents of two regions of Brazil, Rio de Janeiro ($n = 41$) and São Paulo ($n = 41$); few cases were from other states such as Amazonas ($n = 2$), Minas Gerais ($n = 2$), Mato Grosso ($n = 1$), Rio Grande do Sul ($n = 1$) and Pará ($n = 1$) (Figure 1). Skin biopsies were collected according to standard diagnostic procedures and only after informed consent.

DNA EXTRACTION

Two protocols were used for DNA extraction: biopsies from 21 patients (all from Rio de Janeiro) were processed by subjecting them to three freeze/thaw cycles (−70 °C/95 °C), incubation with proteinase K at 60 °C for 18 h, phenol-chloroform-isoamyl alcohol (25:24;1)

Figure 1. State map of Brazil showing the residential origin of the patients included in this study. The states represented in this study are shown in grey and the number of patients is indicated.

extraction and precipitation with isopropanol.[5] The other samples were treated with the DNeasy Blood & Tissue Kit, (Qiagen Inc. Valencia, CA), following the manufacturers recommendations.

MLVA ANALYSIS OF VNTRS

The 16 STR loci were separated into four groups for multiplex PCR using fluorescent labeled 5′ forward primers (Kimura *et al.*;[12] Sakamuri *et al.*; in this series[13]). The amplicons were denatured and submitted to capillary electrophoresis (ABI 3730 Genetic Analyzer, Applied Biosystems), and copy numbers were determined by analysis of the fragment lengths using internal sizing standards with the GeneMapper software v 3.7 (Applied Biosystems).

DIRECT SEQUENCING OF THE VNTRS

When MLVA did not yield satisfactory data to determine the VNTR copy number, single PCR reactions were performed using the HotStartTaq Master Mix (Qiagen Inc.), 0·5 μM of

each primer pair and 2 μl of *M. leprae* DNA extract. After denaturing the DNA at 94 °C for 15 min, PCR was carried out in a thermocycler (Veriti 96 Well Thermal Cycler, Applied Biosystems), by submitting samples to 35 cycles consisting of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 30 s, with a final extension at 72 °C for 10 min. A 5 μl sample of each reaction mixture was run on a 3% agarose gel and PCR products submitted to direct sequencing after purification on Qiaquick Spin Columns (Qiagen Inc.), on an ABI Prism 3730 automated DNA sequencer using the ABI PRISM BigDye Terminator v 3.0 sequencing kit (Applied Biosystems). Both DNA strands were analysed using forward and reverse primers and using the *MEGA* software (Version 4.0).[14]

SNP GENOTYPE ANALYSIS

For differentiation of SNP types 1/2 from 3/4 and characterisation of type 3 and 4, we used a PCR-RFLP-based procedure as described by Sakamuri *et al.*[15] Briefly, differentiation between types 1/2 from 3/4 was obtained by restriction analysis of the locus at nucleotide 2,935,685, by incubating 5 μl of PCR product (generated with primers ATC TGG TCC GGG TAG GAA TC and ACC GGT GAG CGC ACT AAG) with 5 units of *Bst* UI (New England Biolabs, Beverly, MA) at 60 °C for 1 hour. Upon analysis on 3% agarose gel, digestion indicates genotype 3 or 4 while lack of digestion means the presence of genotype 1 or 2. For further differentiation of genotypes 3 and 4, restriction analysis of the PCR product including the locus at nucleotide 14 676 (generated with AAT GGA ATG CTG GTG AGA GC and CAA TGC ATG CTA GCC TTA ATG A) was performed incubating with 5 Units of *Sml* I for 1 hour at 50 °C; digestion indicates SNP type 4 while lack of digestion is typical for type 3. In addition, for differentiation of types 1 and 2, sequencing of the SNP at position 1,642,875 was performed as described by Monot *et al.*[11] Evaluation of association of SNP type with frequency in patients from Rio de Janeiro or São Paulo was assessed by the Chi Square independence test using Epi Info software (CDC, Version 3.5.1).

PHYLOGENETIC ANALYSIS

For phylogenic analysis, we selected 72 of the 89 samples, excluding the 17 that lacked allele definition of one or more VNTR locus (Table 1).

The data were analysed with Unweighted Pair Group Method with Arithmatic Mean (UPGMA), maximum parsimony (MP) and Neighbor Joining (NJ) algorithms in PAUP4.0b software (http://paup.csit.fsu.edu/) and for a minimum spanning tree (MST) in Bionumerics (Version 4.1, Applied Maths, Belgium, http://www.applied-maths.com/bionumerics/bionumerics.htm).

Results

FREQUENCY OF THE *M. LEPRAE* GENOTYPES AS DEFINED BY SNP ANALYSIS

We were able to get conclusive results regarding SNP types in 79 (89%) of the samples while samples from eight patients (four from Rio de Janeiro and São Paulo) either could not be characterised, due either to lack of PCR amplification products or of clinical material (Table 1). Another two samples from RJ could be determined only as being either of type 1 or 2 because there was no more residual template. The SNP type 3 was predominant in the

Table 1. VNTR genotypes of *M. leprae* from RJ, SP and other states (AM, MG, MT, RS, PA)

ID	Form	(AC)8b	(GTA)9	(GGT)5	(AT)17	21-3	(AC)9	(AT)15	(AC)8a	27-5	6-7	(TA)18	(GAA)21	(TA)10	23-3	12-5	18-8	SNP
VNTR genotypes of *M. leprae* from Rio de Janeiro																		
21633[1]	LL	7	11	4	14	2	8	16	8	5	6	19	13	6	2	4	8	1
21850[1]	LL	7	11	4	14	2	8	15	8	5	6	18	13	6	2	4	8	1
21920[1]	LL	7	11	4	14	2	8	15	8	5	6	19	13	6	2	4	8	1 or 2 R
19936	BL	7	11	4	12	2	8	13	9	5	6	13	17	4	2	4	8	1
20255	LL	7	10	4	9	2	8	21	9	5	6	10	19	5	2	4	8	1
21392	LL	7	12	4	15	2	8	14	8	5	6	20	13	6	2	4	8	1
22636	LL	7	12	4	14	2	8	16	7	5	6	18	13	6	2	4	8	1
20825	*BL*	*I*	*10*	*4*	*11*	*2*	*8*	*16*	*9*	*5*	*6*	*12*	*19*	*5*	*2*	*3*	*8*	*1 or 2 R*
20822	BL	7	9	4	12	2	8	17	7	5	5	32	10	7	2	4	5	4
20180	LL	7	15	4	14	2	8	14	9	5	6	17	12	10	2	4	6	4
23353	LL	7	11	4	13	2	7	17	8	5	5	12	9	8	2	4	6	4
MLT/RJ	NA	7	9	4	12	2	8	19	9	5	6	17	12	5	2	4	8	4
21706	BL	7	15	4	14	2	7	13	8	4	6	28 or 18	12	7	2	4	8	4
23825	LL	8	15	4	15	2	8	10	8	4	6	21	16	8	2	3	8	3
21548[2]	*LL*	*7*	*10*	*4*	*11*	*2*	*8*	*10*	*8*	*4*	*7*	*NA*	*11*	*8*	*2*	*4*	*8*	*3*
21866[2]	*BL*	*7*	*10*	*4*	*11*	*2*	*8*	*10*	*8*	*4*	*7*	*NA*	*11*	*12*	*2*	*4*	*8*	*3*
21263	LL	7	10	4	11	2	8	10	8	4	7	18	11	9	2	4	9	3
22368	LL	7	9	4	14	2	8	10	8	4	7	14	15	11	2	4	8	3
21898	LL	7	11	4	11	2	9	18	9	4	6	15	11	6	2	5	8	3
21844	BL	7	12	4	15	2	8	14	9	4	6	14	10	9	2	5	8	3
21372[3]	BL	7	12	4	15	2	8	19	9	4	5	14	10	9	2	5	8	3
21230[3]	BL	7	12	4	15	2	8	20	9	4	5	14	10	9	2	5	8	3
19856	BL	7	8	4	16	2	8	19	9	4	5	16	12	7	2	5	8	3
19914	LL	7	10	4	14	2	8	23	10	4	6	18	13	6	2	5	8	3
20008	LL	7	9	4	12	2	8	20	9	4	6	20	12	8	2	5	6	3
20241	LL	7	10	3	10	2	8	18	10	4	8	14	11	8	2	5	9	3
20368	LL	7	10	4	13	2	8	19	9	4	6	15	10	7	2	5	8	3
21190	LL	7	10	4	11	2	8	25	9	4	6	20	12	10	2	5	8	3
21644	LL	7	10	4	11	2	9	15	9	4	5	15	13	9	2	5	6	3
21726	LL	7	11	4	14	2	8	28	9	4	6	23	10	7	2	5	8	3
22494	LL	7	8	4	13	2	8	20	9	4	7	14	11	6	2	5	8	3
22670	LL	7	11	3	11	2	8	17	10	4	7	14	14	8	2	5	9	3
22830	LL	7	9	4	13	2	8	24	9	4	6	18	11	7	2	5	8	3

Table 1. *continued*

ID	Form	(AC)8b	(GTA)9	(GGT)5	(AT)17	21-3	(AC)9	(AT)15	(AC)8a	27-5	6-7	(TA)18	(GAA)21	(TA)10	23-3	12-5	18-8	SNP
23186	LL	7	9	4	12	2	8	14	9	4	7	15	11	7	2	5	8	3
23321	LL	7	9	4	12	2	8	18	10	4	7	23	10	9	2	5	8	3
23806	LL	7	9	4	11	2	8	15	10	4	6	17	11	6	2	5	9	3
19930[4]	BL	7	9	4	10	2	8	16	10	4	6	21	11	7	2	5	9	3
Genildo (001)	*NA*	*I*	*I*	*I*	*I*	*2*	*7*	*14*	*8*	*4*	*6*	*NA*	*10*	*8*	*2*	*5*	*6*	*I*
Inaldi (002)	*NA*	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*I*	*10*	*2*	*5*	*8*	*I*
20476	LL	7	9	4	9	2	8	15	11	4	6	21	11	9	2	5	8	I
21847	LL	7	11	4	11	2	9	17	9	4	6	15	11	6	2	5	8	NS
VNTR genotypes of *M. leprae* from Sao Paulo																		
AH/SP[4]	LL	7	9	4	10	2	8	14	9	4	6	21	11	7	2	5	9	3
557/SP	LL	7	9	4	10	2	8	18	9	4	6	15	11	10	2	5	9	3
LS/SP[5]	*BT*	*7*	*10*	*4*	*10*	*2*	*8*	*15*	*9*	*4*	*6*	*NA*	*11*	*6*	*2*	*5*	*8*	*3*
ZR/SP[5]	BB	7	10	4	10	2	8	15	8	4	6	15 or 23	11	6	2	5	8	3
501/SP	LL	7	10	4	10	2	8	15	8	4	6	22 or 23	11	9	2	5	6*	3
540/SP	LL	7	7	4	10	2	8	13	8	4	6	16	13	9	2	5	7*	3
548/SP	LL	7	12	4	12	2	8	11	8	4	6	14	15	10	2	5	6*	3
RM/SP	BB	7	12	4	12	2	8	17	9	4	6	15	11	7	2	5	8	3
379/SP	LL	7	10	4	14	2	8	16	9	4	6	18	10	9	2	5	8	3
476/SP	*LL*	*7*	*10*	*4*	*14*	*2*	*8*	*24*	*9*	*4*	*6*	*NA*	*10*	*7*	*2*	*5*	*8*	*3*
478/SP	LL	7	10	4	12	2	8	19	9	4	6	16	9	8	2	5	8	3
484/SP	LL	7	10	4	14	2	8	19	9	4	6	20	12	12	2	5	8	3
510/SP	LL	7	9	4	15	2	8	21	9	4	6	14	10	8	2	5	8	3
512/SP[6]	LL	7	9	4	14	2	8	26	9	4	6	14	10	11	2	5	8	3
567/SP[6]	NA	7	9	4	14	2	8	15	9	4	6	14	10	11	2	5	8	3
518/SP	NA	7	12	4	11	2	9	22	9	3	7	15	14	11	2	5	6*	3
535/SP	LL	7	10	4	14	2	8	14	9	4	6	20	12	10	2	5	3*	3
537/SP	LL	7	13	4	14	2	8	16	9	4	6	14	10	10	2	5	8	3
539/SP	LL	7	9	4	12	2	8	12	9	4	7	16	11	10	2	5	8	3
551/SP	LL	7	10	4	11	2	9	21	9	4	6	17	10	7	2	5	7	3
555/SP	LL	7	9	4	10	2	8	21	9	4	6	21	14	7	2	5	9	3
556/SP	LL	7	9	4	13	2	8	15	9	4	7	15	13	10	2	5	6*	3
558/SP	LL	7	8	4	13	2	8	13	9	4	7	14	11	6	2	5	8	3
564/SP	NA	7	10	4	14	2	8	21	9	4	6	18	10	7	2	5	8	3
569/SP	LL	7	10	4	9	2	8	19	9	4	6	14	12	12	2	5	9	3
570/SP	BL	7	9	4	13	2	8	19	9	4	6	13	10	11	2	5	8	3
571/SP	LL	7	10	4	11	2	8	25	9	4	6	15	10	10	2	5	7	3

Table 1. *continued*

ID	Form	(AC)8b	(GTA)9	(GGT)5	(AT)17	21-3	(AC)9	(AT)15	(AC)8a	27-5	6-7	(TA)18	(GAA)21	(TA)10	23-3	12-5	18-8	SNP
585/SP	*NA*	*7*	*12*	*4*	*13*	*2*	*8*	*17*	*9*	*4*	*6*	*NA*	*11*	*12*	*2*	*5*	***NS***	*3*
588/SP	LL	7	9	4	10	2	7	13	9	4	6	15	10	10	2	5	9	3
PFM/SP	BL	7	10	4	12	2	8	15	9	4	6	16	10	7	2	5	8	3
545/SP	*LL*	*7*	*9*	*4*	*13*	*2*	*7*	*15*	*10*	*4*	*6*	*NA*	*11*	*8*	*2*	*5*	*8*	*3*
479/SP	LL	7	10	4	11	2	8	28	11	4	6	14	10	10	2	5	8	3
578/SP	*NA*	***I***	*10*	*3*	*11*	*2*	*8*	*21*	*11*	*4*	*7*	*15*	*12*	*11*	*2*	*5*	*8**	*3*
582/SP	*NA*	*7*	*10*	*4*	*13*	*2*	*8*	*13*	*9*	*5*	*6*	*NA*	*10*	*8*	*2*	*4*	***I***	*3*
580/SP	*LL*	***NS***	*11*	*4*	*14*	*2*	*9*	*17*	*8*	***NS***	*6*	*14*	*15*	*12*	*2*	***NS***	*2**	*3*
JBF/SP[7]	LL	7	9	4	13	2	8	15	9	4	7	15	13	7	2	5	8	3
581/SP	NA	8	10	4	20	2	6	15	8	5	6	20	14	8	2	4	**3***	4
586/SP	*BL*	***NS***	*11*	*4*	*12*	*2*	*8*	*13 or 17*	*10*	*4*	*5*	*NA*	*10*	*8*	*2*	*5*	*3**	***NS***
JPS/SP	*BL*	*7*	*9*	*4*	*13*	*2*	*8*	*21*	*10*	*5*	*6*	*NA*	*12*	*8*	*2*	*4*	*3**	***I***
AAB/SP	*NA*	***I***	*10*	***I***	*11*	***I***	*8*	*17*	*8*	*4*	*6*	*NA*	*10*	*10*	*2*	*5 or 6*	***I***	***I***
565/SP	*NA*	*7?*	*10*	*4*	*10*	***I***	*8*	*12*	*9*	*4*	*6*	***I***	*12*	*11*	*2*	*5*	*8*	***I***
VNTR genotypes of *M. leprae* from other states (AM, MG, MT, RS, PA)																		
RAS/AM	NA	7	9	4	13	2	8	17	10	4	8	14	13	9	1	5	7	1
577/AM	NA	7	10	4	13	2	8	17	9	5	6	14	13	10	2	4	8	1
JRB/MG[7]	NA	7	9	4	13	2	8	15	9	4	7	15	13	7	2	5	8	3
544/MG	NA	7	11	3	10	2	8	16	10	4	6	14	10	11	2	5	7	3
594/MT	*NA*	*7*	*11*	*4*	*13*	*2*	*8*	*14*	*8*	*4*	*6*	*NA*	*11*	*11*	*2*	*5*	*9*	*3*
566/RS	NA	7	11	4	12	2	8	12	9	4	6	17	11	10	2	5	8	3
568/PA	NA	7	10	4	18	2	7	26	8	5	6	25	13	8	2	4	6*	4

NA: Not available; NS: Unable to perform PCR due to shortage of the DNA template; I: NO PCR product; *Alleles determined by agarose gel; ? Not sure, weak product; Data in italics: not included for phylogenetic analysis; grey shaded fields: relate alleles of loci 27-5 and 12-5 with SNP genotype.

[1,2,3,5,6] *M. leprae* isolates from same state with highly similar VNTR profiles which grouped together in the NJ tree (Figure 2). [4,7] Isolates with similar VNTR profile seen in a different state.

Table 2. Allele types, frequency and allelic diversity of *M. leprae* VNTR loci in Brazilian leprosy patients

Locus name	(AC)8b		(GTA)9		(GGT)5		(AT)17		21-3		(AC)9		(AT)15		(AC)8a		27-5		6-7		(TA)18		(GAA)21		(TA)10		23-3		12-5		18-8	
Population	Brazil- Rio de Janeiro																															
Allelic diversity*	0·053		0·812		0·1		0·85		0		0·28		0·927		0·62		0·43		0·561		0·915		0·833		0·853		0		0·544		0·446	
Allele type and frequency	7	37	8	2	3	2	9	2	2	40	7	3	10	1	7	2	4	28	5	6	10	1	9	1	4	1	2	41	3	2	5	1
	8	1	9	10	4	37	10	2	NA	1	8	34	13	3	8	8	5	12	6	25	12	2	10	8	5	3			4	16	6	5
	NA	3	10	10	NA	2	11	10			9	3	14	5	9	23	NA	1	7	8	13	1	11	12	6	10			5	23	8	30
			11	9			12	6			NA	1	15	5	10	6			8	1	14	7	12	6	7	8					9	5
			12	5			13	4					16	4	11	1			NA	1	15	5	13	7	8	7						
			15	3			14	9					17	5	NA	1					16	1	14	1	9	7						
			NA	2			15	5					18	4							17	3	15	1	10	3						
							16	1					19	4							18	5	16	1	11	1						
							NA	2					20	4							19	2	17	1	12	1						
													21	1							20	3	19	2								
													23	1							21	2	NA	1								
													24	1							23	2										
													25	1							32	1										
													28	1							NA	6										
													NA	1																		
Population	Brazil- São Paulo																															
Allelic diversity*	0·055		0·688		0·05		0·842		0		0·27		0·954		0·475		0·188		0·298		0·853		0·772		0·853		0		0·146		0·614	
Allele type and frequency	7	36	7	1	3	1	9	1	2	39	6	1	11	1	8	7	3	1	5	1	13	1	9	1	6	3	2	41	4	3	3	2
	8	1	8	1	4	39	10	9	NA	2	7	2	12	2	9	29	4	36	6	34	14	9	10	16	7	8			5	36	6	4
	NA	4	9	13	NA	1	11	6			8	35	13	4	10	3	5	3	7	6	15	8	11	10	8	7			NA	2	7	3
			10	19			12	6			9	3	14	2	11	2	NA	1			16	4	12	6	9	3					8	21
			11	2			13	8					15	9							17	1	13	3	10	10					9	5
			12	4			14	9					16	2							18	2	14	3	11	6					NA	6
			13	1			15	1					17	4							20	3	15	2	12	4						
							20	1					18	1							21	2										
													19	4							22	1										
													21	6							NA	10										
													22	1																		
													24	1																		
													25	1																		
													26	1																		
													28	1																		

Table 2. *continued*

Locus name	(AC)8b		(GTA)9		(GGT)5		(AT)17		21-3		(AC)9		(AT)15		(AC)8a		27-5		6-7		(TA)18		(GAA)21		(TA)10		23-3		12-5		18-8	
													NA	1																		
Population	Brazil (Other than Rio de Janeiro and Sao Paolo)																															
Allelic diversity*	0		0·762		0·286		0·715		0		0·286		0·952		0·762		0·477		0·524		0·7		0·667		0·905		0·286		0·477		0·81	
Allele type and frequency	7	7	9	2	3	1	10	1	2	7	7	1	12	1	8	2	4	5	6	5	14	3	10	1	7	1	1	1	4	2	6	1
	NA		10	2	4	6	12	1			8	6	14	1	9	3	5	2	7	1	15	1	11	2	8	1	2	6	5	5	7	2
			11	3			13	4					15	1	10	2			8	1	17	1	13	4	9	1					8	3
							18	1					16	1							25	1			10	2					9	1
													17	2							NA	1			11	2						
													26	1																		
Population	Brazil- General																															
Allelic diversity*	0·05		0·76		0·09		0·86		0		0·27		0·92		0·57		0·43		0·44		0·873		0·81		0·86		0·00		0·41		0·056	

* Simpson's diversity index.[16]

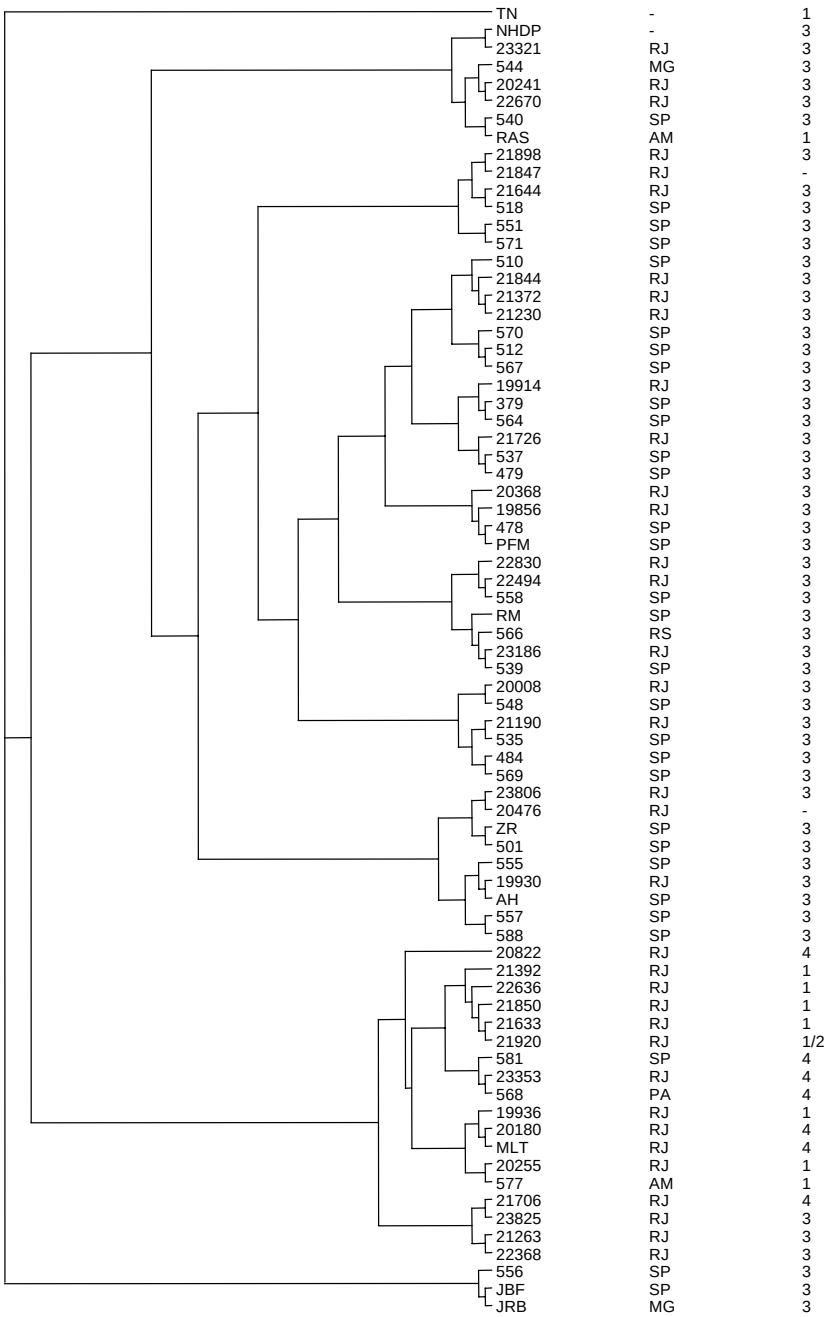

Figure 2. Distance tree generated using the Neighbor joining (NJ) method from VNTR data based on 72 Brazilian *M. leprae* isolates for which allele definition was available for all 16 loci (Table 1). The columns to the right of the tree (left to right) indicate the sample ID, state of origin of patient and *M. leprae* SNP type. TN and NHDP63 are reference strains. Alleles for TN were obtained from the sequenced strain.1 NHDP63 is an armadillo derived clinical isolate used for VNTR strain typing .5,12.

study population ($n = 64$, 79%) and observed in 24 cases (65%) from Rio de Janeiro, 36 (97%) from São Paulo, two from Minas Gerais and one case from Mato Grosso and from Rio Grande do Sul. In addition, the SNP type 4 was detected in five cases (14%) from Rio de Janeiro and one patient from São Paulo and from Pará. Both patients from Amazonas carried bacilli of SNP type 1, observed also among six cases (16%) from Rio de Janeiro. Upon evaluation of eventual association of SNP type with frequency in patients from Rio de Janeiro or São Paulo, SNP type 3 was significantly more frequent in São Paulo than Rio de Janeiro ($X^2 = 10{\cdot}66$, $P = 0{\cdot}001$) while SNP type 1 was more frequently observed in Rio de Janeiro ($X^2 = 4{\cdot}5$, $P = 0{\cdot}03$).

GENOTYPING AS DETERMINED BY ANALYSIS OF 16 VNTRS

In our study population, 15 out of 16 the VNTR loci were observed to be polymorphic while the STR locus 21-3 was invariable, presenting two repeat copies in the amplified DNA from all samples (Table 1). When observing the discriminatory power of the different STRs, highest allelic diversity was presented at locus (AT)15, presenting 18 different alleles that ranged in copy number between 10 and 28 copies, 15 copies being the most frequent type. Allele numbers for the loci (TA)18 and (AT)17 were 14 and 10, respectively, both showing a copy number of 14 as being most common. Less variable were (AC)8b, (GGT)5 and 23-3, presenting two alleles each. The discriminatory power as defined by the Simpson's discriminatory index ranged from 0·92 [(AT)15] to 0 [21-3] and we did not observe any marker that presented region associated variation (Table 2).[16]

When observing both STR and SNP based genotypes, a correlation was observed between the alleles of loci 27-5 and 12-5 and SNP: all except one isolate that presented the 4:5 allele combination was SNP type 3 (Table 1).

PHYLOGENY

Grouping of genotypes was essentially the same when observing the NJ (Figure 2) or MST (data not shown). In general, the samples that could not be analysed for SNP type were also those that had more than one STR locus left undetermined (Table 1) and the fact that 72 samples had an almost complete set of STR typing data and SNP scores available allowed us to determine the relation between these markers for phylogeny studies. When observing clustering as represented in the NJ tree, a close relation was observed between alleles defined by SNP types and by STRs as the type 3 genotypes were separately grouped from the rest of the SNP genotypes (type 1 and 4) (Figure 2). Exceptions were four samples, including one sample of type 1 (RAS/AM) and three of type 3 (23825/RJ, 21263/RJ and 22368 /RJ).

When looking at cluster size and composition, only two samples were found to be identical and were from different States (JBF/SP and JRB/MG). Upon decreasing stringency of cluster definition and observation of similar genotypes (seven clusters with highest similarity that included six with two samples and one with three; Table 1), we could observe a tendency for samples to be clustered according to origin (five of these groups were composed of samples from patients from the same state). Nonetheless, no particular grouping of samples from Rio de Janeiro or São Paulo was observed in the overall NJ tree (Figure 2) or the UPGMA, MP and MST trees (data not shown).

Discussion

Recent advances in the molecular characterisation of isolates of *M. leprae* can lead to alternative approaches for their detection and differentiation, and for the understanding of several other aspects of leprosy.[17] Genome variability was identified on the basis of two types of genetic analysis: SNP identification, for recognition of major genotypes of *M. leprae*, and STR mapping by MLVA, in an attempt to further describe the strain types, their distribution, and the relationship, if any, between the genotypes and disease phenotypes.[11,4,7] As far as we know, data on the genetic variability of isolates from Brazilian leprosy patients are not available. We, therefore, characterised a set of SNPs and VNTRs of *M. leprae* present in skin biopsy samples from patients, residents mainly of Rio de Janeiro and São Paulo. Our analysis demonstrated that this collection of 89 samples, obtained from different patients who had been selected randomly, presented 80 different genotypes as determined by combination of all STRs and related with the three observed SNP defined genotypes, consisting mainly of type 3. We also found a small proportion of SNP type 4, a type postulated to have originated in Western Africa and transmitted to the new world during slave trade.[11] Since both São Paulo and Rio de Janeiro are located in Southeast Brazil, the ethnically most diverse part of the country and populated by Caucasians (59%), Africans and mixed-race (40%), the largest percentage of Asian Brazilians (0·8%) and Amerindians (0·2%), the detection of SNP type 3 and 4 *M. leprae* is plausible.[11] SNP type 1 was noted in six cases from Rio de Janeiro, two from Amazonas and one from São Paulo. Although SNP type 3 is overwhelmingly present in our study population, we still observed a significantly higher frequency of SNP type 3 in São Paulo and of SNP type 1 in Rio de Janeiro; this calls for a well designed study. We imagine, however, that sampling bias could be involved, as most samples from São Paulo are from patients that are residents from the region of Bauru, while the samples from Rio de Janeiro have a more widespread origin (data not shown). We also observed that the expensive [$7 or more per PCR amplicon in a typical sequencing facility] and time-consuming sequencing procedure for SNP type determination can be replaced by PCR-restriction enzymes analysis [less than $4 per sample for reagent (enzyme and agarose) costs for all three SNPs] and was efficiently introduced in our study. Our experimental approach allowed us to determine the alleles for all or most of the 16 STRs in a large fraction of our sampling. The overall Hunter Gaston Discrimination Index (HGDI) considering all 16 STRs was 0·997 and similar to the value of 0·995 observed for the 12 MIRU-VNTR typing system established for *M. tuberculosis*.[18] There is a need for additional sampling in Brazil to evaluate the use of VNTR based typing for different situations including epidemiology and phylogenetics. As observed in earlier studies, the HGDI of the individual STRs is highly variable and ranged from 0·00 [21-3] to 0·92 [(AT)15], yet no significant differences were observed when comparing allelic properties from isolates from Rio and Sao Paulo. We evidenced some links between the genotypes composed either by SNP or VNTRs. The allele combination of the STRs 27-5 and 12-5, that, in a 4-5 copy number combination, were almost exclusive for SNP genotype 3, while that for SNP type 4 was 5-4. This pattern was also observed in samples from Colombia (in this series).[19] The VNTR and SNP association was also clear from the organisation of the phylogenetic trees, being either NJ (Figure 2) or MST (data not shown), where in the SNP type 3 strains were separated from the rest when the grouping was constructed on all (or almost all) VNTRs. This strong association between genotype organisation as defined by different kind of genetic markers, such as DR-spacers, MIRUs or SNPs, has also been observed in *M. tuberculosis*.[20]

A limited number of genotyping assays for *M. leprae* is available and it might be too early to jump to conclusions regarding similarity or difference between data from our study and from others. Nonetheless, we have the impression that regional differences in the allelic range (number of alleles) and frequencies of each of the STR loci might exist. In our study, only two copies of STR 21-3 were observed while Zhang *et al.*[7] observed three different alleles of this locus; interestingly, this difference was not observed when comparing copy number of STR 23-3, exclusively presenting two copies in both studies. Predominance of particular alleles could also be different between countries, as suggested observing the major copy number of markers (GAA)21 and (TA)18 as reported by us and by Weng *et al.*[21] Although region associated genotypic particularities are known to occur for other mycobacterial species, further studies are needed to make clear to what extent this occurs in *M. leprae*.[22–26]

Acknowledgements

We are greatly indebted to the Genomic Platform for DNA Sequencing (PDTIS-Fiocruz). These studies were funded by NIH, NIAID grants AI-047197, AI-063457, and Contract NO1-AI-25469, by the Heiser Program for Research on Leprosy and Tuberculosis of the New York Community Trust through IDEAL – Initiative for Diagnostic and Epidemiological Assays for Leprosy, by the CNPq (401283/2005-4) and the Damien Foundation (ILEP).

References

[1] Cole ST, Eiglmeier KJ, Parkhill KD *et al.* Massive gene decay in the leprosy bacillus. *Nature*, 2001; **409**: 1007–1011.
[2] Matsuoka M, Maeda S, Kai M *et al. Mycobacterium leprae* typing by genomic diversity and global distribution of genotypes. *Int J Leprosy*, 2000; **68**: 121–128.
[3] Shin YC, Lee H, Lee H *et al.* Variable numbers of TTC repeats in *Mycobacterium leprae* DNA from leprosy patients and use in strain differentiation. *J Clin Microbiol*, 2000; **38**: 4535–4538.
[4] Groathouse NA, Rivoire B, Kim H *et al.* Multiple polymorphic loci for molecular typing of strains of *Mycobacterium leprae*. *J Clin Microbiol*, 2004; **42**: 1666–1672.
[5] Truman R, Fontes AB, De Miranda AB *et al.* Genotypic variation and stability of four variable-number tandem repeats and their suitability for discriminating strains of *Mycobacterium leprae*. *J Clin Microbiol*, 2004; **42**: 2558–2565.
[6] Young SK, Taylor GM, Jain S *et al.* Microsatellite mapping of *Mycobacterium leprae* populations in infected humans. *J Clin Microbiol*, 2004; **42**: 4931–4936.
[7] Zhang L, Budiawan T, Matsuoka M. Diversity of potential short tandem repeats in *Mycobacterium leprae* and application for molecular typing. *J Clin Microbiol*, 2005; **43**: 5221–5229.
[8] Matsuoka M, Zhang L, Morris MF *et al.* Polymorphism in the rpoT gene in *Mycobacterium leprae* isolates obtained from Latin American countries and its possible correlation with the spread of leprosy. *FEMS Microbiol Lett*, 2005; **243**: 311–315.
[9] Monot M, Honoré N, Balière N *et al.* Are variable-number tandem repeats appropriate for genotyping *Mycobacterium leprae*? *J Clin Microbiol*, 2008; **46**: 2291–2297.
[10] Young SK, Ponnighaus JM, Jain S *et al.* Use of Short Tandem Repeat Sequences to Study *Mycobacterium leprae* in Leprosy Patients in Malawi and India. *PLoS Negl Trop Dis*, 2008; **2**: e214.
[11] Monot M, Honoré N, Garnier T *et al.* On the origin of leprosy. *Science*, 2005; **308**: 1040–1042.
[12] Kimura M, Sakamuri RM, Groathouse NA *et al.* Rapid Variable-Number Tandem Repeat (VNTR) genotyping for *Mycobacterium leprae* clinical specimens. *J Clin Microbiol*, 2009; **47**: 1757–1766.
[13] Sakamuri RM, Harrison J, Gelber R, *et al.* A continuation: study and characterisation of *M. leprae* short tandem repeat genotypes and transmission of leprosy in Cebu, Philippines. (Submitted in this Series) *Lepr Rev*, 2009; **80**.
[14] Tamura K, Dudley J, Nei M *et al. MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*, 2007; **24**: 1596–1599.

[15] Sakamuri RM, Kimura M, Li W *et al.* Population based molecular epidemiology of leprosy in Cebu, Philippines. *J Clin Microbiol*, 2009; **47**: 2844–2854.
[16] Hunter PR, Gaston MA. Numerical Index of the Discriminatory Ability of Typing Systems: an Application of Simpson's Index pf Diversity. *J Clin Microbiol*, 1988; **26**: 2465–2466.
[17] Matsuoka M. Recent advances in the molecular epidemiology of leprosy. *Jpn J Leprosy*, 2009; **78**: 67–73.
[18] Kremer K, Arnold C, Cataldi A *et al.* Discriminatory power and reproducibility of novel DNA typing methods for Mycobacterium tuberculosis complex strains. *J Clin Microbiol*, 2005; **44**: 4498–4510.
[19] Cardona-Castro N, Beltran-Alzate JC, Romero-Montoya IM *et al.* Identification and comparison of *Mycobacterium leprae* genotypes in two geographical regions of Colombia. (Submitted in this series) *Lepr Rev*, 2009; **80**: 316–321.
[20] Filliol I, Motiwala AS, Cavatore M *et al.* Global Phylogeny of *Mycobacterium tuberculosis* Based on Single Nucleotide Polymorphism (SNP) Analysis: Insights into Tuberculosis Evolution, Phylogenetic Accuracy of Other DNA Fingerprinting Systems, and Recommendations for a Minimal Standard SNP Set. *J Bacteriol*, 2006; **188**: 759–772.
[21] Weng X, Wang Z, Liu J *et al.* Identification and distribution of *Mycobacterium leprae* genotypes in a region of high leprosy prevalence in China: a 3-year molecular epidemiological study. *J Clin Microbiol*, 2007; **45**: 1728–1734.
[22] Penna ML, Wand-Del-Rey-de-Oliveira ML, Penna G. Spatial distribution of leprosy in the Amazon region of Brazil. *Emerg Infect Dis*, 2009; **15**: 650–652.
[23] Penna G, Pinto LF, Soranz D *et al.* High incidence of diseases endemic to the Amazon region of Brazil, 2001–2006. *Emerg Infect Dis*, 2009; **15**: 626–632.
[24] Rodrigues-Júnior AL, do O VT, Motti VG. Spatial and temporal study of leprosy in the state of São Paulo (Southeastern Brazil), 2004–2006. *Rev Saude Publica*, 2008; **42**: 1012–1020 (English, Portuguese).
[25] Amaral EP, Lana FC. Spacial analysis of Leprosy in the microregion of Almenara, MG, Brazil. *Rev Bras Enferm*, 2008; **61**: 701–707 (Portuguese).
[26] Gauy JS, Hino P, dos Santos CB. Spatial distribution of leprosy cases in Ribeirão Preto, Brazil, 2004. *Rev Lat Am Enfermagem*, 2007; **15**: 460–465.