

# A major susceptibility locus for leprosy in India maps to chromosome 10p13

M. Ruby Siddiqui<sup>1</sup>, Sarah Meisner<sup>1</sup>, Kerrie Tosh<sup>1</sup>, Karupiah Balakrishnan<sup>2</sup>, Satish Ghei<sup>3</sup>, Simon E. Fisher<sup>1</sup>, Marina Golding<sup>1</sup>, Nallakandy P. Shanker Narayan<sup>4</sup>, Thiagarajan Sitaraman<sup>5</sup>, Utpal Sengupta<sup>3</sup>, Ramasamy Pitchappan<sup>2</sup> & Adrian V.S. Hill<sup>1</sup>

Leprosy, a chronic infectious disease caused by *Mycobacterium leprae*, is prevalent in India, where about half of the world's estimated 800,000 cases occur<sup>1</sup>. A role for the genetics of the host in variable susceptibility to leprosy has been indicated by familial clustering, twin studies, complex segregation analyses and human leukocyte antigen (HLA) association studies. We report here a genetic linkage scan of the genomes of 224 families from South India, containing 245 independent affected sibpairs with leprosy, mainly of the paucibacillary type. In a two-stage genome screen using 396 microsatellite markers, we found significant linkage (maximum lod score (MLS)=4.09,  $P < 2 \times 10^{-5}$ ) on chromosome 10p13 for a series of neighboring microsatellite markers, providing evidence for a major locus for this prevalent infectious disease. Thus, despite the polygenic nature of infectious disease susceptibility, some major, non-HLA-linked loci exist that may be mapped through obtainable numbers of affected sibling pairs.

*M. leprae* was the first pathogenic bacteria of humans to be identified and remains one of the few that cannot be cultured *in vitro*. It causes leprosy, a disease long recognized to occur in families<sup>2,3</sup>. The more common form, paucibacillary leprosy, is characterized by localized anesthetic skin lesions and low bacillary load in contrast to the disseminated progressive disease and high bacillary load observed in multibacillary leprosy. Twin studies in India have demonstrated higher concordance rates for leprosy in monozygotic (60–85%) than dizygotic (5–20%) twins, indicating a strong genetic component<sup>4,5</sup>. Complex segregation analyses, carried out on several populations, have reported models consistent with one major codominant or recessive gene and perhaps several modifying genes controlling susceptibility to leprosy<sup>6,7</sup>, but the usefulness of this approach has not been validated. Finally, numerous HLA studies have reported associations, but only HLA-DR2 series (for example, *HLA-DRB1\*1501*) alleles have been consistently associated with susceptibility to leprosy<sup>8</sup>. But the reported associations with

this and other candidate genes seemed insufficient to account for the overall genetic component suggested by twin studies<sup>9</sup>.

Semi-automated, fluorescence-based technology and linkage analysis have been used to screen the human genome for genes conferring a major effect on susceptibility to several autoimmune multifactorial diseases<sup>10</sup>. It has been unclear, however, whether a similar approach could identify regions of linkage for common infectious diseases with more modest sibling recurrence risks ( $\lambda_s$ ), an essential environmental infectious agent and a likely highly polygenic basis.

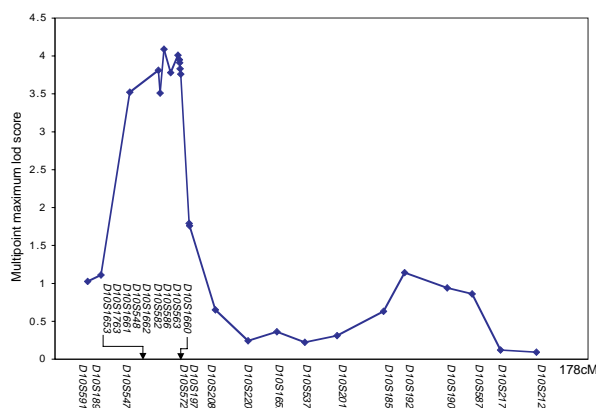
In the first stage of this leprosy genome screen, we genotyped 103 independent, affected sibpairs with mainly paucibacillary leprosy and their parents (Table 1) for 388 highly polymorphic microsatellite markers covering all 22 autosomes and the X chromosome. We calculated probability of linkage for each marker. Regions of the genomes were considered as showing weak suggestive evidence of linkage in this first set of families if an individual marker attained statistical significance at the level of  $P < 0.10$  or a  $MLS \geq 1.0$ . Based on these criteria, we observed weak suggestive evidence of linkage with leprosy for 28 regions on all chromosomes except chromosomes 19 and 21 (Web Table A).

We then assessed 37 markers from these regions in an additional 142 independent affected sibpairs (Table 1). In the overall analysis, only one region on chromosome 10p (*D10S548*) showed a multipoint  $MLS > 3.0$  (Table 2). A further eight markers in this region were typed on all families to increase the marker density. A multipoint lod map was produced for chro-

**Table 1 • Number of families and sibpairs in the two-stage leprosy genome screen**

Number of families	First stage	Second stage
2 affected siblings	84	120
3 affected siblings	8	11
4 affected siblings	1	0
Total families	93	131
Total number of independent sibpairs	103	142

The number of fully independent sibpairs in each family is equal to the number of affected siblings minus one.



**Fig. 1** Maximum likelihood multipoint map for 25 microsatellite markers on chromosome 10, including 8 markers flanking *D10S548*. The maximum multipoint  $MLS$  for the region is 4.09 ( $P$  value=0.000007) and corresponded to *D10S166*.

<sup>1</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>2</sup>Madurai Kamaraj University, Madurai, India. <sup>3</sup>JALMA Institute, Agra, India. <sup>4</sup>Leprosy Project, Sakthi Nagar, India. <sup>5</sup>Hindu Mission Hospital, Kumbakonam, India. Correspondence should be addressed to A.V.S.H. (e-mail: adrian.hill@well.ox.ac.uk).



Table 2 • Chromosome 10p linkage analysis

Microsatellite	Position	MAPMAKER/SIBS		SIBPAIR ANALYZE
		Multipoint MLS	P value	MLS
D10S1653	38.8	3.81	0.000013	2.54
D10S1763	39.5	3.51	0.000014	1.81
D10S1661	40.9	4.09	0.000007	2.02
D10S548	43.4	3.78	0.000015	2.02
D10S1662	46.2	4.01	0.000008	2.64
D10S582	46.7	3.95	0.000010	3.52
D10S586	46.8	3.91	0.000011	0.2
D10S563	47.0	3.83	0.000013	1.64
D10S1660	47.2	3.76	0.000015	2.73

Combined linkage analysis for nine microsatellite markers on chromosome 10p13, with the distance of each marker from the telomere of 10p shown in centimorgans (based on the Genethon map, <ftp://ftp.genethon.fr/pub/Gmap>). The single point MLS values (calculated using SIBPAIR ANALYZE) are shown, as well as the multipoint MLS values and the associated significance level for the latter.

mosome 10 using MAPMAKER/SIBS (Fig. 1 and Table 2). The maximum multipoint lod score for the region corresponding to the marker *D10S1661* was 4.09 ( $P$  value < 0.00002), and thus exceeded the strict level proposed for genome-wide statistical significance<sup>11,12</sup>.

Numerous association studies of candidate genes in leprosy have identified the HLA class II region<sup>8</sup>. We did not observe significant linkage for microsatellite markers within the MHC or flanking the *NRAMP1* locus<sup>13</sup>, indicating that neither of these regions encodes a major susceptibility locus in this population.

So far genetic studies of infectious diseases have largely focused on candidate genes and this approach has identified various genes associated with disease susceptibility. But almost invariably the associated odds ratios are relatively modest or the susceptibility alleles are rare. Partial or more complete genome scans have been reported using as phenotypes parasite burdens or densities as surrogates of disease susceptibility in schistosomiasis and malaria<sup>14–16</sup>. These studies suggest that susceptibility genes for parasitization and disease may often be different. Recently, a genome-wide linkage analysis of sibpair families with clinical tuberculosis addressed susceptibility to tuberculosis, but found only suggestive evidence of susceptibility loci on chromosomes 15q and Xq (ref. 17). We have reported here significant linkage with another mycobacterial disease, leprosy, on chromosome 10p13 providing proof of concept of the usefulness of genome-wide linkage analysis in a prevalent infectious disease.

The extent of familial clustering of a disease may be measured as the ratio of the risk for siblings of patients to that for the general population ( $\lambda_s$ ; ref. 9). The  $\lambda_s$  values for many autoimmune diseases such as juvenile-onset diabetes and multiple sclerosis are high ( $\lambda_s = 15–20$ ; ref. 9), but  $\lambda_s$  values for many infectious diseases may be much lower (for example, for leprosy in another population in south India,  $\lambda_s$  has been estimated as 2.4; refs. 9, 18). The estimated locus-specific  $\lambda_s$  for the peak of linkage on chromosome 10p13 (calculated from the ratio of the expected proportion of affected sibpairs sharing zero alleles identical by descent (0.25) and the observed proportion<sup>9</sup>) is 1.66. Thus, this locus may make an appreciable contribution to the total genetic component.

Our study demonstrates the usefulness of genome-wide linkage studies of affected sibling pairs in the analysis of a prevalent infectious disease. It maps a major locus for susceptibility to an age-old contagion and indicates that the genetic component to infectious disease susceptibility may not always be so polygenic and diffuse that major loci will not be identifiable.

Note: Supplementary information is available on the Nature Genetics web site ([http://genetics.nature.com/supplementary\\_info/](http://genetics.nature.com/supplementary_info/)).

## Methods

**Design of family studies.** Families with two or more siblings affected by leprosy were identified from patient records in Sakthi Nagar, Kumbakonam and Vizag, South India. All patients had been diagnosed based on the classification of Ridley and Jopling and using WHO guidelines<sup>19</sup>. Both parents were available for all families, except five families studied in the second stage. We included 93 families (including 103 independent affected sibpairs) from Sakthi Nagar and Kumbakonam in the first round of the genome screen, using 388 microsatellite markers (Table 1). All siblings were affected by paucibacillary leprosy, except four sibpairs in which one of the siblings was affected by multibacillary leprosy, reflecting the predominance of paucibacillary disease in this area. There were no families in which all affected siblings had multibacillary leprosy. The second stage of the screen using 37 microsatellite markers was conducted on 131 families including 142 independent affected sibpairs with paucibacillary leprosy from Sakthi Nagar, Kumbakonam and Vizag. A further 8 markers on chromosome 10p13 were then typed on all families to increase the marker density at *D10S548*.

**Genotyping.** A total of 396 highly polymorphic microsatellite markers covering all 22 autosomes and the X chromosome were typed in this genome screen. The average interval between adjacent markers was 12 cM according to the Genethon linkage map<sup>20</sup>. Semi-automated fluorescence genotyping was used to type the microsatellite markers<sup>10</sup>. Optimization of PCR conditions for each set of fluorescein-labelled primers was carried out using a range of annealing temperatures (48–61 °C) and magnesium concentrations (1.0–3.0 mM) on MJ Research machines. The 15  $\mu$ l total reaction mix contained template DNA (50 ng), primers (40 ng each), potassium chloride buffer (2.5 mmol), dNTPs (100  $\mu$ M), MgCl<sub>2</sub> (1.0–3.0 mM) and *Taq* polymerase (0.2 U). Pooled PCR products were electrophoresed through 6% acrylamide gels on 373A DNA sequencers (Perkin-Elmer). DNA fragment sizing and microsatellite marker genotyping were carried out using the GENESCAN 672 and Genotyper software programs (Perkin-Elmer).

**Statistical analysis.** We analysed each microsatellite marker using the SIBPAIR ANALYZE program<sup>21</sup>. SIBPAIR ANALYZE provides a likelihood-based test statistic for linkage which is equivalent to the lod score calculated assuming a simple recessive disease model with phase unknown matings. When parental genotypes are not available the likelihood is a sum of terms corresponding to each of the possible parental genotype combinations weighted by the genotype frequencies calculated assuming Hardy-Weinberg equilibrium.

Maximum likelihood multi-point mapping was performed using the MAPMAKER/SIBS program<sup>17</sup>. This is a multi-point method based on the single-point affected relative pair lod score method<sup>22,23</sup>. Only fully independent affected sibpairs were used in this analysis and MLS lod scores were calculated using the possible triangle constraints<sup>24</sup>. The program uses genotype information for each affected sibpair to infer the identity-by-descent distribution for each point along a chromosome<sup>12</sup>. Markers that showed a  $P$  value of < 0.10 on SIBPAIR ANALYZE or a lod score of > 1 on SIBPAIR ANALYZE or MAPMAKER/SIBS in the first stage of the genome screen were followed up in the second stage. Finally additional markers were used to reduce the genetic distance intervals surrounding the chromosome 10 marker that showed a multipoint lod score of > 3.0 after the second stage (that is, *D10S548*).  $P$  values for the multipoint lod scores were calculated as described<sup>12</sup>.

## Acknowledgments

We thank the authorities and staff of the Hindu Mission Hospital and the Sacred Heart Leprosy Centre, Kumbakonam, Sakthi sugar factory voluntary health services, Erode, Madurai Kamaraj University, Madurai, and the JALMA Institute, Agra, for assistance, permissions and facilities and support; the families for their cooperation and consent; J.H. Edwards for facilitation; and the field teams for assistance. S.M. is an MRC clinical research fellow. A.V.S.H. is a Wellcome Trust Principal Research Fellow.



Received 9 January; accepted 1 March 2001.

1. WHO Elimination of leprosy as a public health problem. *Wkly Epidemiol. Rec.* **73**, 308–312 (1998).
2. Aredath, S.P. The occurrence of leprosy in an eight-member family—a case report. *Lepr. Rev.* **55**, 47–50 (1984).
3. Shields, E.D., Russell, D.A. & Pericak-Vance, M.A. Genetic epidemiology of the susceptibility to leprosy. *J. Clin. Invest.* **79**, 1139–1143 (1987).
4. Chakravarti, M.R. & Vogel, F. A twin study on leprosy, in *Topics in Human Genetics* (ed. Becker, P.E.) 1–123 (Georg Thieme, Stuttgart, 1973).
5. Mohammed Ali, P. & Ramanujam, K. Leprosy in twins. *Int. J. Lepr.* **34**, 405–406 (1966).
6. Abel, L. & Demenais, F. Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island. *Am. J. Hum. Genet.* **42**, 256–266 (1988).
7. Feitosa, M.F., Borecki, I., Krieger, H., Beiguelman, B. & Rao, D.C. The genetic epidemiology of leprosy in a Brazilian population. *Am. J. Hum. Genet.* **56**, 1179–1185 (1995).
8. Todd, J.R., West, B.C. & McDonald, J.C. Human leukocyte antigen and leprosy: study in northern Louisiana and review. *Rev. Infect. Dis.* **12**, 63–74 (1990).
9. Risch, N. Assessing the role of HLA-linked and unlinked determinants of disease. *Am. J. Hum. Genet.* **40**, 1–14 (1987).
10. Davies, J.L. *et al.* A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* **371**, 130–136 (1994).
11. Lander, E.S. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* **11**, 241–247 (1995).
12. Kruglyak, L. & Lander, E.S. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet.* **57**, 439–454 (1995).
13. Abel, L. *et al.* Susceptibility to leprosy is linked to the human NRAMP1 gene. *J. Infect. Dis.* **177**, 133–145 (1998).
14. Marquet, S. *et al.* Genetic localization of a locus controlling the intensity of infection by *Schistosoma mansoni* on chromosome 5q31–q33. *Nature Genet.* **14**, 181–184 (1996).
15. Marquet, S., Abel, L., Hillaire, D. & Dessein, A. Full results of the genome-wide scan which localises a locus controlling the intensity of infection by *Schistosoma mansoni* on chromosome 5q31–q33. *Eur. J. Hum. Genet.* **7**, 88–97 (1999).
16. Rihet, P. *et al.* Malaria in humans: *Plasmodium falciparum* blood infection levels are linked to chromosome 5q31–q33. *Am. J. Hum. Genet.* **63**, 498–505 (1998).
17. Bellamy, R. *et al.* Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. *Proc. Natl. Acad. Sci. USA* **97**, 8005–8009 (2000).
18. Haile, R.W., Iselius, L., Fine, P.E. & Morton, N.E. Segregation and linkage analyses of 72 leprosy pedigrees. *Hum. Hered.* **35**, 43–52 (1985).
19. Ridley, D.S. & Jopling, W.H. Classification of leprosy according to immunity. A five-group system. *Int. J. Lepr. Other Mycobact. Dis.* **34**, 255–273 (1966).
20. Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152–154 (1996).
21. Satsangi, J. *et al.* Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nature Genet.* **14**, 199–202 (1996).
22. Risch, N. Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am. J. Hum. Genet.* **46**, 242–253 (1990).
23. Risch, N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am. J. Hum. Genet.* **46**, 229–241 (1990).
24. Holmans, P. Asymptotic properties of affected-sib-pair linkage analysis. *Am. J. Hum. Genet.* **52**, 362–374 (1993).