

## Lemba origins revisited: Tracing the ancestry of Y chromosomes in South African and Zimbabwean Lemba

H Soodyall, BSc (Hons), MSc, PhD

Division of Human Genetics, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand and National Health Laboratory Service, Johannesburg, South Africa

Corresponding author: H Soodyall (hxsood@global.co.za)

**Background.** Previous historical, anthropological and genetic data provided overwhelming support for the Semitic origins of the Lemba, a Bantu-speaking people in southern Africa.

**Objective.** To revisit the question concerning genetic affinities between the Lemba and Jews.

**Methods.** Y-chromosome variation was examined in two Lemba groups: one from South Africa (SA) and, for the first time, a group from Zimbabwe (Remba), to re-evaluate the previously reported Jewish link.

**Results.** A sample of 261 males (76 Lemba, 54 Remba, 43 Venda and 88 SA Jews) was initially analysed for 16 bi-allelic and 6 short tandem repeats (STRs) that resulted in the resolution of 102 STR haplotypes distributed across 13 haplogroups. The non-African component in the Lemba and Remba was estimated to be 73.7% and 79.6%, respectively. In addition, a subset of 91 individuals (35 Lemba, 24 Remba, 32 SA Jews) with haplogroup J were resolved further using 6 additional bi-allelic markers and 12 STRs to screen for the *extended* Cohen modal haplotype (CMH). Although 24 individuals (10 Lemba and 14 SA Jews) were identified as having the original CMH (six STRs), only one SA Jew harboured the *extended* CMH.

**Conclusions.** While it was not possible to trace unequivocally the origins of the non-African Y chromosomes in the Lemba and Remba, this study does not support the earlier claims of their Jewish genetic heritage.

*S Afr Med J* 2013;103(12 Suppl 1):1009-1013. DOI:10.7196/SAMJ.7297



The Lemba have a strong oral history of non-African origins, and a culture similar to that practiced by Jewish/Arabic people. They live among other larger groups of people in southern Africa, mainly the Venda (Limpopo Province, South Africa (SA)) and Northern Sotho or Pedi in Sekhukhuneland (Mpumalanga, SA) and among the Shona (Kalanga) in the southern parts of Zimbabwe. While some Lemba, particularly those from SA, claim Jewish origins,<sup>[1]</sup> the Jewish link is not universally accepted, and there have been several studies suggesting Islamic connections with Arabs.<sup>[2-4]</sup> Moreover, some of the Lemba who live in Zimbabwe, referred to here as Remba (their name for themselves, since there is no 'L' sound in their spoken language Shona), identify with Arabic ancestry, and several clan names in use are Arabic in origin.<sup>[5]</sup>

According to van Warmelo,<sup>[5]</sup> the ancestors of the Lemba came from a huge town somewhere across the seas, where they were skilled in metalwork, pottery, textiles and ship-building – presumably the same place Mathivha<sup>[1]</sup> refers to as Sena or Sa'na. They came to this country to trade their goods, especially for gold. They began leaving some of their men behind with unsold cargo and thus established posts. They moved further and further inland and became well known with the locals, but did not mix with them as they deemed themselves superior.

A few genetic studies have attempted to elucidate the ancestry of the Lemba. When 'classical' serogenetic markers (unpublished) and mitochondrial DNA restriction fragment length polymorphisms (RFLPs) were used, no differences between the Lemba and other southern African populations could be detected.<sup>[6]</sup> However,

*My interest in the history of sub-Saharan African populations was sparked over 25 years ago when listening to the interesting and captivating stories Professor Trefor Jenkins shared about field trips to various parts of the country, about various groups of people he encountered, and how the use of classical genetic markers added value in studies concerning the prehistory of southern African populations. He also challenged us 'youngsters' to read more widely across disciplines, to consider genetic data in conjunction with data from other fields like archaeology, anthropology, history and palaeontology, and to use genetic information judiciously to refine and/or validate theories concerning population history and their affinities put forward by scholars of other disciplines.*

*Thus, when he was told by a friend and colleague, Dr Margaret Nabaro, a musicologist at the University of the Witwatersrand, about the parallels in music she heard played by Lemba people from Limpopo Province with that played by Jewish people, and when he learnt about certain cultural practices that the Lemba shared with Jews, Professor Jenkins took up her offer to visit the Lemba. They attended the Annual Cultural festival at Sweetwater in Limpopo Province in 1987 and 1988 and, following an invitation from Professor Mathiva, a member of the Lemba Cultural Association (LCA), collected blood samples. I was fortunate to have accompanied Professor Jenkins to meet with Professor Mathiva and members of the LCA and to conduct additional field work in 2000.*

*It gives me great pleasure, as a former PhD student of Professor Jenkins, to write this paper in his honour as we celebrate his contribution to science and to present an updated story using new data on the origins of the Lemba.*

Y-chromosome RFLP studies conducted by Spurdle and Jenkins<sup>[7,8]</sup> provided the first definitive evidence that the male gene pool of SA Lemba was derived, in part, from non-African sources. Based on the 49a/*TaqI* system, 53% of Y chromosomes in the Lemba were assigned to haplotypes that were also found at considerable frequencies in SA Indians (38%) and in Jewish populations from SA, Europe and Yemen (>50%). Although Spurdle and Jenkins<sup>[8]</sup> could not distinguish between Jewish and non-Jewish Y chromosomes at this level of resolution, they concluded that the non-African Y chromosomes in the Lemba were of Semitic origin.

Using a combination of 6 short tandem repeat (STR) markers in conjunction with 6 bi-allelic markers, Thomas *et al.*<sup>[9]</sup> were able to further resolve the Y chromosomes in the Lemba. They found that a particular Y-chromosome haplotype – referred to as the Cohen modal haplotype (CMH) – was present in the Lemba at a frequency of 8.8% (12/136). The CMH had previously been reported at frequencies of 44.9% in Ashkenazi and 56.1% in Sephardic Cohanim (descendants of Jewish priests) and among Ashkenazi and Sephardic Israelites at 13.2% and 9.8%, respectively.<sup>[10]</sup> Thomas *et al.*<sup>[10]</sup> estimated that the CMH originated approximately 2 000 - 3 200 years ago and suggested that this Y-chromosome haplotype was useful in tracing Jewish ancestry. The CMH was subsequently shown to segregate on the background of haplogroup J-12f2a and was present in samples of Ashkenazi, Sephardic and Kurdish Jews (without considering religious status) at frequencies of 7.6%, 6.4% and 10.1%, respectively.<sup>[11]</sup>

Increasing the resolution of Y-chromosome analysis that included 75 binary markers and 22 STR markers, Hammer *et al.*<sup>[12]</sup> genotyped 122 Ashkenazi and 93 non-Ashkenazi Cohanim to further resolve the paternal ancestry of Jewish priests. Haplogroup J, defined by the presence of the 12f2a/M304 mutation, was resolved further into 16 sub-haplogroups defined by 15 binary markers. A new mutation, P58 T→C, defined the most common lineage J-P58\* in the Cohanim and it was on this haplogroup background that the original CMH defined on the basis of the 6-STR marker system was found.<sup>[10]</sup> The inclusion of 6 additional STRs extended this haplotype to 12 STRs which was subsequently referred to as the *extended* CMH.<sup>[12]</sup> Of the 99 J-P58\* Y chromosomes they examined, 87 carried the CMH using the 6-STR marker system. However, when the 12-STR marker system was used, only 43 Y chromosomes were found to have the *extended* CMH. Also, whereas the original CMH was found in several Near Eastern populations, the *extended* CMH, together with its two closely related haplotypes, had a much more restricted distribution and was only found among Cohanim (29.8%) and Israelites (1.5%), but not among the Levites and non-Jewish populations surveyed.<sup>[12]</sup>

Given that the higher resolution at both the haplogroup and haplotype level helped refine the definition of the CMH, its frequency among the Cohanim, and its distribution in other Jewish and non-Jewish populations, the present study revisited the question concerning genetic affinities between the Lemba and Jews. This was done following screening for the *extended* CMH in a sample of SA Lemba and Zimbabwean Remba.

## Methods

### Samples

Blood samples were collected from healthy, unrelated adult volunteers with their informed consent and approval from the Ethics Committee for research on human subjects at the University of the Witwatersrand (protocol no. M980553). Two groups of Lemba were examined in this study – one from Limpopo Province, SA (referred to herein as Lemba) and the Remba from Zimbabwe. The Venda from Limpopo, the people among whom the Lemba reside in this region, and a sample of predominantly Ashkenazi Jewish males from

Johannesburg, SA, were included for comparative analysis in the study. Altogether, 261 individuals were examined for Y-chromosome polymorphisms (76 Lemba, 54 Remba, 43 Venda and 88 SA Jews).

### Typing of Y-chromosome polymorphisms

DNA samples were extracted from peripheral blood using the standard salting-out method.<sup>[13]</sup> Initially, 16 bi-allelic Y-chromosome polymorphisms were used to identify the major haplogroups (Fig. 1A).<sup>[14,15]</sup> Haplogroup designation follows the nomenclature proposed by the Y Chromosome Consortium.<sup>[16]</sup> Subsequently, haplogroup J\* defined by the M304 mutation, was resolved further after screening for the mutations M267, M172, P58, M410, M318 and M12<sup>[12]</sup> using a single-base extension method.<sup>[17]</sup>

Haplotypes were resolved using both bi-allelic and STR polymorphisms. The original CMH was defined using 6 STR loci: DYS19, DYS388, DYS390, DYS391, DYS392 and DYS393 (in this order) that were amplified in two separate multiplex polymerase chain reactions (PCR).<sup>[18]</sup> The 12 STRs (DYS19, DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426 and DYS439 that defined the *extended* CMH in this order) were genotyped in a subset of 91 individuals who were found to have haplogroup J-12f2a or J-M172 Y chromosomes. This was accomplished by using AmpFISTR Y-filer (Applied Biosystems) which amplified 17 Y-STRs: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a, DYS385b, DYS437, DYS438, DYS439, DYS456, DYS458, DYS635 and YGATA H4 (underlined loci part of the *extended* CMH) and the Multiplex II PCR amplification kit (Applied Biosystems) that included the two STRs DYS388 and DYS426.

### Statistical analysis

Y-chromosome variation was analysed at the haplogroup level defined by bi-allelic polymorphisms only, and at the haplotype level defined by both bi-allelic and STR markers. Haplogroup and haplotype diversity within populations was determined as described by Nei.<sup>[19]</sup> Population differentiation based on haplogroup and haplotype frequency data was performed using an exact test<sup>[20]</sup> using Arlequin software (version 3.5).<sup>[21]</sup> Variation within haplogroups was determined by calculating haplotype diversity and mean pairwise differences between haplotypes using Arlequin.<sup>[21]</sup>

## Results

### Haplogroup variation using 16 bi-allelic and 6 STR markers

Typing for the 16 bi-allelic polymorphisms in the total sample of 261 males resulted in the derivation of 13 Y-chromosome haplogroups (Fig. 1A). The bi-allelic markers together with the 6 STR loci defined 102 haplotypes (Appendix 1, available online).

Using the distribution of Y-chromosome haplogroups in global populations,<sup>[14,15]</sup> haplogroups B-SRY10831.1, E-M2 and E-M40 were used to trace the African ancestry of Y chromosomes in the combined Lemba/Remba, whereas haplogroups F\*, J, K\*, L, P\* and R were used to trace the non-African contribution. Although haplogroups E-M40 and E-M35 occur in both African and non-African populations, we were able to resolve the ancestry of the E-M40 chromosomes in the Lemba to African sources and the E-M35 chromosomes to non-African sources using the haplotype data from comparative analysis. Thus, the African contribution of Y-chromosome haplogroups in present-day Lemba and Remba was estimated to be 26.3% and 20.4%, respectively.

A comparison of the haplogroup distribution among Remba, Lemba, Venda and Jews revealed that Remba and Lemba had some haplogroups in common with the Venda and others with Jews (Fig. 1A). Of the 11 haplogroups found in the combined Remba/

Lemba sample, only four (E-M2, J-12f2a, J-M172 and L-M11) were shared between them. Although the four haplogroups together constituted a major component of the Y chromosomes in both Lemba (73.7%) and Remba (94.4%), the frequencies of these haplogroups (with the exception of haplogroup E-M2) differed in the two populations (Fig. 1A). The most common haplogroup in the Remba was J-M172, which accounted for 42.6% of their Y chromosomes, whereas J-12f2a was most common in the Lemba (39.5%). Both J haplogroups were observed in Jews at frequencies of 18.1% and 22.7%, respectively. The CMH (14-16-23-10-11-12), which was resolved on haplogroup J-12f2a was only found in the Lemba (13.2%) and Jewish (15.9%) groups, but not in the Remba (H54, Appendix 1).

With the exception of L-M11, which was only found in Lemba and Remba, all haplogroups derived from non-African

ancestry in the combined Lemba/Remba sample were also found in Jews.

Using the exact test of differentiation, all population pairs, including the Remba v. Lemba, were found to be significantly different from each other using both haplogroup ( $p < 0.001$ ) and haplotype ( $p < 0.001$ ) frequency data. Only 5 (H26, H58, H65, H67, H84) of the 42 haplotypes found in the combined Lemba/Remba sample were shared between the two groups (Appendix 1). Altogether, these lineages accounted for 39.2% of the Y chromosomes in the Lemba/Remba and they shared predominantly higher-frequency haplotypes (Table 1).

There was little divergence, as reflected in the small mean pairwise differences and low haplotype diversity values (Table 2), within haplogroups found in the Lemba/Remba. Several haplogroups were represented only by a single haplotype, some of them at high frequency (Table 2). Other haplogroups

comprised one major haplotype from which few one-step neighbours had evolved (Appendix 1).

### Finer resolution of haplogroup J\* and the extended CMH

Screening for the 6 binary markers within haplogroup J\* in 35 Lemba, 24 Remba and 32 SA Jews resolved haplogroup J\* into 4 sub-haplogroups (Fig. 1B). Haplogroup J-P58 which is associated with the CMH was found in the Lemba (57.1%) and SA Jews (56.3%). Haplogroup J-M267\* was found in the Lemba (28.6%) and Remba (16.0%) but not in the SA Jews. Haplogroup J-M172 was further resolved into sub-haplogroups J-M410\* in the Remba (43.5%) and SA Jews (37.5%) and J-M12\* in the Lemba (14.3%), Remba (44.4%) and SA Jews (6.3%).

The 12-STR marker system resolved the 91 M304 Y chromosomes (Fig. 1B) into 46 haplotypes (Appendix 2, available online). The *extended* CMH (14-13-15-16-13-30-23-10-11-12-11-12) on the J-P58 haplogroup background was only found in 1/14 Jewish individuals who were found to have the original CMH (6 STRs). None of the 10 Lemba who were found to have the original CMH had the *extended* CMH. The most common haplotype 14-13-19-16-13-29-23-10-11-12-11-12 found in eight Lemba differed from the *extended* CMH by 5 mutational steps; 4 at the DYS385b locus and 1 at the DYS389II locus. The closest haplotype in the Lemba to the *extended* CMH differed by 4 mutational steps (3 at the DYS385b locus and 1 at the DYS389II locus).

At this higher level of resolution at both haplogroup and haplotype level, the Lemba and Remba only shared 3 haplotypes – 1 on the J-M267\* background and 2 on the J-M12 background (Appendix 2). None of the combined Lemba/Remba haplotypes were shared with the Jews.

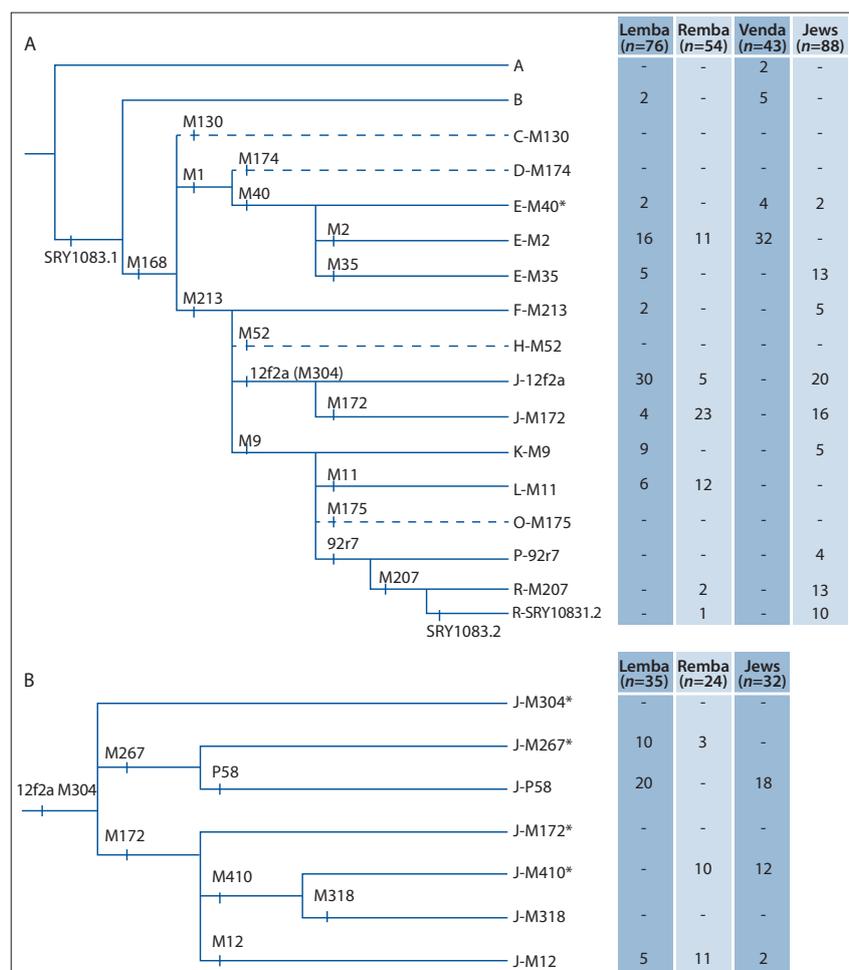


Fig. 1. (A) Phylogeny and distribution of Y chromosome haplogroups defined by the 16 bi-allelic markers used in this study (black bars) in the Lemba, Remba, Venda and South African Jewish populations. The 13 haplogroups found in the total sample are indicated with solid lines, while those haplogroups not found in the sample are shown using dashed lines. (B) Phylogeny showing the finer resolution of haplogroup J chromosomes using 6 additional single nucleotide polymorphisms and the number of individuals from a subset of Lemba, Remba and Jewish samples with these sub-haplogroups.

## Discussion

The Lemba have attracted a great deal of media attention following on Y-chromosome studies that lent support to their claims of having Jewish ancestry.<sup>[7-9]</sup> Given that the refinement in the resolution of Y chromosomes at both a haplogroup and haplotype level delineated an *extended* CMH that was found at a lower frequency than the originally defined CMH among the Jewish priesthood (Cohanim), and had a more restricted global distribution,<sup>[12]</sup> this study sought to revisit the question concerning the origin(s) of non-African Y chromosomes in the Lemba, in particular whether the *extended* CMH was present in the Lemba, and to examine the genetic affinities between the Lemba and Remba.

**Table 1. The 6-STR system haplotypes found at >9% frequency in any of the four populations**

Haplogroup	Haplotype	Lemba (N=76)	Remba (N=54)	Venda (N=43)	Jews (N=88)
E-M2	17-12-21-10-11-14	-	0.093	0.023	-
E-M2	15-12-21-10-11-13	0.079	0.037	0.419 <sup>†</sup>	-
J-12f2a	14-16-24-10-13-12	0.105	-	-	-
J-12f2a*	14-16-23-10-11-12	0.132 <sup>†</sup>	-	-	0.159 <sup>†</sup>
J-12f2a	14-15-24-10-11-12	0.118	0.093	-	-
J-M172	15-15-24-10-11-12	0.039	0.093	-	0.011
L-M11	14-12-22-10-15-12	0.079	0.167 <sup>†</sup>	-	-

STR = short tandem repeat.  
\*Cohen modal haplotype.  
<sup>†</sup>The modal haplotypes in each population.

**Table 2. Haplogroup composition in Remba and Lemba (6-STR system)**

Haplogroup	Population	Haplotypes, <i>n</i>	Individuals, <i>n</i>	Mean pairwise differences*	Haplotype diversity
B-SRY10831.1	Lemba	1	2	0	0
E-M40	Lemba	2	2	4.00	1.0 (±0.5)
E-M2	Lemba	8	16	2.41	0.842 (±0.075)
E-M2	Remba	6	11	3.20	0.8 (±0.114)
E-M35	Lemba	3	5	0.80	0.7 (±0.218)
FI-M213	Lemba	1	2	0	0
J-12f2a	Lemba	6	30	1.92	0.758 (±0.041)
J-12f2a	Remba	1	5	0	0
J-M172	Lemba	2	4	0.50	0.5 (±0.265)
J-M172	Remba	8	23	2.47	0.878 (±0.031)
K,M,N-M9	Lemba	3	9	0.89	0.722 (±0.097)
L-M11	Lemba	1	6	0	0
L-M11	Remba	3	12	0.47	0.439 (±0.158)
R-M207	Remba	1	2	0	0
R-SRY10831.2	Remba	1	1	0	1 (±0)

\*Between haplotypes.

The majority of Y chromosomes found in the Lemba (73.7%) and Remba (79.6%) were traced to non-African origins (Fig. 1). Haplogroups C, D and O were not found, suggesting that populations from geographical regions where haplogroups C and D (Asians) and haplogroup O (Oceanic) are commonly found did not contribute to the Y chromosomes found in the combined Lemba/Remba sample. Haplogroup J was the most common haplogroup in the Lemba/Remba (51.7%) and it was on the background of haplogroup J-12f2a that the original haplotype designation of the CMH based on the 6-STR system was found at frequencies of 9.6% in the Lemba and 15.9% in SA Jews, but not in the Remba (Fig. 1A). When the higher resolution of haplogroup J\* in combination with the 12-STR system was used, the *extended* CMH was only found in one individual of SA Jewish descent (1.1%) and not at all in the Lemba. This finding argues against the claims made previously about possible connections between the Lemba and the Cohanim.<sup>[9]</sup>

Hammer *et al.*<sup>[12]</sup> showed that while haplogroup J-P58\* occurred among both Jewish (18.8%) and non-Jewish (15.5%) populations, the *extended* CMH was restricted to only the Jewish groups from across the range of the Jewish Diaspora. Furthermore, after using the 12-STR marker system to resolve haplotypes associated with haplogroup J-P58 Y chromosomes, the *extended* CMH was found at frequencies of 64.6%

among Cohanim, 10.7% among Israelites and 12.2% in Ashkenazi Jews with unknown castes, but was not found in the Leviim or in 2 099 non-Jewish individuals surveyed.<sup>[12]</sup> Overall, the *extended* CMH was found at a frequency of 5.3% among Jewish populations.<sup>[12]</sup>

Only one of the two closely related haplotypes to the *extended* CMH reported<sup>[12]</sup> was found in a Jewish individual in this study (haplotype 8, Appendix 2). There was no haplotype sharing between the combined Lemba/Remba sample and SA Jews when the 12-STRs were used to derive haplotypes (Appendix 2). Also, there were no matches to any of the J-P58 haplotypes derived in the Lemba/Remba to any of the J-P58 chromosomes reported by Hammer *et al.*<sup>[12]</sup>

The three haplogroup J-M304 chromosome haplotypes shared between the Lemba and Remba (Appendix 2) must have been present in the founding non-African males who contributed to their Y chromosomes. Since none of the haplotypes resolved with the 12 STRs within haplogroups J-M267, J-M410 or J-M12 matched those reported,<sup>[12]</sup> it is still very difficult to unambiguously resolve the origins of these non-African Y chromosomes in the Lemba/Remba. Given the finer resolution of Y chromosomes among the Lemba/Remba, the original claims of close Jewish links are not strongly supported.

More support for the unlikely Jewish link of the Lemba with Jewish populations was reported by Mendez *et al.*<sup>[22]</sup> following the finer resolution of haplogroups L and T. In this study, a novel mutation (P326) was found in both haplogroups L and T. However, additional single nucleotide polymorphisms (SNPs), PS129, PS18 (M70), PS21 and L131 resolved haplogroup T into sub-branches T\*, T1\*, T1a\* (with 6 sub-lineages) and T1b\* (with 3 sub-lineages), respectively. Y chromosomes from Jewish populations clustered in branch T1a\*. In the sample of 34 Lemba included in this study, 6 Lemba Y chromosomes (17.6%) were assigned to the branch T1b\*.<sup>[22]</sup> Although this level of resolution was not genotyped in the present study, T1b\* chromosomes would be represented by haplogroup L-M11 chromosomes found in the Lemba and Remba (Fig 1A). The present-day distribution of haplogroup L-M11 is mostly confined to southern, central and western Asia,<sup>[23]</sup> it has been found only at low frequencies in populations of the Caucasus region and the Middle East and is absent in Palestinian and Syrian Arab.<sup>[11,24-27]</sup>

Overall, this study has shown that Y chromosomes typically linked with Jewish ancestry were not detected by the higher resolution analysis conducted in the present study. It seems more likely that Arab traders, who are known to have established long-distance trade networks involving some thousands of kilometres along the western rim of the Indian Ocean, from Sofala in the south to the Red Sea in the north and beyond to the Hadramut, India and even China from about 900 AD,<sup>[28]</sup> are more likely linked with the ancestry of the non-African founding males of the Lemba/Remba.

Historical data show that after the establishment of a trading hub in Zanzibar Island by Arabs, Indian merchants capitalised on the commercial ventures and settled in East Africa to extend trading ties across the ocean with India.<sup>[28]</sup> By the beginning of the 15th century, Islamic trade and settlement was evident along the East African coast, involving Zanzibar and many other places along the coast, extending to the south to the mouth of the Zambezi River and inland where the borders of Botswana, Zimbabwe and SA meet. This area, in the vicinity of Mapungubwe, was rich in elephant ivory and alluvial gold, and the traditional leaders, already wealthy in cattle, became incredibly wealthy through this coastal trade.<sup>[29,30]</sup> Following a vicious drought in about 1100 AD, the people of Mapungubwe moved north of the Limpopo to found Great Zimbabwe. From here, trade with the East for luxury goods such as textiles (silk and cotton), carpets, glazed pottery from Persia and porcelain bowls from China for African products that included ivory, copper and gold, flourished.<sup>[28]</sup>

## Conclusion

Previous Y chromosome data, particularly the presence of the CMH in SA Lemba, led to claims of genetic links between the Jewish priesthood (Cohanim) and the Lemba. However, higher-resolution genotyping of haplogroup J chromosomes that harboured the CMH led to the delineation of an *extended* CMH that was found at high frequency among the Cohanim and restricted to Jewish groups across the Jewish Diaspora. In this study, the *extended* CMH was only found in a single Jewish individual, and not in the Lemba or the Remba. This finding, together with the lack of matches of haplotypes found in the combined Lemba/Remba group with Jewish populations, suggests that the haplogroup J Y chromosomes in the Lemba/Remba are not closely associated with Jewish ancestry. Rather, this study suggests a stronger link with Middle Eastern populations, probably the result of trade activity in the Indian Ocean. Also, although the SA Lemba and Zimbabwean Remba originated from the same founding population, several differences were observed in the composition and frequencies of their Y chromosome pools, perhaps a consequence of drift and human contacts the groups may have had following their split. However, it should also be stressed that studies on genetic ancestry

makes use of patterns in DNA to look for similarities or differences, and in the present study Y chromosome haplogroups and haplotypes, and that this type of data is in no way aligned with identity testing. Issues of identity are complex and based on many criteria and should not be confused with ancestry testing.

**Acknowledgements.** The author expresses her gratitude to all the participants who donated blood samples that made this study possible; Trefor Jenkins, Bruce Dangerfield, Paul Stidolf and Mr Moeti for their invaluable assistance with fieldwork in the Northern Province and Zimbabwe; Bharti Morar, Carey-Anne Eddy, Almut Nebel, Heeran Makkani, Thijessen Naidoo, Rajeshree Mahabeer and Elize van der Westhuizen for technical assistance and insightful comments to previous drafts of this manuscript. Support for this research was provided jointly by the SA Medical Research Council, National Health Laboratory Service, the University of the Witwatersrand and the National Research Foundation.

## References

- Mathivha MER. The Basena/Vamwenya/Balemba. Toyandu: Morester, 1992.
- Mandivenga EC. The history and 're-conversion' of the Varella of Zimbabwe. *J Relig Afr* 1989;19:98-124.
- Parfitt T. Journey to the Vanished City. The Search for the Lost Tribe of Israel. London: Phoenix, 1997.
- le Roux M. The Lemba. A Lost Tribe of Israel in Southern Africa? Pretoria: UNISA Press, 2003.
- van Warmelo NJ. The Classification of Cultural Groups. In: Hammond-Tooke WD, ed. The Bantu-speaking Peoples of Southern Africa. London: Routledge and Kegan, 1974:56-84.
- Soodyall H. Mitochondrial DNA Polymorphisms in Southern African Populations. PhD thesis. University of the Witwatersrand, 1993.
- Spurdle A, Jenkins T. The search for Y chromosome polymorphism is extended to negroids. *Hum Mol Genet* 1992;1(3):169-170. [<http://dx.doi.org/10.1093/hmg/1.3.169>]
- Spurdle AB, Jenkins T. The origins of the Lemba 'Black Jews' of southern Africa: Evidence from p12F2 and other Y-chromosome markers. *Am J Hum Genet* 1996;59(5):1126-1133.
- Thomas MG, Parfitt T, Weiss DA, et al. Y chromosomes traveling south: The Cohen modal haplotype and the origins of the Lemba - the 'Black Jews of Southern Africa'. *Am J Hum Genet* 2000;66(2):674-686. [<http://dx.doi.org/10.1086/302749>]
- Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB. Origins of Old Testament priests. *Nature* 1998;394(6689):138-140. [<http://dx.doi.org/10.1038/28083>]
- Nebel A, Filon D, Brinkmann B, Majumder PP, Faerman M, Oppenheim A. The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. *Am J Hum Genet* 2001;69(5):1095-1112. [<http://dx.doi.org/10.1086/324070>]
- Hammer ME, Behar DM, Karafet TM, et al. Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood. *Hum Genet* 2009;126(5):707-717. [<http://dx.doi.org/10.1007/s00439-009-0727-5>]
- Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988;16(3):1215. [<http://dx.doi.org/10.1093/nar/16.3.1215>]
- Underhill PA, Passarino G, Lin AA, et al. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 2001;65(Pt 1):43-62. [<http://dx.doi.org/10.1046/j.1469-1809.2001.6510043.x>]
- Underhill PA, Shen P, Lin AA, et al. Y chromosome sequence variation and the history of human populations. *Nat Genet* 2000;26(3):358-361. [<http://dx.doi.org/10.1038/81685>]
- Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 2002;12(2):339-348. [<http://dx.doi.org/10.1101/gr.217602>]
- Naidoo T, Schliebusch CM, Makkani H, et al. Development of a single base extension method to resolve Y chromosome haplogroups in sub-Saharan African populations. *Investig Genet* 2010;1(1):6. [<http://dx.doi.org/10.1186/2041-2223-1-6>]
- Lane AB, Soodyall H, Arndt S, et al. Genetic substructure in South African Bantu-speakers: Evidence from autosomal DNA and Y-chromosome studies. *Am J Phys Anthropol* 2002;119(2):175-185. [<http://dx.doi.org/10.1002/ajpa.10097>]
- Nei M. *Molecular Evolutionary Genetics*. New York: Columbia University Press, 1987.
- Raymond M, Rousset F. An exact test for population differentiation. *Evolution* 1995;49(6):1280-1283. [<http://dx.doi.org/10.2307/2410454>]
- Excoffier L, Lischer HE. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010;10(3):564-567. [<http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x>]
- Mendez FL, Karafet TM, Krahn T, Ostrer H, Soodyall H, Hammer ME. Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa. *Hum Biol* 2011;83(1):39-53. [<http://dx.doi.org/10.3378/027.083.0103>]
- Sengupta S, Zhivotovskiy LA, King R, et al. Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet* 2006 Feb;78(2):202-221. [<http://dx.doi.org/10.1086/499411>]
- Semino O, Passarino G, Oefner PJ, et al. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: A Y chromosome perspective. *Science* 2000;290(5494):1155-1159. [<http://dx.doi.org/10.1126/science.290.5494.1155>]
- Wells RS, Yuldasheva N, Ruzibakiev R, et al. The Eurasian heartland: A continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci* 2001;98(18):10244-10249. [<http://dx.doi.org/10.1073/pnas.171305098>]
- Wilson JB, Weiss DA, Richards M, Thomas MG, Bradman N, Goldstein DB. Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci* 2001;98(9):5078-5083. [<http://dx.doi.org/10.1073/pnas.071036898>]
- Weale ME, Yepiskoposyan L, Jager RF, et al. Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. *Hum Genet* 2001;109(6):659-674. [<http://dx.doi.org/10.1007/s00439-001-0627-9>]
- Oliver R, Atmore A. *Africa Since 1800*. Cambridge: Cambridge University Press, 1981.
- Huffman TN. *Symbols in Stone. Unravelling the Mystery of Great Zimbabwe*. Johannesburg: Witwatersrand University Press, 1987.
- Huffman TN. *Snakes and Crocodiles. Power and Symbolism in Ancient Zimbabwe*. Johannesburg: Witwatersrand University Press, 1996.

Accepted 22 August 2013.