

ORIGINAL ARTICLE

Y-chromosome R-M343 African lineages and sickle cell disease reveal structured assimilation in Lebanon

Marc Haber¹, Daniel E Platt², Simon Khoury¹, Danielle A Badro¹, Miguel Abboud³, Chris Tyler-Smith⁴ and Pierre A Zalloua^{1,5}

We have sought to identify signals of assimilation of African male lines in Lebanon by exploring the association of sickle cell disease (SCD) in Lebanon with Y-chromosome haplogroups that are informative of the disease origin and its exclusivity to the Muslim community. A total of 732 samples were analyzed, including 33 SCD patients from Lebanon genotyped for 28 binary markers and 19 short tandem repeats on the non-recombinant segment of the Y chromosome. Genetic organization was identified using populations known to have influenced the genetic structure of the Lebanese population, in addition to African populations with high incidence of SCD. Y-chromosome haplogroup R-M343 sub-lineages distinguish between sub-Saharan African and Lebanese Y chromosomes. We detected a limited penetration of SCD into Lebanese R-M343 carriers, restricted to Lebanese Muslims. We suggest that this penetration brought the sickle cell gene along with the African R-M343, probably with the Saharan caravan slave trade.

Journal of Human Genetics (2011) 56, 29–33; doi:10.1038/jhgc.2010.131; published online 28 October 2010

Keywords: Lebanon; population genetics; sickle cell disease; Y STRs

INTRODUCTION

Sickle cell disease (SCD) is an inherited autosomal recessive disorder of the β -globin chain, widely distributed throughout equatorial Africa reaching heterozygote frequencies as high as 30–40% in parts of Uganda, West Africa and Zaire,¹ because it confers resistance to malaria. The sickle gene spread to North Africa and the Mediterranean along Saharan caravan trading routes, to North and South America, and the Caribbean by slave trade, and then to Northern Europe by immigration from the Caribbean.² SCD is also found in Asia mainly in the Arabian peninsula and Central India, a region historically hyper-endemic for malaria.³

Linkage analysis of the β -globin sickle mutations suggest homoplasmy,¹ with African and Asian origins distinguished by their haplotypes.⁴ Previous studies of sickle mutations in the Lebanese population showed that 91.5% of the patients were Sunni and Shiite Muslims⁵ (Sunni and Shiites make 58% of the general Lebanese population⁶) and that 90% of SCD subjects carried an African haplotype comprising the Benin (73%), Central African Republic (15%) and Senegal (2%) haplotypes. The remaining 10% of the Lebanese SCD patients carried the Indian–Arab haplotype.⁷

The origin, time of spread or details of the association of SCD with Lebanese Muslims cannot be deduced from the genetic content of sickle haplotypes. The sickle mutation is estimated to have arisen 3000–6000 generations ago,^{8,9} whereas the haplotypes surrounding the

β -globin locus are even older, limiting specificity of Lebanese SCD geographic origin.

We have previously used Y-chromosome markers to study population structure and migration into Lebanon.¹⁰ Male genetic variation within Lebanon was found to be strongly structured by religious affiliation and influenced by recent historical events. The Islamic expansion from the Arabian Peninsula beginning in the seventh century CE appears to have introduced lineages typical of this area into the Lebanese Muslim community,¹⁰ whereas the Crusader activity in the eleventh–thirteenth centuries CE introduced Western European lineages into Lebanese Christians.¹⁰ As SCD is closely affiliated with specific Lebanese communities, which correlate with Y-chromosome diversity, SCD and Y-chromosome polymorphisms could reveal structure in migrant assimilation. Indeed, the time of spread of the sickle mutation across the Sahel has been previously inferred from a Y-chromosome single nucleotide polymorphism that was co-integrated with the sickle gene by West African tribes migrating to the East.¹¹

In this study, we investigated the penetration of the sickle-cell mutation into Lebanon, and explored the association between certain African Y-chromosome lineages and the chromosome 11 sickle cell mutation in order to determine whether there is a sign of persistent population structure characterizing assimilation of Africans in Lebanon. We also investigated the sickle cell penetration into the broader Lebanese population through examination of a set of sickle

¹Medical School, The Lebanese American University, Beirut, Lebanon; ²Computational Biology Center, IBM TJ Watson Research Centre, Yorktown Hgts, NY, USA; ³Department of Pediatrics, American University of Beirut Medical Center, Beirut, Lebanon; ⁴The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambs, UK and ⁵Harvard School of Public Health, Boston, MA, USA

Correspondence: Dr PA Zalloua, Medical School, The Lebanese American University, Chouran, Beirut 1102 2801, Lebanon.

E-mail: pierre.zalloua@lau.edu.lb

Received 20 July 2010; revised 21 September 2010; accepted 1 October 2010; published online 28 October 2010

cell patients bearing non-African Y haplogroups. To achieve these goals, from among a set of candidate Y-chromosome haplogroups, we identified a haplogroup whose short tandem repeat (STR) haplotypes clearly differentiated African from non-African origins. We extended a standard Hardy–Weinberg equilibration argument to show that equilibration among Y-chromosome and chromosome 11 sickle-cell prevalence would be expected to proceed rapidly under panmictic association. Therefore, we sought to test whether there was evidence of population structure by examining Y-haplotype association with sickle cell, and also to test whether there was penetration into the population at large by examining sickle-cell penetration into Lebanese Y-chromosome haplogroups not common in the region of Africa where sickle cell emerged.

MATERIALS AND METHODS

Subjects and comparative datasets

The populations selected for this study were those who had potentially influenced the genetic structure of the Lebanese population¹² in addition to African populations with high incidence of SCD.

In all, 33 SCD patients from Lebanon (28 Muslims, 5 Christians) were genotyped and analyzed. The comparative dataset included 699

samples newly genotyped or obtained from published sources. New samples comprised Libya (33) and Kuwait (8). Data from Lebanon (250), Egypt (84), Iran (56), Malta (20), Central Africa (46), Algeria (16), Italy (103), KSA (50), UAE (18), Uganda (10) and Yemen (5) was obtained from published sources.^{10,13–20} All samples had an unknown SCD status except for the 33 SCD patients from Lebanon. SCD male subjects were used from our database⁷ after they were anonymized and approved for this study by the institutional review board of the Lebanese American University. All other participants provided detailed information on their geographical origin, and gave informed consent approved by the institutional review board of the Lebanese American University.

Genotyping

DNA was extracted from blood using a standard phenol–chloroform method. Samples were genotyped using the Applied Biosystems 7900HT Fast Real-Time PCR System with a set of 28 custom Y-chromosomal binary marker assays (Applied Biosystems, Foster City, CA, USA) from the non-recombining portion of the Y chromosome, which define 21 haplogroups. The new samples were additionally amplified at 19 Y-chromosomal STR loci in two multiplexes and analyzed on an Applied Biosystems 3130xl Genetic Analyzer. The first

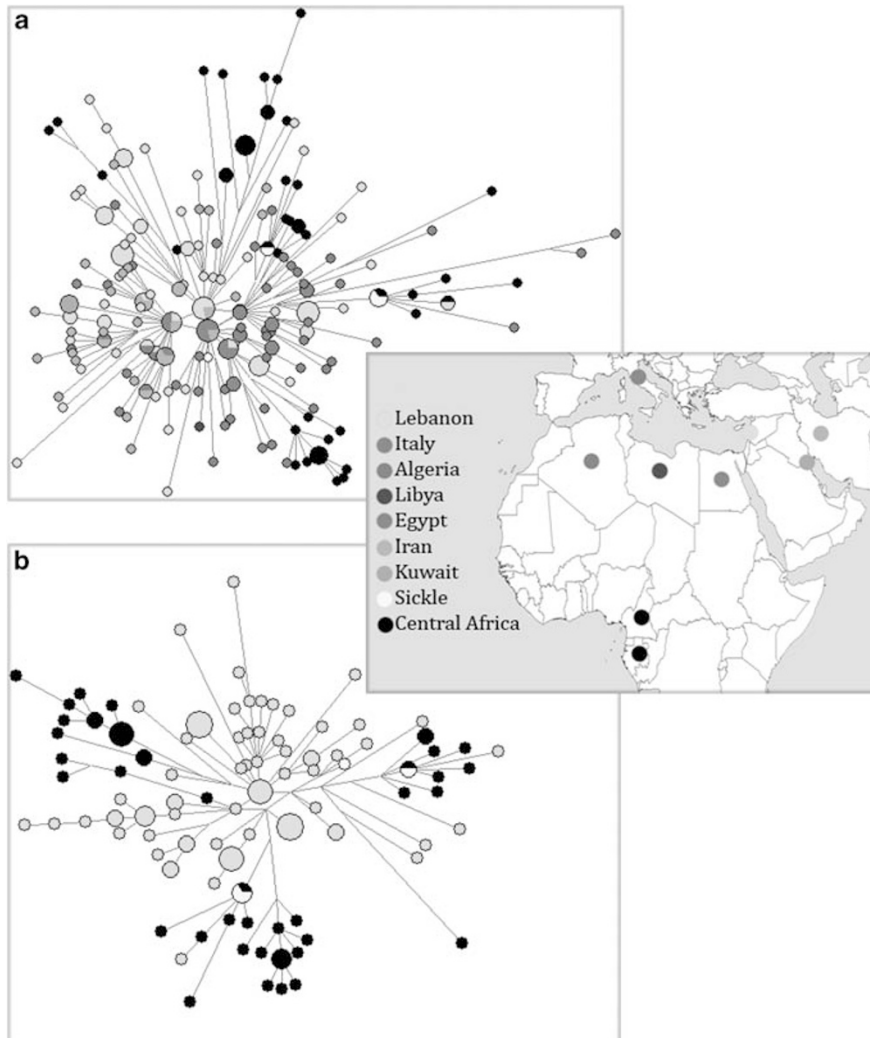


Figure 1 Sample distribution and reduced median network of R-M343. The map shows the location of the samples used in the networks and the respective color for every population. (a) Reduced median network of R-M343 from all indicated populations in the map. (b) Reduced median network of R-M343 from Lebanon, Central Africa and Lebanese sickle-cell disease patients.

multiplex contained the standard 17 loci of the Y-filer PCR Amplification kit (Applied Biosystems). The remaining two loci, DYS388 and DYS426, were genotyped in a separate custom multiplex provided by Applied Biosystems.

Statistical analyses

We sought to identify Y-chromosome haplogroups common in the same region of Africa as sickle cell, which would have likely been brought to Lebanon through the slave trade. Of those, we sought to identify which of those whose STR haplotypes showed clear regional definition. From R-M343, J-M172, J-M267 and E-M35, we found that R-M343 had regionally differentiating Y STR haplotypes identified by reduced median networks²¹ computed with a reduction coefficient of 1, corroborated with multidimensional scaling (MDS)²² applied to Φ_{ST} as distances computed in ARLEQUIN,²³ and validated using analysis of molecular variance.²⁴

The reduced median network and analysis of molecular variance analyses were applied to the Y-STR loci of R-M343 that were common to this study and the literature: DYS393, DYS390, DYS19, DYS391, DYS439, DYS389I, DYS392, DYS389b, DYS437, DYS438. Analysis of molecular variance identifies variance within populations by comparing variation among groups of similar populations via a nested analysis of variance. MDS was executed with SPSS16 in two dimensions with a minimum S-stress of 0.005.

Likewise, non-African haplogroups were identified, marking penetration of African sickle cell mutations into the general Lebanese population.

We sought to test whether there would be a correlation between African haplotypes and sickle cell. As sickle cell is very rare in Lebanon, all of our sickle-cell subjects were selected from patients. However, prevalence statistics are generally not available for the population in general, nor among R-M343 specifically. This limits the possibility of computing a linkage disequilibrium characterizing the population. However, it is possible to test whether there is an association of sickle cell among African versus non-African R-M343 haplotypes under panmictic breeding. To achieve this, an extension to the Hardy–Weinberg²⁵ equilibration model was constructed analytically following standard arguments,²⁶ assuming complete replacement each generation, in which children randomly select both parents (sexes treated distinctly), randomly select a diploid chromosome 11 from each parent and inherit the haploid Y chromosome from the father. Deviations from such random mating imply population structure. The extended Hardy–Weinberg equilibration process showed the heterozygous diploid sickle cell markers among females equilibrating immediately, whereas the deviation from equilibrium for the African Y versus sickle mutations reduced by a factor of 2 each generation. At equilibrium, the Y African markers and the sickle markers are independent. This implies that, if seven or more generations have occurred since African input, the deviation from equilibrium will not exceed 0.8%. Therefore, it is possible to test for population structure, testing whether $P(\text{SCD}|\text{R-M343}=\text{African})=P(\text{SCD}|\text{R-M343})$ by using a Fisher's exact test for dependence between the Y African expansion marker and the sickle.

RESULTS

Y genotyping of the SCD patients revealed six haplogroups (14 J-M267, 5 J-M172, 8 E-M35, 4 R-M343, 1 T-M70, 1 N-LLY22g) all previously reported in the general Lebanese population.^{10,16,20}

Reduced median networks of R-M343 (Figure 1) showed significant regional differentiation between Central Africans, Europeans and Lebanese. The probability that random region assignments would

Table 1 Analysis of molecular variance results comparing variations among Europeans, Central Africans and Lebanese

	$V_a:F_{CT}$ among groups	$V_b:F_{SC}$ within groups	$V_c:F_{ST}$ within populations
Value	0.896; 0.201	0.287; 0.080	3.280; 0.265
Variation (%)	20.07	6.43	73.49
P-value	<0.001	<0.001	<0.001

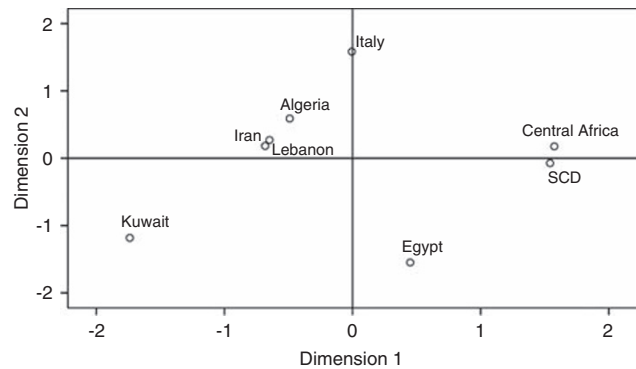


Figure 2 MDS plot of Φ distances between populations derived from Y-STR data of R-M343.

have isolated 46 Africans out of the 48 in the segregated branches, being drawn from 46 out of 124 by chance has a Fisher P -value <0.001. Corollary is that the two Lebanese samples in those branches likely represent African descent. Both samples were identified as Lebanese Muslims. The SCD samples were also contained in African branches. This result is also supported by the MDS plot, as well as significant analysis of molecular variance variation between the groups (20.07%, $P<0.001$, Table 1). R-M343 SCD subjects were all Muslims and clustered with Central African haplotypes. The SCD-Central African Φ_{ST} distances was the lowest ($\Phi_{ST}=0.048$) with no significant differentiation ($P=0.33 \pm 0.03$) compared with all other populations, including the Lebanese ($\Phi_{ST}=0.215$) with significant segregation ($P=0.027 \pm 0.019$). MDS of the Φ_{ST} 's showed strong clustering of the SCD with Central Africans outlying all other populations (Figure 2). It should be noted that, for the most part, the African haplotypes maintained integrity when NETWORK data included European and North African with the Lebanese and African R-M343's, though some variation, including the assignment of one of the SCDs is noted when European and North African sets are excluded. Application of the Fisher exact test for three out of four SCD subjects having African-associated R-M343 STR haplotypes, compared with 2 African of 78 R-M343 samples drawn from the general Lebanese population has a P -value of 0.000443.

Reduced median networks of haplogroups E-M35, J-M267 and J-M172 (Figure 3) show that SCD haplotypes are present only in the general Lebanese population indicating penetration of the sickle gene into the population at large.

DISCUSSION

We have shown that R-M343 STR haplotypes clearly discriminate between Lebanese and African origins, unlike other common haplogroups, which also are not represented in the African populations. This enables identification of African chromosomes that have migrated to Lebanon. Further, SCD among R-M343 Lebanese subjects

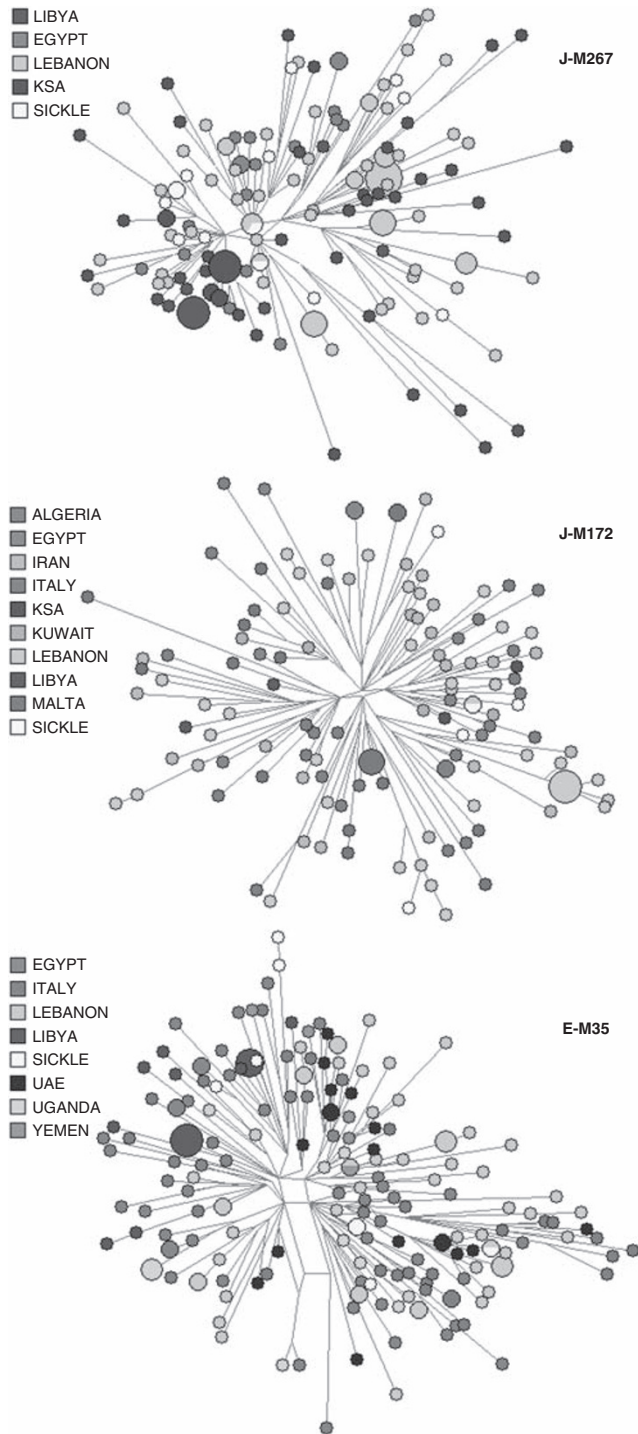


Figure 3 Reduced Median Networks of J-M267, J-M172, and E-M35.

is largely confined to African haplotypes. The strong association of the sickle gene with the African R-M343 would suggest the source populations carried a high frequency of both markers. The sickle mutation is very common to African populations unlike R-M343, which is rare in Africa and found mainly in Europe and Asia. However, R-M343 has been found in high concentration in some populations from Central-West Africa where it reaches very high

frequencies (up to 95%) in populations in the central Sahel; northern Cameroon, northern Nigeria, Chad and Niger.^{27,28} For many centuries African slaves were drawn from those populations and driven across the central Sahara to the Mediterranean ports, a regular trade seems to have been established in the early Islamic era²⁹ and continued until the 1840's^{29,30} confined then to the central and eastern basins of the Mediterranean still under Ottoman influence.²⁹ Deviations from Hardy-Weinberg equilibrium of diploid SCD in both genders with haploid male African markers in a Wright-Fisher model decay by a factor of 2 per generation. Therefore, an unstructured population would have SCD nearly uniformly distributed among Lebanese R-M343 haplotypes, however the R-M343 SCD subjects were mainly of African haplotypes, and found only among Lebanese Muslims.

The probability that 28 or more out of 33 samples drawn from a population would be Muslim when drawn from a population with 58% Muslim is estimated to be 9.50×10^{-4} by binomial test. Conversion to Islam provided a means to better treatment and possible later emancipation,³⁰ and prohibition of non-Muslim slave ownership promoting slave conversions to Islam solely,³¹ explains SCD association with Muslim subjects. However, the limited penetration of SCD into Lebanese R-M343 contradicts a picture of many centuries of conversion and assimilation refreshed by caravan trade, implying a persistently structured population, even allowing for SCD introduction through the female line. Even though equilibration was not observed among R-M343 subjects, mixing was reflected by penetration to non-African haplogroups J-M267, J-M172 and E-M35. In all, this study reveals subtle population structure surviving for over two centuries in Lebanon, with significant medical consequences.

ACKNOWLEDGEMENTS

We thank the sample donors for taking part in this study. CTS is supported by The Wellcome Trust.

- 1 Serjeant, G. R. Geography and the clinical picture of sickle cell disease. An overview. *Ann. NY. Acad. Sci.* **565**, 109–119 (1989).
- 2 Serjeant, G. R. Sickle-cell disease. *Lancet* **350**, 725–730 (1997).
- 3 Kulozik, A. E., Wainscoat, J. S., Serjeant, G. R., Kar, B. C., Al-Awamy, B., Essan, G. J. *et al.* Geographical survey of beta S-globin gene haplotypes: evidence for an independent Asian origin of the sickle-cell mutation. *Am. J. Hum. Genet.* **39**, 239–244 (1986).
- 4 Kamel, K. Heterogeneity of sickle cell anaemia in Arabs: review of cases with various amounts of fetal haemoglobin. *J. Med. Genet.* **16**, 428–430 (1979).
- 5 Inati, A., Jradi, O., Tarabay, H., Moallem, H., Rachkidi, Y., El Accaoui, R. *et al.* Sickle cell disease: the Lebanese experience. *Int. J. Lab. Hematol.* **29**, 399–408 (2007).
- 6 Douaihy, Y. C. Z. *The Lebanese 1907–2006: Confessions, Religions and Demographic Distribution* (Beirut, 2007).
- 7 Inati, A., Taher, A., Bou Alawi, W., Koussa, S., Kaspar, H., Shbaklo, H. *et al.* Beta-globin gene cluster haplotypes and HbF levels are not the only modulators of sickle cell disease in Lebanon. *Eur. J. Haematol.* **70**, 79–83 (2003).
- 8 Kurnit, D. M. Evolution of sickle variant gene. *Lancet* **1**, 104 (1979).
- 9 Solomon, E. & Bodmer, W. F. Evolution of sickle variant gene. *Lancet* **1**, 923 (1979).
- 10 Zalloua, P. A., Xue, Y., Khalife, J., Makhoul, N., Debiante, L., Platt, D. E. *et al.* Y-chromosomal diversity in Lebanon is structured by recent historical events. *Am. J. Hum. Genet.* **82**, 873–882 (2008).
- 11 Bereir, R. E., Hassan, H. Y., Salih, N. A., Underhill, P. A., Cavalli-Sforza, L. L., Hussain, A. A. *et al.* Co-introgression of Y-chromosome haplogroups and the sickle cell gene across Africa's Sahel. *Eur. J. Hum. Genet.* **15**, 1183–1185 (2007).
- 12 Hitti, P. *Lebanon in History from the Earliest Times to the Present* (Macmillan, London, 1957).
- 13 Abu-Amero, K. K., Hellani, A., Gonzalez, A. M., Larruga, J. M., Cabrera, V. M. & Underhill, P. A. Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions. *BMC Genet.* **10**, 59 (2009).
- 14 Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mouguiama-Daouda, P. *et al.* Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Mol. Biol. Evol.* **26**, 1581–1589 (2009).
- 15 Cadenas, A. M., Zhivotovskiy, L. A., Cavalli-Sforza, L. L., Underhill, P. A. & Herrera, R. J. Y-chromosome diversity characterizes the Gulf of Oman. *Eur. J. Hum. Genet.* **16**, 374–386 (2008).

- 16 El-Sibai, M., Platt, D. E., Haber, M., Xue, Y., Youhanna, S. C., Wells, R. S. *et al*. Geographical structure of the Y-chromosomal genetic landscape of the Levant: a coastal-inland contrast. *Ann. Hum. Genet.* **73**, 568–581 (2009).
- 17 Gomes, V., Sanchez-Diz, P., Amorim, A., Carracedo, A. & Gusmao, L. Digging deeper into East African human Y chromosome lineages. *Hum. Genet.* **127**, 603–613 (2010).
- 18 Onofri, V., Alessandrini, F., Turchi, C., Fraternali, B., Buscemi, L., Pesaresi, M. *et al*. Y-chromosome genetic structure in sub-Apennine populations of Central Italy by SNP and STR analysis. *Int. J. Legal. Med.* **121**, 234–237 (2007).
- 19 Robino, C., Crobu, F., Di Gaetano, C., Bekada, A., Benhamamouch, S., Cerutti, N. *et al*. Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. *Int. J. Legal. Med.* **122**, 251–255 (2008).
- 20 Zalloua, P. A., Platt, D. E., El Sibai, M., Khalife, J., Makhoul, N., Haber, M. *et al*. Identifying genetic traces of historical expansions: phoenician footprints in the Mediterranean. *Am. J. Hum. Genet.* **83**, 633–642 (2008).
- 21 Bandelt, H. J., Forster, P., Sykes, B. C. & Richards, M. B. Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743–753 (1995).
- 22 Kruskal, J. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **1**–27 (1964).
- 23 Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50 (2005).
- 24 Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
- 25 Hartl, D. L. & Clark, A. G. *Principles of Population Genetics*, 4th edn. (Sinauer Associates, Sunderland, MA, 2007).
- 26 Felsenstein, J. *Theoretical Evolutionary Genetics* (e-book available from Department of Genome Sciences, University of Washington, Seattle at: <http://evolution.gs.washington.edu/pgbook/pgbook.html> 2009).
- 27 Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A. *et al*. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am. J. Hum. Genet.* **70**, 1197–1214 (2002).
- 28 Cruciani, F., Trombetta, B., Sellitto, D., Massaia, A., Destro-Bisol, G., Watson, E. *et al*. Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *Eur. J. Hum. Genet.* **18**, 800–807 (2010).
- 29 Wright, J. The Mediterranean middle passage: the nineteenth century slave trade between Tripoli and the Levant. *J. North Afr. Stud.* **1**, 42–58 (1996).
- 30 Lovejoy, P. *Transformations in Slavery: A History of Slavery in Africa* (CUP, Cambridge, 1983).
- 31 Ben-Naeh, Y. Blond, tall, with honey-colored eyes: jewish ownership of slaves in the ottoman empire. *Jewish History* **20**, 315–332 (2006).