

ARTICLE

Determination of population origin: a comparison of autosomal SNPs, Y-chromosomal and mtDNA haplogroups using a Malagasy population as example

Micaela Poetsch^{*1}, Aline Wiegand¹, Melanie Harder^{2,3}, Rowena Blöhm¹, Noel Rakotomavo⁴, Sandra Freitag-Wolf⁵ and Nicole von Wurmb-Schwarck^{2,3}

Y-chromosomal and mitochondrial DNA (mtDNA) polymorphisms have been used for population studies for a long time. However, there is another possibility to define the origin of a population: autosomal single-nucleotide polymorphisms (SNPs) whose allele frequencies differ considerably in different populations. In an attempt to compare the usefulness of these approaches we studied a population from Madagascar using all the three mentioned approaches. Former investigations of Malagasy maternal (mtDNA) and paternal (Y chromosome) lineages have led to the assumption that the Malagasy are an admixed population with an African and Asian-Indonesian heritage. Our additional study demonstrated that more than two-third of the Malagasy investigated showed clearly a West African genotype regarding only the autosomal SNPs despite the fact that 64% had an Asian mtDNA and more than 70% demonstrated an Asian-Indonesian heritage in either mtDNA or Y-chromosomal haplogroup or both. Nonetheless, the admixture of the Malagasy could be confirmed. A clear African or Asian-Indonesian heritage according to all the three DNA approaches investigated was only found in 14% and 1% of male samples, respectively. Not even the European or Northern African influences, detected in 9% of males (Y-chromosomal analysis) and 11% of samples (autosomal SNPs) were consistent. No Malagasy in our samples showed a European or Northern African origin in both categories. So, the analysis of autosomal SNPs could confirm the admixed character of the Malagasy population, even if it pointed to a greater African influence as detectable by Y-chromosomal or mtDNA analysis.

European Journal of Human Genetics (2013) 21, 1423–1428; doi:10.1038/ejhg.2013.51; published online 24 April 2013

Keywords: population origin; Madagascar; Y chromosome; mtDNA; SNP

INTRODUCTION

The island of Madagascar in the Indian Ocean was probably the last great island landmass to be settled permanently by humans, probably during the Austronesian diaspora in the sixth century AD.¹ However, human activities may have taken place on Madagascar hundreds of years before this time, probably as early as 300 BC.^{2,3} The present population, the Malagasy, are considered an admixed population demonstrating morphological and cultural traits of Bantu and Austronesian speakers, which has been confirmed by analysis of Y-chromosomal and mitochondrial DNA (mtDNA).^{4–7} The various ethnic groups on Madagascar are generally classified into two main subgroups: the Highlanders (HL), considered being more Asian in culture and phenotype, and the people from the coasts (CT), thought to be more African.⁸ The determination of the phenotype, for example, eye colour, hair colour or skin colour, or the population origin from DNA has been an important research field in recent years, in which considerable progress has been accomplished (see for example^{9–15}). These techniques have already been applied successfully in the characterisation of ancient skeletal human remains.^{16,17}

Regarding these results, we were interested in how the Malagasy would perform in such an assay for the determination of the

phenotype or the population affiliation. We applied our recently published assays for population differentiation¹⁵ in addition to a large set of well known and commonly used binary markers as well as short tandem repeats (STRs) from the Y chromosome and mtDNA single-nucleotide polymorphisms (SNPs) to 168 unrelated individuals from Madagascar.

MATERIALS AND METHODS

Samples

Buccal swabs were collected from 106 men and 62 women from Madagascar (18–72-years-old). They belong to nine different ethnic groups: Antandroy (27 women, 71 men), Antankarana (2 women), Antanosy (1 woman, 1 man), Bara (4 women, 3 men), Betsileo (7 women, 2 men), Betsimisaraka (1 woman, 1 man), Mahafaly (1 woman, 2 men), Merina (12 women, 14 men) and Sakalava (7 women, 12 men). Only the information about age, sex and ethnic origin was collected for each sample.

Samples were obtained and analysed after advice of the Medical Ethics Committees of the University of Duisburg-Essen and the University Hospital of Schleswig-Holstein in accordance with the declaration of Helsinki. All the individuals gave informed consent. The anonymity of the individuals investigated was preserved corresponding to the rules of data protection of the Human Medical Faculties of the Universities Duisburg-Essen and Kiel.

¹Institute of Legal Medicine, University Hospital Essen, Essen, Germany; ²Institute of Legal Medicine, University of Schleswig-Holstein, Kiel, Germany; ³Graduate School 'Human Development in Landscapes', Christian-Albrechts-University Kiel, Kiel, Germany; ⁴Clinic Saint Luc, Toliara, Madagascar; ⁵Institute of Medical Informatics and Statistics, University of Schleswig-Holstein, Kiel, Germany

*Correspondence: Professor Dr rer nat M Poetsch, Institute of Legal Medicine, University Hospital Essen, Hufelandstr. 55, Essen D-45122, Germany. Tel: +49 201 723 3623; Fax: +49 201 723 5940; E-mail: micaela.poetsch@uk-essen.de

Received 25 October 2012; revised 13 February 2013; accepted 20 February 2013; published online 24 April 2013

DNA extraction

DNA was extracted from buccal swabs using the innuPrep DNA Mini Kit (Analytik Jena, Jena, Germany).

SNP analysis

1. Sixteen SNPs in two SNaPshot assays (assay PD-1 comprises SNPs rs885479, rs2228478, rs1545397, rs12913832, rs6119471, rs1426654, rs16891982 and rs1800407, and assay PD-2 comprises rs885479, rs1800404, rs16891982, rs1528460, rs10843344, rs2814778, rs730570, rs1321333, rs5030240, rs4540055) were amplified as described before.¹⁵
2. Forty-four SNPs from the Y chromosome were analysed to differentiate nine major Y-chromosomal haplogroups and their subgroups. Eleven of those have been found in Madagascar. The analysis was done in five different multiplex assays (see Supplementary Table S1b for combination of SNPs) and in a hierarchical system, that is all samples were analysed with multiplex 1 to determine the major haplogroup and only with those subsequent multiplexes which reasonably applied (Supplementary Figure S1).
3. Twelve mtDNA SNPs were analysed to determine all major African, Asian and European mtDNA haplogroups. Moreover, 15 mtDNA SNPs were used to subdifferentiate those haplogroups found in our study in the Malagasy. The analysis was done in three different multiplex assays (see Supplementary Table S1c for combination of SNPs) and in a hierarchical system, that is all samples were analysed with multiplex 1 to determine the major haplogroup and only with those subsequent multiplexes which reasonably applied (Supplementary Figure S2).

For all SNP analyses, multiplex PCR was done in a volume of 12.5 µl in the GeneAmp PCR system 9700 (Applied Biosystems, Darmstadt, Germany) with 2–5 ng of DNA as template in 15 mM Tris/HCl, 50 mM KCl, with 200 µM dNTPs, 1.5 mM MgCl₂, 0.1 µM of each primer and 1.5 Units AmpliTaq Gold Polymerase (Applied Biosystems). PCR conditions were: initial denaturation and activation step of 8 min at 95 °C, 30 cycles of 1 min at 94 °C, 1 min at 55 to 60 °C (see Supplementary Table S1) and 2 min at 72 °C, and a 60 min final elongation step at 60 °C. SNaPshot analyses were performed in accordance with the manufacturer's instructions and evaluated on an ABI310 Genetic Analyser. Electrophoresis results were analysed using the GeneMapper ID Software v3.2 (Applied Biosystems) with self-designed panels and bins sets. The primer sequences for all mtDNA and Y-chromosomal multiplex PCRs and mini sequencing assays (SNaPshot) applied in this study are displayed in Supplementary Table S1.

Y-STR analysis

Analysis of 11 Y-chromosomal STRs was done with the Powerplex Y kit (Promega, Mannheim, Germany) according to the manufacturer's instructions. The haplotypes were then compared with those in the YHRD database (www.yhrd.org), first by entering the results of all 11 analysed STRs. If no match could be found here, an additional search was conducted using only the minimal haplotype comprising eight STRs.

mtDNA sequencing analysis

The amplification of hypervariable regions 1 and 2 was performed in a single PCR assay with 0.4 µM of primers 15938F and 429R using 1 U Immolase (Bioline, Luckenwalde, Germany, GB) with a MgCl₂ concentration of 1.5 mM, a dNTP concentration of 0.2 mM for each dNTP and 1 × Immobuffer in a total reaction volume of 25 µl, including 5–10 ng of DNA extract. The amplifications were performed in the GeneAmp PCR System 2700 (Applied Biosystems) under following cycling conditions: an initial denaturation stage of 95 °C for 10 min, 38 cycles of 94 °C for 1 min, 60 °C for 1 min and 72 °C for 1 min, and a final extension step of 72 °C for 30 min. The clean-up was carried out using the MinElute PCR Purification Kit (Qiagen, Hilden, Germany) following the supplier's protocol.

The sequencing reactions were performed in the GeneAmp PCR System 2700 using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) following the supplier's introduction. For this reaction the primers

15938F, 429R and 16442F were used. For the second clean-up the DyeEx 2.0 Spin Kit (Qiagen) was applied.

Final analysis was carried out using the ABI Prism 3130xl Genetic Analyser (Applied Biosystems) and the Lasergene software (DNA Star, Madison, WI, USA). The haplogroup classifications (with inclusion of mtDNA SNPs as described above) were done using the website HaploGrep (<http://haplogrep.uibk.ac.at>) based on the PhyloTree.org. database release 14.^{18,19}

Statistical analysis

The calculation of population differentiation from the results of the autosomal SNP analysis has been done as described before.¹⁵ The structures of the Y-chromosomal STR loci is illustrated by phylogenetic analysis, which was performed by Network 2.0 with the median joining (MJ)-network method.²⁰ Two different colourings were applied to the networks according to Y haplotype groups and population differentiation SNPs based on correspondence analysis similarities of Y chromosome and mtDNA haplogroup frequencies from African, Asian and European populations are summarised in two-dimensional graphics with the R paket ca.²¹

RESULTS AND DISCUSSION

The samples investigated in this study were grouped into two categories comparable with those chosen by Tofanelli *et al.*⁵ HL and CT. The category HL is composed of samples from the ethnic groups Merina, Bara and Betsileo; the category CT comprises samples from the ethnic groups Antandroy, Sakalava, Mahafaly, Antankarana, Antanosy and Betsimisaraka.

Maternal lineages

In the 168 Malagasy samples analysed in this study, 90 different haplotypes as determined by SNP analysis and sequencing of hypervariable regions 1 and 2 were found (84 haplotypes disregarding the mutation hotspots in 16519, 309.1 and 309.2) (see Supplementary Table S2 for detailed results). All these haplotypes could be assigned clearly to either African heritage (41 haplotypes and 36% of samples) or Asian heritage (43 haplotypes and 64% of samples) (Figure 1), which is in line with results from Hurles *et al.*⁴ and Tofanelli *et al.*⁵ The African maternal lineages were significantly more diverse than the Asian maternal lineages ($P = 0.0007$). Minor differences could be detected between the two chosen categories and the ethnic groups. The African heritage was more prominent in HL samples (45%) than in CT samples (33%). Haplotypes occurring in more than three samples were always shared by people from CT and HL subgroups. The Malagasy motif (a variant of the Polynesian motif, haplogroup B4a1a1a, with the additional polymorphisms 1473 and 3423A⁶) was found in 14 and 19% of HL and CT samples, respectively. No samples were found that demonstrated the Polynesian motif without the additional polymorphisms. We also detected the rare subgroup M23²² in five different haplotypes (2 and 9% of HL and CT samples, respectively). Regarding the African haplogroups, L0 and subtypes were found predominantly in CT samples, whereas L2 and L3 samples were shared between HL and CT samples. The latter is also true for Asian-Indonesian haplogroups M7c3c, M32, E1a1a or F3b1. A comparison of the mtDNA haplogroup distribution in Madagascar with Asian or African populations^{23–28} underlined the special position of the Malagasy (Supplementary Figure S3) who cluster with none of the other populations.

Paternal lineages

Regarding paternal heritage, the prevalence of African lineages (predominantly haplogroups E1b1a, B2a1a and B2b) could clearly be shown in both HL and CT samples (68 and 64%, respectively) (see Supplementary Table S2 for detailed results). However, Asian and Eurasian influences were detectable as well (Figure 2). The most

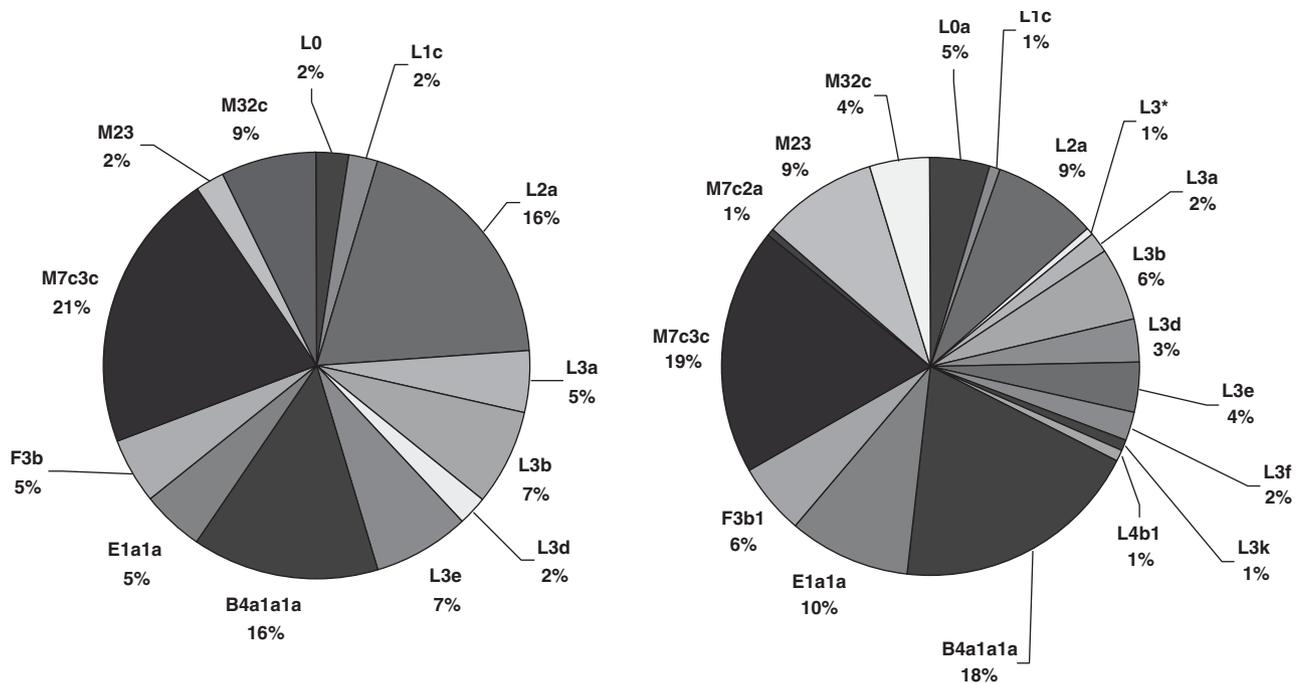


Figure 1 mtDNA haplogroup frequencies in HL (left) and CT (right) samples.

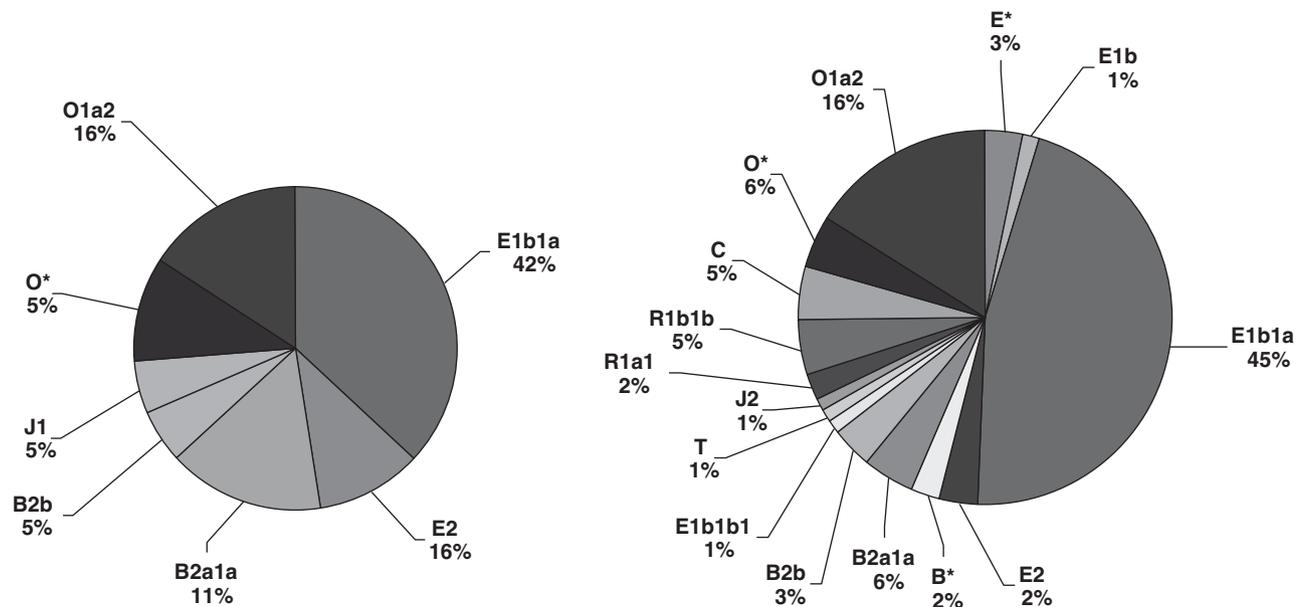


Figure 2 Y-chromosomal haplogroup frequencies in male HL (left) and CT (right) samples.

frequent Y haplogroup was the African haplogroup E1b1a (42% of HL and 45% of CT samples) while the most dominant Asian haplogroup was O1a2 (16% of HL as well as CT samples). No greater differences of the distribution of Y-chromosomal haplogroups could be found between HL and CT ethnic subgroups. Our results differed slightly from those published by Tofaneli *et al*⁵ who reported for nearly 75% of CT, but only 50% of HL samples an African parental origin. This may be due to the rather small number of samples from the Merina HL population in their study ($n = 9$), whereas this ethnic group presents the majority of our HL samples ($n = 26$).

The geographical alignment of each sample regarding Y-SNP analysis and the determination of Y-STR haplotypes – as far as possible – was congruent (Supplementary Table S3 and Supplementary Figure S4). The phylogenetic network (Supplementary Figure S4) showed two distinguishable nodes, one (above) almost exclusively African, while the other (below) has a more admixed configuration. Twenty-five single Y-STR haplotypes and one haplotype occurring twice (based on the eight STRs of the minimal haplotype) could not be found at all in the YHRD database. Y-STR haplotypes occurring more than once were more often than

not shared by people from HL and CT ethnic subgroups. The Y-STR haplotypes of those individuals, in which Eurasian haplogroups E1b1b1, R1a and R1b1b could be established, occurred predominantly in Europe according to the YHRD database. A Eurasian influence could be found in only one HL sample (from the Bara tribe; haplogroup J1; 5%), but in 10% of CT samples (haplogroups E1b1b1, J2, R1a1, R1b1b and T), predominantly in Antandroy. These findings vary significantly from those reported by Tofanelli *et al*,⁵ especially as they detected no Eurasian parental lineage at all in the Antandroy. A corresponding analysis of the Y-chromosomal haplogroup frequencies in Madagascar with African and Asian populations also revealed a higher similarity to Africa (Supplementary Figure S5). The Malagasy did not directly cluster with the African populations, but the genetic distance to them is much smaller than to the widely distributed Asian populations and also smaller than in the comparison analysis for mtDNA haplogroups.

Population differentiation SNPs

The population affiliation to one of the seven populations, which were included in the original project (West Africa, Northern Africa, Near East, Turkey, Balkan states, Northern Europe and East Asia (Japan)) was calculated as described before¹⁵ (see Supplementary Table S4 for calculation results). This assay has been predominantly designed to distinguish African and European/Eurasian populations by phenotype, it is not as powerful in the differentiation of Asian populations. The majority of samples could be assigned to West Africa

(64 and 68% in HL and CT samples, respectively). The allele distribution of about one fifth of HL as well as CT samples lead to a classification to East Asia, whereas 14% of HL and 10% of CT samples grouped with Northern African samples. In our recent publication, we could show that any individual assigned to West Africa was always of African origin.¹⁵ No other sample – be it from Northern Africa, Near East, East Asia, Sri Lanka or India – was identified as West African. No samples from Indonesia were included in our recent study, but samples from the Philippines, Sri Lanka, Thailand or Vietnam were mostly assigned to East Asia (or not identified at all). So, we assumed that the majority of Malagasy people with a predominantly Asian heritage would be identified as Asian or could not be related to any group at all. However, only three CT samples and one HL sample could not be attached to any of the seven chosen populations; they demonstrated for all populations a probability below 50% as outlined in our recent publication.¹⁵ The affiliation of a majority of Malagasy to an African population is not surprising, as the actual phenotype of these people and, therefore, the phenotype of the people included in this study is predominantly African.

Comparison of maternal and paternal lineages

The striking difference in the distribution of the geographical origin of maternal and paternal lineages called for a direct comparison of all male samples (Figure 3, Table 1). About half of the HL samples with African Y-chromosomal lineage had also an African maternal heritage, the other half displayed Asian-Indonesian mtDNA haplogroups.

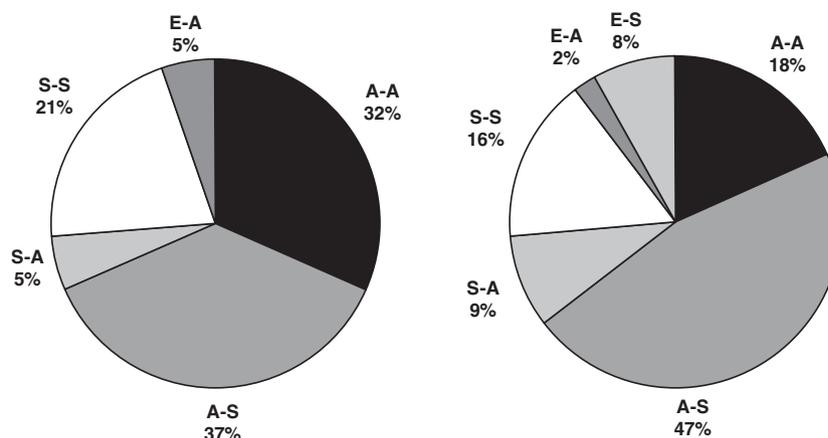


Figure 3 Comparison of maternal and paternal heritage in male HL (left) and CT (right) samples. The order is Y-chromosomal heritage – mtDNA origin. A, Africa; S, Asia; E, Eurasia. The full colour version of this figure is available at *European Journal of Human Genetics* online.

Table 1 Comparison of maternal and paternal heritage in male samples with consideration of the different ethnic groups

	n	African-African	African-Asian	European-African	European-Asian	Asian-African	Asian-Asian
Bara	3	1		1		1	
Betsileo	2		1				1
Merina	14	5 (36%)	6 (43%)				3 (21%)
HL total	19	6 (32%)	7 (37%)	1 (5%)		1 (5%)	4 (21%)
Antandroy	71	13 (18%)	32 (45%)	1 (1%)	7 (10%)	6 (8%)	12 (17%)
Antanosy	1		1				
Betsimisaraka	1		1				
Mahafaly	2	1					1
Sakalava	12	2 (17%)	6 (50%)	1 (8%)		2 (17%)	1 (8%)
CT total	87	16 (18%)	40 (46%)	2 (2%)	7 (8%)	8 (9%)	14 (16%)

Percentages (in brackets) are given only for ethnic groups with more than 10 samples. The order is: Y chromosomal heritage – mtDNA origin.

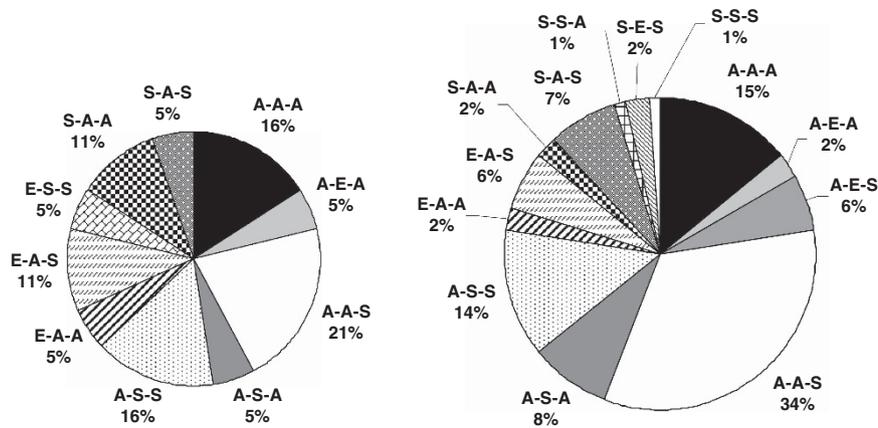


Figure 4 Comparison of maternal and paternal heritage with population affiliation by autosomal SNP analysis in male HL (left) and CT (right) samples. The order is autosomal SNPs – Y-chromosomal heritage – mtDNA origin. A, Africa; S, Asia; E, Eurasia. The full colour version of this figure is available at *European Journal of Human Genetics* online.

Altogether, 53% of HL samples, but only 34% of CT samples had an unambiguous either African or Asian origin. More than 70% of African paternal CT samples showed an Asian-Indonesian maternal heritage. Regarding Asian-Indonesian Y-chromosomal lineage, 25% of HL samples and 36% of CT samples had an African maternal heritage. The CT samples with Eurasian Y-chromosomal heritage mostly displayed an Asian-Indonesian mtDNA origin (78%), whereas the one HL sample with this heritage showed an African mtDNA type.

Admixture studies

In the comparison between the three different DNA approaches analysed in this study a truly heterogeneous picture could be found. Only ~15% of either HL or CT male samples displayed an unambiguous African heritage in all three categories (Figure 4). The greatest number of individuals belonged to an African-African-Asian type (autosomal SNPs – Y-chromosomal lineage – mtDNA lineage) and even those samples that showed a clear Asian-Indonesian heritage in paternal as well as maternal lineages demonstrated – with two exceptions – an African phenotype. The individuals with a European or Northern African influence (in autosomal SNPs or Y-chromosomal haplogroups) also displayed a heterogeneous picture. Not one single sample had a European or Northern African heritage in both categories. In our study, no significant differences were observed between HL and CT samples. The median-joining network of Y-STRs and autosomal SNPs also reflects this heterogeneity and the admixture situation (Figure 5). Although the distribution of mtDNA and Y-chromosomal haplogroups gives an insight in the heritage of the Malagasy, the population affiliation reflects the actual phenotype of people. This may – in addition to the maternal heritage – explain the change of Asian heritage (lower node in Supplementary Figure S4; left node in Figure 5) to African population affiliation.

In comparison of all individuals regarding population affiliation and mtDNA origin a similar picture emerged. About half of the HL samples assigned to a West African population displayed an African mtDNA, while the other half showed an Asian-Indonesian mtDNA. The same distribution is seen looking at HL samples assigned to an East Asian population. Altogether 41 and 39% of HL and CT samples showed an unambiguous either African or Asian-Indonesian heritage, respectively. In contrast, 63% of CT samples determined as West African regarding autosomal SNP results demonstrated an Asian-Indonesian mtDNA, but only 27% of CT samples allocated to an East Asian population had an

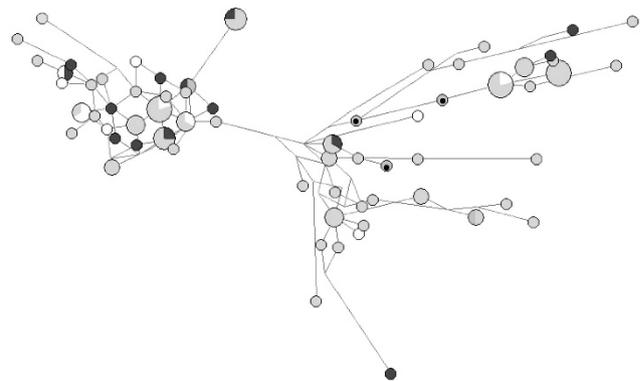


Figure 5 Median-joining network of paternal lineage. The structure is based on the Y-STR haplotypes, the colouring is based on the population affiliation according to autosomal SNPs results (grey: West Africa, white: Northern Africa, black: Asian, grey with a dot: no affiliation possible). Node sizes are proportional to frequencies of individuals. The full colour version of this figure is available at *European Journal of Human Genetics* online.

African mtDNA origin. Individuals assigned to Northern Africa were predominantly Asian-Indonesian in mtDNA heritage.

CONCLUSIONS

Despite the rather small number of samples analysed, our results confirmed the previously reported African and Asian-Indonesian admixture in Malagasy^{4,5} as demonstrated by mtDNA and Y chromosome analysis. However, the calculation of population affiliation after autosomal SNP analysis emphasises that the African heritage is more dominant, which is not surprising as the appearance of the Malagasy is rather African in nature. Moreover, a predominantly Asian origin for people from the HL as proposed by Blechn⁸ could not be confirmed, neither by Y-chromosomal and mtDNA analysis nor by the analysis of the population differentiation SNPs.

We could demonstrate that the analysis of specific autosomal SNPs is a valuable complement to the investigation of Y-chromosomal and mtDNA polymorphisms in population studies and should therefore be included if possible. Nevertheless, the use of all the three approaches might lead to conflicting results in such cases where the actual phenotype of the people investigated differ from their population heritage.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENT

We would like to thank Michael Nothnagel for helpful discussions regarding statistical analysis.

- Dewar R, Wright H: The culture history of Madagascar. *J World Prehist* 1993; **7**: 417–466.
- Burney DA: Pre-settlement vegetation changes at Lake Tritirivakely, Madagascar. *Palaeoecol Africa* 1987; **18**: 357–381.
- Burney DA, Burney LP, Godfrey LR *et al*: A chronology for late prehistoric Madagascar. *J Hum Evol* 2004; **47**: 25–63.
- Hurles ME, Sykes BC, Jobling MA, Forster P: The dual origin of the Malagasy in Island Southeast Asia and East Africa: evidence from maternal and paternal lineages. *Am J Hum Genet* 2005; **76**: 894–901.
- Tofaneli S, Bertoni S, Castri L *et al*: On the origins and admixture of Malagasy: new evidence from high-resolution analyses of paternal and maternal lineages. *Mol Biol Evol* 2009; **26**: 2109–2124.
- Razafindrazaka H, Ricaut F-X, Cox MP *et al*: Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar. *Eur J Hum Genet* 2010; **18**: 575–581.
- Cox MP, Nelson MG, Tumonggor MK, Ricaut F-X, Sudoyo H: A small cohort of island Southeast Asian women founded Madagascar. *Proc R Soc B* 2012; **279**: 2761–2768.
- Blench R: New palaeo-zoogeographical evidence for the settlement of Madagascar. *Azania: J Brit Inst East Africa* 2007; **42**: 69–82.
- Stokowski RP, Pant PV, Dadd T *et al*: A genomewide association study of skin pigmentation in a South Asian population. *Am J Hum Genet* 2007; **81**: 1119–1132.
- Branicki W, Brudnik U, Wojas-Pelc A: Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. *Ann Hum Genet* 2009; **73**: 160–170.
- Kayser M, Schneider PM: DNA-based prediction of human externally visible characteristics in forensics: motivation, scientific challenges, and ethical considerations. *Forensic Sci Int Genet* 2009; **3**: 154–161.
- Spichenok O, Budimlija ZM, Mitchel AA *et al*: Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Sci Int Genetics* 2011; **5**: 472–478.
- Walsh S, Wallstein A, Liu F *et al*: DNA-based eye colour prediction across Europe with the IrisPlex system. *Forensic Sci Int Genetics* 2012; **6**: 330–340.
- Pneuman A, Budimlija ZM, Caragine T, Prinz M, Wurmbach E: Verification of eye and skin color predictors in various populations. *Leg Med* 2012; **14**: 78–83.
- Poetsch M, Blohm R, Harder M, Inoue H, von Wurmb-Schwark N, Freitag-Wolf S: Prediction of people's origin from degraded DNA – presentation of SNP assays and calculation of probability. *Int J Legal Med* 2012; **127**: 347–357.
- Bouakaze C, Keyser C, Crubezy E, Montagnon D, Ludes B: Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. *Int J Legal Med* 2009; **123**: 315–325.
- Rasmussen M, Li Y, Lindgreen S *et al*: Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 2010; **463**: 757–762.
- Kloss-Brandstatter A, Pacher D, Schoenherr S *et al*: HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat* 2011; **32**: 25–32.
- van Oven M, Kayser M: Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 2009; **30**: E386–E394.
- Bandelt H-J, Forster P, Röhl A: Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 1999; **16**: 37–48.
- Nenadic O, Greenacre M: Correspondence analysis in R, with two- and three-dimensional graphics: the ca package. *J Stat Soft* 2007; **20**: 1–13.
- Ricaut F-X, Razafindrazaka H, Cox MP *et al*: A new deep branch of Eurasian mtDNA macrohaplogroup M reveals additional complexity regarding the settlement of Madagascar. *BMC Genomics* 2009; **10**: 605–613.
- Pereira L, Cerny V, Cerezo M *et al*: Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of Tuareg nomads from the African Sahel. *Eur J Hum Genet* 2010; **18**: 915–923.
- Mikkelsen M, Fendt L, Röck AW *et al*: Forensic and phylogeographic characterisation of mtDNA lineages from Somalia. *Int J Legal Med* 2012; **126**: 573–579.
- Plaza S, Salas A, Calafell F *et al*: Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Hum Genet* 2004; **115**: 439–447.
- Hill C, Soares P, Mormina M *et al*: A mitochondrial stratigraphy for island southeast Asia. *Am J Hum Genet* 2007; **80**: 29–43.
- Poloni ES, Naciri Y, Bucho R *et al*: Genetic evidence for complexity in ethnic differentiation and history in East Africa. *Ann Hum Genet* 2009; **73**: 582–600.
- Coudray C, Olivier A, Achilli A *et al*: The complex and diversified mitochondrial gene pool of Berber populations. *Ann Hum Genet* 2009; **73**: 196–214.
- Karafet TM, Hallmark B, Cox MP *et al*: Major east-west division underlies Y chromosome stratification across Indonesia. *Mol Biol Evol* 2010; **27**: 1833–1844.
- Cai X, Qin Z, Wen B *et al*: Human migration through bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum revealed by Y chromosomes. *PLoS One* 2011; **6**: e24282.
- Naidoo T, Schlebusch CM, Makkan H *et al*: Development of a single base extension method to resolve Y chromosome haplogroups in sub-Saharan African populations. *Investig Genet* 2010; **1**: 6.
- de Filippo C, Barbieri C, Whitten M *et al*: Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. *Mol Biol Evol* 2011; **28**: 1255–1269.
- Robino C, Crobu F, De Gaetano C *et al*: Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. *Int J Legal Med* 2008; **122**: 251–255.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)