

# Strong Maternal Khoisan Contribution to the South African Coloured Population: A Case of Gender-Biased Admixture

Lluis Quintana-Murci,<sup>1,2,\*</sup> Christine Harmant,<sup>1,2</sup> H el ene Quach,<sup>1,2</sup> Oleg Balanovsky,<sup>3</sup> Valery Zaporozhchenko,<sup>3</sup> Connie Bormans,<sup>4</sup> Paul D. van Helden,<sup>5</sup> Eileen G. Hoal,<sup>5</sup> and Doron M. Behar<sup>1,6</sup>

The study of recently admixed populations provides unique tools for understanding recent population dynamics, socio-cultural factors associated with the founding of emerging populations, and the genetic basis of disease by means of admixture mapping. Historical records and recent autosomal data indicate that the South African Coloured population forms a unique highly admixed population, resulting from the encounter of different peoples from Africa, Europe, and Asia. However, little is known about the mode by which this admixed population was recently founded. Here we show, through detailed phylogeographic analyses of mitochondrial DNA and Y-chromosome variation in a large sample of South African Coloured individuals, that this population derives from at least five different parental populations (Khoisan, Bantus, Europeans, Indians, and Southeast Asians), who have differently contributed to the foundation of the South African Coloured. In addition, our analyses reveal extraordinarily unbalanced gender-specific contributions of the various population genetic components, the most striking being the massive maternal contribution of Khoisan peoples (more than 60%) and the almost negligible maternal contribution of Europeans with respect to their paternal counterparts. The overall picture of gender-biased admixture depicted in this study indicates that the modern South African Coloured population results mainly from the early encounter of European and African males with autochthonous Khoisan females of the Cape of Good Hope around 350 years ago.

Recently admixed populations provide unique opportunities for studies related to population history, natural selection, and admixture mapping.<sup>1</sup> The mixed ancestries of chromosomal segments in the genomes of admixed populations, such as African Americans, Hispanic/Latinos or Mexican Americans, can be used for the identification of genetic risk factors for complex traits and diseases showing prevalence differences between populations.<sup>2–5</sup> In the African continent, one of the populations displaying the highest levels of recent admixture is the now officially self-designated “Coloured” population, which mostly resides in the Western Cape region of South Africa. The origins of the South African Coloured (SAC) can be traced historically to the shores of Table Bay in the area that is now Cape Town at a date soon after the Dutch East India Company had established a refreshment station there in 1652.<sup>6</sup> A few well-documented social and demographic events over the generations have fused indigenous Khoi and San (Khoisan), tribal Bantu-speaking populations, European settlers, and slaves’ descendants from Java, India, Mozambique, and Madagascar into a highly admixed but discrete ethnic group.<sup>6–8</sup> Currently, the estimated four million SAC make up roughly 9% of the South African population and form a unique population in which the African, European, and south Asian cultures have contributed to the melting pot.<sup>6–8</sup>

From a genetics point of view, only two recent studies have analyzed the genetic composition of this African population.<sup>9,10</sup> The first, which analyzed 1,327 nuclear microsatellite and insertion/deletion markers in a large panel of African populations, found that the SAC cluster in intermediate positions between African and non-African populations.<sup>9</sup> On the basis of STRUCTURE analysis, this population showed nearly equal levels (~25%) of four ancestries, namely southern African Khoisan, Bantu-speaking, Indian, and European.<sup>9</sup> Using high-density genome-wide genotyping in 20 Coloured individuals and comparing their variation with that of other populations worldwide, the second study concluded that the SAC resulted from a complex admixture process involving at least Bantu-speaking populations from South Africa, Europeans, south Asians, and Indonesians.<sup>10</sup>

Although autosomal markers provide genome-wide estimates of the different admixture proportions, they are of limited use for the study of gender-biased demographic events. In this context, exploiting segments of the genome that are inherited through only one sex—the maternally inherited mitochondrial DNA (mtDNA) and the nonrecombining portion of the Y chromosome (NRY)—has proven to be a powerful tool for unraveling gender-specific processes and socio-cultural effects, such as polygyny, matrilocality versus patrilocality, different forms of social

<sup>1</sup>Institut Pasteur, Human Evolutionary Genetics, Department of Genomes and Genetics, F-75015 Paris, France; <sup>2</sup>Centre National de la Recherche Scientifique, URA3012, F-75015 Paris, France; <sup>3</sup>Research Centre for Medical Genetics, Russian Academy of Medical Sciences, 115478 Moscow, Russia; <sup>4</sup>Genomics Research Center, Family Tree DNA, Houston, Texas 77008, USA; <sup>5</sup>Molecular Biology and Human Genetics, Medical Research Council Centre for Molecular and Cellular Biology, Faculty of Health Sciences, Stellenbosch University, Tygerberg, South Africa; <sup>6</sup>Molecular Medicine Laboratory, Rambam Health Care Campus, Haifa 31096, Israel

\*Correspondence: [quintana@pasteur.fr](mailto:quintana@pasteur.fr)

DOI 10.1016/j.ajhg.2010.02.014. ©2010 by The American Society of Human Genetics. All rights reserved.

**Table 1. MtDNA Haplogroup Frequencies and Diversity Indices Observed in the South African Coloured Population**

<b>mtDNA Haplogroup<sup>a</sup></b>	<b>n<sup>b</sup></b>	<b>Percent</b>	<b>k<sup>c</sup></b>	<b>Hg D (SD)<sup>d</sup></b>		<b>Ht D (SD)<sup>e</sup></b>		<b>Pi (SD)<sup>f</sup></b>	
<b>Sub-Saharan African</b>									
Total	445	79.04	96	0.85	(0.01)	0.93	(0.01)	6.81	(3.21)
<i>L0d2a</i>	125	22.20	12			0.28	(0.05)	0.32	(0.32)
<i>L0d1b</i>	95	16.87	18			0.87	(0.02)	2.22	(1.24)
<i>L0d1a1<sup>g</sup></i>	42	7.46	5			0.26	(0.09)	0.28	(0.31)
<i>L2a1</i>	26	4.62	9			0.83	(0.04)	3.34	(1.77)
<i>L0d1a</i>	19	3.37	6			0.70	(0.10)	1.49	(0.94)
<i>L0d2c</i>	19	3.37	4			0.45	(0.13)	0.49	(0.44)
<i>L0d3</i>	19	3.37	6			0.79	(0.06)	1.30	(0.85)
<i>L0a2</i>	18	3.20	3			0.21	(0.12)	0.22	(0.27)
<i>L3d1a1</i>	11	1.95	2			0.54	(0.07)	0.54	(0.48)
<i>L0d1c</i>	10	1.78	3			0.38	(0.18)	0.96	(0.71)
<i>L3e1b</i>	8	1.42	2			0.54	(0.12)	1.07	(0.79)
<i>L0a1b</i>	7	1.24	3			0.52	(0.21)	0.76	(0.62)
<i>L0d2b</i>	6	1.07	1			...	...	...	...
<i>L3e1a</i>	6	1.07	3			0.73	(0.16)	1.13	(0.85)
<i>L1c2</i>	5	0.89	5			1.00	(0.13)	4.40	(2.61)
<i>L3d3</i>	5	0.89	1			...	...	...	...
<i>L4b2</i>	5	0.89	1			...	...	...	...
<i>L1b</i>	3	0.53	1			...	...	...	...
<i>L3e3</i>	3	0.53	1			...	...	...	...
<i>L3f</i>	3	0.53	1			...	...	...	...
<i>L0d2</i>	2	0.36	2			1.00	(0.50)	3.00	(2.45)
<i>L3e2</i>	2	0.36	1			...	...	...	...
<i>L5c</i>	2	0.36	2			1.00	(0.50)	3.00	(2.45)
<i>L0d1c1</i>	1	0.18	1			...	...	...	...
<i>L1c3a</i>	1	0.18	1			...	...	...	...
<i>L2c</i>	1	0.18	1			...	...	...	...
<i>L3e2b</i>	1	0.18	1			...	...	...	...
<b>West Eurasian/European</b>									
Total	26	4.62	11	0.70	(0.07)	0.75	(0.09)	5.52	(2.74)
U5a1	13	2.31	1			...	...	...	...
U2e	6	1.07	4			0.87	(0.13)	1.47	(1.03)
H	3	0.53	2			0.67	(0.31)	0.66	(0.66)
T2b	2	0.36	2			1.00	(0.50)	3.00	(2.45)
HV	1	0.18	1			...	...	...	...
K	1	0.18	1			...	...	...	...
<b>South/southeast Asian</b>									
Total	92	16.34	48	0.69	(0.05)	0.97	(0.01)	5.82	(2.81)
M	49	8.70	28			0.96	(0.01)	3.88	(1.98)
U2c	13	2.31	2			0.46	(0.11)	0.92	(0.68)

**Table 1. Continued**

mtDNA Haplogroup <sup>a</sup>	n <sup>b</sup>	Percent	k <sup>c</sup>	Hg D (SD) <sup>d</sup>	Ht D (SD) <sup>e</sup>	Pi (SD) <sup>f</sup>
U2b	6	1.07	3		0.73 (0.15)	4.13 (2.39)
B4	5	0.89	3		0.70 (0.22)	3.40 (2.08)
R6	5	0.89	1		...	...
U7	4	0.71	1		...	...
F1a	3	0.53	3		1.00 (0.27)	1.33 (1.09)
R	4	0.71	4		1.00 (0.18)	5.67 (3.44)
R5	2	0.36	2		1.00 (0.50)	5.00 (3.87)
B5a	1	0.18	1		...	...

Note: Table S1 provides the full haplogroup and haplotype (HVS-I) dataset.

<sup>a</sup> Khoisan-specific Hgs are in italics; pan-African Hgs are underlined.

<sup>b</sup> Number of individuals.

<sup>c</sup> Number of observed HVS-I haplotypes.

<sup>d</sup> Gene diversity based on Hg profiles (Hg D) and standard deviation (SD).

<sup>e</sup> Gene diversity based on HVS-I haplotypes (Ht D) and standard deviation (SD).

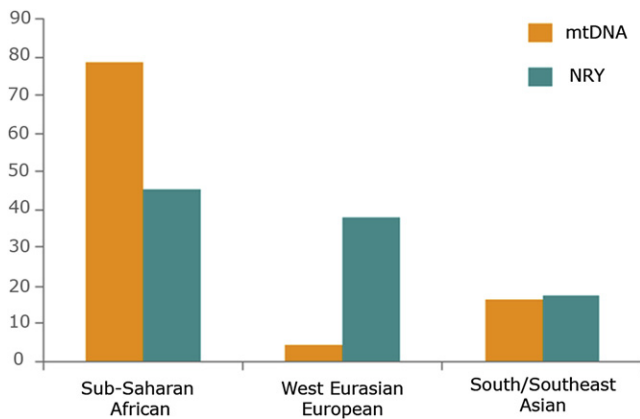
<sup>f</sup> Mean number of pairwise differences (Pi) and standard deviation (SD).

<sup>g</sup> We labeled this new clade, L0d1a1 (see Figure S1), according to the rules previously described.<sup>66</sup>

organization, and gender-biased admixture,<sup>18–27</sup> that might have influenced human evolution.<sup>11–17</sup> An additional advantage of these genetic systems is that they are more sensitive to effects of genetic drift because of their up to four-fold reduction in effective population size relative to the autosomes and their lack of recombination, so they present increased resolution in discriminating between closely related populations.

In the context of SAC, where various continental gene contributions are expected, the use of the uniparentally inherited markers should be optimal given the deep phylogenetic separation of the mtDNA and NRY lineages between the different continents<sup>12,17</sup> from which the SAC population originated.<sup>6</sup> Here, we took advantage of the well-resolved phylogenies and strong geographic clustering of mtDNA and NRY lineages<sup>11,12,14,17</sup> to (1) dissect the maternal and paternal phylogeographic components of the SAC, (2) characterize in detail their contemporary gender-specific proportions of suspected African, European, and Asian parental ancestries, and (3) unmask possible gender-biased admixture during the foundation of this population. To this end, we analyzed mtDNA and NRY diversity in a large sample of 590 unrelated individuals self-identified as South African Coloured. Samples were collected in the Western Cape in areas that are predominantly Christian and therefore do not include people of the group known as “Cape Malays.” Of the 590 donors, 315 were females, and 275 were males. We were technically able to genotype a total of 563 mtDNA genomes and 228 Y chromosomes (Table S1, available online). For all subjects, appropriate informed consent was obtained, and institutional review boards at the Faculty of Health Sciences, Stellenbosch University, South Africa and Institut Pasteur (n° RBM 2008.06) approved all procedures.

We first defined the maternal ancestry of the SAC population by analyzing a total of 64 mtDNA binary markers, selected on the basis of complete mtDNA genome phylogenies,<sup>17,28–30</sup> and sequencing the Hyper Variable Segment-I (HVS-I) (Figure S1; Table S1). The 64 binary markers were tested in a hierarchical order (Figure S1) by fluorescence polarization (VICTOR-2TM Technology; Perkin Elmer) or direct sequencing. Variable positions at the HVS-I were determined from positions 16,024–16,383. Sequencing was performed on a 3730xl DNA Analyzer (Applied Biosystems). We used the resulting dataset to place each of the mtDNA genomes into its respective haplogroup (Hg) and to infer the continental and within-continental geographic origin of each of them. Our analyses revealed 43 different Hgs (Table 1), all known to coalesce prior to the foundation of the SAC population.<sup>17</sup> The vast majority of the maternal gene pool of the SAC is of clear sub-Saharan African origin: 79% of their mtDNA lineages belong to the various sub-Saharan African L(xM,N) Hgs<sup>29,31</sup> (Table 1; Figure 1). It is remarkable that fully 60% of SAC mtDNAs belong to Hg L0d. This Hg, together with L0k, has been shown to be specific to the Khoisan peoples of South Africa (e.g., Ju|'hoansi San, !Xun/Khwe, !Xun, and !Kung), and not to all click speakers; it is virtually absent from other populations across Africa.<sup>29,32–34</sup> Recent complete mtDNA genome sequence data have shown that both L0d and L0k probably evolved within the ancestors of the Khoisan people and are indeed autochthonous to South Africa.<sup>29</sup> To dissect the structure of Hg L0d, we typed 11 markers that are diagnostic of the internal L0d subclades (Figure S1). Our analyses revealed that all SAC L0d lineages can be classified into the three major L0d1, L0d2, and L0d3 subclades; L0d2a and L0d1b are the most frequently observed (22% and 17%, respectively, Table 1). Overall, the exceptionally high frequency of the different L0d



**Figure 1. Diverse Continental Origins of Maternal and Paternal Contributions to the South African Coloured**

Bar sizes reflect the combined frequencies of all haplogroups, indicated separately for mtDNA and NRY, associated with the sub-Saharan African, west Eurasian/European, and south/southeast Asian genetic components (Table S1, Tables 1 and 2).

subclades among SAC indicates a substantial maternal contribution of the Khoisan people to the foundation of the SAC population.

The additional 19% of African maternal lineages observed in SAC comprise the various L0a, L1, L2, L3, L4, and L5 clades (Table 1), which are thought to have been introduced to South Africa during the recent Bantu expansion.<sup>31,35</sup> Because today these lineages are also observed in the Khoisan (~40%),<sup>29,33,34</sup> they should be seen as lineages of “pan-African” origin. Their presence among SAC reflects either direct gene flow from Bantu peoples or indirect Bantu contribution via admixture with the Khoisan, who have received these lineages through their prior admixture with Bantus.<sup>29</sup> In the latter case, the actual maternal contribution of the Khoisan to the SAC population would be up to 79%.

The second most important maternal contribution (16.3%) observed among the SAC is of clear south/southeast Asian origin, as attested by the presence of Hgs B4 and B5a, which are mostly observed among East and Southeast Asians,<sup>36</sup> and Hgs M, R, R5, R6, U2b, U2c, and U7, which are instead characteristic of the Indo-Gangetic region<sup>37,38</sup> (Figure 1; Table 1). The most prevalent lineage (8.7%) accounting for this component is the macro-Hg M, which largely characterizes the Indian subcontinent.<sup>39,40</sup> Despite the extraordinary genetic and geographic variation of macro-Hg M, some inferences can be made from its internal HVS-I haplotype variation. The HVS-I positions 16274–16319, 16126, 16129, 16318T, and 16327–16330 have been suggested to be diagnostic for Hgs M2, M3, M5, M18 and M41, respectively.<sup>39,40</sup> These Hgs have been all observed throughout India at varying frequencies depending on the geographic regions.<sup>37</sup> For example, the M5 samples found among the SAC, which contain the transversion 16265C, have been previously observed in West Bengal, Gujarat, Maharashtra, and Uttar Pradesh among

Brahmin and tribal populations.<sup>37</sup> With respect to non-M Hgs, R5 and R6 clearly characterize Indian and Pakistani populations.<sup>38</sup> In our dataset (Table S1), we observed five samples belonging to Hg R6 and all sharing the HVS-I haplotype (16129-16266-16318-16320-16362), which has been previously observed in Andhra Pradesh tribal and West Bengal Kurmi populations.<sup>37</sup> Likewise, the exact Hg U7 haplotype 16069-16274-16318T, which we found in four SAC samples, has been previously described in Gujarat, Maharashtra, and western India, and one-step derivatives have been described in Andhra Pradesh.<sup>37,41</sup> Overall, the Asian maternal component present in the SAC can largely be traced back to the Indo-Gangetic region, although there is anecdotal presence of lineages originating in Southeast Asia. Because historical records suggest Madagascar as a one of the sources of South African slaves,<sup>7</sup> the presence of Southeast Asian mtDNA lineages (i.e., B4 and B5a) among the SAC may attest, at least partially, admixture with Malagasy via the prior admixture of Southeast Asians and sub-Saharan Africans in Madagascar.<sup>42</sup>

Finally, the least dominant maternal component is the West Eurasian one. Only 4.6% of the SAC variation can be attributed to typical European Hgs, including U5a1, U2e, H, T2b, HV, and K (Figure 1; Table 1). Although these Hgs have a wide West Eurasian distribution,<sup>43,44</sup> their presence among the SAC indicates that their origin is most probably European, and hence they are referred to herein as such. Indeed, half of the observed SAC mtDNA genomes of European ancestry belong to the same Hg U5a1 haplotype (16145-16189-16192-16256-16270) (Table 1 and Table S1), which has been previously observed at low population frequencies.<sup>43,45</sup> This observation might attest to a single female founder’s having harbored this Hg, and the lack of internal variation of Hg U5a1 might be explained by the short time period that has elapsed since the founding event when one considers the mutation rate of the mtDNA HVS-I.<sup>46</sup> Thus, the actual contribution of European females to the founding of the maternal gene pool of the SAC might be even smaller than the percentage observed among contemporary SAC.

To define the paternal ancestry of the SAC, we analyzed the Y chromosomes by means of 46 binary markers and 14 short tandem repeats (Y-STRs) (Table S1). The binary markers were hierarchically typed (Figure S2) by TaqMan assays or direct sequencing, as previously reported.<sup>47</sup> For the microsatellite analysis, the Y-STRs genotyped were DYS19, DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, DYS437, DYS438, and DYS439. PCR products were electrophoresed on a 3730xl Genetic Analyzer (Applied Biosystems), and fragment lengths were converted to repeat number by the use of allelic ladders. Genotyping of binary markers yielded a total of 21 different Hgs (Table 2), and as for mtDNA lineages, all these Hgs are known to coalesce prior to the foundation of the SAC population.<sup>47</sup> In contrast with mtDNA results, the sub-Saharan African paternal component of the SAC was almost half (45.2%) that

**Table 2. Y Chromosome Haplogroup Frequencies and Diversity Indices Observed in the South African Coloured Population**

Y-Chromosome Hg <sup>a</sup>	n <sup>b</sup>	Percent	k <sup>c</sup>	Hg D (SD) <sup>d</sup>	Ht D (SD) <sup>e</sup>	Pi (SD) <sup>f</sup>
<b>Sub-Saharan African</b>						
Total	103	45.18	78	0.68 (0.04)	0.99 (0.00)	7.43 (3.50)
<u>E-M2</u>	55	24.12	43		0.99 (0.01)	5.05 (2.48)
E-P68*	13	5.70	9		0.92 (0.06)	2.83 (1.59)
A-M51	12	5.26	8		0.94 (0.05)	6.97 (3.52)
B-M150*	11	4.82	8		0.89 (0.09)	2.18 (1.30)
<u>E-M35 (xM78, M81, M123, M281)</u>	8	3.51	6		0.93 (0.08)	3.78 (2.13)
<u>B-M112</u>	3	1.32	3		1.00 (0.27)	9.00 (5.72)
E-M123*	1	0.44	1		...	...
<b>West Eurasian/European</b>						
Total	86	37.72	67	0.70 (0.04)	0.99 (0.00)	8.44 (3.94)
R-M343	41	17.98	30		0.98 (0.01)	4.97 (2.46)
I-M170	20	8.77	17		0.99 (0.02)	7.39 (3.61)
R-M17 <sup>g</sup>	12	5.26	10		0.97 (0.04)	3.63 (1.98)
E-M78	5	2.19	5		1.00 (0.13)	4.10 (2.45)
T-M70	4	1.75	1		...	...
G-M201	2	0.88	2		1.00 (0.50)	7.00 (5.29)
J-M304	2	0.88	2		1.00 (0.50)	6.00 (4.58)
<b>South/southeast Asian</b>						
Total	39	17.11	31	0.81 (0.04)	0.98 (0.01)	8.80 (4.15)
F-M89 (xM201, M69, M170, M304, M9)	14	6.14	13		0.99 (0.03)	8.31 (4.09)
O-M122	8	3.51	5		0.79 (0.15)	2.61 (1.55)
H-M69	5	2.19	3		0.83 (0.22)	6.33 (3.80)
O-M119	4	1.75	2		0.50 (0.26)	1.00 (0.82)
R-M124	3	1.32	3		1.00 (0.27)	6.33 (4.12)
K-M9 (XM20, M186, LLY22 g, M175, M45, M70)	3	1.32	3		1.00 (0.27)	6.67 (4.32)
O-M95	2	0.88	2		1.00 (0.50)	8.00 (6.00)

Note: Table S1 provides the full haplogroup and haplotype (STR) dataset.

<sup>a</sup> Khoisan-specific Hg is in italics; pan-African Hgs are underlined; Bantu-specific Hgs are marked with an asterisk. We followed the nomenclature conventions recommended by the Y Chromosome Consortium.<sup>47</sup>

<sup>b</sup> Number of individuals.

<sup>c</sup> Number of observed STR haplotypes.

<sup>d</sup> Gene diversity based on Hg profiles (Hg D) and standard deviation (SD).

<sup>e</sup> Gene diversity based on STR haplotypes (Ht D) and standard deviation (SD).

<sup>f</sup> Mean number of pairwise differences (Pi) and standard deviation (SD).

<sup>g</sup> Hg R-M17 can also be of Indian origin.

observed on the maternal side (79%, Figure 1). This component was clearly dominated by the pan-African Hg E-M2, which constituted 24.1% of the entire variation. Hg E-M2 is restricted to sub-Saharan African populations; it reaches the highest frequencies (up to 90%) across Bantu-speaking groups<sup>48</sup> and varies in its distribution among Khoisan groups (~40%).<sup>33,34,49</sup> Because the multiple Hg E-M2-derived branches do not exhibit a clear population clustering<sup>50</sup> and microsatellite networks for paragroup E-M2 are noninformative as a result of the multiple reticulations observed (<sup>50</sup> and data not shown), the presence of

Hg E-M2 among the SAC, although it could attest to Khoisan contribution, can only be attributed to a component originating south of the Sahara desert. Likewise, the presence of the Hgs E-M35 (xM78, M81, M123, M281) (3.5%) and B-M112 (1.3%), which are shared between Khoisan and Bantu speakers,<sup>33,34</sup> reflect a general pan-African component. Actually, this phylogenetically unresolved pan-African component prevents the ability to test for a prior Bantu-to-Khoisan paternal gene flow, as demonstrated for the maternal side.<sup>29,34</sup> By contrast, the presence of Hgs E-P68 (5.7%), B-M150 (4.8%), and



E-M123 (0.4%), which have never been observed in Khoisan,<sup>33,34</sup> most probably indicates admixture with Bantus. Finally, and in strong contrast with mtDNA results, only a minor fraction (5.3%) of the African paternal variation can be firmly attributed to the Khoisan, as attested by the presence of Hg A-M51, which has been observed only among Khoisan-speaking populations from Namibia and South Africa.<sup>33,34</sup> Thus, although direct Hg frequencies clearly demonstrate both Khoisan and Bantu-specific signatures in contemporary SAC, most of their Y chromosomes can only be attributed to the large pan-African component (Table 2).

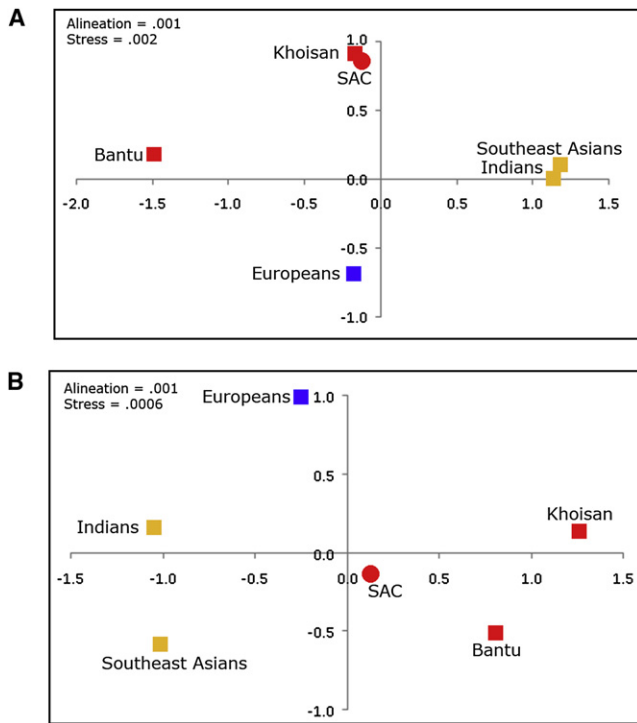
Unlike the negligible European maternal component (<5%), the paternal contribution of West Eurasian lineages<sup>51-53</sup> to the SAC turned out to be very important (>32.5%) (Figure 1, Table 2). This is attested by the presence of Hgs R-M343 (18%), I-M170 (8.8%), R-M17 (5.3%), E-M78 (2.2%), T-M70 (1.7%), G-M201 (0.9%), and J-M304 (0.9%). Specifically, the most predominant West Eurasian lineage was Hg R-M343, which is by far the most frequent among Western Europeans, particularly along the Atlantic coast, and is also sporadically observed in the Near East and north and sub-Saharan Africa.<sup>54</sup> To further localize the geographic origin of the SAC Hg R-M343 chromosomes, we compared their internal STR haplotype variation (Table S1) with that observed at 52,270 haplotypes from 363 populations worldwide (Y Chromosome Haplotype Reference Database, YHRD) by using the 12 overlapping STRs: DYS19, DYS385a, DYS385b, DYS389-I, DYS389-II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439. Our matching analyses revealed that 19 out of the 30 STR haplotypes exhibited by our 41 Hg R-M343 samples matched haplotypes from West Europe, suggesting a major Western European origin for our Hg R-M343 samples. Likewise, an important European Y chromosome contribution is indicated by the high prevalence of Hg I-M170 (~9%), which is virtually restricted to Europe.<sup>55</sup> The third most important West Eurasian contribution to the SAC is illustrated by the presence of Hg R-M17, which is observed at high frequencies in Europe, particularly eastern Europe, but also in central and south Asia, including India.<sup>53,56,57</sup> To characterize more precisely the geographic source of the 12 Hg R-M17 chromosomes, which account for ten STR haplotypes, we searched again for STR haplotype matches in the YHRD database. Most Hg R-M17 STR variation (seven haplotypes) matched with Western or Eastern Europeans, but a minority (three haplotypes) matched also in Asia, India, or north Africa. Finally, low frequencies of other typical West Eurasian Hgs (e.g., T-M70, G-M201) were observed. Altogether, our Y chromosome data revealed an important paternal West Eurasian component among the SAC, along with a predominant European contribution.

The least dominant paternal component observed in the SAC was that originating from south Asia<sup>56,58-61</sup> (17.5%, Figure 1 and Table 2). F-M89 (xM201, M69, M170, M304, M9) (6.1%), H-M69 (2.2), and R-M124 (1.3%) are primarily associated with Indian populations,<sup>58,62</sup> and

O-M122 (3.5%), O-M119 (1.7%), K-M9 (x M20, M186, LLY22 g, M175, M45, M70) (1.3%) and O-M95 (0.9%) are mainly found in southeast Asian populations, including China.<sup>59,61</sup> The STR haplotype profiles associated with each of these Hgs were clearly consistent with Indian and southeast Asian geographic origins, respectively. Similar to the maternal component, the paternal southeast Asian lineages could have been introduced into the SAC population from Madagascar, via their prior existence among Malagasy.<sup>42</sup>

Taken as a whole, both the mtDNA and NRY Hg profiles observed among the SAC population are concordant with a highly admixed population but also reveal exceedingly unbalanced gender-specific continental and within-continental genetic contributions. Indeed, our analyses indicate sharp and significant gender-related differences in the European and African contributions to the SAC (Figure 1) ( $\chi^2$  test,  $p < 10^{-15}$ ). The maternal European contribution turned out to be negligible (~5%), whereas its paternal counterpart is very high (at least 32.4%) and similar in size to the African component (Figure 1 and Tables 1 and 2). Conversely, the maternal African contribution (79%) was almost two-fold the paternal contribution (45%). It is worth elaborating on this African component observed among the SAC. On the maternal side, at least 76% of the African variation (i.e., 60% of the total variation, namely the presence of the different L0d subclades) directly derives from the Khoisan. On the paternal side, although our phylogeographic data attest to some Khoisan and Bantu-specific contributions, a large nonspecific pan-African component cannot be fully resolved because of limitations of the current NRY phylogeny and of comparative databases. An additional intriguing observation is the lack of the Khoisan Hg L0k among the SAC people. Vigilant et al.<sup>32</sup> reported sharp differences in the occurrence of Hg L0k in the !Xun and !Kung populations; it was 26% in the former but only 4% in the latter. Therefore, the possibility that L0d and L0k have a within-Khoisan structure is likely and might suggest that only some of the Khoisan peoples contributed to the SAC population.

To illustrate the gender-biased admixture depicted by our phylogeographic analysis and to understand the position of the SAC in the context of a global set of putative parental populations, we next used multidimensional scaling (MDS) analysis (Figure 2). Using Hg frequencies of mtDNA and NRY in the SAC as well as in five parental metapopulations (Khoisan, Bantu, Europeans, Indians, and southeast Asians; see legend of Figure 2 for details on the specific populations used), we calculated genetic distances by means of the DJ software as previously described<sup>63</sup> and visualized them in the form of an MDS plot by using Statistica 6.0 software. On the maternal side, the SAC population clustered with the Khoisan, and its position is clearly separated from the rest of the populations, including the African Bantus (Figure 2A). In contrast, the NRY-based MDS plot positions the SAC in an intermediate location between African, European,



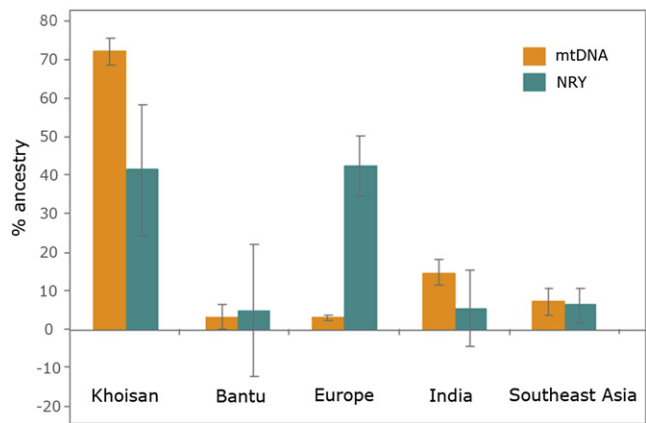
**Figure 2. MDS Plots Illustrating the Relationships between the South African Coloured Population and the Five Putative Parental Populations**

(A) MDS plot based on mtDNA Hg frequencies.

(B) MDS plot based on NRY Hg frequencies.

The five metapopulations used as putative parental populations include Europeans, Indians, Khoisan, Bantu speakers, and southeast Asians. The European dataset included British, French, Dutch, German, and Austrian populations; the Indian dataset included populations from the Bengal Bay coast (Sri Lanka, Bangladesh, and eastern states of India); the Khoisan dataset included all available Khoisan data; the Bantu dataset included Bantu-speaking populations from South Africa and central Africa; and the southeast Asian dataset included populations from the Malay Peninsula, Sumatra, Java, and neighboring smaller isles. Because different individuals and populations have been analyzed for mtDNA and NRY, sample sizes from these metapopulations differ between the two analyses. Sample sizes of European, Indian, Khoisan, Bantu, and southeast Asian populations were, respectively:  $n = 6092$ ,  $n = 1246$ ,  $n = 298$ ,  $n = 459$ , and  $n = 616$  for mtDNA and  $n = 353$ ,  $n = 766$ ,  $n = 122$ ,  $n = 499$ , and  $n = 66$  for NRY. The lower population sample sizes for the NRY data result from the nonoverlapping nature of the markers used in the different studies. These comparative data were obtained from the MURKA database, compiled from published sources by O. Balanovsky, V. Zaporozhchenko, R. Sychev, A. Pshenichnov, and E. Balanovska.

and south/southeast Asian populations (Figure 2B). To formally assess the impact of the various continental and within-continental contributions to the SAC, we carried out admixture analysis with the coalescent-based *mY* estimator, which takes into account allele frequencies and molecular information,<sup>64</sup> as implemented in Admix 2.0,<sup>65</sup> and treated the SAC as a hybrid of five parental populations. Mean admixture proportions and standard deviations for each component are represented schematically in Figure 3. This analysis found that the highest mean proportion of maternal ancestry of the SAC corresponds



**Figure 3. Diverse Continental and within-Continental Admixture Proportions in the South African Coloured Population**

Mean Khoisan, Bantu-speaking, European, Indian, and southeast Asian admixture proportions among the SAC, on the basis of the *mY* estimator, are represented as bars for the mtDNA and NRY data. Error bars indicate standard deviations. The five parental metapopulations used for this analysis, as well as their sample sizes, were the same as those used for the MDS analysis (Figure 2). Time of the initial admixture event was set to 400 years, and standard deviations were calculated on the basis of 10,000 bootstraps.

to the Khoisan, whereas the highest mean proportion of paternal ancestry corresponds to Europeans, followed by the Khoisan. It is worth noting that our MDS and admixture analyses are in contrast to each other with respect to the African paternal origin of the SAC; the former positions the SAC somehow closer to Bantus (Figure 2B), whereas the latter indicates a Khoisan origin (Figure 3). This apparent contradiction most likely results from the current phylogeographic limitations in discriminating between the Khoisan- and Bantu-specific parts within the more general pan-African paternal component, highlighting the general need for studies aiming to improve the resolution of the NRY pan-African phylogeny.

This study reveals several trends, which shed light on the mode in which this admixed population was recently founded. First, three major continental genetic contributions—African, European and Asian—are clearly visible. Second, more European males and African females than European females and African males were involved in the foundation of the population, whereas the south Asian contribution appears to be similar from both sexes. Third, there is an overwhelming contribution of Khoisan peoples to the gene pool of the SAC. Fourth, although the different continental contributions are more equally distributed within the SAC on the paternal side, the maternal counterpart is strongly skewed toward an African, mainly Khoisan, origin. These conclusions add to current genetic knowledge of the SAC by unequivocally identifying the gender-specific differential contributions to this population and by directly linking the contemporary SAC to the Khoisan. Interestingly, contemporary demographic data from the South African population census, combined with the known Hg L0d frequencies observed among the

SAC and Khoisan, suggests that the majority of individuals currently carrying the Khoisan-specific L0d clade actually belong to the SAC rather than to the Khoisan. This demonstrates the successful survival of a globally rare and deep-rooted mtDNA Hg in a recently admixed population.

More generally, our genetic data extend and support current knowledge from historical records describing the common phenomenon of miscegenation due to sexual encounters between European male settlers, mainly from the Netherlands, Britain, Germany, and France, or male slaves and autochthonous Khoisan women.<sup>6</sup> This is not surprising in view of the singular shortage of European women and the reports stating that the proportion of adult male slaves to adult females was up to six to one.<sup>6</sup> A large number of female slaves were given their freedom, and many subsequently married European men after their manumission.<sup>8</sup> Conversely, unions between autochthonous African or southeast Asian slave males and European women were not tolerated.<sup>8</sup> Although Khoisan women were frequently incorporated into the settler households as servants, Khoisan men had no role and were marginalized by the society.<sup>7</sup> Offspring of mixed unions usually became members of the newly emerging community and continued to blend within it, establishing the basis of the contemporary SAC population.<sup>6</sup>

In summary, our results indicate that the gene pool of the contemporary SAC is made up of distinct continental and within-continental components, which have participated in the foundation of this recently admixed population in a strong gender-biased fashion. This pattern within the microcosm of the SAC community reflects the plethora of cultures, languages, and religious beliefs practiced in South Africa, which hosts ethnically diverse populations. Anecdotally, it is noteworthy that the three most frequent combinations of paired paternal/maternal lineages observed among contemporary Coloured are R-M343/L0d, E-M2/L0d, and O-M175/L0d, capturing the essence and the outcome of the first European, pan-African, and south Asian encounters with the original Khoisan-speaking peoples of the Cape of Good Hope approximately 350 years ago.

### Supplemental Data

Supplemental Data include two figures and one table and can be found with this article online at <http://www.cell.com/AJHG>.

### Acknowledgments

We warmly thank all study participants. We are very grateful to Blandine Massonnet for help in genotyping of mtDNA. This work was supported by Institut Pasteur and by the Centre National de la Recherche Scientifique (CNRS). Mitochondrial DNA and Y chromosome databases used for comparative analyses were partly supported by the Russian Fund of Humanities (grant 07-01-12114) and created with the help of Andrey Pshenichnov, Roman Sychev, and Elena Balanovska. P.D.V.H. and E.G.H. are supported by the Department of Science and Technology (DST) and the National

Research Foundation (NRF) Centre of Excellence for Biomedical Tuberculosis Research.

Received: January 5, 2010

Revised: January 27, 2010

Accepted: February 15, 2010

Published online: March 25, 2010

### Web Resources

The URLs for data presented herein are as follows:

Y Chromosome Haplotype Reference Database, YHRD, <http://www.yhrd.org/>

Fluxus Engineering, <http://www.fluxus-technology.com/>

Statistics South Africa 2007, <http://www.statssa.gov.za/>

### References

1. Smith, M.W., and O'Brien, S.J. (2005). Mapping by admixture linkage disequilibrium: Advances, limitations and guidelines. *Nat. Rev. Genet.* 6, 623–632.
2. Smith, M.W., Lautenberger, J.A., Shin, H.D., Chretien, J.P., Shrestha, S., Gilbert, D.A., and O'Brien, S.J. (2001). Markers for mapping by admixture linkage disequilibrium in African American and Hispanic populations. *Am. J. Hum. Genet.* 69, 1080–1094.
3. Tian, C., Hinds, D.A., Shigeta, R., Kittles, R., Ballinger, D.G., and Seldin, M.F. (2006). A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. *Am. J. Hum. Genet.* 79, 640–649.
4. Mao, X., Bigham, A.W., Mei, R., Gutierrez, G., Weiss, K.M., Brutsaert, T.D., Leon-Velarde, F., Moore, L.G., Vargas, E., McKeigue, P.M., et al. (2007). A genomewide admixture mapping panel for Hispanic/Latino populations. *Am. J. Hum. Genet.* 80, 1171–1178.
5. Tian, C., Hinds, D.A., Shigeta, R., Adler, S.G., Lee, A., Pahl, M.V., Silva, G., Belmont, J.W., Hanson, R.L., Knowler, W.C., et al. (2007). A genomewide single-nucleotide-polymorphism panel for Mexican American admixture mapping. *Am. J. Hum. Genet.* 80, 1014–1023.
6. Nurse, G.T., Weiner, S., and Jenkins, T. (1985). *The Peoples of Southern Africa and Their Affinities* (Oxford: Clarendon Press).
7. Mountain, A. (2003). *The First People of the Cape* (Claremont, South Africa: David Philips Publishers).
8. Shell, R. (1994). *Children of bondage* (Johannesburg, South Africa: Witwatersrand University Press).
9. Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.M., Doumbo, O., et al. (2009). The genetic structure and history of Africans and African Americans. *Science* 324, 1035–1044.
10. Patterson, N., Petersen, D.C., van der Ross, R.E., Sudoyo, H., Glashoff, R.H., Marzuki, S., Reich, D., and Hayes, V.M. (2010). Genetic structure of a unique admixed population: Implications for medical research. *Hum. Mol. Genet.* 19, 411–419.
11. Cavalli-Sforza, L.L., and Feldman, M.W. (2003). The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.* 33 (Suppl), 266–275.



12. Jobling, M.A., and Tyler-Smith, C. (2003). The human Y chromosome: An evolutionary marker comes of age. *Nat. Rev. Genet.* 4, 598–612.
13. Pakendorf, B., and Stoneking, M. (2005). Mitochondrial DNA and human evolution. *Annu. Rev. Genomics Hum. Genet.* 6, 165–183.
14. Underhill, P.A., and Kivisild, T. (2007). Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu. Rev. Genet.* 41, 539–564.
15. Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Mirazon Lahr, M., Foley, R.A., Oefner, P.J., and Cavalli-Sforza, L.L. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* 65, 43–62.
16. Zhivotovsky, L.A., Underhill, P.A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T., Scozzari, R., Cruciani, F., Destro-Bisol, G., Spedini, G., et al. (2004). The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* 74, 50–61.
17. Torroni, A., Achilli, A., Macaulay, V., Richards, M., and Bandelt, H.J. (2006). Harvesting the fruit of the human mtDNA tree. *Trends Genet.* 22, 339–345.
18. Bamshad, M.J., Watkins, W.S., Dixon, M.E., Jorde, L.B., Rao, B.B., Naidu, J.M., Prasad, B.V., Rasanayagam, A., and Hammer, M.F. (1998). Female gene flow stratifies Hindu castes. *Nature* 395, 651–652.
19. Destro-Bisol, G., Donati, F., Coia, V., Boschi, I., Verginelli, F., Caglia, A., Tofanelli, S., Spedini, G., and Capelli, C. (2004). Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Mol. Biol. Evol.* 21, 1673–1682.
20. Dupanloup, I., Pereira, L., Bertorelle, G., Calafell, F., Prata, M.J., Amorim, A., and Barbujani, G. (2003). A recent shift from polygyny to monogamy in humans is suggested by the analysis of worldwide Y-chromosome diversity. *J. Mol. Evol.* 57, 85–97.
21. Kayser, M., Brauer, S., Weiss, G., Schiefenhover, W., Underhill, P., Shen, P., Oefner, P., Tommaseo-Ponzetta, M., and Stoneking, M. (2003). Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am. J. Hum. Genet.* 72, 281–302.
22. Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T., and Stoneking, M. (2001). Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nat. Genet.* 29, 20–21.
23. Salem, A.H., Badr, F.M., Gaballah, M.F., and Paabo, S. (1996). The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. *Am. J. Hum. Genet.* 59, 741–743.
24. Wilder, J.A., Kingan, S.B., Mobasher, Z., Pilkington, M.M., and Hammer, M.F. (2004). Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nat. Genet.* 36, 1122–1125.
25. Wilder, J.A., Mobasher, Z., and Hammer, M.F. (2004). Genetic evidence for unequal effective population sizes of human females and males. *Mol. Biol. Evol.* 21, 2047–2057.
26. Wooding, S., Ostler, C., Prasad, B.V., Watkins, W.S., Sung, S., Bamshad, M., and Jorde, L.B. (2004). Directional migration in the Hindu castes: inferences from mitochondrial, autosomal and Y-chromosomal data. *Hum. Genet.* 115, 221–229.
27. Seielstad, M.T., Minch, E., and Cavalli-Sforza, L.L. (1998). Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* 20, 278–280.
28. Achilli, A., Rengo, C., Battaglia, V., Pala, M., Olivieri, A., Fornarino, S., Magri, C., Scozzari, R., Babudri, N., Santachiara-Benerecetti, A.S., et al. (2005). Saami and Berbers—an unexpected mitochondrial DNA link. *Am. J. Hum. Genet.* 76, 883–886.
29. Behar, D.M., VILLEMS, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S., Comas, D., et al. (2008). The dawn of human matrilineal diversity. *Am. J. Hum. Genet.* 82, 1130–1140.
30. Palanichamy, M.G., Sun, C., Agrawal, S., Bandelt, H.J., Kong, Q.P., Khan, F., Wang, C.Y., Chaudhuri, T.K., Palla, V., and Zhang, Y.P. (2004). Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am. J. Hum. Genet.* 75, 966–978.
31. Salas, A., Richards, M., De la Fe, T., Lareu, M.V., Sobrino, B., Sanchez-Diz, P., Macaulay, V., and Carracedo, A. (2002). The making of the African mtDNA landscape. *Am. J. Hum. Genet.* 71, 1082–1111.
32. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., and Wilson, A.C. (1991). African populations and the evolution of human mitochondrial DNA. *Science* 253, 1503–1507.
33. Knight, A., Underhill, P.A., Mortensen, H.M., Zhivotovsky, L.A., Lin, A.A., Henn, B.M., Louis, D., Ruhlen, M., and Mountain, J.L. (2003). African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr. Biol.* 13, 464–473.
34. Tishkoff, S.A., Gonder, M.K., Henn, B.M., Mortensen, H., Knight, A., Gignoux, C., Fernandopulle, N., Lema, G., Nyambo, T.B., Ramakrishnan, U., et al. (2007). History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol. Biol. Evol.* 24, 2180–2195.
35. Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M.J., and Amorim, A. (2001). Prehistoric and historic traces in the mtDNA of Mozambique: Insights into the Bantu expansions and the slave trade. *Ann. Hum. Genet.* 65, 439–458.
36. Kong, Q.P., Bandelt, H.J., Sun, C., Yao, Y.G., Salas, A., Achilli, A., Wang, C.Y., Zhong, L., Zhu, C.L., Wu, S.F., et al. (2006). Updating the East Asian mtDNA phylogeny: A prerequisite for the identification of pathogenic mutations. *Hum. Mol. Genet.* 15, 2076–2086.
37. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T., et al. (2004). Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet.* 5, 26.
38. Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A.S., et al. (2004). Where west meets east: The complex mtDNA landscape of the southwest and Central Asian corridor. *Am. J. Hum. Genet.* 74, 827–845.
39. Thangaraj, K., Chaubey, G., Singh, V.K., Vanniarajan, A., Thanseem, I., Reddy, A.G., and Singh, L. (2006). In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup ‘M’ in India. *BMC Genomics* 7, 151.
40. Sun, C., Kong, Q.P., Palanichamy, M.G., Agrawal, S., Bandelt, H.J., Yao, Y.G., Khan, F., Zhu, C.L., Chaudhuri, T.K., and

- Zhang, Y.P. (2006). The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol. Biol. Evol.* 23, 683–690.
41. Baig, M.M., Khan, A.A., and Kulkarni, K.M. (2004). Mitochondrial DNA diversity in tribal and caste groups of Maharashtra (India) and its implication on their genetic origins. *Ann. Hum. Genet.* 68, 453–460.
  42. Tofanelli, S., Bertoincini, S., Castri, L., Luiselli, D., Calafell, F., Donati, G., and Paoli, G. (2009). On the origins and admixture of Malagasy: New evidence from high-resolution analyses of paternal and maternal lineages. *Mol. Biol. Evol.* 26, 2109–2124.
  43. Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., et al. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* 67, 1251–1276.
  44. Torroni, A., Richards, M., Macaulay, V., Forster, P., Villems, R., Norby, S., Savontaus, M.L., Huoponen, K., Scozzari, R., and Bandelt, H.J. (2000). mtDNA haplogroups and frequency patterns in Europe. *Am. J. Hum. Genet.* 66, 1173–1177.
  45. Behar, D.M., Rosset, S., Blue-Smith, J., Balanovsky, O., Tzur, S., Comas, D., Mitchell, R.J., Quintana-Murci, L., Tyler-Smith, C., and Wells, R.S. (2007). The Genographic Project public participation mitochondrial DNA database. *PLoS Genet.* 3, e104.
  46. Forster, P., Harding, R., Torroni, A., and Bandelt, H.J. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.* 59, 935–945.
  47. Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., and Hammer, M.F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18, 830–838.
  48. Sims, L.M., Garvey, D., and Ballantyne, J. (2007). Sub-populations within the major European and African derived haplogroups R1b3 and E3a are differentiated by previously phylogenetically undefined Y-SNPs. *Hum. Mutat.* 28, 97.
  49. Semino, O., Santachiara-Benerecetti, A.S., Falaschi, F., Cavalliforza, L.L., and Underhill, P.A. (2002). Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am. J. Hum. Genet.* 70, 265–268.
  50. Rosa, A., Brehm, A., Kivisild, T., Metspalu, E., and Villems, R. (2004). MtDNA profile of West Africa Guineans: Towards a better understanding of the Senegambia region. *Ann. Hum. Genet.* 68, 340–352.
  51. Zerjal, T., Beckman, L., Beckman, G., Mikelsaar, A.V., Krumina, A., Kucinkas, V., Hurler, M.E., and Tyler-Smith, C. (2001). Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. *Mol. Biol. Evol.* 18, 1077–1087.
  52. Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., et al. (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: A Y chromosome perspective. *Science* 290, 1155–1159.
  53. Wells, R.S., Yuldashева, N., Ruzibakiev, R., Underhill, P.A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., et al. (2001). The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA* 98, 10244–10249.
  54. Bosch, E., Calafell, F., Comas, D., Oefner, P.J., Underhill, P.A., and Bertranpetit, J. (2001). High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am. J. Hum. Genet.* 68, 1019–1029.
  55. Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., Help, H., Bermisheva, M., Kutuev, I., Barac, L., Pericic, M., Balanovsky, O., et al. (2004). Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am. J. Hum. Genet.* 75, 128–137.
  56. Quintana-Murci, L., Krausz, C., Zerjal, T., Sayar, S.H., Hammer, M.F., Mehdi, S.Q., Ayub, Q., Qamar, R., Mohyuddin, A., Radhakrishna, U., et al. (2001). Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am. J. Hum. Genet.* 68, 537–542.
  57. Underhill, P.A., Myres, N.M., Rootsi, S., Metspalu, M., Zhivotovsky, L.A., King, R.J., Lin, A.A., Chow, C.E., Semino, O., Battaglia, V., et al. (2009). Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur. J. Hum. Genet.* Published online November 4, 2009. 10.1038/ejhg.2009.19.
  58. Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H.V., Stepanov, V., et al. (2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* 72, 313–332.
  59. Xue, Y., Zerjal, T., Bao, W., Zhu, S., Shu, Q., Xu, J., Du, R., Fu, S., Li, P., Hurler, M.E., et al. (2006). Male demography in East Asia: A north-south contrast in human population expansion times. *Genetics* 172, 2431–2439.
  60. Deng, W., Shi, B., He, X., Zhang, Z., Xu, J., Li, B., Yang, J., Ling, L., Dai, C., Qiang, B., et al. (2004). Evolution and migration history of the Chinese population inferred from Chinese Y-chromosome evidence. *J. Hum. Genet.* 49, 339–348.
  61. Karafet, T., Xu, L., Du, R., Wang, W., Feng, S., Wells, R.S., Redd, A.J., Zegura, S.L., and Hammer, M.F. (2001). Paternal population history of East Asia: Sources, patterns, and microevolutionary processes. *Am. J. Hum. Genet.* 69, 615–628.
  62. Watkins, W.S., Thara, R., Mowry, B.J., Zhang, Y., Witherspoon, D.J., Tolpinrud, W., Bamshad, M.J., Tirupati, S., Padmavati, R., Smith, H., et al. (2008). Genetic variation in South Indian castes: Evidence from Y-chromosome, mitochondrial, and autosomal polymorphisms. *BMC Genet.* 9, 86.
  63. Balanovsky, O., Rootsi, S., Pshenichnov, A., Kivisild, T., Churnosov, M., Evseeva, I., Pocheshkhova, E., Boldyreva, M., Yankovsky, N., Balanovska, E., et al. (2008). Two sources of the Russian patrilineal heritage in their Eurasian context. *Am. J. Hum. Genet.* 82, 236–250.
  64. Bertorelle, G., and Excoffier, L. (1998). Inferring admixture proportions from molecular data. *Mol. Biol. Evol.* 15, 1298–1311.
  65. Dupanloup, I., and Bertorelle, G. (2001). Inferring admixture proportions from molecular data: Extension to any number of parental populations. *Mol. Biol. Evol.* 18, 672–675.
  66. Richards, M.B., Macaulay, V.A., Bandelt, H.J., and Sykes, B.C. (1998). Phylogeography of mitochondrial DNA in western Europe. *Ann. Hum. Genet.* 62, 241–260.