Journal of Human Genetics (2012) 57, 65–69

www.nature.com/jhg

SHORT COMMUNICATION

A multiplex SNP assay for the dissection of human Y-chromosome haplogroup O representing the major paternal lineage in East and Southeast Asia

Mannis van Oven, Nathalie van den Tempel and Manfred Kayser

The majority of human Y chromosomes in men from East and Southeast Asia, and a considerable proportion of Oceanian men, especially those from Remote Oceania, belong to haplogroup O, characterized by a 5-bp deletion known as M175 (rs2032678). Recent advances in Y-SNP (single-nucleotide polymorphism) discovery have substantially improved the phylogenetic resolution of haplogroup O sublineages. By taking advantage of this recent knowledge, we hereby introduce a sensitive Y-SNP multiplex genotyping assay for the dissection of haplogroup O into its most significant sublineages. The multiplex assay thus provides an efficient way to infer patrilineal biogeographic ancestry in males of Asian/Oceanian patrilineal descent, and is suitable for applications in human population genetics, anthropological, genealogical, as well as forensic studies.
Journal of Human Genetics (2012) **57,** 65–69; doi:10.1038/jhg.2011.120; published online 3 November 2011

Keywords: East Asia; haplogroup O-M175; patrilineal ancestry; phylogeny; SNP multiplex; Southeast Asia; subhaplogroups; Y chromosome

INTRODUCTION

The paternally inherited human Y chromosome has a well-established phylogeny[1] in which each clade, or haplogroup, is defined by one or more binary markers, such as single-nucleotide polymorphisms and small insertion/deletion polymorphisms (here collectively referred to as SNPs). Each haplogroup has a specific, and sometimes distinct, geographical distribution, shaped by demographic and evolutionary events, such as range expansion, migration and drift. Y-chromosome haplogroup O, characterized by a 5-bp deletion known as M175 (rs2032678), is the dominant haplogroup among males throughout East and Southeast Asia, where its frequency typically ranges between 50 and 100%,[2,3] while being more rare in Central Asia.[4] In addition, haplogroup O is observed in coastal and island parts of Near Oceania,[5] is frequent overall in Remote Oceania,[5] and is present in Madagascar,[6] all due to population movements from East/Southeast Asia. Although several SNPs phylogenetically downstream of M175 are known since more than a decade,[7] recent advances in Y-SNP discovery have further resolved the internal topology of haplogroup O substantially.[1,8,9] To make this accumulated knowledge easily available for future human Y-chromosome studies, we developed a multiplex Y-SNP genotyping tool, based on single-base primer extension (SNaPshot) technology, for the discrimination of the most significant haplogroup O sublineages.

MATERIALS AND METHODS

DNA samples

For testing and validation of the assay, we used DNA from samples of the HapMap 3 reference panel,[10] obtained through the Coriell Institute for Medical Research (http://www.coriell.org/), from samples of the European Collection of Cell Cultures (ECACC) ethnic diversity panel (EDP-1) (http://www.hpacultures.org.uk/products/dna/ethnicdna.jsp), and from samples described previously.[11]

Marker selection

Informative Y-SNP markers were selected by surveying haplogroup frequency data available from the literature as summarized in Figure 1. The final marker selection included M175, M119, P203 (also known as M307), M110, M268, M95, M88, M176 (also known as SRY465 or PS63), M122, M324, KL1 (also known as L465), 002611, P201 (also known as 021354), M7, M134 and PS23.

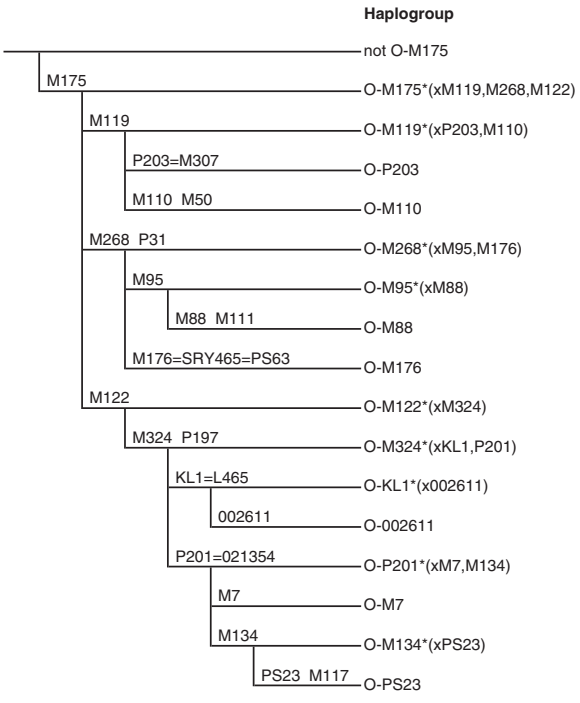
Primer design and genotyping protocol

PCR and extension primers (Table 1) were designed as described previously.[12]

Multiplex PCR amplification was carried out in a reaction volume of 6 µl, containing 1× GeneAmp PCR Gold Buffer (Applied Biosystems, Foster City, CA, USA), 4.5 mM $MgCl_2$ (Applied Biosystems), 100 µM of each dNTP (Roche, Mannheim, Germany), 0.35 units of AmpliTaq Gold DNA polymerase (Applied Biosystems), 1–2 ng of genomic DNA template and PCR primers (desalted; Metabion, Martinsried, Germany) in concentrations as specified in Table 1. The reactions were performed in a Dual 384-well GeneAmp PCR System 9700 (Applied Biosystems) with the following cycling conditions: 10 min at 95 °C, followed by 30 cycles of 94 °C for 15 s, 60 °C for 45 s, and a final extension at 60 °C for 5 min. PCR products were purified by adding 2 µl ExoSAP-IT (USB Corporation, Cleveland, OH, USA) to 6 µl PCR product and incubation at 37 °C for 30 min followed by 80 °C for 15 min.

Multiplex single-base primer extension was carried out in a reaction volume of 6 µl, containing 1 µl SNaPshot Ready Reaction Mix (Applied Biosystems), 1 µl purified PCR product and extension primers (high-performance liquid

Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands
Correspondence: M van Oven, Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, 3000 CA Rotterdam, The Netherlands.
E-mail: m.vanoven@erasmusmc.nl
Received 8 August 2011; revised 19 September 2011; accepted 25 September 2011; published online 3 November 2011



Haplogroup	Northeast Indians n=173	Han Chinese n=361	Koreans n=506	Japanese n=263	Taiwanese aborigines n=48	Western Indonesians[a] n=319	Eastern Indonesians[b] n=903	Admiralty Islanders n=148	Solomon Islanders n=712	Polynesians[c] n=320	Malagasy n=35	U.S. Asian Americans n=62
not O-M175	9.8	19.1	21.1	45.6	2.1	18.5	84.7	81.8	72.5	66.3	65.7	11.3
O-M175*(xM119,M268,M122)	3.5	-	-	-	-	-	-	-	-	-	-	-
O-M119*(xP203,M110)		1.4	2.2	3.4	-	15.4	2.5	-	1.7	0.6	-	17.7
O-P203		13.0			70.8	25.7	4.5					
O-M110		0.8		-	18.8	4.7	2.5	17.6	4.6	0.3	17.1	
O-M268*(xM95,M176)		6.4	1.0	-	-	-	-	-	-	-	-	1.6
O-M95*(xM88)		2.5		0.8	-	16.9	2.1	-	-	-	17.1	6.5
O-M88		-		-	2.1	-	-	-	7.4	1.3	-	
O-M176		0.3	31.4	33.5	-	-	-	-	-	-	-	8.1
O-M122*(xM324)	1.2	1.7	0.4	1.1	-	-	0.8	-	0.6	-	-	25.8
O-M324*(xKL1,P201)		-	15.0		-	-	-	0.7	13.2	30.0		
O-KL1*(x002611)		3.0										
O-002611		16.9	-	3.8	-	-	-					
O-P201*(xM7,M134)		5.3	28.9	4.2	6.3	11.3	2.8					
O-M7		1.9		-	-	7.5	-	-	-	0.6		
O-M134*(xPS23)	85.5	11.4		3.4	-	-	-	-	-	0.9	-	29.0
O-PS23		16.3		4.2								
haplogroup O total	90.2	80.9	78.9	54.4	97.9	81.5	15.3	18.2	27.5	33.8	34.3	88.7

[a] composite of individuals from Borneo (n=86), Sumatra (n=38), Nias (n=60), Mentawai (n=74) and Java (n=61)
[b] composite of individuals from Sumba (n=350), Flores (n=394), Lembata (n=92), Alor (n=28), Timor (n=9) and Moluccas (n=30)
[c] composite of individuals from Futuna (n=50), Tuvalu (n=100), Tonga (n=28), Samoa (n=62), Tokelau (n=6), Niue (n=8) and Cook Islands (n=66)

Figure 1 Marker phylogeny of the Y-SNPs included in the multiplex assay (left part), with previously reported haplogroup frequency data (in percentages) for a range of populations (right part). When the SNP resolution of the available haplogroup frequency data did not match that of the multiplex assay, multiple haplogroups were combined into one class as indicated by the box lines; for example, 2.2% of Koreans fall within haplogroup O-M119, but it is not known how these are distributed over O-M119*(xP203,M110), O-P203 and O-M110, because markers P203 and M110 were not typed in this population sample. In addition, some studies used a different, but phylo-equivalent marker, as compared with the marker included in our assay; in such case, the alternative marker is also shown in the phylogeny; for example, in the Malagasy marker M50 was typed rather than M110. Data sources for haplogroup frequencies are as follows: northeast Indians;[13] Han Chinese;[9] Koreans;[14] Japanese;[8] Taiwanese aborigines, western Indonesians and eastern Indonesians;[3] Admiralty Islanders, Solomon Islanders and Polynesians;[15] Malagasy;[6] US Asian Americans.[16]

chromatography-purified; Metabion) in concentrations as specified in Table 1. The reactions were performed in a Dual 384-well GeneAmp PCR System 9700 (Applied Biosystems) with the following cycling conditions: 2 min at 96 °C, followed by 25 cycles of 96 °C for 10 s, 50 °C for 5 s and 60 °C for 30 s. The reaction products were purified by adding 1 unit of shrimp alkaline phosphatase (USB Corporation) to 6 µl of extension product, and incubation at 37 °C for 45 min followed by 75 °C for 15 min.

The extended fragments were analyzed by capillary electrophoresis using a 3130*xl* Genetic Analyzer (Applied Biosystems) with POP-7 polymer. A mixture of 1 µl purified extension product, 8.7 µl Hi-Di formamide (Applied Biosystems) and 0.3 µl GeneScan-120 LIZ internal size standard (Applied Biosystems) was run with 23 s injection time at 1.2 kV, and 500 s run time at 15.0 kV. Data were analyzed using GeneMapper version 3.7 software (Applied Biosystems).

RESULTS AND DISCUSSION

The assay introduced here includes a total of 16 Y-SNPs, among which there are four recently discovered markers (P203, KL1, 002611 and P201) that prove phylogenetically informative and hence will aid to improve the resolution of future Y-chromosome studies. Figure 1 shows the phylogenetic relationship of the 16 Y-SNPs included and provides frequency data of the respective haplogroups for a range of populations. Figure 2 shows typical electropherograms obtained with the multiplex assay for a range of DNA samples with different haplogroup O subgroups. In all cases, all 16 allele peaks were clearly visible and showed no violation of the established marker topology (Figure 1). The assay is sensitive by virtue of small amplicon sizes (minimum: 45 bp; maximum: 123 bp; average: 90 bp), and is therefore expected to be applicable using small amounts of (degraded) DNA, such as is often encountered in forensic casework and ancient DNA studies. Typically, we obtained good results when using template DNA amounts of 1–2 ng. Furthermore, we tested our assay on a number of commonly available reference DNA samples; we report the determined Y haplogroups in Table 2 so that other researchers can use them as control DNA samples in future studies.

Table 1 Genotyping details of the Y-SNP haplogroup O multiplex

		PCR amplification				*Single-base extension*				
Locus	*Mutation*	*Primer sequences (5′>3′)*		*Conc. (μM)*	*Amplicon size (bp)*	*Primer sequence (5′>3′) (5′-aspecific tail in lowercase italics)*	*Conc. (μM)*	*Length (nt)*	*Orientation*	*Alleles (dye)*
M175	5-bp del	F	CCCAAATCAACTCAACTCCAG	0.500	101/96	*t(gact)$_{10}$* CACATGCCTTCTCACTTCTC	0.500	61	F	a (red), d (green)
		R	TTCTACTGATACCTTTGTTTCTGTTCA	0.500						
M119	A→C	F	CAAACCGCAGTGCTATGTGT	0.600	93	*(gact)$_{17}$gac* CCAATTCAGCATACAGGC	0.200	89	R	A (red), C (blue)
		R	TGGGTTATTCCAATTCAGCA	0.600						
P203	G→A	F	GGCTACATGGAAATGGTTGG	0.100	89	*(gact)$_{7}$gac* GGCTATTGAGTTAGCATAATCA	0.200	53	F	G (blue), A (green)
		R	TTCTCACTTAGCACATATACAAAAGGT	0.100						
M110	T→C	F	CGAGAACGTTCCTGTCACAA	0.200	119	*t(gact)$_{2}$g* GGATGCCGGTACAATGTATT	0.200	30	F	T (red), C (yellow)
		R	AAAATCCAACGACAAATGTGC	0.200						
M268	A→G	F	CATGCCTAGCCTCATTCCTC	0.200	84	*act(gact)$_{3}$* CCTAGCCTCATTCCTCTAAAAT	0.200	37	F	A (green), G (blue)
		R	AACCCTGATCATTCCCCTTC	0.200						
M95	C→T	F	CCTTCTTGGGATCAAATGGA	0.800	86	*act(gact)$_{3}$* GATAAGGAAAGACTACCATATTAGTG	0.400	41	F	C (yellow), T (red)
		R	GCCTACAGGTTGGAAAGGCTA	0.800						
M88	A→G	F	GCTATGGCCTAGGTGCTTTTC	0.200	107	*ct(gact)$_{5}$gac* CTTATTCCTGCTTCTTCTGC	0.200	45	F	A (green), G (blue)
		R	CACAGGCCTTAGAGAGGTAGTCA	0.200						
M176	C→T	F	ATCCCGCTTCGGTACTCTG	0.100	83	*(gact)$_{7}$ga* TGTTGTCCAGTTGCACTTC	0.200	49	R	C (blue), T (green)
		R	TCTTGAGTGTGTGGCTTTCG	0.100						
M122	T→C	F	TTAGTTGCCTTTTGGAAATGAA	0.200	80	*t(gact)$_{9}$ga* AGATTTTCCCCTGAGAGC	0.200	57	R	T (green), C (blue)
		R	TCAGATTTTCCCCTGAGAGC	0.200						
M324	G→C	F	GATTTGATCTACCTGCCCTTTC	0.100	45	*act(gact)$_{11}$ga* ATGGGCTGCAACAAGA	0.200	65	R	G (yellow), C (blue)
		R	ATACATGGGCTGCAACAAGA	0.100						
KL1	G→T	F	TAGATGGTTGAGACACATCTTCA	0.600	116	*(gact)$_{18}$g* GGTTGCTAGAATTGCACAAT	0.400	93	R	G (yellow), T (green)
		R	ACTCCAGGATGTTTGGGAAC	0.600						
002611	C→T	F	GCCCTGCTAGTAGGCACCA	0.600	81	*act(gact)$_{16}$ga* AAGTGCAGCAGTGGCC	0.400	85	R	C (blue), T (green)
		R	AATCCAATGACCCTTTGCAG	0.600						
P201	T→C	F	GTGCTGTGCAAGTTGTGTGA	0.050	96	*(gact)$_{12}$gac* GAGAGCCAGTTAAAGCCC	0.100	69	R	T (green), C (blue)
		R	TGGGTGCAGTTAAGCAATGA	0.050						
M7	C→G	F	CATCACCAAAGGGCATGTAAT	1.200	123	*t(gact)$_{11}$gac* CCTCTCTGTATGTCAGATCTAACAA	0.800	73	F	C (yellow), G (blue)
		R	TTGTCCCTGCAGCCTTGT	1.200						
M134	1-bp del	F	AAGAAAAGGCCCAGGAAAGT	0.050	47/46	*ct(gact)$_{14}$g* CTTTTGATCCCCACCAAT	0.200	77	R	a (yellow), d (green)
		R	GAGATACTTTTGATCCCCACCA	0.050						
PS23	G→A	F	AACAGTGATGGGAGTGTGACC	0.200	97	*t(gact)$_{16}$* GTCTCCCCGAGGAGCC	0.200	81	F	G (blue), A (green)
		R	TTGTCCAGCTCCAAAAGACA	0.200						

Abbreviations: a, ancestral; d, deletion, F, forward; R, reverse; SNP, single-nucleotide polymorphism.

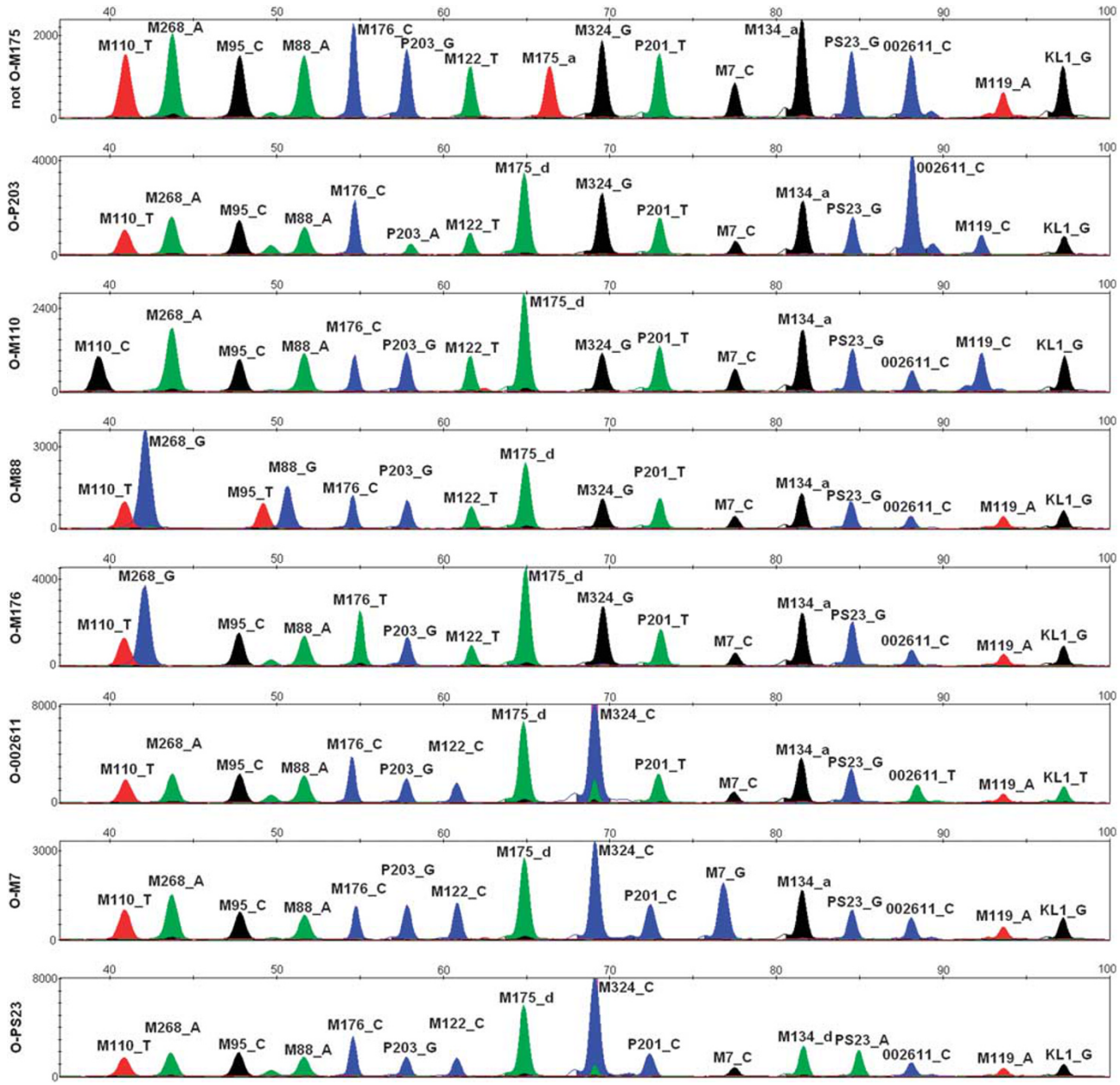


Figure 2 Typical electropherograms obtained with the Y-haplogroup O multiplex assay introduced here. Samples belonging to a range of different haplogroups (indicated at the left of each electropherogram) were selected such that every possible allele can be seen at least once. For each peak, the detected allele is indicated in concordance with Table 1. As is convention, the yellow dye is shown as black for better contrast.

Previous population genetic studies have employed various combinations of the hitherto-known SNPs within Y-chromosome haplogroup O. However, some of the recently discovered Y-SNPs appear to be informative for the breakup of previously undifferentiated clusters. For instance, a considerable fraction of Southeast Asian males that were previously classified as O-M119*(xM110), turned out to belong to O-P203,[3] which is a subhaplogroup of O-M119 and a sister haplogroup to O-M110.[1] We foresee that future studies making use of our assay will benefit from the inclusion of such an increased marker set, allowing to reveal patterns of genetic distribution that would otherwise (at lower phylogenetic resolution) remain unnoticed. Furthermore, such studies will generate further knowledge regarding the precise geographic distribution of each of the haplogroup O sublineages.

In conclusion, we provide a convenient and sensitive multiplex genotyping assay for the dissection of the most significant Y-chromosome haplogroup O sublineages. The assay can be applied to male DNA samples for which prior testing (for example, by the use of a global Y-SNP assay[12]) revealed haplogroup O status, to retrieve more detailed Y-chromosome diversity and patrilineal biogeographic ancestry information, thus being of relevance in human

Table 2 Haplogroup information of commonly available reference samples

DNA Panel	*Sample ID*	*Sample designation*[a]	*Haplogroup*
HapMap	NA18562	Han Chinese	O-M134*(xPS23)
HapMap	NA18572	Han Chinese	O-M134*(xPS23)
ECACC EDP-1	DCH007	Thai	O-PS23
ECACC EDP-1	WATANABE	Oriental	O-M176
ECACC EDP-1	DCH008	Thai	O-M95*(xM88)
ECACC EDP-1	HAU,ML	Oriental	O-M268*(xM95,M176)
ECACC EDP-1	WHONP192	Oriental	O-M134*(xPS23)
ECACC EDP-1	DCH009	Thai	O-002611
ECACC EDP-1	HOKKAIDO	Oriental	O-M176
ECACC EDP-1	SM	Japanese	O-M176
ECACC EDP-1	CHI-007	Thai	O-M134*(xPS23)
ECACC EDP-1	DCH010	Thai	O-M95*(xM88)
ECACC EDP-1	TAB089	Japanese	O-002611
ECACC EDP-1	DCH012	Thai	O-PS23
ECACC EDP-1	KUROIWA	Oriental	O-M176
ECACC EDP-1	PETCH	Oriental	O-M88
ECACC EDP-1	NP369	Oriental	O-P201*(xM7,M134)

[a]Information as supplied by the distributor.

population genetics, anthropological, genealogical, as well as forensic studies.

ACKNOWLEDGEMENTS

We thank Qasim Ayub and Chris Tyler-Smith for help in providing access to HapMap DNA samples, Arwin Ralf for technical assistance, and Susan Walsh for language checking of the manuscript. This work was financially supported in part by the Netherlands Forensic Institute (NFI), and by a grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN).

1 Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L. & Hammer, M. F. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* **18,** 830–838 (2008).
2 Zhong, H., Shi, H., Qi, X. B., Duan, Z. Y., Tan, P. P., Jin, L. *et al.* Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Mol. Biol. Evol.* **28,** 717–727 (2011).
3 Karafet, T. M., Hallmark, B., Cox, M. P., Sudoyo, H., Downey, S., Lansing, J. S. *et al.* Major east-west division underlies Y chromosome stratification across Indonesia. *Mol. Biol. Evol.* **27,** 1833–1844 (2010).
4 Marchani, E. E., Watkins, W. S., Bulayeva, K., Harpending, H. C. & Jorde, L. B Culture creates genetic structure in the Caucasus: autosomal, mitochondrial, and Y-chromosomal variation in Daghestan. *BMC Genet.* **9,** 47 (2008).
5 Kayser, M. The human genetic history of Oceania: near and remote views of dispersal. *Curr. Biol.* **20,** R194–R201 (2010).
6 Hurles, M. E., Sykes, B. C., Jobling, M. A. & 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.* **76,** 894–901 (2005).
7 Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H. *et al.* Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26,** 358–361 (2000).
8 Nonaka, I., Minaguchi, K. & Takezaki, N. Y-chromosomal binary haplogroups in the Japanese population and their relationship to 16 Y-STR polymorphisms. *Ann. Hum. Genet.* **71,** 480–495 (2007).
9 Yan, S., Wang, C. C., Li, H., Li, S. L., Jin, L. & The Genographic Consortium An updated tree of Y-chromosome Haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *Eur. J. Hum. Genet.* **19,** 1013–1015 (2011).
10 Altshuler, D. M., Gibbs, R. A., Peltonen, L., Dermitzakis, E., Schaffner, S. F., Yu, F., International HapMap 3 Consortium *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467,** 52–58 (2010).
11 Kayser, M., Choi, Y., van Oven, M., Mona, S., Brauer, S., Trent, R. J *et al.* The impact of the Austronesian expansion: evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Mol. Biol. Evol.* **25,** 1362–1374 (2008).
12 van Oven, M., Ralf, A. & Kayser, M. An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups. *Int. J. Legal Med.* **125,** 879–885 (2011).
13 Cordaux, R., Weiss, G., Saha, N. & Stoneking, M. The northeast Indian passageway: a barrier or corridor for human migrations? *Mol. Biol. Evol.* **21,** 1525–1533 (2004).
14 Kim, S. H., Kim, K. C., Shin, D. J., Jin, H. J., Kwak, K. D., Han, M. S. *et al.* High frequencies of Y-chromosome haplogroup O2b-SRY465 lineages in Korea: a genetic perspective on the peopling of Korea. *Investig. Genet.* **2,** 10 (2011).
15 Delfin, F., Myles, S., Choi, Y., Hughes, D., Illek, R., van Oven, M. *et al.* Bridging near and Remote Oceania: MtDNA and NRY variation in the Solomon Islands. *Mol. Biol. Evol.* (e-pub ahead of print) (2011).
16 Hammer, M. F., Chamberlain, V. F., Kearney, V. F., Stover, D., Zhang, G., Karafet, T. *et al.* Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases. *Forensic Sci. Int.* **164,** 45–55 (2006).