



**HAL**  
open science

## Larger mitochondrial DNA than Y-chromosome differences between matrilineal and patrilineal groups from Sumatra

Ellen Dröfn Gunnarsdóttir, Madhusudan Nandineni, Mingkun Li, Sean Myles, David Gil, Brigitte Pakendorf, Mark Stoneking

► **To cite this version:**

Ellen Dröfn Gunnarsdóttir, Madhusudan Nandineni, Mingkun Li, Sean Myles, David Gil, et al.. Larger mitochondrial DNA than Y-chromosome differences between matrilineal and patrilineal groups from Sumatra. *Nature Communications*, 2011, 2, pp.228. 10.1038/ncomms1235 . hal-02006798

**HAL Id: hal-02006798**

**<https://hal.univ-lyon2.fr/hal-02006798v1>**

Submitted on 16 Jul 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ARTICLE

Received 12 Jul 2010 | Accepted 9 Feb 2011 | Published 8 Mar 2011

DOI: 10.1038/ncomms1235

# Larger mitochondrial DNA than Y-chromosome differences between matrilineal and patrilineal groups from Sumatra

Ellen Dröfn Gunnarsdóttir<sup>1</sup>, Madhusudan R. Nandineni<sup>2</sup>, Mingkun Li<sup>1</sup>, Sean Myles<sup>3</sup>, David Gil<sup>4</sup>,  
Brigitte Pakendorf<sup>5</sup> & Mark Stoneking<sup>1</sup>

Genetic differences between human populations are typically larger for the Y-chromosome than for mitochondrial DNA (mtDNA), which has been attributed to the ubiquity of patrilineality across human cultures. However, this claim has been disputed, and previous analyses of matrilineal groups give conflicting results. Here we analyse mtDNA variation (complete mtDNA genome sequences via next-generation sequencing) and non-recombining regions of the Y-chromosome variation (Y-single-nucleotide-polymorphisms and Y-short-tandem-repeats (STR)) in a matrilineal group (the Semende) and a patrilineal group (the Besemah) from Sumatra. We find in the Semende significantly lower mtDNA diversity than in the Besemah as expected for matrilineal groups, but unexpectedly we find no difference in Y-chromosome diversity between the groups. We highlight the importance of using complete mtDNA sequences for such analyses, as using only partial sequences (as done in previous studies) can give misleading results.

<sup>1</sup> Department of Evolutionary Anthropology, Max Planck Institute for Evolutionary Anthropology, Leipzig D-04103, Germany. <sup>2</sup> Laboratory of DNA Fingerprinting Services and Laboratory of Genomics and Profiling Applications, Centre for DNA Fingerprinting and Diagnostics, Building no. 7, Gruhakalpa, 5-4-399/B, Nampally, Hyderabad 500001, India. <sup>3</sup> Department of Genetics, Stanford University, Stanford, CA 94305-5120, USA. <sup>4</sup> Department of Linguistics, Max Planck, Max Planck Institute for Evolutionary Anthropology, Leipzig D-04103, Germany. <sup>5</sup> Research Group on Comparative Population Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig D-04103, Germany. Correspondence and requests for materials should be addressed to E.D.G. (email: [gunnarsdottir@eva.mpg.de](mailto:gunnarsdottir@eva.mpg.de)).

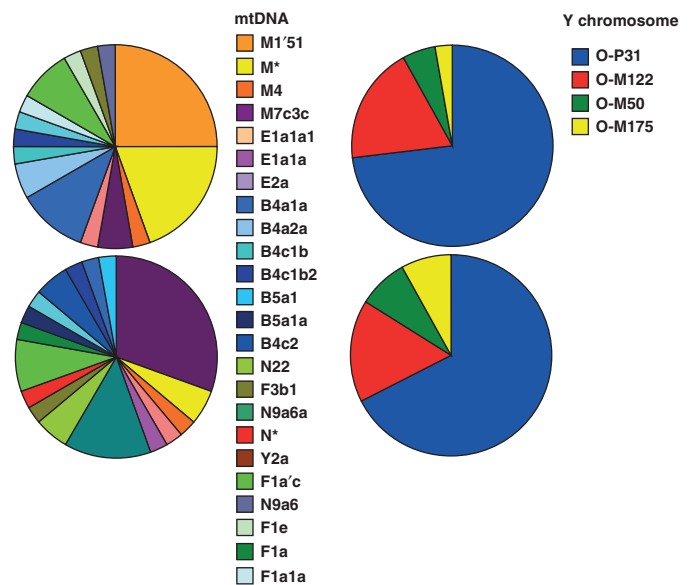
The extent to which cultural practices influence human genetic diversity has been a longstanding question in anthropological genetics. Genetic differences between human populations are usually larger for non-recombining regions of the Y-chromosome (NRY) than for mitochondrial DNA (mtDNA)<sup>1–3</sup>, and this pattern has been attributed to higher female than male migration due to patrilocality<sup>1,4</sup>, which is typical for about 70% of human societies<sup>5</sup>. Higher female than male migration results in a larger effective population size for females than males, which in turn predicts increased mtDNA and decreased Y-chromosome diversity within groups, and larger differences between groups for the Y-chromosome than for mtDNA. An obvious test of this hypothesis is that the genetic differences among matrilineal groups should then be larger for mtDNA than for the NRY and that the mtDNA diversity is lower in matrilineal groups than patrilineal. Indeed, this prediction was fulfilled in the first comparison of patterns of mtDNA and NRY variation in matrilineal and patrilineal groups, among the hill tribes of Thailand<sup>4,6</sup>. However, a subsequent study of matrilineal and patrilineal groups in India failed to find the predicted patterns of mtDNA versus NRY diversity<sup>7</sup>, whereas another study has called into question the original observation of larger differences among human groups in general for the NRY versus mtDNA<sup>8</sup>.

Clearly, studies of additional matrilineal groups are needed to ascertain if there is a general effect of residence pattern on human mtDNA and NRY diversity. We report here an analysis of complete mtDNA genome sequences (determined by next-generation sequencing) and NRY haplogroups and Y-STR haplotypes in a matrilineal group (the Semende) and a patrilineal group (the Besemah) from Sumatra, Indonesia. With respect to mtDNA, we find a lower haplotype diversity (HD) in the Semende, and a significantly large genetic distance between the Semende and the Besemah, as expected if matrilocality is influencing patterns of mtDNA diversity. Unexpectedly, and in contrast to virtually every other study of human mtDNA versus NRY diversity on the local scale, there are no significant differences between these two groups for the NRY. Moreover, our results highlight the importance of obtaining complete mtDNA genome sequences, as there are no significant differences in HD between the Besemah and Semende when only partial sequences are analysed, as was done in previous studies<sup>7,8</sup>.

## Results

**mtDNA sequences.** We obtained 36 complete mtDNA sequences from the Semende (a matrilineal group) and 36 from the Besemah (a patrilineal group) from Sumatra, Indonesia, using high throughput, parallel tagged sequencing<sup>9,10</sup> on the Roche GS/FLX (Roche) and Illumina GAII platforms (Illumina). Sequences were assigned to the closest haplogroup for which all defining mutations were present. For both groups, 24 mtDNA haplogroups were observed (Fig. 1), of which 7 belong to macrohaplogroup M (Supplementary Fig. S1) and 17 to macrohaplogroup N (Supplementary Fig. S1). The majority of the sequences from the Semende (44%) belong to the basal haplogroups M1'51 and a new haplogroup, M\*. For the Besemah, haplogroup M7c3c was at the highest frequency (31%), followed by haplogroup E1a1a at a frequency of 14% (Table 1).

The mtDNA HD for the complete mtDNA genomes (Table 2) is lower in the Semende than in the Besemah, and the difference is highly significant by a permutation test ( $P < 0.001$ ) (Supplementary Fig. S2). Conversely, the mean number of pairwise differences (Table 2) was significantly higher ( $P < 0.03$ , Supplementary Fig. S2) in the Semende ( $k = 36.18$ ) than in the Besemah ( $k = 32.34$ ). Furthermore, the  $F_{ST}$  value for the mtDNA is 0.076 and is significantly different from 0 ( $P = 0$ ). As a previous study that failed to find a difference in patterns of mtDNA diversity between matrilineal and patrilineal groups in India only sequenced the HV1 region<sup>7</sup>, we also analysed only the HV1 sequences from the Semende and Besemah. In contrast to the results based on complete mtDNA genome



**Figure 1 | mtDNA and NRY haplogroup frequencies for the Semende and the Besemah.** Haplogroup composition and frequencies based on complete mtDNA sequences and Y-SNPs.

**Table 1 | Haplogroup frequencies for mtDNA.**

Haplogroup	Semende	Besemah
M1'51	0.25	0
M*	0.19	0.06
M4	0.03	0.03
M7c3c	0.06	0.31
E1a1a1	0.03	0.03
E1a1a	0	0.14
B4a1a	0.11	0.03
B4a2a	0.06	0
B4c1b	0.03	0
B4c1b2	0.03	0.03
B4c2	0	0.03
B5a1	0.03	0.06
B5a1a	0	0.03
F1a1a	0.03	0
F1a'c	0.08	0.03
F1a	0	0.03
F1e	0.03	0
F3b1	0.03	0
N9a6	0.03	0
N9a6a	0	0.06
N22	0	0.03
N*	0	0.03
Y2a	0	0.08
E2a	0	0.03

mtDNA, mitochondrial DNA.

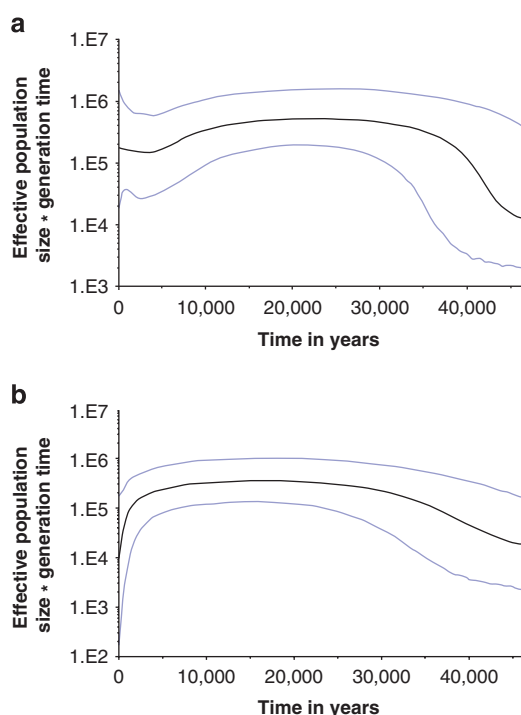
sequences, the difference in mtDNA HD is not significant between the Semende and Besemah (Supplementary Fig. S2), but the mean number of pairwise difference is significantly higher in the Semende than Besemah ( $P < 0.01$ ). Likewise, the  $F_{ST}$  value is 0.088 and is significantly different from 0 ( $P = 0$ ).

To further investigate the maternal history of the Semende and the Besemah, we carried out a Bayesian analysis of changes in population size through time<sup>11</sup>. The Bayesian Skyline Plots (BSPs), based on the complete mtDNA genome sequences, differ between the two groups: the Besemah exhibit a steep increase in population

**Table 2 | Summary statistics for the mtDNA and Y-chromosome.**

	<i>N</i>	<i>h</i>	HD	<i>k</i>
<i>Complete mtDNA genome</i>				
Besemah	36	29	0.984±0.01	32.34
Semende	36	20	0.919±0.03	36.18
<i>mtDNA HV1</i>				
Besemah	36	19	0.886±0.04	6.05
Semende	36	17	0.913±0.03	7.73
<i>Y-STR haplotypes</i>				
Besemah	37	24	0.93±0.04	5.52
Semende	37	23	0.95±0.02	5.65

HD, haplotype diversity; *h*, number of haplotypes; *k*, mean number of pairwise differences; mtDNA, mitochondrial DNA; *N*, sample size.



**Figure 2 | Bayesian Skyline Plots of effective population size through time.** BSP based on the mtDNA coding region, estimated with 30 million MCMC iterations and sampled every 3,000 steps. The y axis for each plot is the product of the effective population size and the generation time and the x axis shows time. A mutation rate of  $1.69 \times 10^{-8}$  per site per year<sup>49</sup> was used. (a) BSP for Besemah, using all 36 sequences. (b) BSP for Semende, using all 36 sequences.

size beginning around 40,000 years ago and a slight decrease around 10,000 years ago (Fig. 2a), whereas the Semende show a more gradual increase beginning around 40,000 years ago, and a sharp decrease beginning around 5,000 years ago (Fig. 2b). The BSPs also indicate that the current estimated effective population size of the Semende is about ten times lower than that of the Besemah.

**NRV variation.** Y-single-nucleotide polymorphisms (SNPs) and Y-STRs were typed for all individuals, however results could not be obtained for all STR loci for one Besemah, who therefore was excluded from the Y-STR analysis. The Y-SNP haplogroup for each individual

**Table 3 | Y-chromosome haplogroup frequencies.**

Haplogroup	Semende	Besemah
O2	0.73	0.71
O3	0.19	0.16
O*	0.03	0.05
O1a2	0.05	0.08

and the Y-STR haplotypes are provided in Supplementary Table S1. Only four NRY haplogroups were observed for both groups, and all individuals belonged to haplogroup O or sublineages thereof (Fig. 1). Both groups had high frequencies of O2 (O-P31) (>70%) and O3 (O-M122) (16–19%), whereas haplogroups O1a2 (O-M50) and O\* were found at low frequencies in both groups (Table 3). In contrast to the mtDNA results, the distribution of Y-SNP haplogroups is very similar in the two groups (Fig. 1) and does not differ significantly ( $P > 0.05$ ). Moreover, the  $F_{ST}$  value for Y-STRs between the Semende and Besemah is only 0.013, and is not significantly different from 0 ( $P > 0.05$ ). Network analysis showed that Y-STR haplotypes are also shared to a large extent between the two groups (Supplementary Fig. S3). Neither HD values nor mean number of pairwise differences (*k*), based on Y-STRs, differ significantly between the Semende (HD = 0.95, *k* = 5.65) and the Besemah (HD = 0.93, *k* = 5.52), based on permutation tests (Supplementary Fig. S2).

## Discussion

If residence pattern influences genetic diversity, then mtDNA HD is expected to be lower in matrilocal than patrilocal groups. This is indeed the case (Table 2): mtDNA diversity is significantly lower (as judged by a permutation test; Supplementary Fig. S2) in the matrilocal Semende than in the patrilocal Besemah. Interestingly, the HD of the HV1 in the Besemah is lower, but not significantly so. No other studies looking at genetic diversity differences between patrilocal and matrilocal groups have used complete mtDNA sequences before; most studies have used only part of the mtDNA genome, usually HV1. These results indicate that it may be insufficient to use only the HV1 to make inferences concerning genetic variation and differences. In particular, perhaps the failure of identifying such differences in previous studies is due to the lack of power using only the HV1 (ref. 7) or only a single gene such as *MT-CO3* (ref. 8).

The higher mean number of pairwise differences in the matrilocal group probably reflects the very different haplogroup composition of this group (Fig. 1 and Table 1): around 20% of mtDNA lineages in the Semende belong to a new haplogroup M\* restricted to this population and another 25% belong to a new subgroup of M1'51. This new subgroup of M1'51 shares 12 mutations with M51 (Supplementary Fig. S1), a recently described haplogroup found in one Cambodian individual<sup>12</sup>. M1'51 is basal to subgroups found in North and West Africa and South Europe and is believed to have arisen in southwestern Asia and to have been brought back to Africa and South Europe via a back-migration<sup>13</sup>. Except for the one Cambodian individual, no other subgroups of M1'51 have been found before in Asia to date. The rest of the mtDNA lineages in the Semende belong to 13 different haplogroups (Fig. 1). Altogether, 53% of the sequences belong to haplogroups frequently found in West and Southeast Asia, for example, subclades of haplogroups B4, B5, R9 and N9 (refs 14–16).

In the Besemah, the mtDNA lineages fall into 17 different haplogroups (Fig. 1 and Table 2); 95% of their mtDNA haplogroups have been previously described in West and Southeast Asia (with some variation at the tips of the branches), including subhaplogroups of N9, M7, F1, E1, E2, B4 and B5 (refs 14–16). Haplogroup M7c3c has the highest frequency (31%), followed by E1a1a (14%)

which are both widespread and found at high frequencies in Southeast Asia<sup>14,15,17</sup>. The Besemah also have one unique M\* lineage; one sample with the same haplotype as two M\* Semende sequences; one sequence that branches off haplogroup M4; and one N\* lineage that shares some mutations with N21 (Supplementary Fig. S1). The M4 lineage shares some mutations with the M4 lineage from the Semende, but each have several unique mutations. Subgroups of M4 have been previously reported in tribal populations in India<sup>18,19</sup>, Nepal<sup>20</sup> and in the Philippines<sup>21</sup>. There is thus a striking dichotomy in the mtDNA lineages between the matrilineal Semende and the patrilineal Besemah. The Semende have high frequencies of M\* and M1'51 lineages not found elsewhere in the world to date, which suggests that these lineages have been maintained in the matrilineal population, perhaps through matrilineal practices, for a long time. By contrast, the majority of the mtDNA lineages in the Besemah are found at high frequency in Southeast Asia and indicate that there has been substantial mtDNA gene flow between this group and surrounding groups, as expected in patrilineal societies.

To further investigate the observed differences in mtDNA diversity, we generated BSPs based on the coding region of the mtDNA genomes (Fig. 2). The BSPs indicate that the matrilineal group has a lower effective population size than the patrilineal group, as expected from their lower genetic diversity. Furthermore, the BSPs indicate different histories for these groups: the Besemah show signatures of population expansion, followed by a slight population reduction, whereas the population size has been relatively more constant for the Semende, with a recent steep population reduction.

Unexpectedly, patterns of NRY variation are very similar, and do not differ significantly, between the two groups. Haplogroup O2 has the highest frequency in both groups (>70%), and this haplogroup is found at high frequency in Southeast Asia<sup>22,23</sup>, and its subhaplogroup O2a (O-M95) in Indonesia<sup>24</sup>. Haplogroups O3 (O-M122) and O1a (O-M119), which are found at low frequency in both groups, have been associated with the Austronesian expansion and are found at high frequency throughout Southeast Asia<sup>22,23,25</sup>. These results are surprising, as other studies have shown that, in general, there is more structure within human populations for Y-chromosome diversity than for mtDNA, which is likely to reflect the high global prevalence of patrilocality<sup>1,5</sup>. Furthermore, patrilineal practices seem to be more tightly regulated than matrilineal practices<sup>4</sup>, resulting in a higher female than male migration rate<sup>2,26–28</sup>. Perhaps matrilocality has been more tightly regulated in the Semende, and patrilocality less tightly regulated in the Besemah, than has been observed previously.

The similarity in NRY diversity for these groups could also be explained by a recent conversion to patrilocality of the Besemah. The current matrilineal and patrilineal residence patterns of the Besemah and Semende are documented since the middle of the 19th century<sup>29–31</sup>, but it is unknown when they were first established. It has been hypothesized that matrilocality is ancestral in Austronesian societies and that descendant groups of Austronesian people in the Pacific adopted a patrilineal residence pattern over time, as a switch from matrilocality to patrilocality is more common than the reverse change<sup>32</sup>. A relatively recent change to patrilocality of the Besemah would explain the low frequency of unique mtDNA lineages as those would have been replaced by new, incoming lineages. The lack of unique NRY types can likewise be explained by a former practice of matrilocality for which in-marrying men would have continuously introduced new Y-chromosomes. The original residence pattern is expected to be reflected in patterns of genetic variation at least 5–6 generations after any switch<sup>28</sup>, but to have disappeared after about 20 generations<sup>33</sup>. Therefore, if the Besemah were previously a matrilineal group, the switch to patrilocality must have happened at least 150 years ago (assuming a generation time of 25 years for females), but not so long ago that there has been time for patterns of NRY variation to reflect the switch to patrilocality. However,

differences in resolution for mtDNA versus the Y-chromosome may also have a role, as we have more detailed information for mtDNA (the complete sequence, compared with a few Y-SNPs and Y-STRs).

In conclusion, it is highly likely that the unique M\* and M'51 mtDNA lineages present in the Semende reflect the initial settlement of the region, and that matrilocality has preserved these lineages. By contrast, patterns of Y-chromosome diversity do not differ between the Besemah and the Semende, suggesting that local groups were more heavily influenced by male gene flow from expanding populations. Notably, the significant differences in mtDNA HD between the Besemah and Semende were only revealed by the complete mtDNA genome sequences, and not by HV1 sequences alone. Thus, previous studies that analysed only a portion of the mtDNA genome and failed to find a difference relating to matrilocality versus patrilocality may have lacked sufficient resolution. Overall, our results confirm the idea that cultural practices can influence genetic variation<sup>34</sup>, but also demonstrate that the expected influence of matrilocality and patrilocality on genetic diversity may not always hold; in particular, in the present case, matrilocality seems more tightly regulated than patrilocality, in contrast to previous results<sup>4</sup>.

## Methods

**DNA samples.** Saliva samples were collected with informed consent by Hengky Firmansyah from nine locations in Sumatra, Indonesia (Supplementary Table S2), consisting of 38 samples from the Besemah (a patrilineal group) and 37 from the Semende (a matrilineal group). DNA was extracted as described previously<sup>35</sup>. These agricultural groups live in very close proximity to each other and are linguistically similar, speaking closely related dialects that are partially mutually intelligible (David Gil, field observation). All samples were collected in villages close to Pagaralam except one that was collected in Padang. The use of these samples in this research was approved by the Ethics Commission of the University of Leipzig Medical Faculty.

**MtDNA genome sequencing.** Complete mtDNA genome sequences were obtained for 36 samples from each group, 27 with the Roche GS/FLX platform (Roche) and 45 with the Illumina GAII platform (Illumina; Supplementary Table S2); coverage for three samples was too low for subsequent analysis and hence these were excluded. All libraries sequenced with the GS/FLX and 12 samples that were sequenced with the GAII were prepared from long-range PCR products. Two overlapping long-range PCR products were amplified for the sequencing of the complete mtDNA genome using primers described previously<sup>21</sup>. The libraries for the remaining 33 samples were prepared using a targeting method designed for the Genome Analyzer platform in which each individual is given its own barcode during the library preparation<sup>10</sup>. The samples were then enriched with a capture method in which mtDNA PCR products were used to capture library mtDNA templates<sup>36</sup>. These samples were sequenced on the GAII analyzer with single reads and 76 cycles (see Supplementary Table S2 for more details). Assembly of the sequences was carried out with a mapping iterative assembler as described previously<sup>37</sup>, using the revised Cambridge Reference Sequence (rCRS) as a reference to which all reads were mapped. A multiple alignment was performed with mafft v6.708b<sup>38</sup>. For the consensus sequences obtained from the mapping iterative assembler, all bases were covered at least two times (bases with <2× coverage were replaced with N's, as missing data; see Supplementary Table S2 for the number of N's in each sequence). A maximum of 1% missing data (N's) was accepted; the number of N's per sequence ranged from 0 to 26 (Supplementary Table S2). Overall, the average coverage was 54-fold, ranging from 9 to 144 with an average minimum coverage of 15.5 (Supplementary Fig. S4 and Supplementary Table S2). Sequences were manually checked and edited because of homopolymer problems occurring with the GS/FLX technology. This problem stems from the inaccuracy in the light signal intensity resulting from runs of three or more identical bases, making it impossible to detect the exact number of bases in such homopolymer regions<sup>39</sup>. Therefore, sequences were manually checked and edited and insertions or deletions were removed in a homopolymer run in a genic region, but not in non-coding regions. These edited positions never occurred at a polymorphic, biallelic site, and all indels were not used in subsequent analyses. These manually edited sequences have been submitted to GenBank (accession numbers: HM596644 to HM596715).

**NRY genotyping.** A total of 12 Y-SNPs (C-RPS4Y, C-M38, C-M208, M-M4, M-P34, M-M104, K-M9, NO-M214, O-M119, O-M122, P-M74, R-M173) were typed using a single-base extension assay with amplicons detected by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using methods described elsewhere<sup>40</sup>. For a higher resolution of specific Y-chromosome haplogroups, further Y-SNPs were detected by hierarchic multiplexes<sup>41</sup> and

genotyping performed with the ABI Prism SnaPshot multiplex kit (Applied Biosystems), with amplicons detected using capillary electrophoresis on an ABI Prism 3100 Genetic Analyzer according to the manufacturer's instructions. Hierarchical SNP typing was done in two SnaPshot multiplexes; in the first one six SNPs were typed (O-M175, M-M5, O-M122, O-P31, N-Lly22g and O-M134), whereas in the second one three SNPs were typed (O-119, O-M101 and O-M50). In addition, 12 Y-STR loci (DYS391, DYS389I, DYS439, DYS389II, DYS438, DYS437, DYS29, DYS392, DYS393, DYS390, DYS385a, DYS385b) were typed using the Promega PowerPlex Y system (Promega Corporation) with amplicons detected on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems), all following the manufacturer's instructions. The phylogenetic relationship of the complete set of SNPs typed in the study is shown in Supplementary Figure S5, following the nomenclature of they Y-chromosome phylogenetic tree<sup>42</sup>.

**Data analysis.** The mtDNA genome sequences were assigned to haplogroups according to Phylotree.org Build 7<sup>43</sup> using a custom Perl script. Positions 309.1C(C), 16182C, 16183C, 16193.1C(C) and 16519 were not used for haplogroup assignment as these are subject to highly recurrent mutations. Y-chromosome haplogroup affiliations were based on the YCC tree<sup>42</sup>.

Basic descriptive diversity statistics were calculated with dnaSP v5 for the complete mtDNA sequences. The Arlequin software package<sup>44</sup>, version 3.5 was used to calculate summary statistics for the NRY data. To test if the diversity values (HD and the mean number of pairwise differences) differed significantly between groups, a custom R script (Supplementary Software) was used to perform a permutation test in which the complete dataset was split randomly into two populations 1,000 times and the relevant diversity statistic was calculated each time, and then the difference between the two randomly generated groups was calculated and compared with the difference between the values obtained from the observed data. The same approach was used to test whether the mean number of pairwise differences (*k*) for the mtDNA data was significantly different between groups, using the function `dist.dna` from the R package APE<sup>45</sup>. As APE only deals with sequence data, a custom R script was used to do the same test based on the mean number of pairwise differences for the Y-STR data.

Before performing the permutation tests, all sites with indels and missing data (N's) were deleted, except for two indel sites, which were recoded as base substitutions as follows: the 9bp deletion in the intergenic region between the *MT-CO2* and lysine tRNA genes was coded as a transitional difference (9-bp deletion = T, absence of deletion = C); and the CA microsatellite beginning at position 520 (five repeats = A, four repeats = T, three repeats (only in one case) = G). In total, 132 sites were deleted, or 0.8%. As this dataset was used for the permutation test, all summary statistics were calculated using this dataset in dnaSP.

The mtDNA coding region (positions 577–16,023) was used to generate BSs using Markov chain Monte Carlo (MCMC) sampling in the program BEAST (version 5.1)<sup>46,47</sup>, using the same parameters as described previously<sup>21</sup>. Each run was analysed using the program Tracer for independence of parameter estimation and stability of MCMC chains<sup>47</sup>.

Network analyses<sup>48</sup> were carried out using version 4.516 of Network and version 1.1.0.7 of Network Publisher. Networks for Y-STR haplotypes used a weighting scheme based on Y-STR locus-specific mutation rates obtained from NIST (<http://www.cstl.nist.gov/biotech/strbase/>).

## References

- Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **20**, 278–280 (1998).
- Kayser, M. *et al.* Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am. J. Hum. Genet.* **72**, 281–302 (2003).
- Nasidze, I. *et al.* Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Ann. Hum. Genet.* **68**, 205–221 (2004).
- Hamilton, G., Stoneking, M. & Excoffier, L. Molecular analysis reveals tighter social regulation of immigration in patrilineal populations than in matrilineal populations. *Proc. Natl Acad. Sci. USA* **102**, 7476–7480 (2005).
- Langergraber, K. E. *et al.* The genetic signature of sex-biased migration in patrilineal chimpanzees and humans. *PLoS ONE* **2**, e973 (2007).
- Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T. & Stoneking, M. Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nat. Genet.* **29**, 20–21 (2001).
- Kumar, V. *et al.* Global patterns in human mitochondrial DNA and Y-chromosome variation caused by spatial instability of the local cultural processes. *PLoS Genet.* **2**, 420–424 (2006).
- Wilders, J. A., Kingan, S. B., Mobasher, Z., Pilkington, M. M. & Hammer, M. F. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nat. Genet.* **36**, 1238 (2004).
- Meyer, M., Stenzel, U. & Hofreiter, M. Parallel tagged sequencing on the 454 platform. *Nat. Protoc.* **3**, 267–278 (2008).
- Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb prot5448, doi:10.1016/pdb.prot5448 (2010).
- Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
- Hartmann, A. *et al.* Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum. Mutat.* **30**, 115–122 (2009).
- Olivieri, A. *et al.* The mtDNA legacy of the levantine early upper palaeolithic in Africa. *Science* **314**, 1767–1770 (2006).
- Hill, C. *et al.* Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol. Biol. Evol.* **23**, 2480–2491 (2006).
- Tabbada, K. A. *et al.* Philippine mitochondrial DNA diversity: a populated viaduct between Taiwan and Indonesia? *Mol. Biol. Evol.* **27**, 21–31 (2010).
- Soares, P. *et al.* Climate change and postglacial human dispersals in Southeast Asia. *Mol. Biol. Evol.* **25**, 1209–1218 (2008).
- Trejaut, J. A. *et al.* Traces of archaic mitochondrial lineages persist in austronesian-speaking Formosan populations (vol 3, pg e376, 2005). *PLoS Biol.* **3**, 1838 (2005).
- Thangaraj, K., Chaubey, G., Reddy, A. G., Singh, V. K. & Singh, L. Unique origin of Andaman Islanders: insight from autosomal loci. *J. Hum. Genet.* **51**, 800–804 (2006).
- Chandrasekar, A. *et al.* Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PLoS ONE* **4**, e7447 (2009).
- Fornarino, S. *et al.* Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. *BMC Evol. Biol.* **9**, 154 (2009).
- Gunnarsdottir, E. D., Li, M., Bauchet, M., Finstermeier, K. & Stoneking, M. High-throughput sequencing of complete human mtDNA genomes from the Philippines. *Genome Res.* **21**, 1–11 (2011).
- Stoneking, M. & Delfin, F. The human genetic history of East Asia: weaving a complex tapestry. *Curr. Biol.* **20**, R188–R193 (2010).
- Karafet, T. M. *et al.* Balinese Y-chromosome perspective on the peopling of Indonesia: genetic contributions from pre-neolithic hunter-gatherers, Austronesian farmers, and Indian traders. *Hum. Biol.* **77**, 93–114 (2005).
- Karafet, T. M. *et al.* Major East-West division underlies Y chromosome stratification across Indonesia. *Mol. Biol. Evol.* **27**, 1833–1844 (2010).
- Kayser, M. *et al.* Melanesian origin of Polynesian Y chromosomes. *Curr. Biol.* **10**, 1237–1246 (2000).
- Salem, A. H., Badr, F. M., Gaballah, M. F. & Paabo, S. The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. *Am. J. Hum. Genet.* **59**, 741–743 (1996).
- Destro-Bisol, G. *et al.* Variation of female and male lineages in sub-saharan populations: the importance of sociocultural factors. *Mol. Biol. Evol.* **21**, 1673–1682 (2004).
- Bolnick, D. A., Bolnick, D. I. & Smith, D. G. Asymmetric male and female genetic histories among native Americans from eastern North America. *Mol. Biol. Evol.* **23**, 2161–2174 (2006).
- Moyer, D. S. Cultural constraints on marriage: anti-exchange behaviour in nineteenth century South Sumatra. *Bijdr. Taal-Land-V.* **139**, 247–259 (1983).
- Bowen, J. R. in *Encyclopedia of World Cultures: East and Southeast Asia* (ed. Paul Hockings) (G.K. Hall, 1993).
- Smith, G. & Bouvier, H. in *Spontaneous Settlements in Indonesia* (eds Muriel Carras & Marc Pain) (Departemen Transmigrasi and the Centre National de la Recherche Scientifique, 1993).
- Jordan, F. M., Gray, R. D., Greenhill, S. J. & Mace, R. Matrilineal residence is ancestral in Austronesian societies. *Proc. R. Soc. B Biol. Sci.* **276**, 1957–1964 (2009).
- Chaix, R. *et al.* From social to genetic structures in central Asia. *Curr. Biol.* **17**, 43–48 (2007).
- Laland, K. N., Odling-Smee, J. & Myles, S. How culture shaped the human genome: bringing genetics and the human sciences together. *Nat. Rev. Genet.* **11**, 137–148 (2010).
- Quinque, D., Kittler, R., Kayser, M., Stoneking, M. & Nasidze, I. Evaluation of saliva as a source of human DNA for population and association studies. *Anal. Biochem.* **353**, 272–277 (2006).
- Maricic, T., Whitten, M. & Paabo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5**, e14004 (2010).
- Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318–321 (2009).
- Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64 (2009).
- Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008).
- Hughes, D. A. *et al.* Parallel selection on TRPV6 in human populations. *PLoS ONE* **3**, e1686 (2008).
- Delfin, F. *et al.* The Y-chromosome landscape of the Philippines: extensive heterogeneity and varying genetic affinities of Negrito and non-Negrito groups. *Eur. J. Hum. Genet.* **19**, 224–230 (2011).
- Karafet, T. M. *et al.* New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* **18**, 830–838 (2008).

43. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–394, doi: 10.1002/humu.20921 (2009).
44. Excoffier, L., Estoup, A. & Cornuet, J.-M. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**, 1727–1738 (2005).
45. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
46. Drummond, A. J., Nicholls, G. K., Rodrigo, A. G. & Solomon, W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**, 1307–1320 (2002).
47. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
48. Bandelt, H. J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
49. Atkinson, Q. D., Gray, R. D. & Drummond, A. J. mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. *Mol. Biol. Evol.* **25**, 468–474 (2008).

### Acknowledgments

We thank Hengky Firmansyah for sample collection, and we gratefully acknowledge all participants who provided samples for this study. We thank Karolin Meyer for assistance with laboratory work, Mark Whitten, Johannes Krause and Tomislav Maricic for help

with technical issues in the laboratory and Cesare deFilippo and David Hughes for their help with the permutation tests. Furthermore, we thank the MPI-EVA Molecular Anthropology Group for helpful discussions. This research was funded by the Max Planck Society.

### Author contributions

M.S. and B.P. initiated the project; D.G. and B.P. arranged the sampling; E.D.G., M.R.N. and S.M. conducted the experiments; M.L. processed the raw high throughput sequences; M.S. supervised the project and E.G. analysed the data and wrote the paper. All authors discussed the paper and gave comments.

### Additional information

**Data deposition:** The complete consensus mtDNA sequences were submitted to GenBank Nucleotide Core database under accession numbers HM596644 to HM596715.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Gunnarsdóttir, E.D. *et al.* Larger mitochondrial DNA than Y-chromosome differences between matrilineal and patrilineal groups from Sumatra. *Nat. Commun.* **2**:228 doi: 10.1038/ncomms1235 (2011).