

## Research Article

**Substitution of Hainan indigenous genetic lineage in the Utsat people, exiles of the Champa kingdom**

<sup>1</sup>Dong-Na LI<sup>†</sup> <sup>2</sup>Chuan-Chao WANG<sup>†</sup> <sup>1</sup>Kun YANG <sup>2</sup>Zhen-Dong QIN <sup>2</sup>Yan LU  
<sup>2</sup>Xue-Jing LIN <sup>2</sup>Hui LI\* the Genographic Consortium<sup>‡</sup>

<sup>1</sup>(Department of Biology, Hainan Medical College, Haikou 571199, China)

<sup>2</sup>(State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China)

**Abstract** The Utsat people do not belong to one of the recognized ethnic groups in Hainan, China. Some historical literature and linguistic classification confirm a close cultural relationship between the Utsat and Cham people; however, the genetic relationship between these two populations is not known. In the present study, we typed paternal Y chromosome and maternal mitochondrial (mt) DNA markers in 102 Utsat people to gain a better understanding of the genetic history of this population. High frequencies of the Y chromosome haplogroup O1a\*-M119 and mtDNA lineages D4, F2a, F1b, F1a1, B5a, M8a, M\*, D5, and B4a exhibit a pattern similar to that seen in neighboring indigenous populations. Cluster analyses (principal component analyses and networks) of the Utsat, Cham, and other ethnic groups in East Asia indicate that the Utsat are much closer to the Hainan indigenous ethnic groups than to the Cham and other mainland southeast Asian populations. These findings suggest that the origins of the Utsat likely involved massive assimilation of indigenous ethnic groups. During the assimilation process, the language of Utsat has been structurally changed to a tonal language; however, their Islamic beliefs may have helped to keep their culture and self-identification.

**Key words** genetic structure, Hainan, mitochondrial DNA, Utsat, Y chromosome.

The Utsat people form a relatively small population that lives mostly in Sanya Prefecture on the southern tip of Hainan Island off the southern coast of mainland China. This population is not considered as a separate ethnic group, and is officially recognized as being of the Hui nationality. This recognition is contentious because of its unclear origin. The folktales of the Utsat name their ancestors as Muslims who originated in Central Asia, as did other Hui Chinese. However, the Utsat are also thought to be descendants of the Champa Kingdom (7th to 18th centuries) who fled their homeland to escape Vietnamese invasion (Olson, 1998). According to oral historical material, a Cham Prince and approximately 1000 Cham moved to Hainan after the Vietnamese completed their conquest of Champa, and the Ming dynasty allowed them to set up a kingdom in exile in Hainan (Tran & Reid, 2006). However, Chinese historical literature records the arrival of the Cham refugees on Hainan even earlier,

in the Song dynasty, after the capital of Cham fell in 982 AD (Andaya, 2008).

The close relationship between the Utsat and Cham is supported by affiliations in their language. Tsat, the language of the Utsat, belongs to the Malayo-Polynesian group within the Austronesian linguistic family, as do the Cham languages (Pang & Maddieson, 1993; Graham, 2006). The Tsat language is very similar to that of the Northern Roglai and is classified as a Northern Cham branch (Lewis, 2009). However, Tsat exists as a “linguistic enclave” in Hainan Island because it is surrounded by non-Austronesian languages (i.e. Tai-Kadai and Chinese). Over the long course of contact with Sinitic dialects and the directionality of internal drift, Tsat has changed structurally to resemble Chinese and Hlai. For example, Tsat has developed into a solidly tonal language, which is seldom observed in Malayo-Polynesian languages (Graham, 1996).

In previous studies, we have found that East Asian languages show a strong association with paternal lineages of Y chromosomes (Wen et al., 2004a, 2004b; Li et al., 2008; Cai et al., 2011). Thus, the structural changes to the Tsat language may be also reflected genetically. Li et al. (2008) reported the Y chromosome data for 31 Utsat. The high frequency of haplogroup

Received: 24 October 2012 Accepted: 5 December 2012

<sup>†</sup>These authors contributed equally to this study.

<sup>‡</sup>Members of the Genographic Consortium are listed in the Appendix.

\* Author for correspondence. E-mail: lhca@fudan.edu.cn. Tel.: 86-21-55664574. Fax: 86-21-55664885.

O1a-M119 (58.1%) and the connection to Daic populations in the Y-short tandem repeat (STR) networks suggests a probable Daic genetic background for the Utsat (Li et al., 2008). These findings suggest that the origin of the Utsat most likely involves assimilation or recent gene flow from neighboring indigenous populations. However, Y chromosome data only offer a paternal perspective and the small sample size in the study of Li et al. (2008) may have contributed to bias. Furthermore, without comparisons to the Cham population, the situation remains contentious. In the present study, we addressed this question by typing the maternally inherited mitochondrial (mt) DNA and relevant Y chromosome markers in 102 samples from the Utsat population (72 men and 30 women) to gain a better understanding of the origin of the Utsat.

## 1 Material and methods

### 1.1 Population samples

The present study was approved by the Ethics Committee of the Fudan School of Life Sciences. Peripheral blood samples were collected from 102 Utsat individuals from Sanya Prefecture, Hainan, China, after adequate information had been provided regarding the study and the subjects had provided written consent. All subjects were all healthy and not related within five generations.

### 1.2 Y chromosome markers

Fourteen single nucleotide polymorphisms (SNPs) in the Y chromosome non-recombining portion as listed in the latest Y chromosome phylogenetic tree (Karafet et al., 2008; Yan et al., 2011; M130, M89, M9, M45, M119, M110, M101, P31, M95, M88, M122, M164, M159, and M7) were typed in the samples collected using polymerase chain reaction (PCR)-restriction fragment length polymorphism (RFLP). Four of SNPs (M48, M8, M217, and M356) were typed using Taqman (Applied Biosystems, Foster City, CA, USA), whereas seven SNPs (YAP, M15, M175, M111, M134, M117, and M121) and seven STR polymorphisms (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393) were typed using fluorescently labeled primers for PCR amplification. Denatured products were separated by acrylamide gel electrophoresis on a 3730xl Genetic Analyzer (Applied Biosystems, Carlsbad, CA, USA) to distinguish the alleles.

### 1.3 Mitochondrial DNA markers

The hypervariable segment I (HVS-I) region of mtDNA was amplified using the primers L15974 and

R16488 (Yao et al., 2002). The PCR products were purified by Shrimp Alkali Enzyme and ExonI (Roche Diagnostics, Shanghai, China). The purified PCR product was sequenced using the Big-Dye Terminator Cycle Sequencing Kit (Applied Biosystems) and an ABI 3730xl Genetic Analyzer (Applied Biosystems). Sequence Analysis 3.3 software (Applied Biosystems) was used to extract the sequences. The HVS-I sequences were edited and aligned against the revised Cambridge reference sequence (Andrews et al., 1999) using DNASTAR software (DNASTAR, Madison, WI, USA). Another 22 polymorphisms in the coding regions of the mtDNA (3010, 7598, 663, 10 400, 10 310, 4216, 4491, 12 308, 10 646, 11 719, 4715, 4833, 8271, 5301, 70 287, 13 263, 14 569, 5417, 5178, 12 705, 15 607, and 9824) were also genotyped hierarchically using SNaPshot (ABI SNaPshot Multiplex Kit; Applied Biosystems). The PCR products were also electrophoresed on the 3730xl Genetic Analyzer (Applied Biosystems). The haplogroup affiliation of each mtDNA sequence was inferred by combined use of the HVS-I motif and diagnostic polymorphisms in the coding regions (Kivisild et al., 2002; Kong et al., 2003).

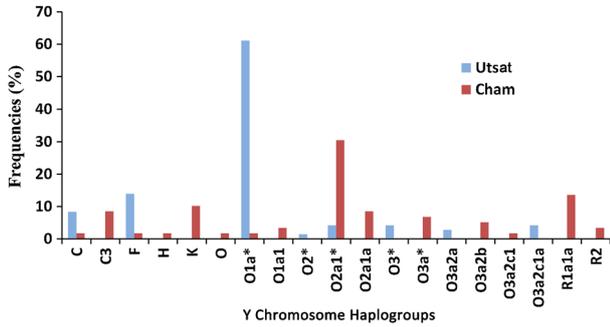
### 1.4 Statistical analyses

Networks of Y chromosome STRs and mtDNA HVS-I motifs were constructed by the median joining method (Bandelt et al., 1999) using Network version 4.5.1.0 (<http://www.Fluxus-engineering.com>, accessed 10 October 2012). Genotype data for the Utsat were generated in the present study, whereas data pertaining to neighboring populations were obtained from the existing literature (Yao et al., 2002; Yao & Zhang, 2002; Wen et al., 2004a, 2004b; Trejaut et al., 2005; Li et al., 2007, 2008, 2010; Gan et al., 2008; Qin et al., 2010; He et al., 2012; Lu et al., 2012). Arlequin 3.11 was used to calculate the Y-STRs Slatkin's linearized fixation index ( $F_{st}$ ) ( $R_{st}$ ) genetic distances (Excoffier et al., 2005). Principal components analysis (PCA) and multidimensional scaling (MDS) were performed using SPSS 18.0 software (SPSS, Chicago, IL, USA).

## 2 Results and Discussion

### 2.1 Y chromosome

According to the nomenclature of the Y Chromosome Consortium (YCC; Karafet et al., 2008; Yan et al., 2011), eight SNP haplogroups were determined from the 72 individual Utsat samples (see Table S1 available as Supplementary Material to this paper). Although the language of the Utsat people is now classified as Northern Cham, their paternal genetic

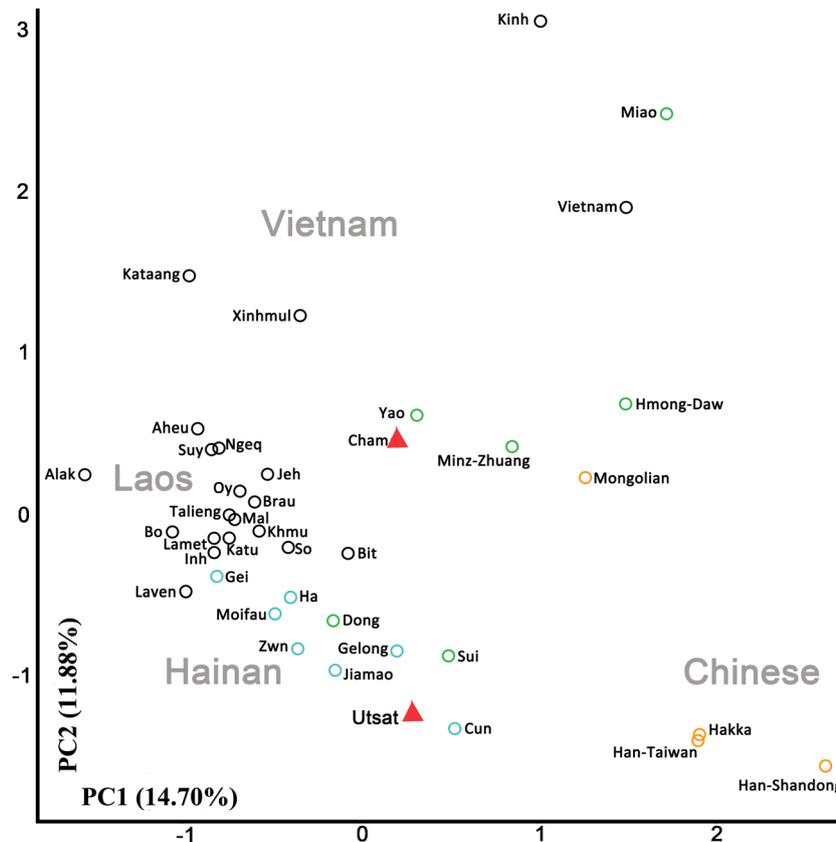


**Fig. 1.** Frequencies of Y chromosome haplogroups for the Utsat and Cham.

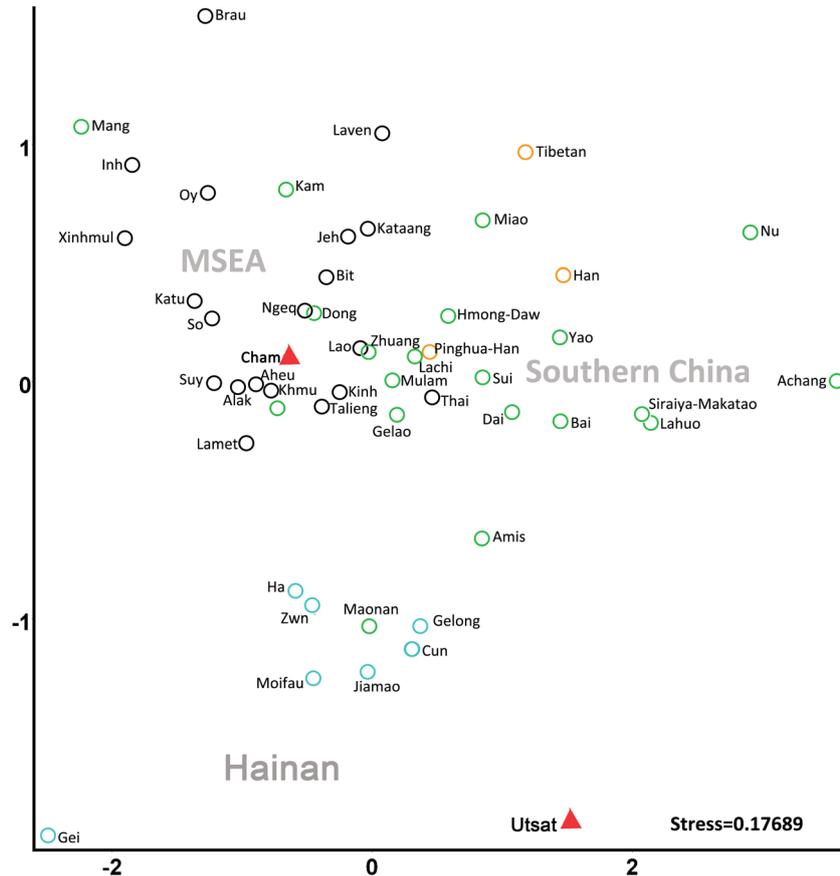
structure, with high frequencies of haplogroup O1a\*-M119, is not similar to that of the Cham populations. O2a1\* and its sub-haplogroup O2a1a, which are dominant haplogroups of the Cham, only account for 4.17% of the Utsat (Fig. 1). The moderate frequencies of ancient southeast Asian lineages C-M130 and F\*-M89 among the Utsat may result from the genetic drift of certain ancestral contributors to the population. One

of the Sino-Tibetan characteristic lineages, namely O3a2c1a-M117, has also been identified in the Utsat at a low frequency of 4.17% (Su et al., 2000; Kang et al., 2012), and probably resulted from recent gene flow from the neighboring Han migrants.

Detailed paternal genetic patterns for the Utsat, Cham, and other East Asian populations were discerned using additional published Y chromosome datasets. We performed a PCA using the Y chromosome haplogroup frequencies of the Utsat and another 43 East Asian populations (Fig. 2). Populations from mainland southeast Asia (MSEA) and from Hainan formed two clusters in the second principal component (PC). The Utsat was clustered into the Hainan group, together with the Hainan aborigines and populations of southern China, such as the Dong and Sui. However, Cham was very closely related to the MSEA group. The MDS plot of 52 populations with  $R_{st}$  genetic distances based on six common Y-STRs (i.e. *DYS19*, *DYS389I*, *DYS390*, *DYS391*, *DYS392*, and *DYS393*) also associated the Utsat with populations from Hainan (Fig. 3). This pattern was due mainly to the high frequency of the

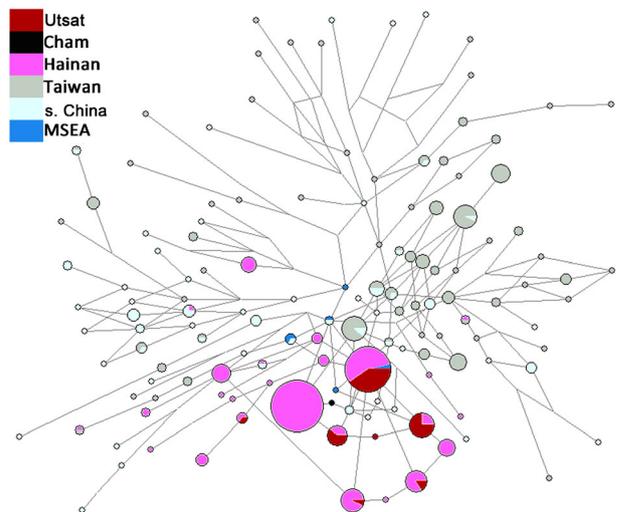


**Fig. 2.** Principal components analysis plot based on Y chromosome haplogroup frequencies of 44 populations. PC1, first principal component; PC2, second principal component.



**Fig. 3.** Multidimensional scaling plot of 52 populations with Slatkin’s linearized  $F_{st}$  ( $R_{st}$ ) genetic distances based on six common Y single tandem repeats (STR; i.e. DYS19, DYS389I, DYS390, DYS391, DYS392, and DYS393).

O1a\*-M119 haplogroup and the low frequency of O2a1\*-M95 in the Utsat. Detailed distances among the Utsat, Hainan aborigines, and MSEA populations within the O1a\*-M119 haplogroup could reveal a clear paternal genetic origin of the Utsat. Therefore, we constructed a median joining network based on six STR haplotypes (i.e. DYS19–DYS389I–DYS390–DYS391–DYS392–DYS393) of O1a\*-M119 individuals in the relevant ethnic groups. In the network (Fig. 4), Hainan aborigines formed several almost exclusive clades with few individuals from other populations, suggesting that Hainan aborigines had been isolated from other Daic populations of southern China and Taiwan aborigines. Almost all the Utsat samples were clustered in those exclusive clades with Hainan aborigines. However, samples from MSEA tended to cluster with southern Chinese populations. These findings suggest that the main paternal haplogroup of the Utsat was introduced from indigenous ethnic groups in Hainan rather than from the Cham or other MSEA populations.



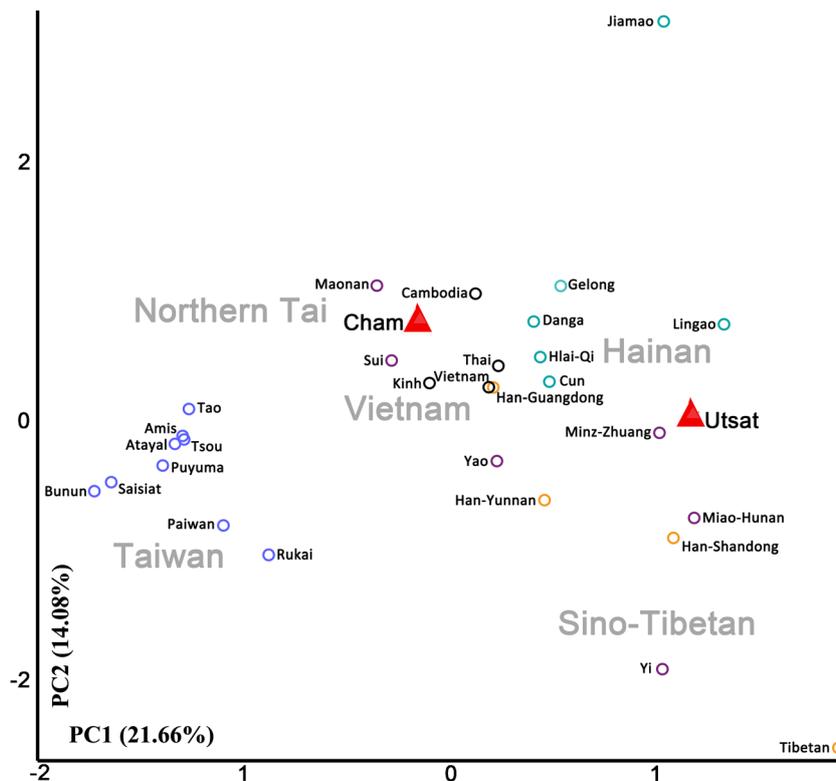
**Fig. 4.** Median joining Y single tandem repeat (STR) network of the O1a\*-M119 haplogroup based on six STR haplotypes (i.e. DYS19–DYS389I–DYS390–DYS391–DYS392–DYS393). The length of the lines between nodes is proportional to the mutation steps. MSEA, mainland southeast Asia; s. China, south China.

## 2.2 Mitochondrial DNA

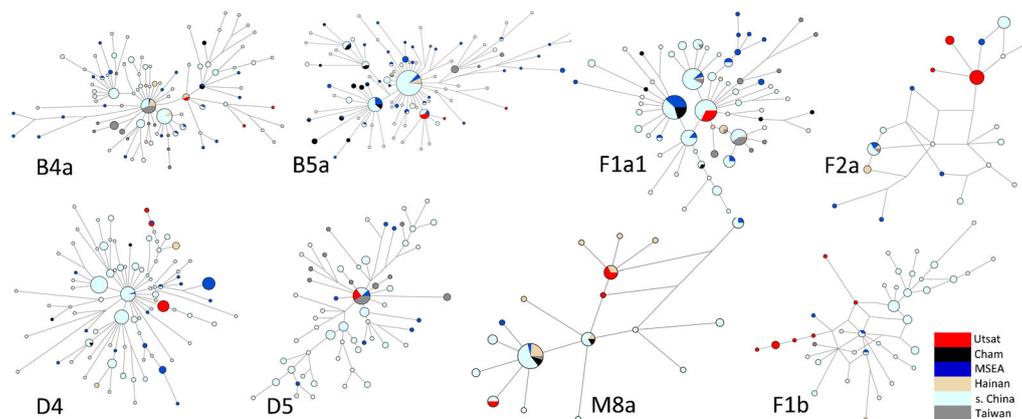
Nineteen mtDNA haplogroups were found in the 102 Utsat samples (Table S2). The mtDNA haplogroups present at high frequencies in the Utsat are (in decreasing order of frequency) D4, F2a, F1b, F1a1, B5a, M8a, M\*, D5, and B4a. The D4 and F2a haplogroups are the two main haplogroups in the Utsat population, accounting for 16.67% and 15.69%, respectively. However, these two haplogroups are absent or occur at very low frequencies in other Hainan aborigines and MSEA populations. We then compared Utsat samples with those two haplogroups at the haplotype level with other relevant populations. Most of the D4 samples from the Utsat have the same HVS-I motif type, namely 16223–16316–16362. However, this haplotype is seldom seen in the populations of East Asia and MSEA. The F2a haplotype in the Utsat is shared exclusively with some Han Chinese and other small ethnic groups (Lahu, Yi, and Mosuo) in Yunnan (Yao & Zhang, 2002; Wen et al., 2004a, 2004b; Lu et al., 2012; Wang et al., 2012). The pattern of high frequencies of B and F haplogroups in the Utsat is very similar to the neighboring populations and other southern populations. Furthermore, we used PCA based on the distribution of mtDNA haplogroup

frequencies to show the overall clustering pattern of the Utsat and another 30 populations (Fig. 5). Taiwan aborigines, MSEA populations, and Hainan aborigines formed three clusters in the first PC. Haplogroups E, F5, and B4 contributed mostly to the Taiwan pole. In contrast, haplogroups G, A, C, M9, and M8 were found to contribute most to the Hainan and Sino-Tibetan pole. The Utsat tended to cluster with Hainan aborigines, whereas the Cham clustered with the MSEA populations. This frequency distribution pattern revealed the genetic affinity between the Utsat and Hainan aborigines.

However, the results based on haplogroup frequency comparisons could be misleading because of the quickly changing frequencies of the mtDNA lineages caused by positive selection or genetic drift (Yang et al., 2011; Lu et al., 2012). A network analysis of individual lineages will most likely offer a better investigation of maternal relationships among the Utsat, Cham, and other populations (Li et al., 2007; Qin et al., 2010). Based on the mtDNA HVS-I motif and SNP-determined haplogroups, the networks of mtDNA haplogroups D4, F2a, F1b, F1a1, M8a, D5, and B4a were analyzed (Fig. 6). Each of these haplogroups has a high or moderate frequency in both the Utsat and



**Fig. 5.** Principal components analysis plot based on mitochondrial DNA haplogroup frequencies in 31 populations. PC1, first principal component; PC2, second principal component.



**Fig. 6.** Mitochondrial DNA hypervariable segment 1 haplotype networks. The length of the lines between nodes is proportional to the mutation steps. MSEA, mainland southeast Asia; s. China, south China.

MSEA populations, with these haplogroups together accounting for 72.55% of the Utsat. In the networks of the D4 haplogroup, only one haplotype was shared between the Utsat and Thai people, with other Utsat samples forming a large, exclusive clade. Exclusive clades were also seen for F2a and F1b. In the networks of B4a and M8a, the Utsat were clustered only with samples of Hainan aborigines. In the F1a1 network, the Utsat were clustered only with populations from southern China. In the B5a and D5 networks, the Utsat, Hainan aborigines, and populations from southern China, Taiwan, and MSEA clustered together in a large clade and in one moderate-sized clade. However, no Utsat samples clustered directly with the Cham. Overall, the maternal lineages of the Utsat are closer to those of the Hainan or southern China ethnic groups than to the MSEA populations.

In the present study, the patterns of paternal Y chromosome and maternal mtDNA diversities indicate that the Utsat are much closer to the Hainan indigenous ethnic groups than to the Cham and other MSEA populations. This suggests that the formation of the Utsat likely involved massive assimilation of indigenous ethnic groups. During the assimilation process, the language of the Utsat has been changed structurally so that it resembles the Sinitic or Tai-Kadai dialects. However, it is most interesting that the culture and self-identification of the Utsat remain consistent with the Cham. The Utsat are Muslims and the religious law of Islam deals with many personal matters, such as hygiene, diet, fasting, and even the exact time for prayer. These Islamic beliefs may have a significant role in maintaining the lifestyle and self-identification of the Utsat. Islam for the Utsat is more about community and probably not about biological descent. We may call it a

“religion dominance” mechanism of genetic replacement: a small invading group is adopted by the larger indigenous resident population, allowing for genetic substitution from the indigenous population to occur. However, the religion of the small invading population may help maintain their religion-based culture and self-identification.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (Grant Nos. 31071098, 30860124, 30890034, 91131002), National Excellent Youth Science Foundation of China (Grant No. 31222030), Shanghai Rising-Star Program (Grant No. 12QA1400300), Shanghai Commission of Education Research Innovation Key Project (Grant No. 11zz04), and Shanghai Professional Development Funding (Grant No. 2010001).

## References

- Andaya LY. 2008. Leaves of the same tree: Trade and ethnicity in the Straits of Melaka. Honolulu: University of Hawaii Press.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics* 23: 147.
- Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37–48.
- Cai X, Qin Z, Wen B, Xu S, Wang Y, Lu Y, Wei L, Wang C, Li S, Huang X, Jin L, Li H. Genographic Consortium. 2011. Human migration through bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum revealed by Y chromosomes. *PLoS One* 6: e24282.

- Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47–50.
- Gan RJ, Pan SL, Mustavich LF, Qin ZD, Cai XY, Qian J, Liu CW, Peng JH, Li SL, Xu JS, Jin L, Li H. 2008. Pinghua population as an exception of Han Chinese's coherent genetic structure. *Journal of Human Genetics* 53: 303–313.
- Graham T. 1996. Language contact and the directionality of internal drift: The development of tones and registers in Chamic. *Language* 72: 1–31.
- Graham T. 2006. Sociolinguistics and contact-induced language change: Hainan Cham, Anong, and Phan Rang Cham. In: Tenth International Conference on Austronesian Linguistics, 17–20 January 2006. Palawan: Linguistic Society of the Philippines and SIL International.
- He JD, Peng MS, Quang HH, Dang KP, Trieu AV, Wu SF, Jin JQ, Murphy RW, Yao YG, Zhang YP. 2012. Patrilineal perspective on the Austronesian diffusion in Mainland Southeast Asia. *PLoS One* 7: e36437.
- Kang L, Lu Y, Wang C, Hu K, Chen F, Liu K, Li S, Jin L, Li H. Genographic Consortium. 2012. Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the eastern Himalayas. *Annals of Human Genetics* 76: 92–99.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research* 18: 830–838.
- Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. *Molecular Biology and Evolution* 19: 1737–1751.
- Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, Zhang YP. 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *American Journal of Human Genetics* 73: 671–676.
- Lewis MP. 2009. *Ethnologue: Languages of the world*. 6th ed. Dallas: SIL International.
- Li D, Li H, Ou C, Lu Y, Sun Y, Yang B, Qin Z, Zhou Z, Li S, Jin L. 2008. Paternal genetic structure of Hainan aborigines isolated at the entrance to East Asia. *PLoS One* 3: e2168.
- Li D, Sun Y, Lu Y, Mustavich LF, Ou C, Zhou Z, Li S, Jin L, Li H. 2010. Genetic origin of Kadai-speaking Gelong people on Hainan island viewed from Y chromosomes. *Journal of Human Genetics* 55: 462–468.
- Li H, Cai X, Winograd-Cort ER, Wen B, Cheng X, Qin Z, Liu W, Liu Y, Pan S, Qian J, Tan CC, Jin L. 2007. Mitochondrial DNA diversity and population differentiation in southern East Asia. *American Journal of Physical Anthropology* 134: 481–488.
- Li H, Wen B, Chen SJ, Su B, Pramoongjago P, Liu Y, Pan S, Qin Z, Liu W, Cheng X, Yang N, Li X, Tran D, Lu D, Hsu MT, Dekka R, Marzuki S, Tan CC, Jin L. 2008. Paternal genetic affinity between Western Austronesians and Daic populations. *BMC Evolutionary Biology* 8: 146.
- Lu Y, Wang C, Qin Z, Wen B, Farina SE, Jin L, Li H. Genographic Consortium. 2012. Mitochondrial origin of the matrilineal Mosuo people in China. *Mitochondrial DNA* 23: 13–119.
- Olson JS. 1998. *An ethnohistorical dictionary of China*. Westport: Greenwood Publishing Group.
- Pang KF, Maddieson I. 1993. Tone in Utsat. In: Edmondson JA, Gregerson KJ eds. *Tonality in Austronesian languages*. Oceanic Linguistics Special Publication 24. Honolulu: University of Hawaii Press. 75–89.
- Qin Z, Yang Y, Kang L, Yan S, Cho K, Cai X, Lu Y, Zheng H, Zhu D, Fei D, Li S, Jin L, Li H. 2010. A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. *American Journal of Physical Anthropology* 143: 555–569.
- Su B, Xiao C, Dekka R, Seielstad MT, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza L, Chakraborty R, Jin L. 2000. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Human Genetics* 107: 582–590.
- Tran NT, Reid A. 2006. *Viet Nam: Borderless histories*. Wisconsin: University of Wisconsin Press.
- Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, Lee ZY, Lin M. 2005. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biology* 3: e247.
- Wang D, Su LY, Zhang AM, Li YY, Li XA, Chen LL, Long H, Yao YG. 2012. Mitochondrial DNA copy number, but not haplogroup, confers a genetic susceptibility to leprosy in Han Chinese from Southwest China. *PLoS One* 7: e38848.
- Wen B, Li H, Lu D, Song X, Zhang F, He Y, Li F, Gao Y, Mao X, Zhang L, Qian J, Tan J, Jin J, Huang W, Dekka R, Su B, Chakraborty R, Jin L. 2004a. Genetic evidence supports demic diffusion of Han culture. *Nature* 431: 302–305.
- Wen B, Xie X, Gao S, Li H, Shi H, Song X, Qian T, Xiao C, Jin J, Su B, Lu D, Chakraborty R, Jin L. 2004b. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *American Journal of Human Genetics* 74: 856–865.
- Yan S, Wang C-C, Li H, Li S-L, Jin L. Genographic Consortium. 2011. An updated tree of Y-chromosome haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *European Journal of Human Genetics* 19: 1013–1015.
- Yang K, Zheng H, Qin Z, Lu Y, Farina SE, Li S, Jin L, Li D, Li H. Genographic Consortium. 2011. Positive selection on mitochondrial M7 lineages among the Gelong people in Hainan. *Journal of Human Genetics* 56: 253–256.
- Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *American Journal of Human Genetics* 70: 635–651.
- Yao YG, Zhang YP. 2002. Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: New data and a reappraisal. *Journal of Human Genetics* 47: 311–318.

### Supplementary Material

The following supplementary materials are available online for this article at <http://onlinelibrary.wiley.com/doi/10.1111/jse.12000/supinfo>:

**Table S1.** Y chromosome single nucleotide polymorphisms and single tandem repeat (STR) data for the Utsat.

**Table S2.** Mitochondrial DNA haplogroups and the hypervariable segment I motif for the Utsat.

### Appendix I Members of the Genographic Consortium

Janet S. ZIEGLE (Applied Biosystems, Foster City, CA, USA)

Li JIN (Fudan University, Shanghai, China)

Hui LI (Fudan University, Shanghai, China)

Shilin LI (Fudan University, Shanghai, China)

Pandikumar SWAMIKRISHNAN (IBM, Somers, NY, USA)

Laxmi PARIDA (IBM, Yorktown Heights, NY, USA)

Daniel E. PLATT (IBM, Yorktown Heights, NY, USA)

Ajay K. ROYYURU (IBM, Yorktown Heights, NY, USA)

Lluis QUINTANA-MURCI (Institut Pasteur, Paris, France)

R. John MITCHELL (La Trobe University, Melbourne, Australia)

Marc HABER (Lebanese American University, Beirut, Lebanon)

Pierre A. ZALLOUA (Lebanese American University, Beirut, Lebanon)

Syama ADHIKARLA (Madurai Kamaraj University, Madurai, India)

ArunKumar GANESHPRASAD (Madurai Kamaraj University, Madurai, India)

Ramasamy PITCHAPPAN (Madurai Kamaraj University, Madurai, India)

Varatharajan Santhakumari ARUN (Madurai Kamaraj University, Madurai, India)

R. Spencer WELLS (National Geographic Society, Washington DC, USA)

Himla SOODYALL (National Health Laboratory Service, Johannesburg, South Africa)

Elena BALANOVSKA (Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia)

Oleg BALANOVSKY (Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia)

Chris TYLER-SMITH (The Wellcome Trust Sanger Institute, Hinxton, UK)

Daniela R. LACERDA (Universidade Federal de Minas Gerais, Belo Horizonte, Brazil)

Fabrcio R. SANTOS (Universidade Federal de Minas Gerais, Belo Horizonte, Brazil)

Pedro Paulo VIEIRA (Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil)

Jaume BERTRANPETIT (Universitat Pompeu Fabra, Barcelona, Spain)

David COMAS (Universitat Pompeu Fabra, Barcelona, Spain)

Begoña MARTÍNEZ-CRUZ (Universitat Pompeu Fabra, Barcelona, Spain)

David F. SORIA-HERNANZ (Universitat Pompeu Fabra, Barcelona, Spain)

Christina J. ADLER (University of Adelaide, Adelaide, Australia)

Alan COOPER (University of Adelaide, Adelaide, Australia)

Clio S. I. Der SARKISSIAN (University of Adelaide, Adelaide, Australia)

Wolfgang HAAK (University of Adelaide, Adelaide, Australia)

Matthew E. KAPLAN (University of Arizona, Tucson, AZ, USA)

Nirav C. MERCHANT (University of Arizona, Tucson, AZ, USA)

Colin RENFREW (University of Cambridge, Cambridge, UK)

Andrew C. CLARKE (University of Otago, Dunedin, New Zealand)

Elizabeth A. MATISOO-SMITH (University of Otago, Dunedin, New Zealand)

Matthew C. DULIK (University of Pennsylvania, Philadelphia, PA, USA)

Jill B. GAIESKI (University of Pennsylvania, Philadelphia, PA, USA)

Amanda C. OWINGS (University of Pennsylvania, Philadelphia, PA, USA)

Theodore G. SCHURR (University of Pennsylvania, Philadelphia, PA, USA)

Miguel G. VILAR (University of Pennsylvania, Philadelphia, PA, USA)