

## Research Article

## Three phases for the early peopling of Hainan Island viewed from mitochondrial DNA

<sup>1</sup>Dong-Na LI<sup>†</sup> <sup>2</sup>Chuan-Chao WANG<sup>†</sup> <sup>2</sup>Yan LU <sup>2</sup>Zhen-Dong QIN <sup>1</sup>Kun YANG  
<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** Hainan, an island linking mainland East Asia and Southeast Asia, lay in one of the routes of early migration to East Asia. The largest indigenous group of Hainan is called Hlai, possibly direct descendants of the earliest migrants. However, there are no sufficient genetic data to assess the population history of Hainan Island. Here, we have analyzed mitochondrial DNA control-region and coding-region sequence variations in 566 Hlai individuals from all five subgroups, Ha, Gei, Zwn, Moifau, and Jiamao. Our results suggest three phases for the peopling of Hainan. The first phase represents the initial settlement of the island as part of the African dispersal approximately 50 000 years ago. The second phase reflects colonization events from mainland Asia before the Last Glacial Maximum, which was recorded by wide distributed lineages, such as F\*, B4a, and D4a. The third phase reflects population expansions under lineages F1b, M7b, and R9b after the Last Glacial Maximum and Neolithic migrations in and out of Hainan Island. Selection also started to play a role during the last phase. We also detected different distributions between paternal Y chromosome and maternal mitochondrial DNA among isolated Hlai populations, which might be caused by sex-biased cultural practices.

**Key words** Hainan, Hlai, mtDNA, population expansion, selection, sex-biased cultural practices.

Hainan Island is located at the southernmost tip of China. It is geographically unique, in that it sits between East Asia and Southeast Asia. Therefore, populations may show characters of both regions. During the last Ice Age, when the sea level was much lower than present, Hainan Island was connected to the mainland (Shi et al., 1989; Xing, 2004), and was in one of the main routes of modern human migration from Southeast Asia to East Asia.

The ethnic groups of Hainan include Han Chinese, Hlai, Miao (here called Kimmun), Lingao, Gelong, and Utsat. Hlai people, distributed in the mountains of central and southern Hainan, are probably the earliest settlers of the island. Nowadays, Hlai is the largest indigenous population on the island with a population of more than 1.2 million (2000 census), comprising 14.7% of the whole population in Hainan. The language of Hlai people belongs to the most primordial branches of the Tai-Kadai linguistic family (Thurgood, 1994;

Wang et al., 2004). According to linguistic and cultural diversity, Hlai can be divided into five subgroups, Ha, Gei, Zwn, Moifau, and Jiamao.

Knowledge of the history of Hlai populations is limited as very few archaeological studies have been carried out in Hainan. However, genetic studies allow an attempt to reveal the origin of Hlai and the peopling of Hainan Island. From the paternal side, analyses of Y chromosomal diversity of Hlai populations resolved that haplogroups O1a\* and O2a\* are dominant among Hainan aborigines. The frequency of the mainland East Asia dominant haplogroup O3 is quite low, suggesting that the Hainan aborigines have been isolated from the mainland for approximately 20 000 years (Li et al., 2008, 2010). From the maternal side, high frequencies of mitochondrial DNA (mtDNA) lineages B4, M\*, and R9b was observed in 34 Hlai-Qi and 27 Jiamao individuals in Hainan. Most Hlai haplotypes were shared with Daic populations from southern China (Li et al., 2007). The genetic affinity between Hlai and populations of southern China has also been supported by the study of Peng et al. (2011). This study analyzed mtDNA variations of hypervariable sequence I and II (HVS-I and II) as well as partial coding regions of 285 Hlai individuals. They found that Hainan islanders

Received: 21 December 2012 Accepted: 10 April 2013

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

<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.

showed a close relationship with the populations in mainland southern China, especially from Guangxi, suggesting that the recent gene flow from the mainland might play important roles in shaping the maternal pool of Hainan islanders (Peng et al., 2011). However, because of the relatively poor representation of Hlai subgroups in the sample collection, the genetic structure of the Hlai populations remains poorly understood. No research on the mtDNA variations of the Hlai populations has been reported elsewhere. Therefore, we analyzed mtDNA variations of 566 individuals collected from all five Hlai populations, Gei, Ha, Jiamao, Moifau, and Zwn, to uncover more clues to the origin of the Hlai people.

## 1 Material and methods

### 1.1 Population samples

We collected blood samples of 566 healthy and unrelated individuals from five Hlai populations (Gei, Ha, Jiamao, Moifau, and Zwn) in Hainan province, China. All individuals signed informed consent. Genomic DNA was extracted through the classic phenol–chloroform method.

### 1.2 Mitochondrial DNA markers

The HVS-I of the control region was amplified by primers L15974 and H16488 (Yao et al., 2002). Purified polymerase chain reaction products were sequenced using the BigDye terminator cycle sequencing kit and an ABI 3130XL genetic analyzer (Applied Biosystems, Foster City, CA, USA). A SNaPshot assay was used for typing single nucleotide polymorphisms (SNPs) in the coding regions to confirm haplogroup identity. This assay was designed as a multiplex panel including 21 coding region SNPs and one length variation marker (Qin et al., 2010). Both the HVS-1 motif and the coding region variations were used to infer haplogroups according to the classification of Kivisild et al. (2002).

### 1.3 Statistical analyses

Molecular diversity, population structure estimates, and genetic distances between populations were calculated using Arlequin version 3.11 (Excoffier & Lischer, 2010). Classical frequency spectrum tests, such as Tajima's  $D$ , Fu & Li's  $D$ ,  $D^*$ ,  $F$ , and  $F^*$ , were calculated using DnaSP5.0 to detect deviation from neutrality (Tajima, 1989; Fu & Li, 1993; Fu, 1997; Librado & Rozas, 2009). Neighbor-joining unrooted trees based on  $F_{ST}$  statistics were carried out using MEGA 5.1 (Saitou & Nei, 1987; Tamura et al., 2011) to

visualize relationships among Hlai populations. Networks of the mtDNA HVS-1 motifs were constructed by the median-joining method (Bandelt et al., 1999) using Network version 4.5.1.6 (Fluxus-engineering.com). Coalescence times of haplogroups of interest were calculated by the  $\rho$  statistic method (Forster et al., 1996; Saillard et al., 2000) using recently corrected calibrated mutation rate: 19 171 years per mutation in HVS-I (16090–16365) (Soares et al., 2009). It is worthy of note that the age generated could be only intended as a rough guide for relative haplogroup ages. The reasons are that: (i) the mutation rates used in time estimation are in debate; and (ii) some haplogroups had already differentiated before their arrival in Hainan Island.

Reference population data on the Y chromosomes of Hlai groups (Li et al., 2008) and mtDNA (Fucharoen et al., 2001; Tsai et al., 2001; Oota et al., 2002; Tajima et al., 2004; Wen et al., 2004a, 2004b, 2005; Black et al., 2006; Hill et al., 2006, 2007; Trivedi et al., 2006; Li et al., 2007; Wong et al., 2007; Lertrit et al., 2008; Irwin et al., 2008; Dancause et al., 2009; Jin et al., 2009; Mona et al., 2009; Zimmermann et al., 2009; Maruyama et al., 2010; Peng et al., 2010, 2011; Tabbada et al., 2010; Yang et al., 2011; Wang et al., 2012) were retrieved from published works.

## 2 Results and Discussion

### 2.1 Population summary statistics

We typed 566 individuals from five Hlai populations, Gei, Ha, Jiamao, Moifau, and Zwn. Mitochondrial DNA HVS-I sequences showed high levels of diversity, as measured by the number of polymorphic sites, number of haplotypes, the mean number of pairwise differences, and haplotype diversity (Table 1). For the entire Hlai population sample, a total of 203 different haplotypes were defined by 122 variable sites in the HVS-I dataset. All five studied populations showed moderate genetic diversities, close to the average of East Asia and Southeast Asia populations. The haplotype diversity of the Hlai groups ranged from 0.944 to 0.986, with the lowest haplotype diversity observed in Moifau (0.944) and the highest in Zwn (0.986). Measures of population growth (Tajima's  $D$ , Fu's  $F_s$ , and Fu & Li's  $D^*$  and  $F^*$ ) all gave negative values for each population, with the exception of Jiamao and Zwn for Fu & Li's  $D^*$ . However, all of the negative values except Fu's  $F_s$  of each population were not significant. When we combined those five populations into one Hlai group, all the four tests yielded significant negative values. This result might

**Table 1** Molecular diversity indices and growth summary statistics for Hlai population groups, Hainan Island, China

	N	h	S	Diversity indices				Growth statistics						
				Haplotype diversity (SD)	MNPD (SD)	Nucleotide diversity (SD)	Tajima's D	<i>P</i>	Fu's <i>F<sub>s</sub></i>	<i>P</i>	Fu & Li's <i>D*</i>	<i>P</i>	Fu & Li's <i>F*</i>	<i>P</i>
Gei	112	52	69	0.970±0.006	8.621±4.012	0.0196±0.0101	-1.085	0.132	-24.606	0.0000	-0.773	>0.10	-1.118	>0.10
Ha	122	60	67	0.985±0.003	8.464±3.943	0.0192±0.0099	-1.017	0.145	-24.598	0.0000	-1.100	>0.10	-1.372	>0.10
Jiamao	101	55	73	0.982±0.004	9.090±4.218	0.0207±0.0106	-1.149	0.110	-24.580	0.0000	0.256	>0.10	-0.426	>0.10
Moifau	103	45	72	0.944±0.012	8.301±3.877	0.0189±0.0098	-1.293	0.073	-24.689	0.0000	-1.243	>0.10	-1.545	>0.10
Zwn	128	75	81	0.986±0.004	8.357±3.895	0.0190±0.0098	-1.403	0.051	-24.598	0.0001	0.151	>0.10	-0.688	>0.10
Total	566	203	122	0.987±0.001	8.715±4.029	0.0198±0.0101	-1.471	0.023	-23.947	0.0030	-2.507	<0.05	-2.429	<0.05

h, number of haplotypes; MNPD, mean number of pairwise differences; N, number of sequences; *P*, probability value; S, number of polymorphic sites; SD, standard deviation.

suggest very low levels of growth across the Hlai groups.

### 2.2 Mitochondrial DNA haplogroup profiles

All the samples were assigned to mtDNA haplogroups using a combination of HVS-I sequence motifs and SNPs distributed around the coding region of the mtDNA genome. A total of 43 haplogroups or paragroups (unclassified lineages within a clade marked with an asterisk (\*)) were identified (Table S1). Among the five groups, Ha shows the highest degree of heterogeneity, with 30 lineages, followed by Zwn (29) and Jiamao (26), and the other two populations have 24 lineages each. Almost all the mtDNA lineages in Hlai groups are East Asian-specific haplogroups, including those from Northeast Asia (A, D, G, M8a, C, and Z) and southern China or Southeast Asia (B, F, M7, and R9). The frequencies of Northeast Asian haplogroups in Gei, Ha, Jiamao, Moifau, and Zwn are 16.07%, 13.12%, 10.89%, 14.56%, and 21.88%, respectively. However, Jiamao, Ha, Moifau, Gei and, to a lesser extent, Zwn, show a considerable proportion of southern China or Southeast Asia lineages (81.19%, 77.87%, 77.67%, 75.00%, and 62.50%, respectively) (Table S2). The high frequencies of southern haplogroups are consistent with our previous paternal Y chromosome study (Li et al., 2008) which suggested that the specific lineages of southern areas of mainland East Asia (haplogroups O1a-M119 and O2a1-M95) are also dominant in Hlai groups.

### 2.3 Phylogeography of macrohaplogroup N

However, mtDNA haplogroup frequencies vary a lot among the Hlai populations (Table S2). Haplogroup R and its subhaplogroups (F and B) represent the majority of the lineages branching from the superhaplogroup N. Haplogroup F1a1 and F1b comprise 20.39% and 18.45% of Moifau people, respectively. In the other four Hlai populations, however, those two haplogroups are observed at moderate or relatively low levels (Zwn: 0, 2.34%; Gei: 9.82%, 2.68%;

Jiamao: 3.96%, 3.96%; Ha: 5.74%, 2.46%). At the haplotype level, one haplotype (16129-16162-16172-16304-16399) was carried by 90.48% of Moifau F1a1 individuals and the other haplotype (16189-16304-16311) also contributed 89.47% of Moifau F1b, suggesting a strong bottleneck or founder effect in the maternal population history of Moifau. It is noteworthy that haplogroup F1a4, which was previously suggested as one of the candidate lineages for a mid-Holocene Taiwanese dispersal (Hill et al., 2007; Tabbada et al., 2010), was also detected, albeit limited, in Hlai groups except Moifau (from 1.79% in Gei to 0.82% in Ha). Haplogroup B4 is characterized by subhaplogroup B4b1 in Hlai groups but its frequency decreases northward, from Ha (10.66%) and Jiamao (8.91%) to Gei (5.36%), Zwn (3.91%) and Moifau (2.91%). Haplogroup B5 is largely represented by B5a in Gei (7.14%), Zwn (4.69%), Ha (4.10%), Jiamao (2.97%), and Moifau (1.94%).

One of the basal mtDNA lineages R22 was reported in the Nicobar Islands (Trivedi et al., 2006), Nias islands (van Oven et al., 2011), and shows a concentration in eastern Indonesia (Hill et al., 2007). R22 was also detected in Cham, Thailand, and Cambodia (4.2%, 1.1%, and 6.5%, respectively) (Peng et al., 2010). R22 might represent indigenous Southeast Asian lineages. However, R22 was also found in Gei at a frequency of 0.89% in this study, which might reflect very recent gene flow, or probably represent ancient maternal lineages tracing back to the first settlers in Hainan. Haplogroup N\*, a sister clade of haplogroup R, is largely represented by N9a in Hlai populations. Specifically, haplotype N9a-16311-16390, which predominates in Maonan (12.5%) and other populations at the border of Guangxi and Guizhou (Li et al., 2007), comprises 63.64% of all the N9a samples in Hlai.

### 2.4 Phylogeography of macrohaplogroup M

The frequencies of superhaplogroup M in Hlai populations also vary heavily, ranging from 54.46% in

Gei to 21.36% in Moifau. Within this superhaplogroup, haplogroup M7 and its subhaplogroups (M7b, M7b1, M7c1, and M7e) were found to predominate in the Hlai maternal gene pool, which represents nearly a quarter of individuals in Zwn and Jiamao, and one third of Gei and Ha. Correspondingly, these haplogroups are rare or absent in Moifau. It is noteworthy that haplogroup M7\* and M7b show geographical differential distributions among the four Hlai populations (Zwn, Gei, Jiamao, and Ha). Higher proportions of M7b are observed in the northern populations (Zwn and Gei: 11.72%, 9.82%, respectively) compared to that in the southern groups (Jiamao and Ha: 5.94%, 5.74%, respectively). In contrast, M7\* comprises 11.88% and 12.30% of Jiamao and Ha, but only accounts for 3.13% and 4.46% of Zwn and Gei. At the haplotype level, one haplotype (16129-16223-16297) comprises 64.10% of Hlai M7b samples. Haplogroup M7e represents roughly 10% of Gei and Ha, but only comprises 6.25% of Zwn and less than 1% of Moifau and Jiamao. Within M7e, only one haplotype was found among 12 Gei individuals, whereas five were

found among the Hlai. Haplogroup D, the sister clade of haplogroup M\*, accounts for more than 10% of Zwn and Gei, but only comprises 3.96% of Jiamao. Specifically, subhaplogroup D5 represents the majority of haplogroup D samples in Gei (8.04%) and Moifau (7.77%), as well as Ha (4.92%). In contrast, D4a accounts for 6.25% of Zwn.

## 2.5 Population structure

The detailed genetic patterns among Hlai groups and other East Asian and Southeast Asian populations were discerned with the aid of additional published mtDNA datasets. We carried out a principal component analysis using the mtDNA haplogroup frequencies of Hlai groups in this study and another 65 populations (Fig. 1). The first principal component revealed a clear geographic division between ISEA (islands of Southeast Asia, including Taiwan) and mainland East and Southeast Asia. Hlai groups were clustered together within the MSEA (mainland Southeast Asia) and MEA (mainland East Asia) group. Haplogroups M8, M7b, C,

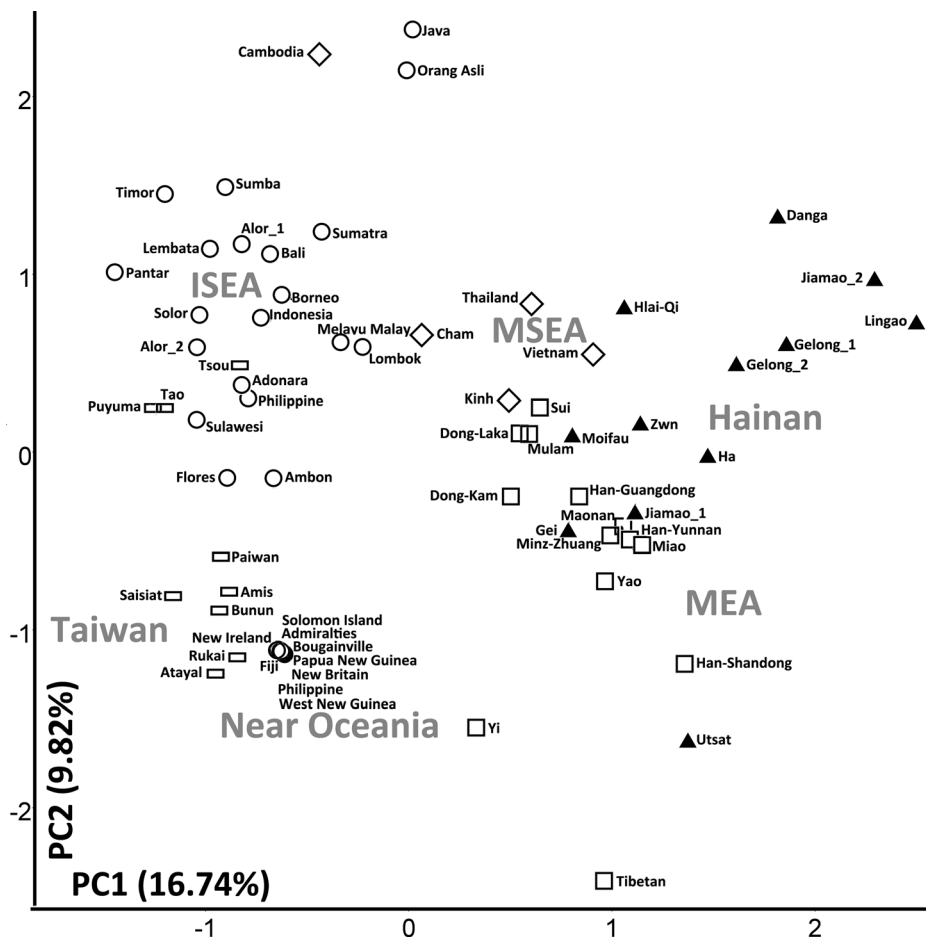
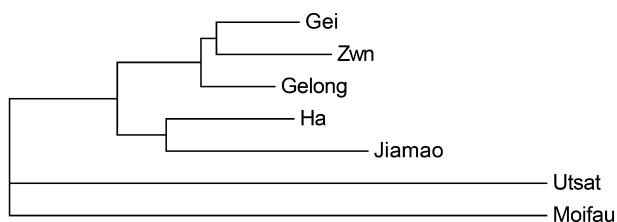


Fig. 1. Principal component analysis plot based on mitochondrial DNA haplogroup frequencies of 70 populations from East Asia and Southeast Asia.



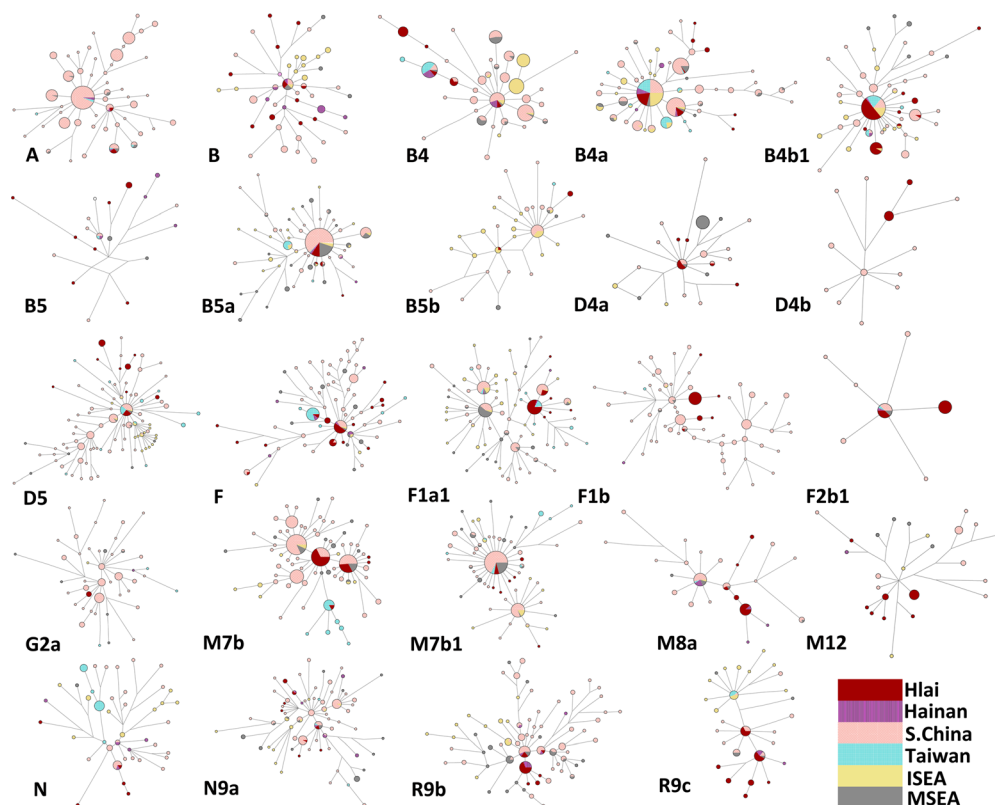
**Fig. 2.** Neighbor-joining tree of Hlai groups and other populations in Hainan, China.

and G contributed most to the MSEA and MEA pole. In contrast, haplogroups E, P, Q, and M7c were found to contribute most to the ISEA pole. To visualize the detailed relationships among Hlai populations and other Hainan populations, a neighbor-joining unrooted tree was constructed using the HVS-I sequences using Slatkin's linearized  $F_{ST}$  as the genetic distance between populations using MEGA 5.1 (Saitou & Nei, 1987; Tamura et al., 2011) (Fig. 2). An obvious geographical division was observed in the neighbor-joining tree, in that northern Hainan groups (Gei, Zwn, and Gelong) cluster away from southern Hainan populations (Ha and

Jiamao). However, Moifau fell out of the Hlai group, probably because of the high frequencies of its founder lineages, F1a1 and F1b.

### 2.6 Network analysis

However, the analysis based on the haplogroup frequency comparison could be misleading due to 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 relations among Hlai groups and other populations (Li et al., 2007; Qin et al., 2010). Based on the mtDNA HVS-1 motif and SNP-determined haplogroups, the networks of mtDNA haplogroups A, B, B4, B4a, B4b1, B5, B5a, B5b, D4a, D4b, D5, F\*, F1a1, F1b, F2b1, G2a, M7b, M7b1, M8a, M12, N\*, N9a, R9b, and R9c were analyzed (Fig. 3). More than half of all the Hlai samples were found sharing haplotypes with individuals from southern China and MSEA. In the networks of B4, D4b, D5, F1b, F2b1, and M12, some Hlai samples formed almost exclusive clades with few individuals



**Fig. 3.** Networks of mitochondrial DNA showing the gene flow between Hlai and other populations in Hainan, South China, Southeast Asia, and Taiwan. The Network program was run at the default settings with the same weight for each single nucleotide polymorphism. Node sizes are proportional to the individual numbers of haplotypes. The lengths of the lines between nodes are proportional to the step numbers of the mutations. ISEA, islands of Southeast Asia; MSEA, mainland Southeast Asia; S.China, southern China.

from other populations. Most Hlai samples were also clustered in exclusive clades with other Hainan aborigines in the networks of M8a and R9b. Those exclusive haplotypes might represent the first settlers on this island. Hlai groups and other Hainan aborigines, descendants of the first settlers, probably had undergone bottlenecks or founder events in their history, and remained genetically isolated for a long period of time. Some Hlai samples were clustered with Taiwan individuals in exclusive clades in the networks of F\*, F1a1, and M7b. Those shared lineages might be ancestral haplotypes, and their descendent haplotypes depicted star-like networks surrounding those exclusive clades, which probably indicated population expansions.

## 2.7 Neutrality tests

To validate the authenticity of these star-like expansion signals in the networks, we used population growth statistics (Fu's Fs) and frequency spectrum tests for deviation from neutrality (Tajima's D and Fu & Li's D, D\*, F, and F\*) to show the differences of the expansion and selection effects among lineages (Table 2). Significant negative values of these tests indicate positive selection, purifying selection, or demographic expansion effect. Significant negative Fu's Fs values were detected in B\*, B4a, F\*, F1b, M7b, and R9b. It is particularly noteworthy that F1b also

showed significant negative values for frequency spectrum tests, together with another two lineages, B4b1 and F1a1.

## 2.8 Time estimation

Time estimates of the approximate ages of several high-frequency branches in specific haplogroups of Hlai were also carried out (Table 3). The mtDNA lineage history generally falls into three phases, the first at approximately 50–40 thousand years ago (kya), the second from 40 to 15 kya, and the third from 15 kya. The first phase of Hainan mtDNA lineage history represents the initial settlement of the island as part of the African dispersal ~50 kya. During this time, which coincides with the last Ice Age when the sea level was much lower than it is today, the island was still connected with mainland Southeast Asia and East Asia. Hainan acted as one of the routes of modern human migration from Southeast Asia to East Asia. The second phase reflects recurrent colonization events from mainland Asia before the Last Glacial Maximum (LGM), which was recorded by many wide distributed lineages across Southeast Asia and East Asia, such as F\*, B4a, D4a. The third phase reflects population expansions after the LGM and Neolithic migrations in and out of Hainan Island. Strong population expansions were detected under lineages F1b, M7b, and R9b. Low levels of population growth were also observed in

**Table 2** Growth summary statistics and frequency spectrum tests for deviation from neutrality

Haplogroup	Tajima's D	Fu's Fs	Fu & Li's D	Fu & Li's D <sup>a</sup>	Fu & Li's F	Fu & Li's F <sup>a</sup>
A	0.379	-2.262 <sup>b</sup>	0.796	0.762	0.822	0.753
B <sup>a</sup>	-0.823	<b>-4.388<sup>a</sup></b>	-0.438	-0.226	-0.671	-0.44
B4	0.498	0.587	-0.266	-0.131	-0.029	0.058
B4a	-1.238	<b>-9.448<sup>a</sup></b>	-1.828	-1.948	-1.971	-2.024
B4b1	<b>-1.909<sup>a</sup></b>	-2.346 <sup>b</sup>	-1.718	<b>-3.060<sup>a</sup></b>	-2.147 <sup>b</sup>	<b>-3.163<sup>a</sup></b>
B5	-0.607	-0.024	-1.012	-0.972	-1.093	-0.994
B5a	-1.727 <sup>b</sup>	-2.268 <sup>b</sup>	-2.002 <sup>b</sup>	-2.052 <sup>b</sup>	-2.304 <sup>b</sup>	-2.286 <sup>b</sup>
B5b	-0.797	2.598	0.513	-0.797	0.339	-0.753
D4a	-1.422	-1.549	-1.044	-1.375	-1.375	-1.549
D4b	0.334	0.536	0.792	0.888	0.803	0.825
D5	-1.03	-1.504 <sup>b</sup>	-1.876	-1.942	-1.92	-1.943
F <sup>a</sup>	-1.471	<b>-5.175<sup>a</sup></b>	-2.002 <sup>b</sup>	-1.92	-2.200 <sup>b</sup>	-2.092 <sup>b</sup>
F1a1	-1.546	-1.865 <sup>b</sup>	-1.074	<b>-2.744<sup>a</sup></b>	-1.468	<b>-2.773<sup>a</sup></b>
F1b	<b>-1.814<sup>a</sup></b>	<b>-3.658<sup>a</sup></b>	<b>-2.715<sup>a</sup></b>	<b>-2.980<sup>a</sup></b>	<b>-2.895<sup>a</sup></b>	<b>-3.066<sup>a</sup></b>
F2b1	1.289	-0.317	0.855	0.866	1.131	1.125
M7 <sup>a</sup>	-0.94	-2.172 <sup>b</sup>	-2.207 <sup>b</sup>	-1.86	-2.117 <sup>b</sup>	-1.833
M7b	-1.271	<b>-3.193<sup>a</sup></b>	-0.407	-0.342	-0.827	-0.752
M7b1	-1.346	-2.422 <sup>b</sup>	-0.725	-1.782	-1.093	-1.902
M7c1	-0.302	0.263	-0.803	-0.519	-0.814	-0.507
M7e	0.084	-2.077 <sup>b</sup>	-0.44	-0.376	-0.329	-0.277
M8a	0.69	-0.935	1.291	1.236	1.319	1.25
M12	-0.649	-0.293	-0.881	-1.352	-0.978	-1.332
N <sup>a</sup>	-0.504	-0.714	-0.473	-0.504	-0.575	-0.539
N9a	0.073	-2.639 <sup>b</sup>	0.12	-0.096	0.128	-0.059
R9b	-1.23	<b>-3.432<sup>a</sup></b>	-0.217	-0.822	-0.606	-1.079
R9c	0.751	-2.286 <sup>b</sup>	1.399 <sup>b</sup>	0.704	1.425	0.838

<sup>a</sup> $P < 0.05$  (numerals in bold).

<sup>b</sup> $0.05 < P < 0.1$ . Haplotype L3e was used as an out group when calculating Fu & Li's D and F.

**Table 3** Estimates of the coalescence time of selected mitochondrial DNA haplogroups inferred from the  $\rho$  statistic within Hainan

Haplogroup	B*	B4	B4a	B4b1	B5a	D4a	D4b	D5	F	F1a1
Time (years BP $\pm$ SD)	49670 $\pm$ 13158	50768 $\pm$ 13364	27036 $\pm$ 7357	13560 $\pm$ 6079	12405 $\pm$ 3383	19171 $\pm$ 6390	4793 $\pm$ 4793	46275 $\pm$ 16580	34601 $\pm$ 7214	8335 $\pm$ 3997
Haplogroup	F1b	F2b1	M7b	M7b1	M8a	M12	N*	N9a	R9b	R9c
Time (years BP $\pm$ SD)	9876 $\pm$ 3897	8594 $\pm$ 8594	15976 $\pm$ 12441	14378 $\pm$ 3929	31751 $\pm$ 17822	37214 $\pm$ 16689	76684 $\pm$ 17040	51413 $\pm$ 14685	13693 $\pm$ 6495	16775 $\pm$ 7459

BP, before present; SD, standard deviation.

B4b1, B5a, F1a1, M7b1, M7e, and R9c. As a consequence of environment change after the LGM period, selection also started to play a role in the population adaptation process (Table 2).

## 2.9 Comparison between paternal and maternal lineages

Our previous studies into Y chromosome diversity of Hlai populations resolved that haplogroups O1a\* and O2a\* are dominant among Hainan aborigines. The exclusive clades in the short tandem repeat (STR) networks of those two haplogroups suggest that the Hainan aborigines have been isolated at the entrance to East Asia for about 20 000 years (Li et al., 2008). In this study, we have also detected some isolated mtDNA lineages in Hlai. The genetic patterns in human societies are often influenced by their cultural practices, such as residence patterns and subsistence strategies (Forster & Colin, 2011). Therefore, the mtDNA results of this study give us a good opportunity to test whether the sex-biased cultural practices also have an influence on relatively isolated populations. We used AMOVA to reveal the possible genetic differences of paternal Y chromosome and maternal mtDNA between Hainan aborigines. The Y chromosome STRs ( $F_{ST} = 0.08022$ ) have distinctly higher diversity among populations than mtDNA HVS-I ( $F_{ST} = 0.02313$ ), suggesting that women have moved more widely between communities in the past than men. Such sex-biased dispersal and admixture patterns remain even after correcting for the different mutation rates of the Y chromosome STRs and the mtDNA HVS-I. Although the society of the Hlai is patrilineal nowadays, matriarchy probably dominated ancestral Hlai societies for a very long time. Frog grain and LiMu worship, tattoos of Hlai women, and Neolithic pottery pattern of reproduction worship are all thought to be vestiges of matriarchal society. That is to say, this sex-biased genetic pattern began to emerge a few thousand years ago, after the transition from matriarchy to patriarchy occurred in those isolated populations. Cultural practices probably play an important role in shaping human population genetic structures. However, we still could not rule out other possible interpretations, for example, that there might be a higher effective population size in ancient Hlai for females than for males, so that the original mtDNA variation of Hlai was less affected by genetic drift.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (Grant Nos. 31071098, 30860124, 30890034, 91131002), the National Excellent Youth Science Foundation of China

(Grant No. 31222030), the Shanghai Rising-Star Program (Grant No. 12QA1400300), the Shanghai Commission of Education Research Innovation Key Project (Grant No. 11zz04), and Shanghai Professional Development Funding (Grant No. 2010001).

## References

- Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37–48.
- Black ML, Dufall K, Wise C, Sullivan S, Bittles AH. 2006. Genetic ancestries in northwest Cambodia. *Annals of Human Biology* 33: 620–627.
- Dancause KN, Chan CW, Arunotai NH, Lum JK. 2009. Origins of the Moken Sea Gypsies inferred from mitochondrial hypervariable region and whole genome sequences. *Journal of Human Genetics* 54: 86–93.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.
- Forster P, Colin R. 2011. Mother tongue and Y chromosomes. *Science* 333: 1390–1391.
- Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: A reappraisal. *American Journal of Human Genetics* 59: 935–945.
- Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.
- Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133: 693–709.
- Fucharoen G, Fucharoen S, Horai S. 2001. Mitochondrial DNA polymorphisms in Thailand. *Journal of Human Genetics* 46: 115–125.
- Hill C, Soares P, Mormina M, Macaulay V, Clarke D, Blumbach PB, Vizuete-Forster M, Forster P, Bulbeck D, Oppenheimer S, Richards M. 2007. A mitochondrial stratigraphy for Island Southeast Asia. *American Journal of Human Genetics* 80: 29–43.
- Hill C, Soares P, Mormina M, Macaulay V, Meehan W, Blackburn J, Clarke D, Raja JM, Ismail P, Bulbeck D, Oppenheimer S, Richards M. 2006. Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Molecular Biology and Evolution* 23: 2480–2491.
- Irwin JA, Saunier JL, Strouss KM, Diegoli TM, Sturk KA, O’Callaghan JE, Paintner CD, Hohoff C, Brinkmann B, Parsons TJ. 2008. Mitochondrial control region sequences from a Vietnamese population sample. *International Journal of Legal Medicine* 122: 257–259.
- Jin HJ, Tyler-Smith C, Kim W. 2009. The peopling of Korea revealed by analyses of mitochondrial DNA and Y-chromosomal markers. *PLoS ONE* 4: e4210.
- 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.
- Lertrit P, Poolsuwan S, Thosarat R, Sanpachudayan T, Boonyarit H, Chinpaisal C, Suktitipat B. 2008. Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis. *American Journal of Physical Anthropology* 137: 425–440.
- 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.
- Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
- 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: 13–19.
- Maruyama S, Nohira-Koike C, Minaguchi K, Nambiar P. 2010. mtDNA control region sequence polymorphisms and phylogenetic analysis of Malay population living in or around Kuala Lumpur in Malaysia. *International Journal of Legal Medicine* 124: 165–170.
- Mona S, Grunz KE, Brauer S, Pakendorf B, Castri L, Sudoyo H, Marzuki S, Barnes RH, Schmidtke J, Stoneking M, Kayser M. 2009. Genetic admixture history of eastern Indonesia as revealed by Y chromosome and mitochondrial DNA analysis. *Molecular Biology and Evolution* 26: 1865–1877.
- Oota H, Kitano T, Jin F, Yuasa I, Wang L, Ueda S, Saitou N, Stoneking M. 2002. Extreme mtDNA homogeneity in continental Asian populations. *American Journal of Physical Anthropology* 118: 146–153.
- van Oven M, Hämmerle JM, van Schoor M, Kushnick G, Pennekamp P, Zega I, Oscar Lao O Brown L, Kennerknecht I, Kayser M. 2011. Unexpected island effects at an extreme: Reduced Y chromosome and mitochondrial DNA diversity in Nias. *Molecular Biology and Evolution* 28: 1349–1361.
- Peng MS, He JD, Liu HX, Zhang YP. 2011. Tracing the legacy of the early Hainan Islanders—A perspective from mitochondrial DNA. *BMC Evolutionary Biology* 11: 46.
- Peng MS, Quang HH, Dang KP, Trieu AV, Wang HW, Yao YG, Kong QP, Zhang YP. 2010. Tracing the Austronesian footprint in Mainland Southeast Asia: A perspective from mitochondrial DNA. *Molecular Biology and Evolution* 27: 2417–2430.
- 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.
- Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S. 2000. mtDNA variation among Greenland Eskimos: The edge of the Beringian expansion. *American Journal of Human Genetics* 67: 718–726.



- Saitou N, Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406–425.
- Shi YF, Cui ZJ, Li JJ. 1989. Quaternary glacier in Eastern China and the Climate Fluctuation. Beijing: Science Press.
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for purifying selection: An improved human mitochondrial molecular clock. *American Journal of Human Genetics* 84: 740–759.
- Tabbada KA, Trejaut J, Loo JH, Chen YM, Lin M, Mirazón-Lahr M, Kivisild T, De Ungria MC. 2010. Philippine mitochondrial DNA diversity: A populated viaduct between Taiwan and Indonesia? *Molecular Biology and Evolution* 27: 21–31.
- Tajima F. 1989. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- Tajima A, Hayami M, Tokunaga K, Juji T, Matsuo M, Marzuki S, Omoto K, Horai S. 2004. Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *Journal of Human Genetics* 49: 187–193.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.
- Thurgood G. 1994. Tai-Kadai and Austronesian: The nature of the historical relationship. *Oceanic Linguistics* 33: 345–368.
- Trivedi R, Sitalaximi T, Banerjee J, Singh A, Sircar PK, Kashyap VK. 2006. Molecular insights into the origins of the Shompen, a declining population of the Nicobar archipelago. *Journal of Human Genetics* 51: 217–226.
- Tsai LC, Lin CY, Lee JC, Chang JG, Linacre A, Goodwin W. 2001. Sequence polymorphism of mitochondrial D-loop DNA in the Taiwanese Han population. *Forensic Science International* 119: 239–247.
- 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.
- Wang XP, Wang JC, Xing GY. 2004. China Hlai. Beijing: The Ethnic Publishing House.
- Wen B, Li H, Gao S, Mao X, Gao Y, Li F, Zhang F, He Y, Dong Y, Zhang Y, Huang W, Jin J, Xiao C, Lu D, Chakraborty R, Su B, Deka R, Jin L. 2005. Genetic structure of Hmong–Mien speaking populations in East Asia as revealed by mtDNA lineages. *Molecular Biology and Evolution* 22: 725–734.
- 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, Deka 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.
- Wong HY, Tang JS, Budowle B, Allard MW, Syn CK, Tan-Siew WF, Chow ST. 2007. Sequence polymorphism of the mitochondrial DNA hypervariable regions I and II in 205 Singapore Malays. *Legal Medicine (Tokyo)* 9: 33–37.
- Xing GY. 2004. Discursion from the earliest governmentalization of Hainan Island. *Hainan Archives* 3: 43–47.
- 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.
- Zimmermann B, Bodner M, Amory S, Fendt L, Rock A, Horst D, Horst B, Sanguanserm Sri T, Parson W, Brandstätter A. 2009. Forensic and phylogeographic characterization of mtDNA lineages from northern Thailand (Chiang Mai). *International Journal of Legal Medicine* 123: 495–501.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at <http://onlinelibrary.wiley.com/doi/10.1111/jse.12024/supinfo>:

**Table S1.** Mitochondrial DNA haplogroups and hypervariable sequence I (HVS-I) motif of Hlai populations, Hainan Island, China

**Table S2.** Mitochondrial DNA haplogroup frequencies of Hlai populations, Hainan Island, China

## 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)