Molecular Biology and Evolution

Research article

# Tracing the Austronesian Footprint in Mainland Southeast Asia: A Perspective from Mitochondrial DNA

Min-Sheng Peng,†[,1,2,6] Huy Ho Quang,†[,4] Khoa Pham Dang,[4] An Vu Trieu,[4] Hua-Wei Wang,[3] Yong-Gang Yao,[5] Qing-Peng Kong,*[,1,2] and Ya-Ping Zhang*[,1,2,3]

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, P.R. China

[2]KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming, P.R. China

[3]Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming, P.R. China

[4]Department of Immunology, Ha Noi Medical University, Ha Noi, Vietnam

[5]Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Kunming, P.R. China

[6]Graduate School of the Chinese Academy of Sciences, Beijing, P.R. China

†These authors contributed equally to this work.

***Corresponding author:** E-mail: zhangyp@mail.kiz.ac.cn; kongqp@yahoo.com.cn.

**Associate editor**: Lisa Matisoo-Smith

The sequence data for this paper appear in GenBank under accession numbers GQ301556–GQ301862 for HVS sequences and GQ301863–GQ301886 and GU592216–GU592219 for complete mtDNA genomes.

## Abstract

As the relic of the ancient Champa Kingdom, the Cham people represent the major Austronesian speakers in Mainland Southeast Asia (MSEA) and their origin is evidently associated with the Austronesian diffusion in MSEA. Hitherto, hypotheses stemming mainly from linguistic and cultural viewpoints on the origin of the Cham people remain a welter of controversies. Among the points of dissension is the muddled issue of whether the Cham people arose from demic or cultural diffusion from the Austronesians. Addressing this issue also helps elucidate the dispersal mode of the Austronesian language. In the present study, we have analyzed mitochondrial DNA (mtDNA) control-region and coding-region sequence variations in 168 Cham and 139 Kinh individuals from Vietnam. Around 77% and 95% matrilineal components in the Chams and the Kinhs, respectively, could be assigned into the defined mtDNA haplogroups. Additionally, three common East Eurasian haplogroups B, R9, and M7 account for the majority (>60%) of maternal components in both populations. Entire sequencing of 20 representative mtDNAs selected from the thus far unclassified lineages, together with four new mtDNA genome sequences from Thailand, led to the identification of one new haplogroup M77 and helped to re-evaluate several haplogroups determined previously. Comparing the Chams with other Southeast Asian populations reveals that the Chams had a closer affinity with the Mon–Khmer populations in MSEA than with the Austronesian populations from Island Southeast Asia (ISEA). Further analyses failed to detect the potential homelands of the Chams in ISEA. Therefore, our results suggested that the origin of the Cham was likely a process of assimilation of massive local Mon–Khmer populations accompanied with language shift, thus indicating that the Austronesian diffusion in MSEA was mainly mediated by cultural diffusion, at least from the matrilineal genetic perspective, an observation in agreement with the hypothesis of the Nusantao Maritime Trading and Communication Networks.

**Key words:** mtDNA, Cham, Vietnam, Austronesian, NMTCN.

## Introduction

The vast majority of Austronesian languages are distributed on islands from Madagascar to Easter Island; the exceptions are Moken and Chamic, which are spoken in Mainland Southeast Asia (MSEA) by two minority groups, Moken and Cham, respectively (Bellwood et al. 2006). In contrast to the Moken people living as "Sea Gypsies" with a relatively small population size, the Cham people established the thriving Kingdom of Champa, which lasted as a major early Southeast Asian historic civilization for more than 1 millennium. Champa reached its zenith from the sixth to the tenth century AD and once ruled over the coastal plains and the interior highlands in South-Central Vietnam. Thereafter, Champa began a gradual decline because of the growing invasions of the Kinh people from northern Vietnam as well as the long-drawn-out wars with Khmer Empire, finally being forced to merge with the Kinhs in 1832 AD (Southworth 2004; Thurgood 2005; He 2006). Thus, a deeper insight into the origin of the Cham people, who harbor this Austronesian linguistic relic in MSEA, can help to clarify the question of how the Austronesian language had dispersed into MSEA. In this context, two prevalent hypotheses have been proposed to explain the origin of the Cham people as well as the dispersal mode of the Austronesian language. Under the "Out-of-Taiwan"





http://mbe.oxfordjournals.org/

hypothesis, the Cham ancestors were regarded as the Austronesian immigrants from Island Southeast Asia (ISEA)—especially from southwest Borneo (cf. Blust 1994) around 500 BC (Thurgood 1999; Higham 2002; Southworth 2004; Bellwood 2007). An alternative hypothesis known as "Nusantao Maritime Trading and Communication Networks (NMTCN)" proposes that the origin of the Chams was primarily mediated by cultural diffusion through the agency of the NMTCN in which case the direction of diffusion and its explicit source could not be determined (Solheim et al. 2007). For instance, the coastal strip of the middle and southern part of Vietnam had served as an especially important hub in the extensive maritime commerce networks around South China Sea since 500 BC or even earlier (Higham 1989, 2002; Southworth 2004; Bellwood 2007; Hung et al. 2007; Lam 2009).

In contrast with linguistic and archaeological works, genetic study allows an attempt to distinguish the influence of demic and cultural diffusion for (pre-) historic events (Cavalli-Sforza et al. 1994; Chikhi et al. 2002; Wen, Li, et al. 2004). However, in a recent Y-chromosome study, 11 Cham individuals were assigned into haplogroups O* (1/11) and O1a* (10/11) (Li et al. 2008). Because of the small sample size and the relatively low resolution of Y-chromosome phylogeny, the genetic structure of the Cham population remains poorly understood. Hitherto, no research on the mitochondrial DNA (mtDNA) variation in the Cham population has been reported elsewhere. By contrast, mtDNA data from Southeast Asia, especially from ISEA, are accumulating quickly (Fucharoen et al. 2001; Prasad et al. 2001; Tajima et al. 2004; Black et al. 2006; Hill et al. 2006, 2007; Trivedi et al. 2006; Li et al. 2007; Wong et al. 2007; Irwin et al. 2008; Lertrit et al. 2008; Dancause et al. 2009; Mona et al. 2009; Zimmermann et al. 2009; Maruyama et al. 2010; Tabbada et al. 2010). Furthermore, the appearance of complete mtDNA genomes has improved the resolution of mtDNA phylogeny of this region (Kong et al. 2003, 2006; Macaulay et al. 2005; Tabbada et al. 2010). Therefore, to shed more light on the origin of the Cham people, we first analyzed the whole control-region and partial coding-region sequence variations of mtDNA in 168 Cham individuals collected from southern Vietnam. One hundred and thirty-nine Kinh individuals from northern Vietnam were also sampled for comparison as they represent the majority ethnic group in Vietnam. Our results not only help further understand the mtDNA phylogeny in Southeast Asia but also provide deeper insights into the origin of the Chams—the major Austronesian relic in MSEA.

## Material and Methods

### Sampling and Data Collecting

We have collected 168 unrelated Cham samples (62 males and 106 females) from Binh Thuan Province that was a "sanctuary" for Chams at the point of mergence by the expanding Kinhs and is the only part of the coastal strip retaining significant numbers of Chamic speakers in southern Vietnam (Higham 2002). Additionally, 139 unrelated Kinh samples (79 males and 60 females) were collected from Hanoi, northern Vietnam (fig. 1). All subjects were interviewed to ascertain their ethnic affiliations and to obtain informed consent before blood collection.

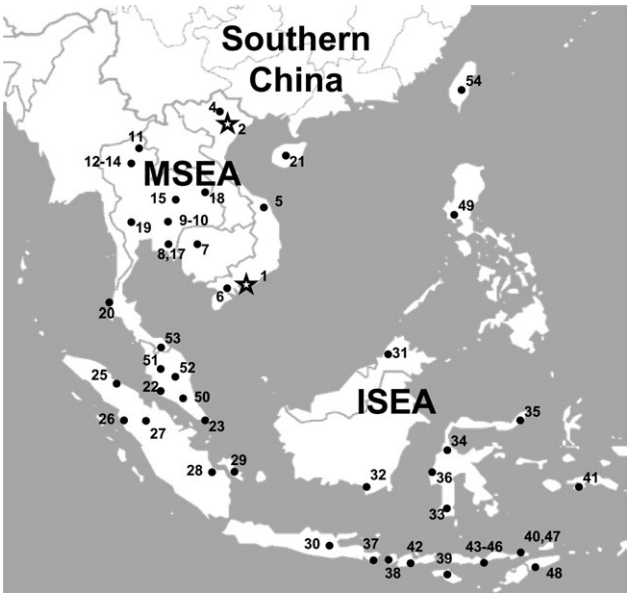


FIG. 1. Map of Southeast Asia and southern part of East Asia showing the locations of the samples considered in this study. The Cham and the Kinh are indicated by the star symbol. The other reported populations are indicated by bold circles (table 1). MSEA, Mainland Southeast Asia (not including Malay Peninsula); ISEA, Island Southeast Asia (including Malay Peninsula).

Comparative mtDNA data from Southeast Asia (fig. 1 and table 1) and southern China were taken from previously published literature (Sykes et al. 1995; Melton et al. 1998; Prasad et al. 2001; Qian et al. 2001; Kivisild et al. 2002; Oota et al. 2002; Yao, Kong, et al. 2002; Yao, Nie, et al. 2002; Yao and Zhang 2002; Kong et al. 2003; Wen, Li, et al. 2004; Wen, Xie, et al. 2004; Macaulay et al. 2005; Trejaut et al. 2005; Wen et al. 2005; Black et al. 2006; Hill et al. 2006, 2007; Kong et al. 2006; Trivedi et al. 2006; Li et al. 2007; Wong et al. 2007; Gan et al. 2008; Irwin et al. 2008; Lertrit et al. 2008; Soares et al. 2008; Dancause et al. 2009; Jin et al. 2009; Mona et al. 2009; Zimmermann et al. 2009; Maruyama et al. 2010; Tabbada et al. 2010; Wang et al. 2010). Additional comparative data were taken from Laos and China (authors' unpublished data).

### mtDNA Sequence Variation Screening

Genomic DNA was extracted from whole blood samples by the standard phenol/chloroform methods. The mtDNA control-region sequences were amplified by the polymerase chain reaction (PCR) method reported previously (Yao et al. 2003). Then, an 832-bp segment within the control region, including the hypervariable segment I (HVS-I) (16038–16569) and partial HVS-II (1–300), was sequenced in all samples as described elsewhere (Yao et al. 2003). By a first round of haplogroup-specific control-region motif

**Table 1.** General Information about the Populations in Southeast Asia, Hainan, and Taiwan.

| Group | No. | Population | Size | Language | Location | Reference |
|---|---|---|---|---|---|---|
| MSEA | 1 | Cham | 168 | Austronesian | Binh Thuan, Vietnam | This study |
| | 2 | Kinh | 139 | Austro-Asiatic | Hanoi, Vietnam | This study |
| | 3 | Viet_Jin[a] | 42 | Austro-Asiatic | Vietnam | Jin et al. (2009) |
| | 4 | Viet_N[a] | 187 | Austro-Asiatic | Hanoi area, Vietnam | Irwin et al. (2008) |
| | 5 | Viet_M[b] | 63 | Austro-Asiatic | Middle Vietnam | Li et al. (2007) |
| | 6 | Viet_S[a] | 35 | Austro-Asiatic | First-generation South Vietnamese immigrants, USA | Oota et al. (2002) |
| | 7 | Cambodia | 31 | Austro-Asiatic | Siem Reap, Cambodia | Black et al. (2006) |
| | 8 | Khmer | 22 | Austro-Asiatic | Chanthaburi, Thailand | Lertrit et al. (2008) |
| | 9 | Chao-Bon | 20 | Austro-Asiatic | Nakhon Ratchasima, Thailand | Lertrit et al. (2008) |
| | 10 | Thai_Korat | 32 | Tai-Kadai | Nakhon Ratchasima, Thailand | Lertrit et al. (2008) |
| | 11 | Thai_Yao | 34 | Tai-Kadai | Northern Thailand | Yao, Nie, et al. (2002) |
| | 12 | Mussur | 21 | Tibeto-Burman | Chiang Mai, Thailand | Fucharoen et al. (2001) |
| | 13 | Lisu | 25 | Tibeto-Burman | Chiang Mai, Thailand | Fucharoen et al. (2001) |
| | 14 | Thai_CM | 220 | Tai-Kadai | Chiang Mai, Thailand | Fucharoen et al. (2001); Zimmermann et al. (2009) |
| | 15 | Thai_KK | 44 | Tai-Kadai | Khon Kaen,Thailand | Fucharoen et al. (2001) |
| | 16 | Thai_Jin | 40 | Tai-Kadai | Thailand | Jin et al. (2009) |
| | 17 | Chong | 25 | Austro-Asiatic | Chanthaburi, Thailand | Fucharoen et al. (2001) |
| | 18 | Phu-Thai | 25 | Tai-Kadai | Mukdahan, Thailand | Fucharoen et al. (2001) |
| | 19 | Lao-Song | 25 | Tai-Kadai | Suphanburi, Thailand | Fucharoen et al. (2001) |
| | 20 | Moken[c] | 12 | Austronesian | Surin Islands, Thailand | Dancause et al. (2009) |
| Hainan | 21 | Hainan | 159 | Tai-Kadai | Hainan | Li et al. (2007); Tajima et al. (2004); Hill et al. (2006); |
| ISEA | 22 | Malay_KL | 183 | Austronesian | Kuala Lumpur, Malaysia | Maruyama et al. (2010) |
| | 23 | Malay_SG | 205 | Austronesian | Singapore | Wong et al. (2007) |
| | 24 | Indonesia | 54 | Austronesian | Indonesia | Tajima et al. (2004) |
| | 25 | Medan | 42 | Austronesian | Medan, Sumatra | Hill et al. (2006) |
| | 26 | Padang | 24 | Austronesian | Padang, Sumatra | Hill et al. (2006) |
| | 27 | Pekanbaru | 52 | Austronesian | Pekanbaru, Sumatra | Hill et al. (2006) |
| | 28 | Palembang | 28 | Austronesian | Palembang, Sumatra | Hill et al. (2006) |
| | 29 | Bangka | 34 | Austronesian | Bangka-Belitung Islands | Hill et al. (2006) |
| | 30 | Javanese | 46 | Austronesian | Java | Hill et al. (2007) |
| | 31 | Kota Kinabalu | 68 | Austronesian | Kota Kinabulu, Borneo | Hill et al. (2007) |
| | 32 | Banjarmasin | 89 | Austronesian | Banjarmasin. Borneo | Hill et al. (2007) |
| | 33 | Ujung Padang | 46 | Austronesian | Ujung Padang, Sulawesi | Hill et al. (2007) |
| | 34 | Palu | 38 | Austronesian | Palu, Sulawesi | Hill et al. (2007) |
| | 35 | Manado | 89 | Austronesian | Manado, Sulawesi | Hill et al. (2007) |
| | 36 | Toraja | 64 | Austronesian | Toraja, Sulawesi | Hill et al. (2007) |
| | 37 | Bali | 82 | Austronesian | Bali | Hill et al. (2007) |
| | 38 | Lombok | 44 | Austronesian | Lombok | Hill et al. (2007) |
| | 39 | Sumba | 50 | Austronesian | Sumba | Hill et al. (2007) |
| | 40 | Alor-1 | 45 | Austronesian | Alor | Hill et al. (2007) |
| | 41 | Ambon | 43 | Austronesian | Ambon | Hill et al. (2007); Hill et al. (2007); |
| | 42 | Flores | 77 | Austronesian | Flores | Mona et al. (2009) |
| | 43 | Adonara | 77 | Austronesian | Adonara | Mona et al. (2009) |
| | 44 | Solor | 41 | Austronesian | Solor | Mona et al. (2009) |
| | 45 | Lembata | 34 | Austronesian | Lembata | Mona et al. (2009) |
| | 46 | Pantar | 38 | Papuan[d] | Pantar | Mona et al. (2009) |
| | 47 | Alor-2 | 27 | Papuan[e] | Alor | Mona et al. (2009) |
| | 48 | E-Timor | 38 | Austronesian[f] | East Timor | Mona et al. (2009); Tajima et al. (2004); Hill et al. (2007); |
| | 49 | Philippine | 543 | Austronesian | Philippine; Immigrants in Taiwan | Tabbada et al. (2010) |
| | 50 | Aboriginal Malay[g] | 96 | Austronesian | West Malaysia | Hill et al. (2006) |
| | 51 | Semang[g] | 112 | Austro-Asiatic | West Malaysia | Hill et al. (2006) |
| | 52 | Senoi[g] | 52 | Austro-Asiatic | West Malaysia | Hill et al. (2006) |
| | 53 | Sakai[g] | 20 | Austro-Asiatic | Trang, Thailand | Fucharoen et al. (2001) |

**Table 1.** Continued.

| Group | No. | Population | Size | Language | Location | Reference |
|---|---|---|---|---|---|---|
| Taiwan | 54 | Formosa | 718 | Austronesian | Taiwan | Sykes et al. (1995); Melton et al. (1998); Trejaut et al. (2005) |

NOTE.—[a] Because there is no ethnolinguistic information, we referred that this population might be mainly composed of Kinh, belonging to the Austro-Asiatic population.
[b] Although three individuals were Tai-Kadai, we considered the population as Austro-Asiatic.
[c] The Moken were removed in PCA and frequency estimates due to small sample size and severe genetic drift (Dancause et al. 2009).
[d] Although ten individuals were Austronesian, we considered the population as Papuan.
[e] Although three individuals were Austronesian, we considered the population as Papuan.
[f] Although five individuals were Papuan, we considered the population as Austronesian.
[g] As outliers in previous PCA caused by severe genetic drift (Hill et al. 2006), these Orang Asli populations were removed in PCA.

recognition and (near-) matching search with the published mtDNA data (Yao et al. 2004), we were able to tentatively assign each mtDNA under study into specifically named haplogroups. Certain coding-region fragments containing diagnostic sites were further amplified and sequenced to confirm the predicted haplogroup status of each mtDNA (supplementary tables S1 and S2, Supplementary Material online). For the remaining samples that could not be classified into the available mtDNA phylogeny at the time, their phylogenetic status was fully recognized by adopting a strategy that has been used to pinpoint East Asian basal lineages (Kong et al. 2006; authors' unpublished data). Specifically, all unclassified mtDNAs were tentatively assigned into different groups based on their control region–specific variations, then at least one representative was selected from each group for completely mtDNA sequencing. In this way, a total of 24 representatives were selected for complete mitochondrial genome sequencing: 16 from the Chams, 4 from the Kinhs, and 4 from the Thais whose HVS-I sequences had been reported in our previous work (Yao, Nie, et al. 2002). Then, based on this newly obtained mtDNA genomic information, at least one coding-region specific mutation was selected for typing among the remaining unclassified mtDNAs to further ascertain its haplogroup affiliation (supplementary table S1 and S2, Supplementary Material online). Moreover, to further understand the phylogeny of haplogroup B4c2, which presented significant distribution differences from the other haplogroups, four additional sequences belonging to this haplogroup, as judged from their control-region motif (16147-16184A-16189-16217-16235), were selected for complete sequencing as well.

During the process of mtDNA genome sequencing, we adopted sequencing protocols reported elsewhere (Wang et al. 2008; Fendt et al. 2009; Zhao et al. 2009) and followed caveats for quality control in mtDNA genome study (Kong et al. 2008; Yao et al. 2008, 2009). To avoid any nomenclature conflicts, we followed the criterion of PhyloTree (http://www.phylotree.org, mtDNA tree Build 7 [10 November 2009]; van Oven and Kayser 2009) as well as the newly proposed haplogroup naming scheme (Kong et al. 2010). Sequences were edited and aligned by DNASTAR software (DNAStar Inc., Madison, WI), and mutations were scored relative to the revised Cambridge reference sequence (rCRS) (Andrews et al. 1999). For the length variants in the control region, we followed the rules proposed by Bandelt and Parson (2008). The transition at 16519 and the C-length polymorphisms in regions 16180–16193 and 303–315 were disregarded in the analyses.

### Phylogenetic Tree Construction and Data Analysis

In total, 48 complete mtDNA genome sequences (including 28 new mtDNAs obtained in this study as well as 20 eastern Eurasian complete mtDNA genomes retrieved from the literature; Ingman et al. 2000; Macaulay et al. 2005; Reddy et al. 2007; Dancause et al. 2009; Hartmann et al. 2009; Tabbada et al. 2010) were employed for the phylogenetic tree reconstruction. Median-joining network was reconstructed manually and checked by the program NETWORK 4.510 (www.fluxus-engineering.com/sharenet.htm) (Bandelt et al. 1999). We also constructed the reduced median network (Bandelt et al. 1995) of HVS sequences in the Chams using the NETWORK 4.510.

The coalescent age of a haplogroup of interest was estimated by statistics $\rho \pm \sigma$ (Forster et al. 1996; Saillard et al. 2000), and recently corrected calibrated mutation rates were adopted: 7,884 years per synonymous mutation and 18,845 years per transition in HVS-I (16090–16365) (Soares et al. 2009). Principal components analysis (PCA) followed the method developed by Richards et al. (2002) with SPSS13.0 software (SPSS). Analysis of molecular variance (AMOVA) was computed with the package Arlequin 3.11 (Excoffier et al. 2005). To detect the potential Austronesian source of the Chams, haplotype-sharing analyses between the Cham and other Austronesian populations were carried out based on phylogeny (Achilli et al. 2007), and the $D_{HS}$ distances were calculated as described by Tofanelli et al. (2009).

## Results

### Classification of mtDNA Sequences in the Chams and the Kinhs

In terms of the combined information from HVS and partial coding-region segments (supplementary tables S1 and S2, Supplementary Material online), 130 of 168 (~77.4%) of the Cham and 132 of 139 (~95.0%) of the Kinh samples were unambiguously assigned into the previously defined haplogroups in East and Southeast Asian (Macaulay et al. 2005; Kong et al. 2006; Hill et al. 2007), among which the

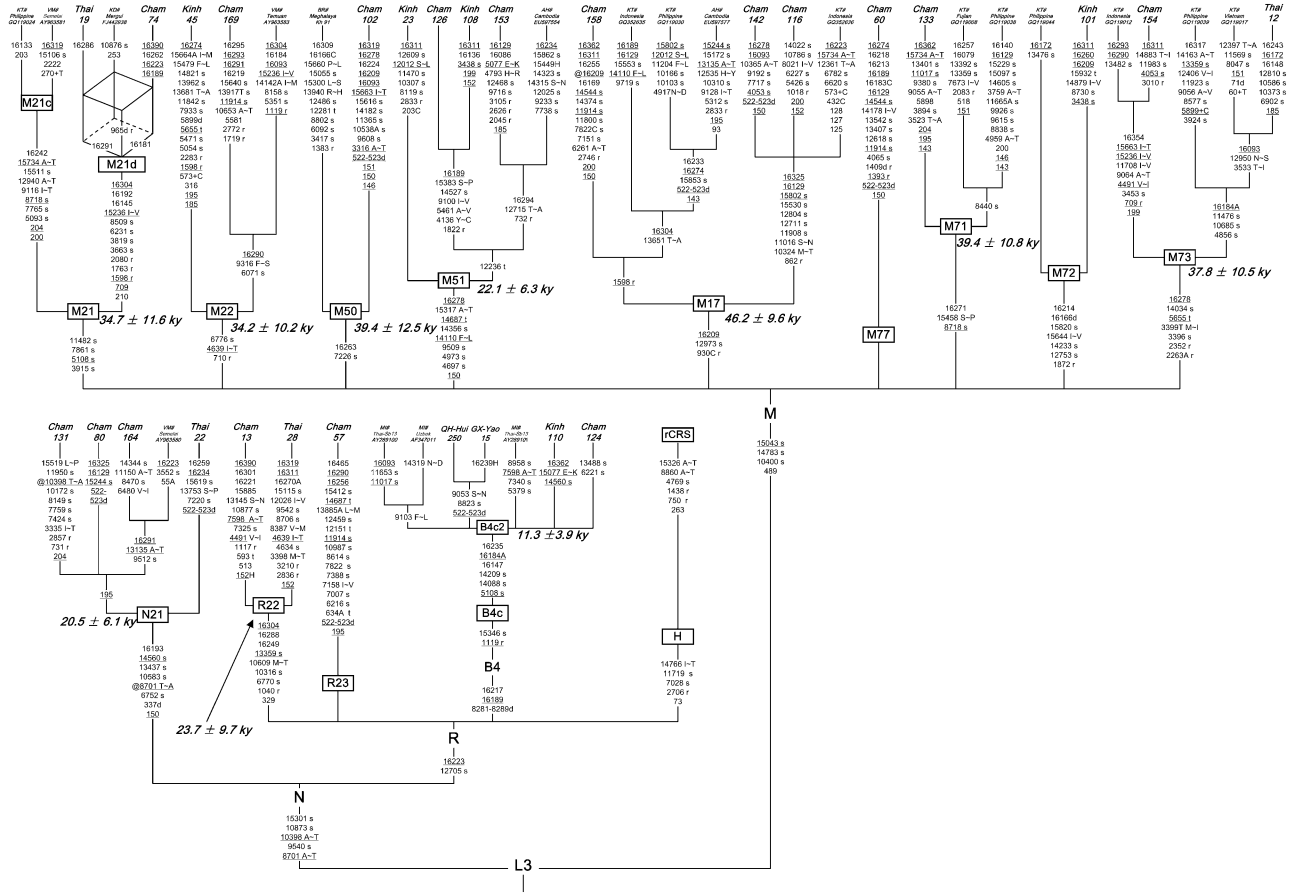


**Fig. 2.** Reconstructed phylogenetic tree of 48 complete mtDNA genome sequences from haplogroups M17, M21, M22, M50, M51, M71, M72, M73, M77, N21, R22, R23, and B4c2. The 20 reported sequences were taken from the literature and were further labeled by the symbols MI (Ingman et al. 2000), VM (Macaulay et al. 2005), BR (Reddy et al. 2007), KD (Dancause et al. 2009), AH (Hartmann et al. 2009), and KT (Tabbada et al. 2010), followed by "#," the geographic locations, and the sample code or the accession numbers in GenBank. Haplogroup age estimates (± standard errors) are indicated at the branch roots in terms of the calibrated mutation rate of 7,884 years per synonymous mutation in coding-region (ky: 1,000 years; Soares et al. 2009). Mutations are transitions at the respective nucleotide position unless otherwise specified. Letters following positions indicate transversions; others are transitions. Recurrent mutations are underlined. +, insertion; d, deletion; @, back mutation; H, heterogeneity. Amino acid replacements are specified by single-letter code; s, synonymous replacements; t, change in transfer RNA; r, change in ribosomal RNA gene.

most common haplogroups are B, R9, and M7, which all together account for 60.1% and 75.5% of the maternal gene pools of the Chams and the Kinhs, respectively.

As displayed in figure 2, the mtDNA phylogeny in Southeast Asia is largely improved by incorporating our 24 new mtDNA genomes, which in return helps recognize the remaining unclassified mtDNAs pinpointed in our study. Now it becomes evident that most of these unclassified mtDNAs observed in the Chams and the Kinhs cluster with certain previously reported mtDNA genomes and can be allocated into haplogroups M17, M21, M22, M50, M51, M71, M72, M73, and N21 (PhyloTree: http://www.phylotree.org). Meanwhile, haplogroups R22 and R23, which were previously defined based merely on HVS-I variation (Hill et al. 2007), are now substantiated by the complete mtDNA genomes (R22: *Cham13* and *Thai28*; R23: *Cham57*). One sequence (*Cham60*) with specific HVS-I motif 16129-16189-16213-16218-16223 was found to share transitions 12618 and 1393 with haplogroups M23 (consisting of sequences GQ389777 and FJ543102) and M46 (FJ442939), respectively, and was tentatively assigned to a newly defined basal haplogroup—M77.

Haplogroups M17, M21, M22, M51, M71, M72, M73, N21, R22, and R23 were initially observed in Southeast Asia and might represent the ancient maternal components in this region (Macaulay et al. 2005; Hill et al. 2006, 2007; Tabbada et al. 2010). In light of the updated mtDNA phylogeny of Southeast Asian (fig. 2), we performed (near-) match searches with HVS-I motifs among the published eastern Eurasian mtDNA data sets to evaluate the distributions of these haplogroups (supplementary table S3, Supplementary Material online). Subsequently, the reduced median networks were reconstructed to display the internal phylogeny within each haplogroup (fig. 3). Our results support the previous observations that haplogroups M17, M21d, M22, M50, M51, M72, M73, M77, N21, R22, and R23 were distributed almost exclusively in Southeast Asia and could trace their origins back to this region (Macaulay et al. 2005; Hill et al. 2006, 2007). The exception is haplogroup M71, which was recently suggested to have a Philippine

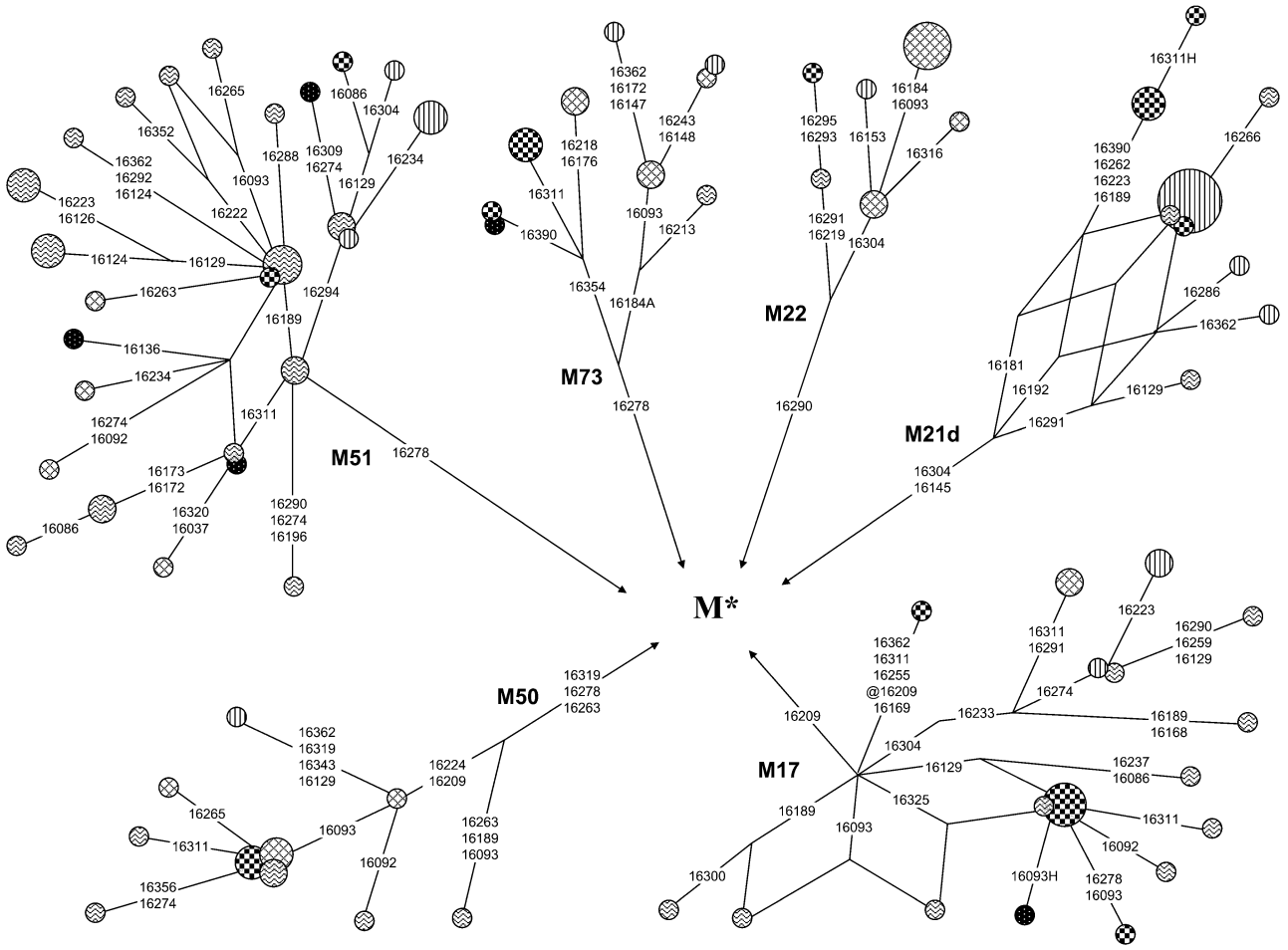


**Fig. 3.** Reduced median network of HVS-I sequences of haplogroups N21, R22, M17, M21d, M22, M50, M51, M71, and M73 in the region 16085–16365. The circles represent mtDNA HVS-I sequence types, shaded according to region with an area proportional to their absolute frequency. Subclades are labeled, and the N*, R*, and M* ancestors are indicated (arrow). Mutations are transitions unless the base change is explicitly indicated. Heteroplasmic positions are indicated by an "H" after the nucleotide position. The A–C transversions at nps 16181, 16182, and 16183 were ignored in interpopulation analyses.

origin (Tabbada et al. 2010) but now seems more likely to trace its root to southern China (this study and authors' unpublished data).

## Comparison of the Chams and the Kinhs with the Other Southeast Asian Populations

To discern the relationships of the Cham and the Kinh populations with other Southeast Asian populations, we have employed PCA based on the haplogroup distribution frequency (supplementary table S4, Supplementary Material online). The first principal component (PC) revealed a clear division between ISEA (including Taiwan) and MSEA, corresponding to Austronesians and non-Austronesians, respectively. The Cham population, as well as the Kinh, was clustered within the MSEA group, which reflected a geographic clustering pattern rather than linguistic affinity (fig. 4*a*). Haplogroups E1, M7c3c, Q, and Y contributed most to the ISEA pole (Austronesians). Contrastingly, haplogroups M7b*, C, F1a1a, and B5a were concentrated at the MSEA pole (non-Austronesians) (fig. 4*b*). Although the variation between the ISEA and MSEA groups only accounted for 1.6% of the total variation, the island–mainland patterning was significant ($P < 0.05$) in AMOVA based on haplogroup profiles (supplementary table S4, Supplementary Material online).

The second PC showed an east–west pattern in ISEA as described before (Hill et al. 2007; Mona et al. 2009; Karafet et al. 2010) and uncovered a large division between the south and the north in MSEA. The Cham population fell within the group of populations in southern MSEA and furthermore positioned more closely to the Mon–Khmer populations such as Chong, Khmer, and Cambodian from southern Thailand and Cambodia. Haplogroups B4c2, M51, N21, R22, and M* were found to contribute most to the southern pole of MSEA. In contrast, the Kinh, together with other Vietnamese populations converged into a northern MSEA group, with high frequency of the haplogroups D*, B4*, G, M7c1*, and A. The south–north patterning in MSEA seen in PCA was small, but significant, in AMOVA ($P < 0.05$).

## Dissecting the mtDNA Variation in the Chams

To trace the recent gene flow from ISEA, we have dissected the matrilineal pool of the Chams at the haplotype level

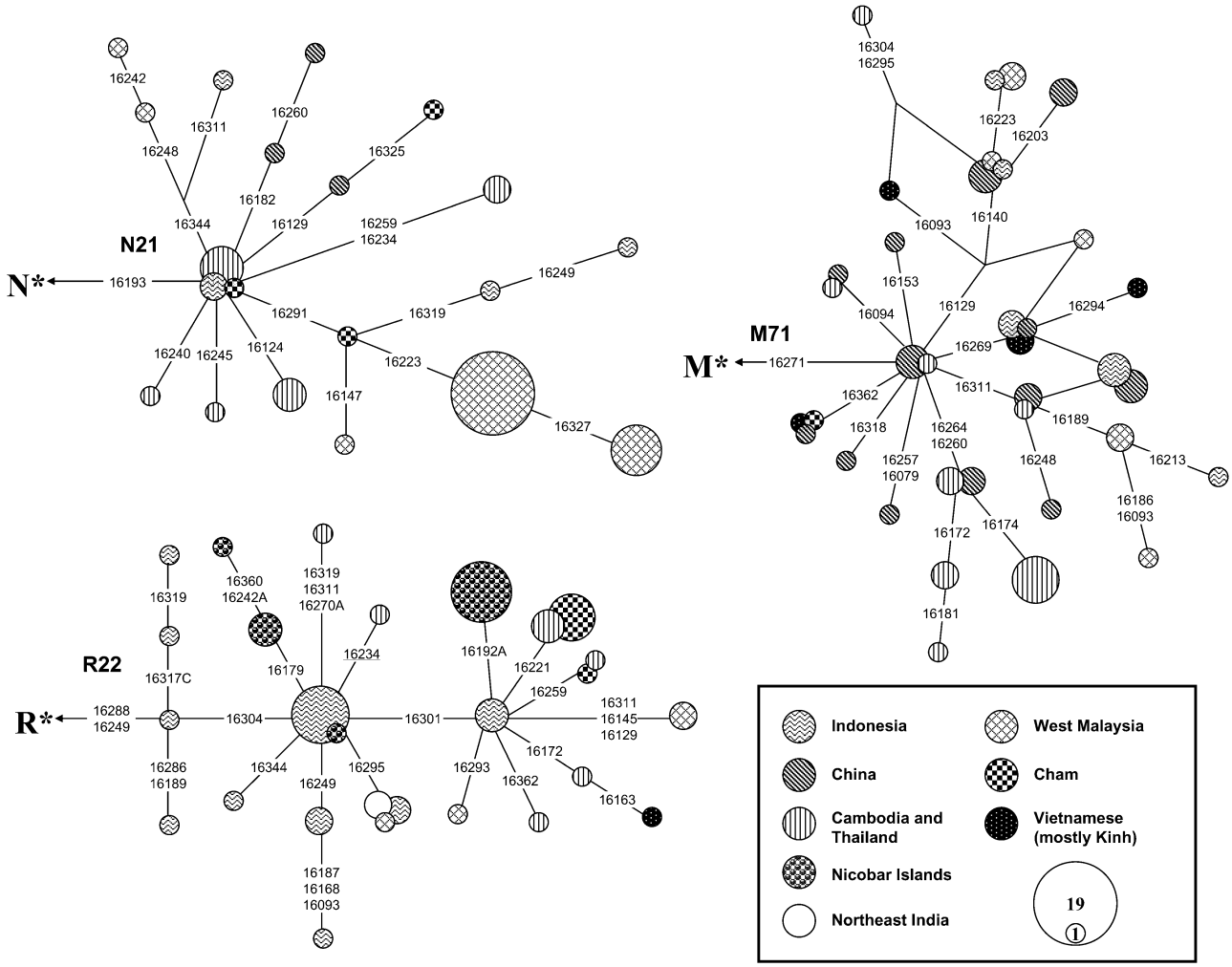


**Fig. 3.** (Continued).

and compared the Cham mtDNA haplotypes to more than 3,000 HVS-I sequences from ISEA (Melton et al. 1998; Tajima et al. 2004; Trejaut et al. 2005; Hill et al. 2006, 2007; Wong et al. 2007; Soares et al. 2008; Maruyama et al. 2010; Tabbada et al. 2010). Twenty-eight Cham haplotypes (16080-16370) belonging to different (sub-)haplogroups have identical counterparts in ISEA (fig. 5). Among them, eight haplotypes (consisting of 17 sequences; ~10.1%, 17/168) belonging to haplogroups B5a, D, E, M17, M50, M51, and R23 were not observed in a number of 10,572 samples from China (9,633) and MSEA (939) but shared exclusively between the Chams and some ISEA populations (fig. 5), thus indicative of certain direct and recent genetic links with ISEA. To control for the possibility of recent back migration from MSEA to ISEA, we performed the founder analysis with *f1* criterion (Richards et al. 2000) on the 28 haplotypes. Three haplotypes within haplogroups D, M7b, and M8a, which failed to relate to candidate founders in ISEA, were removed in further analyses (fig. 5). The haplotype (with motif 16147-16184A-16189-16217-16235) within haplogroup B4c2 was not considered as well because it likely represents a more ancient migration event (see below). As a result, 24 haplotypes consisting of 51 sequences (~30.4%, 51/168) in the Chams shared commonality with ISEA populations (table 2 and fig. 5), thus potentially identifying recent gene flow from ISEA. Then, the $D_{HS}$ distances between the Chams and other ISEA populations were calculated based on haplotype-sharing analyses (table 2). Samples from Lombok had the lowest $D_{HS}$ distance (0.850); the highest value (0.959) was observed in Formosan samples from Taiwan.

## Potential Marker for a Postglacial Dispersal

Haplogroup B4c2, first defined by Tanaka et al. (2004), was found with relatively high frequency (~10.1%, 17/168) in the Chams. This haplogroup is distributed widely in southern China and Southeast Asia and reaches a peak frequency (~15–16%) in Cambodia and its neighboring area in Thailand (fig. 6). In spite of close geographic distance between southern Vietnam and Cambodia, haplogroup B4c2 in the Chams was unlikely the result of recent gene flow from Cambodia and Thailand because most B4c2 types (9/13) from both countries were located in different branches with the Cham-specific haplotypes (fig. 7). The reduced–median network of B4c2 presented an obvious division (as distinguishable by transversion 16184A) between the lineages from ISEA and the MSEA as well as

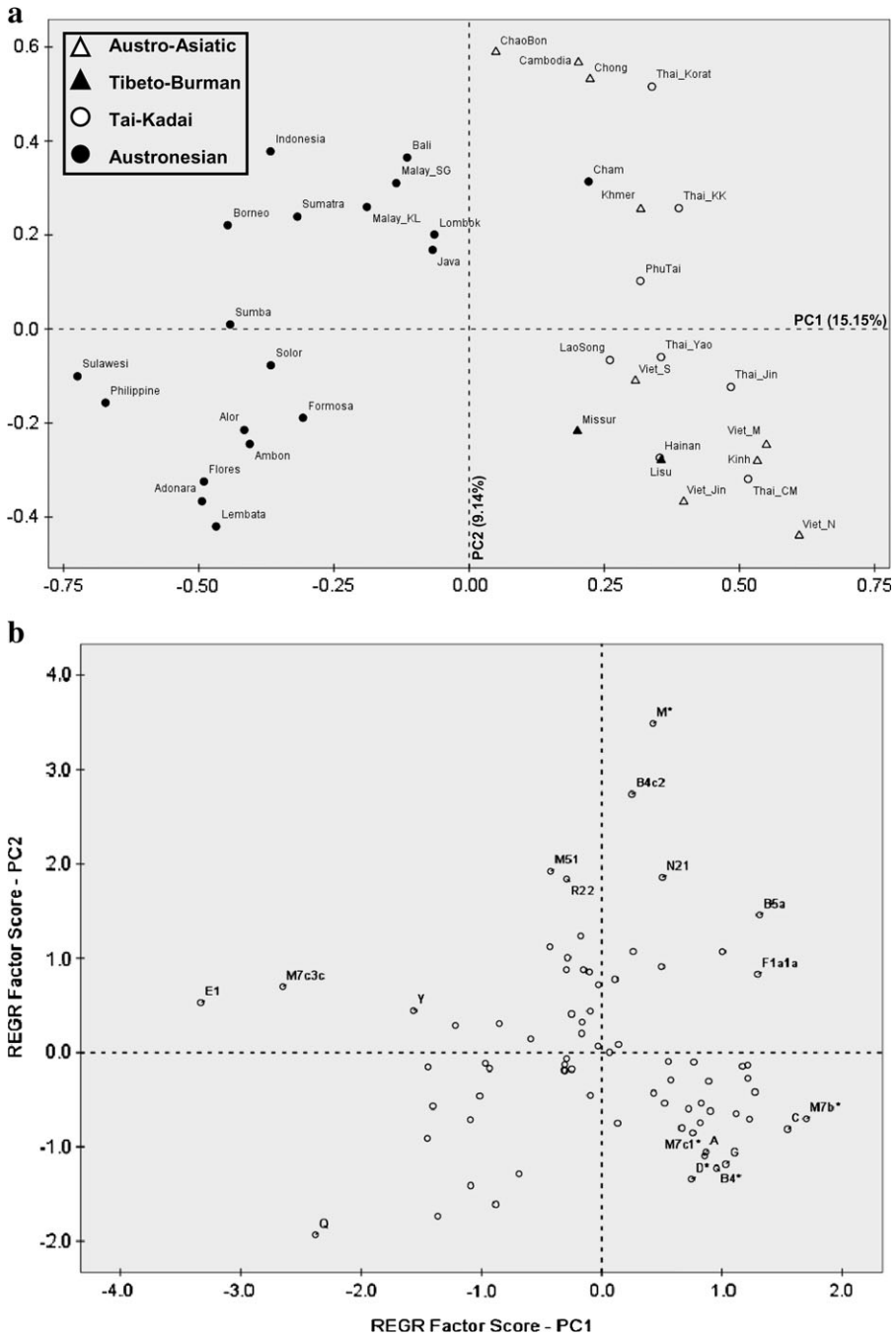


**Fig. 4.** PCA of populations in Southeast Asia (table 1). (*a*) PC map of populations based on mtDNA haplogroup frequencies. The original absolute frequencies (supplementary table S4, Supplementary Material online) were transformed as Richards et al. (2002) suggested to standardize against the different effect of genetic drift on haplogroups of different frequencies. Sumatra: Medan, Padang, Pekanbaru, Palembang, and Bangka; Borneo: Kota Kinabaru and Banjarmasin; Sulawesi: Ujung Padang, Palu, Manado, and Toraja. (*b*). Plot of the haplogroup contribution of the first and second PC. The contribution of each haplogroup was calculated as the factor scores for PC1 and PC2 with regression (REGR) method in SPSS13.0 software (SPSS).

southern China. Meanwhile, the starlike phylogeny of the branch characteristic of 16184A suggests that it underwent a population expansion dating to the beginning Holocene (fig. 7), which was compatible with the results based on the mtDNA genomes (fig. 2). During this period, most of the Sunda Shelf region was submerged because of the rise of the sea level which then formed the geographic division between MSEA and ISEA known as the current Gulf of Thailand (Hanebuth et al. 2000; Lambeck and Chappell 2001; Sathiamurthy and Voris 2006). As this eustatic change due to climatic oscillation was suggested to play an important role in shaping the modern maternal pools of populations in ISEA (Hill et al. 2007; Soares et al. 2008; Karafet et al. 2010), our results were further examined to test for this influence. It is evident that this change had also affected the maternal pools in MSEA and southern China,

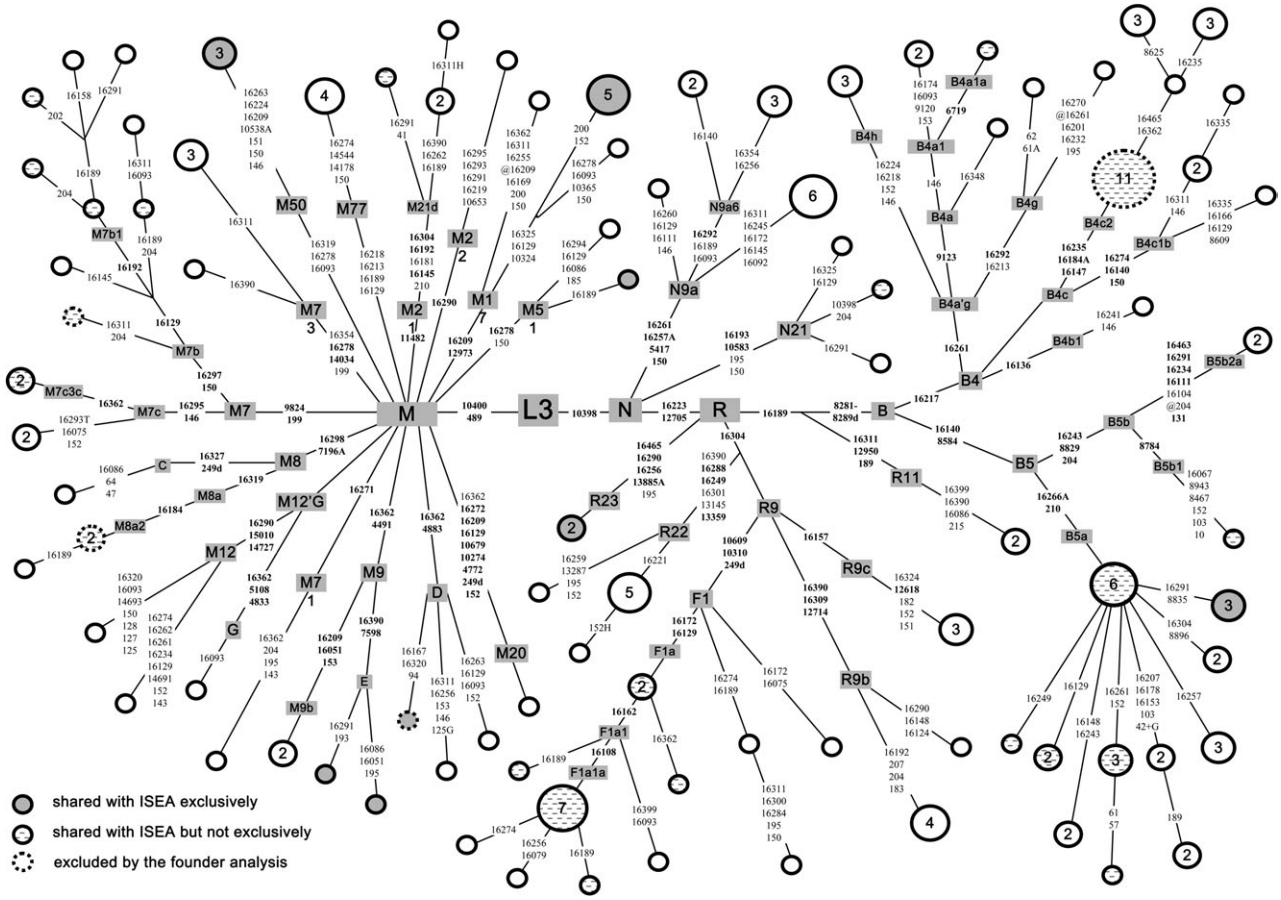


FIG. 5. Network profile of the 91 mtDNA haplotypes observed in the 168 Chams. This tree was constructed manually by comparison with the available mtDNA data sets and the basal East Asian and Southeast Asian mtDNA classification trees (Kong et al. 2006; Hill et al. 2007). Diagnostic sites genotyped in this work are indicated in bold. Mutations are transitions unless the base change is explicitly indicated. Insertions are suffixed with a plus sign (+) and the inserted nucleotide and deletions have a "d" suffix. Heteroplasmic positions are indicated by a "H" after the nucleotide position. Back mutations are indicated by a "@" before the nucleotide position. The A–C transversions at nps 16181, 16182, and 16183 were ignored in interpopulation analyses.

an observation in agreement with previous suggestions (Wen et al. 2005; Ricaut et al. 2006).

## Discussion

The mtDNA phylogeny in Southeast Asia reconstructed in the present study (fig. 2) helps identify phylogenetic status of all mtDNAs from 168 Cham and 139 Kinh individuals. Most previously uncharacterized mtDNA lineages in the Chams and Kinhs could be allocated into the indigenous haplogroups (i.e., M17, M21, M22, M50, M51, M73, M77, N21, R22, and R23) in Southeast Asia. The distribution (supplementary table S3, Supplementary Material online; fig. 3) and age estimates (fig. 2) of these haplogroups show patterns of long-term in situ evolution (Macaulay et al. 2005; Hill et al. 2006, 2007; Tabbada et al. 2010). However, in the Chams, most previously uncharacterized lineages were found to be shared with the other populations in Southeast Asia or located at the tips of the networks with the derived states (fig. 3). These patterns suggest that these lineages in the Chams were likely to be introduced by the other populations, especially in southern MSEA (Thailand and Cambodia) and ISEA, via recent gene flow.

Although the Chams show tight links with Austronesian speakers from ISEA in language and culture, analysis of the Cham mtDNA variation revealed that the genetic links between the Chams and the ISEA populations were much weaker. To illustrate, the ISEA characteristic and prevailing haplogroup E (Hill et al. 2007; Soares et al. 2008) was only detected in two Cham individuals. Likewise, except for haplogroup M7c3c (~1.2%, 2/168 in the Chams), the other "Out-of-Taiwan" candidate lineages (e.g., D5, M7b3, Y2, F1a3, and F1a4) (Hill et al. 2007; Tabbada et al. 2010) that were distributed widely in ISEA were not observed in the Chams. As a result, the mtDNA profiles of the Chams showed significant difference with populations in ISEA, an observation in accordance with the PCA clustering results: the Cham displayed a closer relationship with populations (especially the Mon–Khmers) in southern MSEA rather than with the Austronesian populations from ISEA (fig. 4).

The discordance between the linguistic and the genetic evidence in the Chams indicates that the Austronesian

**Table 2.** The $D_{HS}$ distances based on haplotype sharing between the Cham and other Austronesian populations from ISEA.

| Halogroup | Haplotype HVS-I (16,000 +)[a] | Cham | Philippine | Formosa | Java | Bali | Lombok | Sumba | Alor | Ambon | Orang-Asli | Indonesia | Malay_KL | Malay_S | Medan | Pekanbaru | Bangka | Padang | Palembang | Banjarmasin | Kota Kinabalu | Manado | Palu | Ujung Padang | Toraja | Indonesia /East Malaysia |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B4a | 189 217 261 | 1 | 43 | 28 | | 1 | 2 | | 2 | | | 1 | 3 | | | 7 | | 2 | | 5 | 3 | 1 | | 4 | 3 | |
| B5a | 129 140 189 266A | 2 | | | | | | | | | | | | 1 | | | | | | | | | | | | |
| B5a | 140 189 249 266A | 1 | | | | | | | 1 | | | | | | | | | | | | | | | | | |
| B5a | 140 189 261 266A | 4 | | | | | | | | | | | 4 | 3 | | | | | 1 | | | | | 2 | | |
| B5a | 140 189 266A | 6 | 1 | | 1 | 4 | 3 | | | 2 | 3 | 3 | 4 | 7 | 2 | | | | 1 | 3 | 1 | 2 | | 1 | | |
| B5a | 140 189 266A 291 | 3 | | | | | | | 1 | | | | | | | | | | | | | | | | | |
| B5b | (067) 140 189 243 | 1 | 26 | | | | 1 | | | | | | | | 2 | | | | | 2 | | | 1 | 1 | 1 | |
| E1a | 223 291 362 390 | 1 | 38 | 24 | | 3 | | 1 | 2 | 2 | | 5 | 6 | 3 | | | 2 | 1 | | 1 | 2 | 15 | 3 | 5 | 10 | |
| E2 | (051) 086 223 362 (390) | 1 | | | | | | | | | | | | | | | | | | | | | | | | 1 |
| F1a | 129 172 304 | 2 | | | | | 1 | 1 | 1 | | | 2 | 5 | 4 | 1 | 1 | 1 | 2 | | | 1 | 1 | | 1 | | |
| F1a | 129 172 304 362 | 1 | 1 | | | | 1 | | | | | | | | | | | | | | | | | | | |
| F1a1 | 129 162 172 189 304 | 1 | | 6 | | | | | | | | | | | | | | | | | | | | | | |
| F1a1a | 108 129 162 172 189 304 | 1 | | | | | | | | | 4 | 1 | | 5 | | | | | | | | | | | | |
| F1a1a | 108 129 162 172 304 | 7 | | | 2 | 1 | 3 | 1 | 1 | 1 | 17 | 2 | 2 | 4 | 1 | 2 | 1 | | 1 | | | 1 | 1 | | 1 | |
| M21d | 145 181 192 223 291 304 | 1 | | | | 1 | | | | | | | | | | | | | | | | | | | | |
| M17 | 129 209 223 325 | 5 | | | | 1 | | | | | | | | | | | | | | | | | | | | |
| M51 | 189 223 278 | 1 | | | | | 1 | | | | | 1 | | 1 | | | | | 1 | | | 1 | 1 | | 1 | |
| M50 | 093 209 223 224 263 278 319 | 3 | | | | | | | | | | | | 3 | | 2 | | | | | | | | | | |
| M7b | 129 189 223 297 | 1 | | | | 1 | | | | 1 | | | | | 1 | | | | | | | | | | | |
| M7b1 | 129 189 192 223 297 | 1 | | | | | 1 | | | | | 1 | 3 | 1 | 3 | | | | | | | | 1 | | | |
| M7b1 | 129 192 223 297 | 2 | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| M7c3c | 223 295 362 | 2 | 37 | 20 | 5 | 3 | 1 | 3 | 1 | 1 | 8 | 3 | 6 | 9 | 1 | 3 | 1 | | 6 | 1 | 3 | 8 | 6 | 1 | 4 | |
| N21 | 193 223 | 1 | | | | 1 | | | | | | | | | | | | | | | | | 1 | | | |
| R23 | 256 290 (465) | 2 | | | | 3 | | 1 | | | | | | | | | | | | | | | | | | |
| | No. different haplotypes | 86 | 173 | 111 | 23 | 60 | 36 | 34 | 34 | 29 | 49 | 42 | 125 | 133 | 32 | 37 | 24 | 19 | 22 | 62 | 47 | 48 | 28 | 32 | 27 | — |
| | Population size | 168 | 543 | 718 | 46 | 82 | 44 | 50 | 45 | 43 | 280 | 54 | 183 | 205 | 42 | 52 | 34 | 24 | 28 | 89 | 68 | 89 | 38 | 46 | 64 | — |
| | $D_{HS}$ | | 0.945 | 0.959 | 0.943 | 0.861 | 0.850 | 0.915 | 0.882 | 0.911 | 0.940 | 0.857 | 0.923 | 0.898 | 0.879 | 0.917 | 0.926 | 0.941 | 0.905 | 0.931 | 0.923 | 0.908 | 0.892 | 0.862 | 0.892 | — |

[a] The fragment of HVS-I (16080-16370) was considered in the haplotype-sharing analyses, so the sites out of this fragment were noted in parentheses.

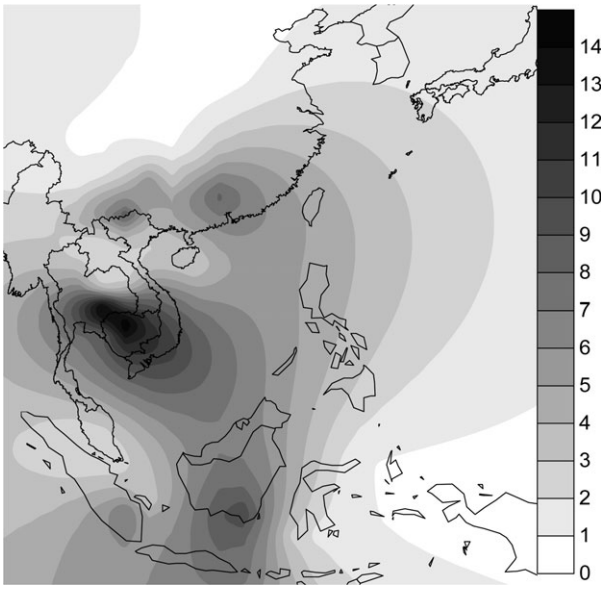


FIG. 6. Spatial frequency distribution of haplogroup B4c2. The figure was created by using the Kriging algorithm of the Surfer 8.0 package.

diffusion in MSEA cannot be simply explained as a demic diffusion. Given the fact that the Mon–Khmer speakers had already occupied the middle and southern part of Vietnam before the arrival of the Austronesian immigrants (Bellwood 2006, 2007), contact between both populations might have involved extensive genetic admixture. During the process of admixture, one expanding Austronesian language (proto-Chamic) was imposed on or adopted by certain immigrants from ISEA with only a relative minor genetic contribution from the expanding Austronesian people. Conversely, the indigenous Mon–Khmers—the major genetic donors to the Chams, only contributed minor linguistic components as loan words to the Chamic language (Thurgood 1999; Southworth 2004). Thus, the Austronesian diffusion in MSEA was mainly a process of language shift (Cavalli-Sforza et al. 1994; Diamond and Bellwood 2003) by indigenous populations.

When we focused on the potential recent Austronesian components in the Chams that were shared with the ISEA populations (fig. 5 and table 2), similar $D_{HS}$ values were observed across several Austronesian populations from ISEA. However, the populations (Kota Kinabalu and Banjarmasin) from Borneo—the potential source of Chamic (Blust 1994)—were relatively distant ($D_{HS}$ values: 0.923 and 0.931, respectively) from the Chams. Consequently, the Austronesian homelands of the Chams as well as the possible migration route of pioneer Austronesian people (and the language) to southern Vietnam cannot be pinpointed for the time being, although the bias caused by incomplete sampling of the "correct" source populations cannot be excluded. Taken together, our results support that cultural diffusion had played a dominate role during the spread of the Austronesian language into MSEA as suggested by the NMTCN hypothesis (Solheim et al. 2007). However, because our current study focuses merely on mtDNA, which reflects in fact the maternal history of the Cham people, extreme caution shall be taken during the explanation of our results. For instance, the paucity of Y-chromosome data in MSEA may hamper the deeper understanding of the origin of the Chams. In brief, we cannot exclude the possibility of existence of asymmetric sexual gene flow for a genetic link between the contemporary Cham and other Austronesian people may exist on the paternal side. Meanwhile, it also should be noted that potential minor gene flow entering MSEA was further diluted by admixture

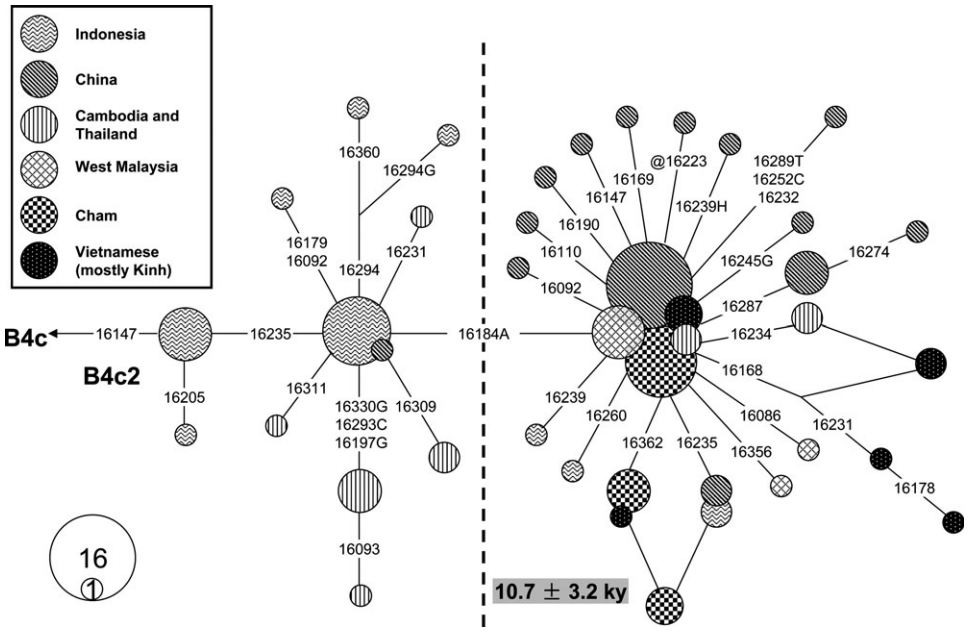


FIG. 7. Reduced median network of haplogroup B4c2 based on HVS-I sequences in the region 16080–16365. Labels described as above.

such as the one suggested with Mon–Khmer ancestral population to a level that may not be appreciable. In regards to this point, multidisciplinary data are essential to uncover the entire past history of the Cham population.

In addition, the comparison between the Chams and the Kinhs reveal a significant difference (PCA and AMOVA) in the maternal gene pools, which was consistent with the different ethnohistories of the Chams and the Kinhs that, respectively, represent the different cultures in southern and northern Vietnam. Southern Vietnam, which was historically a highly Indianized kingdom known as Champa, harbored a culture similar to that of the Khmer Empire. In contrast, northern Vietnam, known as Jiaozhi or Annam, was under Chinese domination for more than 1,000 years. Besides the deeply Sinicized culture, substantial Chinese immigration would make a hefty contribution to the modern gene pool of the Kinhs in northern Vietnam (Higham 2002; He 2006). Although masses of Cham people were suggested to have been assimilated into the Kinh group after the annihilation of Champa (He 2006), it seems that the remaining group of Chams has retained its own characteristics not only in culture but also in genetics, which well distinguish this unique population from the Kinhs, notwithstanding waves of Kinhs influxing into southern Vietnam in the past several hundred years.

## Supplementary Material

Supplementary tables S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

## Acknowledgments

We thank Chun-Ling Zhu, Gui-Mei Li, Wen-Zhi Wang, Yang Yang, Shi-Fang Wu, and Shi-Kang Gou for technical assistance. We are grateful to all volunteers for providing DNA samples. We thank the two anonymous referees for their valuable comments and Miss Diana Chen for language editing. This research was supported by grants from the Natural Science Foundation of Yunnan Province, National Natural Science Foundation of China (30900797 and 30621092), and the Chinese Academy of Sciences.

## References

Achilli A, Olivieri A, Pala M, et al. (22 co-authors). 2007. Mitochondrial DNA variation of modern Tuscans supports the near eastern origin of Etruscans. *Am J Hum Genet*. 80(4): 759–768.

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. *Nat Genet*. 23(2):147.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 16(1):37–48.

Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141(2):743–753.

Bandelt HJ, Parson W. 2008. Consistent treatment of length variants in the human mtDNA control region: a reappraisal. *Int J Legal Med*. 122(1):11–21.

Bellwood P. 2006. Austronesian prehistory in southeast Asia: homeland, expansion and transformation. In: Bellwood P, Fox JJ, Tryon D, editors. The austronesians: historical and comparative perspectives. Canberra (ACT): ANU E Press. p. 103–114.

Bellwood P. 2007. Prehistory of the Indo-Malaysian Archipelago. Canberra (ACT): ANU E Press.

Bellwood P, Fox JJ, Tryon D. 2006. The Austronesians in history: common origins and diverse transformations. In: Bellwood P, Fox JJ, Tryon D, editors. The Austronesians: historical and comparative perspectives. Canberra (ACT): ANU E Press. p. 1–14.

Black ML, Dufall K, Wise C, Sulliva S, Bittles AH. 2006. Genetic ancestries in northwest Cambodia. *Ann Hum Biol*. 33(5–6): 620–627.

Blust RA. 1994. The Austronesian settlement of mainland Southeast Asia. In: Adams KL, Hudak TJ, editors. Papers from the Second Annual Meeting of the Southeast Asian Linguistics Society, 1992. Tempe (AZ): Arizona State University. p. 25–83.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton (NJ): Princeton University Press.

Chikhi L, Nichols RA, Barbujani G, Beaumont MA. 2002. Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci U S A*. 99(17):11008–11013.

Dancause KN, Chan CW, Arunotai NH, Lum JK. 2009. Origins of the Moken Sea Gypsies inferred from mitochondrial hypervariable region and whole genome sequences. *J Hum Genet*. 54(2): 86–93.

Diamond J, Bellwood P. 2003. Farmers and their languages: the first expansions. *Science* 300(5619):597–603.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online*. 1:47–50.

Fendt L, Zimmermann B, Daniaux M, Parson W. 2009. Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. *BMC Genomics*. 10:139.

Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of native American mtDNA variation: a reappraisal. *Am J Hum Genet*. 59(4):935–945.

Fucharoen G, Fucharoen S, Horai S. 2001. Mitochondrial DNA polymorphisms in Thailand. *J Hum Genet*. 46(3):115–125.

Gan RJ, Pan SL, Mustavich LF, et al. (13 co-authors). 2008. Pinghua population as an exception of Han Chinese's coherent genetic structure. *J Hum Genet*. 53(4):303–313.

Hanebuth T, Stattegger K, Grootes PM. 2000. Rapid flooding of the Sunda Shelf: a late-glacial sea-level record. *Science* 288(5468): 1033–1035.

Hartmann A, Thieme M, Nanduri LK, Stempfl T, Moehle C, Kivisild T, Oefner PJ. 2009. Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum Mutat*. 30(1):115–122.

He P. 2006. The origin and evolution of nationalities in the Indochina Peninsula. Beijing (China): The Ethnic Publishing House.

Higham C. 1989. The archaeology of mainland Southeast Asia: from 10,000 BC to the fall of Angkor. Cambridge (UK): Cambridge University Press.

Higham C. 2002. Early cultures of Mainland Southeast Asia. Bangkok (Thailand): River Books.

Hill C, Soares P, Mormina M, et al. (12 co-authors). 2006. Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol Biol Evol*. 23(12):2480–2491.

Hill C, Soares P, Mormina M, et al. (11 co-authors). 2007. A mitochondrial stratigraphy for Island Southeast Asia. *Am J Hum Genet*. 80(1):29–43.

Hung HC, Iizuka Y, Bellwood P, Nguyen KD, Bellina B, Silapanth P, Dizon E, Santiago R, Datan I, Manton JH. 2007. Ancient jades

map 3,000 years of prehistoric exchange in Southeast Asia. *Proc Natl Acad Sci U S A*. 104(50):19745–19750.

Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature*. 408(6813):708–713.

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. *Int J Legal Med*. 122(3):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(1):e4210.

Karafet TM, Hallmark B, Cox MP, Sudoyo H, Downey S, Lansing JS, Hammer MF. 2010. Major east-west division underlies Y chromosome stratification across Indonesia. *Mol Biol Evol*. Advance Access published March 5, 2010, doi:10.1093/molbev/msq063.

Kivisild T, Tolk HV, Parik J, Wang YM, Papiha SS, Bandelt HJ, Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol*. 19(10):1737–1751.

Kong QP, Bandelt HJ, Sun C, et al. (12 co-authors). 2006. Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet*. 15(13):2076–2086.

Kong QP, Bandelt HJ, Zhao M, Zhang YP, Yao YG. 2010. Reply to van Oven: suggestions and caveats for naming mtDNA haplogroup. *Proc Natl Acad Sci U S A*. 107(11):E40-E41.

Kong QP, Salas A, Sun C, Fuku N, Tanaka M, Zhong L, Wang CY, Yao YG, Bandelt HJ. 2008. Distilling artificial recombinants from large sets of complete mtDNA genomes. *PLoS One*. 3(8):e3016.

Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, Zhang YP. 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet*. 73(3):671–676.

Lam MD. 2009. Sa Huynh regional and inter-regional interactions in the Thu Bon valley, Quang Nam province, central Vietnam. *IPPA Bull*. 29:68–75.

Lambeck K, Chappell J. 2001. Sea level change through the last glacial cycle. *Science*. 292(5517):679–686.

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. *Am J Phys Anthropol*. 137(4):425–440.

Li H, Cai XY, Winograd-Cort ER, et al. (12 co-authors). 2007. Mitochondrial DNA diversity and population differentiation in Southern East Asia. *Am J Phys Anthropol*. 134(4):481–488.

Li H, Wen B, Chen SJ, et al. (19 co-authors). 2008. Paternal genetic affinity between western Austronesians and Daic populations. *BMC Evol Biol*. 8:146.

Macaulay V, Hill C, Achilli A, et al. (21 co-authors). 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308(5724):1034–1036.

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. *Int J Legal Med*. 124(2):165–170.

Melton T, Clifford S, Martinson E, Batzer M, Stoneking M. 1998. Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *Am J Hum Genet*. 63(6):1807–1823.

Mona S, Grunz KE, Brauer S, et al. (11 co-authors). 2009. Genetic admixture history of eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. *Mol Biol Evol*. 26(8):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. *Am J Phys Anthropol*. 118(2):146–153.

Prasad BVR, Ricker CE, Watkins WS, Dixon ME, Rao BB, Naidu JM, Jorde LB, Bamshad M. 2001. Mitochondrial DNA variation in Nicobarese Islanders. *Hum Biol*. 73(5):715–725.

Qian YP, Chu ZT, Dai Q, Wei CD, Chu JY, Tajima A, Horai S. 2001. Mitochondrial DNA polymorphisms in Yunnan nationalities in China. *J Hum Genet*. 46(4):211–220.

Reddy BM, Langstieh BT, Kumar V, Nagaraja T, Reddy ANS, Meka A, Reddy AG, Thangaraj K, Singh L. 2007. Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. *PLoS One*. 2(11):e1141.

Ricaut FX, Bellatti M, Lahr MM. 2006. Ancient mitochondrial DNA from Malaysian hair samples: some indications of southeast Asian population movements. *Am J Hum Biol*. 18(5):654–667.

Richards M, Macaulay V, Hickey E, et al. (37 co-authors). 2000. Tracing European founder lineages in the near eastern mtDNA pool. *Am J Hum Genet*. 67(5):1251–1276.

Richards M, Macaulay V, Torroni A, Bandelt HJ. 2002. In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet*. 71(5):1168–1174.

Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet*. 67(3):718–726.

Sathiamurthy E, Voris HK. 2006. Maps of Holocene sea level transgression and submerged lakes on the Sunda Shelf. Nat Hist J Chulalongkorn Univ. (Suppl 2):1–43.

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet*. 84(6):740–759.

Soares P, Trejaut JA, Loo JH, et al. (14 co-authors). 2008. Climate change and postglacial human dispersals in Southeast Asia. *Mol Biol Evol*. 25(6):1209–1218.

Solheim WG, Bulbeck D, Flavel A. 2007. Archaeology and culture in Southeast Asia: unraveling the Nusantao. Quezon City (Philippines): University of the Philippines Press.

Southworth WA. 2004. The coastal states of Champa. In: Glover I, Bellwood P, editors. Southeast Asia: from prehistory to history. London: RoutledgeCurzon. p. 209–233.

Sykes B, Leiboff A, Lowbeer J, Tetzner S, Richards M. 1995. The origins of the Polynesians—an interpretation from mitochondrial lineage analysis. *Am J Hum Genet*. 57(6):1463–1475.

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? *Mol Biol Evol*. 27(1):21–31.

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. *J Hum Genet*. 49(4):187–193.

Tanaka M, Cabrera VM, González AM, et al. (28 co-authors). 2004. Mitochondrial genome variation in Eastern Asia and the peopling of Japan. *Genome Res*. 14(10A):1832–1850.

Thurgood G. 1999. From ancient Cham to modern dialects: two thousand years of language contact and change. Honolulu (HI): University of Hawaii Press.

Thurgood G. 2005. A preliminary sketch of Phan Rang Cham. In: Adelaar KA, Himmelmann N, editors. The Austronesian languages of Asia and Madagascar. London: Routledge. p. 489–512.

Tofanelli S, Bertoncini S, Castri L, Luiselli D, Calafell F, Donati G, Paoli G. 2009. On the origins and admixture of Malagasy: new evidence from high resolution analyses of paternal and maternal lineages. *Mol Biol Evol*. 26(9):2109–2124.

Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, Li ZY, Lin M. 2005. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol*. 3(8):1362–1372.

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. *J Hum Genet*. 51(3):217–226.

van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*. 30(2):E386–E394.

Wang HW, Jia XY, Ji YL, Kong QP, Zhang QJ, Yao YG, Zhang YP. 2008. Strikingly different penetrance of LHON in two Chinese families with primary mutation G11778A is independent of mtDNA haplogroup background and secondary mutation G13708A. *Mutat Res*. 643(1–2):48–53.

Wang WZ, Wang CY, Cheng YT, Xu AL, Zhu CL, Wu SF, Kong QP, Zhang YP. 2010. Tracing the origins of Hakka and Chaoshanese by mitochondrial DNA analysis. *Am J Phys Anthropol*. 141(1):124–130.

Wen B, Li H, Gao S, Mao XY, et al. (18 co-authors). 2005. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Mol Biol Evol*. 22(3):725–734.

Wen B, Li H, Lu DR, et al. (18 co-authors). 2004. Genetic evidence supports demic diffusion of Han culture. *Nature* 431(7006): 302–305.

Wen B, Xie XH, Gao S, et al. (13 co-authors). 2004. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet*. 74(5):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. *Leg Med (Tokyo)*. 9(1):33–37.

Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet*. 70(3):635–651.

Yao YG, Kong QP, Man XY, Bandelt HJ, Zhang YP. 2003. Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. *Mol Biol Evol*. 20(2):214–219.

Yao YG, Kong QP, Salas A, Bandelt HJ. 2008. Pseudomitochondrial genome haunts disease studies. *J Med Genet*. 45(12):769–772.

Yao YG, Kong QP, Wang CY, Zhu CL, Zhang YP. 2004. Different matrilineal contributions to genetic structure of ethnic groups in the Silk Road region in China. *Mol Biol Evol*. 21(12):2265–2280.

Yao YG, Nie L, Harpending H, Fu YX, Yuan ZG, Zhang YP. 2002. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am J Phys Anthropol*. 118(1): 63–76.

Yao YG, Salas A, Logan I, Bandelt HJ. 2009. mtDNA data mining in GenBank needs surveying. *Am J Hum Genet*. 85(6):929–933.

Yao YG, Zhang YP. 2002. Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: new data and a reappraisal. *J Hum Genet*. 47(6):311–318.

Zhao M, Kong QP, Wang HW, et al. (15 co-authors). 2009. Mitochondrial genome evidence reveals successful late Paleolithic settlement on Tibetan Plateau. *Proc Natl Acad Sci U S A*. 106(50):21230–21235.

Zimmermann B, Bodner M, Amory S, Fendt L, Rock A, Horst D, Horst B, Sanguansermsri T, Parson W, Brandstatter A. 2009. Forensic and phylogeographic characterization of mtDNA lineages from northern Thailand (Chiang Mai). *Int J Legal Med*. 123(6):495–501.