



DA DẠNG DI TRUYỀN VÀ NGUỒN GỐC NGƯỜI VIỆT NAM

Nông Văn Hải

**Viện Nghiên cứu hệ gen
Viện Hàn lâm Khoa học và Công nghệ Việt Nam**
vhnong@igr.ac.vn

Hà Nội, 12-6/2020

Đế Minh là cháu ba đời của vua **Thần Nông**, đi tuần thú phương Nam, núi Ngũ Lĩnh (nay thuộc tỉnh Hồ Nam, Trung Quốc) gặp một nàng tiên, lấy u đẻ ra người con tên là **Lộc Tục**. Sau Đế Minh truyền ngôi lại cho con trưởng **Ế Nghi** làm vua phương Bắc (từ núi Ngũ Lĩnh về phía Bắc), phong cho Lộc làm vua phương Nam (từ núi Ngũ Lĩnh về phía Nam), xưng là **Kinh Dương**, quốc hiệu là **Xích Quỷ**.

Kinh Dương Vương làm vua **Xích Quỷ** vào năm **Nhâm Tuất 2879 TCN**, con gái **Đông Đình Hồ quân** (còn có tên là **Thần Long**) là **Long nữ** sinh **Long Lâm**, nối ngôi làm vua, xưng là **Lạc Long Quân**. **Lạc Long Quân lấy Âu** (một nàng tiên ở phương Bắc), sinh một lần trăm người con.

Một hôm, Lạc Long Quân bảo Âu Cơ rằng: "Ta là giống **rồng**, nàng là **tiên**, thủy hỏa khắc nhau, chung hợp thật khó". Bèn từ biệt nhau, chia 50 con theo mẹ về núi, 50 con theo cha về ở miền biển (có bản chép là về **Nam Hải**), phong cho con trưởng làm **Hùng Vương**, nối ngôi vua.

(Theo **Đại Việt Sử ký Toàn**

TRUYỀN THUYẾT



Về sau, **Xích Quỷ** chia ra thành những nước nhỏ, gọi là **Bách Việt**. Đầu **thời kỳ đồ đồng**: **15 nhóm Lạc Việt** khác nhau vùng núi **miền Bắc** & **châu thổ sông Hồng** **12 nhóm Âu Việt** sống ở **vùng Đông Bắc**, **một số nhóm người khác** sinh sống trên các lưu vực sông Thanh - Nghệ - Tĩnh ngày nay. Để trao đổi buôn bán, phòng chống lụt lội, chống lại kẻ thù... các bộ lạc Lạc Việt dần hình thành một nước lấy tên **Văn Lang**, người đứng đầu tự xưng là **Hùng Vương**. Có 18 đời Hùng Vương cai trị ở thời đại Hồng Bàng (đến năm **258 TCN**). Các bằng chứng **khảo cổ học** như **trống đồng Đông Sơn** được tìm thấy ở **miền bắc Việt Nam** có cùng niên đại với thời kỳ **Hồng Bàng**, thể hiện một nền **văn hóa đồ đồng** rất phát triển (**văn hóa Đông Sơn**).

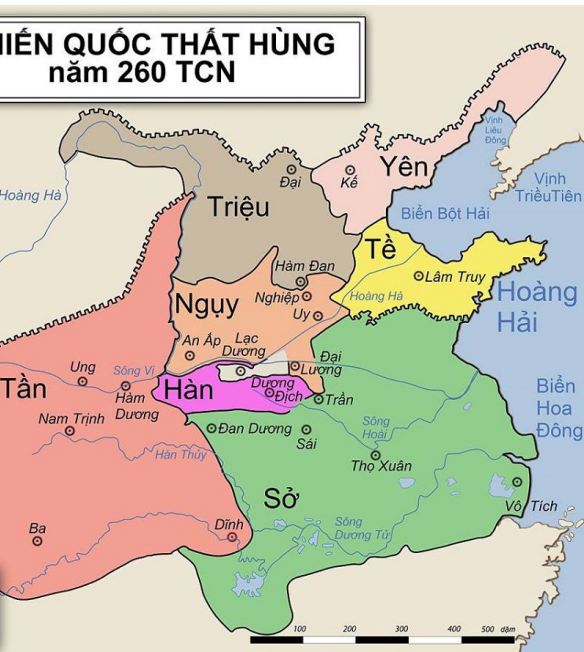


BẢN ĐỒ TRUNG QUỐC VÀ BẮC VIỆT NAM THỜI XUÂN THU, CHIẾN QUỐC, TẦN, HÁN

Trung Nguyên cuối thời Xuân Thu (thế kỷ 5 TCN)



482



DÂN SỐ VIỆT NAM QUA CÁC THỜI KỲ

Thời kỳ Văn Lang - Âu Lạc

Thời Văn Lang: ~500 ngàn (nhiều bộ tộc)

Âu Lạc (An Dương Vương (người Tày? Tai-Kadai?): ~70-80 vạn, bao gồm 2 tộc người: Âu Việt (Tày-Nùng cổ?), Lạc Việt (Việt-Mường cổ) và 10 vạn tù binh Tần bị đày đi khai hoang các miền xa xôi)

Thời kỳ Bắc thuộc

Nước Nam Việt (Triệu Đà): 2 quận Giao Chỉ và Cửu Chân: ~40 vạn (đa số người Bách Việt cũ và 10 vạn người Hán di cư), không tính các quận khác thuộc đất TQ ngày nay

Đất Vũ Đế (Tây Hán): Giao Chỉ ~75 vạn; Cửu Chân ~17 vạn; Nhật Nam 7 vạn, tổng ~99 vạn

Bà Trưng: ~1 triệu

Đất Hán: Giao Chỉ: ~70 vạn, Cửu Chân: ~21 vạn; Nhật Nam ~14 vạn, tổng ~1,05 triệu

Đường: Chỉ thấy ghi số hộ dân, chưa thấy ghi số dân

Thời kỳ độc lập và tự chủ

Đầu Nguyên: ~2 triệu

Đầu Minh: ~1,5 triệu

Đầu Đại Việt: ~2,2 triệu

Đầu Lê: ?

Đầu Nguyễn: ~4-5 triệu

Đầu Trung Hưng (1750): Đàng trong ~1,5 triệu, đàng ngoài ~4 triệu, tổng 5,5 triệu

Đầu Nguyễn: Gia Long ~5,7 triệu, Minh Mạng ~7,8 triệu

Đầu Pháp (Pháp thuộc): ~10 triệu

Hiện đại

1955/1976: Miền Bắc ~30 triệu; Miền Nam ~19 triệu, tổng ~49 triệu

Hiện nay: >96 triệu

(Nguồn: Wikipedia)

“Thế là từ đây Đông Nê châu, 83 quận, 313 huyện, hộ khẩu 52 vạn 3 ngàn, quan lại 3 vạn 2 ngàn, quân 23 vạn, trai gái già cả thấy 230 vạn, thóc 280 vạn斛, thuyền hàng nghìn chiếc, cung nữ hàng nghìn người đều về cả Đại Tấn”. Như vậy, toàn dân già trẻ gái trai hơn có 3 triệu!

(Tam Quốc Diễn Nghĩa)

NGŨ HỆ NAM Á (AUSTROASIATIC LANGUAGES, AA)

Nhóm Việt-Mường (Vietic), 4 dân tộc: Kinh, Mường, Thổ, Chứt
Nhóm Mon-Khmer, 21 dân tộc: Ba Na, Brâu, Bru-Vân Kiều, Chơ Ro, Co, Cơ Ho, Cơ Tu, Giẻ Triêng, Hrê, Kháng, Kơ Me, Kơ Mú, Mạ, Mảng, M'Nông, Ơ Đu, Rơ Măm, Tà Ôi, Xinh Mun, Xơ Đăng, X'Tiêng

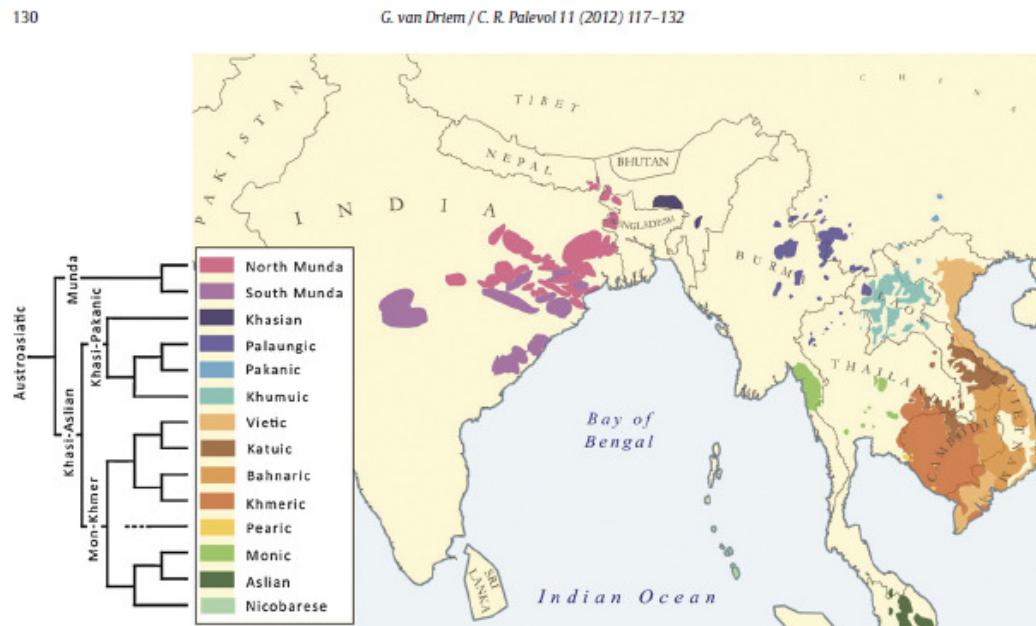
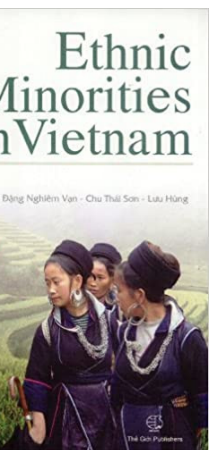
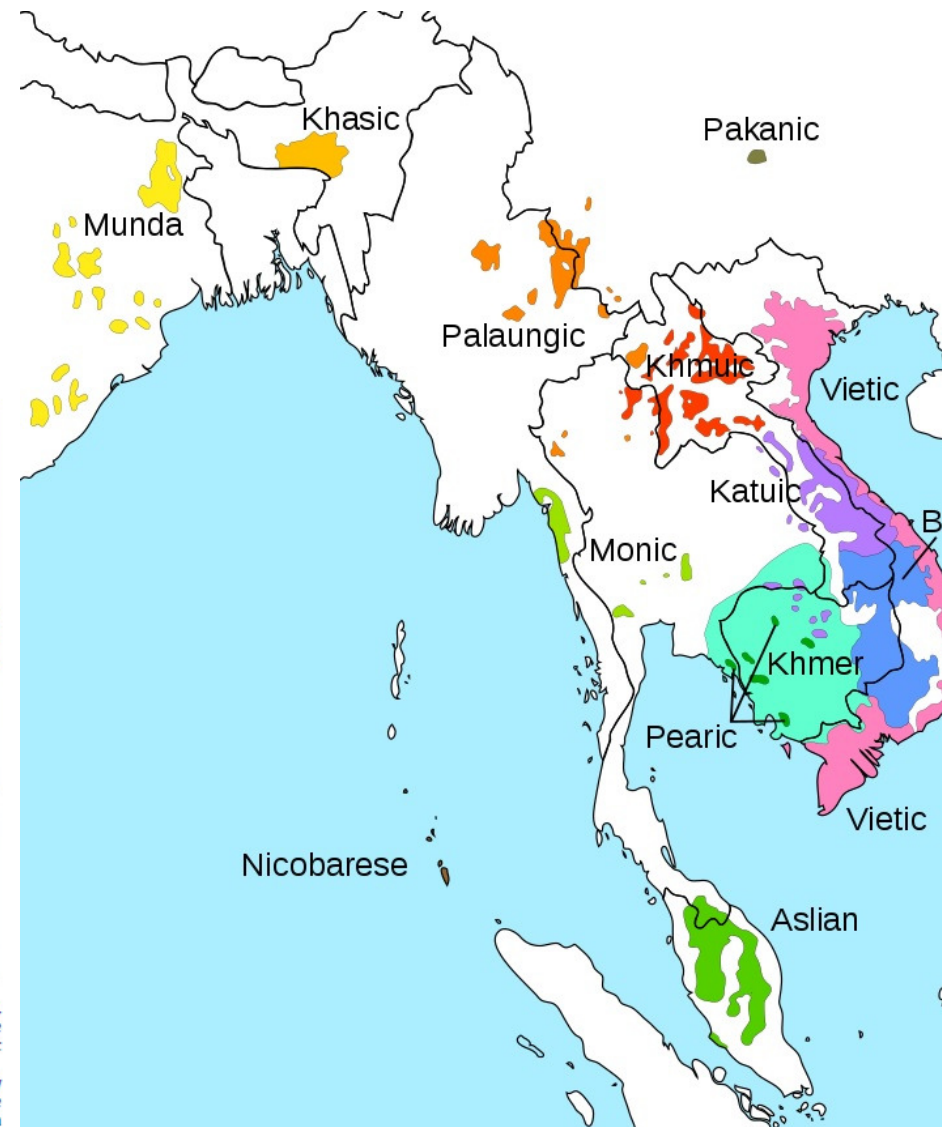


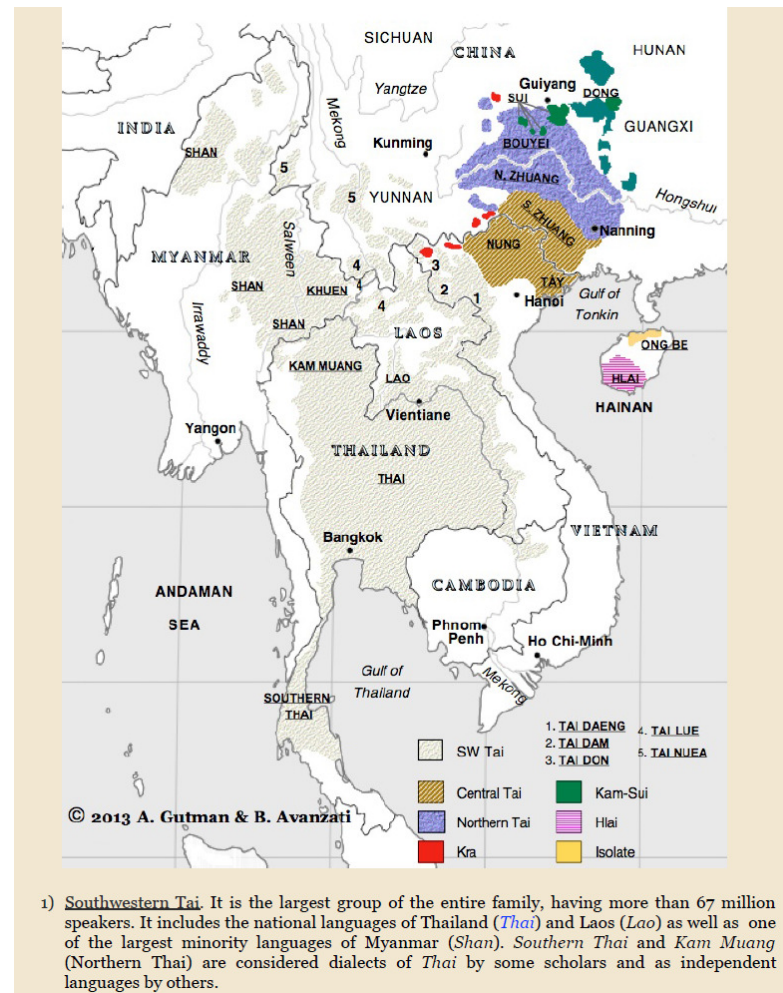
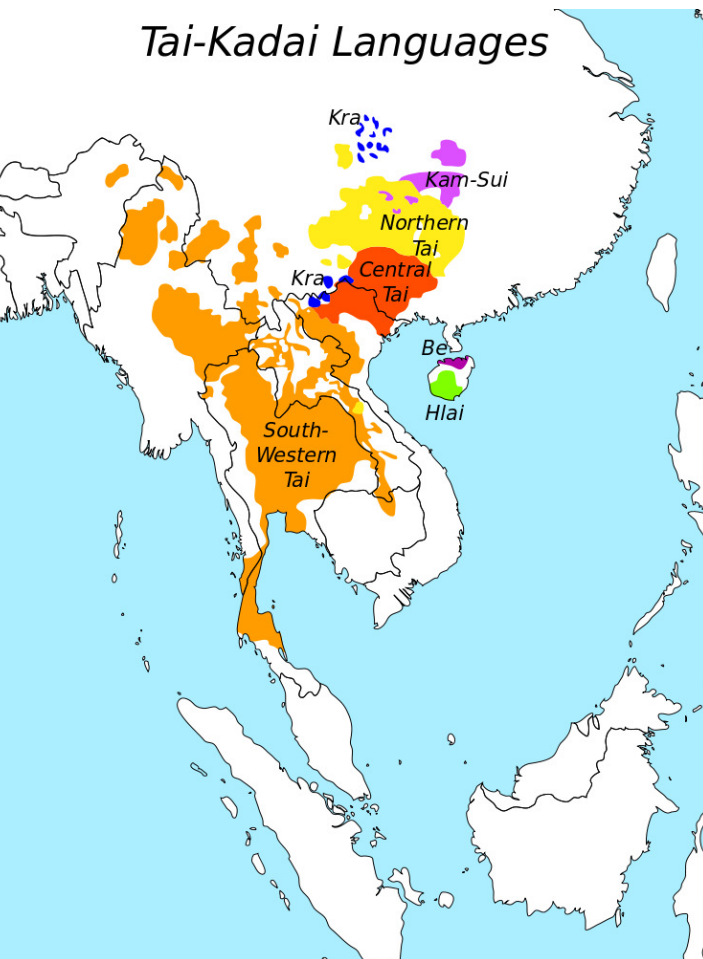
Fig. 19. Gérard Diffloth's revised (2009) Austroasiatic phylogeny and the geographical distribution of Austroasiatic languages, adapted from Chaubey et al. (2010). The two trunks of the Austroasiatic phylum are Munda, in Eastern, Northeastern and Central India, and Khasi-Aslian, which stretches from the Meghalaya in the northeast of the subcontinent to the Nicobars, the Malay peninsula and the Mekong delta in Southeast Asia. The precise phylogenetic position of Pearic within Mon-Khmer remains uncertain.

Fig. 19. Phyllogénie austroasiatique révisée par Gérard Diffloth (2009) et distribution géographique des langues austroasiatiques, adaptées de Chaubey et al. (2010). Les deux branches du phylum austroasiatique sont Munda, à l'est, au nord-est et au centre de l'Inde, et Khasi-Aslian, qui s'étend depuis le Meghalaya dans le nord-est du sous-continent jusqu'aux îles nicobaraises, la péninsule malaise et le delta du Mékong dans le sud-est de l'Asie. La position phyllogénétique précise du Pearic parmi le Mon-Khmer reste incertaine.



NGŨ' HỆ TAI – KADAI (TAI-KADAI LANGUAGES, TK)

Nhóm Tày Thái, 8 dân tộc: BỐ Y, Giáy, Lào, Lự, Nùng, Sán Chay, Tày, Thái
Nhóm Kadai (Kra), 4 dân tộc: Cơ Lao, La Chí, La Ha, Pu Páo



Speakers. The total number of native speakers of Tai-Kadai languages is around 93 million, which more than 67 million in the Southwestern Tai group alone. Most speakers live in Thailand (around 57 million) and the next largest number in China (about 24 million). Smaller numbers reside in northern Vietnam, Laos and Myanmar. The largest Tai-kadai languages are the Thai languages (speakers in millions):

Language	Speakers	Branch	Group	Area
Thai ¹	56.0	Tai	SW	Thailand
Northern Zhuang	10.5	Tai	North	S. China
Southern Zhuang	5.5	Tai	Central	S. China
Lao	4.3	Tai	SW	Laos
Shan	4.0	Tai	SW	North and East Myanmar
Bouyei	3.0	Tai	North	S. China
Dong	1.6	Kam Sui	-----	S. China
Tày	1.5	Tai	Central	N. Vietnam
Tai Dam	0.9	Tai	SW	N. Vietnam
Tai Lue	0.9	Tai	SW	S. China, Myanmar, Laos
Nung	0.9	Tai	Central	N. Vietnam
Tai Nuea	0.7	Tai	SW	S. China
Hlai (Li)	0.7	Hlai	-----	SE. China (Hainan Island)
Ong Be	0.6	Isolate	-----	SE. China (Hainan Island)

1. Thai speakers include here those of Thai proper plus those of Southern Thai and Kam Muang (Northern Thai).

- 1) **Southwestern Tai.** It is the largest group of the entire family, having more than 67 million speakers. It includes the national languages of Thailand (*Thai*) and Laos (*Lao*) as well as one of the largest minority languages of Myanmar (*Shan*). *Southern Thai* and *Kam Muang* (Northern Thai) are considered dialects of *Thai* by some scholars and as independent languages by others.

NGỮ HỆ H'MONG – MIỀN (H'MONG-MIEN LANGUAGES, HM)

nhóm H'Mong-Dao, 3 dân tộc: Dao, H'Mong (Mông), Pà Thẻn

Hmong–Mien languages



120

G. van Driem / C. R. Palevol 11 (2012) 117–132

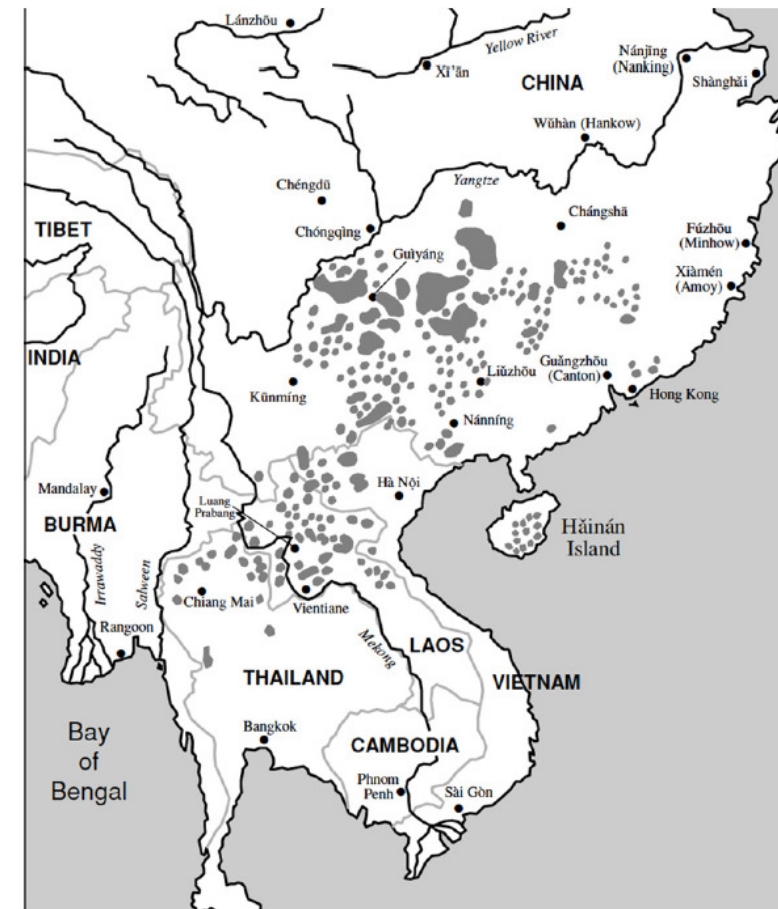
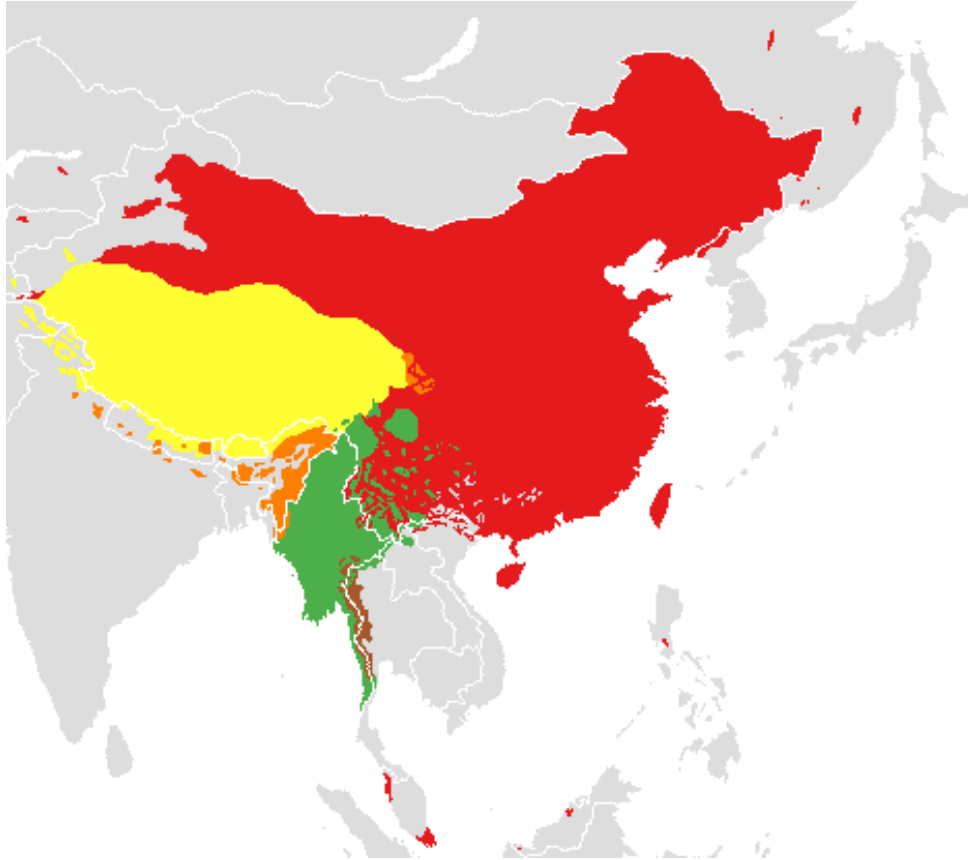


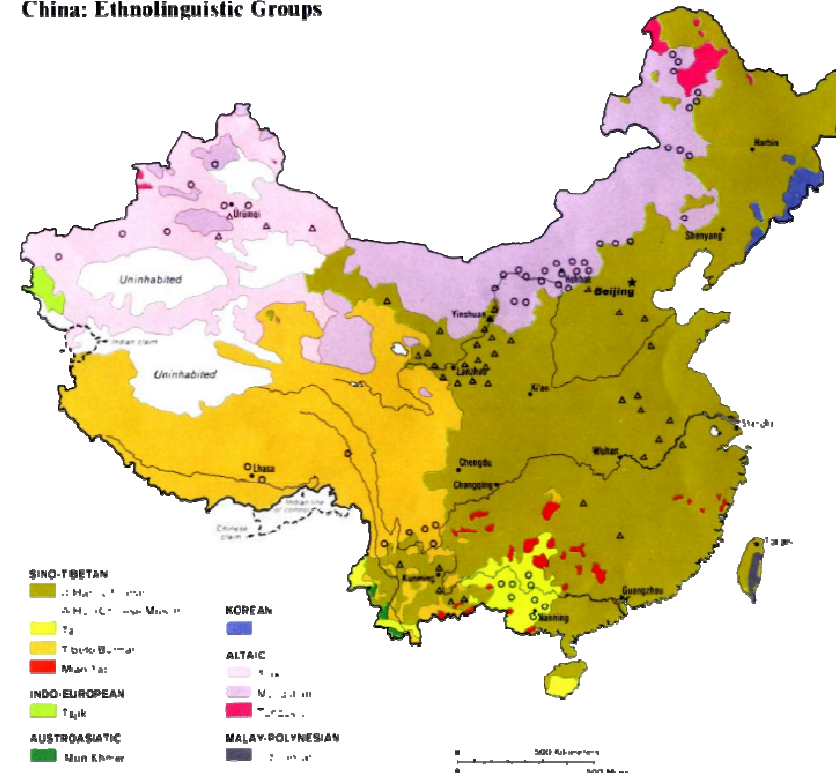
Fig. 1. The geographical distribution of modern Hmong-Mien language communities (van Driem, 2001: 319).
Fig. 1. Distribution géographique des communautés de langues modernes hmong-mien (van Driem, 2001: 319).

NGỮ HỆ HÁN - TẠNG (SINO-TIBETAN LANGUAGES, ST)



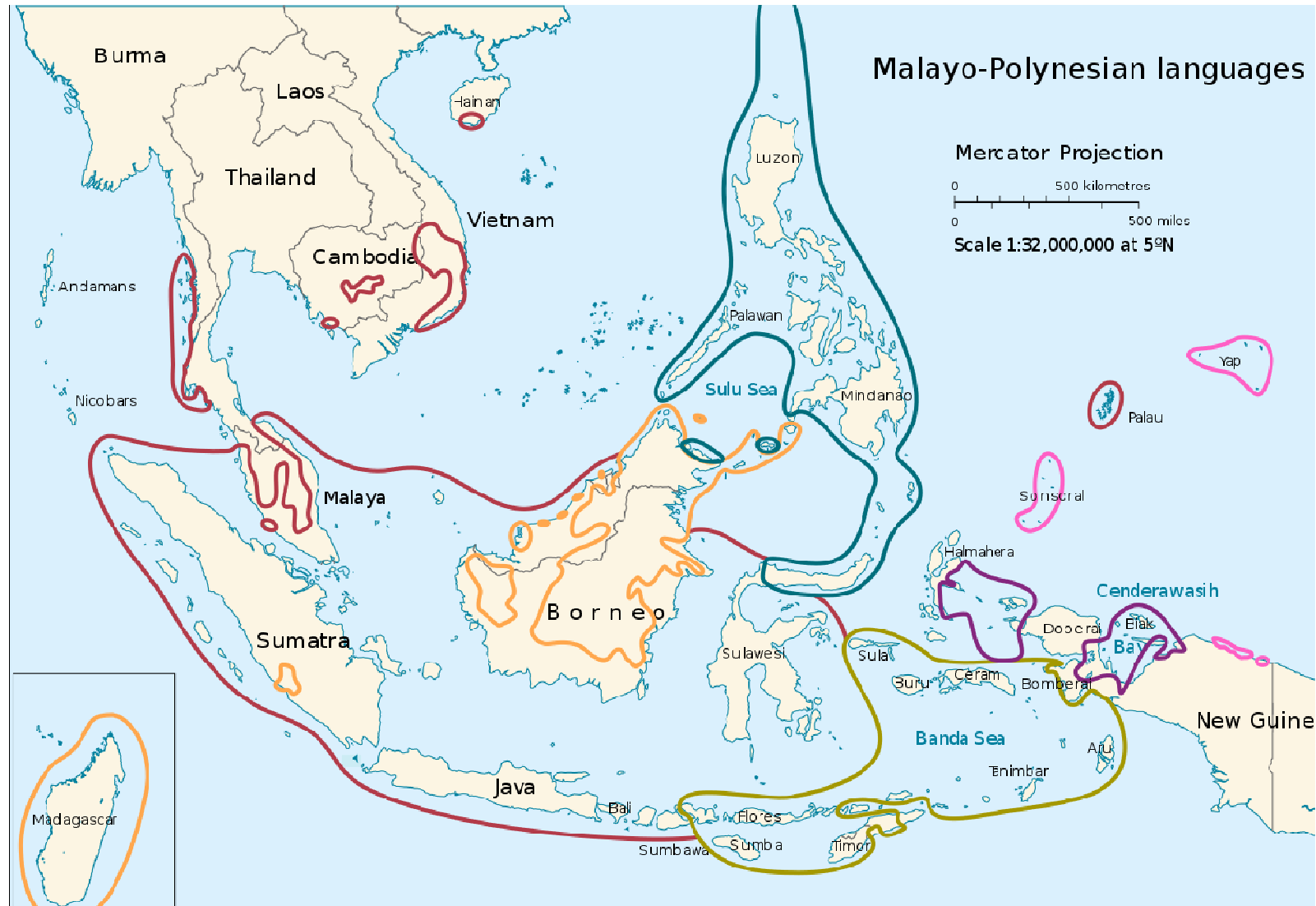
- **Nhóm Hán (Sinitic), 3** dân tộc: Hoa, Ngái (Hakka), Sán D
- **Nhóm Tạng-Miến (Tibet-Burma), 6** dân tộc: Cống, Hà Nhì, La Hủ, Lô Lô, Phù Lá, Si La

China: Ethnolinguistic Groups



NGŨ HỆ NAM ĐẢO (AUSTRONESIAN LANGUAGES, AN)

- **Nhóm Nam Đảo (Malayo-Polynesia), 5** dân tộc: Chăm (Chàm), Chu Ru, Ê Đê, Gia Lai, Ra Glai



NGHIÊN CỨU LỊCH SỬ DI TRUYỀN CÁC DÂN TỘC/ TỘC NGƯỜI

- DNA người cổ
- DNA người hiện đại

HỆ GEN TY THỂ
mtDNA GENOME
Di truyền theo dòng MẸ
(Mẫu hệ)

VÙNG ĐẶC HIỆU GIỚI TÍNH NAM
TRÊN NHIỄM SẮC THỂ Y
MALE SPECIFIC REGION
OF Y CHROMOSOME (MSY)
Di truyền theo dòng BỐ (Phụ hệ)

PHÂN TÍCH TRÊN
TOÀN BỘ HỆ GEN:
SNP CHIP, WES, WGS
(Di truyền theo cả dòng BỐ
và dòng MẸ, mức độ
lai hỗn hợp/ hoà huyết)

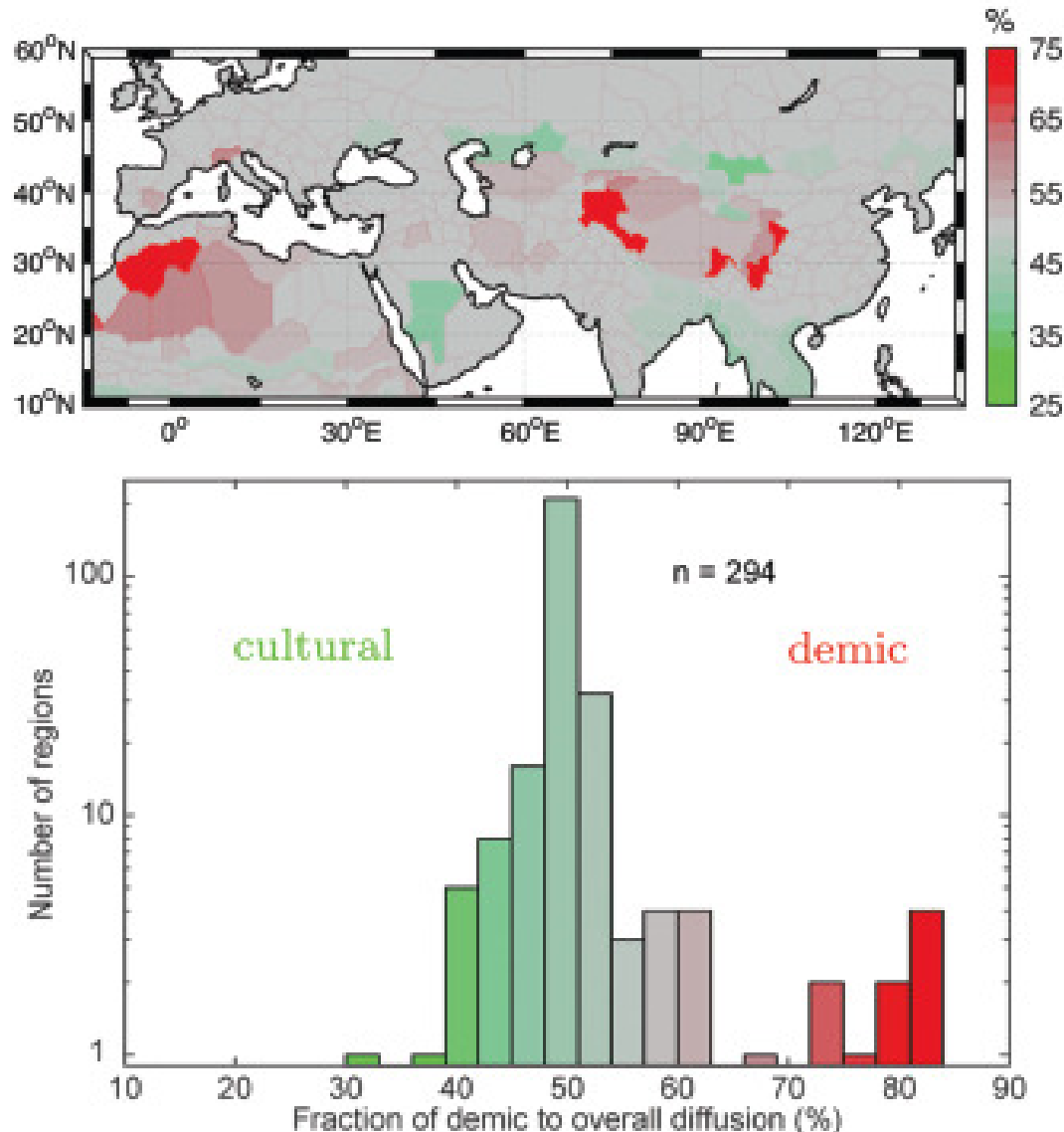


Fig. 3. Time integrated contribution of predominantly demic (red) and cultural (green) diffusion represented geographically (top panel) and as a histogram (bottom panel). For most regions, no predominance (grey) of either mechanism is found.

Cultural and demic diffusion of first farmers, herders, and their innovations across Eurasia

Carsten Lemmen

C. L. Science Consult; Helmholtz-Zentrum Geesthacht, Lüneburg, Geesthacht, DE
science@carsten-lemmen.de

ABSTRACT - Was the spread of agro-pastoralism from the Eurasian founder regions dominated by demic or by cultural diffusion? This study employs simulations that unfold a complex inter-regional and time varying pattern of demic and diffusive exchange processes during the Neolithisation, and thus supports from a modelling perspective the hypothesis that there is no simple or exclusive demic or cultural diffusion, but that in most regions of Eurasia a combination of demic and cultural processes were important.

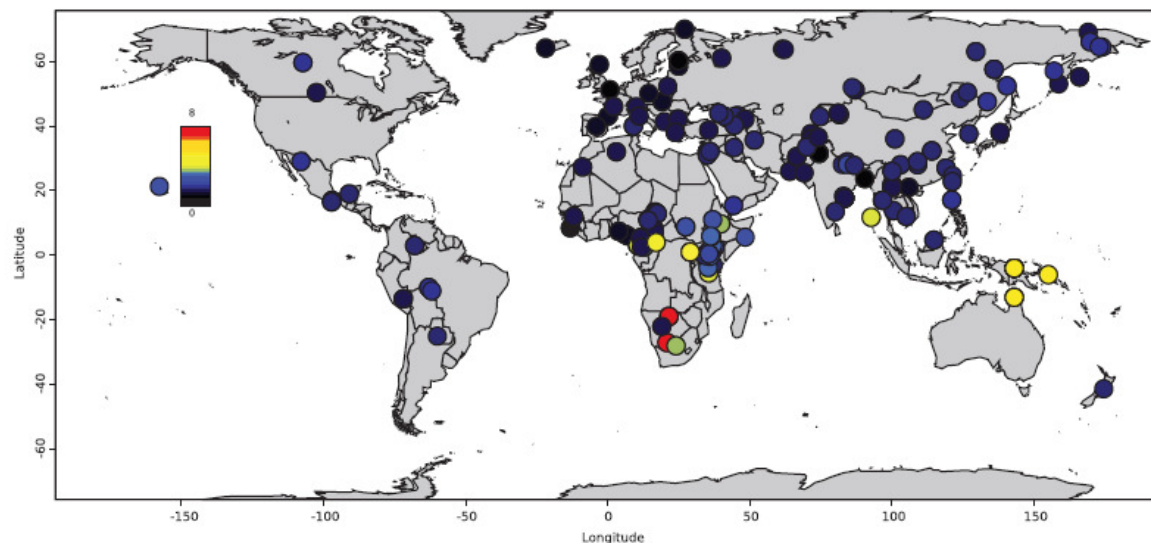
Conclusion

I presented a numerical simulation study on the diffusion processes during Neolithisation in Eurasia, using an adaptive model of prehistoric societies in their environmental context that is able to resolve local innovation, cultural diffusion and demic diffusion. Although a mixed diffusion process was already suggested long ago, the analysis of simulations with either cultural or demic diffusion, and with mixed diffusion reveals an even more complex spatio-temporal pattern of the expansion of agro-pastoralism throughout Eurasia than has previously been found: demic and cultural processes occur contemporaneously, or multiple times iteratively or intermittently in most regions of Eurasia. There is no simple demic or cultural explanation, but a very complex and rich interplay of both processes in time and space. The polarised debate of either demic or cultural diffusion should give way to acknowledging again this more complex picture and to studying and appreciating the richness of mechanisms.

Simons Genome Diversity Project: genomes from 142 diverse populations

Authors and affiliations appears at the end of the paper.

We report the Simons Genome Diversity Project data set: high quality genomes from 300 individuals from 142 populations. These genomes include at least 5.8 million base pairs that are not present in the human reference genome. Our analysis reveals key features of the landscape of human genome variation, including that the rate of mutations has accelerated by about 5% in non-Africans compared to Africans since divergence. That the ancestors of some pairs of present-day human populations were substantially separated by 100,000 years well before the archaeologically attested onset of behavioural modernity. We also demonstrate that indigenous populations, New Guineans and Andamanese do not derive substantial ancestry from an early dispersal of modern humans. Instead, their modern human ancestry is consistent with coming from the same source as that of other non-



Extended Data Figure 1 | Heat map of fraction of heterozygous sites missed in the 1000 Genomes project. For each sample, we examine all heterozygous sites passing filter level 1, and compute the fraction included as known polymorphisms in the 1000 Genomes project.

ARTICLE RESEARCH

RESEARCH ARTICLE

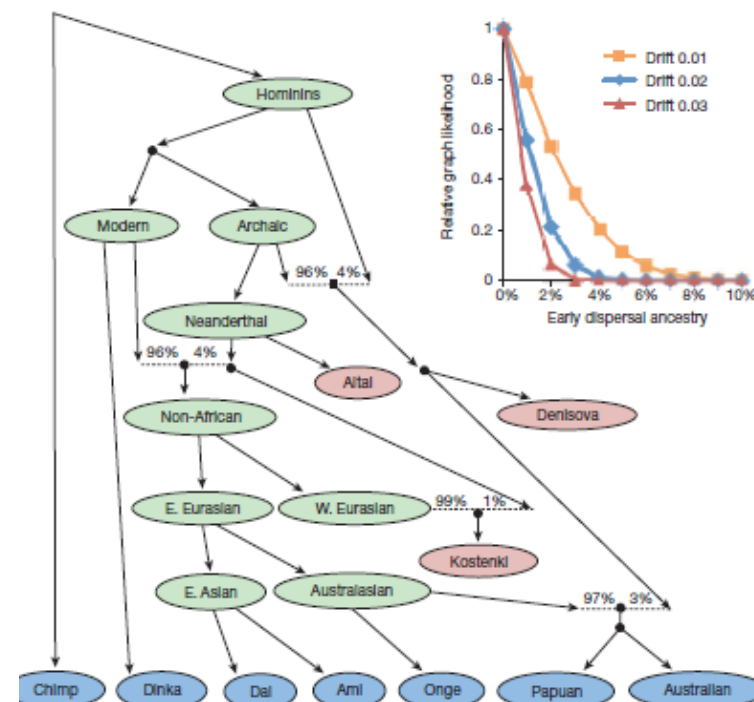
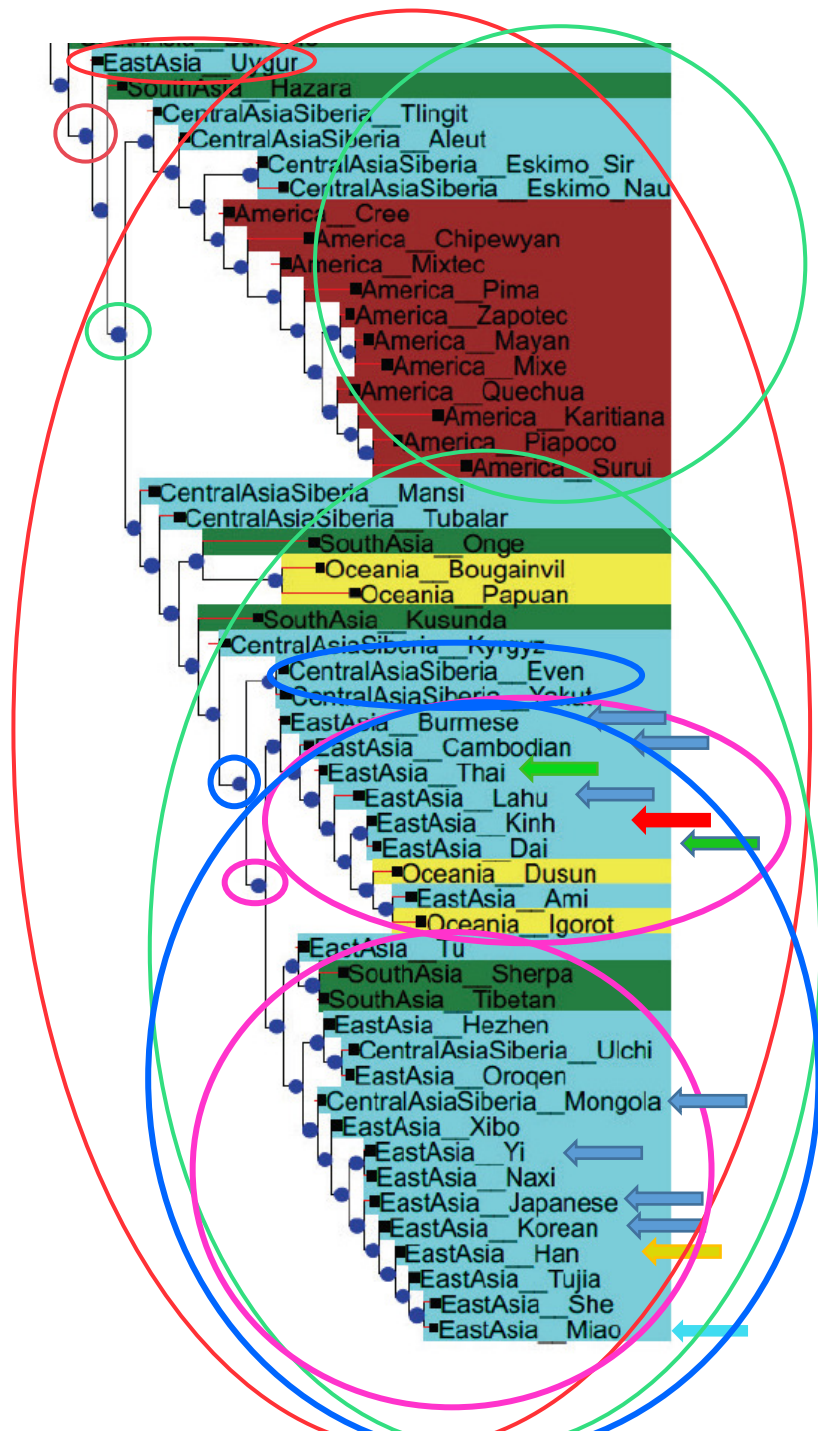
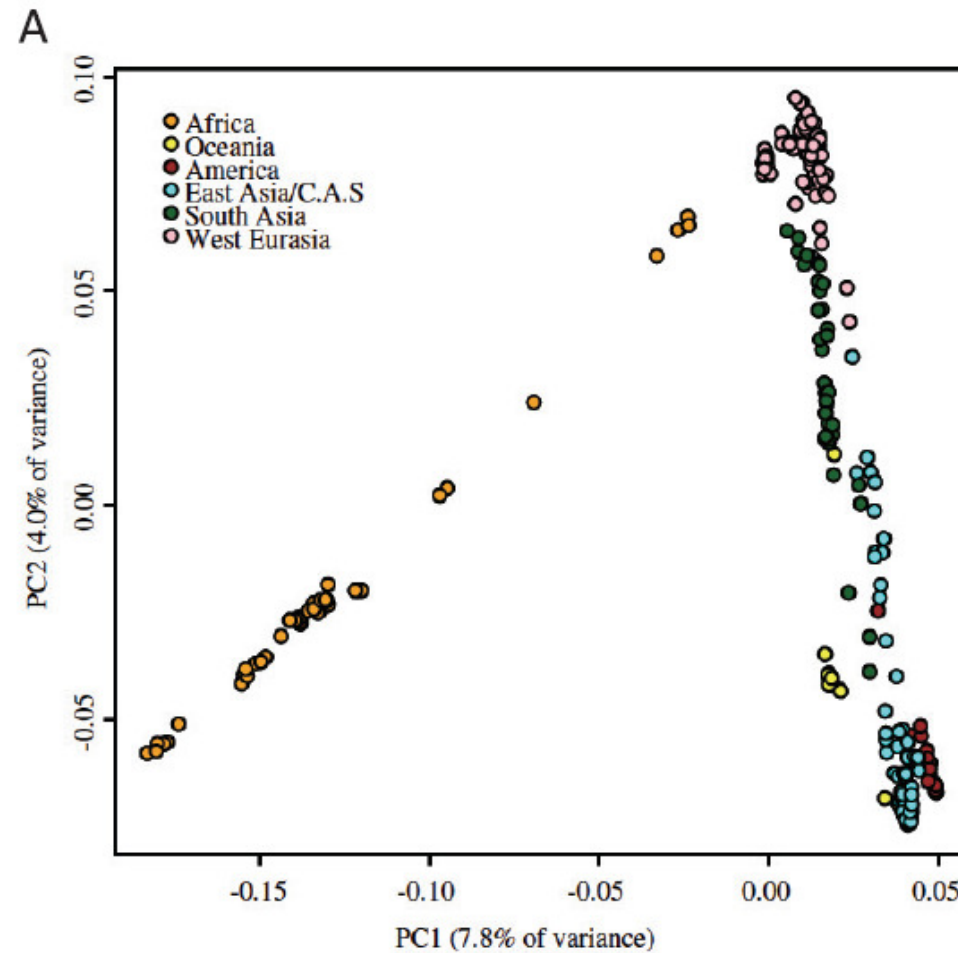


Figure 3 | Present-day populations have negligible ancestry from an early dispersal of modern humans out of Africa. Best-fitting admixture graph model of relationships among Australians, New Guineans, Andamanese and other diverse populations. Present-day populations are shown in blue, ancient samples in red, and select inferred ancestral nodes in green. Dotted lines indicate admixture events, all of which involve archaic humans. All f -statistic relationships are accurately fit to within 2.1 standard errors. Inset, results of adding putative early dispersal admixture to the graph model for different assumptions about when the early lineage split off. We specify the split time in terms of the genetic drift above the 'Non-African' node, with 0.01 units of drift representing on the order of ten thousand years. The (approximate) model likelihood is maximized with zero early dispersal ancestry, and no more than a few per cent is consistent with the data.

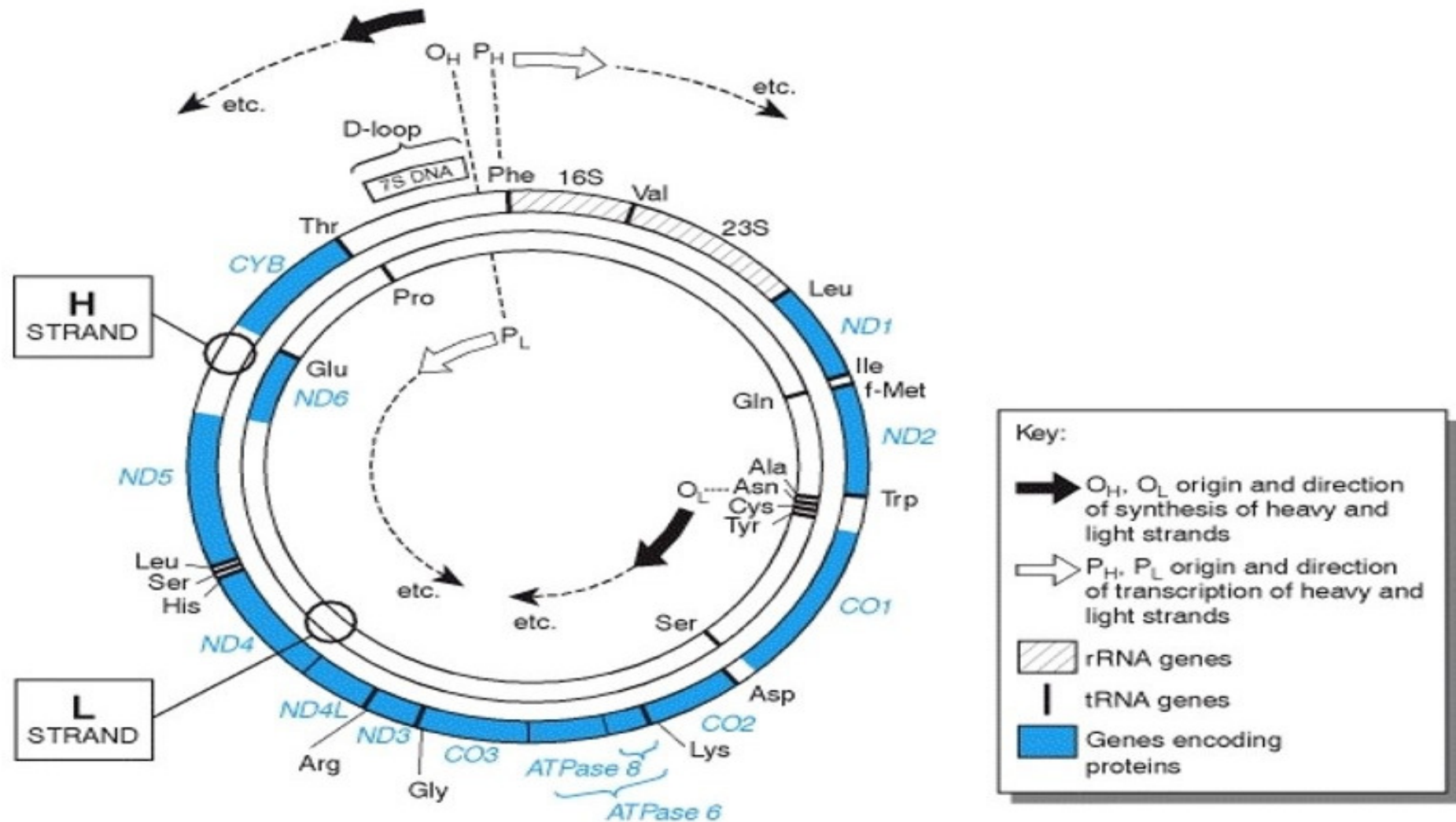


RESEARCH ARTICLE



Extended Data Figure 4 | Principal component analysis and neighbour joining tree. a, Principal component analysis. b, Neighbour-joining tree. c, Neighbour-joining tree on F_{ST} values for all populations with at least two samples.

Mitochondrial Genome (mtDNA): ~ 16.000 bp, 37 genes



Human (nuclear) Genome:

Size: ~ 3.2 Gb , ~21,000 genes

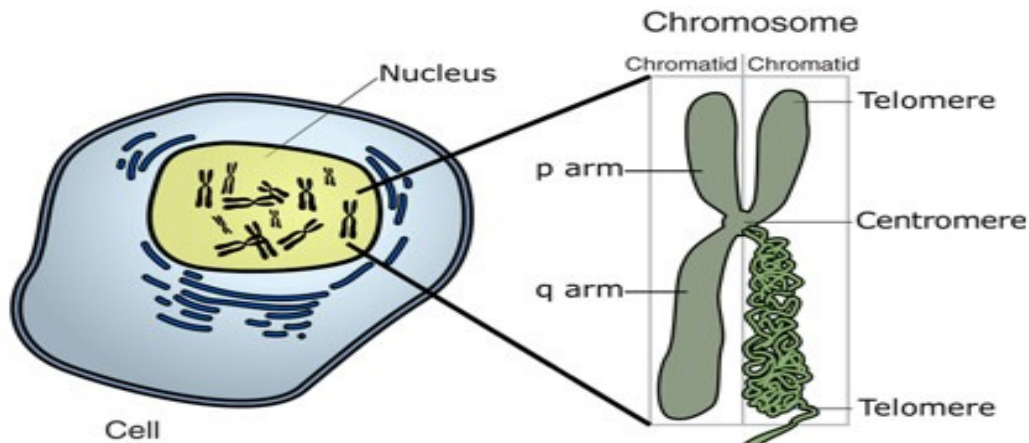
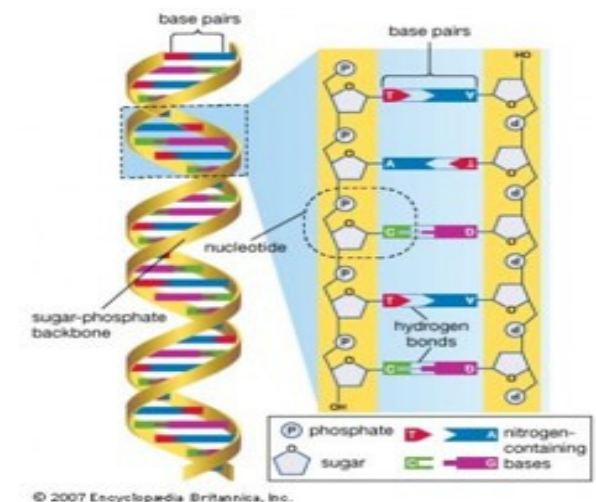
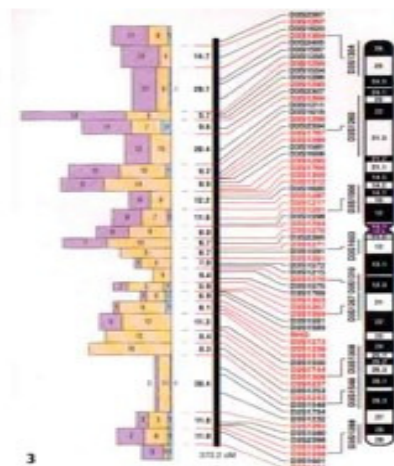
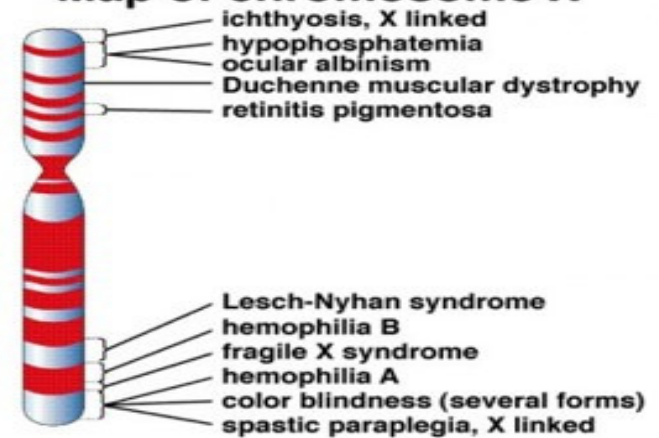


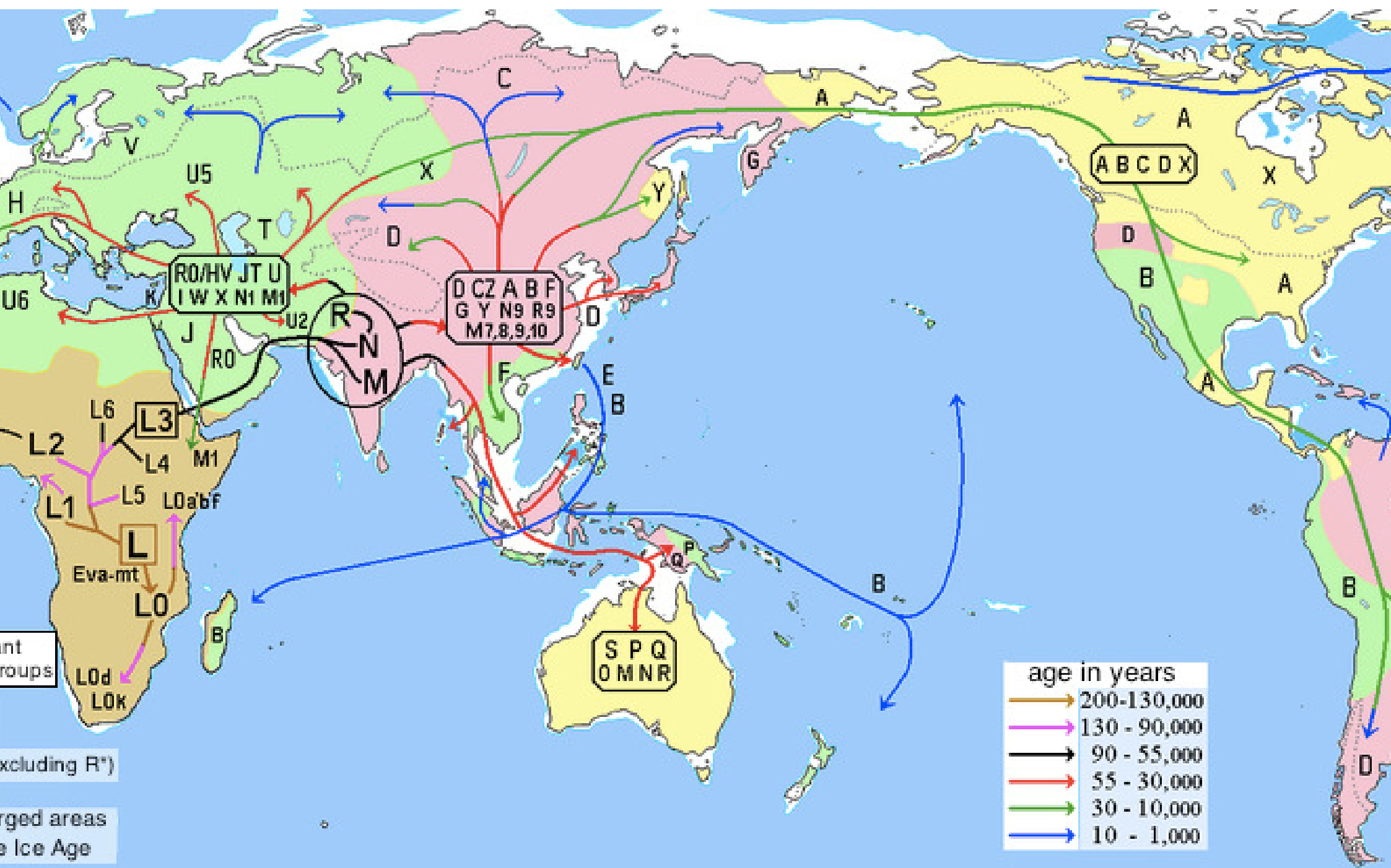
Image adapted from: National Human Genome Research Institute.

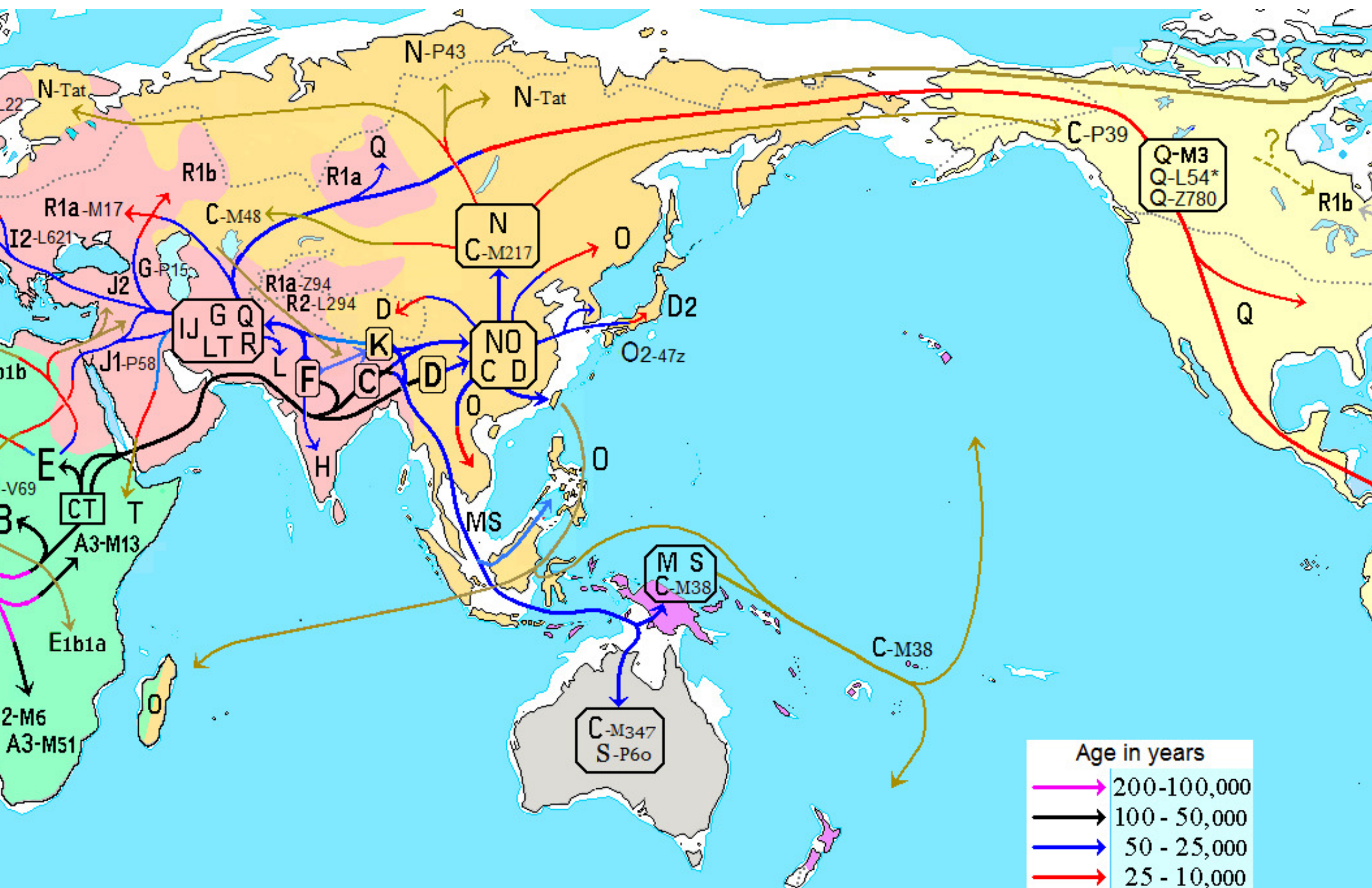
Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.

Map of chromosome X

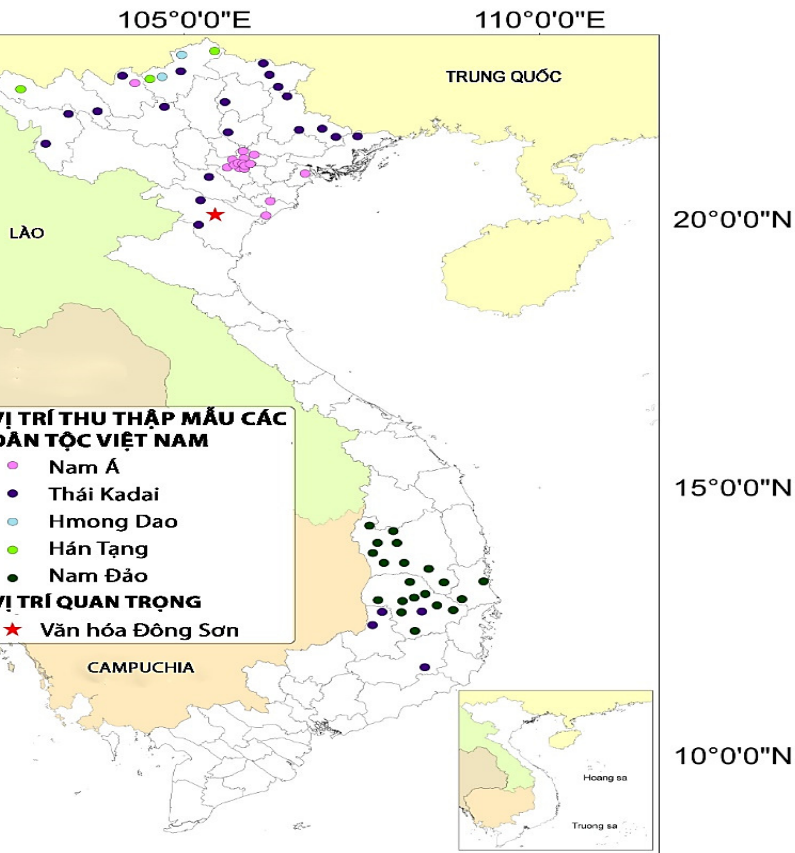


© 2007 Encyclopedia Britannica, Inc.





Giải trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Việt Nam



- Đã thu được mẫu máu của 609 cá thể nam thuộc 17 dân tộc 5 ngữ hệ
- Đã tách chiết DNA tổng số và giải trình tự toàn bộ hệ gen ty thể trên hệ thống NextSeq500
- Trình tự hoàn chỉnh của hệ gen ty thể được sử dụng để tiến hành cách phân tích tiếp theo

Hình 11. Vị trí lấy mẫu 17 dân tộc người Việt Nam

Bảng trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

Thống kê số đa hình của 609 cá thể thuộc 17 dân tộc theo 5 ngữ hệ

STT	Dân tộc	Ngữ hệ	Số lượng mẫu	Số đa hình
1	Gia Rai	Nam Đảo	30	203
2	Ê Đê		24	172
3	Phù Lá	Hán Tạng	35	235
4	Si La		31	135
5	Lô Lô		36	151
6	La Hủ		32	219
7	Hà Nhì		33	246
8	Pà Thẻn	Mông – Dao (H'mông-Miền)	36	159
9	Dao		44	231
10	H'mông		41	181
11	Thái	Thái - Kadai	24	242
12	Tày		48	336
13	Nùng		37	251
14	Cờ Lao		34	261
15	La Chí		36	159
16	Mảng	Nam Á	37	142
17	Kinh		51	394

Giải trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Nam (tiếp)

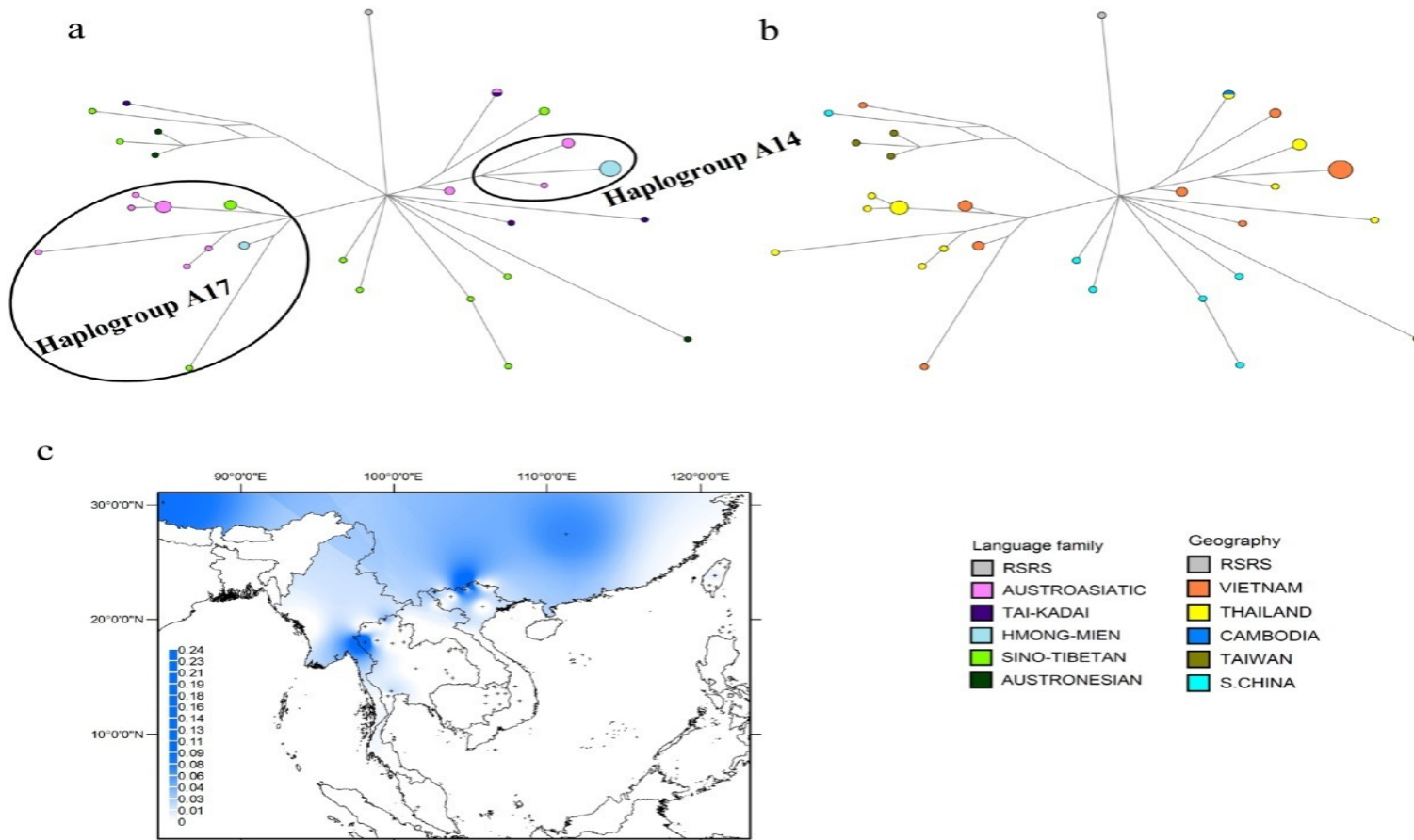
Tần suất các nhóm đơn bội chính của 2742 cá thể theo 5 ngữ hệ

Đơn vị	Nam Á (657)	Thái-Kadai (932)	Mông-Dao (154)	Hán-Tạng (399)	Nam Đảo (600)	Tổng số (2742)	Nhóm đơn bội	Nam Á (657)	Thái-Kadai (932)	Mông-Dao (154)	Hán-Tạng (399)	Nam Đảo (600)	Tổng (2742)
17	0,91	0,5	5,84	2,26	0,50	1,17	M12	5,94	1,18	3,25	0,25	0,17	2,08
	1,52	0,00	1,30	1,00	0,00	0,58	M20	2,59	1,29	0,00	0,25	0,67	1,24
	0,46	2,68	2,60	4,51	0,50	1,93	M21	1,22	0,21	0,00	0,50	3,17	1,13
	0,46	2,47	9,09	3,01	7,83	3,61	M24	1,98	0,97	0,00	1,00	0,33	1,02
	0,30	2,90	0,00	0,75	5,33	2,33	M51	1,98	0,43	0,00	0,50	0,50	0,80
	0,30	3,22	1,30	8,52	3,00	3,14	M61	0,15	1,18	0,00	0,00	0,00	0,44
	10,05	11,91	8,44	4,26	5,17	8,68	M62b	0,00	0,00	0,00	2,76	0,00	0,40
	0,46	0,21	0,00	1,00	0,17	0,36	M68a	2,13	0,11	0,00	0,00	0,50	0,66
	4,11	1,93	2,60	0,00	0,17	1,82	M71	3,81	1,07	0,00	0,25	3,50	2,08
7)	0,30	0,21	5,19	0,50	0,00	0,51	M73	0,76	0,11	0,00	0,00	0,50	0,33
	1,83	3,97	7,79	1,25	0,83	2,59	M74	3,35	1,82	1,30	0,75	0,50	1,71
	5,63	2,90	4,55	7,02	0,67	3,76	M7b	3,96	12,23	1,30	5,01	9,50	7,99
	0,61	0,97	1,30	3,01	6,50	2,41	M7c	1,07	3,65	0,00	2,76	5,67	3,14
	0,00	0,00	0,00	0,75	0,33	0,18	M8	0,15	0,43	0,00	0,50	0,17	0,29
	0,00	0,00	0,00	0,25	8,83	1,97	M9	0,46	0,97	0,00	2,01	0,17	0,77
F1a)	6,54	6,44	5,84	5,01	0,17	4,85	M91	0,76	0,43	0,00	0,75	0,00	0,44
	15,07	12,88	11,69	6,27	9,17	11,56	N	1,83	1,07	0,00	0,25	1,50	1,17
	0,00	1,39	5,19	2,76	0,50	1,28	N21	1,07	0,43	1,30	0,75	1,67	0,95
	0,15	2,04	5,84	1,00	6,00	2,52	N9a	1,07	3,54	4,55	3,26	3,83	3,03
	0,15	0,75	0,00	0,00	4,33	1,24	R	3,35	1,39	0,00	2,51	1,33	1,93
	0,30	0,54	5,84	2,01	0,33	0,95	R9	2,28	3,43	1,95	1,50	3,83	2,88
	8,52	4,51	1,30	17,54	1,00	6,42	Other	2,44	0,64	0,65	1,25	0,83	1,20
	0,00	0,97	0,00	0,50	0,33	0,47							

Nội dung 3: Giải trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

- Các dòng nhóm đơn bội mtDNA ở lục địa Đông nam Á

Nhóm đơn bội A

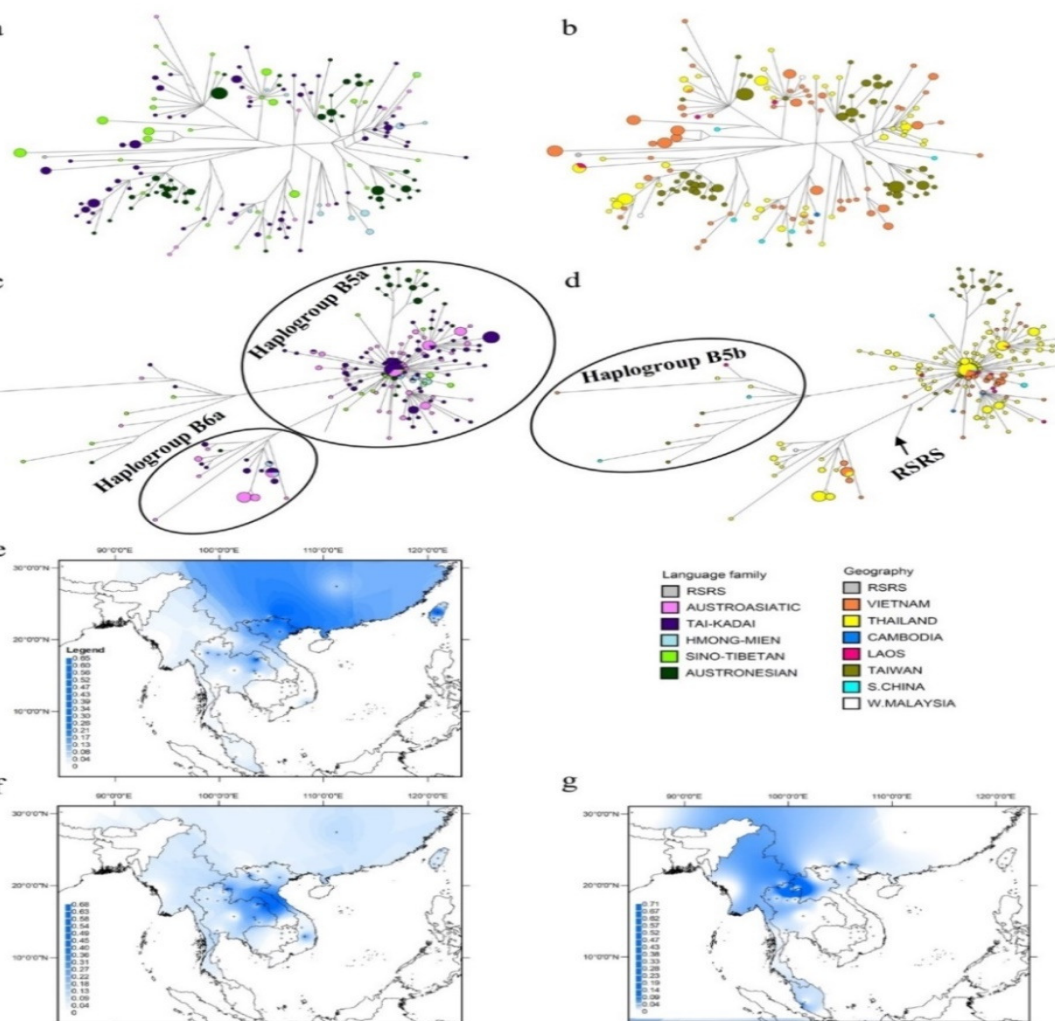


(a, b) Network nhóm đơn bội ty thể A, được mã hóa màu theo ngữ hệ và nguồn gốc địa lý;
(c) Bản đồ phân bố theo tần suất của nhóm đơn bội A: màu càng đậm, tần suất trong dân số càng cao; Các dấu chữ thập đánh dấu các vị trí của 68 quần thể Đông Nam Á có trong phân tích.

Giải trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

➤ Các dòng nhóm đơn bội mtDNA ở lục địa Đông nam Á

Nhóm đơn bội B



(a, b) Network nhóm đơn bội ty thể B4

(c, d) Network nhóm đơn bội ty thể B5 và B6a, được mã hóa màu theo ngữ hệ và nguồn gốc địa lý

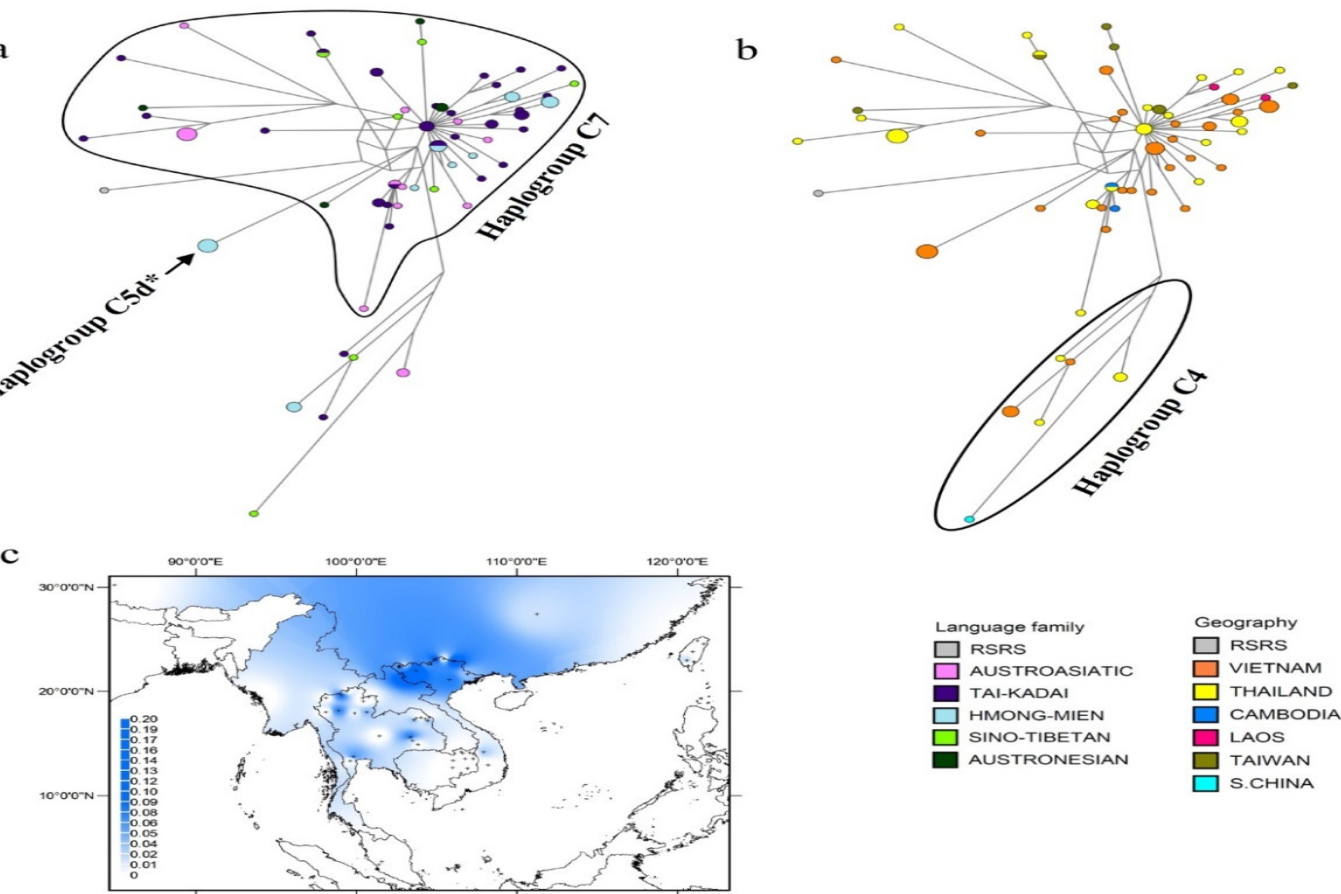
(e-g) Bản đồ phân bố theo tần suất của nhóm đơn bội B4, B5 và B6a, màu càng đậm, tần suất trong dân số càng cao

Các dấu chữ thập đánh dấu các vị trí của 68 quần thể dân tộc Việt Nam Á có trong phân tích.

Chương 3: Giải trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

➤ Các dòng nhóm đơn bội mtDNA ở lục địa Đông nam Á

Nhóm đơn bội C

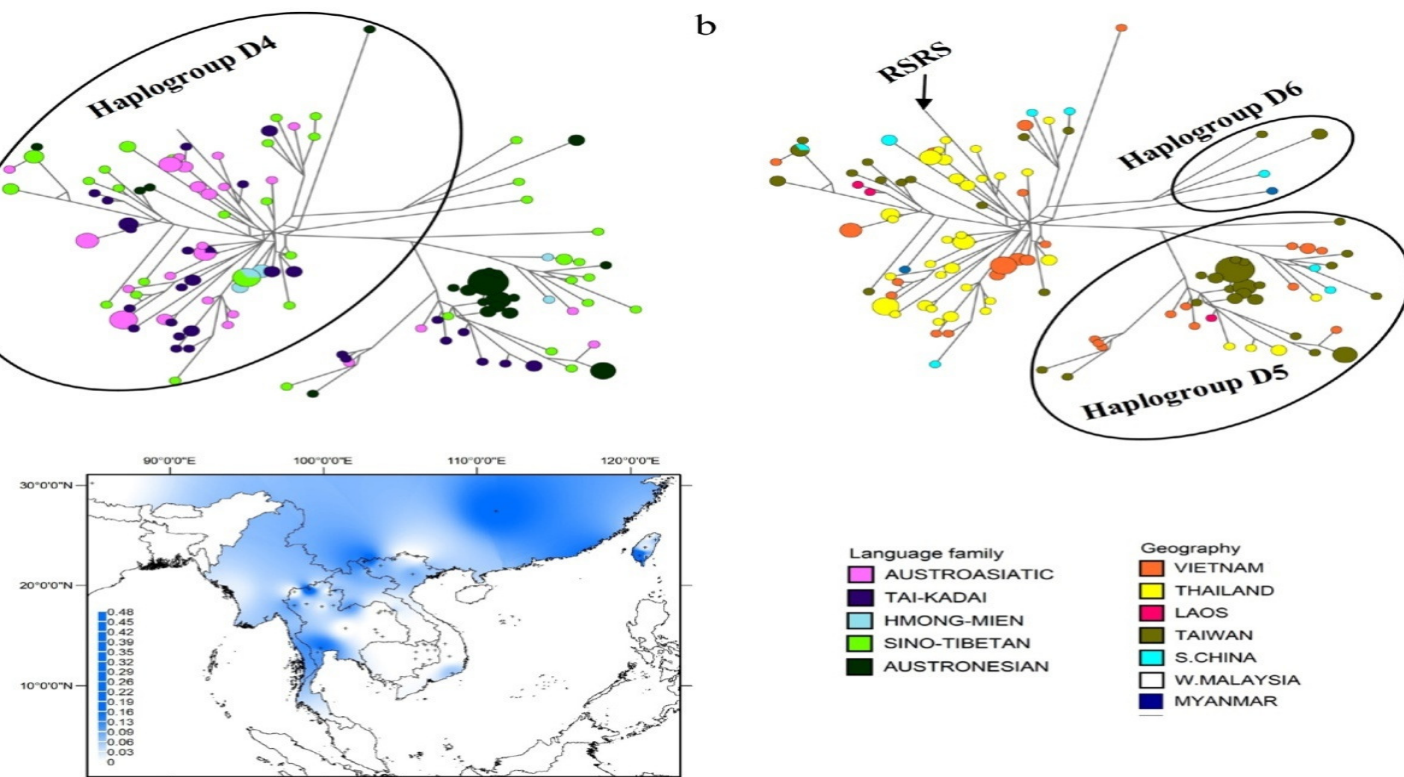


(a, b) Network nhóm đơn bội ty thể B4, được mã hóa màu theo ngữ hệ và nguồn gốc địa lý
(c) Bản đồ phân bố theo tần suất của nhóm đơn bội C: màu càng đậm, tần suất trong dân số càng cao
Các dấu chữ thập đánh dấu các vị trí của 68 thể Đông Nam Á

Giải trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

➤ Các dòng nhóm đơn bội mtDNA ở lục địa Đông Nam Á

Nhóm đơn bội D

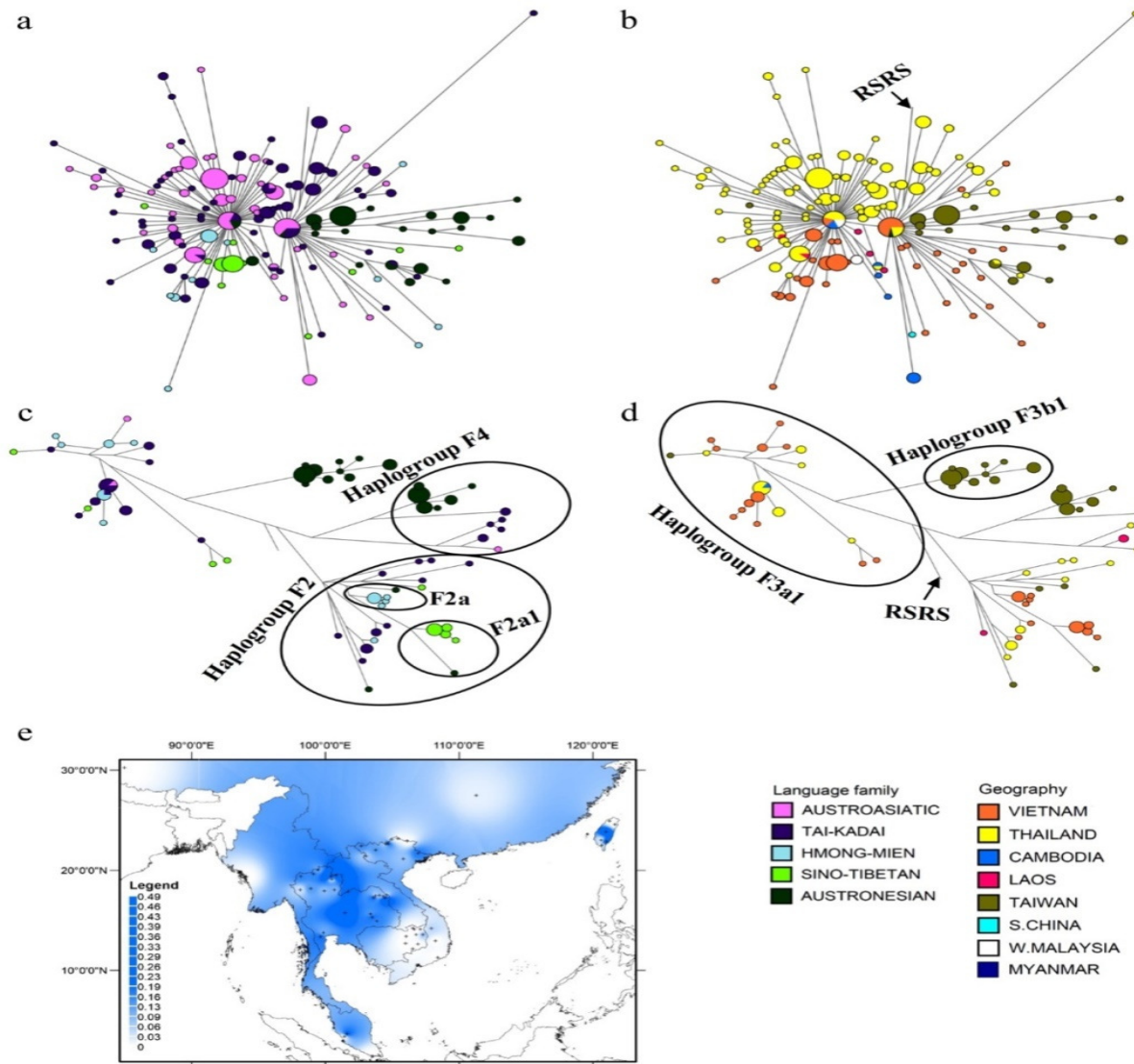


(a, b) Network nhóm đơn bội ty thể B, được mã hóa màu theo ngữ hệ và nguồn gốc địa lý.
(c) Bản đồ phân bố theo tần suất của nhóm đơn bội D: màu càng đậm, tần suất trong dân cư càng cao.
Các dấu chữ thập đánh dấu các vị trí của quần thể MSEA có trong phân tích.

Giải trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

Các dòng nhóm đơn bội mtDNA ở lục địa Đông Nam Á

nhóm đơn bội F



(a, b) Network nhóm đơn bội ty thể F1a; (c,d) Network nhóm đơn bội ty thể F2-4, được mã hóa màu theo ngữ hệ và nguồn gốc địa lý

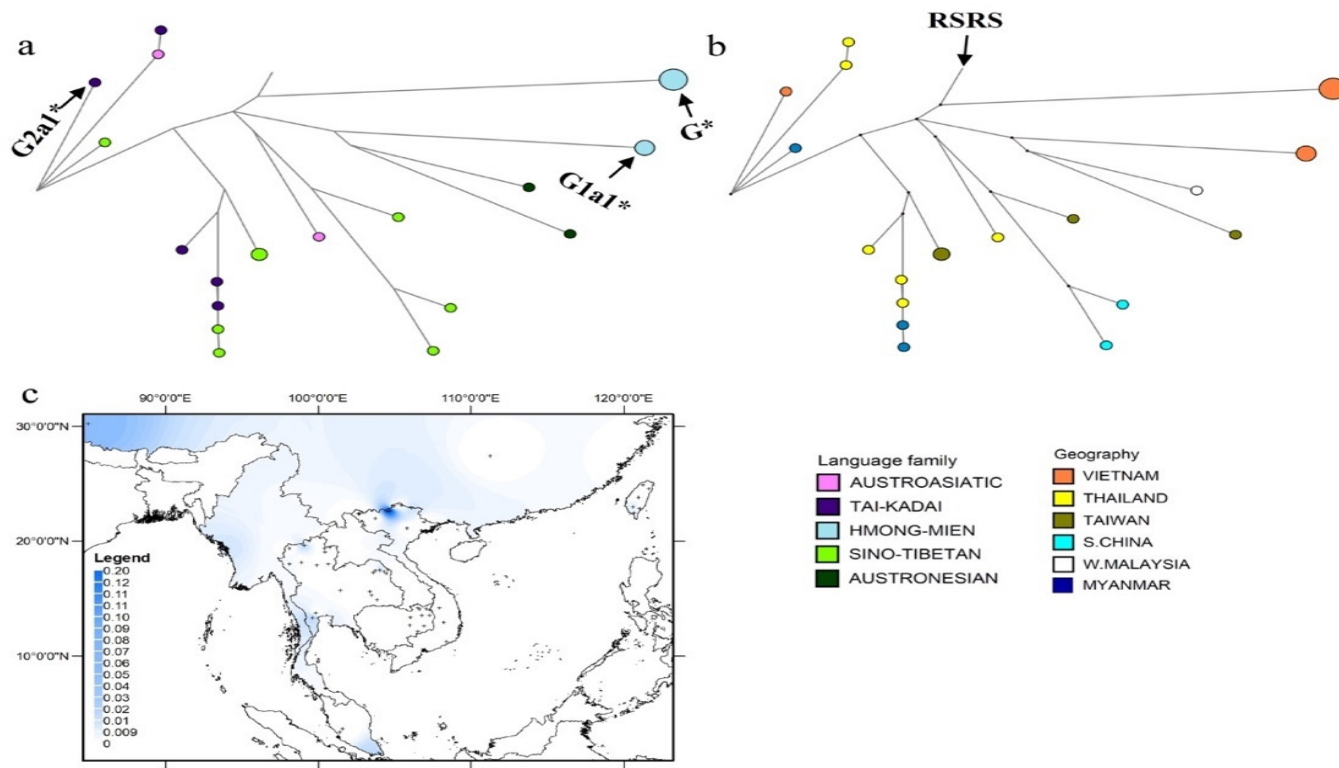
(e) Bản đồ phân bố theo tần suất của nhóm đơn bội F: màu càng đậm, tần suất trong dân số càng cao

Các dấu chữ thập đánh dấu vị trí của 68 quần thể MSEA trong phân tích

Quá trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

Các dòng nhóm đơn bội mtDNA ở lục địa Đông Nam Á

Nhóm đơn bội G

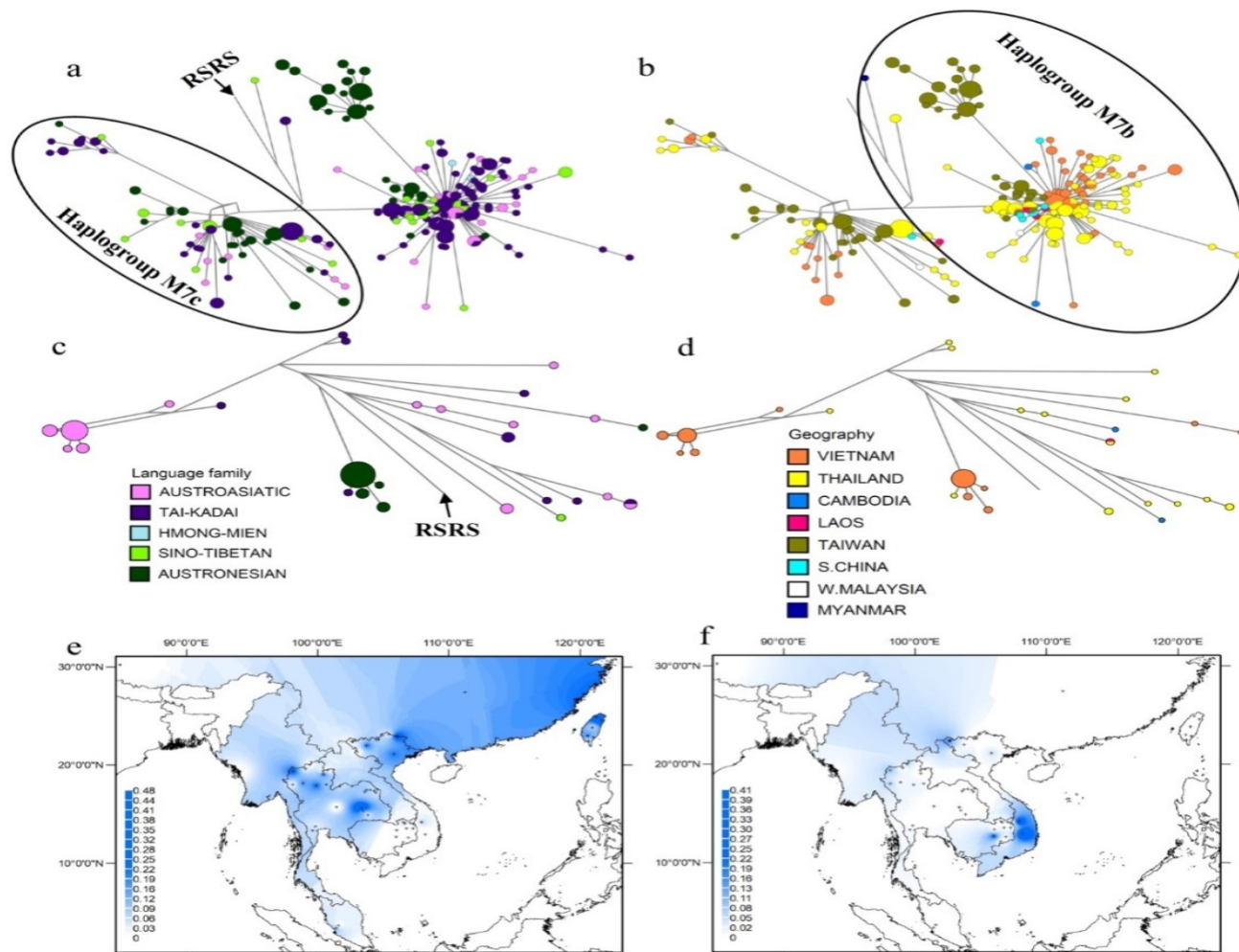


(a, b) Network nhóm đơn bội ty thể G được mã hóa màu theo ngữ hệ và nguồn gốc địa lý
(c) Bản đồ phân bố theo tần suất của nhóm đơn bội F: màu càng đậm, tần suất trong dân số càng cao
Các dấu chữ thập đánh dấu các vị trí của 68 quần thể MSEA có trong phân tích

Quá trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

Các dòng nhóm đơn bội mtDNA ở lục địa Đông Nam Á

Nhóm đơn bội M



(a, b) Network nhóm đơn bội ty thể M7

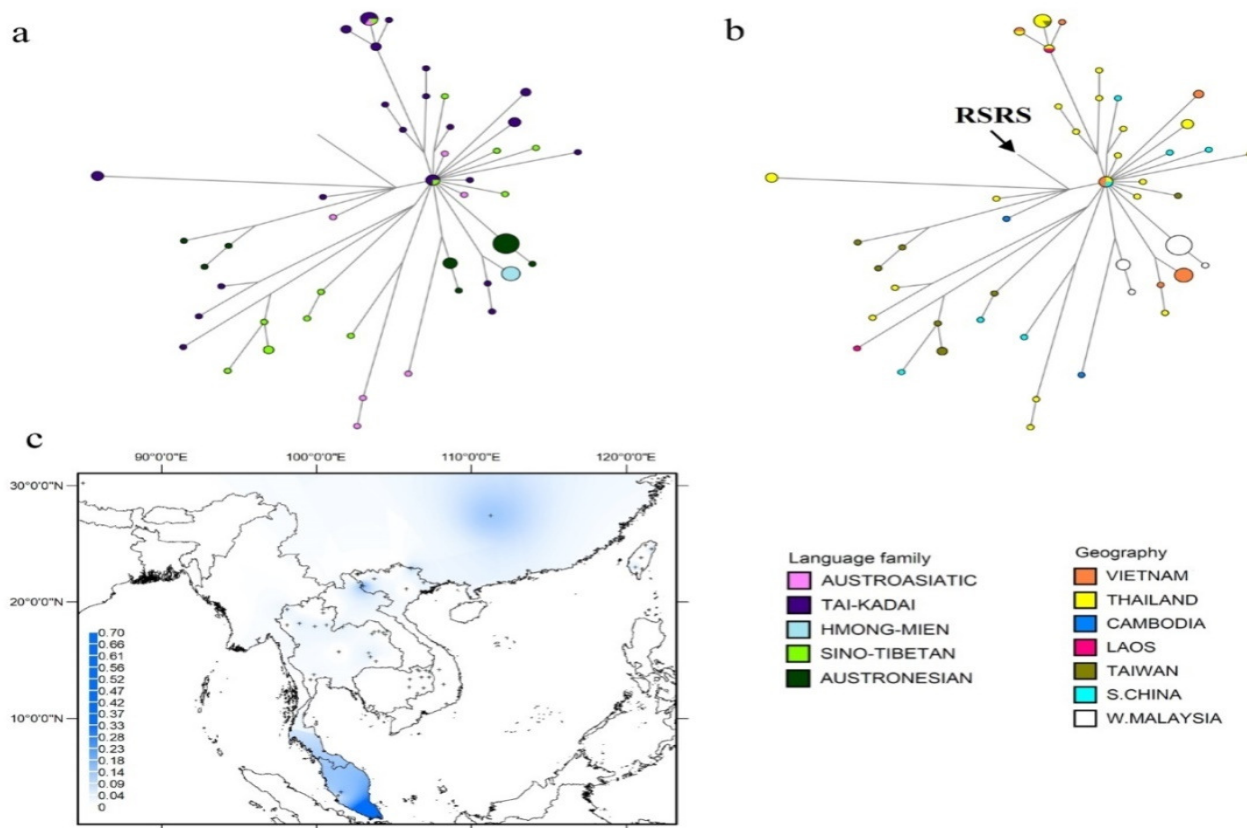
(c,d) Network nhóm đơn bội ty thể M71, được mã hóa màu theo ngôn ngữ và nguồn gốc địa lý

(e,f) Bản đồ phân bố theo tần suất của nhóm đơn bội F: màu càng tối tần suất trong dân số càng cao. Các dấu chữ thập đánh dấu các vị trí của 68 quần thể MSEA phân tích

Giải trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

Các dòng nhóm đơn bội mtDNA ở lục địa Đông nam Á

Nhóm đơn bội N9a

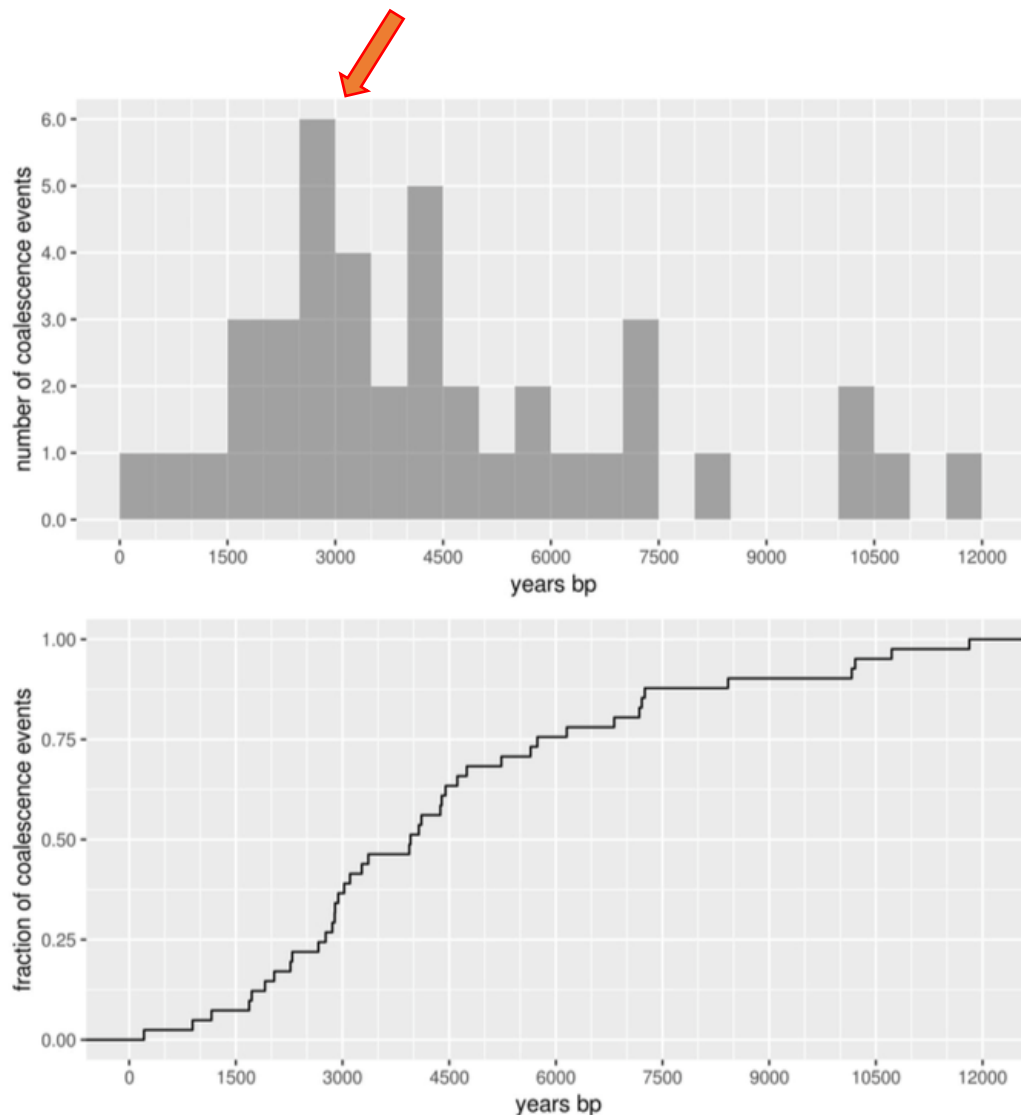


(a, b) Network nhóm đơn bội ty thể N9a, được mã hóa màu theo ngữ hệ và nguồn gốc địa lý
(c) Bản đồ phân bố theo tần suất của nhóm đơn bội N9a: màu càng đậm, tần suất trong dân số càng cao
Các dấu chữ thập đánh dấu các vị trí của 68 quần thể MSEA phân tích.

ội dung 3: Giải trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

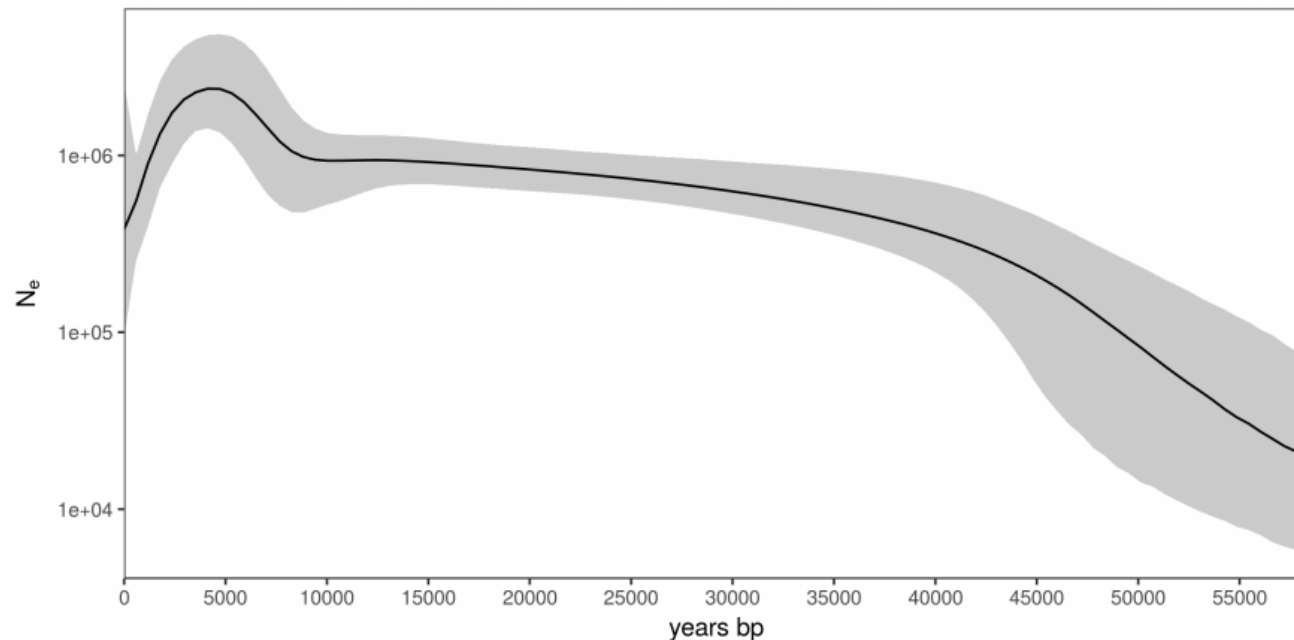
Biểu đồ các khoảng thời gian hợp nhất các nhóm đơn bội mtDNA mới

- Phát hiện được 111 dòng nhánh mới của mtDNA
- Phát hiện được đỉnh cao của sự đa dạng mtDNA khoảng 2,5-3 ngàn năm về trước, trùng với nền văn hoá Đông Sơn, có thể liên quan đến sự mở rộng nền văn hoá theo xu hướng nông nghiệp.
- Đây là bằng chứng khoa học mới, đầu tiên trên thế giới về di truyền phân tử người liên quan đến khoảng thời gian ra đời của nền văn hoá Đông Sơn.



Nội dung 3: Giải trình tự và phân tích hệ gen ty thể của một số dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

Biểu đồ kích thước quần thể dựa tính theo mtDNA



- Phát hiện sự mở rộng kích thước quần thể khoảng 4-5 ngàn năm trước

Phân bố các đa hình đại diện cho nhóm đơn bội O1 ở 17 dân tộc

[illegible]

Sự phân bố các đa hình đại diện cho nhóm đơn bội O1 ở 17 dân tộc

[illegible]

hiên cứu xác định đa hình nucleotide đơn trên vùng không trao đổi chéo của nhi
c thể Y của các cá thể thuộc 10 dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp

Xác định các SNP 600 cá thể thuộc nhóm đơn bội O ở 17 dân tộc Việt Nam

Sự phân bố các đa hình đại diện cho nhóm đơn bội O1 ở 17 dân tộc

Nhóm đơn bội	Dân tộc																
	Kinh (50)	Mông (37)	Tày (46)	Nùng (37)	Thái (24)	Cờ Lao (34)	La Chí (36)	Hmông (41)	Dao (43)	Pà Thẻn (36)	Hà Nhì (33)	La Hủ (32)	Lô Lô (35)	Sì La (30)	Phù Lá (35)	Ê Đê (24)	Gia Rai (27)
O1b1a2			1	6								2					
O1b1a2a1	1										3			24			

Nghiên cứu xác định đa hình nucleotide đơn trên vùng không trao đổi chéo của nhiễm sắc thể Y của các cá thể thuộc 10 dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

Xác định các SNP 600 cá thể thuộc nhóm đơn bội O ở 17 dân tộc Việt Nam

Sự phân bố các đa hình đại diện cho nhóm đơn bội O2 ở 17 dân tộc

Nhóm đơn bội	Dân tộc																
	Kinh (50)	Mảng (37)	Tày (46)	Nùng (37)	Thái (24)	Cờ Lao (34)	La Chí (36)	Hmong (41)	Dao (43)	Pà hên (36)	Hà Nhi (33)	La Hủ (32)	Lô Lô (35)	Sì La (30)	Phù Lá (35)	Ê Đê (24)	Gia Rai (27)
O2					2												
O2a1c1a	1																
O2a1c1a1						1											
O2a1c1a1a1b										1							
O2a1c1a5	2																
O2a1c1a6a1	4			1									1				
O2a1c1b2										3							
O2a2a1				3													
O2a2a1a2a			2														
O2a2a1a2a1a					1				2							2	1
O2a2a1a2a1a2						1		6	1	16			2		1		

ội dung 4: Nghiên cứu xác định đa hình nucleotide đơn trên vùng không trao đổi chéo của nhiễm sắc thể Y của các cá thể thuộc 10 dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

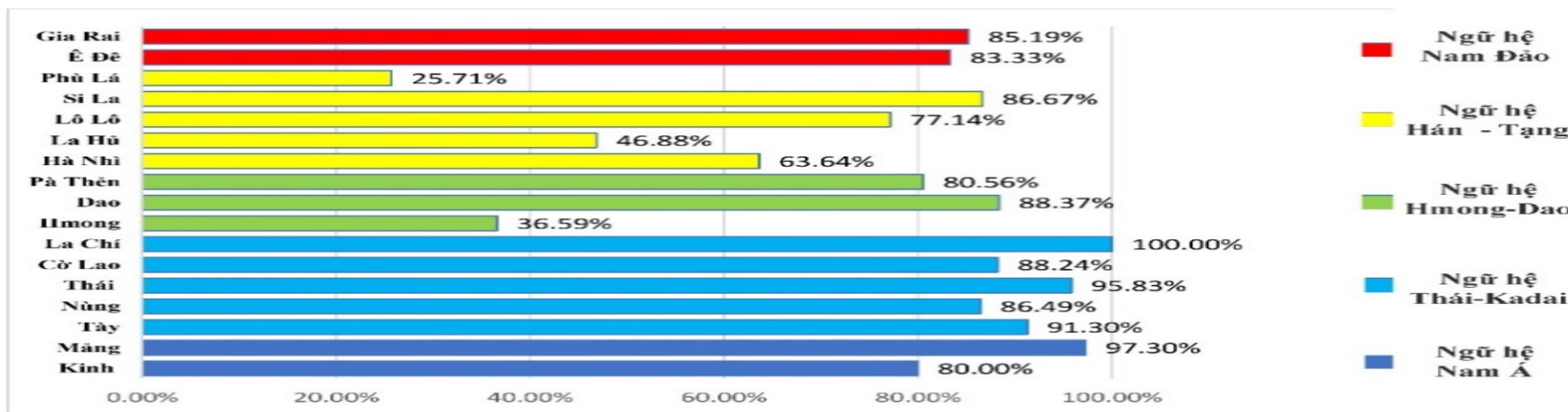
Xác định các SNP 600 cá thể thuộc nhóm đơn bội O ở 17 dân tộc Việt Nam

Sự phân bố các đa hình đại diện cho nhóm đơn bội O2 ở 17 dân tộc

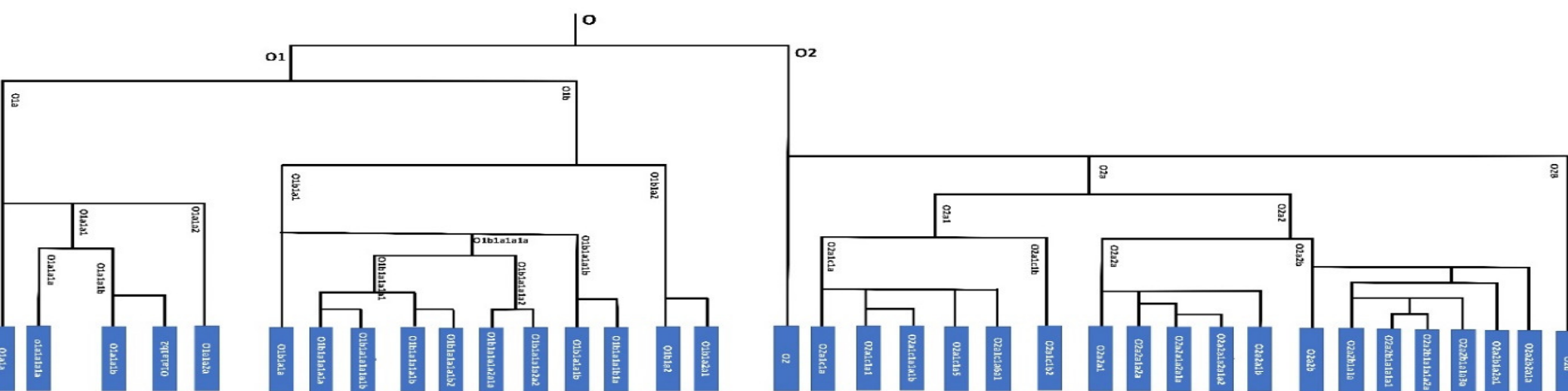
SNP	Nhóm đơn bội	Dân tộc																	Tổng số
		Kinh (50)	Mảng (37)	Tày (46)	Nùng (37)	Thái (24)	Cờ Lao (34)	La Chí (36)	Hmông (41)	Dao (43)	Pà hên (36)	Hà Nhi (33)	La Hủ (32)	Lô Lô (35)	Sì La (30)	Phù Lá (35)	Ê Đê (24)	Gia Rai (27)	
F155	O2a2b1a1a1a1			1															1
F2137	O2a2b1a1a1a2a											12	4						16
F317	O2a2b1a1a3a			1															1
F46	O2a2b1a2a1	2								12			1						15
F4110	O2a2b2a1a									6									6
F742	O2b															1			1
F1863	O2a2a1a2a2	4		1														3	8
F1837	O2a2a1b					2													2
P164	O2a2b			1															1
F8	O2a2b1a1a	2		5	1	2		8			1					3		1	23

Nghiên cứu xác định đa hình nucleotide đơn trên vùng không trao đổi chéo của nhiễm sắc thể Y của các cá thể thuộc 10 dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

Xác định các SNP 600 cá thể thuộc nhóm đơn bội O ở 17 dân tộc Việt Nam



Tần suất nhóm đơn bội O ở 17 dân tộc Việt Nam



Cây phát sinh n...
đơn bội O ở 17
tộc Việt Nam

Nghiên cứu xác định đa hình nucleotide đơn trên vùng không trao đổi chéo của nhiễm sắc thể Y của các cá thể thuộc 10 dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

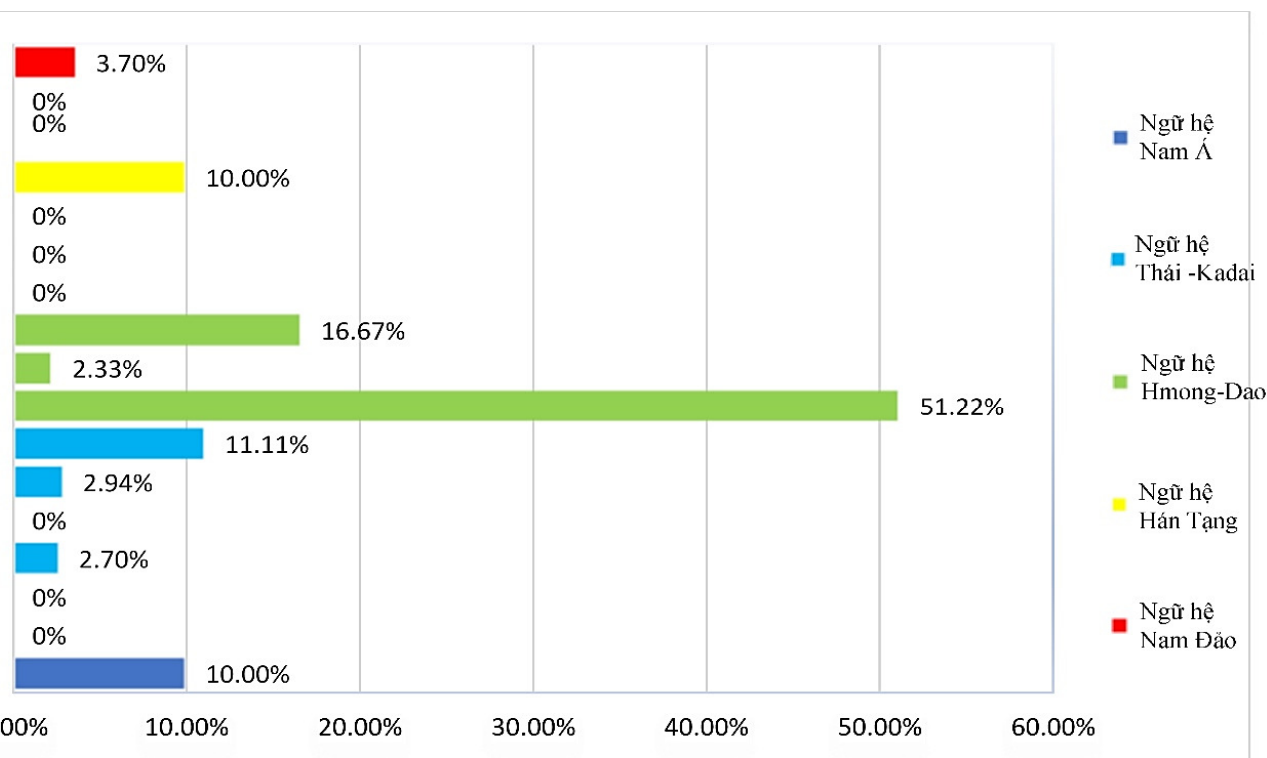
Xác định các SNP 600 cá thể thuộc nhóm đơn bội C ở 17 dân tộc Việt Nam

Sự phân bố các đa hình đại diện cho nhóm đơn bội C ở 17 dân tộc

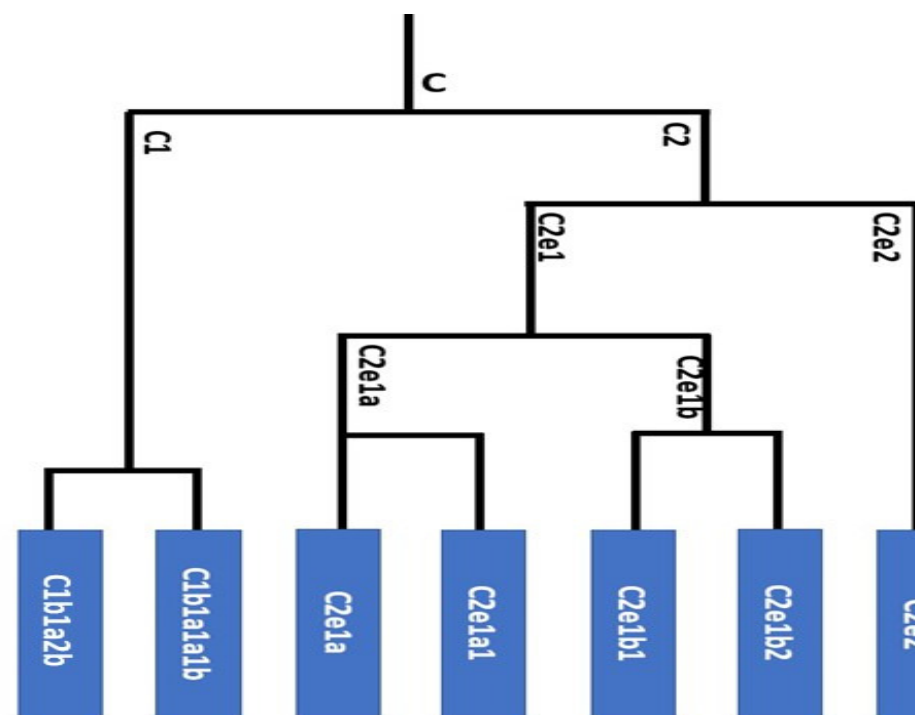
	Nhóm đơn bội	Dân tộc								
		Cờ Lao	Hà Nhi	Hmong	Gia Rai	Kinh	Nùng	Pathen	Sila	Dao
	C1b1a2b	1		10			1	1		1
	C1b1a1a1a				1					
657	C2e1a		1						3	
579	C2e1a1		3			1				
9	C2e1b1			8						
385	C2e1b2					3				
	C2e2			3		1		5		

Nghiên cứu xác định đa hình nucleotide đơn trên vùng không trao đổi chéo của nhiễm sắc thể Y của các cá thể thuộc 10 dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

Xác định các SNP 600 cá thể thuộc nhóm đơn bội O ở 17 dân tộc Việt Nam



Tần suất nhóm đơn bội C ở 17 dân tộc Việt Nam



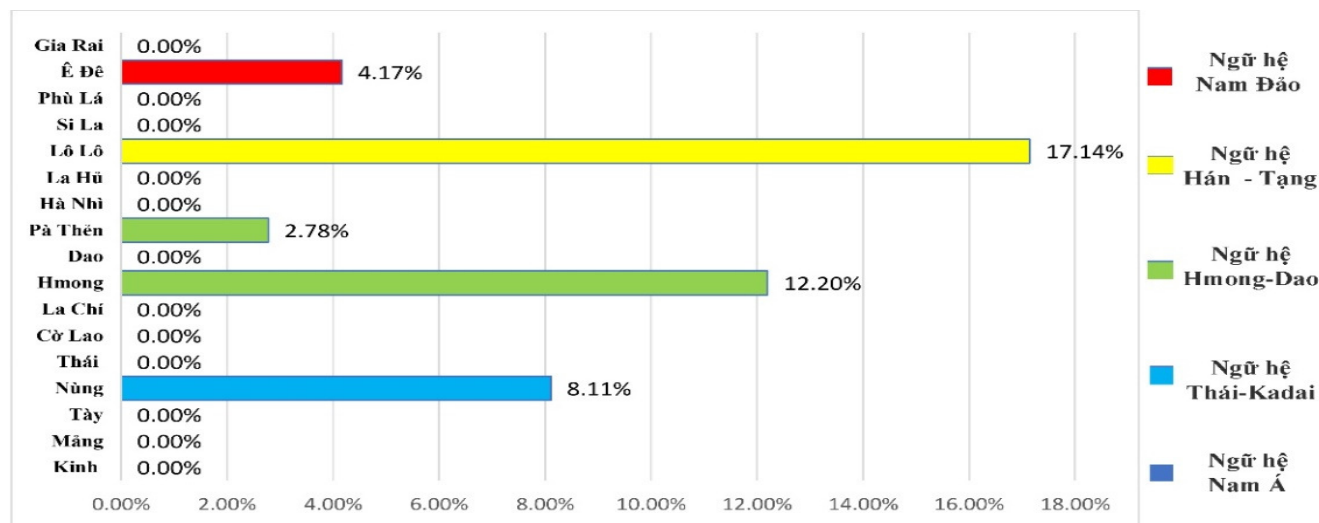
Cây phát sinh nhóm đơn bội C ở 17 dân tộc Việt Nam

hiên cứu xác định đa hình nucleotide đơn trên vùng không trao đổi chéo của nhi
c thể Y của các cá thể thuộc 10 dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp

Xác định các SNP 600 cá thể thuộc nhóm đơn bội D ở 17 dân tộc Việt Nam

Sự phân bố các đa hình đại diện cho nhóm đơn bội D ở 17 dân tộc

SNP	Nhóm đơn bội	Dân tộc					Tổng số
		Ê Đê	Hmong	Pà Thẻn	Lô Lô	Nùng	
M174	D*	1					1
CTS11577	D1		5	1			6
N1	D1a1a				6	3	9



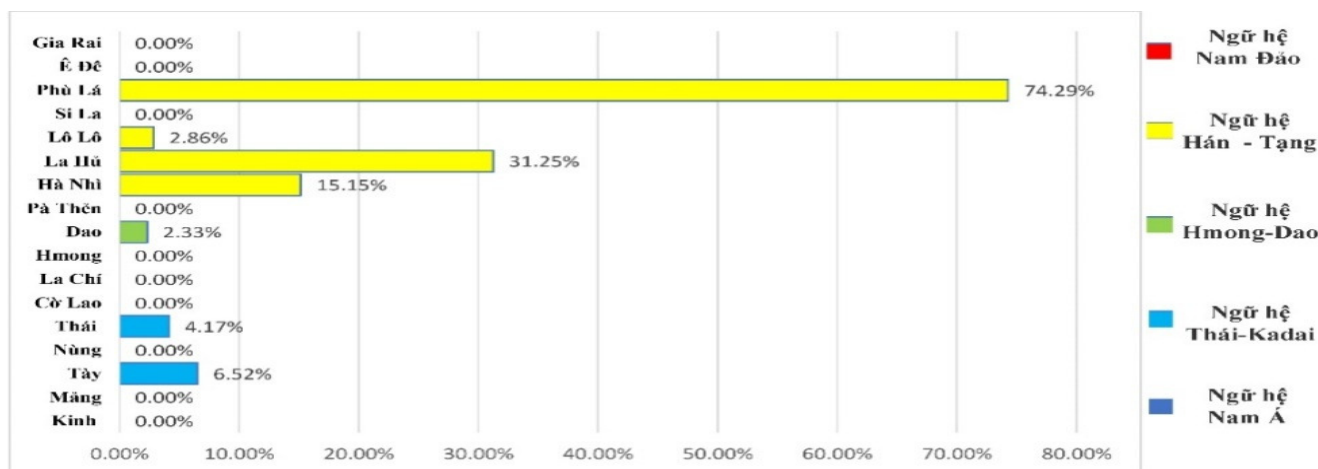
Tần suất nhóm đơn bội D ở 17 dân tộc Việt Nam

Nghiên cứu xác định đa hình nucleotide đơn trên vùng không trao đổi chéo của nhiễm sắc thể Y của các cá thể thuộc 10 dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

Xác định các SNP 600 cá thể thuộc nhóm đơn bội F ở 17 dân tộc Việt Nam

Sự phân bố các đa hình đại diện cho nhóm đơn bội F ở 17 dân tộc

SNP	Nhóm đơn bội	Dân tộc							Tổng số
		Hà Nhì	La Hủ	Lô Lô	Phù Lá	Tày	Thái	Dao	
M89	F*	5	10	1	26	3	1	1	47



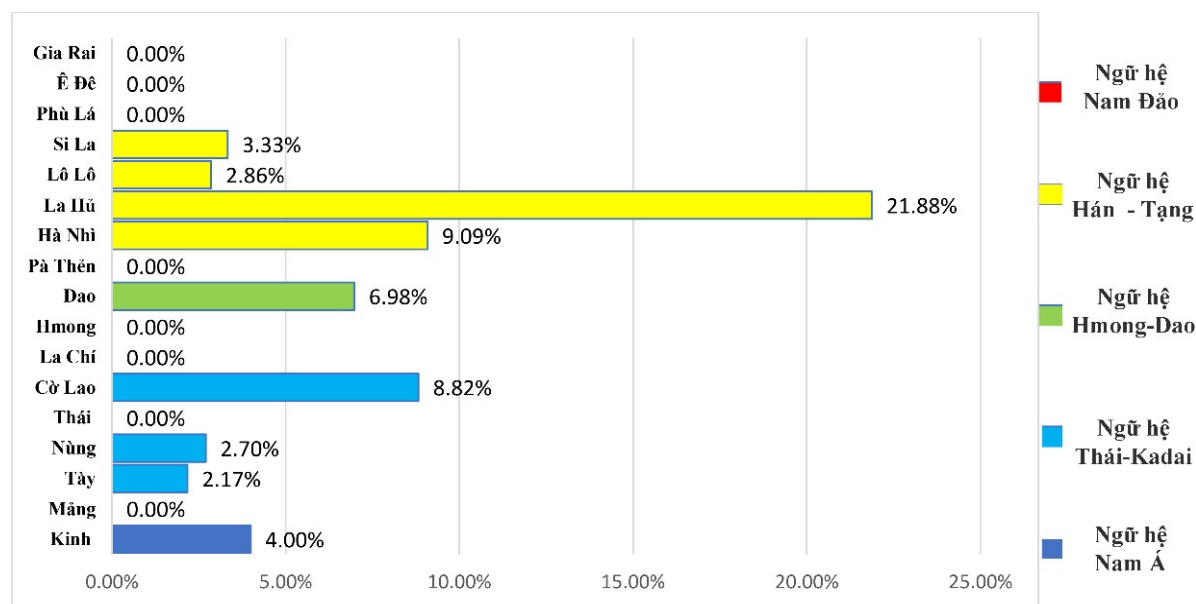
Tần suất nhóm đơn bội F ở 17 dân tộc Việt Nam

hiên cứu xác định đa hình nucleotide đơn trên vùng không trao đổi chéo của nhi
c thể Y của các cá thể thuộc 10 dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp

Xác định các SNP 600 cá thể thuộc nhóm đơn bội N ở 17 dân tộc Việt Nam

Sự phân bố các đa hình đại diện cho nhóm đơn bội N ở 17 dân tộc

SNP	Nhóm đơn bội	Dân tộc									Tổng số
		Cờ Lao	Hà Nhì	Kinh	La Hủ	Lô Lô	Nùng	Si La	Tày	Dao	
L665	N1c2b2	3	3	2	7	1	1	1	1	3	22



Tần suất nhóm đơn bội N ở 17 dân tộc Việt Nam

Nghiên cứu xác định đa hình nucleotide đơn trên vùng không trao đổi chéo của nhiễm sắc thể Y của các cá thể thuộc 10 dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

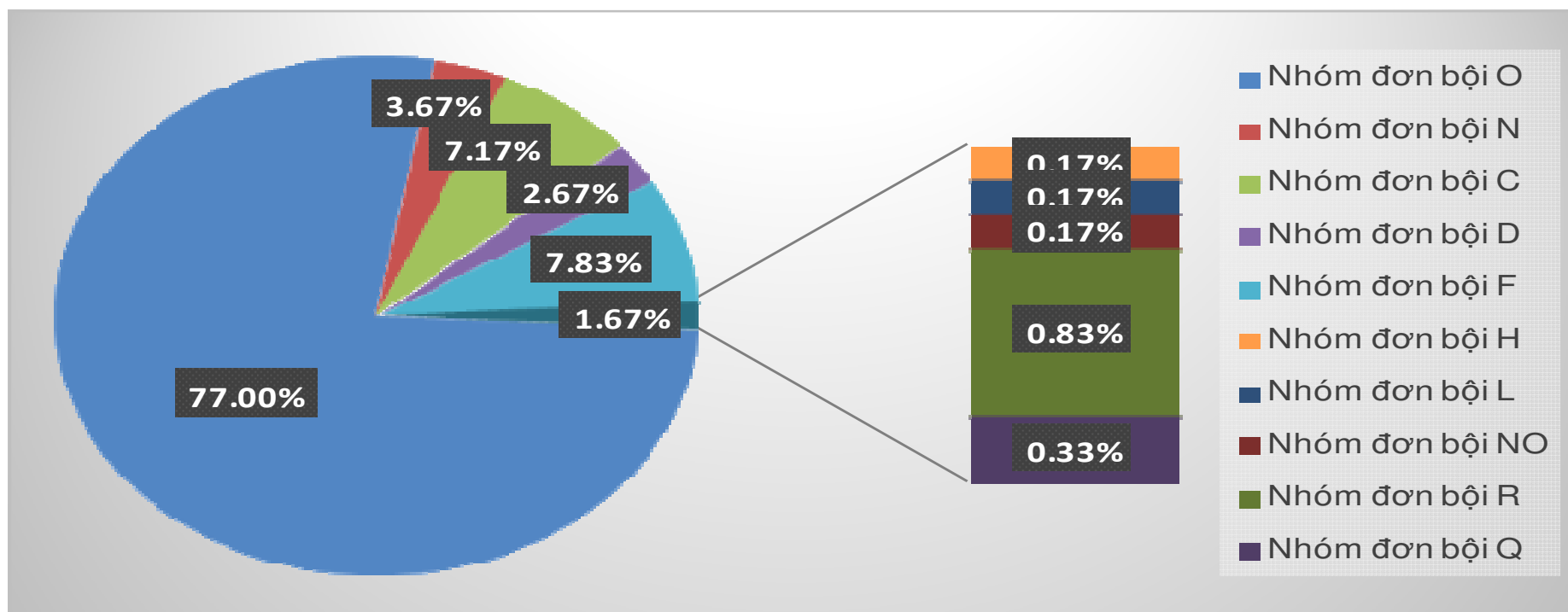
Xác định các SNP 600 cá thể thuộc nhóm đơn bội khác ở 17 dân tộc Việt Nam

Sự phân bố các đa hình đại diện cho nhóm đơn bội khác ở 17 dân tộc

SNP	Nhóm đơn bội	Dân tộc				Tổng số
		Kinh	Mảng	Gia Rai	Ê Đê	
487	H1a2a	1				1
27	L1a1			1		1
2313	NO		1			1
120	Q1a1a1	2				2
645	R1a1a1b			1	1	3
2123	R1a1a1b2a2a				1	
479	R2			1	1	2

ên cứu xác định đa hình nucleotide đơn trên vùng không trao đổi chéo của nhiễm / của các cá thể thuộc 10 dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

Tần suất các nhóm đơn bội NST Y ở 600 cá thể thuộc 17 dân tộc



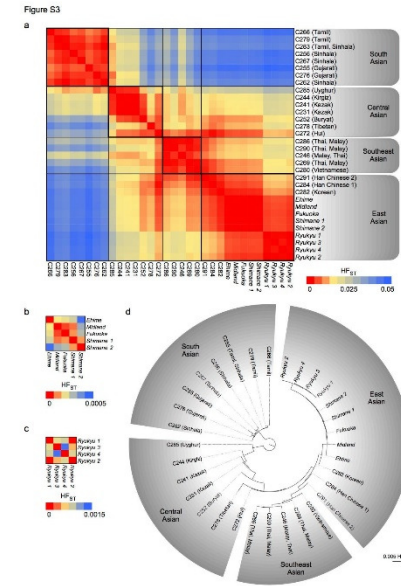
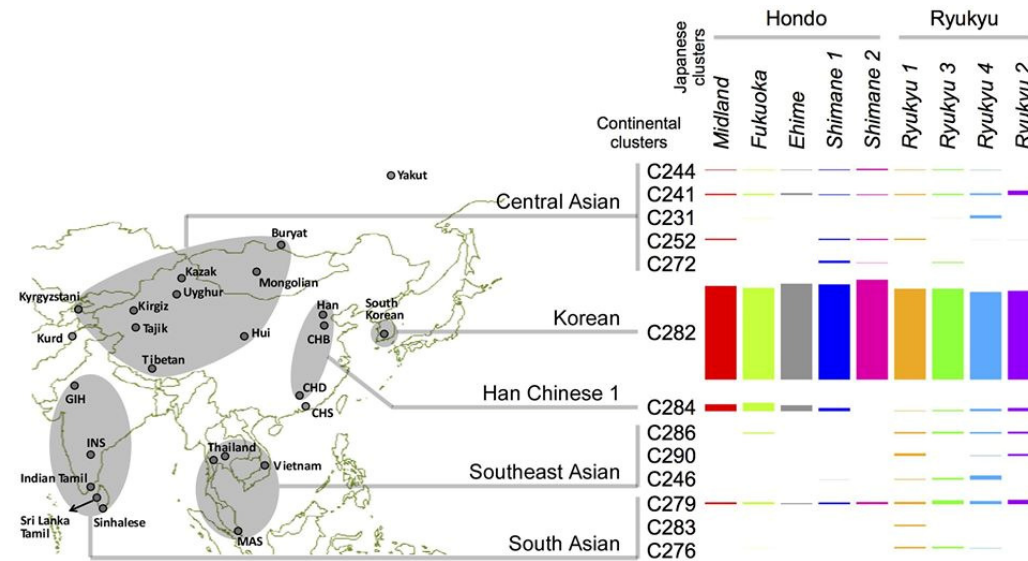
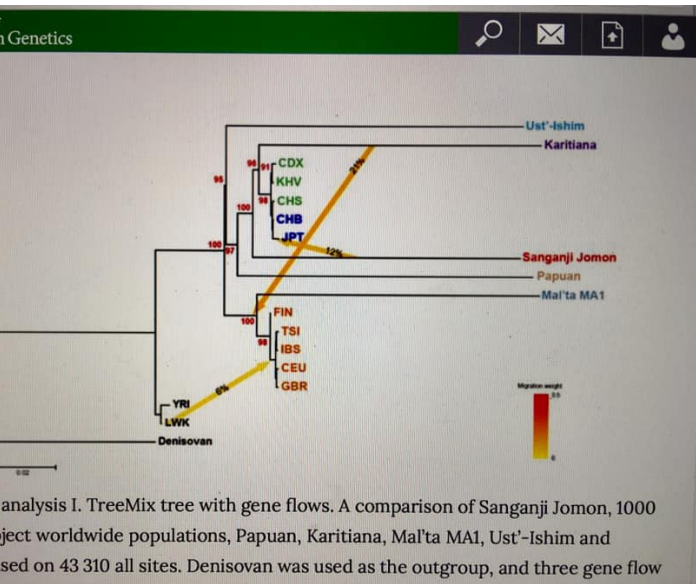
Nghiên cứu xác định đa hình nucleotide đơn trên vùng không trao đổi chéo của nhiễm sắc thể Y của các cá thể thuộc 10 dân tộc trong cộng đồng các dân tộc Việt Nam (tiếp)

Kết quả xác định các SNP của 600 cá thể thuộc nhóm đơn bội khác nhau ở 17 dân tộc Việt Nam

Số lượng các phân nhóm đơn bội NST Y ở 17 dân tộc

Nhóm đơn bội	Số lượng phân nhóm																	Tổng số
	Nam Á		Thái-Kadai			Hmong-Dao					Hán Tạng					Nam Đảo		
	Kinh	Mảng	Tày	Nùng	Thái	Cờ Lao	La Chi	Hmong	Dao	Pà Thẻn	Hà Nhi	La Hủ	Lô Lô	Sì La	Phù Lá	Ê Đê	Gia Rai	
	15	1	15	10	9	5	4	3	9	7	5	6	3	3	4	5	6	38
	1		1	1		1			1		1	1	1	1				1
	3			1		1		3	1	2	2			1			1	7
				1				1		1			1			1		3
			1		1				1		1	1	1		1			1
	1																	1
																	1	1
		1																1
																3	2	3
	1																	1
Tổng số	36	3	32	23	19	12	8	10	21	17	14	14	9	8	9	14	16	57

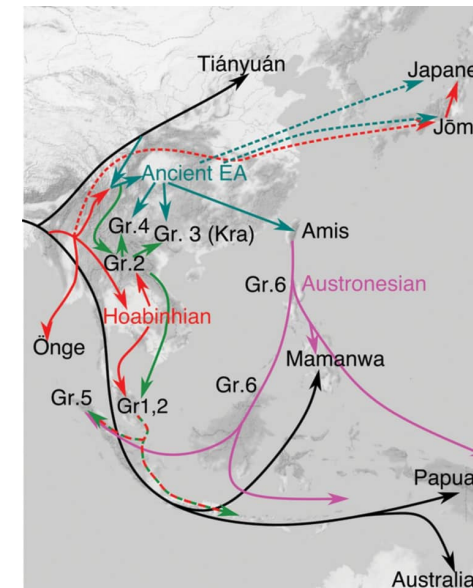
Quan hệ di truyền giữa người Nhật Bản và người Việt Nam



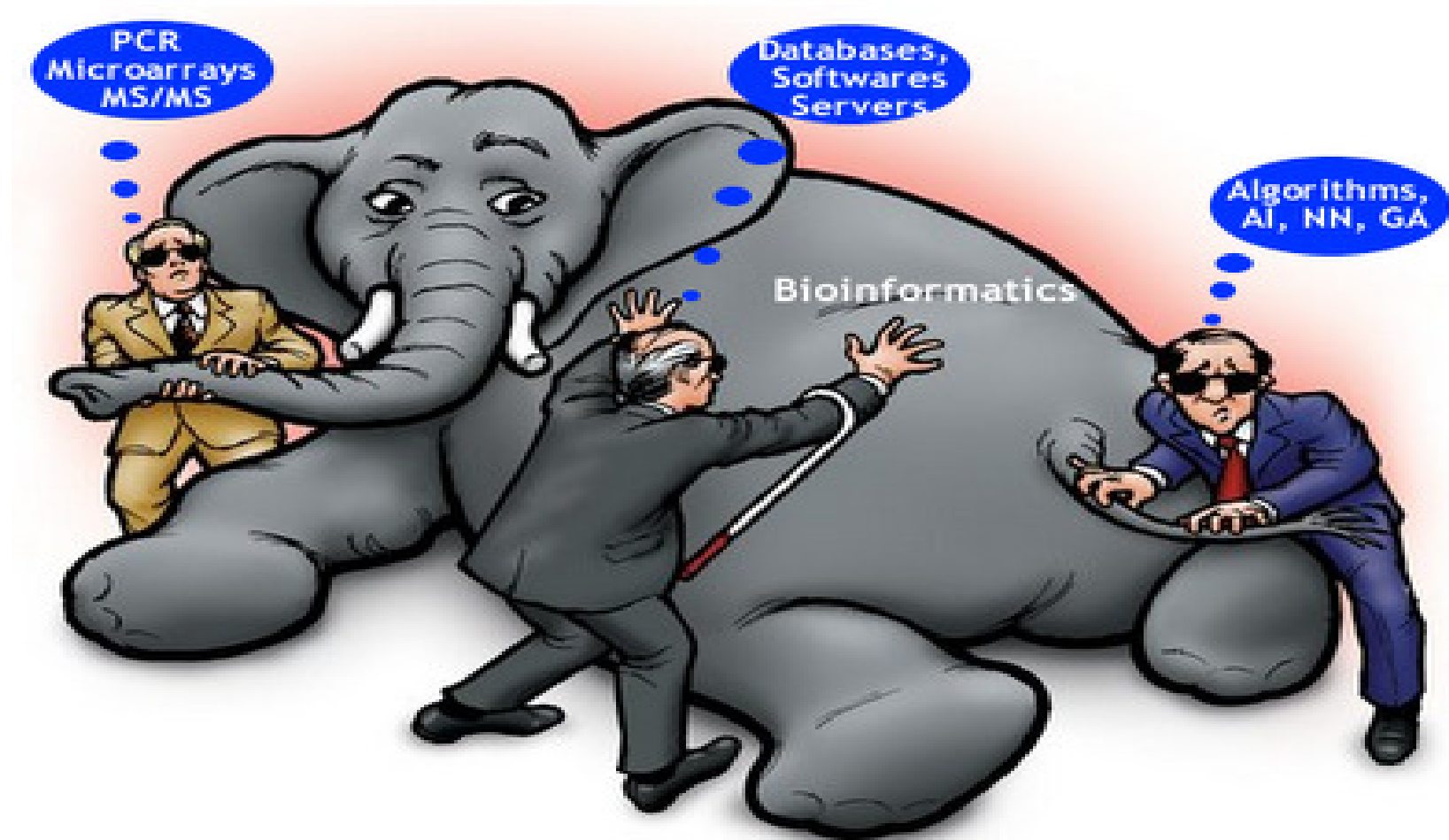
wa-Kiriyama et al. (2017). Hum Genet. 2017 Feb;62(2):213-221. doi: 10.1038/jhg.2016.110.

uchi et al. (2017). PLoS ONE 12(11): e0185487 (Hình 2,3)

ll et al. (2018). Science 361(6397):88-92. doi: 10.1126/science.aat3628.

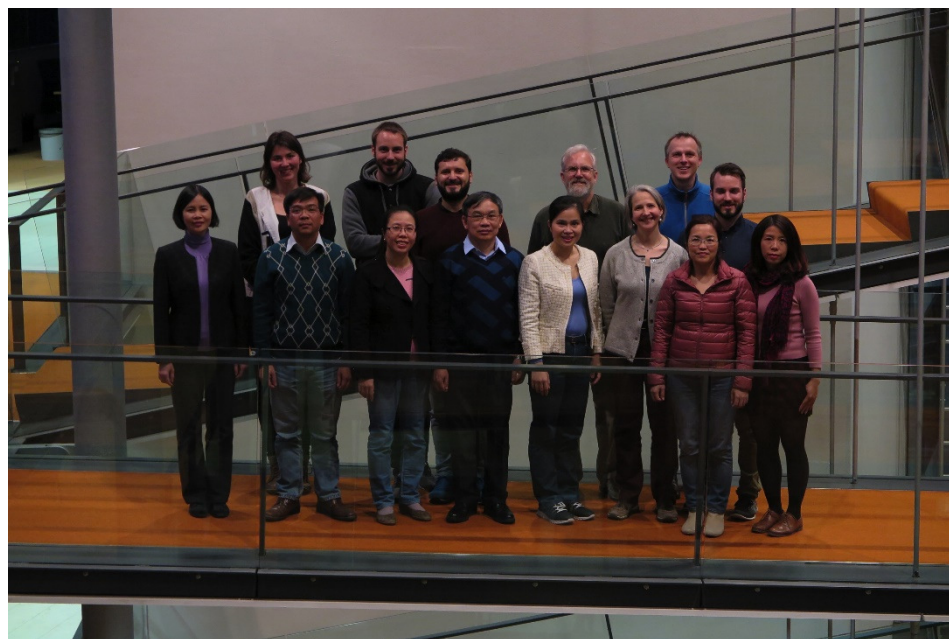


What are Genomics, Proteomics and Bioinformatics?





XIN TRÂN TRỌNG CẢM ƠN!



Công bố chấn động về gen: Người Việt có nguồn gốc từ Châu Phi

Công bố quốc tế về giải trình tự gen cho thấy người Việt có nguồn gốc từ châu Phi, không liên quan đến người Hán ở phía bắc Trung Quốc”.

VTC NEWS
Hội thảo cuộc sống

Cuộc thi Khám phá vẻ đẹp Việt Nam

THỜI SỰ VIDEO KINH TẾ THỂ THAO THẾ GIỚI GIÁO DỤC PHÁP LUẬT VĂN HÓA - GIẢI TRÍ SỨC KHỎE

BỆNH VÀ THUỐC DINH DƯỠNG TIN TỨC KHỎE ĐẸP TỰ VẤN AN TOÀN THỰC PHẨM

Công bố chấn động về gen: Người Việt có nguồn gốc từ châu Phi?

SỨC KHỎE Thứ Tư, 17/07/2019 14:51:00 +07:00 > 2

Like 12 Share Chia sẻ Tweet

Công bố quốc tế về giải trình tự gen cho thấy người Việt có nguồn gốc từ châu Phi, không liên quan đến người Hán ở phía bắc Trung Quốc.

Tạp chí di truyền quốc tế uy tín Human Mutation (IF 4.5) vừa công bố "Nghiên cứu về bộ gen của người Việt" do các nhà khoa học thuộc Viện nghiên cứu Tế bào gốc - công nghệ gen Vinmec (VRISG) thực hiện.

Là nghiên cứu lớn và công phu nhất về bộ gen người Việt từ trước đến nay, công trình mở ra cơ sở tham chiếu cho nhiều nghiên cứu Y - Sinh, từ đó góp phần giải đáp nhiều vấn đề sức khỏe của người Việt Nam hiện tại.

Công bố nghiên cứu bộ gen người Việt: Bất ngờ về nguồn gốc

17/07/2019 06:25 GMT+7

TTO - Nghiên cứu này có nhiều điểm mới được phát hiện về y sinh và nguồn gốc người Việt, trong đó có những điểm rất bất ngờ là có thể người Việt chúng ta có khởi nguồn từ... châu Phi và độc lập với người Hán.

- Xác định loại gen mới ra hy vọng cho bệnh nhân ung thư phổi
- Xác định được loại gen mới có liên quan đến chứng tâm thần phân liệt
- Phát hiện kiểu gen tác động đến hạnh phúc vợ chồng

992 32 f Lưu



VINMEC
INTERNATIONAL HOSPITAL

Tim kiếm...

Tất cả địa điểm English

Tin tức > Thông tin sức khỏe > Sức khỏe tổng quát

Sau kết quả nghiên cứu về bộ gen người Việt, Vinmec sẽ làm gì?

31/12/2019



Bài viết được tư vấn chuyên môn bởi **Tiến sĩ Trần Trung Kiên** - Chuyên viên nghiên cứu di truyền tại Phòng nghiên cứu, Viện nghiên cứu Tế bào gốc và Công nghệ gen Vinmec.

Theo **GS. TS Nguyễn Thanh Liêm** thì bộ gen người Việt khác rất nhiều gen người Hán Trung Quốc. Điều này chứng minh dù nước ta có bị đô hộ hàng nghìn năm, dân tộc ta không những gìn giữ được ngôn ngữ riêng, hệ gen không bị ảnh hưởng đáng kể. Đó là biểu hiện của sức đề kháng vô cùng mãnh liệt của người Việt.

XEM THÊM:

CÔNG TRÌNH “CƠ SỞ DỮ LIỆU BIẾN THỂ DI TRUYỀN NGƯỜI VIỆT NAM” CỦA VINMEC

Human Mutation
Variation, Informatics, and Disease

OFFICIAL JOURNAL
HGV
HUMAN GENOME VARIATION SOCIETY

DATABASES

A Vietnamese human genetic variation database

Vinh S. Le✉, Kien T. Tran, Hoa T. P. Bui, Huong T. T. Le, Canh D. Nguyen, Duong H. Do, Ha T. T. Ly, Linh T. D. Pham, Lan T. M. Dao, Liem T. Nguyen✉

First published: 10 June 2019 | <https://doi.org/10.1002/humu.23835> | Citations: 4

Read the full text >

PDF TOOLS SHARE

Human Mutation
Variation, Informatics, and Disease

OFFICIAL JOURNAL
HGV
HUMAN GENOME VARIATION SOCIETY

RESEARCH ARTICLE

Whole genome sequencing and mutation rate analysis of trios with paternal dioxin exposure

Nguyen Dang Ton, Hidewaki Nakagawa, Nguyen Hai Ha, Nguyen Thuy Duong, Vu Phuong Nhung, Le Thi Thu Hien, Huynh Thi Thu Hue, Nguyen Huy Hoang, Jing Hao Wong, Kaoru Nakano, Kazuhiro Maejima, Aya Sasaki-Oku, Tatsuhiko Tsunoda, Akihiro Fujimoto✉, Nong Van Hai✉ ... See fewer authors ^

First published: 03 July 2018 | <https://doi.org/10.1002/humu.23585> | Citations: 4

Communicating by George P. Patrinos

Human Mutation

Journal Abbreviation: HUM MUTAT
Journal ISSN: 1059-7794

About Human Mutation

Human Mutation is a peer-reviewed journal that offers publication of original Research Articles, Methods, Updates, Reviews, Database Articles, Mutations in Brief (MIBs), Rapid Communications, and Letters on broad aspects of mutation research in humans. Reports of novel DNA variations and their phenotypic consequences, reports of SNPs demonstrated as valuable for genomic analysis, descriptions of new molecular detection methods, and novel approaches to clinical diagnosis are welcomed. Novel reports of gene organization at the genome level, reported in the context of mutation investigation, may be considered.

Year	BioRx Journal Impact*	IF	Total Articles	Total Citations
2018/2019	-	4.453	185	1,100
2017	-	5.359	178	1,100
2016	-	4.601	142	1,100
2015	-	5.089	153	1,100
2014	-	5.144	174	1,100
2013	-	5.050	218	1,100
2012	-	5.213	224	1,100
2011	-	5.686	200	1,100
2010	-	5.956	209	1,100

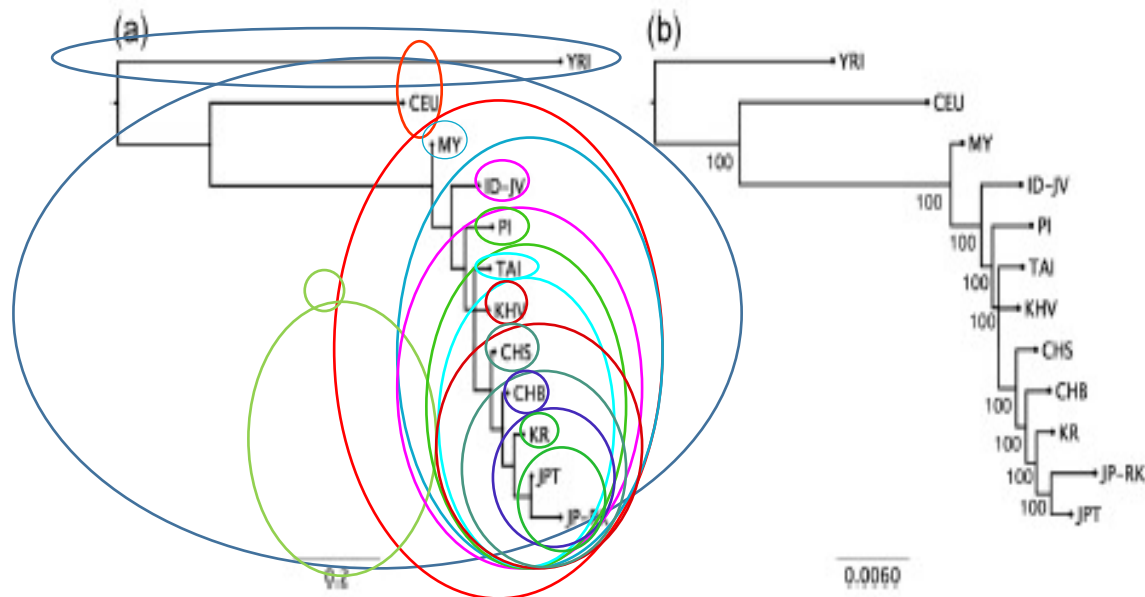


FIGURE 4 The phylogenetic trees of 12 populations where YRI is considered as the outgroup. (a) The neighbor-joining tree based on F_{ST} distances between populations. (b) The TreeMix tree where numbers on branches are bootstrap support values indicating the reliability of the tree. The scale bar reflects the amount of genetic drift between populations. CHB, Northern Han Chinese; CHS, Southern Han Chinese; ID-JV, Indonesian; JP-RK, Ryukyuan Japanese; JPT, Japanese; KHV, Kinh Vietnamese; KR, Korean; MY, Malay Malaysia; PI, Filipino Philippine; TAI, Thai Thailand; YRI, African

ance with the South-to-North geographical locations of these populations. The phylogenetic tree represents the evolutionary relationships between the populations. Figure 4 represents the constructed phylogenetic tree of 12 populations where YRI is considered as the outgroup. The tree topologies constructed by TreeMix and neighbor-joining methods are identical. All branches of the TreeMix tree have bootstrap support values of 100 indicating that the tree structure is

highly reliable. The tree structures show that SEA populations are closer to the YRI and CEU than EA populations. The positions of Asian populations in the tree agree with the South-to-North ordering of their geographical locations. The results from both phylogenetic tree reconstruction and PCA support the hypothesis that a population migration from Africa entered Asia along a South-to-North route (Abdulla et al., 2009; Chu et al., 1998).

mainly derived from the SEA ancestries and partly from the EA ancestries.

Finally, we examined the gene flows from other Asian populations to the KHV population using the F_3 statistic test. A significant negative F_3 (KHV; B, C) value indicates the existence of significant gene flows from populations B and C to KHV. The F_3 statistic tests did not reveal any significant gene flow from Asian populations to the KHV population. The findings explain to some extent the difference between allele frequencies of KHV and other populations.

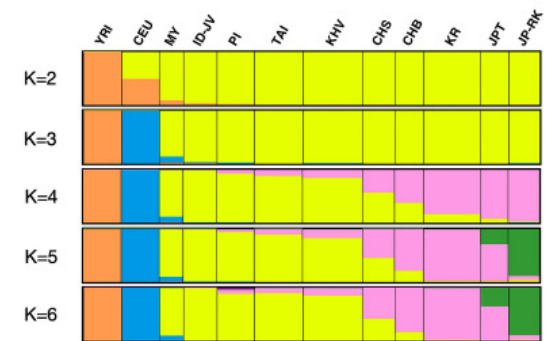


FIGURE 5 The ancestral population structure analysis of 12 populations with fastSTRUCTURE. The additional meaning of the colors: CEU, European; CHS, Southern Han Chinese; CHB, Northern Han Chinese; ID-JV, Javanese Indonesian; JP-RK, Ryukyuan Japanese; JPT, Japanese; KR, Korean; MY, Malay Malaysia; PI, Filipino Philippine; TAI, Thai Thailand; YRI, African

We compared phylogenetic trees in the study with two neighbor-joining trees reported by Simons Genome Diversity Project (Simons et al., 2016). Generally, our trees are concordant with the neighbor-joining trees. Note that the two neighbor-joining trees are not identical. Particularly, KHV and TAI populations are adjacent in the tree based on F_{ST} distances, but not adjacent in the tree based on pairwise divergence per nucleotide distances. Our trees are concordant with the neighbor-joining trees.

Vietnam has a complex history of thousands of years. Large-scale genomic analyses of KHV together with other Asian populations elucidated that KHV and other SEA populations mainly derived from the same SEA ancestry. The results from different analyses are generally consistent and support the hypothesis of population migration from Africa to Asia following the South-to-North route. Interestingly, we discovered that KHV and other SEA populations had similar genomic structures and close relationships. The findings suggest the usefulness of KHV for Vietnamese as well as other closely related populations.

Abstract

Large-scale human genome projects have created tremendous human genome databases for some well-studied populations. Vietnam has about 95 million people (14th largest country by population in the world) of which more than 86% are Kinh people. To date, genetic studies for Vietnamese people mostly rely on genetic information from other populations. Building a Vietnamese human genetic variation database is a must for properly interpreting Vietnamese genetic variants. To this end, we sequenced 105 whole genomes and 200 whole exomes of 305 unrelated Kinh Vietnamese (KHV) people. We also included 101 other previously published KHV exomes to build a Vietnamese human genetic variation database of 406 KHV people. The KHV database contains 24.81 million variants (22.47 million single nucleotide polymorphisms (SNPs) and 2.34 million indels) of which 0.71 million variants are novel. It includes more than 99.3% of variants with a frequency of >1% in the KHV population. Noticeably, the KHV database revealed 107 variants reported in the human genome mutation database as pathological mutations with a frequency above 1% in the KHV population. The KHV database (available at <https://genomes.vn>) would

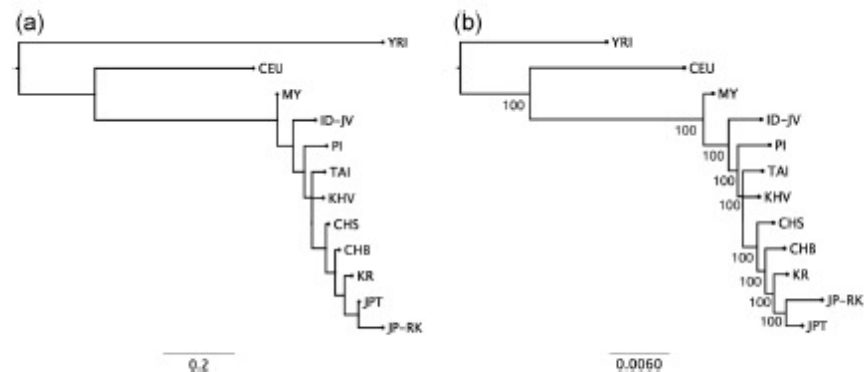


FIGURE 4 The phylogenetic trees of 12 populations where YRI is considered as the outgroup. (a) The neighbor-joining tree based on F_{ST} distances between populations. (b) The TreeMix tree where numbers on branches are bootstrap support values indicating the reliability of the nodes. The scale bar reflects the amount of genetic drift between populations. CHB, Northern Han Chinese; CHS, Southern Han Chinese; ID-JV, Indonesian Javanese; ID-RK, Indonesian Ryukyuan; JPT, Japanese; KHV, Kinh Vietnamese; KR, Korean; MY, Malay Malaysia; PI, Filipino Philippines; TAI, Thai

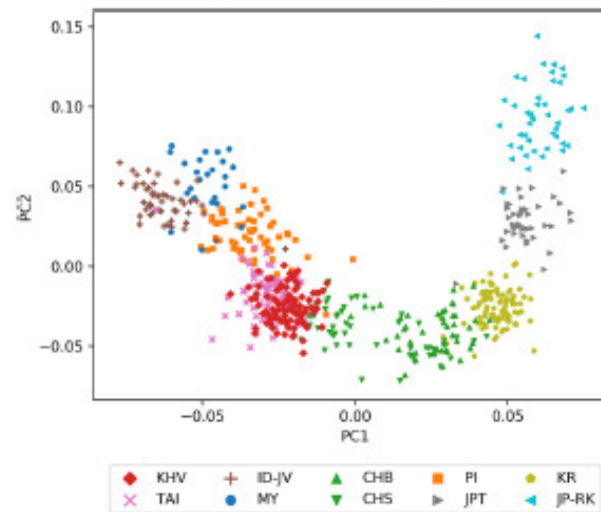


FIGURE 3 Principal component analysis of KHV and other Asian populations. CHB, Northern Han Chinese; CHS, Southern Han Chinese; ID-JV, Javanese Indonesia; JP-RK, Ryukyuan Japanese; JPT, Japanese; KHV, Kinh Vietnamese; KR, Korean; MY, Malay Malaysia; PI, Filipino Philippines; TAI, Thai

3.3.1 | Population relationship analysis

We performed PCA and phylogenetic tree reconstruction to investigate the genomic relationships among populations. The KHV and 11 other populations were classified into four main groups: (a) YRI; (b) CEU; (c) South East Asian (SEA) including Malay Malaysia (MY), Filipino Philippines (PI), Javanese Indonesia (ID-JV), Thai Thailand (TAI), Kinh Vietnamese (KHV); and (d) East Asian (EA) consisting of Southern Han Chinese (CHS), Northern Han Chinese (CHB), Korean (KR), Japanese (JPT), Ryukyuan Japanese (JP-RK). Figure 2 shows the geographical locations of the 12 populations.

The PCA result displays relationships among individuals of the 12 populations in Figure 3. The plot of the first two principal components shows that individuals from the same population were clustered into one group. The KHV and TAI populations were considerably overlapped and separated from other populations. We also observe an overlap between Southern and Northern Han Chinese populations. The CHS population is close to the SEA populations, while the CHB population is close to the EA populations.

The concordance with the South-to-North geographical locations of the 12 populations.

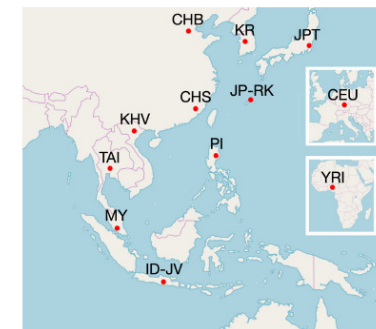


FIGURE 2 The geographical locations of 12 populations under the study African (YRI), European (CEU), Malay Malaysia (MY), Filipino Philippines (PI), Javanese Indonesia (ID-JV), Thai Thailand (TAI), Kinh Vietnamese (KHV), Southern Han Chinese (CHS), Northern Han Chinese (CHB), Korean (KR), Japanese (JPT), and Ryukyuan Japanese (JP-RK)

HUMAN GENOMICS

Ancient genomes document multiple waves of migration in Southeast Asian prehistory

Mark Lipson^{1*}, Olivia Cheronet^{2,3,4}, Swapan Mallick^{1,5}, Nadin Rohland¹, Marc Oxenham⁶, Michael Pietrusewsky⁷, Thomas Oliver Pryce^{8,9,10}, Anna Willis¹¹, Hirofumi Matsumura¹², Hallie Buckley¹³, Kate Domett¹⁴, Giang Hai Nguyen¹⁵, Hoang Hiep Trinh¹⁵, Aung Aung Kyaw¹⁶, Tin Tin Win¹⁶, Baptiste Pradier⁹, Nasreen Broomandkhoshbacht^{1,17}, Francesca Candilio^{18,19}, Piya Changmai²⁰, Daniel Fernandes^{2,3,21}, Matthew Ferry^{1,17}, Beatriz Gamarra^{3,4}, Eadaoin Harney^{1,17}, Jatupol Kampuansai^{22,23}, Wibhu Kutanan²⁴, Megan Michel^{1,17}, Mario Novak^{3,25}, Jonas Oppenheimer^{1,17}, Kendra Sirak^{3,26}, Kristin Stewardson^{1,17}, Zhao Zhang¹, Pavel Flegontov^{20,27†}, Ron Pinhasi^{2,3,*†}, David Reich^{1,5,17,*†}

Lipson *et al.*, *Science* **361**, 92–95 (2018) 6 July 2018

Southeast Asia is home to rich human genetic and linguistic diversity, but the details of past population movements in the region are not well known. Here, we report genome-wide ancient DNA data from 18 Southeast Asian individuals spanning from the Neolithic period through the Iron Age (4100 to 1700 years ago). Early farmers from Man Bac in Vietnam exhibit a mixture of East Asian (southern Chinese agriculturalist) and deeply diverged eastern Eurasian (hunter-gatherer) ancestry characteristic of Austroasiatic speakers, with similar ancestry as far south as Indonesia providing evidence for an expansive initial spread of Austroasiatic languages. By the Bronze Age, in a parallel pattern to Europe, sites in Vietnam and Myanmar show close connections to present-day majority groups, reflecting substantial additional influxes of migrants.

HUMAN GENOMICS

The prehistoric peopling of Southeast Asia

Hugh McColl^{1*}, Fernando Racimo^{1*}, Lasse Vinner^{1*}, Fabrice Demeter^{1,2*}, Takashi Gakuhari^{3,4}, J. Victor Moreno-Mayar¹, George van Driem^{5,6}, Uffe Gram Wilken¹, Andaine Seguin-Orlando^{1,7}, Constanza de la Fuente Castro¹, Sally Wasef⁸, Rasmi Shoocongdej⁹, Viengkeo Souksavatdy¹⁰, Thongsas Sayavongkhamdy¹⁰, Mohd Mokhtar Saidin¹¹, Morten E. Allentoft¹, Takehiro Sato¹², Anna-Sapfo Malaspinas¹³, Farhang A. Aghakhanian¹⁴, Thorfinn Korneliussen¹, Ana Prohaska¹⁵, Ashot Margaryan^{1,16}, ~~Peter de Barros Damgaard¹, Supannee Kaewsutthi¹⁷, Patcharee Lertrit¹⁷, Thi Mai Huong Nguyen¹⁸, Hsiao-chun Hung¹⁹, Thi Minh Tran¹, Huu Nghia Truong¹⁸, Giang Hai Nguyen¹~~, Shaiful Shahidan¹¹, Ketut Wiradnyana²⁰, Hiromi Matsumae⁴, Nobuo Shigehara²¹, Minoru Yoneda²², Hajime Ishida²³, Tadayuki Masuyama²⁴, Yasuhiro Yamada²⁵, Atsushi Tajima¹², Hiroki Shibata²⁶, Atsushi Toyoda²⁷, Tsunehiko Hanihara⁴, Shigeki Nakagome²⁸, Thibaut Deviese²⁹, Anne-Marie Bacon³⁰, Philippe Düringer^{31,32}, Jean-Luc Ponche³³, Laura Shackelford³⁴, Elise Patole-Edoumba³⁵, Anh Tuan Nguyen¹⁸, Bérénice Bellina-Pryce³⁶, Jean-Christophe Galipaud³⁷, Rebecca Kinaston^{38,39}, Hallie Buckley³⁸, Christophe Pottier⁴⁰, Simon Rasmussen⁴¹, Tom Higham²⁹, Robert A. Foley⁴², Marta Mirazón Lahr⁴², Ludovic Orlando^{1,7}, Martin Sikora¹, Maudé E. Phipps¹⁴, Hiroki Oota⁴, Charles Higham^{43,44}, David M. Lambert⁴, Eske Willerslev^{1,15,45†}

The human occupation history of Southeast Asia (SEA) remains heavily debated. Current evidence suggests that SEA was occupied by Hòabìnhian hunter-gatherers until ~4000 years ago, when farming economies developed and expanded, restricting foraging groups to remote habitats. Some argue that agricultural development was indigenous; others favor the “two-layer” hypothesis that posits a southward expansion of farmers giving rise to present-day Southeast Asian genetic diversity. By sequencing 26 ancient human genomes (25 from SEA, 1 Japanese Jōmon), we show that neither interpretation fits the complexity of Southeast Asian history: Both Hòabìnhian hunter-gatherers and East Asian farmers contributed to current Southeast Asian diversity, with further migrations affecting island SEA and Vietnam. Our results help resolve one of the long-standing controversies in Southeast Asian prehistory.

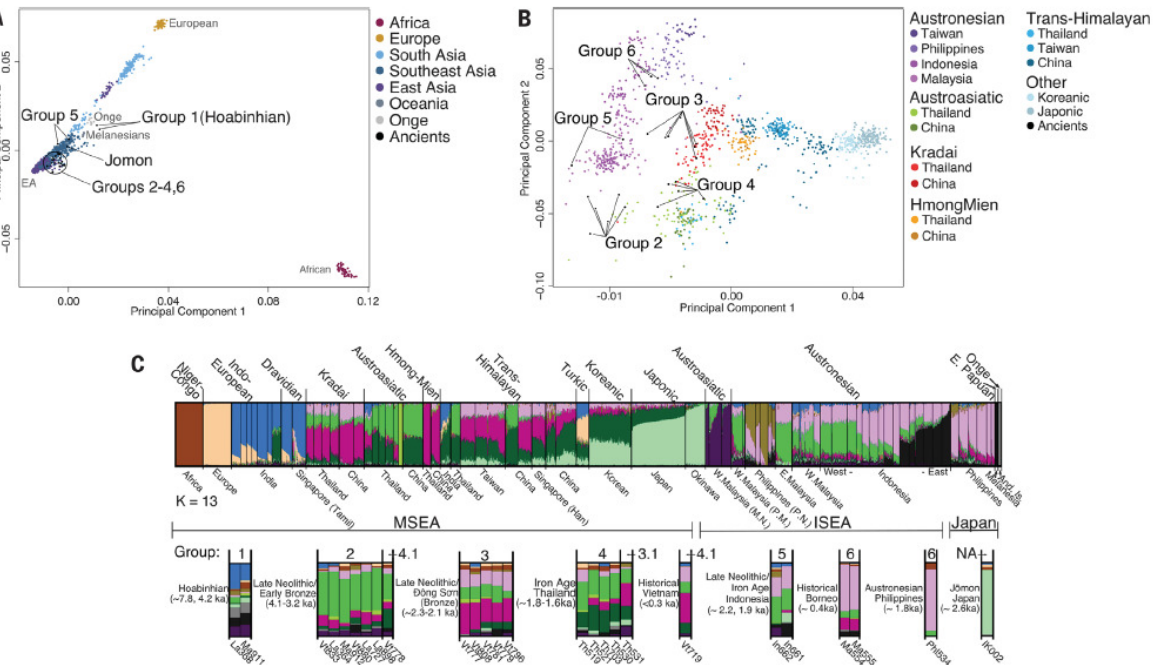


Fig. 2. Exploratory analyses of relationships of ancient Southeast Asian genomes to those of present-day populations. Ancient samples are projected on the first two components of PCAs for (A) worldwide populations and (B) a subset of populations from EA and SEA. (C) fastNGSadmix plot at $K = 13$ (II). We refer to the following present-day

language-speaking groups in relation to our ancient samples: Austroasiatic (bright green), Austronesian (pink), and Hmong-Mien (dark pink), along with a broad East Asian component (dark green). P.M., proto-Malay; M.N., Malaysian negrito; P.N., Philippines negrito; And. Is., Andaman Islands; NA, not applicable.

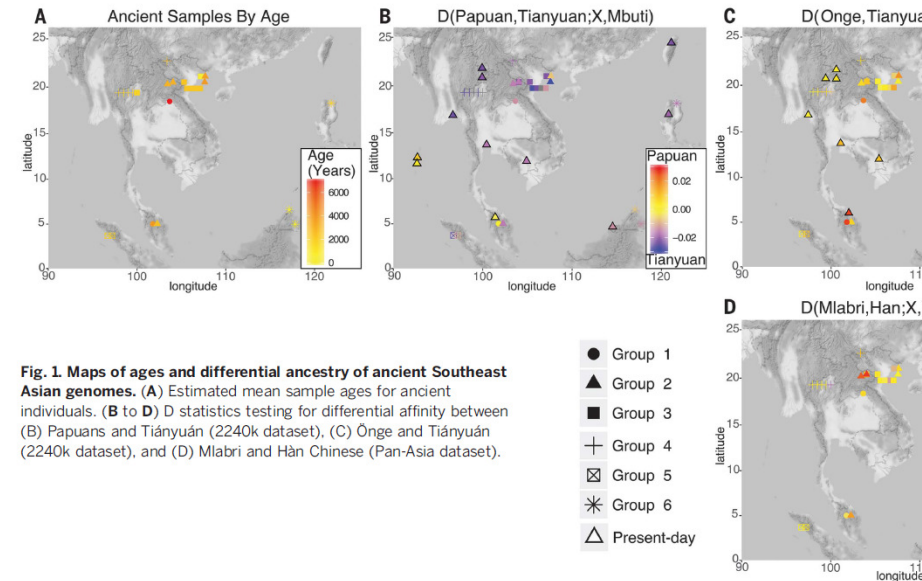
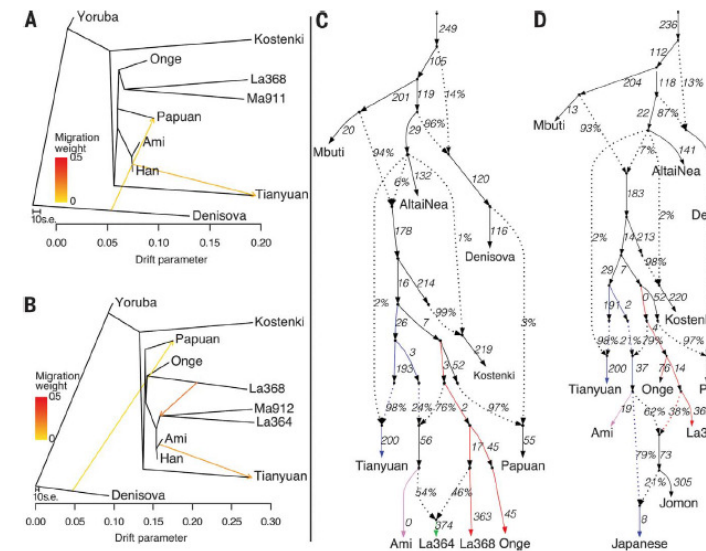


Fig. 1. Maps of ages and differential ancestry of ancient Southeast Asian genomes. (A) Estimated mean sample ages for ancient individuals. (B to D) D statistics testing for differential affinity between (B) Papuans and Tianyuan (2240k dataset), (C) Onge and Tianyuan (2240k dataset), and (D) Mlabri and Han Chinese (Pan-Asia dataset).

Fig. 3. Admixture graphs fitting ancient Southeast Asian genomes. TreeMix and qpGraph admixture graphs combining present-day populations and selected ancient samples with high single-nucleotide polymorphism coverage (II). (A) A graph including group 1 samples (Ma911 and La368) fits them as sister groups to present-day Onge. (B) A graph including the highest-coverage group 1 (La368) and group 2 (La364, Ma912) samples shows that group 2 receives ancestry from both group 1 and the East Asian branch. (C) Using qpGraph, we modeled present-day East Asians (represented by Amis) as a mixture of an Onge-like population and a population related to the Tianyuan individual. (D) The Jōmon individual is modeled as a mix of Hoabinhian (La368) and East Asian ancestry.



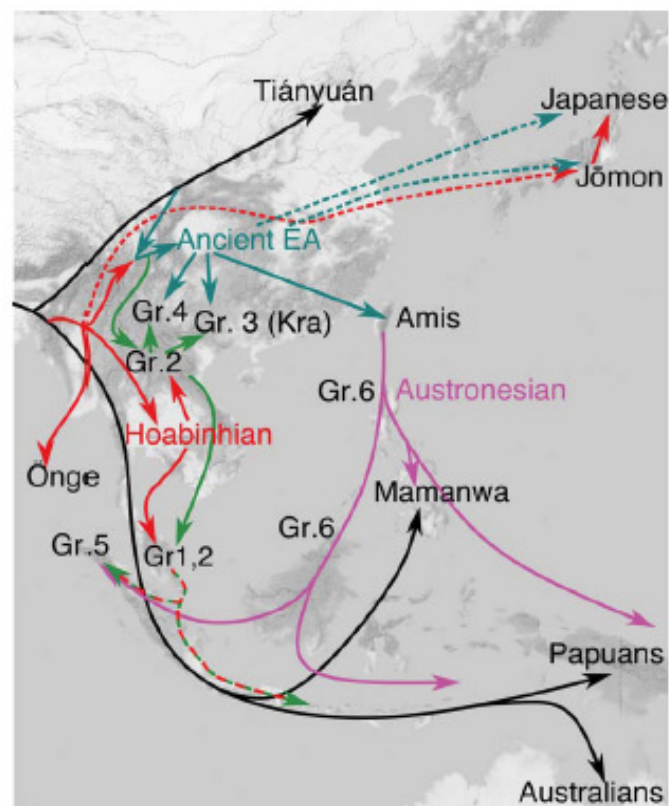


Fig. 4. Model for plausible migration routes into SEA. This schematic is based on ancestry patterns observed in the ancient genomes. Because we do not have ancient samples to accurately resolve how the ancestors of Jōmon and Japanese populations entered the Japanese archipelago, these migrations are represented by dashed arrows. A mainland component in Indonesia is depicted by the dashed red-green line. Gr, group; Kra, Kra-dai.

Present-day Southeast Asian populations derive ancestry from at least four ancient populations (Fig. 4). The oldest layer consists of mainland Hòabìnhians (group 1), who share ancestry with present-day Andamanese Önge, Malaysian Jehai, and the ancient Japanese Ikawazu Jōmon. Consistent with the two-layer hypothesis in MSEA,

we observe a change in ancestry by ~4 ka ago, supporting a demographic expansion from EA into SEA during the Neolithic transition to farming. However, despite changes in genetic structure coinciding with this transition, evidence of admixture indicates that migrations from EA did not simply replace the previous occupants. Additionally, late Neolithic farmers share ancestry with present-day Austroasiatic-speaking hill tribes, in agreement with the hypotheses of an early Austroasiatic farmer expansion (20). By 2 ka ago, Southeast Asian individuals carried additional East Asian ancestry components absent in the late Neolithic samples, much like present-day populations. One component likely represents the introduction of ancestral Kra-dai languages in MSEA (11), and another the Austronesian expansion into ISEA reaching Indonesia by 2.1 ka ago and the Philippines by 1.8 ka ago. The evidence described here favors a complex model including a demographic transition in which the original Hòabìnhians admixed with multiple incoming waves of East Asian migration associated with the Austroasiatic, Kra-dai, and Austronesian language speakers.