Southeast Asian Mitochondrial DNA Analysis Reveals Genetic Continuity of Ancient Mongoloid Migrations

S. W. Ballinger,* Theodore G. Schurr,* Antonio Torroni,* Y. Y. Gan,† J. A. Hodge,* K. Hassan,‡ K.-H. Chen§,[1] and Douglas C. Wallace*,,††,[2]**

*Departments of *Biochemistry, **Pediatrics, and ††Anthropology, Emory University School of Medicine, Atlanta, Georgia 30322, †Department of Biotechnology, Universiti Pertanian Malaysia, Serdang, 434000 Selangor, Malaysia, ‡Institute of Medical Research, Kuala Lumpur, Selangor, Malaysia and §Department of Mathematics, University of California, Long Beach, California*

Manuscript received December 10, 1990
Accepted for publication August 29, 1991

ABSTRACT

Human mitochondrial DNAs (mtDNAs) from 153 independent samples encompassing seven Asian populations were surveyed for sequence variation using the polymerase chain reaction (PCR), restriction endonuclease analysis and oligonucleotide hybridization. All Asian populations were found to share two ancient *Alu*I/*Dde*I polymorphisms at nps 10394 and 10397 and to be genetically similar indicating that they share a common ancestry. The greatest mtDNA diversity and the highest frequency of mtDNAs with *Hpa*I/*Hinc*II morph 1 were observed in the Vietnamese suggesting a Southern Mongoloid origin of Asians. Remnants of the founding populations of Papua New Guinea (PNG) were found in Malaysia, and a marked frequency cline for the COII/tRNA^{Lys} intergenic deletion was observed along coastal Asia. Phylogenetic analysis indicates that both insertion and deletion mutations in the COII/tRNA^{Lys} region have occurred more than once.

PREHISTORIC migrations in Southeast Asia have been the subject of much speculation (ISKANDAR 1976; BELLWOOD 1979, 1985; TURNER 1987; ZHAO and LEE 1989). Using dental morphological traits, TURNER (1983, 1987) hypothesized that two migrations originated from central China about 20,000–30,000 YBP (years before present). One group, the Sinodonts, expanded northward into China, Siberia and across the Bering land bridge to the New World. The second group, the Sundadonts, moved southward into Southeast Asia and Indonesia, and later through Melanesia, Micronesia, and Polynesia.

Using linguistic comparisons, BELLWOOD (1985) proposed two major prehistoric migrations into Southeast Asia. The first was an ancient "Australoid" migration from the Indo-Malaysian Archipelago which settled Australia and New Guinea about 40,000 YBP. The second was the more recent "Southern Mongoloid" or "Austronesian" migration that originated from the Fujian or Zhejian provinces of contemporary China and settled throughout much of island and mainland Southeast Asia about 4,000–6,000 YBP.

It is believed that the remnant Australoid populations in Southeast Asia were essentially replaced or assimilated by the southern Mongoloid migration (BELLWOOD 1985), whereas the major Australoid populations in Australia and New Guinea were relatively unaffected by the second migration (TURNER 1987). Assuming that these interpretations are correct, we anticipate that Southeast Asia may be a clinal zone between the Australoid (south and east) and Mongoloid (north and west) genotypes.

We have investigated the regional sequence variation of mtDNAs from Southeast Asian populations. By virtue of its matrilineal transmission (GILES *et al.* 1980; CASE and WALLACE 1981) and high mutation rate (BROWN, GEORGE and WILSON 1979; MIYATA *et al.* 1982; WALLACE *et al.* 1987), the mtDNA rapidly accumulates sequence changes along radiating female lineages, thus providing a detailed record of ancient migration patterns.

Previous analyses of mtDNAs from the Southeast Asian Aeta (Philippine Negritos) and other Asians [Japanese, Ainu (northern Japanese) and Koreans] has confirmed the Mongoloid affinity of these populations (HARIHARA *et al.* 1988). A detailed analysis of coastal and highland PNG mtDNAs has also confirmed the Asian association of these populations (STONEKING *et al.* 1990). Moreover, the coastal and highland PNG populations have become genetically differentiated (STONEKING *et al.* 1990), with coastal populations having the 9-bp COII/tRNA^{Lys} intergenic deletion (CANN and WILSON 1983; WRISCHNIK *et al.* 1987) in about 40% of their mtDNAs while the highland populations lack this marker (HERTZBERG *et al.* 1989; STONEKING *et al.* 1990). This marker is also associated with Pacific coastal and island populations, appearing at high frequencies in Melanesia and Polynesia, reaching fixation

[1] Present address: Institute of Public Health, National Taiwan University College of Medicine, Taipei, Taiwan.
[2] To whom correspondence should be addressed.

(100%) on some islands (HERTZBERG *et al.* 1989).

In an effort to integrate these Asian mtDNA studies into a coherent view of Southern Mongoloid migrations, we have conducted a detailed analysis of the mtDNAs from seven East Asian populations. The data provide evidence that: (1) the Vietnamese are the most diverse and, hence, the oldest population; (2) Malaysians retain remnants of haplotypes found in PNG; (3) coastal Asians have a striking frequency cline for the 9-bp deletion; and (4) both insertion and deletion mutations in the COII/tRNA^{Lys} intergenic region have occurred more than once.

MATERIALS AND METHODS

Populations: Asian blood samples were collected from 153 independent maternal pedigrees (unrelated through at least one generation). These included 14 Malaysian Chinese descendant from the Fujian/Guangdong region of south China; 14 Malays and 32 Malay Aborigines or "Orang Asli" [7 Temiar, 5 Semai, 1 Jakun, 2 Jeni, and 17 others of unidentified tribal origin] from the Malay peninsula; 30 Aborigines [24 Kadazan (Dusun), 2 Berungei, 3 Rungus, 1 Murut], and 2 Bisaya (Northern Borneo) from Sabah state (Borneo), Malaysia; 20 Han Chinese from Taiwan originating from central (Hunan) China; 28 Vietnamese; and 13 Koreans from South Korea (Seoul, Taejon, and Tamyang).

Methods: DNA was extracted from platelets of lymphoblastoid cell pellets and buffy coats (WALLACE, GARRISON and KNOWLER 1985). The mtDNA for each sample was PCR amplified (SAIKI *et al.* 1985) with AmpliTaqI polymerase (Perkin Elmer-Cetus) into 9 overlapping fragments that encompassed the entire mtDNA molecule (SCHURR *et al.* 1990) (APPENDIX A). All PCR fragments were digested with 18 restriction endonucleases [*Alu*I, *Ava*II, *Bam*HI, *Dde*I, *Hae*II, *Hae*III, *Hha*I, *Hin*cII, *Hin*fI, *Hpa*I, *Hpa*II, *Mbo*I, *Pst*I, *Pvu*II, *Rsa*I, *Taq*I, *Xba*I, *Xho*II] and electrophoresed on 1.0–4.0% NuSieve + 1.0% SeaKem agarose gels (FMC Bio-Products) containing 1 μg/ml ethidium bromide, to determine their respective restriction patterns through UV fluorescence.

In addition, each sample was screened for the COII/tRNA^{Lys} length mutations (CANN and WILSON 1983; WRISCHNIK *et al.* 1987) by differential oligonucleotide hybridization (SCHURR *et al.* 1990). The results were confirmed by restriction analysis.

Sequence divergence: Inter- and intrapopulational comparisons of mtDNA haplotype divergences were estimated using maximum likelihood estimates based upon nucleotide counting (NEI and TAJIMA 1983), using the computer program DREST (generously provided by L. JIN). This procedure considers the ratio of shared sites to the total number of sites between two haplotypes, and the mean length of the restriction enzyme recognition sequences to calculate an initial estimate of π (the probability that the two mtDNAs have different nucleotides at a given nucleotide position). Using this initial estimate, π is solved iteratively using Equation 28 and sequence divergence (δ) estimated by Equation 21 (NEI and TAJIMA 1983).

Phylogenetic analysis: Phylogenetic trees were inferred from the restriction site data under the principle of maximum parsimony using PAUP (Version 3.0m, Swofford 1990). A variety of rooting techniques using hypothetical ancestors inferred from hypothesized African haplotypes [*e.g.*, Ancestor "a" from CANN, STONEKING and WILSON (1987)] and Caucasian haplotypes (unpublished data) were used to determine genealogical relationships of the Asian mtDNA haplotypes. All methods produced similar results.

RESULTS

Restriction analysis: A total of 191 polymorphic sites and 106 haplotypes were observed in these mtDNAs (APPENDIX B), with an average of 390 independent sites per genome or approximately 10% of the mtDNA sequence being screened. Only 8 haplotypes (25, 28, 33, 51, 54, 55, 62 and 83) were found to be shared between 2 or more populations. Twenty "haplotype groups" were identified, distinguished by shared polymorphisms. Table 1 lists each sample and its corresponding haplotype and haplotype group. Figure 1 shows the distribution and frequencies of these haplotype groups in Southeast Asia. Of all the haplotype groups, "D" was the most prevalent type observed, with "A," "E," and "F" the next most frequent. These four groups are distributed in nearly every population sampled.

Differences at *Hin*cII/*Hpa*I sites also revealed several informative groupings of haplotypes. MtDNAs with the *Hin*cII/*Hpa*I site loss at np 12406 [*Hpa*I/*Hin*cII morph-1 (DENARO *et al.* 1981; BLANC *et al.* 1983); group A] occurred at the highest frequencies in the Vietnamese and Malay Aborigines (32.1% and 28.1%, respectively), and were found in almost all of the populations surveyed. A significant proportion of these mtDNAs (80.0%) also had a combined *Hae*II np 9052/*Hha*I np 9053 site loss. The *Hin*cII np 207 site gain previously observed in PNG populations (STONEKING, BHATIA and WILSON 1986; STONEKING *et al.* 1990) was also found in a Malay peninsula population (MM9, haplotype 90). Another *Hin*cII site loss at np 7853 [morph-5 (BLANC *et al.* 1983)] was seen in Vietnamese, Taiwanese Han and Sabah Aborigine mtDNAs and defined haplotype group B. In addition, several Sabah Aborigines had a new *Hin*cII site loss at np 7937. Lastly, the *Hin*cII np 12026 site gain in the Vietnamese [seen previously in one Australian aborigine (CANN, BROWN and WILSON 1984)] and the np 1004 site loss observed in the Vietnamese and Taiwanese Han [previously seen in one Japanese (HORAI, GOJOBORI and MATSUNAGA 1984)] create separate haplotype groups P and Q, respectively.

The combined *Alu*I np 10397 and *Dde*I np 10394 sites essentially split all haplotypes within these Asian populations into two major clusters (Figure 2). Several haplotypes had only the *Dde*I site, which creates a semisite for *Alu*I at np 10397. Thus the *Dde*I site is not only necessary, but precedes the creation of the *Alu*I site at np 10397. Both sites may be lost concurrently via a single base substitution in the overlapping recognition sequences.

TABLE 1

Sample haplotypes and haplotype groupings

Sample	Haplotype	Group	Sample	Haplotype	Group	Sample	Haplotype	Group	Sample	Haplotype	Group
MC01	17	N	VN28	53	A	MM04	86	O	KN08	100	C
MC02	18	N	VN29	54	D	MM05	87	E	KN09	101	D
MC03	19	N	VN30	55	C	MM06	88	E	KN10	102	G
MC04	20	N	MA01	69	A	MM07	62	F	KN11	103	G
MC05	21	F	MA02	70	F	MM08	89	G	KN12	104	K
MC06	22	E	MA03	71	A	MM09	90	S	KN13	105	M
MC07	23	T	MA04	62	F	MM10	91	A	SA01	106	G
MC08	24	O	MA05	62	F	MM11	49	D	SA02	107	B
MC10	25	L	MA06	72	D	MM12	92	F	SA03	108	E
MC11	26	F	MA07	73	I	MM13	93	E	SA04	109	G
MC12	27	N	MA08	74	D	MM14	94	G	SA05	110	D
MC13	25	L	MA09	75	I	TW01	56	H	SA06	111	D
MC14	28	H	MA10	76	J	TW02	28	H	SA07	83	G
MC15	29	O	MA11	77	F	TW03	54	D	SA08	112	N
VN01	30	A	MA12	78	I	TW04	57	D	SA09	106	G
VN02	31	E	MA13	33	A	TW05	57	D	SA10	113	E
VN04	32	Q	MA14	33	A	TW06	58	D	SA11	114	E
VN05	33	A	MA15	33	A	TW07	59	C	SA12	114	E
VN06	34	M	MA16	33	A	TW08	55	C	SA13	115	D
VN07	35	T	MA17	75	I	TW09	60	C	SA14	112	N
VN08	36	C	MA18	72	D	TW10	61	A	SA15	109	G
VN09	37	B	MA19	79	D	TW11	62	F	SA16	111	D
VN10	37	B	MA20	33	A	TW12	62	F	SA17	106	G
VN11	38	E	MA21	76	J	TW13	25	L	SA18	116	D
VN12	39	A	MA22	72	D	TW14	63	B	SA19	117	E
VN13	40	F	MA23	74	D	TW15	64	F	SA20	118	E
VN14	41	E	MA24	62	F	TW16	65	R	SA21	111	D
VN15	37	B	MA25	33	A	TW17	66	T	SA22	109	G
VN16	42	B	MA27	80	D	TW18	51	A	SA23	106	G
VN17	43	B	MA28	62	F	TW19	67	B	BS24	115	D
VN18	44	P	MA29	81	I	TW20	68	Q	SA25	119	G
VN19	45	A	MA30	82	I	KN01	25	L	SA26	54	D
VN21	46	A	MA31	82	I	KN02	95	F	SA27	120	A
VN22	47	A	MA32	33	A	KN03	96	K	SA28	121	D
VN23	48	A	MA33	72	D	KN04	97	L	SA29	115	D
VN24	54	D	MM01	83	G	KN05	98	A	SA30	114	E
VN25	50	T	MM02	84	D	KN06	99	A	SA31	55	C
VN26	51	A	MM03	85	A	KN07	96	K	SA32	122	E
VN27	52	B									

Sample abbreviations: MC, Malaysian Chinese; VN, Vietnamese; MA, Malay Aborigines; MM, Malays; TW, Taiwanese Han; KN, Koreans; SA, Sabah Aborigines. Haplotype groupings were classified according to polymorphic sites that were shared within each group. The site gains and losses for each haplotype group relative to the published sequence (ANDERSON *et al.* 1981) are bold face and non-bold face, respectively; slashes between enzyme letters or sites indicate non-independent events. Restriction sites enclosed in brackets indicate sites that frequently accompany the definitive sites with that haplotype group. The letter designation of restriction enzymes are as follows: a, *Alu*I; b, *Ava*II; c, *Dde*I; e, *Hae*III; f, *Hha*I; g, *Hin*fI; h, *Hpa*I; i, *Hpa*II; j, *Mbo*I; k, *Rsa*I; l, *Taq*I; m, *Bam*HI; n, *Hae*II; o, *Hin*cII. Haplotype groups are indicated by capital letters and consist of sets of polymorphic restriction sites: [A]-12406h/12406o, **16517e,** [9052n/9053f]; [B]-7853o, **10394c, 10397a;** [C]-3534c/3537a, **10394c,** 15234g/**15235j, 16517e;** [D]-**16517e;** [E]-**10394c, 10397a, 16517e;** [F]-**10394c, 10397a;** [G]-**16389g**/16390b, **10394c, 10397a,** 7598f, **16517e;** [H]-**663e, [16517e];** [I]-**10143a, 9326n/9329f, 10394c, 10397a,** [951j, **1063e**]; [J]-4711i, 11403g, **131180j,** 1715c, **10394c;** [K]-**4830n/4831f, 10394c, 10397a;** [L]-5176a, **10394c, 10397a;** [M]-**10394c, [16517e];** [N]-16517e; [O]-13366b/**13367j/13368m, [16517e];** [P]-**12026h/12026o;** [Q]-1002q/1004o; [R]-13259o/**13261a, 10394c, 10397a, 16517e;** [S]-**207h/207o, 15606a;** [T]-**16389g**/16390b, **16517e.** The COII/tRNA^{Lys}9-bp deletion occurs within haplotype groups A–D and F; the 4-bp addition occurs within haplotype groups A and M.

Asian COII/tRNA^{Lys} intergenic length mutation haplotypes: The 9-bp COII/tRNA^{Lys} deletion (CANN and WILSON 1983; WRISCHNIK *et al.* 1987) was observed in 25 individuals from all seven populations, comprising 16.3% of the samples within this study. Table 2 shows the overall distribution of the deletion in the populations analyzed. Haplotype groups C and D include 14 of the 17 deletion haplotypes (Table 1; APPENDIX B). Although these haplotype groups form 2 distinct clades distinguished by the *Dde*I np 10394 site, they all have the same *Hin*cII sites [morph-2 (BLANC *et al.* 1983)], and can be derived from haplotype 54. In contrast, the Vietnamese deletion haplotype 43 shares several sites with Group B haplotypes

Haplotype Group

	A B C D E F G H I J K L M N O P Q R S T	n
Koreans	2 0 1 1 0 1 1 1 0 0 3 2 1 0 0 0 0 0 0	13
Malay	2 0 0 2 3 2 3 0 0 0 0 0 0 0 1 0 0 0 1 0	14
Malay Aborigines	9 0 0 8 0 6 0 0 7 2 0 0 0 0 0 0 0 0 0	32
Malaysian Chinese	0 0 0 0 1 2 0 1 0 0 0 2 0 5 2 0 0 0 0 1	14
Sabah Aborigines	1 1 1 [10] 8 0 9 0 0 0 0 0 2 0 0 0 0 0	32
Taiwanese Han	2 2 3 4 0 3 0 2 0 0 0 1 0 0 0 0 1 1 0 1	20
Vietnamese	9 6 2 2 3 1 0 0 0 0 0 1 0 0 1 1 0 0 2	28

FIGURE 1.—Map of Southeast Asia Showing mtDNA Sample Localities. Haplotype groups are described in Table 1. The number of mtDNAs within each haplotype group are indicated below each corresponding letter, and the total number of individuals in each population indicated under "*n*."

and must have been derived from an unrelated mtDNA (*i.e.*, haplotype 37), indicating that the COII/tRNA^{Lys} deletion has occurred more than once.

In addition, two individuals (VN6 and SA27) had an insertion of approximately 4-bp in the COII/tRNA^{Lys} region, yet these haplotypes (34-Vietnamese and 120-Sabah Aborigine) are distinct and differ by at least 8 mutational events. Consequently, insertion mutations must also have occurred at least twice in Southeast Asian populations.

Genetic divergence: Table 3 presents the genetic divergence estimates for intra- and interpopulational comparisons (NEI and TAJIMA 1983). The highest intrapopulational divergence was observed within the Vietnamese at 0.00236 (0.236%). The lowest was within the Malay Aborigines and Taiwanese Han at 0.148% and 0.145%, respectively.

Phylogenetic analysis: A tree generated using a hypothesized ancestor (haplotype HYPANC) is presented in Figure 2. The two branches of this tree are defined by the *Dde*I and *Alu*I sites at nps 10394 and 10397, respectively. The majority of the deletion haplotypes cluster together (groups D* and C) except for haplotypes 21, 43 and 61. Most of the distinct branches within the tree encompass multiple populations, indicating that some haplotype groups (A–E and G) may represent common ancient Asian lineages. Overall there are few population-specific groupings of haplotypes within the network. One exception is haplotype group I in which five Malay aboriginal haplotypes (73, 75, 78, 81, 82) are associated.

Combining the Southeast Asian mtDNA data with that of PNG indicates that specific haplotype groupings are more characteristic of isolated populations. A tree of both populations (STONEKING *et al.* 1990) is presented in Figure 3. Two additional haplotype groups now stand out, group S, defined by the *Hinc*II/*Hpa*I np 207 site gain, and group U, a subgroup of F. Like the Orang Asli group I, groups S and U are isolated, occurring predominately in highland PNG. Other PNG haplotypes are dispersed within the other Southeast Asian haplotype groups. These include type P150 which falls within group A and deletion types P119–P130 which fall within group D*.

DISCUSSION

Similarity of mongoloid types: Analysis of Southeast Asian mtDNA variation indicates that all extant populations were derived from a common ancestral population which encompassed most of the variation. The mean of the intrapopulational divergence is 0.182%, while the mean interpopulational divergence corrected for intrapopulational divergence (NEI and TAJIMA 1983) is about one-sixth this value or 0.030%, with a range of 0.019% to 0.053% (Table 3). Thus, it would appear that most of the mtDNA variation is shared between the Southeast Asian populations and predated the present geographic subdivision. Of the current populations, the Vietnamese have the greatest intrapopulational genetic divergence (0.236%) suggesting that it is the oldest. Since Vietnam was colonized by a southeast China migration, this would imply a southern Chinese origin of Mongoloid people about 59,000 to 118,000 YBP (assuming that mtDNA divergence is 2–4% per million years, CANN, BROWN and WILSON 1984; CANN, STONEKING and WILSON 1987; NECKELMANN *et al.* 1987, 1989; WALLACE *et al.* 1987).

Haplotype group A (Table 1), which was present in six of the populations, further substantiates our previous proposal that the *Hinc*II/*Hpa*I morph 1 poly-

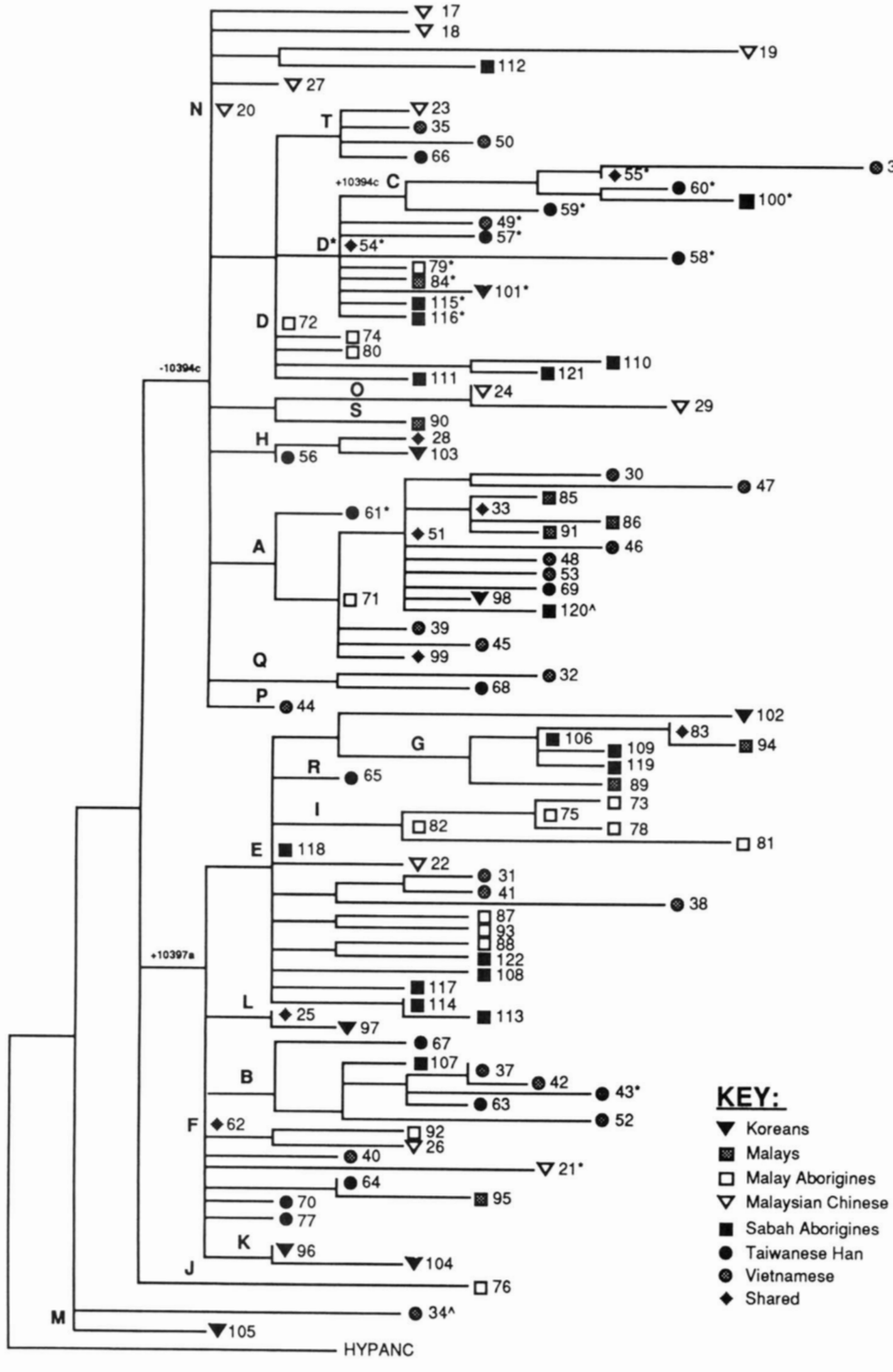

FIGURE 2.—Phylogeny of Southeast Asian mtDNA haplotypes. The length of this tree is 237 steps. It was rooted from hypothetical ancestor "a" from CANN, STONEKING and WILSON (1987). Although no shorter trees were found, shorter trees may exist, and the number of trees equal in length is probably large. Letters indicate the major haplotype groupings, whereas symbols designate haplotype origin. Asterisks (*) or hats (Λ) by a haplotype number indicate the presence of the 9-bp deletion or 4-bp insertion in region V, respectively. Haplotype 46, originally designated within group O, falls within group A in this tree due to the presence of the HpaII/HhaI site losses at np 9052/9053.

morphism has been associated with some of the earliest Asian mtDNAs (BLANC *et al.* 1983). This haplotype group is most frequent in the Vietnamese, (32.1%) and the Malay Aborigines (28.1%). In light of their language affiliation, [Austro-Asiatic family (BELLWOOD, 1979)] these populations seem to be derived from a common stock. MtDNAs from haplotype group A were also found in the Taiwanese Han (10.0%), Malays (14.3%), Koreans (15.4%), and Sabah Aborigines (3.1%) substantiating the early appearance of this haplotype group.

The aboriginal populations of the Malay peninsula (Senoi and Proto-Malays) and Borneo (Sabah) show a degree of genetic substructure. The majority of Malay Aborigine samples were taken from the Senoi, a group of tribes living in the mountainous jungles of peninsular Malaysia. This group is believed to have arrived with a "second wave" of migration occurring about 4,000–8,000 YBP (BELLWOOD 1985). It has been postulated that the islands of Borneo and Indonesia received the spillover from this migration in Southeast Asia and the Malay peninsula (TAN *et al.* 1979). The

TABLE 2

COII/tRNA^{Lys} 9-bp deletion frequencies

Population	n_d	N	%
Malaysian Chinese	1	14	7.1
Malays	2	14	14.3
Malay Aborigines	1	32	3.1
Sabah Aborigines	6	32	18.75
Taiwanese Han	8	20	40.0
Vietnamese	5	28	17.9
Koreans	2	13	15.4
Coastal PNG	23	55	41.8 (STONEKING *et al.* 1990)
Coastal PNG	4	28	14.2 (HERTZBERG *et al.* 1989)
PNG Highlanders	0	64	0.0 (STONEKING *et al.* 1990)
PNG Highlanders	0	30	0.0 (HERTZBERG *et al.* 1989)
Aust. Aborigines	1	31	3.2 (HERTZBERG *et al.* 1989)
Aust. Aborigines	0	20	0.0 (CANN, STONEKING and WILSON 1987)
Japanese	19	116	16.4 (HORAI and MATSUNAGA 1986)
East Asians*	6	34	17.6 (CANN, STONEKING and WILSON 1987)
Polynesians	139	150	92.7 (HERTZBERG *et al.* 1989)
Fijians	23	28	82.1 (HERTZBERG *et al.* 1989)
Amerindians			
Pima	14	31	45.2 (SCHURR *et al.* 1990)
Maya	8	37	21.6 (SCHURR *et al.* 1990)
Ticuna	0	31	0.0 (SCHURR *et al.* 1990)

Malaysian Chinese, Malays, Malay Aborigines, Sabah Aborigines, Taiwanese Han, Vietnamese, and Korean populations represent the groups from this report. n_d = no. of deleted mtDNAs; N = total sample size; % = percentage of deleted mtDNAs within a population. * East Asians = 1 Japanese, 1 Taiwanese, 1 Vietnamese, 2 Philippino and 1 Tongan.

TABLE 3

Percent sequence divergence

	MC	MM	MA	SA	TW	VN	KN
MC	**0.196**	**0.229**	**0.207**	**0.241**	**0.193**	**0.255**	**0.219**
MM	*0.04*	**0.182**	**0.188**	**0.200**	**0.193**	**0.236**	**0.205**
MA	*0.035*	*0.023*	**0.148**	**0.196**	**0.177**	**0.211**	**0.194**
SA	*0.053*	*0.019*	*0.032*	**0.180**	**0.195**	**0.254**	**0.220**
TW	*0.022*	*0.029*	*0.031*	*0.032*	**0.145**	**0.215**	**0.189**
VN	*0.039*	*0.027*	*0.019*	*0.046*	*0.024*	**0.236**	**0.243**
KN	*0.028*	*0.021*	*0.028*	*0.037*	*0.024*	*0.032*	**0.185**

Intrapopulational divergences are along the diagonal (underlined), and interpopulational divergence and interpopulational divergences corrected for intrapopulational variation (NEI and TAJIMA 1983) are above (bold type) and below (italics) the diagonal, respectively.

Semai tribe of the Senoi had the unique group I haplotypes (73, 75, 78, 81, and 82) defined by an *Alu*I np 10143 site gain and a *Hae*II/*Hha*I nps 9326/9329 site gain. The Jeni (MA10, MA21) had haplotype 76 (group J), with several previously unreported polymorphisms. The populations of the Malay peninsula also showed close affinities to the Sabah Aborigines, sharing haplotype groups A, D, E and G.

The Kadazan represent the largest ethnic group in Sabah and are thought to have originated from an Austronesian migration originating in South China (TAN *et al.* 1979). The other ethnic groups from Sabah (Berungei, Murut, Rungus) and the Northern Borneo Bisaya, had mtDNAs similar to the Kadazan. For the Bisaya (106, 122), this was unexpected since their mtDNAs did not resemble previously reported Philippino haplotypes (CANN 1982; CANN, STONEKING and WILSON 1987). This probably reflects the partial assimilation of these minority groups into the Kadazan.

Papua New Guinea *vs.* Southeast Asia: There are several similarities between the haplotypes of PNG and those of the Malay Aborigines, Malays and Sabah Aborigines. Based on shared haplotype character states, the Southeast Asians appear closest to the coastal PNG populations. The combined *Hinc*II/*Hpa*I site gain at np 207 observed in coastal and highland PNG was found in our haplotype 90 (MM9), which is virtually identical to the PNG type 94 (group S, Figures 2 and 3) (STONEKING *et al.* 1990). Additionally, the deletion haplotype found in Southeast Asia (haplotype 54, Figures 2 and 3) is also found in coastal New Guinea [type P119 (STONEKING *et al.* 1990)]. All other PNG deletion haplotypes fall within the Southeast Asian haplotype group D*. The group A haplotypes frequently observed in the Malay Aborigines and Vietnamese are also found on the southern coast of New Guinea in type P150 (STONEKING *et al.* 1990). The presence of these Southeast Asian mtDNAs in coastal PNG is consistent with a postulated Southeast Asian origin of these populations (BELLWOOD 1987; STONEKING *et al.* 1990). Additional Sabah and Malay peninsula haplotypes shared with coastal PNG include haplotype 83 (haplotype group G) which is essentially

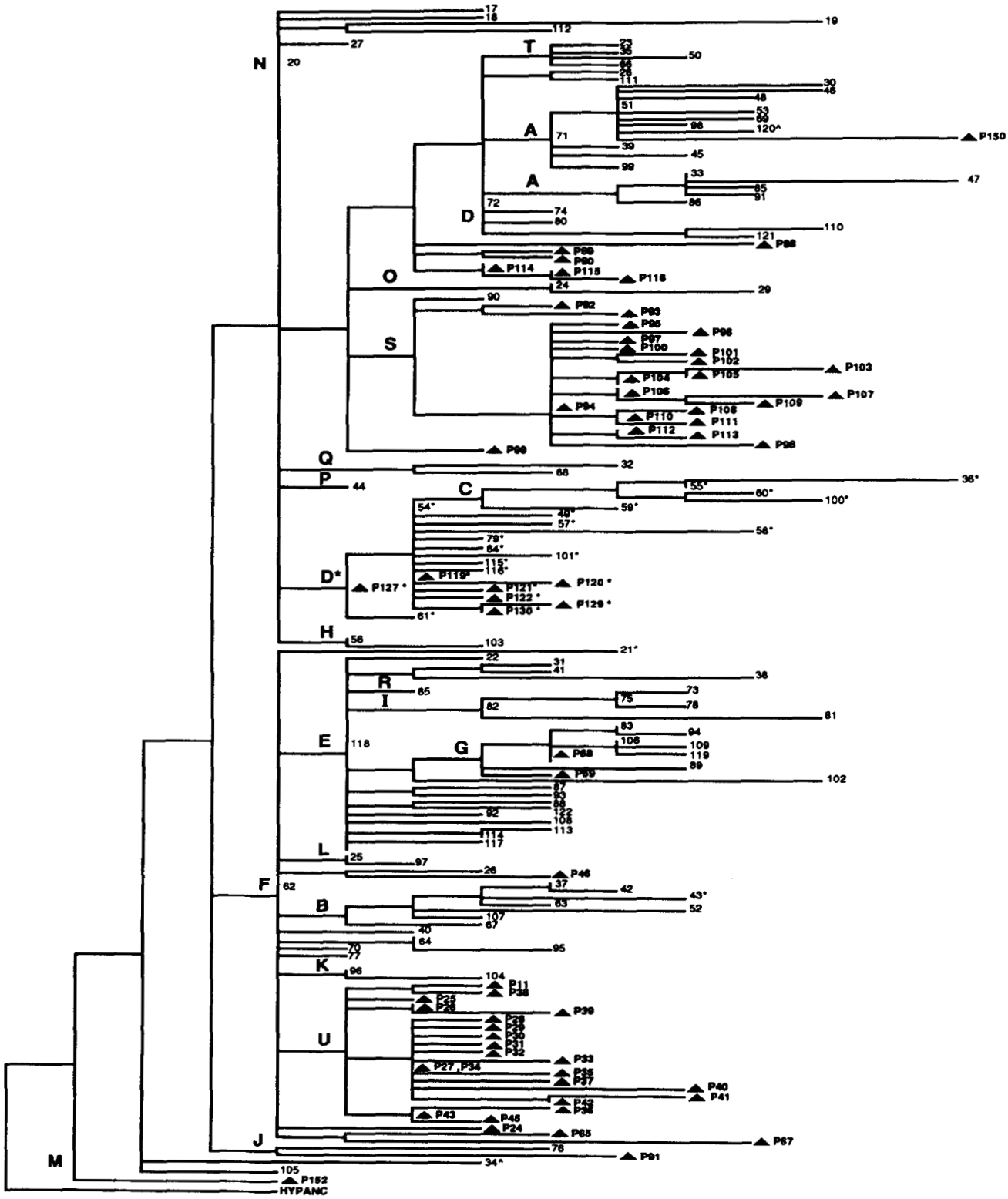

FIGURE 3.—A phylogeny of Southeast Asian and Papua New Guinea (PNG) mtDNA haplotypes. The rooting and branching algorithm for this tree are identical to that of Figure 2, and it has a length of 351 steps. PNG haplotypes are indicated by solid upright triangles, with the numbering of types matching the nomenclature of Stoneking *et al.* (1990). A new haplotype group (U) is formed by the PNG highland mtDNAs (types 11, 25–45), but otherwise the original branching structure is preserved. PNG types with asterisks signify mtDNAs with the 9-bp intergenic deletion.

type P68 of the southern coast of New Guinea and related group G haplotypes 89, 94, 106, 109, and 119 which share the *Hha*I site loss at np 7598 and are similar to PNG types P67–P69 (STONEKING *et al.* 1990). Thus, it appears that both the Malay peninsular and Borneo (Sabah) populations retain remnants of the Austronesian migration that expanded into the Pacific Basin and coastal PNG.

Certain Southeast Asian mtDNAs also show affinities with those of the PNG highlands. Some Sabah Aboriginal mtDNAs share the combined *Dde*I site gain at np 8569 and *Hae*III np 8572 site loss with PNG highlanders. The Malay peninsular populations share a *Dde*I site loss at np 1715 with the PNG highlands, and Vietnamese haplotype 40 shares site gains at nps 15882 (*Ava*II), 10394 (*Dde*I) and 10397 (*Alu*I) with PNG type P37. However, these populations often lack other mutations which may be associated with site losses or gains, making clear associations between Borneo (Sabah), the Malay peninsula, and highland PNG difficult.

***Alu*I/*Dde*I np 10397/10394 sites:** The overlapping *Alu*I and *Dde*I sites at nps 10397 and 10394 appear to be ancient mutations. This pair of sites was prevalent in every Southeast Asian population and divided each of them into two major groups (Figures 2 and 3). The *Dde*I site has been found in mtDNAs from every racial group (CANN, STONEKING and WILSON 1987; BROWN *et al.* 1992), and is present in the most divergent African haplotypes reported (CANN, STONEKING and WILSON 1987), indicating its antiquity. The *Alu*I site has not been previously reported, but correlates highly with a reported *Alu*I site at np 1403 which is consistently associated with the *Dde*I site at np 10394 (CANN, STONEKING and WILSON 1987; STONEKING *et al.* 1990). It seems likely that the putative *Alu*I site at np 1403 was previously misplaced, and is in fact at np 10397. If this is the case, these sites also subdivide the PNG mtDNAs into two major groups (STONEKING *et al.* 1990) indicative of a common Mongoloid origin.

COII/tRNA^{Lys} length polymorphism: The COII/tRNA^{Lys} deletion appears to have originated in a mtDNA similar to haplotype 54, probably in central China. As migrating populations radiated out from this region successive founder events then resulted in the increased frequency of the deletion haplotypes in some populations. Today the deletion is distributed among Pacific coastal or island populations (HORAI and MATSUNAGA 1986; CANN, STONEKING and WILSON 1987; HERTZBERG *et al.* 1989; STONEKING *et al.*1990) as well as the Amerindians (SCHURR *et al.* 1990; TORRONI *et al.* 1992) (Table 2).

The deletion appears to have been associated with at least two major migrations. One migration moved south along the Asian coastline, eastward into Indonesia, and out into the Pacific islands (HERTZBERG *et al.* 1989). The other migration went north into Siberia and eventually crossed the Bering land bridge into the New World, yielding the Amerindians (SCHURR *et al.* 1990).

Subsequent nucleotide substitutions occurring in ancestral mtDNA haplotypes have resulted in two distinct clades creating haplotype groups C and D* (Figure 2). The *Dde*I site at np 10394 is present in haplotype group C but absent in D*. In addition, deletion haplotypes within group C appear more divergent (0.089%) than those of haplotype group D* (0.067%). Consequently, the deletion haplotypes within group C appear to be older than that of haplotype group D*. Haplotype group D* consists of haplotype 54 and closely associated haplotypes. All the deletion haplotypes seen in aborigines from PNG (STONEKING *et al.* 1990) and Amerindians (SCHURR *et al.* 1990) fall within haplotype group D*. Thus, it would appear that the recent migrants from Asia that carried the COII/tRNA^{Lys} deletion belonged to haplotype group D*.

While groups C and D* are probably derived from a single COII/tRNA^{Lys} deletion event, deletions associated with two other haplotypes probably are the result of independent events. Haplotypes 43 in a Vietnamese and 21 in a Malaysian Chinese do not fit into either group C or D* (Figure 2). Both differ by at least 6 mutational events from the next closest deletion haplotype, 54. Haplotype 21 has very little similarity to any other haplotype in our study. In contrast, 43 shares all the same restriction sites with haplotype 37 (haplotype group B) which does not have the deletion. To directly verify the presence of the deletion in haplotype 43, we sequenced the mtDNA through region V and found the reported deletion.

The marked differences in haplotypes 43 and 21 relative to each other and groups C and D* can best be explained by parallel independent deletions. To further test this possibility, we tried to weight the deletion in phylogenetic analyses and thereby force the tree to assume a structure in which the deletion could arise only once. This resulted in increasing the tree length by 12 additional steps (data not shown). Consequently, it appears that there must have been three deletion events, one in a haplotype similar to 54 giving the major deletion clades, one in a haplotype similar to 37 to create haplotype 43, and one in an undefined haplotype to yield haplotype 21. Deletion haplotypes 58 and 61 may also have occurred independently, or alternatively, 58 may be a derivative of 54, and haplotype 61 a derivative of 54 via parallel site losses at np 12406 (*Hinc*II/*Hpa*I) and np 16517 (*Hae*III).

The two insertion mutations may also have had independent origins since they are associated with

very different haplotypes, 34 and 120. Our haplotype 120 and CANN's insertion type 92 (CANN 1982; CANN, STONEKING and WILSON 1987) are similar to haplotype group A. However, haplotype 34 is quite distinct, retaining three phylogenetically important sites (*Hinc*II and *Hpa*I np 12406, *Hae*II np 9052, and *Ava*II np 8249) and five other point mutations creating unique restriction sites.

In summary, all Southeast Asian populations analyzed in this study appear to have common origins, consistent with a hypothesized southern Mongoloid origin of the peoples in this region (BELLWOOD 1985, and references therein; TURNER 1987). These mtDNAs are divided into two major branches by the *Alu*I/*Dde*I nps 10397/10394 polymorphisms. The populations from the Malay peninsula and Borneo (Sabah) appear to have genetic ties to those of coastal PNG. The high sequence diversity of the Vietnamese and the high frequency of the *Hinc*II/*Hpa*I morph 1 haplotypes suggest that Southern China is the center of Asian mtDNA radiation (BLANC *et al.* 1983) and, it appears that the deletion and insertion mutations have occurred multiple times in Asian mtDNA lineages. The high frequencies of the deletion haplotype group D* mtDNAs in Southeast Asia, the Pacific islands, and the New World implies that the migrants carrying this marker were descendant from a single founder population.

We are grateful to A. R. A. GREGORY, LILY LAWRENCE, P. S. DAVID, J. JOHNSON (Blood Bank, General Hospital, Seramban, Malaysia), MUHAMMED AWANG (Universiti Pertanian Malaysia) for their efforts in collecting Malaysian bloods and DARREN KASTIN and CHI-CHUAN YANG for assistance with the statistical analysis. The assistance of the Clinical Research Center of Emory University School of Medicine (supported by National Institutes of Health grant M01-RR-00039) is greatfully acknowledged. This work was supported by a National Science Foundation grant BNS8718775 awarded to D.C.W. and a Fullbright award to Y.Y.G.

LITERATURE CITED

ANDERSON, S., A. T. BANKIER, B. G. BARRELL, M. H. L. DE BRUIJN, A. R. COULSON, J. DROUIN, I. C. EPERON, D. P. NIERLICH, B. A. ROE, F. SANGER, P. SCHREIER, A. J. H. SMITH, R. STADEN and I. G. YOUNG, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457–465.

BELLWOOD, P., 1979 *Man's Conquest of the Pacific: The Prehistory of Southeast Asia and Oceania.* Oxford University Press.

BELLWOOD, P., 1985 *Prehistory of the Indo-Malaysian Archipelango.* Academic Press, Sydney.

BELLWOOD, P., 1987 *The Polynesians, Prehistory of an Island People.* Thames & Hudson, London.

BLANC, H., K. H. CHEN, M. D'AMORE and D. C. WALLACE, 1983 Amino acid change associated with the major polymorphic *Hinc*II site of Oriental and Caucasian mitochondrial DNAs. Am. J. Hum. Genet. **235:** 167–176.

BROWN, M. D., A. S. VOLJAVEC, M. T. LOTT, A. TORRONI, C.-C. YANG and D. C. WALLACE, 1992 Mitochondrial DNA complex I and III mutations associated with Leber's hereditary optic neuropathy. Genetics **130:** 163–173.

BROWN, W. M., M. GEORGE, JR. and A. C. WILSON, 1979 Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76:** 1967–1971.

CANN, R. L., 1982 The evolution of human mitochondrial DNA. Ph.D. thesis, University of California, Berkeley.

CANN, R. L., W. M. BROWN and A. C. WILSON, 1984 Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. Genetics **106:** 479–499.

CANN, R. L., M. STONEKING and A. C. WILSON, 1987 Mitochondrial DNA and human evolution. Nature **325:** 31–36.

CANN, R. L., and A. C. WILSON, 1983 Length mutations in human mitochondrial DNA. Genetics **104:** 699–711.

CASE, J. T., and D. C. WALLACE, 1981 Maternal inheritance of mitochondrial DNA polymorphisms in cultured human fibroblasts. Somatic Cell Genet. **7:** 103–108.

DENARO, M., H. BLANC, M. J. JOHNSON,-K. H.-CHEN, E. WILMSEN, L. L. CAVALLI-SFORZA and D. C. WALLACE, 1981 Ethnic variation in *Hpa*I endonuclease cleavage patterns of human mitochondrial DNA. Proc. Natl. Acad. Sci. USA **78:** 5768–5772.

GILES, R. E., H. BLANC, H. M. CANN and D. C. WALLACE, 1980 Maternal inheritance of human mitochondrial DNA. Proc. Natl. Acad. Sci. USA **77:** 6715–6719.

HARIHARA, S., N. SAITOU, M. HIRAI, T. GOJOBORI, K. S. PARK, S. MISAWA, S. B. ELLEPOLA, T. ISHIDA and K. OMOTO, 1988 Mitochondrial DNA polymorphism among five Asian populations. Am. J. Hum. Genet. **43:** 134–143.

HERTZBERG, M., K. N. P. MICKELSON, S. W. SERJEANTSON, J. F. PRIOR and J. TRENT, 1989 An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. Am. J. Hum. Genet. **44:** 504–510.

HORAI, S., T. GOJOBORI and E. MATSUNAGA, 1984 Mitochondrial DNA polymorphism in Japanese. Hum. Genet. **68:** 324–332.

HORAI, S., AND E. MATSUNAGA, 1986 Mitochondrial DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. Hum. Genet. **72:** 105–177.

ISKANDAR, C., 1976 *Orang Asli: The Aboriginal Tribes of Peninsular Malaysia.* Oxford University Press.

MIYATA, T., H. HAYASHIDA, R. KIKUNO, M. HASEGAWA, M. KOBAYASHI and K. KOIKE, 1982 Molecular clock of silent substitution: at least a six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J. Mol. Evol. **19:** 28–35.

NECKELMANN, N., K. LI, R. H. WADE, R. SHUSTER and D. C. WALLACE, 1987 cDNA sequence of a human skeletal muscle ADP/ATP translocator: lack of a leader peptide, divergence from a fibroblast translocator cDNA and coevolution with mitochondrial DNA genes. Proc. Natl. Acad. Sci. USA **84:** 7480–7584.

NECKELMANN, N., C. K. WARNER, A. CHUNG, J. KUDOH, S. MINOSHIMA, R. FUKUYAMA, M. MEAKAWA, Y. SHIMIZU, N. SHIMIZU, J. D. LIU and D. C. WALLACE, 1989 The human ATP synthase β subunit gene: sequence analysis, chromosome assignment, and differential expression. Genomics **5:** 829–843.

NEI, M., and F. TAJIMA, 1983 Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics **105:** 207–217.

SAIKI, R. K., S. SCHARF, F. FALOONA, K. B. MULLIS, G. T. HORN, H. A. ERLICH and N. ARNHEIM, 1985 Enzymatic amplification of β-globin genomic sequences and restriction site analyses for diagnoses of sickle cell anemia. Science **230:** 1350–1354.

SCHURR, T. G., S. W. BALLINGER, Y. Y. GAN, J. A. HODGE, D. A.

MERRIWETHER, D. N. LAWRENCE, W. C. KNOWLER, K. M. WEISS and D. C. WALLACE, 1990 Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am. J. Hum. Genet. **46:** 613–623.

STONEKING, M., K. BHATIA and A. C. WILSON, 1986 Mitochondrial DNA variation in eastern highlands of New Guinea, pp. 87–100 in *Genetic Variation and Its Maintenance in Tropical Populations*, edited by D. F. ROBERTS and G. DE STEFANO. Cambridge University Press, Cambridge.

STONEKING, M., L. B. JORDE, K. BHATIA and A. C. WILSON, 1990 Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics **124:** 717–733.

SWOFFORD, D. L., 1990 *PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0.* Computer Program distributed by the Illinois Natural History Survey, Champaign, Ill.

TAN, S. G., Y. S. TENG, J. GANESAN, K. Y. LAU and L. E. LIE-INJO, 1979 Biochemical genetic markers in the Kadazans of Sabah, Malaysia. Hum. Genet. **49:** 349–353.

TORRONI, A., T. G. SCHURR, C.-C. YANG, E. J. E. SZATHMARY, R. C. WILLIAMS, M. S. SCHANFIELD, G. A. TROUP, W. C. KNOWLER, D. N. LAWRENCE, K. M. WEISS and D. C. WALLACE, 1992 Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics **130:** 153–162.

TURNER, C. G., II, 1983 Dental evidence for the peopling of the Americas, pp 147–157 in *Early Man in the New World*, edited by R. SHUTLER, JR. Sage, Beverly Hills, Calif.

TURNER, C. G., II, 1987 Late Pleistocene and Holocene population history of East Asia based on dental variation. Am. J. Phys. Anthropol. **73:** 305–321.

WALLACE, D. C., K. GARRISON and W. C. KNOWLER, 1985 Dramatic founder effects in Amerindian mitochondrial DNAs. Am. J. Phys. Anthropol. **68:** 149–155.

WALLACE, D. C., J. YE, S. N. NECKELMANN, G. SINGH, K. A. WEBSTER and B. D. GREENBERG, 1987 Sequence analysis of cDNAs for the human and bovine ATP synthase β subunit: mitochondrial genes sustain seventeen times more mutations. Curr. Genet. **12:** 81–90.

WALLACE, D. C., G. SINGH, M. T. LOTT, J. A. HODGE, T. G. SCHURR, A. M. S. LEZZA, L. J. ELSAS and E. K. NIKOSKELAINEN, 1988 Mitochondrial DNA mutations associated with Leber's hereditary optic neuropathy. Science **242:** 1427–1430.

WRISCHNIK, L. A., R. G. HIGUCHI, M. STONEKING, H. A. ERLICH, N. ARNHEIM and A. C. WILSON, 1987 Length mutations in human mitochondrial DNA; direct sequencing of enzymatically amplified DNA. Nucleic Acids Res. **15:** 529–542.

ZHAO, T., and T. LEE, 1989 Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. Hum. Genet. **83:** 101–110.

Communicating editor: V. G. FINNERTY

APPENDIX A

The oligonucleotide primers for PCR amplication of Southeast Asian mtDNAs are given in Table 4.

TABLE 4

Oligonucleotide primers for PCR amplifications of Southeast Asian mtDNAs

5′ → 3′ coordinates (forward, reverse)	T_H
1562-1581, 3717-3701	51
3007-3023, 5917-5898	55
5317-5333, 7608-7588	57
7392-7410, 8921-8902	57
8282-8305, 10107-10088	57
9911-9932, 11873-11851	69
11673-11691, 13950-13932	57
13914-13930, 16547-16527	47
16453-16472, 1696-1677	61

Primer pair coordinates are positioned according to ANDERSON *et al.* (1981). The coordinates before the comma correspond to the forward primer and those after the comma to the reverse primer. The T_H used for annealing was the lowest for the primer pair as calculated from the nucleotide sequence of each primer, $T_H = 4(C + G) + 2(T + A) - 5°$.

APPENDIX B

Figure 4 presents the polymorphic restriction sites observed in Southeast Asian mtDNA haplotypes. mtDNA haplotypes are numbered according to Table 1. A "1" indicates the presence of a site and a "0" indicates the absence of a site except for region V where "1" indicates a single copy of the 9-bp repeat, "2" indicates two copies of the repeat, and "3" indicates the presence of the 4-bp insertion. Sites are numbered from the first nucleotide of the recognition sequence according to the published sequence (ANDERSON *et al.* 1981); bold face numbers indicate site gains relative to the published sequence and non-bold face numbers indicate site losses. The 18 restriction enzymes are designated by the following single-letter code: *Alu*I, a; *Ava*II, b; *Dde*I, c; *Hae*III, e; *Hha*I, f; *Hin*f I, g; *Hpa*I, h; *Hpa*II, i; *Mbo*I, j; *Rsa*I, k; *Taq*I, l; *Bam*HI, m; *Hae*II, n; *Hin*cII, o; *Pst*I, p; *Pvu*II, q; *Xba*I, r; *Xho*I, s (after CANN and WILSON 1984; CANN, STONEKING and WILSON 1987). Sites separated by a diagonal line indicate either simultaneous site gains or site loss for two different enzymes or a site gain for one enzyme and a site loss for another enzyme because of a single inferred nucleotide substitution; these sites are considered to be only one restriction site polymorphism in the analysis. Sites marked with an asterisk were found to be present or absent in all samples (except where polymorphic) contrary to the published sequence and were confirmed by sequencing (WALLACE *et al.* 1988; J. BROWN *et al. 1992*).

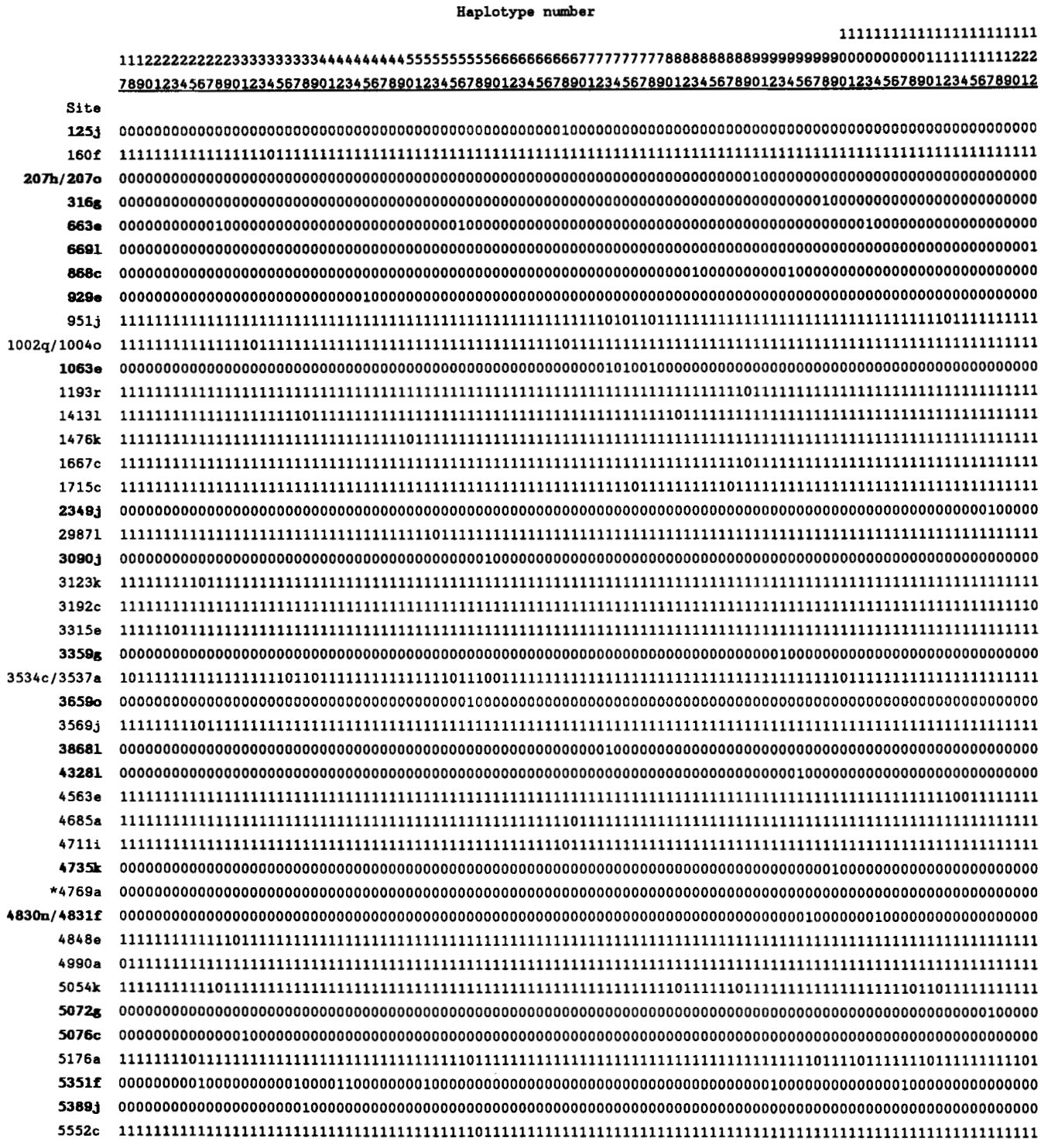

FIGURE 4.—Polymorphic restriction sites observed in Southeast Asian mtDNA haplotypes.

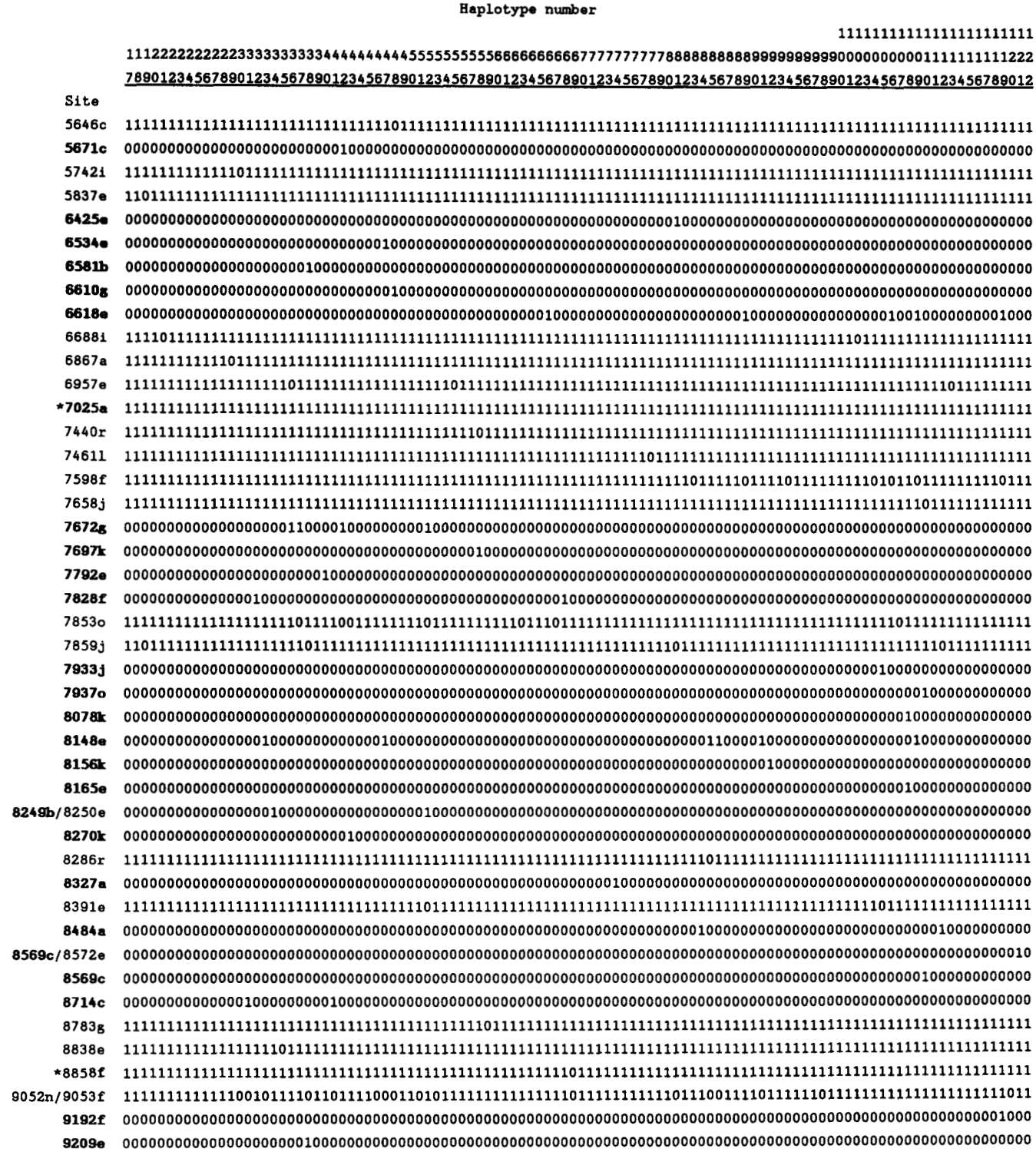

FIGURE 4.—*Continued*

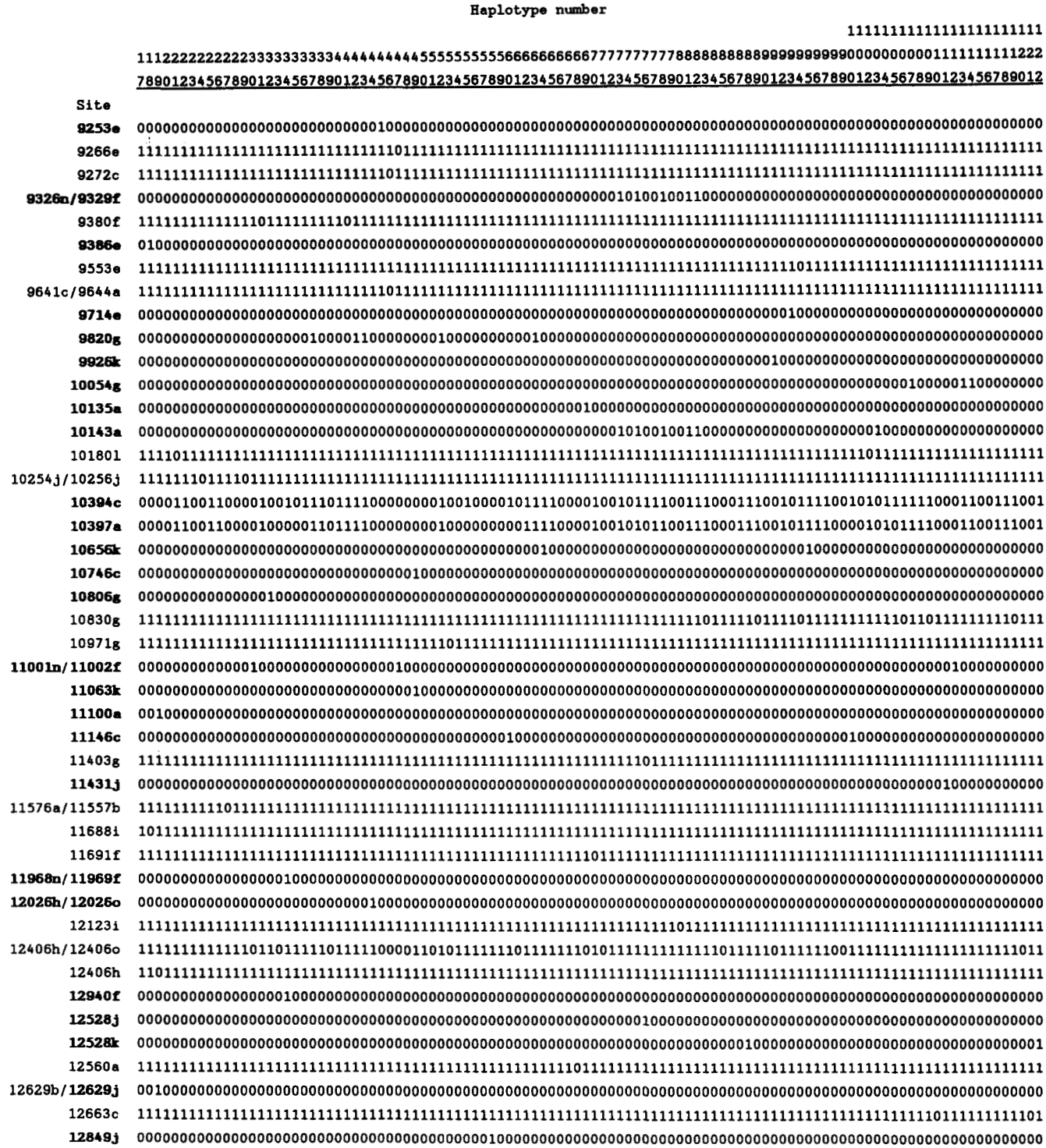

FIGURE 4.—*Continued*

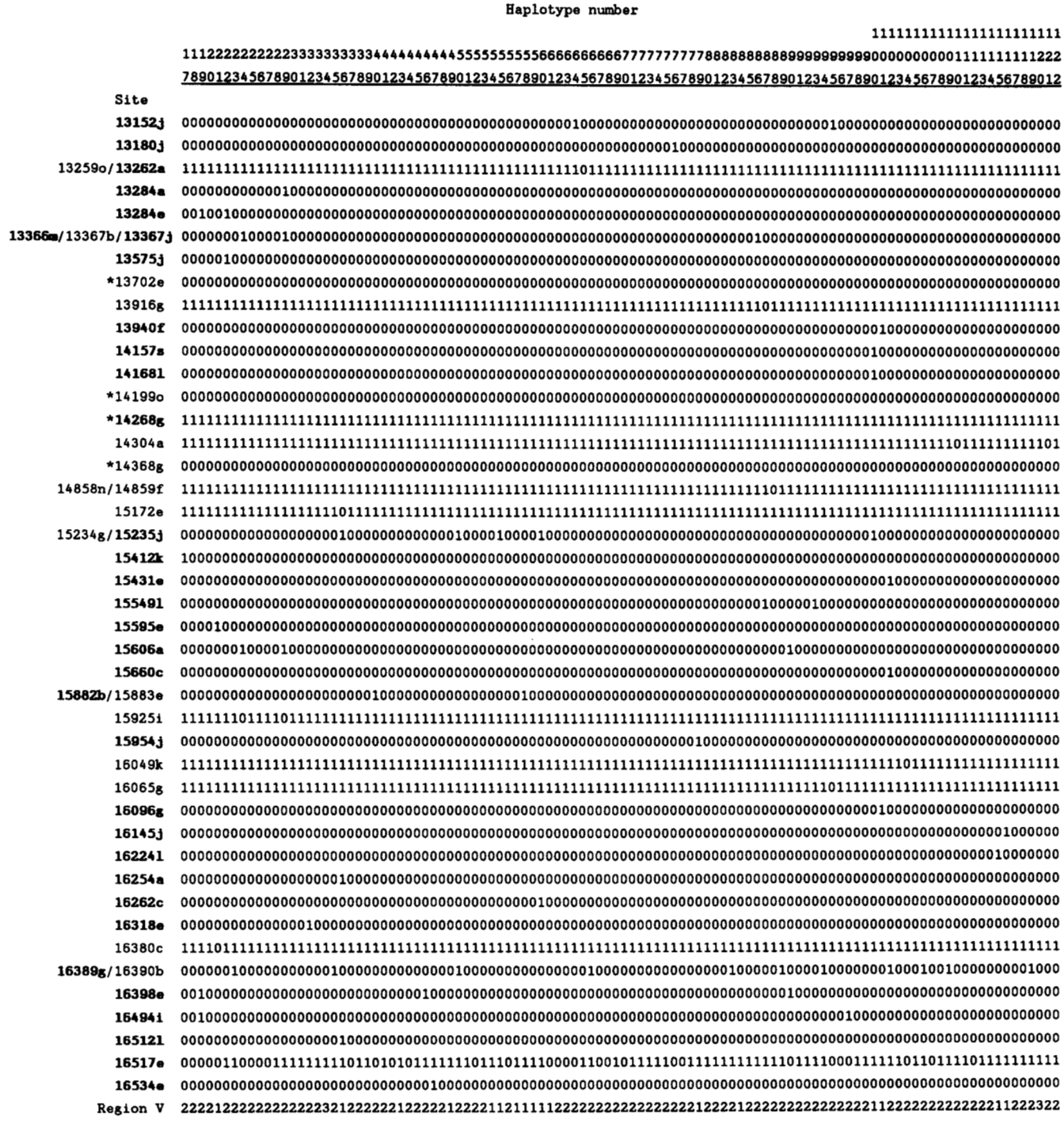

FIGURE 4.—*Continued*