Table 2. *Continued*

Group / Sample	JPN 1312	RYU 50	AIN 51	Ch1 213	Ch2 435	Ch3 32	Ch4 72	Ch5 757	Ch6 67	CA1 204	CA2 93	TWA 208	MAN 98	ITE 46	FIU 38	ALU 56	KAM 91	CHU 60	TUV 36	BUR 40	KOR 537	TIB 65	SAK 20	FIL 32	IND 40	SAB 34
F2	–	–	–	0.47	0.46	3.13	6.94	0.4	–	–	–	–	–	–	5.26	–	–	–	–	–	–	–	–	–	–	–
F2a	0.15	–	–	0.47	0.92	–	1.39	0.79	7.46	–	–	–	–	–	–	–	–	–	–	–	0.19	–	–	–	–	–
F2a1	0.08	–	–	–	–	–	–	0.66	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	2.5	–
F3	0.08	–	–	–	0.46	–	–	0.53	–	–	–	–	–	–	–	–	–	–	–	–	0.19	–	–	–	–	–
F4	–	–	–	0.47	–	–	–	0.92	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	3.13	2.5	14.7
F4b	–	–	–	–	–	–	–	0.13	–	–	–	10.1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
M	0.38	2	3.92	–	–	–	–	0.26	–	0.49	–	–	–	–	–	–	–	–	–	–	0.19	–	–	–	–	–
M (PNG)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	10	–
M5/D4a/G1	0.46	–	13.7	0.94	0.23	18.8	–	0.26	–	0.49	2.15	–	–	–	–	–	–	3.33	–	2.5	1.68	–	5	–	2.5	2.94
M7a	7.39	12	15.7	–	0.23	–	–	0.53	4.48	–	–	–	–	–	–	–	–	–	–	2.5	3.35	–	–	6.25	–	–
M7a1	0.08	14	–	1.41	1.38	–	1.39	2.25	–	0.98	–	0.48	–	–	–	–	–	–	–	–	–	–	–	3.13	–	–
M7b	–	–	–	0.47	0.46	–	1.39	2.11	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
M7b1	0.08	–	–	2.35	2.99	–	–	5.02	–	0.49	–	12	–	–	–	–	–	–	–	–	0.56	–	–	–	2.5	–
M7b2	4.73	8	3.92	0.47	0.69	–	1.39	0.13	–	–	–	–	1.02	–	–	–	–	–	–	–	2.79	–	–	–	–	–
M7c	0.76	2	–	1.88	2.07	–	2.78	2.51	–	1.96	–	4.33	–	–	–	–	–	–	–	2.5	3.72	3.08	–	18.8	2.5	20.6
M8	0.15	–	–	0.47	–	–	–	0.13	–	–	–	–	–	–	–	–	–	1.67	–	–	0.37	–	–	–	–	–
M8a	1.22	–	–	4.23	0.92	–	4.17	1.59	–	1.47	–	–	–	–	–	–	–	–	–	2.5	1.68	–	–	–	–	–
M8a2	–	–	–	–	0.23	–	–	–	–	–	1.08	–	–	13	–	–	20.9	1.67	2.78	–	0.37	–	–	–	–	–
C	–	–	–	1.41	6.67	–	2.78	4.1	7.46	3.92	9.68	–	12.2	–	–	16.1	26.4	5	16.7	10	0.37	1.54	–	–	–	–
C1	0.3	–	–	–	0.46	–	–	–	–	–	1.08	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
C4a	0.08	–	–	0.47	0.92	–	–	–	–	0.98	1.08	–	4.08	–	–	–	–	–	13.9	–	0.37	–	–	–	–	–
C5	0.08	–	–	–	0.23	–	–	–	–	1.96	2.15	–	1.02	–	–	–	–	–	2.78	–	0.74	–	–	–	–	–
Z	1.3	–	–	2.35	2.76	–	4.17	0.53	10.4	0.49	9.68	–	–	6.52	–	–	8.79	–	–	–	1.49	9.23	–	–	–	–
G	0.53	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.37	–	–	–	–	–
G1a1/D	2.13	–	5.88	–	0.23	–	–	0.4	–	–	–	–	–	–	–	–	–	–	–	–	2.05	–	10	–	–	–
G2a	1.68	4	–	1.88	–	–	–	0.26	–	0.49	–	–	–	–	–	–	–	–	2.78	5	1.86	–	–	–	–	–
G2a1	2.52	–	3.92	5.16	0.92	–	1.39	0.66	–	7.35	1.08	–	6.12	–	–	–	–	–	5.56	–	1.49	–	–	–	–	–
G5	–	–	–	–	–	–	–	–	–	–	–	–	–	69.6	–	67.9	27.5	6.67	–	–	–	–	–	–	–	–
M9	–	–	–	–	–	–	2.78	1.59	2.99	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
M9a	2.44	4	1.96	5.63	1.61	–	1.39	1.98	–	2.45	1.08	–	–	–	–	–	–	–	–	5	3.17	10.8	–	–	–	–
M10	1.3	–	–	2.35	1.84	–	–	1.72	–	0.49	1.08	–	–	–	–	–	–	–	–	2.5	4.66	7.69	–	–	–	2.94
M12	0.08	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.19	6.15	–	–	–	–
D	0.08	–	–	0.47	0.46	–	1.39	0.26	–	–	–	–	2.04	–	–	–	–	–	–	–	0.19	–	–	–	–	–
D4	18.9	2	7.84	12.7	20.9	6.25	1.39	6.74	16.4	11.3	18.3	13.5	3.06	–	18.4	–	–	1.67	13.9	17.5	18.4	12.3	–	18.8	7.5	11.8
D4a	7.39	–	1.96	4.23	–	–	1.39	0.92	–	2.45	1.08	0.48	–	–	–	–	–	11.7	–	–	5.77	–	80	–	–	–
D4a1	0.53	4	–	0.94	1.38	–	2.78	1.06	–	–	–	–	–	–	–	–	–	–	–	–	2.42	1.54	–	–	–	–
D4b	2.36	6	–	1.41	1.61	–	1.39	0.92	–	0.98	3.23	–	1.02	–	–	3.57	–	6.67	2.78	2.5	0.93	18.5	–	–	–	–
D4d	2.67	–	–	0.94	0.69	3.13	–	–	–	1.96	1.08	–	–	–	–	–	–	–	2.78	2.5	0.93	–	–	–	–	–
D4k	0.15	–	–	–	1.15	–	1.39	0.13	22.4	–	–	–	–	–	–	–	–	–	–	–	0.19	3.08	–	–	–	–
D4n	0.61	–	3.92	–	0.23	–	1.39	–	2.99	–	–	–	–	–	–	–	–	–	–	–	0.19	–	–	–	2.5	–
D5	3.73	2	–	1.88	3.22	–	4.17	2.64	–	0.98	–	6.25	1.02	–	–	–	–	–	–	2.5	2.98	1.54	–	3.13	2.5	2.94
D5a	1.07	–	3.92	1.41	3.91	–	4.17	1.59	–	0.49	–	–	–	–	–	–	–	–	2.78	–	0.56	1.54	–	–	–	–

transitions. Yao et al. (2002a) assigned mutations 199 and 9824 as basic for M7. However, our phylogenetic tree points to 6455 and 9824 as the basal mutations for this group, whereas 199 is only common to the M7b and M7c subgroups (Fig. 1A), which coincides with the phylogeny proposed by Kivisild et al. (2002). M7 can be RFLP-diagnosed by the lack of the 6451 MboII restriction site. The M7a subgroup can be defined by several coding-region positions (Fig. 1A; Kivisild et al. 2002). The M7b classification remains as proposed in Kivisild et al. (2002); but M7c has, in addition to 146 and 16295, three more coding-region substitutions (4850, 5442, and 12091) in its basal branch (Fig. 1A). At this point, it is worthwhile pointing out that the ambiguously assigned sequence 536 in Herrnstadt et al. (2002) belongs to M7c, as it has the five identifying coding-region mutations distinctive of this subhaplogroup. As for the geographic distribution, M7a1 has its highest frequencies (14%) and diversities (86%) in the Ryukyuans, and it is also very common in the whole of China, with a mean diversity of ~76%. But, curiously, it has not been detected in Koreans or in Ainu, and is rare in mainland Japanese. In a similar way, M7a has its highest diversity in Ryukyuans (83%). Both groups are rather common in the Philippines. Although M7b has its greatest diversity in northern China (75%–62%), its derivative M7b2, has it again in Ryukyuans (100%), Koreans (53%), and mainland Japanese (45%). On the contrary, M7c is absent in Ainu and rare in mainland Japanese but very common in Sabah and the Philippines, although its highest diversity is in the whole of China (76% ± 11%).

Haplogroup M10

This haplogroup has been defined by substitutions 10646 and 16311 (Yao et al. 2002a). In addition, Kong et al. (2003) have found several new mutations in its basal branch that we confirm here (Fig. 1A). Minor modifications are that a new Japanese lineage shares with M10 only the 8793 mutation, and that a new mutation, 13152, seems to be basal for our M10 Japanese lineages. Although its highest frequency is in Tibetans (8%), the largest diversities are found in China. It is present in Koreans and mainland Japanese but has not been detected in either Ainu or Ryukyuans (Table 2).

Haplogroup M11

This haplogroup has been defined by Kong et al. (2003) by seven coding-region mutations (1095, 6531, 7642, 8108, 9950, 11969, and 13074) and four mutations in HVS-II (146, 215, 318, and 326). We confirm the same characterization for our M11 Japanese lineages. A subclade defined by mutation 14340 was found in Chinese (Kong et al. 2003), but it has not been detected in Japanese. In turn, Japanese have a new subclade characterized by mutation 14790. Finally, our data suggest that mutation 15924 is at the root of M11 and the new clade M12.

Haplogroup M12

This haplogroup has been defined in the present study. It harbors a characteristic motif (16145–16188–16189–16223–16381) in its noncoding region and several unique mutations in its coding region (Fig. 1A). Overall, it is a rare haplogroup, being detected only in mainland Japanese, Koreans, and Tibetans, the last-mentioned sample showing its highest frequency (8%) and diversity (50%).

Haplogroup M1

Although not present in eastern Asia, this haplogroup has been included in the phylogenetic tree of macrohaplogroup M to ascertain its hierarchical level with respect to other M clades. It was first detected in Ethiopia (Quintana-Murci et al. 1999) and defined by four transitions in the HVSI region (16129, 16189, 16249, and 16311). After this, M1 was also detected in the Mediterranean basin including Jordan (Maca-Meyer et al. 2001). Several mutations in the coding region are distinctive of this haplogroup (Fig. 1A). Its RFLP diagnosis is possible by an MnlI site loss at position 12401.

Subdivisions Within Macrohaplogroup N

Representatives of two major superhaplogroup N migratory branches are present in Japan. Two main clades, that directly sprout from the basal N trunk (A and N9), have a prevailing northern Asia dispersion, whereas the other two (B and F), having a southern radiation focus, belong to the derivative R clade, characterized by the loss of 16223 and 12705 mutations. Although not detected in Japan, to compare their hierarchical levels with those of the Asian branches, we have included the rCRS sequence and a N1b sequence (Kivisild et al. 1999) as representatives of the western Eurasian R and N clades, respectively.

Haplogroup A

This haplogroup was defined by an HaeIII site gain at 663 (Torroni et al. 1992). It was subdivided on the basis of HVSI motifs in A1 (16223–16290–16319) and A2 (16111–16223–16290–16319) by Forster et al. (1996). In our Japanese sample, we have detected several A1 representatives characterized by two substitutions (8563, 11536). Two of these lineages (ON67 and ND218) have been ascribed to the A1a subgroup that is defined by 4655, 11647, and 16187 substitutions. Two additional A1 Japanese clusters (A1b and A1c) have also been phylogenetically defined (Fig. 2). The A2 subgroup is represented in the tree by a Chukchi (6971) and two (KA21 and ON125) Japanese lineages, all sharing the 16362 mutation. As the Chukchi harbors the 16111 and 16265 mutations, it has been labeled as an A2a representative, as tentatively proposed by Saillard et al. (2000), having four additional mutations (152, 153, 8027, and 12007) in its basal branch. Owing to their phylogenetic position, three more Japanese lineages (ND28, TC48, and J42) should be classified as representatives of three new A subhaplogroups, respectively named A3, A4, and A5 (Fig. 2). Geographically, whereas A1 has a wide northern and central Asian distribution, subclade A1a is confined to Korea and mainland Japan. The greatest diversity for A1 is in central Asia (79%). In Japan it is present in both mainland and indigenous populations. Subhaplogroup A2 is mainly present in northeast Siberia including the Kamchatka peninsula, although a lineage has also been detected in Tibet. The main diversity (30%) and frequency (60%) for this subhaplogroup are in the Chukchi.

Subhaplogroups Y, N9a, and N9b

Haplogroup N9 characterized by the 5417 substitution (Yao et al. 2002a) phylogenetically comprises three subhaplogroups. Subhaplogroup N9a was mentioned as another N subcluster with a distinctive HVSI motif (16223, 16257A, 16261) by Richards et al. (2000). It appears named as N9a in Yao et al. (2002a), who added as basal substitutions 150 and 5231. Recently, Kong et al. (2003) added mutations 12358 and 12372 at the basal branch of N9a, which is according to our Japanese phylogeny (Fig. 2). A Japanese N9a1 lineage (TC2) shares mutations 4386, 12007, 16111, and 16129 with the Chinese lineage GD7834 of Kong et al. (2003). Three more N9a Japanese clusters sharing 16172 as their basal mutation have been considered distinct N9a2 branches (Fig. 2). Subhaplogroup Y was first identified by a set of HVSI polymorphisms (16126, 16189, 16231, 16266, 16519), an HaeIII site loss at 8391 and MboI and DdeI site gains at 7933 and 10394, respectively (Schurr et al. 1999). However, according to the classification of Kong et al. (2003), all these mutations define the Y1a1 branch specifically. Our Japanese (Fig. 2) and the Chinese (Kong et al. 2003) phylogenies characterize Y by seven mutations (8392, 10398, 14178, 14693, 16126, and 16231 gains and a 16223 loss).

The branch Y1 would be identified by mutations 3834 and 16266, and the Y1a subcluster by 7933 (Fig. 2; Kong et al. 2003). In Japan we have found a new subclade (Y1b) characterized by four mutations (146, 10097, 15221, 15460). Furthermore, a new branch (Y2) with the same phylogenetic consideration as Y1, and distinguished by six basal mutations must be aggregated to the Y phylogeny (Fig. 2). Finally, we have detected a sister branch of Y in Japan. This new lineage, named N9b, shares two basal mutations (5147 and 16519) with Y and is further characterized by four (10607, 11016, 13183, 14893) additional mutations in its basal branch. All N9b1 representatives seem to have the 16189 mutation, and three branches of this trunk (a, b, and c) have been provisionally defined (Fig. 2). The geographic distribution of subhaplogroup Y is predominantly in Northeast Asia. The highest frequency (22%) is in the Ainu, although only one lineage accounts for this frequency. The greatest diversities are in northern China (80%), and this group is also very diverse in the Nivkhs from northeast Siberia (Torroni et al. 1993a). As for N9a, it has a great diversity in the whole of China (83%) and Korea (79%). In Japan, only mainland Japanese have N9a representatives. Finally, N9b is very scarce, being detected in southern China and Korea. Surprisingly, it is most abundant in the Japanese including the indigenous Ryukyans and Ainu.

Haplogroup F

This haplogroup was first defined as group A by Ballinger et al. (1992), and later renamed as F by Torroni et al. (1994). This group was characterized by the lack of HincII and HpaI sites at 12406. According to the newly proposed nomenclature (Kivisild et al. 2002; Kong et al. 2003), 12406 is now one of the six mutations that specifically define subhaplogroup F1. Recently, haplogroup F has been phylogenetically included as a subcluster of haplogroup R9 (Yao et al. 2002a). Besides F1, two new subgroups (F2 and F3) have been defined by Kong et al. (2003). We have found a new subcluster, named F4 (Fig. 2), that is characterized by three coding-region mutations (5263, 12630, 15670). This group has a particularly high incidence in Southeast Asia (Ballinger et al. 1992), but only subhaplogroup F1b is well represented in the Japanese, including the indigenous Ainu and Ryukyuan. The highest diversities for this subgroup are in eastern China including Taiwan (100%).

Haplogroup B

Renamed as B after Torroni et al. (1992), this haplogroup was identified by the presence of a 9-bp deletion in the COII/tRNA^{Lys} intergenic region of mtDNA. This polymorphism was first detected in Asia by RFLP analysis (Cann and Wilson 1983). It was used to classify Japanese on the basis of the presence/absence of this deletion (Horai and Matsunaga 1986). Even in Asia, the monophyletic status of this cluster has been repeatedly questioned (Ballinger et al. 1992; Yao et al. 2000b); but although the 9-bp deletion has a high recurrence, it seems that together with transition 16189 it defines fairly well a monophyletic cluster, at least in eastern Asia. Recently, a sister clade of B, keeping the 16189 mutation but lacking the 9-bp deletion, has been detected in China, being designated as R11 (Kong et al. 2003). Asian subhaplogroups of B have been named as B4, identified by the 16217 mutation and B5, characterized by 10398 and 16140 mutations (Yao et al. 2002a). It has been deduced from analysis of complete sequences that transitions 709, 8584, and 9950 are also in the basal branch defining B5 (Fig. 2; Kong et al. 2003). Lower-level subdivisions have also been proposed. Three subclades (B4a, B4b, and B4c) were defined within B4 (Kong et al. 2003). At the same phylogenetic level are our Japanese branches named B4d, B4e, and B4f; and several new secondary clusters have also been detected in Japan within B4a, B4b, and B4c (Fig. 2). It is worthwhile to mention that those lineages harboring 16189, 16217, 16247, and 16261, also known as the Polynesian motif (Soodyall et al. 1995), belong to a branch of B4a, having in addition to 16247, 146, 6719, 12239, 14022, and 15746 as basic mutations. The B5 cluster was also subdivided in B5a and B5b on the basis of the HVSI mutations 16266A and 16243, respectively (Yao et al. 2002a), and reinforced with several additional positions after the analysis of complete Chinese (Kong et al. 2003) and Japanese (Fig. 2) sequences. Within B5b, new subdivisions are necessary to accurately classify the Japanese sequences (Fig. 2). Finally, on the basis of characteristic HVSI motifs, we had tentatively defined as B4a3 those lineages with 16189, 16217, 16261, and 16292 transitions. However, the phylogenetic position of a Chinese complete sequence (GD7812) belonging to this HVSI group (Kong et al. 2003) shows that a future redefinition of B4a might be necessary. The geographic distribution of haplogroup B is very complex. As expected from its age, the ancestral motif is widely distributed in Asia excluding Koryacks and other Siberians. The likewise old subhaplogroup B4 has mainly a central–eastern Asian distribution with diversities near 100% from central Asia to Japan. B4a shows a similar distribution as B4, having branches prevalent in Ryukyuans, Lahu of Yunnan, and aborigine Taiwanese (Table 2). In a similar vein, some branches of B4c are more abundant in southern areas (B4c2), whereas others (B4c1) are mainly detected in Korea and Japan, with derivatives in Taiwan (B4c1b). On the other hand, subhaplogroup B5a has its greatest diversity in southern–eastern China (89%), including Taiwan aborigines (67%), but its B5a1 derivative shows the greatest diversity in northern China (71%), being present in mainland Japanese. In turn, subhaplogroup B5b has its major diversity in Korea (83%) and also reached the Philippines (50%). Curiously, the B5b1 derivative shows its highest diversity (67%) and frequency (1%) in mainland Japanese.

Lineage Sorting and Population Pooling

A total of 110 clades with different phylogenetic range have been proposed on the basis of the pool of the eastern Asian complete sequences (Figs. 1A,B and 2). Of these subdivisions (Table 2), 83 have been used to classify all Asian partial sequences analyzed in this study. As a test of accuracy in the sorting of partial sequences into haplogroups, we classified our 672 Japanese complete sequences by using only their HVSI motifs and found that 34 of them (5%) had an ambiguous status or were misclassified. The main sources of errors were those sequences that differed from CRS in only one or two mutations. For instance, the 16223 mutation was found in M and N backgrounds. The 16189, 16223 motif can be D6 or N9b. Within M, sorting into D or G was one of the main sources of ambiguity. Some 16223, 16325, 16362 lineages were D4 and some G1. The motif 16114A, 16223, 16362, classified as D4, was in reality G3. Sometimes further subdivision within a haplogroup is rather difficult; for example, there are 16189, 16223, 16362 representatives in D4 and in D5. Because of recurrency and isolation, it can be expected that this uncertainty level increases with geographic distance. For instance, we have found that several 16129, 16223 Japanese lineages belong to D4, but to infer from this that southern Asian sequences with the same HVSI motif are also D4 would be inappropriate. From a total of 4713 sequences analyzed, 9.2% had an ambiguous status. In spite of this percentage there are enough sequences left to carry out population analysis with statistical confidence.

In a first approach, Japanese, Ainu, and Ryukyuan samples were compared with the rest of Asian samples shown in Table 3 by means of F_{ST}. The closest affinities of mainland Japanese were to three population groups. The first include Korean and Han from Shandong (mean P-value = 0.29 ± 0.06), the second Han from Liaoning and Xinjiang, and the Tu ethnic minority

(0.20 ± 0.06), and the third Han from Xi'an and the Sali, a branch of the Yi ethnic group (0.15 ± 0.06). Ryukyuans and Ainu behave as outliers with significant differences with all the samples. Population groups resulting from the F_{ST} and CLUSTER analysis are defined in Table 3. Although mainland Japanese from Aichi were significantly different from other mainland Japanese because of their high frequency of haplogroup B, they were merged with them as JPN for comparisons with other areas. Control of the conglomerate number expected in CLUSTER analysis allows for a hierarchical grouping of populations. With two conglomerates, the first distinguished isolate was the aboriginal Sakai from Thailand (Fucharoen et al. 2001). This group was unique among other Thai people owing to its lack of lineages with the 9-bp deletion that characterizes haplogroup B, and to the high frequency of the authors' C6 cluster (included in our D4a). The lack of any representative of macrohaplogroup N in a population anthropologically considered one of the oldest groups in Thailand, if not caused by genetic drift, is compatible with the hypothesis that derivatives of macrohaplogroup N had, in southern Asia, a different route from macrohaplogroup M (Maca-Meyer et al. 2001). Also striking is the presence in Sakai of an unequivocal representative (16223–16274–16278–16294–16309) of the sub-Saharan African L2a haplogroup (Torroni et al. 2001), which again is compatible with the physical characteristics of this Negrito group. Although the suggestion that the first spreading out of Africa of modern humans could have carried some L2 lineages in addition to the L3 ancestors (Watson et al. 1997) is a tempting explanation, a recent admixture is more in consonance with the phylogenetic proximity of this lineage to the present African ones. The next outsiders were the majority of the Siberian isolates, which could not be pooled because of big differences in the frequency of distinctive haplogroups (Table 2). This considerable differentiation was already emphasized (Schurr et al. 1999), with strong genetic drift being its most probable cause. Subsequent isolates belong to some Chinese minorities such as those of Lisu and Nu, Lahu, and Taiwanese aborigines. Unexpectedly, other Chinese minorities (Bai, Sali, and Tu) were left in Han Chinese northern clusters. The Bai belong to the Sino-Tibetan Tibeto-Burman ethnic linguistic group and have been strongly influenced by Han. The Sali are a minority within the Yi ethnic group whose most probable ancestors were the Qiang from northwest China. Finally, the Tu, although belonging to the Mongolian branch of the Altaic Family, show their main genetic affinities to the Han from Xi'an ($P = 0.95$), Xinjiang ($P = 0.89$), and Shanghai ($P = 0.79$), all of them clustered in the Ch2 group. On the other hand, Thais, Vietnamese, and Cambodians joined with southern Chinese. As already observed (Chunjie et al. 2000; Yao et al. 2002a), the Han Chinese do not comprise a homogeneous group. With the exception of cluster Ch4, that includes samples from Hubei and Guandong (Table 3), they appear geographically differentiated. The two central Asian groups detected mainly differ in their frequencies for A1b, Z, and G2a. With less than 14 conglomerates, the Japanese, including Ainu and Ryukyuans, were part of a big group formed by Korean, Buryat, Tibetans, and northern Chinese. Ainu was the first differentiated Japanese sample. Ryukyuans separated later, when mainland Japanese and Koreans still comprised a single group. The lack of homogeneity between Ainu and Ryukyuans was pointed out by Horai et al. (1996), who questioned that they shared a recent common ancestor. The main differences between them were attributed to two dominant clusters (C1 and C16, corresponding to our Y and M5/D4a/G1, respectively) present in Ainu but absent in Ryukuyans, and two Ryukyuan dominant clusters (C3 and C13, belonging to our R and M, respectively) absent in Ainu. In addition, applying the present haplogroup nomenclature to the same data, the high frequency of M7a1 and D4a1/D4b in Ryukyuans, but their absence in Ainu, stands out. The MDS plot (Fig. 3A), based on F_{ST} haplogroup frequency distances between final groups (data not shown), only partially reflects the sequential process described above, as only Sakai and Siberians are well differentiated from the rest. On the contrary, relationships obtained from haplotype matches (Fig. 3B) show populations highly structured by geography with the only exceptions being the Ainu and Tuvinian isolates.

The Peopling of Japan

To further know the relative affinities of the Japanese between themselves and with the different Asian groups formed, the data obtained from the global approaches based on haplogroup frequency distances and on sequence match identities are presented in Table 4. Both values are moderately correlated in the comparisons involving the mainland Japanese ($r = -0.479$; two-tail probability 0.012) but not at all in those involving aborigine Ryukyuans ($r = -0.310$; two-tail probability 0.115) and Ainu ($r = 0.087$; two-tail probability 0.667). This result can be explained by assuming that these aboriginal people have suffered important genetic drift effects with substantial changes in haplogroup frequencies and lineage losses or, less probably, that these populations have been isolated long enough to have accumulated new variation. Results based on haplogroup frequencies by far relate mainland Japanese to Koreans followed by northern Chinese. Ryukyuans present the smallest distances to Buryats from South Siberia, followed in short by southern Chinese. In turn, the Ainu have their closest affinities with mainland Japanese, Koreans, and northern Chinese. As regards sequence matches, mainland Japanese also joins first to Koreans and second to Buryats. Aborigine Ryukyuans are closest to Buryats and then to Koreans. Finally, Ainu show comparatively less shared sequences, their greater affinities being toward Chukchi and Koryaks of Kamchatka. This global picture is congruent with an important influence on mainland Japanese from northern Asian populations through Korea, that the Ryukyuans had a dual northern and southern Asian background previous to the new northern influences acquired by admixture with mainland Japanese, and that the Ainu represent the most isolated group in Japan in spite of the genetic input received from Kamchatka. Also noticeable is the great distance and low identity values obtained for the Ainu–Ryukyuan pair compared with those obtained in their respective comparison to mainland Japanese, which is another hint of its notable maternal isolation.

The distance and identity statistics used above are based on frequencies of haplogroups and haplotypes, respectively; however, frequencies are more affected by genetic drift than the number of different haplotypes present in a population. To measure the relative affinities of Japanese populations between them and to Continental Asia in a frequency-independent way, we chose a haplotype-sharing approach calculating the relative contribution of lineages shared with other areas to the number of different haplotypes present in each Japanese population. In these comparisons all other Asians were merged. Table 5 shows the results of this analysis. Note that despite the difference in sample size the haplotype frequency in mainland Japanese and Ainu is ~50%, whereas in Ryukyuans it is 84%; which means that, if there was not a bias in the sampling process, in spite of its small size, the Ainu sample seems to be representative of that population. However, it would be desirable to enlarge that of the Ryukyuans (Helgason et al. 2000). Haplotypes present only in a given population account for 13% in Ainu but ~50% in mainland Japanese (60%) and Ryukyuans (45%). This finding once more points to the existence of important drift effects in Ainu. Mainland Japanese exclusively share with Ryukyuans and Ainu only 3% and 2%, respectively, of its lineages, which could reach 6% and 3% if those

Table 3. Asian Populations Used in This Study

Population	Locality	Ethnic group	Group	Sample	HVRI	HVRII	Other[a]	References
Japan	Tokyo	Japanese	JPN	373	16024–16569	1–648	649–16023	This work
Japan	Nagoya	Japanese	JPN	299	16024–16569	1–648	649–16023	This work
	Japan	Japanese	JPN	20	1600–16413	—		Bamshad et al. 2001
				19	—	71–270		Jorde et al. 1995
Japan	Tokyo	Japanese	JPN	162	16051–16365	73–340		Imaizumi et al. 2002
Japan	Tokyo	Japanese	JPN	150	16030–16481			Nishimake et al. 1999
Japan	Tokyo	Japanese	JPN	13	16024–16569	1–648	RFLPs	Abe et al. 1998
Japan	Miyazaki	Japanese	JPN	100	15998–16400	30–407		Seo et al. 1998
Japan	Tottori	Japanese	JPN	89	16026–16396			Oota et al. 2002
Japan	Shizuoka	Japanese	JPN	62	16129–16569	1–41		Horai et al. 1996
Japan	Aichi	Japanese	JPN	.50	16040–16375	20–430		Koyama et al. 2002
Japan	Okinawa	Ryukyuan	RYU	50	16129–16569	1–41		Horai et al. 1996
Japan	Hokkaido	Ainu	AIN	51	16129–16569	1–41		Horai et al. 1996
Korea		Korean	KOR	306	16020–16400	1–70		Lee et al. 1997
Korea		Korean	KOR	4	16024–16370			Torroni et al. 1993a,b
Korea		Korean	KOR	60	16024–16365	73–340		Pfeiffer et al. 1998
Korea		Korean	KOR	2	16000–16413	—		Bamshad et al. 2001
Korea					—	71–270		Jorde et al. 1995
Korea		Korean	KOR	64	16129–16569	1–41		Horai et al. 1996
Korea		Korean	KOR	3	16128–16408			Horai and Hayasaka 1990
Korea		Korean	KOR	98	16075–16362	73–315	14747–15887	Lee et al. 2002
China	Liaoning	Han	Ch1	51	16001–16497	30–47	10171–10659 and RFLPs	Yao et al. 2002a
China	Shandong	Han	Ch1	50	16001–16497	30–47	10171–10659 and RFLPs	Yao et al. 2002a
China	Yunnan	Bai	Ch1	31	16001–16495			Yao et al. 2002b
China	Changsha	Han	Ch1	82	16026–16396			Oota et al. 2002
China	Xinjiang	Han	Ch2	47	16001–16497	30–47	10171–10659 and RFLPs	Yao et al. 2002a
China	Yunnan	Sali	Ch2	31	16001–16495			Yao et al. 2002b
China	Qinghai	Tu	Ch2	35	16001–16495			Yao et al. 2002b
China	Xi'an	Han	Ch2	84	16026–16396			Oota et al. 2002
China	Shanghai	Han	Ch2	120	13030–16481			Nishimake et al. 1999
Mongolia		Mongolian	Ch2	103	16020–16400		RFLPs	Kolman et al. 1996
Mongolia		Mongolian	Ch2	15	16001–16495			Yao et al. 2002b
China	Yunnan	Lahu	Ch3	32	16048–16569	1–49		Qian et al. 2001
China	Hubei	Han	Ch4	42	16001–16497	30–47	10171–10659 and RFLPs	Yao et al. 2002a
China	Guangdong	Han	Ch4	30	16001–16497	30–47	10171–10659 and RFLPs	Yao et al. 2002a
China	Yunnan	Han	Ch5	43	16001–16497	30–47	10171–10659 and RFLPs	Yao et al. 2002a
China	Taiwan		Ch5	6	16024–16370			Torroni et al. 1993a,b
China	Taiwan		Ch5	3	15999–16413			Bamshad et al. 2001
China	Taiwan		Ch5	9	16065–16375			Sykes et al. 1995
China	Taiwan		Ch5	66	16129–16569	1–41		Horai et al. 1996
China	Taiwan	Han	Ch5	155	15997–16569	1–407		Tsai et al. 2001
China	Yunnan	Dai	Ch5	21	16048–16569	1–49		Qian et al. 2001
China	Yunnan	Wa	Ch5	22	16048–16569	1–49		Qian et al. 2001
China	Yunnan	Dai	Ch5	38	16001–16495			Yao et al. 2002b
China	Guangxi	Zhuang	Ch5	83	16001–16495			Yao et al. 2002b
China	South China	Han	Ch5	28	16024–16399			Betty et al. 1996
Thailand			Ch5	32	16001–16495			Yao et al. 2002b
Thailand		See ref.	Ch5	121	16048–16569	1–41		Fucharoen et al. 2001
Thailand	See ref.	Native	Ch5	74	16048–16569	1–41		Fucharoen et al. 2001
Vietnam			Ch5	35	16026–16396			Oota et al. 2002
Vietnam			Ch5	9	15999–16413	—		Bamshad et al. 2001
					—	71–270		Jorde et al. 1995
Cambodia			Ch5	12	15999–16413	—		Bamshad et al. 2001
					—	71–270		Jorde et al. 1995
China	Yunnan	Lisu	Ch6	37	16001–16495			Yao et al. 2002b
China	Yunnan	Nu	Ch6	30	16001–16495			Yao et al. 2002b
China	Taiwan	Native	TWA	28	15997–16400	30–407		Melton et al. 1998
China	Taiwan	Native	TWA	180	16048–16569	1–41		Tajima et al. 2003
Central Asia		Uygur	CA1	46	16001–16495			Yao et al. 2000a
Kazagstan		Kazakh	CA1	55	15997–16400			Comas et al. 1998
Kirgizistan	Talas	Kirghiz	CA1	48	15997–16400			Comas et al. 1998
Kazagstan		Uygur	CA1	55	15997–16400			Comas et al. 1998
Central Asia		Kazak	CA2	30	16001–16495			Yao et al. 2000a
Kirgizistan	Sary-Tash	Kirghiz	CA2	46	15997–16400			Comas et al. 1998
Siberia	See ref.	Altai	CA2	17	16024–16383			Shields et al. 1993
Tibet		Tibetan	TIB	1	16024–16370			Torroni et al. 1993a,b
Tibet		Tibetan	TIB	40	16001–16495			Yao et al. 2000b
Tibet		Tibetan	TIB	24	16048–16569	1–41		Qian et al. 2001
Russia	East Ural	Mansi	MAN	98	16039–16519	64–295	RFLPs	Derbeneva et al. 2002a

(continued)

Table 3. *Continued*

Population	Locality	Ethnic group	Group	Sample	HVRI	HVRII	Other*	References
Siberia		Finno-Ugrian	FIU	38	13021–16505			Voevoda Accession nos. AF214068–AF214105
South Siberia		Tuvinian	TUV	36	16000–16400		RFLPs	Derenko et al. 2000
South Siberia		Buryat	BUR	40	16000–16400		RFLPs	Derenko et al. 2000
Siberia		Chukchi	CHU	60	16001–16405			Voevoda et al. 1994
Siberia	Aluitor	Koryak	ALU	56	16000–16525			Schurr et al. 1999
Siberia	Karagin	Koryak	KAM	37	16000–16525			Schurr et al. 1999
Siberia	Palan	Koryak	KAM	54	16000–16525			Schurr et al. 1999
Siberia	Kovran	Itel men	ITE	46	16000–16525			Schurr et al. 1999
Philippine			FIL	32	16065–16375			Sykes et al. 1995
Thailand	Trang	Sakai	SAK	20	16048–16569	1–41		Fucharoen et al. 2001
Malaysia			IND	6	15999–16413	—		Bamshad et al. 2001
					—	71–270		Jorde et al. 1995
Indonesia			IND	34	16024–16400	31–407		Redd and Stoneking 1999
Borneo		Sabah	SAB	34	16065–16375			Sykes et al. 1995

*RFLPs and additional sequences.

also shared with Continental Asian populations are added. In comparison they shared 21% of its lineages with other Asians. On the contrary, Ryukyuans and Ainu share about 50% of their lineages with mainland Japanese and only 10% and 21%, respectively, with Continental populations, which may reflect other independent Asian influences on Japan. With respect to those lineages exclusively shared by Japanese and Continental Asian populations, it is worth mentioning that, again, Korea is the main contributor, participating in ~50% of the haplotype sharing with mainland Japanese (55%), as much as with Ryukyuans (50%) and Ainu (50%). However, differences exist in the provenance of the rest of the shared lineages. Whereas in Ainu (northern China and Siberia) and in Ryukyuans (northern China and central Asia) they are from northern areas, the second region contributing to mainland Japanese is southern China (17.5%), followed, at the same level (12.5%), by northern China and central Asia. In addition, there exists a minor percentage of exclusive sharing with Indonesia (2.5%). On the other hand, all the matches with Siberia and Tibet are also shared with other populations. From these results, it can be deduced that the ancient Japanese inhabitants came from northern Asia and that southern areas affected the Japanese by later immigration. Nevertheless, it must be borne in mind that older influences could be undetectable by lineage sharing. With respect to the haplogroup affiliation of those lineages that Ainu and Ryukyuans exclusively shared with no Japanese samples, new differences appear between them. Ainu share derived lineages of haplogroups A, G, M9, and D5, all of them compatible with a rather recent Siberian influence. In contrast, those shared by Ryukyuans are basical M lineages, more congruent with an older radiation from southern China. These dual influences are also detected when the haplogroup affiliation of the Ainu and Ryukyuan unique lineages is studied. First, the percentage of lineages belonging to macrohaplogroup N is larger in Ainu (50%) than in Ryukyuans (15%) and from a different provenance, as those in Ainu are from haplogroups N, N9b, and Y, whereas those of Ryukyuans belong to the southern haplogroups F and B. The remaining 50% of the Ainu lineages equitably belong to different M haplogroups (M, M7c, G1, and D5a), but in Ryukyuans the remainder are mainly concentrated in M7a (41%) and M7b2 (18%), two groups that have their greatest Asian diversities precisely in Ryukyuans. Although an indigenous focus of radiation cannot be discarded, it is more conservative to suppose that the most probable origin of these lineages is again southern China. Thus, Ainu and Ryukyuans are not only largely isolated populations, but they most probably had different maternal origins.

Although no matches are involved, the geographic distribution of haplogroup frequency and diversities for some groups present in Japan and in other distinct Asian areas are also relevant to trace these older connections. For instance, haplogroups M9, M10, M12, D4b, and F1c have correlated geographic frequencies with a peak in an area that comprises Tibet (Table 2). Curiously, one of these haplogroups (M12) is today absent in China but present in Korea and Japan.

DISCUSSION

Although the recent out-of-Africa origin for all modern humans (Cann et al. 1987) is being widely supported (Takahata et al.

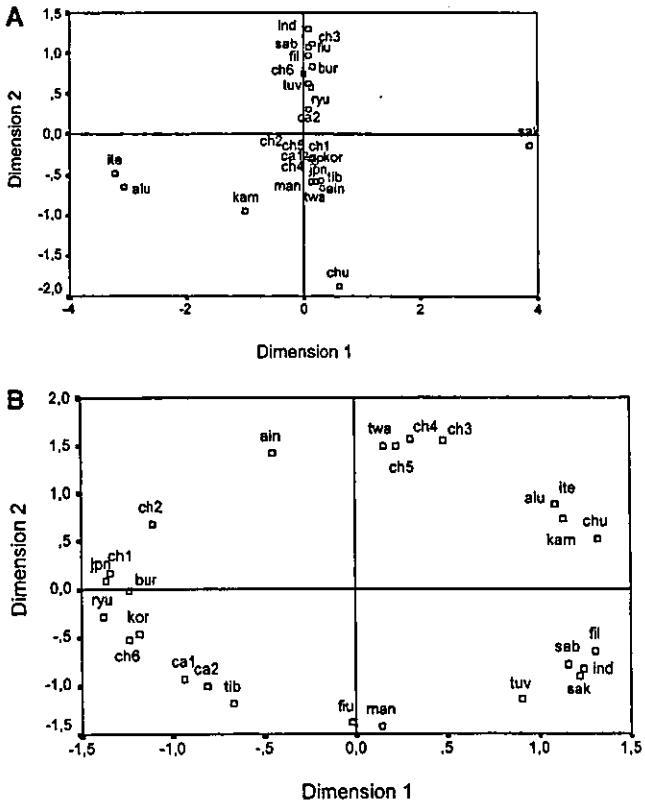

Figure 3 MDS plots based on (*A*) F_{ST} and (*B*) D match distances. Population groups are as detailed in Table 3.

Table 4. Frequency-Based F_{ST} and Sequence Match Identities (in Percentage) Between Japanese Samples and With Other Asian Populations

	JPN		RYU		AIN	
	F_{ST}	Matches	F_{ST}	Matches	F_{ST}	Matches
RYU	0.04	0.41				
AIN	0.04	0.33	0.05	0.04		
KOR	0.00	1.10	0.04	0.57	0.04	0.25
CH1	0.01	0.59	0.04	0.11	0.04	0.18
CH2	0.01	0.51	0.05	0.19	0.05	0.21
CH3	0.07	0.01	0.10	0.00	0.08	0.00
CH4	0.03	0.06	0.03	0.00	0.05	0.03
CH5	0.03	0.16	0.03	0.09	0.05	0.08
CH6	0.04	0.01	0.08	0.00	0.08	0.09
TWA	0.04	0.23	0.07	0.08	0.08	0.04
TIB	0.04	0.36	0.04	0.18	0.08	0.06
CA1	0.02	0.58	0.04	0.25	0.05	0.16
CA2	0.04	0.73	0.07	0.20	0.08	0.19
ITE	0.29	0.00	0.39	0.00	0.40	0.26
FIU	0.06	0.50	0.08	0.32	0.10	0.10
MAN	0.06	0.24	0.06	0.24	0.08	0.04
ALU	0.29	0.01	0.39	0.00	0.39	0.46
KAM	0.14	0.01	0.16	0.00	0.15	0.45
CHU	0.17	0.01	0.21	0.00	0.22	0.00
TUV	0.03	0.09	0.07	0.17	0.07	0.05
BUR	0.03	0.97	0.02	2.75	0.07	0.15
FIL	0.03	0.11	0.05	0.13	0.06	0.00
IND	0.09	0.04	0.09	0.00	0.11	0.00
SAK	0.29	0.00	0.44	0.00	0.43	0.00
SAB	0.06	0.09	0.05	0.29	0.08	0.12

2001), the most probable time and routes chosen by these earliest migrants to reach eastern Asia is an open issue. In the following discussion we weigh the different alternatives proposed in light of the phylogenetic tree obtained from complete mtDNA sequences. One of the first questions raised was whether there was more than one out-of-Africa dispersion. All the mtDNA lineages detected in Old World populations belong to one of two M and N macrohaplogroups with only secondary representatives in Africa. The proposed radiation ages for both, 30,000 to 58,000 years ago and 43,000 to 53,000 years ago, respectively (Maca-Meyer et al. 2001), give a temporal frame compatible with only one main dispersion or two successive dispersions, in which case the M precursor is the most probable candidate for the older exit. Even if the one dispersion option is chosen, more than one geographical route to eastern Asia is possible. In fact, a northern Continental route through the Near East and western-central Asia and a southern coastal route through the Arabian and Indian peninsulas have been proposed (Cavalli-Sforza et al. 1994; Kivisild et al. 1999). The geographical distribution of these two macrohaplogroups, with lack of ancient M representatives and the presence of deep N lineages in western Asia, and the abundance of basal M lineages in India and southwestern Asia and concomitant lack of equivalent-age N clades, gave rise to the hypothesis that N represents the main footprint of the northern Continental expansion, whereas M is the equivalent footprint for the southern coastal expansion. The presence of N and M lineages in alternative areas has been explained to have been the result of secondary migrations (Maca-Meyer et al. 2001). However, another plausible explanation is that both M and N reached southern Asia at the same time, quickly expanding to Papua New Guinea (PNG) during maximal glacial ages when the permafrost boundary precluded a northern human occupation. During postglacial ages, subsequent migrations northward carried derivatives of both macrohaplogroups to northern Asia (Forster et al. 2001). Nevertheless, under this second hypothesis, the presence of basal N clusters should be expected in India, southern Asia, and PNG; but this is not the case. All N representatives in India belong to R, a clade derived from N by the loss of 16223 and 12705 mutations (Fig. 2). In addition, the bulk of these Indian lineages belong to western Caucasian haplogroups that, most probably, reached India as the result of secondary immigrations, as has already been proposed (Kivisild et al. 1999; Bamshad et al. 2001). Similarly, the N representatives in southern Asia belong to haplogroups F and B, two sister clades also derived from R (Fig. 2). Furthermore, when totally sequenced PNG N lineages (Ingman et al. 2000; Ingman and Gyllensten 2003) are added to the N phylogenetic tree (data not shown), they form three monophyletic clades that have their roots in the derived R trunk. On the contrary, the geographically northern Asian clades A, N9a, N9b, and Y (Fig. 2) and the western Eurasian clades W, N1b, I, and X all split from the basal N root (Maca-Meyer et al. 2001), although A, N9a, N9b, and Y radiations were delayed congruent with subsequent northern Asian expansions. Therefore, at present, mtDNA data are compatible with the supposition that the northern route, harboring mainly N precursors, met climatic difficulties and when they finally reached Southeast Asia, the M representatives, brought by the southern route, had already colonized the area. This southern expansion of N derivatives has, as a lower temporal boundary, the coalescence ages of F, B, and PNG R haplogroups being ~46,000 ± 10,000 years ago. However, when recently published (Ingman et al. 2000; Ingman and Gyllensten 2003) Australian N lineages are taken into account, it seems evident that the real situation could be far more complex than the one migration–one lineage hypothesis. Australian N lineages directly sprout from the basal trunk (data not shown). They most probably differentiated in that continent, supporting the idea that ancestral N lineages reached Australia but not PNG, although the undemonstrable possibility of lineage extinctions and subsequent recolonization events in PNG can be an argument. Both hypotheses have difficulties to explain the presence of ancient N lineages in Australia. If the two, M and N lineages, were brought with the southern coastal dispersion, the lack of primitive N in India, southern Asia, and PNG has to be explained by the subsequent loss of all N lineages carried to Australia; if the northern Continental route of N is favored, the loss of N representatives in all populations formed in route to Australia has also to be explained. Recently, an N lineage has been detected in Chenchus, a southern Indian tribal group (Kivisild et al. 2003). From the information published, it can be deduced that this lineage only shares mutation 1719 with the western Eurasian Nb1/I and X clades. More extensive studies of populations in southern India

Table 5. Distribution of Unique and Shared Haplotypes in Japanese Populations

	Japanese populations		
	JPN	RYU	AIN
Sample	1318	50	51
Haplotypes	626	42	24
Haplotype frequency	0.48	0.84	0.47
Singleton + Unique	377 (0.60)	19 (0.45)	3 (0.13)
Shared	249 (0.40)	23 (0.55)	21 (0.87)
JPN	137 (0.22)	20 (0.48)	13 (0.54)
RYU	20 (0.03)	1 (0.02)	1 (0.04)
AIN	13 (0.02)	1 (0.02)	5 (0.21)
Other*	130 (0.21)	4 (0.10)	5 (0.21)

*Other Asians.

and southern and central Asia would add empirical support to any of these theories.

Concerning macrohaplogroup M, it has already been commented that the star radiation of all the main Indian and southeast Asian M clades strongly suggests that this wide geographic colonization could have happened in a relatively short time (Maca-Meyer et al. 2001). This star radiation includes the Australian and PNG M complete sequences recently published (Ingman et al. 2000; Ingman and Gyllensten 2003). However, for those clades and subclades with later northward expansions, long radiation delays are observed. For instance, whereas M7 and M8 have coalescence ages ~35,000 to 45,000 years ago, other groups such as G, D4, M7a, or M7c have coalescence ages ~15,000 to 30,000 years ago, more in frame with those calculated for A, Y, and N9 derivates, which, although belonging to macrohaplogroup N, share with them a central–northern Asian geographic distribution (see Supplemental material). It seems that the simultaneous lineage bursts ~60,000 to 70,000 years ago from Africa (Maca-Meyer et al. 2001), ~30,000 to 55,000 years ago for macrohaplogroups M and N, and ~15,000 to 30,000 years ago for clusters with prominent central–northern Asian radiations were related to main climatic changes. The role of selection in these expansions is an open question (Elson et al. 2004; Ruiz-Pesini et al. 2004).

The application of global pairwise-distance and detailed phylogeographic methods to the peopling of Japan shows that both approaches have different grasps but together demonstrate that the actual Japanese population is the result of a complex demographic history, from which the different theories proposed to explain it only emphasize partial aspects. Global distances and detailed haplotype comparisons confirm that Ainu and Ryukyuans are heterogeneous populations (Horai et al. 1996) and that both are well differentiated from the mainland Japanese. In spite of this, they have common peculiarities such as having the highest frequencies in Asia for M7a, M7b2, and N9b, shared with mainland Japanese. Furthermore, for both, their closest relatives are northern populations. At first sight, these results are against a supposed southern origin for the Paleolithic Japanese, favoring the replacement theory or even that the Paleolithic inhabitants of Japan came from northeastern Asia (Nei 1995). Although based on a single locus, our results are strikingly coincident with the previously proposed northern origin and influences received by the Japanese. In an early study using serum gammaglobulin polymorphisms, it was concluded that the homeland of all Japanese could have been in the Lake Baikal area in Siberia (Matsumoto 1988), which agrees with the close proximity found here between Buryats and Ryukyuans or mainland Japanese. More recently, classical markers (Omoto and Saitou 1997) and mtDNA (Horai et al. 1996) studies demonstrated that the Japanese are most closely related to the Koreans, which is also true in our global analysis. It can be added that a substantial part of this common maternal pool has recent roots, as Korea specifically shares with Ainu, mainland Japanese, and Ryukyuans 10%, 7%, and 5%, respectively, of their haplotypes. This particular affinity is increased with the existence of derived lineages only detected (A1a, B4c1, B4f) or mainly detected (N9b, B4a1, B4b1, G1a, M7b2, M12) in Japanese and Koreans. This Korean influence has been attributed to the archeologically well-documented Continental immigration to Japan during the Yayoi period (Horai et al. 1996). However, specific haplotype matches with other areas increases the geographic range of these recent influences. Thus, mainland Japanese share part of their haplotypes exclusively with South China (2.5%), North China (1.5%), Central Asia (1.5%), and Indonesia (0.3%); and, also, Ryukyuans have specific affinities with North China (2.4%) and Central Asia (2.4%). The recent Siberian input on the Ainu has also been stressed (Schurr et al. 1999). At least, another independent migratory wave from central Asia also affected mainland Japanese. It was first detected by the peculiar distribution of the Y-chromosome marker YAP+, and seems to have originated in an area including Tibet (Su et al. 2000). Haplogroup M12 is its mitochondrial counterpart. As with the Y-chromosome marker, its punctual presence in Tibet and eastern Asia might be explained as the result of subsequent migrations in the Continent that erased the route followed by the people harboring these markers. In addition, there are clues, at least in Ryukyuans, that a substantial part of their maternal pool had an ancient southern Asian provenance. This fraction is represented by the M, M7a, and M7a1 basic lineages (31%), which the Ryukyuans do not share with northern populations. This southern signal is, in part, congruent with the southern Asian origin for the Paleolithic Japanese proposed by the dual structure model (Hanihara 1991). Furthermore, the fact that the highest diversities for M7a, M7a1, and M7b2 have been found in Ryukyuans and for N9b and B5b2 in Japan raises the possibility that this area was within a focus of migratory radiations to northern and southern isles and even to the mainland from Paleolithic to recent times. The significant latitudinal clines detected in Japan for some genetic markers (Orito et al. 2001; Takeshita et al. 2001) could also be explained as the result of southern and northern influences on Japanese. Finally, some mtDNA results obtained from ancient Jomon remains (Horai et al. 1991; Shinoda and Kanai 1999; K.-I. Shinoda, unpubl.) are congruent with a genetically diverse background for the Paleolithic Japanese population (Horai et al. 1996). A tentative comparison of Jomon with present-day Japanese populations based on shared lineages (data not shown) significantly relates Jomon first to the indigenous Ainu and then to Ryukyuans and last to mainland Japanese. In summary, Japan could have received several northern and southern Asian maternal inputs since Paleolithic times, with notable northern Asian immigrations through Korea in the late Neolithic and more specific gene flows from western Asia, Siberia, and southern islands.

METHODS

Samples

Complete mtDNA sequences were obtained from a total of 672 unrelated Japanese including 373 from Tokyo and 299 from the Nagoya area. All subjects gave their written consent to participate in this study, which was approved by the Ethical Committees of the Gifu International Institute of Biotechnology and collaborative institutions. The sources of 11 additional complete sequences used to build the final phylogenetic trees are in Table 1. For the analysis of the peopling of Japan, we used a total of 1438 Japanese and 3275 central and eastern Asian HVI sequences, as detailed in Table 3.

Isolation and Amplification of DNA

Total DNA was extracted from the blood with either Dr. Gen TLE (Takara) or MagExtractor System MFX-2000 (Toyobo). The entire mitochondrial genome was amplified as six fragments (~3000–3400 bp) by the first PCR and 60 overlapping segments (~600–1000 bp) by the second PCR. The primer pairs and their nucleotide sequences were described previously (Tanaka et al. 1996). The conditions for the first and second PCR were the same: an initial denaturation step for 5 min at 94°C, followed by 40 cycles of denaturation for 15 sec at 94°C, annealing for 15 sec at 60°C, and extension for 3 min at 72°C, with a final extension for 10 min at 72°C. The amplified fragments were analyzed by electrophoresis on a 1% agarose gel and visualized by staining with ethidium bromide. These second PCR products were purified by use of the MultiScreen-PCR Plates (Millipore). The quality of DNA templates was examined by electrophoresis on a 1.2% agarose gel after staining with ethidium bromide by use of a Ready-To-Run Separation Unit (Amersham Pharmacia Biotech).

Sequence Analysis of Mitochondrial DNA

Sequence reactions were carried out with a BigDye terminator cycle sequencing FS ready reaction kit (Applied Biosystems). After excess dye terminators had been removed with MultiScreen-HV plates (Millipore) packed with Sephadex G50 superfine (Pharmacia), the purified DNA samples were precipitated with ethanol, dried, and suspended in the template suppression reagent (TSR) or formamide from Applied Biosystems. The dissolved DNA samples were heated for 2 min at 95°C for denaturation, then immediately cooled on ice. Sequences were analyzed with automated DNA sequencers 377 and 310 by use of Sequencing Analysis Program version 4.1 (Applied Biosystems). A computer program, Sequencher version 4.1 (Gene Codes Co.), was used to indicate possible single nucleotide polymorphism (SNP) loci. For verification, visual inspection of each candidate SNP was carried out. At least two overlapping DNA templates amplified with different primer pairs were used for identification of each SNP. Mitochondrial SNPs (mtSNPs) were identified by comparison with the revised Cambridge sequence (rCRS) reported by Andrews et al. (1999).

Phylogenetic Analysis of Complete Coding-Region mtDNA Sequences

In this present study, nucleotide positions were numbered as in the Cambridge Reference Sequence (CRS; Anderson et al. 1981), nucleotide substitutions were expressed as differences from the revised CRS (Andrews et al. 1999), transitions were denoted only by their nucleotide positions, and transversions were designated by their nucleotide positions followed by the changed base. A total of 942 complete coding-region mtDNA sequences, including our 672 Japanese; one additional Japanese (GenBank accession no. AB055387); 53 worldwide sequences (Ingman et al. 2000); 42 worldwide sequences (Maca-Meyer et al. 2001); two Finnish sequences having Asian relatives (Finnilä et al. 2001); 17 Asian sequences without concrete geographic assignation (Herrnstadt et al. 2002); 37 sequences from the Bering area (Derbeneva et al. 2002b); 70 Asian, New Guinean, and Australian sequences (Ingman and Gyllensten 2003); and 48 Chinese sequences (Kong et al. 2003) were aligned with the rCRS by CLUSTAL V software, and the coding region was used to construct a phylogenetic network (Bandelt et al. 1999) rooted with a chimpanzee sequence (GenBank accession no. D38113) as implemented in the Network 3.1 program (Fluxus Engineering; http://www.fluxus-engineering.com). The noncoding positions were added by hand using molecular weighted parsimony criteria (Bandelt et al. 2000). The phylogenetic relationships obtained were also confirmed by means of a neighbor-joining tree (1000× bootstrapped; Saitou and Nei 1987), built using MEGA2 (Kumar et al. 2001). From this network (see Supplemental material) we chose 102 Japanese and nine Asiatic sequences that represented the main clusters and subclusters within the two macrohaplogroups M and N that colonized Asia. To define these groups we followed the most generalized cladistic nomenclature actually used to classify mtDNA lineages (Richards et al. 1998). For the haplogroups previously detected, we maintained the same notation as their authors proposed (Richards et al. 2000; Bamshad et al. 2001; Kivisild et al. 2002; Yao et al. 2002a; Kong et al. 2003). Those haplogroups introduced here for the first time were named according to their phylogenetic range deduced from the tree of complete sequences.

Haplogroup Assorting of Published Partial mtDNA Sequences

The unambiguously classified complete mtDNA sequences were used as an initial pool that was hierarchically enlarged by the successive addition of those published partial mtDNA sequences with the largest coding information, ending with those for which information on only control-region sequences for both mtDNA hypervariable segments or just one (HVS-I and/or HVS-II) was available, always following sequence matches or, as default, sequence-relatedness criteria. Some of those partial sequences that could be assigned to more than one haplogroup were tentatively assorted in the most probable one deduced from their geographic origin and the relative haplogroup distribution.

Pooling Small Size Samples and Rare Clades

To avoid small sample sizes and rare alleles in population comparisons, samples with <20 individuals were pooled with others from the same geographic and ethnic group. Within populations, individuals belonging to rare clades were pooled with those classified in the nearest branch. Pairwise sample distances were calculated as linearized F_{ST} distances as implemented in the ARLEQUIN program (Schneider et al. 2000), taking mtDNA as one locus with as many alleles as the different subhaplogroups considered.

Quantitative Affinities of Japanese Samples

Relative affinities of Japanese samples to the other Asiatic populations were assessed by linearized F_{ST} distances, using subhaplogroup frequencies, and haplotype matches' distances (D) estimated simply as $D = 1 - \Sigma(x_i y_i)$, x_i and y_i being the frequency of haplotype i in the two compared populations. To be statistically robust, these analyses require large sample sizes, thus further pooling was necessary. Previous studies in the area prevented us from pooling populations by geographic proximity (Schurr et al. 1999) and/or ethno-linguistic relationship (Comas et al. 1998; Chunjie et al. 2000; Yao et al. 2002a). For this reason, a genetic affinity criterion was chosen. Two approaches were used. In the first, all samples with no significant F_{ST} distances between them and with a similar behavior to the rest of the samples studied, were grouped. In the second, pooling was carried out by means of the CLUSTER algorithm implemented in the SPSS ver 9 package. We followed an iterative method specifying the number of conglomerates from 2 to 30. Different groupings were tested by AMOVA, and that with the least assigned variance within areas was chosen. The data were graphically represented by multidimensional scaling (MDS) plots (Kruskal and Wish 1978) using SPSS.

Qualitative Affinities of Japanese Samples

Particular sharing of subhaplogroups and particular haplotype matches of Japanese samples with concrete Continental areas were phylogeographically analyzed by taking into account the relative genetic diversities of the clades involved in the different areas, measured as relative haplotypic frequencies, and their minimum estimates of coalescence ages based on mean divergence among lineages for the coding region (Saillard et al. 2000). A constant evolutionary rate of 1.7×10^{-8} per site per year (Ingman et al. 2000) was used.

ACKNOWLEDGMENTS

This work was supported in part by the Support Project for Database Development from the Japan Science and Technology Corporation (to M.T.), Grants-in-Aid for Scientific Research (C2-10832009, A2-15200051) and for Priority Areas from the Ministry of Education, Science, Sports and Culture of Japan (to M.T.), and by grants BMC2001-3511 and COF2002-015 (to V.M.C.).

REFERENCES

Abe, S., Usami, S., Shinkawa, H., Weston, M.D., Overbeck, L.D., Hoover, D.M., Kenyon, J.B., Horai, S., and Kimberling, W.J. 1998. Phylogenetic analysis of mitochondrial DNA in Japanese pedigrees of sensorineural hearing loss associated with the A1555G mutation. *Eur. J. Hum. Genet.* **6:** 563–569.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290:** 457–465.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23:** 147.

Ballinger, S.W., Schurr, T.G., Torroni, A., Gan, Y.Y., Hodge, J.A., Hassan,

K., Chen, K.H., and Wallace, D.C. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics* **130:** 139–152.
Bamshad, M., Kivisild, T., Watkins, W.S., Dixon, M.E., Ricker, C.E., Rao, B.B., Naidu, J.M., Prasad, B.V., Reddy, P.G., Rasanayagam, A., et al. 2001. Genetic evidence on the origins of Indian caste populations. *Genome Res.* **11:** 994–1004.
Bandelt, H.-J., Forster, P., and Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16:** 37–48.
Bandelt, H.-J., Macaulay, V., and Richards, M. 2000. Median networks: Speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol. Phylogenet. Evol.* **16:** 8–28.
Betty, D.J., Chin-Atkins, A.N., Croft, L., Sraml, M., and Easteal, S. 1996. Multiple independent origins of the COII/tRNA(Lys) intergenic 9-bp mtDNA deletion in aboriginal Australians. *Am. J. Hum. Genet.* **58:** 428–433.
Cann, R.L. and Wilson, A.C. 1983. Length mutations in human mitochondrial DNA. *Genetics* **104:** 669–711.
Cann, R.L., Stoneking, M., and Wilson, A.C. 1987. Mitochondrial DNA and human evolution. *Nature* **325:** 31–36.
Cavalli-Sforza, L.L., Menozzi, P., and Piazza, A. 1994. *The history and geography of human genes.* Princeton University Press, Princeton, NJ.
Chunjie, X., Cavalli-Sforza, L.L., Minch, E., and Ruofu, D.U. 2000. Principal component analysis of gene frequencies of Chinese populations. *Science in China Ser. C* **43:** 472–481.
Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E., Martinez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., et al. 1998. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am. J. Hum. Genet.* **63:** 1824–1838.
Derbeneva, O.A., Starikovskaya, E.B., Wallace, D.C., and Sukernik, R.I. 2002a. Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. *Am. J. Hum. Genet.* **70:** 1009–1014.
Derbeneva, O.A., Sukernik, R.I., Volodko, N.V., Hosseini, S.H., Lott, M.T., and Wallace, D.C. 2002b. Analysis of mitochondrial DNA diversity in the Aleuts of the commander islands and its implications for the genetic history of Beringia. *Am. J. Hum. Genet.* **71:** 415–421.
Derenko, M.V., Malyarchuk, B.A., Dambueva, I.K., Shaikhaev, G.O., Dorzhu, C.M., Nimaev, D.D., and Zakharov, I.A. 2000. Mitochondrial DNA variation in two South Siberian Aboriginal populations: Implications for the genetic history of North Asia. *Hum. Biol.* **72:** 945–973.
Elson, J.L., Turnbull, D.M., and Howell, N. 2004. Comparative genomics and the evolution of human mitochondrial DNA: Assessing the effects of selection. *Am. J. Hum. Genet.* **74:** 229–238.
Finnilä, S., Lehtonen, M.S., and Majamaa, K. 2001. Phylogenetic network for European mtDNA. *Am. J. Hum. Genet.* **68:** 1475–1484.
Forster, P., Harding, R., Torroni, A., and Bandelt, H.J. 1996. Origin and evolution of Native American mtDNA variation: A reappraisal. *Am. J. Hum. Genet.* **59:** 935–945.
Forster, P., Torroni, A., Renfrew, C., and Röhl, A. 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol. Biol. Evol.* **18:** 1864–1881.
Fucharoen, G., Fucharoen, S., and Horai, S. 2001. Mitochondrial DNA polymorphisms in Thailand. *J. Hum. Genet.* **46:** 115–125.
Glover, I.C. 1980. Agricultural origins in East Asia. In *The Cambridge encyclopedia of archaeology* (ed. A. Sherratt), pp. 152–161. Crown, New York.
Hammer, M.F. and Horai, S. 1995. Y chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.* **56:** 951–962.
Hanihara, K. 1991. Dual structure model for the population history of the Japanese. *Japan Review* **2:** 1–33.
Helgason, A., Sigureth Ardottir, S., Gulcher, J.R., Ward, R., and Stefansson, K. 2000. mtDNA and the origin of the Icelanders: Deciphering signals of recent population history. *Am. J. Hum. Genet.* **66:** 999–1016.
Helgason, A., Hickey, E., Goodacre, S., Bosnes, V., Stefánsson, K., Ward, R., and Sykes, B. 2001. mtDNA and the islands of the North Atlantic: Estimating the proportions of Norse and Gaelic ancestry. *Am. J. Hum. Genet.* **68:** 723–737.
Herrnstadt, C., Elson, J.L., Fahy, E., Preston, G., Turnbull, D.M., Anderson, C., Ghosh, S.S., Olefsky, J.M., Beal, M.F., Davis, R.E., et al. 2002. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am. J. Hum. Genet.* **70:** 1152–1171.
Horai, S. and Hayasaka, K. 1990. Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am. J. Hum. Genet.* **46:** 828–842.
Horai, S. and Matsunaga, E. 1986. Mitochondrial DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. *Hum. Genet.* **72:** 105–117.
Horai, S., Kondo, R., Murayama, K., Hayashi, S., Koike, H., and Nakai, N. 1991. Phylogenetic affiliation of ancient and contemporary humans inferred from mitochondrial DNA. *Phil. Trans. R Soc. Lond. B* **333:** 409–417.
Horai, S., Murayama, K., Hayasaka, K., Matsubayashi, S., Hattori, Y., Fucharoen, G., Harihara, S., Park, K.S., Omoto, K., and Pan, I.H. 1996. mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. *Am. J. Hum. Genet.* **59:** 579–590.
Imaizumi, K., Parsons, T.J., Yoshino, M., and Holland, M.M. 2002. A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals. *Int. J. Legal. Med.* **116:** 68–73.
Ingman, M. and Gyllensten, U. 2003. Mitochondrial genome variation and evolutionary history of Australian and New Guinean Aborigines. *Genome Res.* **13:** 1600–1606.
Ingman, M., Kaessmann, H., Pääbo, S., and Gyllensten, U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* **408:** 708–713.
Jorde, L.B., Bamshad, M.J., Watkins, W.S., Zenger, R., Fraley, A.E., Krakowiak, P.A., Carpenter, K.D., Soodyall, H., Jenkins, T., and Rogers, A.R. 1995. Origins and affinities of modern humans: A comparison of mitochondrial and nuclear genetic data. *Am. J. Hum. Genet.* **57:** 523–538.
Kivisild, T., Kaldma, K., Metspalu, M., Parik, J., Papiha, S., and Villems, R. 1999. The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the Old World. In *Genomic diversity: Applications in human population genetics* (eds. S. Papiha et al.), pp. 135–152. Plenum Press, New York.
Kivisild, T., Tolk, H.-V., Parik, J., Wang, Y., Papiha, S.S., Bandelt, H.-J., and Villems, R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. *Mol. Biol. Evol.* **19:** 1737–1751.
Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H.V., Stepanov, V., et al. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* **72:** 313–332.
Kolman, C.J., Sambuughin, N., and Bermingham, E. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* **142:** 1321–1334.
Kong, Q.-P., Yao, Y.-G., Sun, C., Bandelt, H.-J., Zhu, C.-L., and Zhang, Y.-P. 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am. J. Hum. Genet.* **73:** 671–676.
Koyama, H., Iwasa, M., Maeno, Y., Tsuchimochi, T., Isobe, I., Seko-Nakamura, Y., Monma-Ohtaki, J., Matsumoto, T., Ogawa, S., Sato, B., et al. 2002. Mitochondrial sequence haplotype in the Japanese population. *Forensic Sci. Int.* **125:** 93–96.
Kruskal, J.B. and Wish, M. 1978. *Multidimensional scaling.* Sage Publications, Beverly Hills, CA.
Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* **17:** 1244–1245.
Lee, S.D., Shin, C.H., Kim, K.B., Lee, Y.S., and Lee, J.B. 1997. Sequence variation of mitochondrial DNA control region in Koreans. *Forensic Sci. Int.* **87:** 99–116.
Lee, S.D., Lee, Y.S., and Lee, J.B. 2002. Polymorphism in the mitochondrial cytochrome B gene in Koreans. An additional marker for individual identification. *Int. J. Legal Med.* **116:** 74–78.
Maca-Meyer, N., González, A.M., Larruga, J.M., Flores, C., and Cabrera, V.M. 2001. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.* **2:** 13–20.
Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B., and Torroni, A. 1999. The emerging tree of West Eurasian mtDNAs: A synthesis of control-region sequences and RFLP. *Am. J. Hum. Genet.* **64:** 232–249.
Malyarchuk, B.A. and Derenko, M.V. 2001. Mitochondrial DNA variability in Russians and Ukrainians: Implication to the origin of the Eastern Slavs. *Ann. Hum. Genet.* **65:** 63–78.
Matsumoto, H. 1988. Characteristics of Mongoloid and neighboring populations based on the genetic markers of human immunoglobulins. *Hum. Genet.* **80:** 207–218.
Melton, T., Clifford, S., Martinson, J., Batzer, M., and Stoneking, M. 1998. Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *Am. J. Hum. Genet.* **63:** 1807–1823.
Nei, M. 1995. The origins of human populations: Genetic, linguistic, and archeological data. In *The origin and past of modern humans as viewed from DNA* (eds. S. Brenner and K. Hanihara), pp. 71–91. World Scientific, Singapore.
Nishimaki, Y., Sato, K., Fang, L., Ma, M., Hasekura, H., and Boettcher, B. 1999. Sequence polymorphism in the mtDNA HV1 region in Japanese and Chinese. *Legal Med.* **1:** 238–249.

Omoto, K. and Saitou, N. 1997. Genetic origins of the Japanese: A partial support for the dual structure hypothesis. *Am. J. Phys. Anthropol.* **102:** 437–446.
Oota, H., Kitano, T., Jin, F., Yuasa, I., Wang, L., Ueda, S., Saitou, N., and Stoneking, M. 2002. Extreme mtDNA homogeneity in continental Asian populations. *Am. J. Phys. Anthropol.* **118:** 146–153.
Orito, E., Ichida, T., Sakugawa, H., Sata, M., Horiike, N., Hino, K., Okita, K., Okanoue, T., Iino, S., Tanaka, E., et al. 2001. Geographic distribution of hepatitis B virus (HBV) genotype in patients with chronic HBV infection in Japan. *Hepatology* **34:** 590–594.
Pfeiffer, H., Steighner, R., Fisher, R., Mornstad, H., Yoon, C.L., and Holland, M.M. 1998. Mitochondrial DNA extraction and typing from isolated dentin-experimental evaluation in a Korean population. *Int. J. Legal Med.* **111:** 309–313.
Qian, Y.P., Chu, Z.T., Dai, Q., Wei, C.D., Chu, J.Y., Tajima, A., and Horai, S. 2001. Mitochondrial DNA polymorphisms in Yunnan nationalities in China. *J. Hum. Genet.* **46:** 211–220.
Quintana-Murci, L., Semino, O., Bandelt, H.-J., Passarino, G., McElreavey, K., and Santachiara-Benereccetti, A.S. 1999. Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat. Genet.* **23:** 437–441.
Redd, A.J. and Stoneking, M. 1999. Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. *Am. J. Hum. Genet.* **65:** 808–828.
Richards, M., Macaulay, V., Bandelt, H.-J., and Sykes, B. 1998. Phylogeography of mitochondrial DNA in western Europe. *Ann. Hum. Genet.* **62:** 241–260.
Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **67:** 1251–1276.
Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., and Wallace, D.C. 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* **303:** 223–226.
Saillard, J., Forster, P., Lynnerup, N., Bandelt, H.J., and Norby, S. 2000. mtDNA variation among Greenland Eskimos: The edge of the Beringian expansion. *Am. J. Hum. Genet.* **67:** 718–726.
Saitou, N. and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4:** 406–425.
Schneider, S., Roessli, D., and Excoffier, L. 2000. *Arlequin ver. 2000: A software for population genetics data analysis.* Genetic and Biometry Laboratory, University of Geneva, Switzerland.
Schurr, T.G., Sukernik, R.I., Starikovskaya, Y.B., and Wallace, D.C. 1999. Mitochondrial DNA variation in Koryaks and Itel'men: Population replacement in Okhotsk Sea–Bering Sea region during the Neolithic. *Am. J. Phys. Anthropol.* **108:** 1–39.
Seo, Y., Stradmann-Bellinghausen, B., Rittner, C., Takahama, K., and Schneider, P.M. 1998. Sequence polymorphism of mitochondrial DNA control region in Japanese. *Forensic Sci. Int.* **97:** 155–164.
Shields, G.F., Schmiechen, A.M., Frazier, B.L., Redd, A., Voevoda, M.I., Reed, J.K., and Ward, R.H. 1993. mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am. J. Hum. Genet.* **53:** 549–562.
Shinoda, K.-I. and Kanai, S. 1999. Intracemetry genetic analysis at the Nakazuma Jomon site in Japan by mitochondrial DNA sequencing. *Anthropol. Sci.* **107:** 129–140.
Shiraishi, T. 2002. Wakoku tanjou (The formation of ancient Japanese society). In *History of Japan 1* (ed. T. Shiraishi et al.), pp. 8–94. Yoshikawakobunkan, Tokyo, Japan (in Japanese).
Snäll, N., Savontaus, M.-L., Kares, S., Lee, M.S., Cho, E.K., Rinne, J.O., and Huoponen, K. 2002. A rare mitochondrial DNA haplotype observed in Koreans. *Hum. Biol.* **74:** 253–262.
Soodyall, H., Jenkins, T., and Stoneking, M. 1995. 'Polynesian' mtDNA in the Malagasy. *Nat. Genet.* **10:** 377–378.
Stoneking, M., Jorde, L.B., Bhatia, K., and Wilson, A.C. 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* **124:** 717–733.
Su, B., Xiao, C., Deka, R., Seielstad, M.T., Kangwanpong, D., Xiao, J., Lu, D., Underhill, P., Cavalli-Sforza, L., Chakraborty, R., et al. 2000. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum. Genet.* **107:** 582–590.
Sykes, B., Leiboff, A., Low-Beer, J., Tetzner, S., and Richards, M. 1995. The origins of the Polynesians: An interpretation from mitochondrial lineage analysis. *Am. J. Hum. Genet.* **57:** 1463–1475.
Tajima, A., Sun, C.-S., Pan, I.-H., Ishida, T., Saitou, N., and Horai, S. 2003. Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: Implications for the population history of aboriginal Taiwanese. *Hum. Genet.* **113:** 24–33.
Takahata, N., Lee, S.-H., and Satta, Y. 2001. Testing multiregionality of modern human origins. *Mol. Biol. Evol.* **18:** 172–183.
Takeshita, T., Yasuda, Y., Nakashima, K., Mogi, K., Kishi, H., Shiono, K., Sagisaka, I., Yuasa, H., Nishimukai, H., and Kimura, H. 2001. Geographical north-south decline in *DNASE*2* in Japanese populations. *Hum. Biol.* **73:** 129–134.
Tanaka, M., Hayakawa, M., and Ozawa, T. 1996. Automated sequencing of mitochondrial DNA. *Methods Enzymol.* **264:** 407–421.
Torroni, A., Schurr, T.G., Yang, C.C., Szathmary, E.J., Williams, R.C., Schanfield, M.S., Troup, G.A., Knowler, W.C., Lawrence, D.N., Weiss, K.M., et al. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* **130:** 153–162.
Torroni, A., Sukernik, R.I., Schurr, T.G., Starikorskaya, Y.B., Cabell, M.F., Crawford, M.H., Comuzzie, A.G., and Wallace, D.C. 1993a. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* **53:** 591–608.
Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M., and Wallace, D.C. 1993b. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* **53:** 563–590.
Torroni, A., Miller, J.A., Moore, L.G., Zamudio, S., Zhuang, J., Droma, T., and Wallace, D.C. 1994. Mitochondrial DNA analysis in Tibet: Implications for the origin of the Tibetan population and its adaptation to high altitude. *Am. J. Phys. Anthropol.* **93:** 189–199.
Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.-L., and Wallace, D.C. 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144:** 1835–1850.
Torroni, A., Rengo, C., Guida, V., Cruciani, F., Sellitto, D., Coppa, A., Calderon, F.L., Simionati, B., Valle, G., Richards, M., et al. 2001. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am. J. Hum. Genet.* **69:** 1348–1356.
Tsai, L.C., Lin, C.Y., Lee, J.C., Chang, J.G., Linacre, A., and Goodwin, W. 2001. Sequence polymorphism of mitochondrial D-loop DNA in the Taiwanese Han population. *Forensic Sci. Int.* **119:** 239–247.
Voevoda, M.I., Avksentyuk, A.V., Ivanova, A.V., Astakhova, T.I., Babenko, V.N., Kurilovich, S.A., Duffy, L.K., Segal, B., and Shields, G.F. 1994. Molecular genetic studies in the population of native inhabitants of Chukchee Peninsula. Analysis of polymorphism of mitochondrial DNA and of genes controlling alcohol metabolizing enzymes. *Sibirskii Ekolog. Z.* **1:** 139–151.
Watson, E., Forster, P., Richards, M., and Bandelt, H.J. 1997. Mitochondrial footprints of human expansions in Africa. *Am. J. Hum. Genet.* **61:** 691–704.
Yao, Y.G., Lu, X.M., Luo, H.R., Li, W.H., and Zhang, Y.P. 2000a. Gene admixture in the silk road region of China: Evidence from mtDNA and melanocortin 1 receptor polymorphism. *Genes Genet. Syst.* **75:** 173–178.
Yao, Y.G., Watkins, W.S., and Zhang, Y.P. 2000b. Evolutionary history of the mtDNA 9-bp deletion in Chinese populations and its relevance to the peopling of east and southeast Asia. *Hum. Genet.* **107:** 504–512.
Yao, Y.G., Kong, Q.P., Bandelt, H.J., Kivisild, T., and Zhang, Y.P. 2002a. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am. J. Hum. Genet.* **70:** 635–651.
Yao, Y.-G., Nie, L., Harpending, H., Fu, Y.-X., Yuan, Z.-G., and Zhang, Y.-P. 2002b. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am. J. Phys. Anthropol.* **118:** 63–76.

WEB SITE REFERENCES

http://www.fluxus-engineering.com; Network 3.1 program, Fluxus Engineering.
http://www.glib.or.jp/mtsnp/index_e.html; authors' data.

Received December 17, 2003; accepted in revised form June 14, 2004.

Association of a −1997G → T Polymorphism of the Collagen Iα1 Gene with Bone Mineral Density in Postmenopausal Japanese Women

YOSHIJI YAMADA,[1] Fujiko Ando,[2] Naoakira Niino,[2] and Hiroshi Shimokata[2]

Abstract Genetic variants that affect collagen Iα1 metabolism may be important in the development of osteoporosis or osteoporotic fractures. A −1997G → T polymorphism in the promoter of the collagen Iα1 gene (COL1A1) was shown to be associated with bone mineral density (BMD) for the lumbar spine in postmenopausal Spanish women. The relation of this polymorphism to BMD in Japanese women or men has now been examined in a population-based study. The subjects (1,110 women, 1,126 men) were 40 to 79 years of age and were randomly recruited for a population-based prospective cohort study of aging and age-related diseases. BMD for the lumbar spine, right femoral neck, right trochanter, and right Ward's triangle was measured using dual-energy x-ray absorptiometry. Genotypes for the −1997G → T polymorphism of COL1A1 were determined with a fluorescence-based allele-specific DNA primer assay system. When all women were analyzed together, BMD for the lumbar spine and trochanter was significantly lower in subjects with the *COL1A1*G/*G* genotype than in those in the combined group of *COL1A1*G/*T* and *COL1A1*T/*T* genotypes. When postmenopausal women were analyzed separately, BMD for the femoral neck and trochanter was also significantly lower in those with the *COL1A1*G/*G* genotype than in those with the *COL1A1*G/*T* genotype or those in the combined group of *COL1A1*G/*T* and *COL1A1*T/*T* genotypes. BMD was not associated with −1997G → T genotype in premenopausal women or in men. Multivariate regression analysis revealed that −1997G → T genotype affected BMD at various sites with a variance of 0.46–0.62% for all women and 0.61–1.01% for postmenopausal women. The −1997G → T genotype was not related to the serum concentration of osteocalcin, the serum activity of bone-specific alkaline phosphatase, or the urinary excretion of deoxypyridinoline or cross-linked N-telopeptides of type I collagen in men or in premenopausal or postmenopausal women. These results suggest that COL1A1 is a susceptibility locus for reduced BMD in postmenopausal Japanese women.

[1]Department of Human Functional Genomics, Life Science Research Center, Mie University, 1515 Kamihama, Tsu, Mie 514-8507, Japan.
[2]Department of Epidemiology, National Institute for Longevity Sciences, Obu, Aichi, Japan.

Human Biology, February 2005, v. 77, no. 1, pp. 27–36.
Copyright © 2005 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: BONE MINERAL DENSITY, COLLAGEN Iα1, COLA1A, OSTEOPOROSIS, JAPAN.

11214$ $CH3 01-04-05 13:53:53 PS PAGE 27

Osteoporosis, a major health problem of the elderly, is characterized by a reduction in bone mineral density (BMD) and a deterioration in the microarchitecture of bone, both of which result in predisposition to fractures (Kanis et al. 1994). Although several environmental factors, including diet, smoking, and physical exercise, influence BMD, a genetic contribution to this parameter has been recognized (Peacock et al. 2002). Genetic linkage analyses (Morrison et al. 1994; Johnson et al. 1997; Devoto et al. 1998; Koller et al. 1998, 2000; Niu et al. 1999) and candidate gene association studies (Morrison et al. 1994; Kobayashi et al. 1996; Uitterlinden et al. 1998; Yamada et al. 2001; Ishida et al. 2003) have implicated several loci and candidate genes in the regulation of bone mass and the prevalence of osteoporosis or osteoporotic fractures. However, the genes that contribute to genetic susceptibility to osteoporosis remain to be identified definitively.

Type I collagen is the most abundant protein of bone matrix. Mutations in the coding regions of the genes for the two type I collagen chains (COL1A1 and COL1A2) result in a severe autosomal dominant pediatric condition known as osteogenesis imperfecta (Sykes 1990). A G → T single nucleotide polymorphism (SNP) at the first base of a consensus binding site for the transcription factor Sp1 in the first intron of COL1A1 was associated not only with BMD in white women (Grant et al. 1996) but also with osteoporotic fractures in postmenopausal women (Langdahl et al. 1998; Uitterlinden et al. 1998). The *COL1A1*T* allele of this polymorphism affects collagen gene regulation in such a manner that it increases the production of the α1(I) collagen chain relative to that of the α2(I) chain and leads to reduced bone strength by a mechanism that is partly independent of bone mass (Mann et al. 2001). These observations thus implicate genetic variants that affect collagen Iα1 metabolism as important determinants of the development of osteoporosis and osteoporotic fractures. Other studies, however, have shown only a weak association of the Sp1 binding site polymorphism with BMD or osteoporotic fractures in premenopausal French women (Garnero et al. 1998) or a lack of association in postmenopausal women in Sweden (Liden et al. 1998), in American women (Hustmyer et al. 1999), or in postmenopausal Danish women (Heegaard et al. 2000).

A −1997G → T SNP in the promoter of COL1A1 was also associated with BMD for the lumbar spine in postmenopausal Spanish women, and this SNP and the G → T SNP of the Sp1 binding site of COL1A1 were shown to be in linkage disequilibrium (Garcia-Giralt et al. 2002). Given the ethnic divergence of gene polymorphisms, it is important to examine polymorphisms potentially related to BMD in each ethnic group. We have now examined whether the −1997G → T SNP of COL1A1 is associated with BMD in Japanese women or men in a population-based study.

Materials and Methods

Study Population. The National Institute for Longevity Sciences Longitudinal Study of Aging (NILS-LSA) is a population-based prospective cohort study

of aging and age-related diseases (Shimokata et al. 2000). The subjects of the NILS-LSA are stratified by both age and sex and are randomly selected from resident registrations in the city of Obu and the town of Higashiura in central Japan (Yamada et al. 2003a, 2003b). Individuals with disorders known to cause abnormalities of bone metabolism, including diabetes mellitus, renal diseases, rheumatoid arthritis, and thyroid, parathyroid, and other endocrinologic diseases, were excluded from the study. Women who had taken drugs such as estrogen, progesterone, glucocorticoids, and bisphosphonates were also excluded.

We examined the relation of BMD at various sites to the $-1997G \rightarrow T$ SNP of COL1A1 in 2,236 participants (1,110 women, 1,126 men). All analyses were performed separately for men and for women. In addition, to uncover potential differences between women according to menopausal status, we conducted all analyses separately for premenopausal and postmenopausal women. Menopausal status was evaluated by a detailed questionnaire, and menopause was defined as complete cessation of menstruation. Furthermore, the relation of biochemical markers of bone turnover to $-1997G \rightarrow T$ genotype of COL1A1 was examined for men or premenopausal or postmenopausal women separately. The study protocol was approved by the Committee on Ethics of Human Research of National Chubu Hospital and the NILS, and written informed consent was obtained from each subject.

Measurement of BMD. BMD for the lumbar spine (L2–L4), right femoral neck, right trochanter, and right Ward's triangle was measured using dual-energy x-ray absorptiometry (DXA) (QDR 4500; Hologic, Bedford, Mass.). The coefficients of variation (CVs) of the DXA instrument were 0.9% (L2–L4), 1.3% (femoral neck), 1.0% (trochanter), and 2.5% (Ward's triangle); these values were determined by measurement of BMD three times at each site in 10 healthy subjects (mean age ± SE, 38.7 ± 2.4 years).

Determination of Genotypes. Genotypes were determined with a fluorescence-based allele-specific DNA primer assay system (Toyobo Gene Analysis, Tsuruga, Japan) (Yamada et al. 2002). The polymorphic region of COL1A1 was amplified using the polymerase chain reaction with allele-specific sense primers labeled at the 5′ end with either fluorescein isothiocyanate (5′-TGGGTCAGTTCCAAGAGXCC-3′) or Texas red (5′-TGGGTCAGTTCCAAGAGXAC-3′) and with an antisense primer labeled at the 5′ end with biotin (5′-TCTAAATGTCTGTTCCCTCCAA-3′). The reaction mixture (25 μL) contained 20 ng of DNA, 5 pmol of each primer, 0.2 mmol/L of each deoxynucleoside triphosphate, 3.5 mmol/L $MgCl_2$, and 1 U of rTaq DNA polymerase (Toyobo, Osaka, Japan) in polymerase buffer. The amplification protocol was initial denaturation at 95°C for 5 min; 35 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30 s, and extension at 68°C for 30 s; and a final extension at 68°C for 2 min.

The amplified DNA was incubated in a solution containing streptavidin-conjugated magnetic beads in the wells of a 96-well plate at room temperature.

11214$ SCH3 01-04-05 13:53:54 PS PAGE 29

The plate was then placed on a magnetic stand, and the supernatants from each well were transferred to the wells of a 96-well plate containing 0.01 mol/L NaOH and measured for fluorescence with a microplate reader (Fluoroscan Ascent; Dainippon Pharmaceutical, Osaka, Japan) at excitation and emission wavelengths of 485 nm and 538 nm, respectively, for fluorescein isothiocyanate and of 584 nm and 612 nm, respectively, for Texas red.

Measurement of Biochemical Markers of Bone Turnover. Venous blood and urine samples were collected in the early morning after the subjects had fasted overnight. Blood samples were centrifuged at 1,600 × *g* for 15 min at 4°C, and the serum fraction was separated and stored at −80°C until analysis. The serum concentration of intact osteocalcin was measured with an immunoradiometric assay kit (Mitsubishi Chemical, Tokyo, Japan). The activity of bone-specific alkaline phosphatase in serum was measured with an enzyme immunoassay kit (Metra Biosystems, Mountain View, Calif.). Urine samples were collected in plain tubes and stored at −80°C. Urinary deoxypyridinoline was measured with an enzyme immunoassay kit (Metra Biosystems); the values were corrected for urinary creatinine and expressed as picomoles per micromole of creatinine. The urinary concentration of cross-linked N-telopeptides of type I collagen (NTx) was measured with an enzyme-linked immunosorbent assay kit (Mochida Pharmaceutical, Tokyo, Japan); the values were expressed as picomoles of bone collagen equivalents per micromole of creatinine. Urinary creatinine was enzymatically measured with a creatinine test kit (Wako Chemical, Osaka, Japan).

Statistical Analysis. Quantitative data were compared among the three groups using one-way analysis of variance and the Tukey-Kramer post hoc test and between two groups using the unpaired Student's *t* test. BMD values were analyzed with adjustment for age and body mass index (BMI) using the least-squares method in a general linear model. The effect of −1997G → T genotype on BMD at various sites was evaluated using multivariate regression analysis; R^2 and *P* values were calculated from the analysis including age, BMI, and COL1A1 genotype (0 = *COL1A1*G/*G*, 1 = *COL1A1*G/*T* = *COL1A1*T/*T*) . Allele frequencies were estimated using the gene-counting method, and the chi-square test was used to identify significant departure from Hardy-Weinberg equilibrium. A *P* value less than 0.05 was considered statistically significant.

Results

The distribution of −1997G → T genotypes was in Hardy-Weinberg equilibrium, and age and BMI did not differ among genotypes for all women (Table 1). BMD for the lumbar spine and trochanter was significantly lower in women with the **G/*G* genotype than in those in the combined group of **G/*T* and

Table 1. BMD and Other Characteristics of All Women (n = 1,110) According to the −1997G → T Genotype of COL1A1

Characteristic	**G/*G*	**G/*T*	**T/*T*	**G/* T + *T/*T*
Number (%)	407 (36.7)	526 (47.4)	177 (15.9)	703 (63.3)
Age (years)	60.0 ± 0.5	58.9 ± 0.5	58.4 ± 0.8	58.8 ± 0.4
BMI (kg/m^2)	22.9 ± 0.2	22.8 ± 0.1	23.0 ± 0.2	22.9 ± 0.1
BMD values (g/cm^2)				
L2–L4	0.855 ± 0.006	0.870 ± 0.006	0.878 ± 0.010	0.872 ± 0.005[a]
Femoral neck	0.672 ± 0.004	0.681 ± 0.004	0.680 ± 0.007	0.681 ± 0.003
Trochanter	0.564 ± 0.004	0.575 ± 0.004	0.574 ± 0.006	0.575 ± 0.003[b]
Ward's triangle	0.500 ± 0.006	0.512 ± 0.005	0.508 ± 0.009	0.511 ± 0.004

Data are means ± SE. BMD values are adjusted for age and BMI.
a. P = 0.039 vs. **G/*G*.
b. P = 0.033 versus **G/*G*.

**T/*T* genotypes; the difference in BMD between the **G/*G* genotype and the combined group of **G/*T* and **T/*T* genotypes (expressed as a percentage of the corresponding larger value) was 1.9% for both the lumbar spine and the trochanter.

We also analyzed BMD and other characteristics for premenopausal and postmenopausal women independently. The distributions of −1997G → T genotypes were in Hardy-Weinberg equilibrium, and age and BMI did not differ among genotypes for premenopausal or postmenopausal women (Table 2). For postmenopausal women there was no difference in years after menopause among genotypes. For premenopausal women BMD was not associated with −1997G → T genotype. In contrast, BMD for the femoral neck or trochanter was significantly lower in postmenopausal women with the **G/*G* genotype than in those with the **G/*T* genotype or those in the combined group of **G/*T* and **T/*T* genotypes; the differences in BMD between the **G/*G* genotype and the combined group of **G/*T* and **T/*T* genotypes were 2.5% for the femoral neck and 2.2% for the trochanter.

The distribution of −1997G → T genotypes was in Hardy-Weinberg equilibrium, but BMD did not differ among these genotypes in men (Table 3).

The effect of −1997G→T genotype on BMD was evaluated using multivariate regression analysis (Table 4). The analysis revealed that the −1997G → T genotype affected BMD at various sites with a variance of 0.46–0.62% for all women and of 0.61–1.01% for postmenopausal women.

The relation of biochemical markers of bone turnover to −1997G → T genotype of COL1A1 was also examined. No association of −1997G → T genotype with the serum concentration of intact osteocalcin, serum activity of bone-specific alkaline phosphatase, or urinary excretion of deoxypyridinoline or NTx was apparent for men or premenopausal or postmenopausal women (Table 5).

11214$ $CH3 01-04-05 13:53:56 PS PAGE 31

Table 2. BMD and Other Characteristics of Women (n = 1,093) According to Menopausal Status and the −1997G → T Genotype of COL1A1

	Premenopausal Women (n = 278)				*Postmenopausal Women* (n = 815)			
Characteristic	**G/*G*	**G/*T*	**T/*T*	**G/*T + *T/*T*	**G/*G*	**G/*T*	**T/*T*	**G/*T + *T/*T*
Number (%)	94 (33.8)	140 (50.4)	44 (15.8)	184 (66.2)	306 (37.5)	377 (46.3)	132 (16.2)	509 (62.5)
Age (years)	45.9 ± 0.4	46.3 ± 0.4	45.8 ± 0.6	46.2 ± 0.3	64.6 ± 0.5	63.8 ± 0.4	62.8 ± 0.7	63.5 ± 0.4
Years after menopause					15.4 ± 0.5	15.1 ± 0.5	13.7 ± 0.8	14.7 ± 0.4
BMI (kg/m^2)	22.8 ± 0.3	22.9 ± 0.3	22.4 ± 0.5	22.8 ± 0.2	23.0 ± 0.2	22.8 ± 0.2	23.2 ± 0.3	22.9 ± 0.1
BMD values (g/cm^2)								
L2–L4	1.018 ± 0.012	1.026 ± 0.010	1.044 ± 0.018	1.030 ± 0.009	0.798 ± 0.007	0.813 ± 0.007	0.821 ± 0.011	0.815 ± 0.006
Femoral neck	0.782 ± 0.009	0.767 ± 0.008	0.771 ± 0.014	0.768 ± 0.007	0.634 ± 0.005	0.650 ± 0.004[a]	0.647 ± 0.007	0.650 ± 0.004[b]
Trochanter	0.656 ± 0.009	0.661 ± 0.007	0.656 ± 0.013	0.660 ± 0.006	0.532 ± 0.005	0.544 ± 0.004	0.545 ± 0.007	0.544 ± 0.004[c]
Ward's triangle	0.663 ± 0.012	0.657 ± 0.010	0.658 ± 0.018	0.657 ± 0.009	0.444 ± 0.007	0.459 ± 0.006	0.455 ± 0.010	0.458 ± 0.005

Data are means ± SE. BMD values are adjusted for age and BMI.

a. P = 0.034 vs. **G/*G*.

b. P = 0.011 vs. **G/*G*.

c. P = 0.033 vs. **G/*G*.

11214$ $CH3 01-04-05 13:53:56 PS PAGE 32

Table 3. BMD and Other Characteristics of Men (n = 1,126) According to the −1997G → T Genotype of COL1A1

Characteristic	**G/*G*	**G/*T*	**T/*T**	*G/*T + *T/*T*
Number (%)	457 (40.6)	511 (45.4)	158 (14.0)	669 (59.4)
Age (years)	58.5 ± 0.5	59.7 ± 0.5	59.2 ± 0.9	59.6 ± 0.4
BMI (kg/m^2)	22.9 ± 0.1	23.0 ± 0.1	22.9 ± 0.2	22.9 ± 0.1
BMD values (g/cm^2)				
L2–L4	0.990 ± 0.007	0.975 ± 0.007	0.983 ± 0.012	0.977 ± 0.006
Femoral neck	0.754 ± 0.005	0.754 ± 0.004	0.744 ± 0.008	0.751 ± 0.004
Trochanter	0.672 ± 0.005	0.665 ± 0.004	0.667 ± 0.008	0.665 ± 0.004
Ward's triangle	0.557 ± 0.006	0.552 ± 0.005	0.540 ± 0.010	0.549 ± 0.005

Data are means ± SE. BMD values are adjusted for age and BMI.

Discussion

The −1997G → T SNP of the COL1A1 promoter has previously been associated with BMD for the lumbar spine and, to a lesser extent, with BMD for the femoral neck in postmenopausal Spanish women, and with the **T/*T* genotype, which represents a risk factor for reduced BMD (Garcia-Giralt et al. 2002). We have now shown that the −1997G → T SNP is associated with BMD for the femoral neck and trochanter in postmenopausal Japanese women and with the **G/*G* genotype, which represents a risk factor for reduced BMD. The −1997G → T genotype affected BMD at various sites with a variance of 0.61–1.01% for postmenopausal women, although this SNP was not associated with biochemical markers of bone turnover.

The alleles of the −1997G → T polymorphism associated with reduced

Table 4. Effects of the −1997G → T Genotype of COL1A1 on BMD for All Women (n = 1,110) or Postmenopausal Women (n = 815)

Site	R^2	P
All women		
L2–L4	0.0061	0.0102
Femoral neck	0.0047	0.0243
Trochanter	0.0062	0.0093
Ward's triangle	0.0046	0.0262
Postmenopausal women		
L2–L4	0.0061	0.0263
Femoral neck	0.0101	0.0044
Trochanter	0.0076	0.0137
Ward's triangle	0.0071	0.0172

The R^2 and P values were derived from multivariate regression analysis including age, BMI, and COL1A1 genotype (0 = **G/*G*, 1 = **G/*T* = **T/*T*).

11214$ $CH3 01-04-05 13:53:56 PS PAGE 33

Table 5. Biochemical Markers of Bone Turnover for Women or Men According to the −1997G → T Genotype of COL1A1

Marker	**G/*G*	**G/*T*	**T/*T*
Premenopausal women			
Osteocalcin (ng/mL)	6.35 ± 0.29	6.46 ± 0.24	6.93 ± 0.42
Bone-specific alkaline phosphatase (U/L)	19.6 ± 0.5	20.3 ± 0.5	19.0 ± 0.8
dPyr (pmol/μmol Cr)	5.54 ± 0.15	5.35 ± 0.12	5.50 ± 0.22
NTx (pmol BCE/μmol Cr)	33.5 ± 1.5	33.6 ± 1.3	38.3 ± 2.3
Postmenopausal women			
Osteocalcin (ng/mL)	10.53 ± 0.21	10.30 ± 0.19	10.25 ± 0.32
Bone-specific alkaline phosphatase (U/L)	31.6 ± 0.6	31.5 ± 0.6	30.5 ± 0.9
dPyr (pmol/μmol Cr)	4.08 ± 0.06	4.01 ± 0.05	3.99 ± 0.10
NTx (pmol BCE/μmol Cr)	60.4 ± 1.6	60.2 ± 1.5	59.4 ± 2.5
Men			
Osteocalcin (ng/mL)	7.67 ± 0.11	7.64 ± 0.11	7.64 ± 0.20
Bone-specific alkaline phosphatase (U/L)	26.3 ± 0.4	25.6 ± 0.4	26.2 ± 0.7
dPyr (pmol/μmol Cr)	4.08 ± 0.06	4.01 ± 0.05	3.99 ± 0.10
NTx (pmol BCE/μmol Cr)	36.6 ± 0.7	36.2 ± 0.7	36.4 ± 1.2

Data are means ± SE. dPyr, deoxypyridinoline; Cr, creatinine; NTx, cross-linked N-telopeptides of type I collagen; BCE, bone collagen equivalents.

BMD thus differ between the present study (**G* allele) and the previous study (**T* allele) (Garcia-Giralt et al. 2002). Although the reason for this apparent discrepancy is unclear, there are three major differences between the two studies. First, the subjects were older in our study (mean age of 64 years for postmenopausal women) than in the previous study (mean age, 51 years), and years since menopause were significantly greater in our study (mean, 15.0 years) than in the previous study (mean, 3.6 years). Given that bone resorption markedly increases during 10 years after menopause, genetic effects on BMD might differ between women for short and long time after menopause. Second, the number of subjects in which the association was detected was greater in our study ($n = 815$ for postmenopausal women) than in the previous study ($n = 256$). The results of association studies with small sample sizes are prone to bias compared with those with large sample sizes. Finally, the distribution of −1997G → T genotypes differed significantly ($P < 0.0001$; chi-square test) between our study (postmenopausal women: **G/*G*, 38%; **G/*T*, 46%; **T/*T*, 16%) and the previous study (**G/*G*, 76%; **G/*T*, 22%; **T/*T*, 2%), possibly reflecting the difference in ethnicity. The difference in genetic influences on BMD between different ethnic groups might be attributable, at least in part, to the difference in the distribution of genotypes. It is also possible that the −1997G → T SNP of COL1A1 is in linkage disequilibrium with other polymorphisms of COL1A1 or with polymorphisms of other nearby genes that are actually responsible for the observed association with BMD. Given the multiple comparisons of genotype performed, we cannot completely exclude the possible occurrence of statistical errors such as

false positives, although we observed a significant association of this SNP with BMD at different sites.

Evidence suggests that the −1997G → T SNP of COL1A1 may affect promoter function (Garcia-Giralt et al. 2002). A double-stranded oligonucleotide containing the −1997G → T site bound osteoblast nuclear factors; however, the extent of factor binding was even more pronounced with a single-stranded antisense DNA probe, suggesting the involvement of a protein selective for single-stranded DNA. The extent of factor binding observed with a probe corresponding to the **G* allele was greater than that apparent with a probe based on the **T* allele. The effect of this SNP on COL1A1 transcription, however, remains to be determined.

In conclusion, our present results suggest that the −1997G → T SNP of COL1A1 is associated with BMD for the femoral neck and trochanter in postmenopausal Japanese women and that the alleles associated with reduced BMD differ between postmenopausal Japanese (**G* allele) and Spanish (**T* allele) women, although the contribution of this SNP to bone mass appears relatively small.

Acknowledgments This study was supported in part by a Research Grant for Longevity Sciences (grant 12C-01 awarded to Y. Yamada and H. Shimokata) and by a Health and Labor Sciences Research Grant for Comprehensive Research on Aging and Health (grant H15-Chojyu-014 awarded to Y. Yamada, F. Ando, N. Niino, and H. Shimokata) from the Ministry of Health, Labor, and Welfare of Japan.

Received 22 March 2004; revision received 16 July 2004.

Literature Cited

Devoto, M., K. Shimoya, J. Caminis et al. 1998. First-stage autosomal genome screen in extended pedigrees suggests genes predisposing to low bone mineral density on chromosomes 1p, 2p, and 4p. *Eur. J. Hum. Genet.* 6:151–157.

Garcia-Giralt, N., X. Nogués, A. Enjuanes et al. 2002. Two new single-nucleotide polymorphisms in the COL1A1 upstream regulatory region and their relationship to bone mineral density. *J. Bone Miner. Res.* 17:384–393.

Garnero, P., O. Borel, S. F. A. Grant et al. 1998. Collagen Iα1 Sp1 polymorphism, bone mass, and bone turnover in healthy French premenopausal women: The OFELY Study. *J. Bone Miner. Res.* 13:813–817.

Grant, S. F. A., D. M. Reid, G. Blake et al. 1996. Reduced bone density and osteoporosis associated with a polymorphic Sp1 binding site in the collagen type Iα1 gene. *Nat. Genet.* 14:203–205.

Heegaard, A., H. L. Jorgensen, A. W. Vestergaard et al. 2000. Lack of influence of collagen type Iα1 Sp1 binding site polymorphism on the rate of bone loss in a cohort of postmenopausal Danish women followed for 18 years. *Calcif. Tissue Int.* 66:409–413.

Hustmyer, F. G., G. Liu, C. C. Johnston et al. 1999. Polymorphism at an Sp1 binding site of COL1A1 and bone mineral density in premenopausal female twins and elderly fracture patients. *Osteoporosis Int.* 9:346–350.