both loci. Thus, it is difficult to distinguish the families based on each locus. However, after identification of *DJ-1* mutations, it became possible to separate PARK6 from PARK7 using *DJ-1* mutation studies.

In the PARK6 region, several genes have been mapped so far. The genes for USP31 (ubiquitin specific protease 31), HTR6 (serotonin receptor), and ECE1 (neprilysin activity) are located in this locus. Considering the pathogenesis of PD, these genes may be candidate genes for PARK6. USP31 is related to the proteasome system that may be involved in the pathogenesis of PD. HTR6 gene is highly expressed in the striatum and C267T polymorphism in this gene was associated with increased risk for the development of PD.[8] Reduced neprilysin activity may cause Alzheimer disease[9] and the similar mechanism may be related to PD.

These potential candidate genes and numerous other genes located in PARK6 region should be investigated in PARK6-linked families.

Acknowledgment

The authors thank the patients and their families.

References

1. Kitada T, Asakawa S, Hattori N, et al. Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. Nature 1998;392:605–608.
2. Lücking CB, Dürr A, Bonifati V, et al. Association between early-onset Parkinson's disease and mutations in the *parkin* gene. N Engl J Med 2000;342:1560–1567.
3. Bonifati V, Rizzu P, van Baren MJ, et al. Mutations in the *DJ-1* gene associated with autosomal recessive early-onset parkinsonism. Science 2003;299:256–259.
4. Hedrich K, Djarmati A, Schäfer N, et al. *DJ-1* (*PARK7*) mutations are less frequent than *Parkin* (*PARK2*) mutations in early-onset Parkinson disease. Neurology 2004;62:389–394.
5. Valente EM, Bentivoglio AR, Dixon PH, et al. Localization of a novel locus for autosomal recessive early-onset parkinsonism, PARK6, on human chromosome 1p35–36. Am J Hum Genet 2001;68:895–900.
6. Valente EM, Brancati F, Ferraris A, et al. PARK6-linked parkinsonism occurs in several European families. Ann Neurol 2002;51:14–18.
7. Bentivoglio AR, Cortelli P, Valente EM, et al. Phenotypic characterisation of autosomal recessive PARK6-linked parkinsonism in three unrelated Italian families. Mov Disord 2001;16:999–1006.
8. Messina D, Annesi G, Serra P, et al. Association of the 5-HT6 receptor gene polymorphism C267T with Parkinson's disease. Neurology 2002;58:828–829.
9. Iwata N, Tsubuki S, Takaki Y, et al. Metabolic regulation of brain Aβ by neprilysin. Science 2001;292:1550–1552.

Novel *PINK1* Mutations in Early-Onset Parkinsonism

Yasuko Hatano, MD,[1] Yuanzhe Li, MD,[1]
Kenichi Sato, MD, PhD,[1] Shuichi Asakawa, PhD,[2]
Yasuhiro Yamamura, MD,[3] Hiroyuki Tomiyama, MD,[1]
Hiroyo Yoshino, BS,[1] Masato Asahina, MD,[4]
Susumu Kobayashi, MD,[5] Sharon Hassin-Baer, MD,[6]
Chin-Song Lu, MD,[7] Arlene R. Ng, MD,[8]
Raymond L. Rosales, MD, PhD,[9]
Nobuyoshi Shimizu, PhD,[2] Tatsushi Toda, MD, PhD,[10,11]
Yoshikuni Mizuno, MD,[1] and
Nobutaka Hattori, MD, PhD[1,11]

***PINK1* was recently found to be associated with PARK6 as the causative gene. We performed mutation analysis in eight inbred families whose haplotypes link to the PARK6 region. We identified six pathogenic mutations (R246X, H271Q, E417G, L347P, and Q239X/R492X) in six unrelated families. All sites of mutations were novel, suggesting that *PINK1* may be the second most common causative gene next to *parkin* in parkinsonism with the recessive mode of inheritance.**

Ann Neurol 2004;56:424–427

The primary cause of Parkinson's disease (PD) is still unknown despite recent progress in research on the molecular mechanism of loss of dopaminergic neurons. Although most patients with PD are sporadic, identification of causative genes of the rare monogenic forms of PD or parkinsonism could provide important insights into the understanding of disease pathogenesis. To date, four genes have been identified as the causative genes for familial parkinsonism: mutations of *α-synuclein* and *UCH-L1* in autosomal dominant forms of parkinsonism and mutations of *parkin* and *DJ-1* in autosomal recessive forms. Among the monogenic forms of parkinsonism, mutations of *parkin* have been detected in approximately 50% of cases with autosomal recessive early-onset parkinsonism (AREP).[1] Although *DJ-1* mutations responsible for PARK7 were reported to cause another type of AREP,[2] it is unlikely to be of numerical significance in clinical practice.[3] Thus, it is possible that other loci are responsible in the remaining patients with AREP.

Recently, mutations of *PINK1* were detected as the causative gene for PARK6.[4] We also performed linkage analysis in 39 families with AREP who were negative for *parkin* and *DJ-1* mutations. Eight of these families showed evidence of linkage with PARK6. *PINK1* is located only 324kb from the D1S2732 at which we obtained multipoint log of the odds score of 9.88.[5] To define the genotype–phenotype relationship, we performed mutation analysis for *PINK1* in these families.

Patients and Methods

Eight families were chosen for *PINK1* mutation screening. Three families were Japanese; two Taiwanese; and one each from Israel, Turkey, and the Philippines. Families A, B, C, D, and E showed homozygosity at the PARK6 region, whereas compound heterozygosity was suggested in Families F, G, and H in our linkage analysis.[5] The clinical characteristics of affected subjects are described in the previous study (mean age at onset ±SD, 30 ± 10.7 years; range, 18–33 years).[5] The study was approved by the ethics review committee of Juntendo University. After obtaining informed consent, we performed mutation analysis of *PINK1* by direct sequencing of the polymerase chain reaction products using the following primers: Ex2 forward 5′-CTGACCTCTCAGATCATTGAGTATTGT-3′, Ex2 reverse 5′- AATCTGTCTTTTCCTACCTACTTCCTG-3′, Ex3 forward 5′-GTTAAGACAGGTCATCTT-ATCTCGAAG-3′, Ex3 reverse 5′-CTACTGTCATA-TCAGACACTGTACCAGG-3′, Ex4 forward 5′-GTACAGTACCTGGCACATAGCAAATCTA-3′, Ex4 reverse 5′- CACTATAGCAAAGTTAGGGGATACA-GAG-3′, Ex5 forward 5′- CTCTTACTTCCTAATT-TGAGGATGGTG-3′, Ex5 reverse 5′- ACTTAGAACA-CAAAACCAGAGAGGAC-3′, Ex6 forward 5′- AAAT-CAAAGTCTCCTGGGGTATAAG-3′, Ex6 reverse 5′- GTTTATGTGACAGGACTTGCATTCT-3′, Ex7 forward 5′- AGAATGCAAGTCCTGTCACATAAAC-3′, Ex7 reverse 5′- GTAACTAGCCTTTACCTTCCTAACACAG-3′, Ex8 forward 5′- ATAGAGGAGACTACTTACCT-GGTTCAAG-3′, and Ex8 reverse 5′- AGACTGAACTCT-CACTCAAGTTCTTCC. Primers for exon 1 were used as reported previously.[4] Dideoxy cycle sequencing was performed with Big Dye Terminator Chemistry (Applied Biosystems, Foster City, CA). This was followed by exon sequencing on ABI377 and 310 automated DNA sequence analyzers (Applied Biosystems). Although the haplotypes of

From the [1]Department of Neurology, Juntendo University School of Medicine; [2]Department of Molecular Biology, Keio University School of Medicine, Tokyo; [3]Institute of Health Science, Hiroshima University School of Medicine, Hiroshima; [4]Department of Neurology, Chiba University Graduate School of Medicine, Chiba; [5]Department of Neurology, Kitano Hospital, The Tazuke Kofukai Medical Research Institute, Osaka, Japan; [6]Parkinson's Disease and Movement Disorders Clinic, Department of Neurology, Chaim Sheba Medical Centre, Tel Hashomer, Israel; [7]Movement Disorder Unit, First Department of Neurology, Chang Gung Memorial Hospital, Taipei, Taiwan; [8]Third Department of Internal Medicine, Kagoshima University School of Medicine, Kagoshima, Japan; [9]Department of Neurology and Psychiatry, University of Santo Tomas Faculty of Medicine and Surgery; Manila, Philippines; [10]Division of Functional Genomics, Osaka University Graduate School of Medicine, Suita; and [11]CREST, Japan Science and Technology Corporation, Kawaguchi, Saitama, Japan.

Received Jun 8, 2004, and in revised form Jul 14. Accepted for publication Jul 23, 2004.

Published online Aug 31, 2004, in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/ana.20251

Address correspondence to Dr Hattori, Department of Neurology, Juntendo University School of Medicine, 2-1-1 Hongo, Bunkyo, Tokyo 113-0033, Japan. E-mail: nhattori@med.juntendo.ac.jp

the affected members of Family F showed compound heterozygotes, we identified a homozygous point mutation in exon 5. Considering this finding, it is possible that an exonic deletion in the same exon takes place in other alleles of the affected members of this family. Therefore, we performed gene dosage assay in Family F to exclude this possibility using TaqMan real-time quantitative polymerase chain reaction. Primers and probes were designed by Assay-by-Design Service (Applied Biosystems). Sequences of primers and probes and the protocols are available upon request.

Results

We identified four types of homozygous point mutations (R246X, H271Q, E417G, and L347P) involving exons 3, 4, 5, and 6 in *PINK1* of patients from five unrelated families (Fig 1). We also detected two nonsense mutations (Q239X and R492X) as a compound heterozygote in a Taiwanese family (Family G) (Table). All mutations cosegregated with the disease phenotype. In addition, the mutations were not found in 200 normal Japanese chromosomes.

The site of nonsense mutation (c.736 C-to-T transition) was not identical to that reported recently,[4] suggesting a novel mutation site. In addition, although the same mutation was detected in different ethnic groups (one in a Japanese and the other in an Israeli), these families did not share a common haplotype, thus excluding the possibility of a single founder effect. This finding indicates that the point mutation (R246X) may be a hot spot in *PINK1* mutations. Premature termination by this mutation could lead to a truncated protein that lacks 336 amino acids, including a highly conserved protein kinase domain. Two Japanese and one Filipino families carried missense mutations (c.813 C-to-A transversion, c.1040 T-to-C transition, and c.1250 A-to-G transition) in exons 4, 5, and 6 resulting in the substitution of highly conserved amino acids in the putative kinase domain, suggesting that this domain is of functional importance (Fig 2).

Although the affected members of Family F were compound heterozygous, we identified a homozygous missense mutation (c.1040 T-to-C). This finding suggests that the affected members of this family are compound heterozygotes with both a missense and an exonic deletion in the same exon 5. However, we could exclude this possibility because we could not detect the heterozygous exonic deletion in exon 5 using the gene dosage technique. Thus, we conclude that the affected members of this family had a homozygous mutation. For this mutation, we could not exclude the possibility

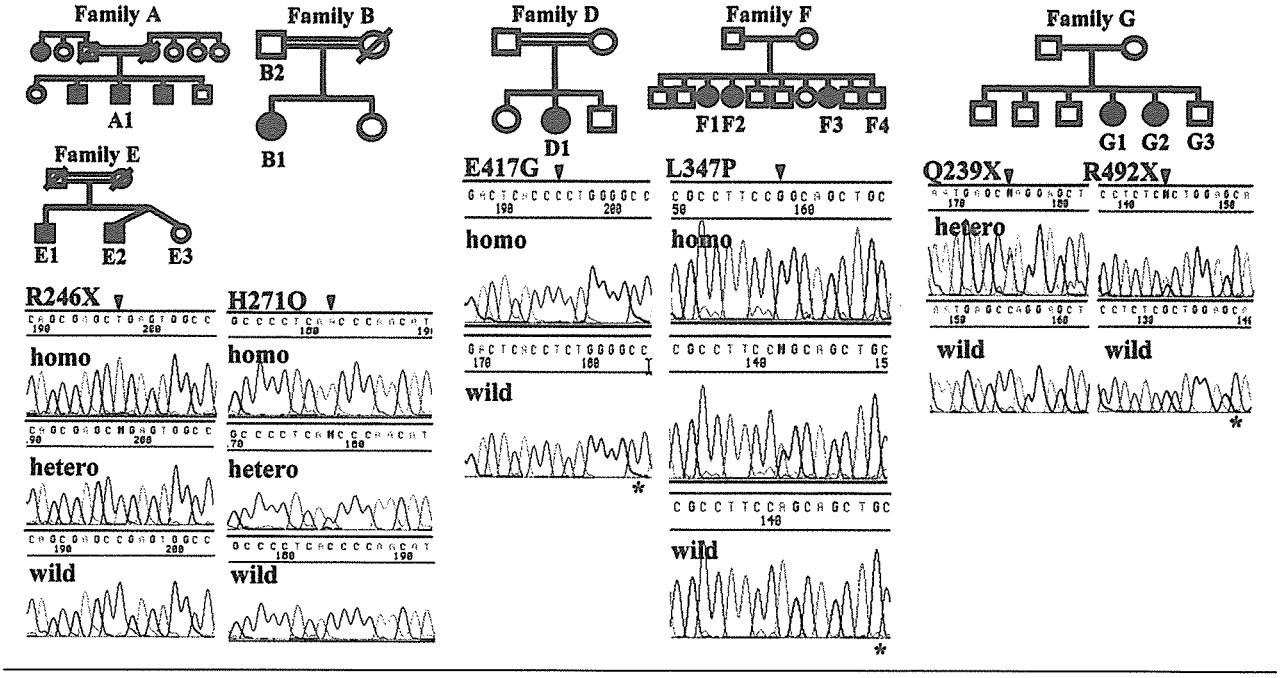

Fig 1. Pedigree and chromatograms illustrating nonsense, missense, and compound heterozygote mutations. Homozygous nonsense mutations (R246X) in exon 3 of affected members (A1, E1, and E2). Homozygous missense mutations (H271Q, E417G, and L347P) in exons 4, 5, and 6 from Families B, D, and F. Compound heterozygote mutation (Q239X/R492X) in exons 3 and 7 of affected members (G1 and G2). Heterozygote states were identified in healthy individuals in Families E (E3), B (B2), and F (F4). One of the unaffected members in Family G (G3) had only a heterozygote mutation (Q239X). (circles) *Women;* (squares) *men;* (solid symbols) *homozygous affected individuals;* (open symbols) *healthy individuals.* (asterisk) *Complementary sequences are presented in exons 5, 6, and 7.*

Table. Mutations in the PINK1 *gene*

	Origin	Nucleotide change	amino acid change	Exon	Zygosity	Mutation type	AAO	DD
Family A	Japan	c.736 C-to-T	R246X	3	homo	nonsense	30	17
Family B	Japan	c.813 C-to-A	H271Q	4	homo	missense	23	15
Family D	Japan	c.1250 A-to-G	E417G	6	homo	missense	33	8
Family E	Israel	c.736 C-to-T	R246X	3	homo	nonsense	25,33	17,21
Family F	Philippines	c.1040 T-to-C	L347P	5	homo	missense	27,27,32	18,19,23
Family G	Taiwan	c.715 C-to-T/ c.1474 C-to-T	Q239X/R492X	3/7	com/hetero	nonsense/ nonsense	18,19	22,24

homo = homozygous; com/hetero = compound heterozygote; AAO = age at onset (years); DD = disease duration (years)

that this alteration is a rare polymorphism because we could not screen for the mutation among the same races such as normal Filipino controls. However, we consider this mutation to be pathogenic because of the significance linkage to PARK6 of this family,[5] absence of its mutation in 100 normal Japanese controls, and the alteration of highly conserved amino acid among several species.

Several polymorphic variants were identified in normal Japanese controls. In exon 5, a homozygous c.1018G→A substitution (frequency: 10%, n = 100) and a heterozygous c.1018G→A substitution (frequency: 44%, n = 100) were found. Another variant, a C→T homozygous substitution (c.914C→T, P305L), was found in all Japanese controls (frequency: 100%, n = 100) and IVS4-5 G→A was found as homozygous (frequency: 68%, n = 100) and heterozygous (frequency: 29%, n = 100).

All patients with *PINK1* mutations showed early age at onset (mean age at onset ±SD, 26.7±5.9 years; range, 18–33 years), long disease duration (mean, 18.4 ±4.67 years), and good response to L-dopa.[5] There were no distinct clinical signs that could distinguish patients of homozygous mutation from those with compound heterozygous mutation.

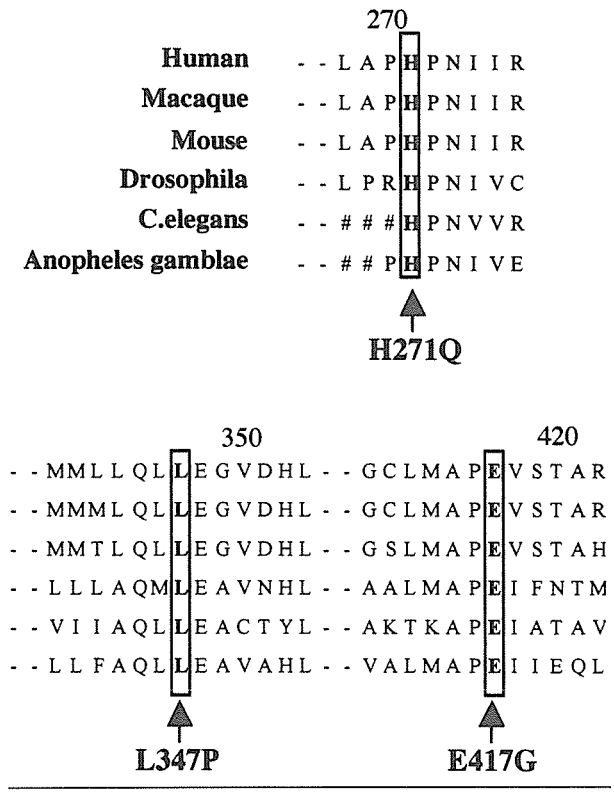

Fig 2. Alignment of PINK1 *homologs showing the conserved amino acid mutated in Families B, D, and F.*

Discussion

Our results indicate that pathogenic mutations in *PINK1*-positive AREP are not limited to Europeans but occur also in Asians, suggesting that *PINK1* mutation is the second most frequent next to *parkin*. Different point mutations seem to be more frequently responsible for the disease phenotype than are deletions.

A homozygous mutation (L347P) was detected in the affected members of the Filipino family of which haplotypes at the PARK6 region showed compound heterozygotes, indicating that the frequency of *PINK1* mutations could be high next to the *parkin* mutations.[5]

In this study, we could not identify the *PINK1* mutation in the protein coding regions including the splicing sites in a Turkish (Family C) and the other Taiwanese families (Family H). Although we cannot exclude the possibility that the patients may have homozygous mutation in the regulatory regions or intron sequences that cause exon skipping, these families may be linked to other loci. Indeed, homozygosity in the *PINK1*-negative families spanned the PARK6 and 9 regions. Thus, these families may have an allelic disorder in the PARK9 gene because the clinical phenotype of PARK9 is a distinct entity from PARK6. We found, based on the comparison between the *PINK1*-positive and -negative families, that the clinical features are very

similar. It is difficult to distinguish *PINK1*-positive AREP from the *PINK1*-negative one. In this regard, the discovery of *PINK1* helps us to provide key clinical information based on the differential diagnosis of AREP. The characteristic clinical features of our *PINK1*-positive families included slow progression and lack of dystonia at onset except for two patients (D1 and G2), indicating similarity to the Italian families described in the original report.[4] Although further studies are needed to determine the frequency of dystonia in *PINK1*-positive AREP, the lack of dystonia might be a distinct clinical sign for differentiating this form from *parkin*- or *DJ-1*–positive AREP. Furthermore, two affected members of Family E showed some psychiatric problems at the onset of the disease. The disease onset is slightly earlier than in patients of the original report, indicating phenotypic variability.

Although PINK1 function is unknown, it originally was reported to be upregulated by the tumor suppressor gene, *PTEN*, in cancer cells.[6] Preliminary results showed that the loss-of-function effect of PINK1 might be associated with mitochondrial dysfunction. There has been considerable progress in our understanding of the molecular mechanisms of nigral degeneration; mitochondrial respiratory failure and oxidative stress appear to play important roles in the progression of the disease.[7] DJ-1 acts as an antioxidant protein, and oxidative stress can damage the 26S proteasome in which parkin acts as an ubiquitin ligase. Thus, all the gene products in AREP may form a common cascade. In summary, the novel mutations identified in this study indicate that *PINK1* is a pathogenic gene in AREP.

This study was supported by the Ministry of Education, Science, Sports, and Culture of Japan, by the Fund for "Research for the Future" Program from the Japan Society for the Promotion of Science.

We are grateful to the patients and their families. We thank J. Fukae, T. Shimazaki, and A. Shimizu for their assistance with the mutation analysis; H. Shinotoh, B. Elibol, F. Belgin ATAç, H. C. Chang, Y. H. Wu-Chou, and Y. Shinar for obtaining the clinical samples; and E. M. Valente and coworkers for exchanging manuscripts prior to publication.

References

1. Lücking CB, Dürr A, Bonifati V, et al. Association between early-onset Parkinson's disease and mutations in the *parkin* gene. N Engl J Med 2000;342:1560–1567.
2. Bonifati V, Rizzu P, van Baren M, et al. Mutations in the *DJ-1* gene associated with autosomal recessive early-onset parkinsonism. Science 2003;299:256–259.
3. Hedrich K, Djarmati A, Schäfer N, et al. *DJ-1 (PARK7)* mutations are less frequent than *Parkin (PARK2)* mutations in early-onset Parkinson disease. Neurology 2004;62:389–394.
4. Valente EM, Abou-Sleiman PM, Caputo V, et al. Hereditary early-onset Parkinson's disease caused by mutations in *PINK1*. Science 2004;304:1158–1160.
5. Hatano Y, Sato K, Elibol B, et al. PARK6-linked autosomal recessive early-onset parkinsonism in Asian populations. Neurology (in press).
6. Unoki M, Nakamura Y. Growth-suppressive effects of *BPOZ* and *EGR2*, two genes involved in the *PTEN* signaling pathway. Oncogene 2001;20:4457–4465.
7. Jenner P. Oxidative stress in Parkinson's disease. Ann Neurol 2003;53:S26–S38.

Published by Wiley-Liss, Inc., through Wiley Subscription Services

THE LANCET
32 Jamestown Road London NW1 7BY UK

Pathogenetic mechanisms of *parkin* in Parkinson's disease

Nobutaka Hattori
Yoshikuni Mizuno

Reprinted from THE LANCET
21 August 2004 Vol. 364
No. 9435 Pages 722-724

Pathogenetic mechanisms of *parkin* in Parkinson's disease

Lancet 2004; 364: 722–24

Nobutaka Hattori, Yoshikuni Mizuno

Department of Neurology, Juntendo University School of Medicine, Tokyo 113-8421, Japan (N Hattori MD, Prof Y Mizuno MD)

Correspondence to: Prof Nobutaka Hattori nhattori@med.juntendo.ac.jp

Context The cause and pathogenesis of Parkinson's disease remain unknown; mitochondrial dysfunction, oxidative damage, environmental factors, and genetic predisposition might all be involved. Identification of the causative genes for familial Parkinson's diseases allow study of the pathogenesis of the disease at the molecular level.

Starting point Katja Hedrich and colleagues studied 75 Serbian patients with early-onset Parkinson's disease for *DJ-1* mutations (*Neurology* 2004; 62: 389–94). One patient was a compound heterozygote and another had a heterozygous exon deletion. *DJ-1* mutations seem to be rare in this European population. By contrast, *parkin* mutations have been found in about 50% of familial cases and in 10–20% of cases without a positive family history.

Where next The fact that parkin is a ubiquitin ligase gives special meaning to the molecular mechanism of neurodegeneration in general. In Parkinson's disease, Lewy bodies are immunoreactive for ubiquitin. Accumulation of abnormal proteins has also been seen in other neurodegenerative disorders. Disturbance of protein degradation by the ubiquitin-proteasome system might have a critical role in neurodegeneration. Although α-synuclein mutations are infrequent, α-synuclein accumulates in Lewy bodies, and α-synuclein fibrils impair the 26S proteasome function. UCH-L1 is also an abundant deubiquitylating enzyme, and its mutation is linked to *PARK5*. Furthermore, *DJ-1* might interact with SUMO-1 (small ubiquitin-like modifier), which can counteract ubiquitin and stabilise proteins against degradation by the 26S proteasome. Uncovering the mechanisms of protein degradation should add importantly to understanding the neurodegenerative process in these neurodegenerative diseases.

Parkinson's disease is characterised pathologically by selective degeneration of neurons in the substantia nigra and locus coeruleus, and the presence of Lewy bodies in the remaining neurons. Although the cause of the disease remains unknown, the identification of genetic mutations in several forms of familial Parkinson's disease has provided new insights into the mechanism of neuronal loss in the substantia nigra.

In familial disease, *parkin* mutations are the most common cause of autosomal recessive early-onset parkinsonism, including the autosomal recessive juvenile disease. The frequency of the mutation is estimated at 50% in families with autosomal recessive early-onset parkinsonism.[1] The clinical features of the autosomal recessive early-onset disease with *parkin* mutations are highly variable compared with the juvenile disease. Autopsies on patients with parkin-related diseases, except for one case, commonly show lack of Lewy bodies,[2] suggesting that the normal function of parkin is essential for the formation of Lewy bodies. Also, the discovery that parkin is an ubiquitin ligase suggests that the ubiquitin-proteasome system might have an important role in maintaining dopaminergic neurons.[3]

Mutations in *parkin*

Autosomal recessive juvenile parkinsonism is characterised by early-onset parkinsonism (before 40 years of age, average 26·1 years), mild dystonia, diurnal fluctuation, transient improvement of motor disability after sleep or rest, a good response to levodopa, and less frequent resting tremor compared with sporadic Parkinson's disease. With the recognition of the various *parkin* mutations in autosomal recessive early-onset parkinsonism, the clinical features of parkin-related diseases can be similar to those of late-onset sporadic Parkinson's disease, including the age of onset and cardinal features such as postural instability, resting tremor, rigidity, and akinesia.

There are 12 exons spanning 1·4 Mb in *parkin*, encoding a 465-aminoacid protein with moderate homology to ubiquitin at its aminoterminus (ubiquitin-like domain, Ubl) and two RING finger motifs (RINGs) at the carboxy-terminus. Various *parkin* mutations have been identified, including exonic deletion, insertions, and several missense mutations.[4–7] Mutations in *parkin* are not limited to a particular area or race. In fact, early-onset parkinsonism with *parkin* mutations is the most frequent form of familial Parkinson's disease. In addition to homozygous mutations, compound heterozygous states, representing different mutations in each allele, are also not uncommon in young patients with apparently sporadic disease. Compound heterozygotes are difficult to detect by conventional PCR because of the large size of this gene, but can be quantitatively identified by the gene-dosage technique. Point mutations are infrequent in Japanese patients compared with whites, although exonic deletions are common. Katja Hedrich and colleagues[8] studied 75 Serbian patients with early-onset Parkinson's disease for *DJ-1* mutations. One patient was a compound heterozygote and another had a heterozygous exon deletion. Thus *DJ-1* mutations seem to be rare in this European population compared with *parkin* mutations.

The common hot spots for exonic deletions are from exons 2 to 5. By contrast, point mutations have been found from exons 6 to 12, where the two RING finger motifs and the in-between RINGs are located. The clinical phenotypes of parkin-related diseases are expanding. For instance, slowly progressive cerebellar ataxia has been reported.[7] Also, more than a few patients who had psych-

iatric or behavioural symptoms were found to have *parkin* mutations. These atypical clinical symptoms appear before or after the onset of parkinsonism.[7]

Carriers of *parkin* mutations may have a Parkinson's disease phenotype. Indeed positron-emission tomography with fluorodopa revealed reduction of uptake even in carriers.[9] Furthermore, heterozygous *parkin* mutations that seem to be transmitted dominantly have been identified in multiple generations,[10] which suggests that parkin-related diseases might be sometimes dominantly inherited. Single heterozygous mutations in exon 7 might also act as susceptible alleles for the late-onset form of parkin-related diseases.[11] And the association[12] of *parkin* haploinsufficiency in sporadic Parkinson's disease further implicates the role of parkin in the more common form of the disease. The single heterozygous state might exert haploinsufficiency effects rather than dominant negative effects. By contrast, some mis-sense mutations might have a dominant negative effect as mis-sense mutations in functional domains (Ubl or RINGS),[13] resulting in an earlier onset than with mutations in other regions.

Function of parkin and pathogenesis of parkin-related diseases

The ubiquitin-proteasome pathway is important in protein processing and degradation, and contributes to quality control of proteins in cells. Ubiquitin is attached covalently to target proteins. Protein ubiquitination is catalysed by three enzymes, E1 (ubiquitin-activating enzyme), E2 (ubiquitin-conjugating enzyme), and E3 ubiquitin ligase. Mutations in *parkin* result in a loss-of-function of E3. Subsequently, substrates for parkin accumulate within dopaminergic neurons, potentially leading to nigral neuronal death, and it is important to identify such substrates.

Nine candidate proteins are degraded by parkin (table).[14–22] Other proteins also interact with parkin, such as E2s, multiprotein ubiquitin-ligase complex (eg, cullin-1),[18] CASK/Lin2, which acts as a scaffolding protein containing postsynaptic PDZ,[23] actin filaments,[24] γ-tubulin,[25] and Rpn 10,[26] which is the binding site for proteasomal proteins. Although further research is needed to elucidate the functional role of the RINGs of parkin, the presence of the proteins listed in the table suggests that parkin might interact with various proteins, including substrates.

Animal *parkin*-knockout models can help to elucidate the mechanism of parkin-related diseases. Mice that are *parkin*-null have motor and cognitive deficits and inhibition of amphetamine-induced dopamine release and glutamate neurotransmission.[27] These mice have high concentrations of dopamine in the limbic areas of the brain, and have a shift of dopamine metabolism towards monoamine oxidase. The finding on monoamine oxidase suggests the presence of oxidative stress in parkin-related diseases (figure). Additionally, *parkin*-null mice show accumulation of parkin substrates, but steady-state concentrations of CDCrel-1, synphilin-1, and α-synuclein, which are substrates for parkin, are not altered.[28] A drosophila model with inactivated orthologue of human *parkin* showed muscular degeneration and mitochondrial pathology,[29] which suggests a physiological role for parkin at a mitochondrial level.[30] However, it remains to be seen why muscular degeneration but not neurodegeneration is observed in this model, although it may represent differences between mice and drosophila.

Most importantly, while it is clear that Lewy-body formation generally does not occur in parkin-related diseases, Lewy bodies were identified in one patient with compound heterozygous mutations,[31] which suggests that some *parkin* mutants with mis-sense mutations might

Substrate	Proposed function
CDCrel-1	Exocytosis (dopamine storage?)
CDCrel-2	
Pael receptor	Stress in endoplasmic reticulum (unfolded protein response)
O-glycosylated α-synuclein	Lewy-body formation
Synphilin-1	Lewy-body formation
Cyclin E	Apoptosis (kainate excitoxication)
α/β tubulin	Microtubules (assembly dysfunction)
p38 subunit	aminoacyl-tRNA synthesis (protein biosynthesis)
Synaptotagmin XI	Fusion or docking, synaptic functions
Parkin-interacting proteins	
UbcH7, UbcH8, Ubc6/7, Ubc4	E2
Actin filament	Morphology
CASK/Lin2	PDZ-containing scaffolding protein
Cullin-1	Multiprotein ligase
γ-tubulin	Centrosome
Rpn 10	Binding of parkin to proteasomal proteins

PDZ=postsynaptic density-95, disc large, zona occludens.

Table: **Candidate substrates for parkin**

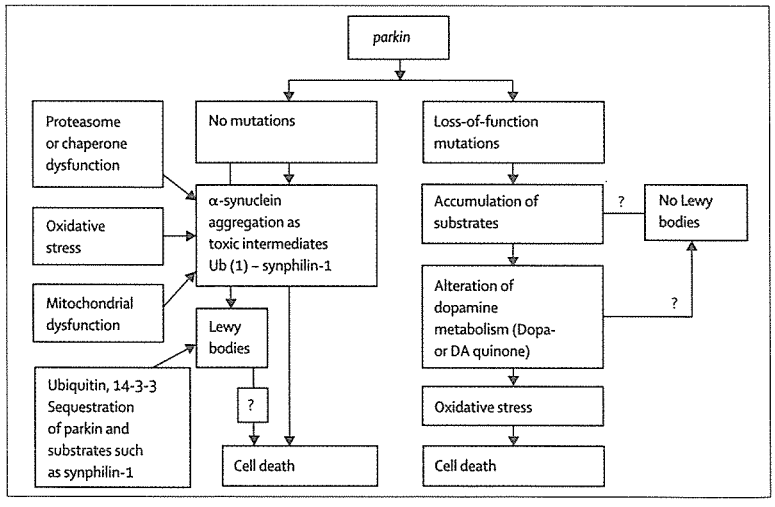

Figure: **Model for pathogenesis of Park 2 and sporadic Parkinson's disease**
Lewy bodies need parkin for thier formation. By contrast, Park2 lacks Lewy-body formation, suggesting that accumulation of substrates inhibits aggregation of α-synuclein (Lewy bodies). In Parkinson's disease, proteasome dysfunction, oxidative stress, and mitochondrial dysfunction could induce aggregation of α-synuclein, with subsequent formation of intermediate filaments of α-synuclein. Finally, Lewy bodies are formed although whether their formation is cytotoxic is debatable. UB=ubiquitin, Dopa=3,4-dihydroxyphenylalanine, DA=dopamine.

have ubiquitin-ligase activities. Although the pathological findings in the brains of patients with parkin-related diseases include severe neuronal loss and gliosis in the substantia nigra, several atypical findings have been reported, such as accumulation of tau protein as neurofibrillary tangles in the substantia nigra, locus coeruleus, red nucleus, and posterior hypothalamus, and neurofibrillary tangles and thorn-shaped astrocytes in frontal, temporal, and parietal cortices.[2] Accumulation of tau protein as tufted astrocytes, but not neurofibrillary tangles, was reported in a patient with compound heterozygous mutations.[32] In this respect, part of the pathology of parkin-related diseases resembles that of progressive supranuclear palsy, which suggests that tau may be one of the parkin-interacting proteins.

To investigate the toxicity of the substrates for parkin, drosophilia is a suitable model for elucidating the mechanism of dopaminergic neuronal loss. Transgenic flies that express human PAEL receptors (PAEL-R), a parkin substrate, under conditions of altered parkin activity show age-dependent selective degeneration of dopaminergic neurons, despite equal levels of expression of PAEL-R in all neurons.[33] This Pael-R-mediated neurotoxicity in dopaminergic neurons was attenuated by co-expression of human parkin, and exacerbated by blocking the activity of endogenous parkin by RNA interference. Overexpression of parkin can also suppress α-synuclein-induced toxicity.[33] However, drosophila has neither PAEL-R nor α-synuclein proteins. Thus the null background of the ectopic expression of both proteins induces neuropathological changes in the dopaminergic system, which is corrected to some extent by co-expression of parkin. These findings suggest that parkin probably has a central role in maintaining dopaminergic neurons. In other words, parkin is an essential protein for the survival of dopaminergic neurons.

We have no conflict of interest to declare.

References

1 Kitada T, Asakawa S, Hattori N, et al. Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. *Nature* 1998; **392:** 605–08.

2 Mori H, Kondo T, Yokochi M, et al. Pathologic and biochemical studies of juvenile parkinsonism linked to chromosome 6q. *Neurology* 1998; **51:** 890–92.

3 Shimura H, Hattori N, Kubo S, et al. Familial Parkinson disease gene product, *parkin*, is a ubiquitin-protein ligase. *Nat Genet* 2000; **25:** 302–05.

4 Hattori N, Matsumine H, Kitada T, et al. Molecular analysis of a novel ubiquitin-like protein (*PARKIN*) gene in Japanese families with AR-JP. *Ann Neurol* 1998; **44:** 935–41.

5 Lücking CB, Durr A, Bonifati V, et al. Association between early-onset Parkinson's disease and mutations in the *parkin* gene. *N Engl J Med* 2000; **342:** 1560–67.

6 Periquet M, Latouche M, Lohmann E, et al. *Parkin* mutations are frequent in patients with isolated early-onset parkinsonism. *Brain* 2003; **126:** 1271–78.

7 Khan NL, Graham E, Critchley P, et al. Parkin disease: a phenotypic study of a large case series. *Brain* 2002; **126:** 1279–92.

8 Hedrich K, Djarmati A, Schafer N, et al. *DJ-1* (*PARK7*) mutations are less frequent than *Parkin* (*PARK2*) mutations in early-onset Parkinson disease. *Neurology* 2004; **62:** 389–94.

9 Hilker R, Klein C, Ghaemi M, et al. Positron emission tomographic analysis of the nigrostriatal dopaminergic system in familial parkinsonism associated with mutations in the *parkin* gene. *Ann Neurol* 2001; **49:** 367–76.

10 Klein C, Pramstaller PP, Kis B, et al. *Parkin* deletions in a family with adult-onset, tremor-dominant parkinsonism: expanding the phenotype. *Ann Neurol* 2000; **48:** 65–71.

11 Oliveira SA, Scott WK, Martin ER, et al. Parkin mutations and susceptibility alleles in late-onset Parkinson's disease. *Ann Neurol* 2003; **53:** 624–29.

12 West AD, Maraganore D, Crook J, et al. Functional association of the *parkin* gene promoter with idiopathic Parkinson's disease. *Hum Mol Genet* 2002; **11:** 2787–92.

13 Lohmann E, Periquet M, Bonifati V, et al. How much phenotypic variation can be attributed to *parkin* genotype? *Ann Neurol* 2003; **54:** 176–85.

14 Zhang Y, Gao J, Chung KK, et al. Parkin functions as an E2 dependent ubiquitin-protein ligase and promotes the degradation of the synaptic vesicle associated protein, CDCrel-1. *Proc Natl Acad Sci USA* 2000; **97:** 13354–59.

15 Imai Y, Soda M, Inoue H, et al. An unfolded putative transmembrane polypeptide, which can lead to endoplasmic reticulum stress, is a substrate of Parkin. *Cell* 2001; **105:** 891–02.

16 Chung KK, Zhnag Y, Lim KL, et al. Parkin ubiquitinates the α-synuclein-interacting protein synphilin-1. *Nat Med* 2001; **7:** 1144–50.

17 Shimura H, Schlossmacher MG, Hattori N, et al. Ubiquitination of a new form of alpha-synuclein by parkin from human brain: implication for Parkinson's disease. *Science* 2001; **293:** 263–69.

18 Staropoli JF, McDermott C, Martinat C, et al. Parkin is a component of an SCF-like ubiquitin ligase complex and protects postmitotic neurons form kainite excitotoxicity. *Neuron* 2003; **37:** 735–49.

19 Ren Y, Zhao J, Feng J. Parkin binds to α-tubulin and increases their ubiquitination and degradation. *J Neurosci* 2003; **23:** 3316–24.

20 Corti O, Hampe C, Koutnikova H, et al. The p38 subunit of the aminoacyl-tRNA synthase complex is a parkin substrate. *Hum Mol Genet* 2003; **12:** 1427–37.

21 Huynh DP, Scoles DR, Neuyen D, et al. The autosomal recessive juvenile Parkinson disease gene product, *parkin*, interacts with and ubiquitinates synaptotagmin XI. *Hum Mol Genet* 2003; **12:** 2587–97.

22 Choi P, Snyder H, Petrucelli L, et al. SEPT5_v2 is a parkin-binding protein. *Brain Res Mol Brain Res* 2003; **117:** 179–89.

23 Fallon L, Moreau F, Croft BG, et al. Parkin and CASK/LIN-2 associate via a PDZ-mediated interaction and are co-localized in lipid rafts and postsynaptic densities in brain *J Biol Chem* 2002; **277:** 486–91.

24 Huynh DP, Scoles DR, Ho TH, et al. Parkin is associated with actin filaments in neuronal and nonneuronal cells. *Ann Neurol* 2000; **48:** 737–44.

25 Zhao J, Ren Y, Jiang Q, Feng J. Parkin is recruited to the centrosome in response to inhibition of proteasomes. *J Cell Sci* 2003; **116:** 4011–19.

26 Sakata E, Yamaguchi Y, Kurimoto E, et al. Parkin binds the Rpn10 subunit of 26S proteasomes through its ubiquitin-like domain. *EMBO Rep* 2003; **4:** 301–06.

27 Itier J-M, Ibanez P, Mena MA, et al. Parkin gene inactivation alters behaviour and dopamine neurotransmission in the mouse. *Hum Mol Genet* 2003; **12:** 2277–91.

28 Goldberg MS, Fleming SM, Palacino JJ, et al. Parkin-deficient mice exhibit nigrostriatal deficits but not loss of dopaminergic neurons. *J Biol Chem* 2003; **278:** 43628–35.

29 Greene JC, Whitworth AJ, Kuo I, et al. Mitochondrial pathology and apoptotic muscle degeneration in *Drosophila parkin* mutants. *Proc Natl Acad Sci USA* 2003; **100:** 4078–83.

30 Darios F, Corti O, Lucking CB, et al. Parkin prevents mitochondrial swelling and cytochrome c release in mitochondria-dependent cell death. *Hum Mol Genet* 2003; **12:** 517–26.

31 Farrer M, Chan P, Chen R, et al. Lewy bodies and parkinsonism in families with *parkin* mutations. *Ann Neurol* **50:** 293–300, 2001.

32 van de Warrenburg BP, Lammens M, Lucking CB, et al. Clinical and pathologic abnormalities in a family with parkinsonism and *parkin* gene mutations. *Neurology* 2001; **56:** 555–57.

33 Yang T, Nishimura I, Imai Y, et al. parkin suppressed dopaminergic neuron-selective neurotoxicity induced by Pael-R in *Drosophila*. *Neuron* 2003; **37:** 911–24.

The Lancet is a weekly subscription journal. For further information on how to subscribe please contact our Subscription Department
Tel: +44 (0) 1865 843077 Fax: +44 (0) 1865 843970 Email: custserv@lancet.com
North America Tel: +1 (800) 462 6198 Fax: +1 (800) 327 9021 Email: USLancetCS@elsevier.com

Printed in UK
by Lynhurst Press Ltd

Mitochondrial Genome Variation in Eastern Asia and the Peopling of Japan

Masashi Tanaka,[1,15] Vicente M. Cabrera,[2] Ana M. González,[2] José M. Larruga,[2] Takeshi Takeyasu,[1,3] Noriyuki Fuku,[1,4] Li-Jun Guo,[1,3] Raita Hirose,[1] Yasunori Fujita,[1] Miyuki Kurata,[1] Ken-ichi Shinoda,[5] Kazuo Umetsu,[6] Yoshiji Yamada,[7,1] Yoshiharu Oshida,[3] Yuzo Sato,[3] Nobutaka Hattori,[8] Yoshikuni Mizuno,[8] Yasumichi Arai,[10] Nobuyoshi Hirose,[10] Shigeo Ohta,[11] Osamu Ogawa,[9] Yasushi Tanaka,[9] Ryuzo Kawamori,[9] Masayo Shamoto-Nagai,[1,4,12] Wakako Maruyama,[12] Hiroshi Shimokata,[13] Ryota Suzuki,[14] and Hidetoshi Shimodaira[14]

[1]Department of Gene Therapy, Gifu International Institute of Biotechnology, Kakamigahara, Gifu 504-0838, Japan; [2]Department of Genetics, Faculty of Biology, University of La Laguna, Tenerife 38271, Spain; [3]Department of Sports Medicine, Graduate School of Medicine, Nagoya University, Nagoya 464-8601, Japan; [4]Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan; [5]Department of Anthropology, National Science Museum, Tokyo 169-0073, Japan; [6]Department of Forensic Medicine, Yamagata University School of Medicine, Yamagata 990-9585, Japan; [7]Department of Human Functional Genomics, Life Science Research Center, Mie University, Tu-shi, Mie 514-8507, Japan; [8]Department of Neurology and [9]Department of Medicine, Metabolism and Endocrinology, Juntendo University School of Medicine, Tokyo 113-8421, Japan; [10]Department of Geriatric Medicine, Keio University School of Medicine, Tokyo 160-8582, Japan; [11]Department of Biochemistry and Cell Biology, Institute of Gerontology, Nihon Medical School, Kawasaki 211-8533, Japan; [12]Laboratory of Biochemistry and Metabolism, Department of Basic Gerontology, and [13]Department of Epidemiology, National Institute for Longevity Sciences, Obu 474-8522, Japan; [14]Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo 152-8552, Japan

To construct an East Asia mitochondrial DNA (mtDNA) phylogeny, we sequenced the complete mitochondrial genomes of 672 Japanese individuals (http://www.giib.or.jp/mtsnp/index_e.html). This allowed us to perform a phylogenetic analysis with a pool of 942 Asiatic sequences. New clades and subclades emerged from the Japanese data. On the basis of this unequivocal phylogeny, we classified 4713 Asian partial mitochondrial sequences, with <10% ambiguity. Applying population and phylogeographic methods, we used these sequences to shed light on the controversial issue of the peopling of Japan. Population-based comparisons confirmed that present-day Japanese have their closest genetic affinity to northern Asian populations, especially to Koreans, which finding is congruent with the proposed Continental gene flow to Japan after the Yayoi period. This phylogeographic approach unraveled a high degree of differentiation in Paleolithic Japanese. Ancient southern and northern migrations were detected based on the existence of basic M and N lineages in Ryukyuans and Ainu. Direct connections with Tibet, parallel to those found for the Y-chromosome, were also apparent. Furthermore, the highest diversity found in Japan for some derived clades suggests that Japan could be included in an area of migratory expansion to Continental Asia. All the theories that have been proposed up to now to explain the peopling of Japan seem insufficient to accommodate fully this complex picture.

[Supplemental material is available online at www.genome.org.]

Recent analysis of global mitochondrial DNA diversity in humans based on complete mtDNA sequences has provided compelling evidence of a human mtDNA origin in Africa (Ingman et al. 2000). Less than 100,000 years ago, at least two mtDNA human lineages began to rapidly spread from Africa to the Old World (Maca-Meyer et al. 2001). The archaeological records attest that humans reached Japan, at the eastern edge of Asia, around 30,000 years ago (Glover 1980). At that time, Japan was connected to the Continent by both northern and southern land bridges, enabling two migratory routes. As early as 13,000 years ago, pottery appeared in Japan and Siberia for the first time in the world (Shiraishi 2002). Subsequent technical improvements gave rise to the Japanese Neolithic period known as the Jomon period, in which the population growth was considerable. Later, Continental people arrived in Japan from the Korean peninsula, initiating the Yayoi period, with this migration reaching its maximum at the beginning of the first millennium.

With this archaeological framework in mind, it was of an-

[15]Corresponding author.
E-MAIL mtanaka@giib.or.jp; FAX 81-583-71-4412.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.2286304.

14:1832–1850 ©2004 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/04; www.genome.org

thropological interest to us to know whether the modern Japanese are the result of an admixture between the Paleolithic–Neolithic aborigines and more recent immigrant populations, whether the indigenous population gradually evolved to give rise to the modern Japanese, with subsequent colonizations having strong cultural influences but only minor demographic impact, or even whether the late Neolithic waves entirely replaced the indigenous residents. Morphometric data obtained from the remains of Japanese Paleolithic people are more in accordance with a southern origin for these first immigrants. Subsequent morphological studies on modern indigenous (northern Ainu and southern Ryukyuans) and mainland Japanese favored an admixture model in which the former would be descendants of the Paleolithic Japanese and the latter derived from the Continental immigrants who gave rise to the Yayoi period (Hanihara 1991). Genetic analysis using classical markers assigned a definitive northern origin to the Upper Paleolithic inhabitants of Japan; but whereas some authors favored a homogeneous background for all modern Japanese (Nei 1995), others claimed that although Upper Paleolithic and Yayoi period immigrants had probably a northern Asian origin, they were genetically differentiated (Omoto and Saitou 1997). The application of molecular markers to define maternal and paternal lineages to the peopling of Japan confirmed the dual admixture model but added some interesting novelties. For example, the study of Y-chromosome markers led to the discovery of remarkable Korean and Tibetan influences on the Japanese population (Hammer and Horai 1995); and mtDNA HVS-I sequences also confirmed the Korean input (Horai et al. 1996) and closer affinities of the Japanese to Tibetans than to southern Asians (Qian et al. 2001). In quantitative estimations of maternal admixture, it was found that ~65% of the mainland Japanese gene pool was derived from Continental gene flow after the Yayoi period. However, the indigenous Ainu from the northern island of Hokkaido and the Ryukyuans from southern Okinawa showed <20% Continental specificity, pointing to them as the most probable descendants of the Jomon people. The fact that these indigenous groups were, in turn, genetically well differentiated indicated a notable degree of heterogeneity and/or isolation among the early Japanese immigrants (Horai et al. 1996). However, two handicaps of these studies are the incomplete representation of Asian populations and the relatively small sample size of those analyzed, which weakens the reliance on the relative affinities found by genetic distance methods (Helgason et al. 2001). For mtDNA there are currently enough HVI/HVII data from eastern Asia, including Japan, to test the validity of the above-mentioned results. However, these sequences have been assorted into different clades following different insufficient criteria or even have not been classified at all. Furthermore, the phylogenetic confidence of results based only on sequences from the noncoding region (HVI, HVII) has been recently questioned (Bandelt et al. 2000). This is mainly due to the frequent occurrence of parallel mutations in independent lineages that confuse the correct classification, a source of error that is increased because the basal motif in the noncoding region for the two macrolineages that expanded throughout Asia is the same (16223). In addition, as the noncoding region has not evolved at a constant rate across all human lineages, it is considered inappropriate to use this region for dating evolutionary events (Ingman et al. 2000; Finnilä et al. 2001).

To make reliable use of this important source of available data on the mtDNA noncoding region to contrast the maternal structure and to determine the most probable origin of the modern Japanese, we have undertaken the following approach: First, we used a set of complete mtDNA sequences of 672 Japanese individuals to create a phylogenetic network (Bandelt et al. 1999) that related them to other complete sequences, already published, belonging to the major haplogroups proposed by others (Torroni et al. 1992, 1996; Macaulay et al. 1999; Yao et al. 2002a). Discriminative positions in the noncoding region, defining additional Asian subhaplogroups, were then used to further classify 766 previously published Japanese partial sequences. For this purpose we also included other unambiguously assorted sequence data reported by other research groups (Derbeneva et al. 2002b; Yao et al. 2002a). These HVI sequences thus pooled were then compared with other published Asian sequences. Finally, using all of these classified sequences, we tested the relative affinities of modern Japanese and Continental Asians using global distance methods and phylogeographic approaches framed at different age levels.

RESULTS

Eastern Asia Phylogeny Based on Complete mtDNA Sequences

The phylogenetic network constructed with the complete mtDNA sequences fully coincides with those previously published at worldwide (Maca-Meyer et al. 2001; Herrnstadt et al. 2002) or regional scale (Kong et al. 2003). Moreover, their main branches are well supported by high bootstrap values on a neighbor-joining tree (Supplemental material, condensed by more than 40% bootstrap values).

From the L3 African trunk, two early branches came out of Africa and radiated extensively, originating superhaplogroups M and N, which were defined by the basic mutations depicted in Figures 1A and 2, respectively. Representatives of both superhaplogroups reached Japan. The construction of these phylogenetic trees by using our Japanese complete sequences and other published Asian sequences (Table 1) resulted in a better definition of the known haplogroups and in the identification of new clades at different phylogenetic levels. Characteristic HVI motifs and diagnostic RFLPs in the coding region, and coalescence ages for these haplogroups and subhaplogroups are given in Supplemental Tables A and B. To contribute to the unification of the mitochondrial nomenclature, we revised the previously proposed haplogroups by adding the following new information.

Subdivisions Within Macrohaplogroup M

Haplogroup D

Haplogroup D has been defined by the specific RFLP −5176 AluI (Torroni et al. 1992). Studies on Native American HVI sequences permitted further subdivision of D into subgroups D1 by mutation 16325 and D2 by mutation 16271 (Forster et al. 1996). Additional subdivisions into subhaplogroups D4 and D5 have been proposed for Asian lineages (Yao et al. 2002a). These investigators characterized D4 by position 3010. Two additional mutations, 8414 and 14668, have been proposed to define D4 (Fig. 1B; Kivisild et al. 2002). Whereas these two latter mutations seem to be rare events, 3010 has also been independently detected in haplogroups H and J. A new branch at the same phylogenetic level as D4 and D5 has been detected in Japan (Fig. 1B). It is characterized by mutations 709, 1719, 3714, and 12654 and was named D6. The subdivision of D4 into subgroups D4a and D4b was proposed on the basis of the distinctive mutational motif 152, 3206, 14979, and 16129 for the first and 10181 and 16319 for the second (Kivisild et al. 2002). Both subclades have been detected in our Japanese sample. From our data it can be deduced that mutation 8473 is also basal for D4a. In relation to D4b it seems that its ancestral branch is defined by the 8020 substitution (Fig. 1B). Consequently, the D4b subgroup proposed by Yao et al. (2002a) should be renamed D4b1 harboring 15440 and 15951 as additional basic mutations. A new subgroup character-

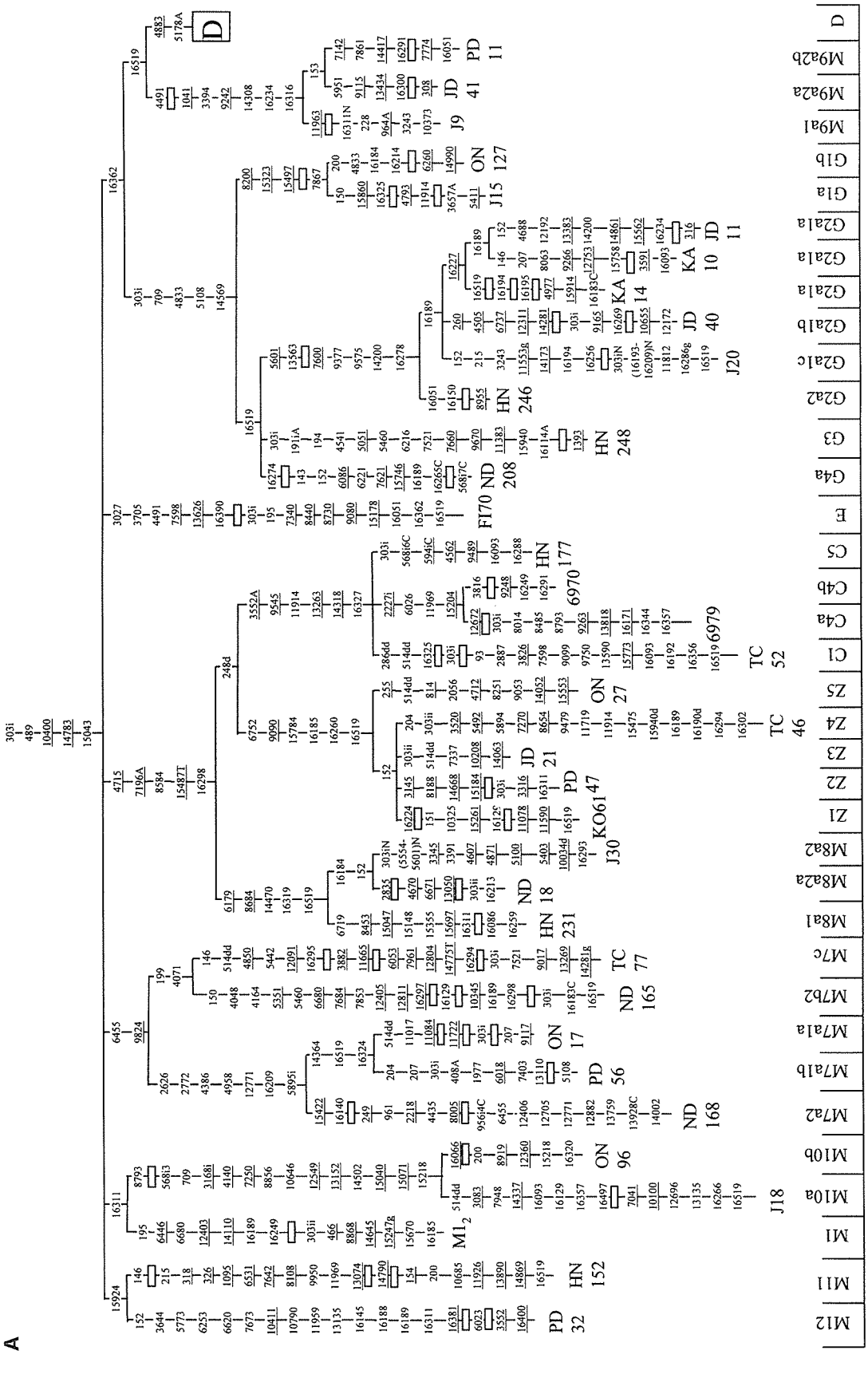

Figure 1 (Continued on next page)

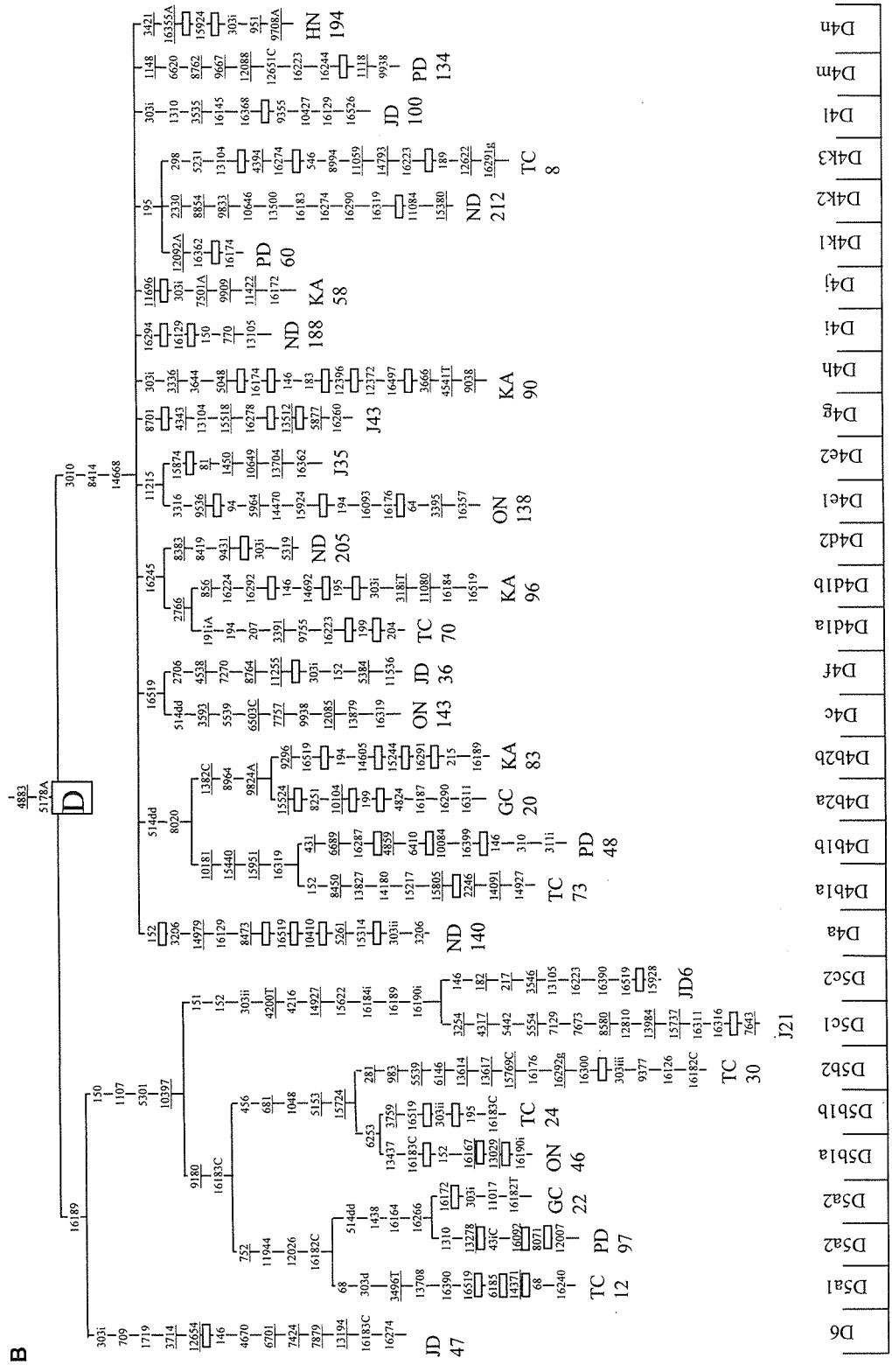

Figure 1 Phylogenetic tree, based on complete mtDNA sequences, for macrohaplogroup M in general (*A*) and for subhaplogroup D (*B*) in particular. Subject origins are given in Table 1. The numbers along the links refer to nucleotide positions, arbitrarily written in ascending order. Open boxes are nodes from which other (not shown) sequences branch. A, C, G, and T indicate transversions; whereas "d" indicates deletions and "i" insertions. Nonrecurrent mutations are underlined.

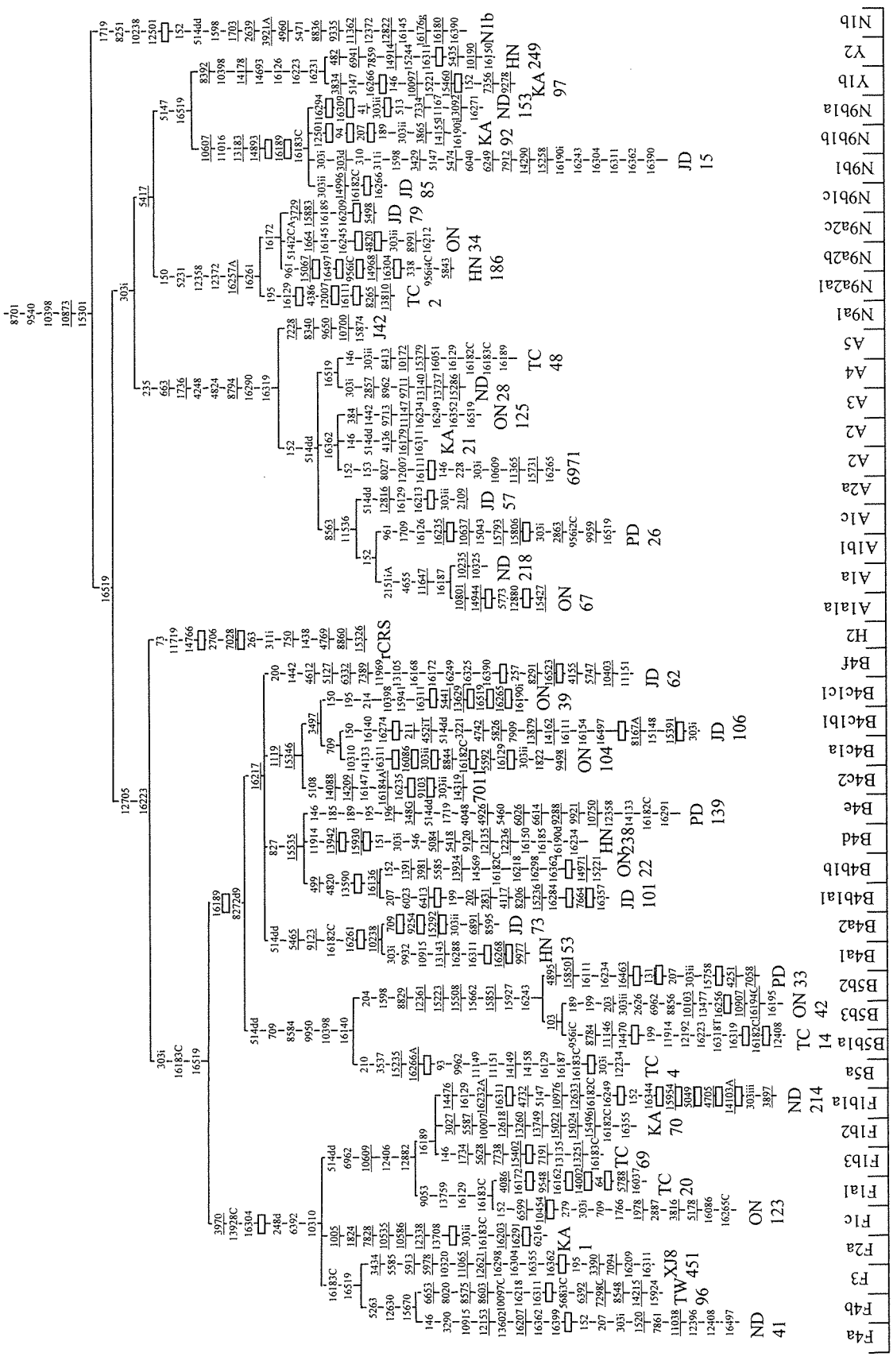

Figure 2 Phylogenetic tree, based on complete mtDNA sequences, for macrohaplogroup N. Origins of subjects are explained in Table 1. The numbers along the links refer to nucleotide positions, arbitrarily written in ascending order. Open boxes are nodes from which other (not shown) sequences branch. A, C, G, and T indicate transversions; whereas "d" indicates deletions and "i" insertions. Nonrecurrent mutations are underlined.

Table 1. List of Individuals Used to Build Up the Networks Shown in Figures 1 and 2

Sample	Haplogroup	Origin	References
PD32	M12	Japanese	This work
HN152	M11	Japanese	This work
$M1_2$	M1	Jordanian	Maca-Meyer et al. 2001
J18	M10a	Japanese	This work
ON96	M10b	Japanese	This work
ND168	M7a2	Japanese	This work
PD56	M7a1b	Japanese	This work
ON17	M7a1a	Japanese	This work
ND165	M7b2	Japanese	This work
TC77	M7c	Japanese	This work
HN231	M8a1	Japanese	This work
ND18	M8a2a	Japanese	This work
J30	M8a2	Japanese	This work
KO61	Z1	Koryac	Ingman and Gyllensten 2003
PD47	Z2	Japanese	This work
JD21	Z3	Japanese	This work
TC46	Z4	Japanese	This work
ON27	Z5	Japanese	This work
TC52	C1	Japanese	This work
6979	C4a	Evenki	Ingman et al. 2000
6970	C4B	Buryat	Ingman et al. 2000
HN177	C5	Japanese	This work
F170	E	Philippine	Ingman and Gyllensten 2003
ND208	G4a	Japanese	This work
HN248	G3	Japanese	This work
HN246	G2a2	Japanese	This work
J20	G2a1c	Japanese	This work
JD40	G2a1b	Japanese	This work
KA14	G2a1a	Japanese	This work
KA10	G2a1a	Japanese	This work
JD11	G2a1a	Japanese	This work
J15	G1a1	Japanese	This work
ON127	G1a2	Japanese	This work
J9	M9a1	Japanese	This work
JD41	M9a2a	Japanese	This work
PD11	M9a2b	Japanese	This work
JD47	D6	Japanese	This work
TC12	D5a2	Japanese	This work
PD97	D5a1a	Japanese	This work
GC22	D5a1b	Japanese	This work
ON46	D5b1a	Japanese	This work
TC24	D5b1b	Japanese	This work
TC30	D5b2	Japanese	This work
J21	D5c1	Japanese	This work
JD6	D5c2	Japanese	This work
ND140	D4a	Japanese	This work
TC73	D4b1a	Japanese	This work
PD48	D4b1b	Japanese	This work
GC20	D4b2a	Japanese	This work
KA83	D4b2b	Japanese	This work
ON143	D4c	Japanese	This work
JD36	D4f	Japanese	This work
TC70	D4d1a	Japanese	This work
KA96	D4d1b	Japanese	This work
ND205	D4d2	Japanese	This work
0N138	D4e1	Japanese	This work
J35	D4e2	Japanese	This work
J43	D4g	Japanese	This work
KA90	D4h	Japanese	This work
ND188	D4i	Japanese	This work
KA58	D4j	Japanese	This work
PD60	D4k1	Japanese	This work
ND212	D4k2	Japanese	This work
TC8	D4k3	Japanese	This work
JD100	D4l	Japanese	This work
PD134	D4m	Japanese	This work
HN194	D4n	Japanese	This work
ND41	F4a	Japanese	This work
TW96	F4b	Indigenous Taiwanese	Ingman and Gyllensten 2003
XJ8451	F3	Chinese	Kong et al. 2003
KA1	F2a	Japanese	This work
ON123	F1c	Japanese	This work
TC20	F1a1	Japanese	This work
TC69	F1b3	Japanese	This work
KA70	F1b2	Japanese	This work
ND214	F1b1a	Japanese	This work
TC4	B5a	Japanese	This work
TC14	B5b1a	Japanese	This work
ON42	B5b3	Japanese	This work
PD33	B5b2	Japanese	This work
HN153	B4a1	Japanese	This work
JD73	B4a2	Japanese	This work
JD101	B4b1a1	Japanese	This work
ON22	B4b1b	Japanese	This work
HN238	B4d	Japanese	This work
PD139	B4e	Japanese	This work
7011	B4c2	Uzbek	Ingman et al. 2000
ON104	B4c1a	Japanese	This work
JD106	B4c1b1	Japanese	This work
ON39	B4c1c1	Japanese	This work
JD62	B4f	Japanese	This work
rCRS	H2	English	Andrews et al. 1999
N1b	N1b	Jordanian	Maca-Meyer et al. 2001
ON67	A1a1a	Japanese	This work
ND218	A1a	Japanese	This work
PD26	A1b1	Japanese	This work
JD57	A1c	Japanese	This work
6971	A2a	Chukchi	Ingman et al. 2000
KA21	A2	Japanese	This work
ON125	A2	Japanese	This work
ND28	A3	Japanese	This work
TC48	A4	Japanese	This work
J42	A5	Japanese	This work
TC2	N9a1	Japanese	This work
HN186	N9a2a1	Japanese	This work
ON34	N9a2b	Japanese	This work
JD79	N9a2c	Japanese	This work
JD85	N9b1c	Japanese	This work
JD15	N9b1	Japanese	This work
KA92	N9b1b	Japanese	This work
ND153	N9b1a	Japanese	This work
KA97	Y1b	Japanese	This work
HN249	Y2	Japanese	This work

ized by 1382C, 8964, and 9824A mutations and named D4b2, is represented by lineages GC20 and KA83 in Figure 1B. Furthermore, 12 new branches at the same phylogenetic level as subhaplogroups D4a and D4b can be identified in the network. Accordingly, they have been successively named from D4c to D4n. On the other hand, D5 was defined by mutations 150, 10397, and 16189 (Yao et al. 2002a); however, 16189 is not present in all D5 lineages. We have named D5a and D5b those lineages that share this mutation and 9180 and D5c those lacking them. Consequently, we propose to rename D5a of Yao et al. (2002a) as D5a1. Additional mutations (1107 and 5301) define D5 (Fig. 1B), as has been recently confirmed (Kong et al. 2003). Of the four mutations at the basal branch of this group, 10397 seems to be a unique event; and the group can be diagnosed by the RFLP polymorphism +10396 BsrI. Recently, the phylogeny of haplogroup D has been revised in the light of complete sequences from Aleuts (Derbeneva et al. 2002b). By comparing their nomenclature to ours, it is possible to equate their D2 lineage to our D4e1 and their D3 lineage to our D4b1. As a total, D is the most abundant haplogroup in people of central and eastern Asia including mainland Japanese but not in the Ainu and Ryukyuans. However, the geographic distributions of some subhaplogroups are peculiar.

For example, D5 is prevalent in southern areas. D4a is abundant in Chukchi of northeast Siberia, but D4a1 has its highest frequency in the Ryukyuans and clade D4n in the Ainu (Table 2).

Haplogroup M9

It is confirmed that haplogroup M9 is characterized by mutation 4491 (Fig. 1A), as recently proposed (Kong et al. 2003). Subhaplogroup M9a, as redefined by Kong et al. (2003), was identified by positions 153, 3394, 14308, 16234, and 16316 (Yao et al. 2002a). Nevertheless, not all lineages have 153. Although M9 could be RFLP-diagnosed by +1038 NlaIII and +3391 HaeIII polymorphisms, the latter one should be avoided; as 3391 is also present in some D4d1 lineages (Fig. 1B) and thus could produce misclassification. We have grouped lineages with 11963 as M9a1 and those with 153 as M9a2. M9 has a central and eastern Asian geographic distribution, and it reaches its greatest frequency (11%) and diversity (87%) in Tibet. In Japan, in addition to mainland Japanese it has been detected in the indigenous Ainu and Ryukyuans (Horai et al. 1996).

Haplogroup G

This haplogroup was first detected by Ballinger et al. (1992) and later named G by Torroni et al. (1994). It was defined by the presence of the combined RFLP polymorphism +4830 HaeII/ +4831 HhaI. In addition, the basal branch has mutations 709, 5108, and 14569 (Fig. 1; Kivisild et al. 2002). Subhaplogroup G1 was defined by transition 16017 (Schurr et al. 1999) and G2 by mutations 7600 and 16278 (Yao et al. 2002a). Recently, mutations 8200, 15323, and 15497 have been used for G1 status (Kong et al. 2003). This is confirmed with our Japanese sequences; consequently, we have defined G1a by 7867 (Fig. 1A). To avoid repetitions, the G1 group of Schurr et al. (1999) has been provisionally renamed as G5 (Table 2). At least two mutations (5601 and 13563) characterize G2; and five more, G2a (Fig. 1A; Kong et al. 2003). We have defined subclade G2a1 by the presence of 16189 and the derivative G2a1a by the addition of 16227, whereas 16051 and 16150 identify G2a2 lineages. Furthermore, two new subclades, G3 and G4, are also apparent in Japanese (Fig. 1A). Subgroup G5 is dominant in northeastern Siberia, but we have not detected it in our set of Japanese complete sequences. However, G1a1 has its highest frequencies in a cluster embracing Japanese, Ainu, Ryukyuan, and Koreans. On the contrary, G2 is relatively abundant in northern China and central Asia, reaching notable frequencies in the Mansi and in Tuvinians at the respective west and east ends of South Siberia (Table 2).

Haplogroup E

Haplogroup E was first RFLP-defined as having +16389 HinfI and −7598 HhaI by Ballinger et al. (1992), who named it G, and then later it was renamed E by Torroni et al. (1994). As a loss of restriction sites can be produced by different nucleotide mutations within the recognition sequence, since the beginning, some G2 sequences characterized by the 7600 transition were erroneously classified as belonging to haplogroup E. Recently, based on the complete sequences of coding regions, Herrnstadt et al. (2002) defined three Asiatic lineages as E, although only one (sequence 214) seems to be a genuine representative. It possesses transition 7598, which, similar to 7600, is also detectable with HhaI as a site loss; and it also harbors mutations 10834 and 869, which were found by Ballinger et al. (1992) as −10830 HinfI and +868 DdeI in all and some individuals respectively classified as E. However, the inclusion of a Philippine complete sequence (Ingman and Gyllensten 2003) in our global tree clearly demonstrates that the last two mutations might only define a branch of E, as the Philippine sequence lacks both of them. On the contrary, in addition to 7598 and 16390, some of the four E mutations represented in Figure 1A before the branching point might be basic mutations. In Herrnstadt et al. (2002), sequence 169 belongs to Haplogroup M9 because it has all coding-region positions defining this haplogroup; and sequence 287 to M1 because it has 6446 and 6680, the coding-region mutations that define the basic branch of M1 (Fig. 1). It must be mentioned that the ambiguous Korean lineage classified as E/G by Schurr et al. (1999), because it had both the −7598 HhaI characteristic E site and the +4830 HhaI characteristic G site, has been recently found again in a Korean sample (Snäll et al. 2002). All of them are, in fact, members of subhaplogroup G2. It seems that haplogroup E has a southern Asia distribution. Until now it has been detected in the Malay peninsula populations and in the Sabah of Borneo (Ballinger et al. 1992); and it is also present in coastal Papua New Guinea (Stoneking et al. 1990) as well as in some Pacific islands such as Guam (Herrnstadt et al. 2002) and the Philippines (Ingman and Gyllensten 2003). However, until now, it has not been detected in more northern Continental populations or islands such as the Japanese archipelago.

Haplogroup M8

A monophyletic clade (Fig. 1A) groups M8a, C, and Z lineages. Mutations 4715, 15487T, and 16298 have been proposed as diagnostic for this clade (Yao et al. 2002a). The transversion 7196A and the transition 8584 should also be included in its definition (Fig. 1A; Kivisild et al. 2002). However, as the 248d is also shared by all Z and C lineages (Fig. 1A), a basal node defined by this deletion and named CZ has been recently proposed (Kong et al. 2003). Subhaplogroup C was RFLP-defined by Torroni et al. (1992) by +13262 AluI. Yao et al. (2002a) added 248d, 14318, and 16327 as characteristic of C. In addition, positions 3552A, 9545, and 11914 are also diagnostic of this clade (Fig. 1A; Kivisild et al. 2002). The Japanese TC52 has the C1 status and the Buryat 6970 and the Evenky 6979 have the C4 status proposed by Kong et al. (2003). Subhaplogroup Z was defined by Schurr et al. (1999) by the presence of the following noncoding motifs: 16185, 16223, 16224, 16260, and 16298. Recently, it was considered that only 16185 and 16260 mutations should be counted as basic for the group (Yao et al. 2002a). However, in full agreement with the characterization proposed on the basis of complete Chinese Z sequences (Kong et al. 2003), three additional mutations (6752, 9090, and 15784) have been placed on the basal branch of Z (Fig. 1A). We detected four Japanese Z clades that, in addition, shared mutation 152 and another without it. Tentatively, they have been named from Z1 to Z5 (Fig. 1A). Yao et al. (2002a) defined M8a by 14470, 16184, and 16319 transitions. Two more mutations (6179 and 8684) are also characteristic of this subhaplogroup (Kong et al. 2003). In Japanese we have found that 16184 is not harbored by all M8a members. Consequently, lineages with this mutation have M8a2 status and those lacking it M8a1 status (Fig. 1A). The largest diversities for C are in Korea (100%), central Asia (86%), and northern China (78%–74%). Therefore, C can be considered a clade with a Northeast Asian radiation. Representatives of subhaplogroup Z extend from the Saami (Finnilä et al. 2001) and Russians (Malyarchuk and Derenko 2001) of west Eurasia to the people of the eastern peninsula of Kamchatka (Schurr et al. 1999). Its largest diversities are found in Koreans (88%), northern China (73%), and central Asia (67%), compatible with a central–East Asian origin of radiation for this group. Finally, M8a has its highest diversity in Koreans (100%), and southern (100%) and eastern Chinese, including Taiwanese (73%). Thus, southeastern China was a potential focus of radiation of this group. All these subhaplogroups are present in mainland Japanese but neither in Ryukyuans nor in Ainu.

Haplogroup M7

This haplogroup was defined by Bamshad et al. (2001) as having two branches, M7a characterized by 16209 and M7b by 16297

Table 2. Frequency (in Percentage) of Each Haplogroup in Each Group of Populations

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
L2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	5	–	–	–
M/N	1.3	18	1.96	3.29	1.61	–	4.17	5.55	–	1.96	1.08	2.4	1.02	–	2.63	–	–	–	–	25	5.96	9.23	–	6.25	10	17.6
I/W/N	–	–	–	–	–	–	–	–	–	2.45	–	–	–	–	2.63	–	–	–	–	–	0.37	–	–	–	–	–
A1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1.67	–	–	0.19	–	–	–	–	–
A1a	2.13	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	3.54	–	–	–	–	–
A1b	4.57	2	3.92	5.16	7.36	–	5.56	2.77	7.46	6.86	2.15	0.48	3.06	6.52	–	1.79	3.3	–	5.56	5	2.98	6.15	–	–	–	–
A1b1	0.15	–	–	0.47	0.46	–	1.39	0.92	–	–	–	–	–	–	–	–	–	–	–	–	0.74	–	–	–	–	–
A2a	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	10	–	–	–	–	–	–	–	–
A2b	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	3.57	2.2	50	–	–	0.37	3.08	–	–	–	–
N9a	4.57	–	–	3.76	1.38	–	6.94	2.91	–	1.96	–	–	–	–	–	–	–	–	–	–	3.91	–	–	–	2.5	–
N9b	2.13	2	2	–	–	–	–	0.13	–	–	–	–	–	–	–	–	–	–	–	–	0.19	–	–	–	–	–
Y	0.38	–	21.6	3.29	1.38	–	–	0.53	–	0.49	–	0.96	–	4.35	–	7.14	9.89	–	–	–	2.23	–	–	3.13	2.5	2.94
HV	0.91	4	–	0.47	1.84	–	–	1.85	–	13.7	20.5	–	18.4	–	36.79	–	1.1	–	2.78	5	1.12	–	–	3.13	5	–
JT	–	–	–	2.82	1.15	–	–	–	–	5.4	8.61	–	19.4	–	23.65	–	–	–	–	–	–	–	–	–	–	–
UK	–	–	–	–	1.38	–	–	0.39	–	10.3	5.39	–	25.5	–	10.52	–	–	–	2.78	–	–	–	–	–	5	–
R9a	0.08	–	–	1.88	0.69	–	–	1.85	–	–	–	7.69	–	–	–	–	–	–	–	–	–	–	–	3.13	–	2.94
R11	–	–	–	0.94	0.46	12.5	–	1.85	–	0.49	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B	1.3	–	–	0.47	1.84	–	4.17	1.19	–	0.98	1.08	–	–	–	–	–	–	–	–	–	–	–	–	3.13	–	2.94
B4	0.76	2	–	1.88	2.53	9.38	1.39	1.98	5.97	1.96	1.08	–	–	–	–	–	–	–	–	2.5	0.74	1.54	–	–	–	–
B4a	0.84	–	–	2.35	1.61	–	1.39	4.36	–	0.49	1.08	14.9	–	–	–	–	–	–	2.78	–	0.56	–	–	–	–	–
B4a1	0.84	2	–	–	–	–	–	–	–	–	–	0.48	–	–	–	–	–	–	–	–	0.56	–	–	–	–	–
B4a2	–	–	–	0.47	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B4a3	–	–	–	0.47	–	15.6	2.78	1.85	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B4b	0.53	–	–	1.88	1.61	–	6.94	1.72	–	–	–	4.33	–	–	–	–	–	–	11.1	–	0.74	–	–	–	–	–
B4b1	2.13	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.56	–	–	–	–	–
B4c1	1.6	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.74	–	–	–	–	–
B4c1b	0.61	–	–	1.88	–	–	–	1.06	–	–	–	4.81	–	–	–	–	–	–	–	–	0.56	–	–	–	–	–
B4c2	0.08	–	–	–	–	–	–	1.19	–	0.49	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B4f	0.3	4	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.38	–	–	–	–	–
B5	–	–	–	–	–	–	–	0.13	–	–	–	0.48	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B5a	–	–	–	1.88	0.69	–	1.39	6.74	5.97	0.98	1.08	6.25	–	–	–	–	–	–	–	–	–	–	–	3.13	–	11.8
B5a1	0.61	–	–	0.47	1.38	–	–	–	–	0.49	1.08	–	–	–	–	–	–	–	–	–	3.54	–	–	–	–	–
B5b	0.3	2	–	–	1.38	–	–	0.4	–	0.49	–	–	–	–	–	–	–	–	–	–	0.19	–	–	12.5	–	–
B5b1	0.99	–	–	–	–	–	–	0.13	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B5b2	2.29	–	–	0.47	1.15	–	1.39	0.26	–	0.49	–	–	–	–	–	–	–	–	–	–	0.93	–	–	–	–	–
B5b3	0.08	–	–	–	–	–	–	0.13	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
F	0.23	–	–	0.94	2.3	–	–	2.25	1.49	1.96	2.15	1.92	–	–	–	–	–	–	–	–	0.37	1.54	–	–	–	2.94
F1a	0.15	–	–	0.47	1.61	15.6	4.17	4.62	1.49	–	1.08	–	–	–	–	–	–	–	–	–	0.74	–	–	12.5	35	2.94
F1a1	1.52	–	–	2.82	1.84	15.6	2.78	7.93	–	1.47	–	7.21	–	–	–	–	–	–	–	–	1.12	–	–	–	5	–
F1b	3.13	2	2	2.82	2.99	–	2.78	1.19	2.99	3.92	–	–	1.02	–	–	–	–	–	8.33	5	2.05	–	–	–	–	–
F1c	–	–	–	1.41	1.15	–	1.39	0.4	–	0.49	–	–	–	–	–	–	–	–	–	–	0.19	1.54	–	–	–	–

(continued)

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