Table 1. RT patients analysed in this study

All patients received a transplanted kidney from a living related donor.

Patient	Gender, age (years)	Time after transplant (months)	Serum creatinine (mg dl^{-1})*
1	F (51)	4	1·4
2	M (30)	13	1·5
3	M (42)	82, 83	1·4
4	F (36)	12	1·7
5	M (55)	3	1·1
6	M (49)	7	1·3

*To convert to μmol l^{-1} multiply by 88.

coding region were detected in six individuals, but not in five individuals. The detected variations were mostly single-nucleotide substitutions, and only three of 10 nt substitutions caused amino acid substitutions. Furthermore, the authors detected possible prototypal sequences at the nodes of family specific clusters of phylogenetic trees.

We examined the stability of the BKPyV genome in RT patients without PVAN, as a basis of future studies analysing possible genetic changes of BKPyV associated with the pathogenesis or progression of PVAV (see above). We established multiple full-sized BKPyV DNA clones from the urine of each of six RT patients with surviving renal allografts, by using the standard method of molecular cloning (Sambrook *et al.*, 1989). In each patient, three to five complete BKPyV DNA clones were sequenced and the resultant sequences were compared in each patient.

RT patients analysed in this study are shown in Table 1. Entire BKPyV DNAs were cloned into pUC19 at the unique *Bam*HI site by the standard method (Sambrook *et al.*, 1989) as described previously (Yogo *et al.*, 1991). The complete BKPyV DNA clones were prepared using a Qiagen Plasmid Maxi kit. Purified plasmids were sequenced as described

Table 2. Complete BKPyV DNA sequences detected in patients 1–6

Nucleotides (amino acids) at positions of the Dunlop genome (Seif *et al.*, 1979) are shown, when differences were detected among sequences from the same patients. The numbers of clones with individual sequences and the GenBank/EMBL/DDBJ accession numbers are also shown. Six clones (TW-1, TW-3, TW-4, TW-5, TW-8 and THK-9) will be reported elsewhere (T. Takasaka and others, unpublished data).

Sequence	Subtype*	nt 1744†	nt 1807†	No. clones	Accession no.‡
Patient 1					
TW-1§	Ic	G (E)	G (E)	2	AB211381
TW-1a	Ic	G (E)	C (Q)	1	AB217917
TW-1b	Ic	A (K) nt 410‖	G (E)	1	AB217918
Patient 2					
TW-3	IV	T (L)	–	1	AB211391
TW-3a§	IV	G (R)	–	4	AB217919
Patient 3					
TW-4		–	–	3	AB211382
Patient 4					
TW-5	Ic	– nt 1156¶	–	3	AB211383
Patient 5					
TW-8§	Ic	C (S)	–	3	AB211385
TW-8a	Ic	A (stop codon) nt 1127¶	–	1	AB217920
Patient 6					
THK-9§	Ic	T (V)	–	2	AB211379
THK-9a	Ic	G (V)	–	1	AB217921

*Previously classified using phylogenetic analysis (T. Takasaka and others, unpublished data).
†Located in the VP1 gene.
‡GenBank/EMBL/DDBJ accession numbers.
§Ancestral sequence (see text).
‖Located in the agnogene.
¶Located in the VP2/3 gene.

elsewhere (T. Takasaka and others, unpublished data). The determined and reference sequences were aligned using the CLUSTAL W program (Thompson *et al.*, 1994). Translation of nucleotide sequences into amino acid sequences was performed with GENETYX-MAC version 11.10 (Genetyx). A neighbour-joining (NJ) phylogenetic tree (Saitou & Nei, 1987) was constructed using the CLUSTAL W program (Thompson *et al.*, 1994). Divergences were estimated with the two-parameter method (Kimura, 1980). The phylogenetic tree was visualized using TREEVIEW (Page, 1996). The confidence of branching patterns of the NJ trees was assessed based on 1000 bootstrap replicates (Felsenstein, 1985).

When we recently established full-sized BKPyV DNA clones from the urine of RT patients, we obtained multiple clones in six patients, but sequenced only a single clone for each patient (T. Takasaka and others, unpublished data). In the present study, we determined the clones left unsequenced. A comparison of the sequences determined in this and the other study (T. Takasaka and others, unpublished data) identified one to three unique sequences in each patient (Table 2). (i) Three sequences, TW-1, TW-1a and TW-1b, were detected in patient 1. In reference to TW-1, TW-1a and TW-1b carried single nucleotide substitutions at nt 1807 and 1744, respectively, within the VP1 gene [nucleotide numbers are those of the strain Dunlop (Seif *et al.*, 1979)]. Both nucleotide substitutions resulted in amino acid substitutions within a predicted outer loop (the BC loop) of the VP1 protein (Chang *et al.*, 1996) (Table 2). (ii) Two sequences, TW-3 and TW-3a, were detected in patient 2. In reference to TW-3, TW-3a carried a single nucleotide substitution at nt 410 within the agnogene. This nucleotide substitution resulted in an amino acid substitution. (iii) A single sequence, TW-4, was detected in patient 3. (iv) Similarly, a single sequence, TW-5, was detected in patient 4. (v) Two sequences, TW-8 and TW-8a, were detected in patient 5. In reference to TW-8, TW-8a carried a single nucleotide substitution at nt 1156 within the VP2/3 gene. This nucleotide substitution inserted a stop codon. (vi) Finally, two sequences, THK-9 and THK-9a, were detected in patient 6. In reference to THK-9, THK-9a carried a single nucleotide substitution at nt 1127 within the VP2/3 gene. This nucleotide substitution did not result in any amino acid substitution.

The TCR sequences of representative BKPyV clones in patients 1–6 were previously clarified, and all had the archetypal configuration (Takasaka *et al.*, 2004). In the

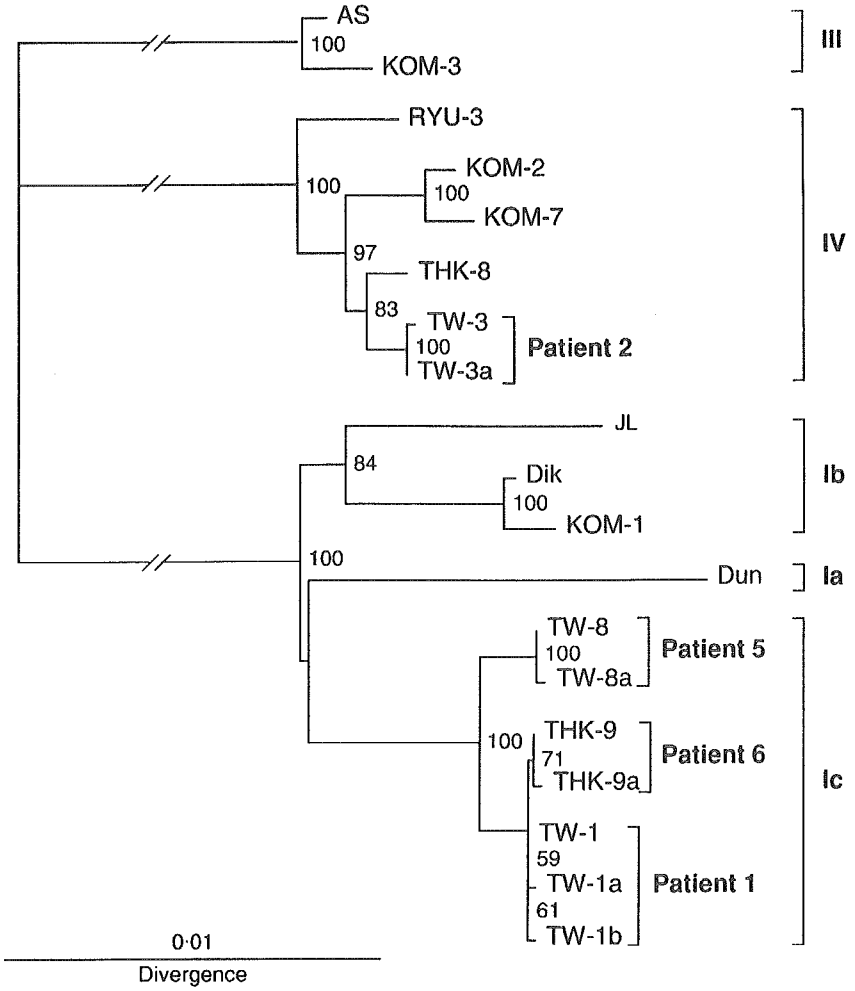

Fig. 1. NJ phylogenetic tree relating complete BKPyV DNA sequences detected in patients 1, 2, 5 and 6. An NJ phylogenetic tree was constructed from nine complete BKPyV sequences detected in patients 1, 2, 5 and 6, and 10 complete BKPyV sequences detected in unrelated individuals and belonging to subtype Ia, Ib, III or IV (Seif *et al.*, 1979; Tavis *et al.*, 1989; T. Takasaka and others, unpublished data) (the non-coding regulatory region of the BKPyV genome was excluded from this phylogenetic analysis). The phylogenetic tree was visualized using TREEVIEW. The tree was rooted using subtype III isolates (i.e. AS and KOM-3) as the outgroup. The numbers at nodes in the tree indicate the bootstrap confidence levels (percentage) obtained with 1000 replications (only values ⩾50% are shown). Subtypes Ia, Ib, Ic, III and IV and sequences detected in patients 1, 2, 5 and 6 are indicated.

present study, we did not detect any difference in the TCR among clones derived from each patient.

To elucidate the evolutionary relationships among several unique sequences detected in patients 1, 2, 5 and 6, we constructed an NJ phylogenetic tree from the BKPyV DNA sequences detected in these patients together with reference sequences reported previously (Seif *et al.*, 1979; Tavis *et al.*, 1989; T. Takasaka and others, unpublished data). On the resultant tree (Fig. 1), the BKPyV DNA sequences in patients 1, 2, 5 and 6 formed individual clusters. We detected a sequence (TW-1, TW-3a, TW-8 or THK-9) at the node of each cluster, probably representing the prototypal sequence that generated variant sequences in each patient. It may be worth noting that the prototypal sequences were usually the major ones (Table 2).

In this study, we investigated the stability of the BKPyV genome in RT recipients without PVAN. We found that the genome of BKPyV is rather stable in these patients, with only minor nucleotide substitutions in the coding region. This finding forms the basis of further study on the genetic changes possibly involved in the progression of PVAN.

The finding noted immediately above is contradictory to a high intra-strain genetic diversity in BKPyV suggested by Chen *et al.* (2004). This discrepancy may be related to the difference in the methods used to obtain a full-sized genome. We used standard molecular cloning (Sambrook *et al.*, 1989), whereas Chen *et al.* (2004) used PCR amplification. Standard molecular cloning warrants the isolation of intact complete viral genomes, while PCR amplification inevitably involves replication errors, even though the frequency of errors may be reduced by using a thermostable DNA polymerase with proofreading activity. The frequently detected variations in BKPyV (Chen *et al.*, 2004) (see above) could have been introduced by the authors during PCR. Nevertheless, it remains to be elucidated whether the BKPyV genome undergoes a high variability in a specific disease, i.e. $BKPyV_{CAP}$.

In this study, we detected four non-synonymous nucleotide substitutions and one synonymous substitution. Of the four non-synonymous substitutions, three resulted in amino acid changes in VP1 and the agnoprotein, and one generated an incomplete VP2/3 protein due to the insertion of a stop codon. While viruses with incomplete VP2/3 proteins may not be infectious, it remains to be elucidated whether the amino acid changes in VP1 and the agnoprotein cause alterations in the properties of BKPyV.

Acknowledgements

This study was supported in part by grants from the Ministry of Health, Labor and Welfare, Japan.

References

Baksh, F. K., Finkelstein, S. D., Swalsky, P. A., Stoner, G. L., Ryschkewitsch, C. F. & Randhawa, P. (2001). Molecular genotyping of BK and JC viruses in human polyomavirus-associated interstitial nephritis after renal transplantation. *Am J Kidney Dis* **38**, 354–365.

Chang, D., Liou, Z.-M., Ou, W.-C., Wang, K.-Z., Wang, M., Fung, C.-Y. & Tsai, R.-T. (1996). Production of the antigen and the antibody of the JC virus major capsid protein VP1. *J Virol Methods* **59**, 177–187.

Chen, C.-H., Wen, M.-C., Wang, M., Lian, J.-D., Wu, M.-J., Cheng, C.-H., Shu, K.-H. & Chang, D. (2001). A regulatory region rearranged BK virus is associated with tubulointerstitial nephritis in a rejected renal allograft. *J Med Virol* **64**, 82–88.

Chen, Y., Sharp, P. M., Fowkes, M., Kocher, O., Joseph, J. T. & Koralnik, I. J. (2004). Analysis of 15 novel full-length BK virus sequences from three individuals: evidence of a high intra-strain genetic diversity. *J Gen Virol* **85**, 2651–2663.

de Bruyn, G. & Limaye, A. P. (2004). BK virus-associated nephropathy in kidney transplant recipients. *Rev Med Virol* **14**, 193–205.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120.

Page, R. D. M. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.

Randhawa, P. S., Khaleel-Ur-Rehman, K., Swalsky, P. A., Vats, A., Scantlebury, V., Shapiro, R. & Finkelstein, S. (2002). DNA sequencing of viral capsid protein VP-1 region in patients with BK virus interstitial nephritis. *Transplantation* **73**, 1090–1094.

Randhawa, P., Zygmunt, D., Shapiro, R., Vats, A., Weck, K., Swalsky, P. & Finkelstein, S. (2003). Viral regulatory region sequence variations in kidney tissue obtained from patients with BK virus nephropathy. *Kidney Int* **64**, 743–747.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning, a Laboratory Manual*, 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory.

Seif, I., Khoury, G. & Dhar, R. (1979). The genome of human papovavirus BKV. *Cell* **18**, 963–977.

Takasaka, T., Goya, N., Tokumoto, T. & 14 other authors (2004). Subtypes of BK virus prevalent in Japan and variation in their transcriptional control region. *J Gen Virol* **85**, 2821–2827.

Tavis, J. E., Walker, D. L., Gardner, S. D. & Frisque, R. J. (1989). Nucleotide sequence of the human polyomavirus AS virus, an antigenic variant of BK virus. *J Virol* **63**, 901–911.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Yogo, Y., Iida, T., Taguchi, F., Kitamura, T. & Aso, Y. (1991). Typing of human polyomavirus JC virus on the basis of restriction fragment length polymorphisms. *J Clin Microbiol* **29**, 2130–2138.

Zheng, H. Y., Kitamura, T., Takasaka, T., Chen, Q. & Yogo, Y. (2004). Unambiguous identification of JC polyomavirus strains transmitted from parents to children. *Arch Virol* **149**, 261–273.

AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 129:465–472 (2006)

Phylogenetic Analysis of Major African Genotype (Af2) of JC Virus: Implications for Origin and Dispersals of Modern Africans

Tomokazu Takasaka,[1] Tadaichi Kitamura,[1] Chie Sugimoto,[2] Jing Guo,[3] Huai-Ying Zheng,[1] and Yoshiaki Yogo[1]*

[1]*Department of Urology, Faculty of Medicine, University of Tokyo, Tokyo 113-8655, Japan*
[2]*Tsukuba Primate Center for Medical Science, National Institute of Infectious Diseases, Tsukuba 305-0843, Japan*
[3]*Membrane Protein Research Group, Canadian Institutes of Health, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada*

KEY WORDS JC virus; complete DNA sequences; phylogenetic analysis; population history; dispersals of modern Africans

ABSTRACT Both mtDNA and the Y chromosome have been used to investigate how modern humans dispersed within and out of Africa. This issue can also be studied using the JC virus (JCV) genotype, a novel marker with which to trace human migrations. Africa is mainly occupied by two genotypes of JCV, designated Af1 and Af2. Af1 is localized to central/western Africa, while Af2 is spread throughout Africa and in neighboring areas of Asia and Europe. It was recently suggested that Af1 represents the ancestral type of JCV, which agrees with the African origin of modern humans. To better understand the origin of modern Africans, we examined the phylogenetic relationships among Af2 isolates worldwide. A neighbor-joining phylogenetic tree was constructed based on the complete JCV DNA sequences of 51 Af2 isolates from Africa and neighboring areas. According to the resultant tree, Af2 isolates diverged into two major clusters, designated Af2-a and -b, with high bootstrap probabilities. Af2-a contained isolates mainly from South Africa, while Af2-b contained those from the other parts of Africa and neighboring regions of Asia and Europe. These findings suggest that Af2-carrying Africans diverged into two groups, one carrying Af2-a and the other carrying Af2-b; and that the former moved to southern Africa, while the latter dispersed throughout Africa and to neighboring regions of Asia and Europe. The present findings are discussed with reference to relevant findings in genetic and linguistic studies. Am J Phys Anthropol 129:465–472, 2006.

Recently, nonrecombining genetic loci, mtDNA, and the Y chromosome were used to clarify the population history of our species. These studies, along with classical studies of nuclear genes, supported the recent theory of an African origin of modern humans (Cavalli-Sforza and Feldman, 2003, and references therein). In this study, we focused on human dispersal within and out of Africa occurring after the emergence of anatomically modern humans in Africa.

It is thought that two haplogroups (L0 and L1) of human mtDNA and two (A and B) of the human Y chromosome represent the ancestral human lineages, based on phylogenetic trees relating various worldwide haplogroups (Cavalli-Sforza and Feldman, 2003; Forster, 2004) (note that the former paraphyletic L1 group split into haplogroups L0 and L1; Mishmar et al., 2003). The L0 and L1 haplogroups occur mainly in African hunter-gatherers (Watson et al., 1997). Haplogroup A is frequent in eastern and southern Africa, while haplogroup B is spread throughout sub-Saharan Africa (Hammer et al., 2001). These ancestral haplogroups of mtDNA and the Y chromosome are now a minority, and instead the L2 and L3 haplogroups of mtDNA and the E haplogroup of the Y chromosome are most prevalent in Africa (Watson et al., 1997: Salas et al., 2004). These major haplogroups are thought to represent the African reexpansion about 60,000–80,000 years ago, leading to the first successful modern human migration out of Africa (Forster, 2004).

Language is a useful basis for classification of African peoples. The major linguistic families of Africa are Afroasiatic, Niger-Kordofanian (including Niger-Congo), Nilo-Saharan, and Khoisan (Greenberg, 1963). Niger-Kordofanian, Nilo-Saharan, and Khoisan are spoken exclusively in Africa, while Afroasiatic languages are also spoken in the Middle East. Languages of the Khoisan family are spoken by hunter-gatherers in southwesternmost parts of the continent. The present-day distribution of African languages would have been associated, at least partly, with human dispersals within Africa.

Human dispersals in Africa can also be studied based on the JC virus (JCV) genotype, a novel marker with

Grant sponsor: Ministry of Education, Culture, Sports, Science and Technology, Japan; Grant sponsor: Ministry of Health, Labor and Welfare, Japan.

*Correspondence to: Yoshiaki Yogo, Department of Urology, Faculty of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. E-mail: yogo-tky@umin.ac.jp

Received 10 May 2004; accepted 14 September 2004.

DOI 10.1002/ajpa.20208
Published online 5 December 2005 in Wiley InterScience (www.interscience.wiley.com).

which to trace human migrations (Sugimoto et al., 1997; Agostini et al., 1997b; Yogo et al., 2004). JCV is a member of the *Polyomaviridae* family. Its genome is a single molecule of covalently closed, circular double-stranded DNA about 5,100 bp in length (Cole and Conzen, 2001). JCV is ubiquitous in the human population, infecting children asymptomatically, and then persisting in renal tissue throughout life (Padgett and Walker, 1973; Chesters et al., 1983; Tominaga et al., 1992; Kitamura et al., 1997). Indeed, JCV DNA can be detected in the urine of more than 50% of healthy individuals aged 40 years or older (Kitamura et al., 1990, 1994; Agostini et al., 1996). JCV is usually transmitted from parents to children during cohabitation (Kunitake et al., 1995; Suzuki et al., 2002; Zheng et al., 2004), but is rarely transmitted among human populations, unless the populations are intermixed (Kato et al., 1997). JCV strains in the world can be classified into more than 10 major genotypes (Sugimoto et al., 1997; Guo et al., 1998; Agostini et al., 2001b), with each genotype occupying a unique geographical domain, suggesting that the evolution of JCV occurred in association with the division of human populations. Furthermore, the rate of synonymous substitutions is two orders of magnitude higher in JCV than in human nuclear genes (Hatwell and Sharp, 2000; Sugimoto et al., 2002). The high mutation rate, together with the host-linked mode of evolution, appears to make JCV a highly sensitive and reliable marker for tracing the history of human populations (Yogo et al., 2004).

Two genotypes (Af1 and Af2) of JCV are mainly spread in Africa (Sugimoto et al., 1997). Af1 (also named type 6; Agostini et al., 2001b) is localized to central/western Africa (Sugimoto et al., 1997; Chima et al., 1998), while Af2 (also named type 3; Agostini et al., 2001b) is widespread in not only all of Africa but also western, southern, and central-western Asia and southern Europe (Sugimoto et al., 1997; Agostini et al., 1997a, 2001a; Chima et al., 1998; Dubois et al., 2001; Saruwatari et al., 2002; Pagani et al., 2003). Af1 probably represents the ancestral type of JCV (Pavesi, 2003), and its occurrence in Africa is consistent with the African origin of our species. Af2 was generated by the first split in the type-B supercluster about 50,000 years ago (Af1 was assumed to have emerged about 100,000 years ago; Sugimoto et al., 2002). Thus, Af2 appears to represent more recent dispersals of modern humans in Africa and neighboring areas. In the present study, we performed a phylogenetic study of Af2 isolates worldwide, to gain some insights into the population history of modern humans in Africa and neighboring areas.

Jobes et al. (1998) analyzed 22 JCV isolates, including five belonging to type 3 (Af2 according to our designation system). These authors found that the five type-3 isolates split into two subgroups, named type 3A and type 3B. Four isolates (two Tanzanian and two African-American isolates) belonged to type 3A, while a single African American isolate belonged to type 3B. As the number of JCV isolates belonging to Af2 was too small in the study noted above, we attempted to examine genetic variations among a larger number of Af2 isolates obtained in Africa and neighboring areas.

MATERIALS AND METHODS

Geographic origins of JCV isolates

JCV isolates were recovered in this and a previous study (Guo et al., 1996) by molecular cloning (see below) from urine samples collected from native volunteers or patients at sites indicated in Table 1.

DNA analysis

Entire JCV DNAs were cloned into pUC19 at the unique *Bam*HI site, as described previously (Yogo et al., 1991). The resultant complete JCV DNA clones were prepared using a QIAGEN Plasmid Midi kit (QIAGEN GmbH, Hilden, Germany). Purified plasmids were used for a cycle-sequencing reaction set up using the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences, Buckinghamshire, UK). Primers used were a set of primers reported previously (Agostini et al., 1997a), excluding JIG-9, JIG-10, JIG-53, and SEC-11, and four additional primers: 11AF (5′-TTTTTGAGGGGACAGAGCAACTTCC-3′, corresponding to nucleotides 2445–2469 in the JCV (Mad-1) genome (Frisque et al., 1984), with a few mismatches), W-1a (5′-CTTCGCCAGCTGTCACGTAAGGCTTCTG-3′, corresponding to nucleotides 283–310), and the M13 universal forward and reversal primers. Primers were added to a final concentration of 0.25 pmol/μl in a final reaction volume of 20 μl. The cycling conditions were 25 cycles of 30 sec at 96°C, 15 sec at 50°C, and 60 sec at 60°C. The reaction was terminated at 4°C. Cycle-sequencing products were purified by Centri-Sep columns (Princeton Separations, Adelphia, NJ). DNA sequencing was performed using an automated sequencer (ABI Prism 373S DNA sequencer, Applied Biosystems, Foster City, CA).

Phylogenetic analysis

The noncoding regulatory region of the JCV genome was excluded from phylogenetic analysis, as this region is hypervariable (Yogo and Sugimoto, 2001). Rates of synonymous substitution were estimated using the Diverge program in the GCG Wisconsin package (Accelrys Inc., San Diego, CA). DNA sequences were aligned using CLUSTAL W (Thompson et al., 1994), with a gap-opening penalty of 15.00 and gap-extension penalty of 6.66. To evaluate phylogenetic relationships among DNA sequences, we performed a neighbor-joining (NJ) analysis (Saitou and Nei, 1987), using the CLUSTAL W program. Divergences were estimated with the two-parameter method of Kimura (1980). To assess the confidence of branching patterns of the NJ tree, bootstrap probabilities (BPs) were estimated with 1,000 bootstrap replicates (Felsenstein, 1985), using CLUSTAL W. A phylogenetic tree was visualized using the TREEVIEW program (Page, 1996).

RESULTS

Phylogenetic analysis of Af2 isolates worldwide

We sequenced 47 complete JCV (Af2) DNA clones, including 28 established previously (Guo et al., 1996) and 19 established in this study (the origins of these clones are shown in Table 1). We confirmed that these complete sequences were not recombinant, using the method described previously (Sugimoto et al., 2002) (data not shown). We constructed an NJ phylogenetic tree from these sequences plus four complete Af2 sequences (numbers 308, 311, 312, and ET-3) reported previously (Table 1). The latter included three Af2 isolates (numbers 308, 311, and 312) which were previously classified into two groups (types 3A and 3B) by Jobes

TABLE 1. JCV isolates used to construct the NJ phylogenetic tree (Fig. 1)

Geographic origin	Linguistic origin[1]	Isolate	Genotype	Reference[2]	Accession no.[3]
Africa					
CAR (Bangui)	Niger-Kordofanian	CA-8	Af2-b1	a/b	AB126981
CAR (Bangui)	Niger-Kordofanian	CA-9	Af2-b1	a/b	AB126982
Ethiopia (Addis Ababa)	Afroasiatic	ET-1	Af2-b[4]	a/b	AB126983
Ethiopia (Addis Ababa)	Afroasiatic	ET-2	Af2-b[4]	a/b	AB126984
Ethiopia (Addis Ababa)	Afroasiatic	ET-3	Af2-b2	b/c	AB048547
Ethiopia (Addis Ababa)	Afroasiatic	ET-5	Af2-b2	a/b	AB126985
Ethiopia (Addis Ababa)	Afroasiatic	ET-8	Af2-b2	a/b	AB126986
Kenya (Nairobi)	Niger-Kordofanian	KE-1	Af2-b2	a/b	AB126993
Kenya (Nairobi)	Niger-Kordofanian	KE-2	Af2-b2	a/b	AB126994
Kenya (Nairobi)	Niger-Kordofanian	KE-4	Af2-b2	a/b	AB126995
Kenya (Nairobi)	Niger-Kordofanian	KE-5	Af2-b[4]	a/b	AB126996
Kenya (Nairobi)	Niger-Kordofanian	KE-7	Af2-b2	a/b	AB126997
Mauritania (Nouakchott)	Afroasiatic	MA-1	Af2-b1	b/b	AB126998
Mauritania (Nouakchott)	Afroasiatic	MA-8	Af2-b1	b/b	AB126999
Morocco (Fes/Ifrane)	Afroasiatic	MR-6	Af2-b1	b/b	AB127001
Morocco (Fes/Ifrane)	Afroasiatic	MR-9	Af2-b1	b/b	AB127002
Morocco (Fes/Ifrane)	Afroasiatic	MR-10	Af2-b1	b/b	AB127003
Niger (Tessaoua)	Afroasiatic	NG-2	Af2-b1	b/b	AB127004
Niger (Tessaoua)	Afroasiatic	NG-3	Af2-b1	b/b	AB127005
Niger (Tessaoua)	Afroasiatic	NG-4	Af2-b1	b/b	AB127006
Niger (Tessaoua)	Afroasiatic	NG-5	Af2-b1	b/b	AB127007
South Africa (Welkom)	Niger-Kordofanian	SO-1	Af2-a	a/b	AB127012
South Africa (Welkom)	Niger-Kordofanian	SO-2	Af2-a	a/b	AB127013
South Africa (Welkom)	Niger-Kordofanian	SO-3	Af2-a	a/b	AB127014
South Africa (Welkom)	Niger-Kordofanian	SO-4	Af2-a	a/b	AB127015
South Africa (Welkom)	Niger-Kordofanian	SO-5	Af2-a	a/b	AB127016
South Africa (Welkom)	Niger-Kordofanian	SO-6	Af2-a	a/b	AB127017
Sudan (Khartoum)	Afroasiatic	SU-4	Af2-b2	b/b	AB127018
Sudan (Khartoum)	Afroasiatic	SU-5	Af2-b3	b/b	AB127019
Tanzania (Shirati)	Niger-Kordofanian	No. 308	Af2-b2	–[5]/d	U73500
Zambia (Lusaka)	Niger-Kordofanian	ZA-1	Af2-b2	a/b	AB127026
Zambia (Lusaka)	Niger-Kordofanian	ZA-2	Af2-b2	a/b	AB127027
Asia					
India (Varanasi)	Indo-European	IN-1	Af2-b3	a/b	AB126988
India (Varanasi)	Indo-European	IN-3	Af2-b3	a/b	AB126989
India (Varanasi)	Indo-European	IN-4	Af2-b3	a/b	AB126990
India (Varanasi)	Indo-European	IN-7	Af2-b3	a/b	AB126991
India (Varanasi)	Indo-European	IN-8	Af2-b3	a/b	AB126992
Saudi Arabia (Riyadh)	Afroasiatic	SA-1	Af2-b2	a/b	AB127008
Saudi Arabia (Riyadh)	Afroasiatic	SA-6	Af2-b2	a/b	AB127009
Saudi Arabia (Riyadh)	Afroasiatic	SA-7	Af2-b2	a/b	AB127010
Saudi Arabia (Riyadh)	Afroasiatic	SA-8	Af2-b[4]	a/b	AB127011
Turkey (Ankara)	Afroasiatic	TU-16	Af2-b3	b/b	AB127020
Turkey (Ankara)	Afroasiatic	TU-18	Af2-b3	b/b	AB127021
Uzbekistan (Tashkent)	Altaic	UZ-4	Af2-b3	b/b	AB127025
Uzbekistan (Tashkent)	Altaic	UZ-19	Af2-b3	b/b	AB127022
Uzbekistan (Tashkent)	Altaic	UZ-20	Af2-b3	b/b	AB127023
Uzbekistan (Tashkent)	Altaic	UZ-24	Af2-b3	b/b	AB127024
Europe					
Greece (Athens)	Indo-European	GR-17	Af2-b3	b/b	AB126987
Macedonia (Bitola)	Indo-European	MC-18	Af2-b3	b/b	AB127000
Americas					
USA (MD)	Indo-European	No. 311	Af2-a	–/d	U73501
USA (CA)	Indo-European	No. 312	Af2-b1	–/d	U73502

[1] Linguistic groups prevailing in areas where isolates were obtained (Greenberg, 1963).
[2] References for complete JCV DNA clones and complete JCV DNA sequences are indicated. a, Guo et al., 1996; b, this study; c, Sugimoto et al., 2002; d, Agostini et al., 1997a.
[3] GSDB (Genome Sequence Data Base), DDBJ (DNA Data Bank of Japan), EMBL (EMBL Nucleotide Sequence Database), and NCBI (National Center for Biotechnology Information) accession numbers.
[4] Not subclassified.
[5] No complete JCV DNA was isolated (PCR-amplified fragments were sequenced).

et al. (1998). Two Tanzanian Af2 isolates, numbers 309 and 310, were not included, as their reported sequences had too many mutations (Jobes et al., 1998). According to the resultant tree (Fig. 1), we found that Af2 isolates worldwide are classified into two major clusters, designated Af2-a and -b, with high BPs (95% and 99%). Af2-a contained six South African (SO-1–6) and one African-American isolate (number 311), while Af2-b contained all

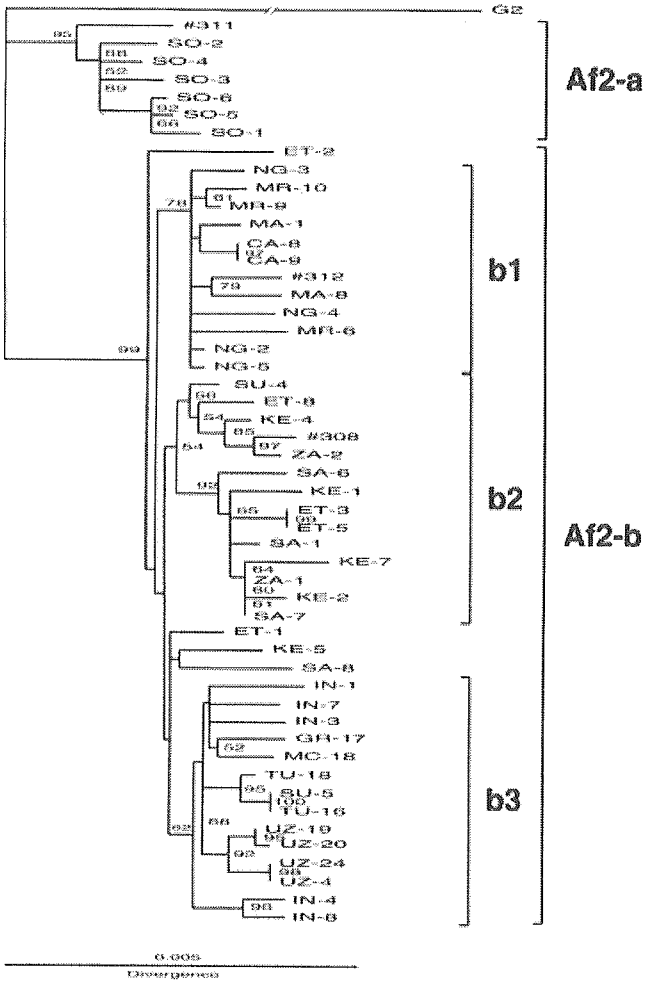

Fig. 1. NJ phylogenetic tree relating 51 complete Af2 sequences. Phylogenetic tree was constructed from complete sequences, excluding regulatory sequences, using NJ method. An isolate (G2; Kato et al., 2000) belonging to type A was used as outgroup. Phylogenetic tree was visualized using TREEVIEW program (Page, 1996). Symbols for sequences are shown in Table 1 and elsewhere (Sugimoto et al., 2002). Numbers at nodes in tree indicate BPs (percent) obtained with 1,000 replicates (only those ≥50% are shown). Af2 subgroups are indicated. A scale for nucleotide divergence is given at the bottom.

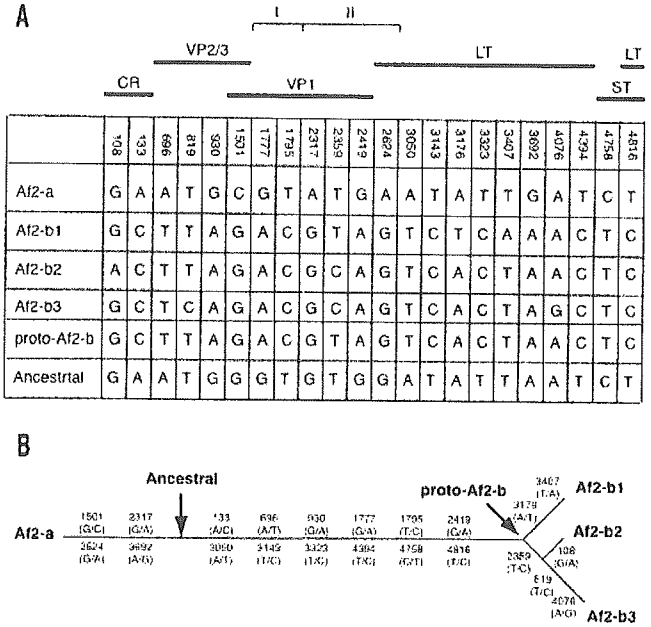

Fig. 2. A: Nucleotide variations in viral genome among four Af2 subgroups. Nucleotides shown are those at positions where a difference was found among consensus sequences of Af2-a, -b1, -b2, and -b3. Proto-Af2-b is a hypothetical sequence that has at most positions common nucleotides among Af2-b subgroups, with consensus nucleotides among all four Af2 subgroups at positions showing variations. Ancestral state for each position was given by consensus of five non-Af2 complete sequences, including G2 (genotype EU-a), MR-7 (genotype EU-b), N4 (genotype B1-c), and GH-1 (genotype Af1). Nucleotide numbers are those of Af2 isolates starting at midpoint of origin of replication and proceeding clockwise. I, type-determining region; II, IG region. **B:** Schematic representation of divergence of Af2. Based on data in A, possible pathways generating Af2-a and various Af2-b subgroups (Af2-b1, -b2, and -b3) are shown. Nucleotide substitutions were located along each branch. Ancestral Af2 and proto-Af2-b are indicated by arrows.

other isolates, including numbers 308 and 312. As number 311 was previously classified as type 3B and numbers 308 and 312 as type 3A, Af2-a and -b corresponded to types 3B and 3A, respectively.

Af2-b further split into three major subclusters (designated Af2-b1, -b2, and -b3) with medium BPs (78%, 54%, and 62% for Af2-b1, -b2, and -b3, respectively), although four isolates (SA-8, ET-1, ET-2, and KE-5) were not included in these subclusters. Af2-b2 and -b3 were weakly combined, with a lower BP (<50%).

Subclassification of Af2 isolates based on single-nucleotide polymorphisms

Partial DNA sequences (i.e., the VT-intergenic sequences and the type-determining sequences within the VP1 gene) were reported for many Af2 isolates, although their complete genomic sequences remain undetermined. To detect SNPs with which to subclassify these isolates, we aligned the 53 complete Af2 DNA sequences (Table 1) using CLUSTAL W. We identified nucleotide variations among Af2 subgroups at 22 positions distributed throughout the genome (Fig. 2A). Sixteen nucleotide substitutions were found between Af2-a and -b, and 3–5 substitutions among Af2-b1, -b2, and -b3. A comparison at each position between the Af2-b subgroups generated the proto-Af2-b sequence. The ancestral state for each position was obtained from a consensus of four complete non-Af2 sequences, including G2 (genotype EU-a), MR-7 (genotype EU-b), N4 (genotype B1-c), and GH-1 (genotype Af1) (the hypothetical sequence with ancestral states at all positions was designated the ancestral Af2 sequence). A comparison between the ancestral sequence and Af2-a (or proto-Af2-b) allowed us to locate not only the origin of Af2 but also nucleotide substitutions on branches extending from the origin to Af2-a and proto-Af2-b (Fig. 2B). We found that four and 12 substitutions occurred along the branches extending from the origin to Af2-a and proto-Af2-b, respectively (Fig. 2B). Furthermore, it was suggested that proto-Af2-b first diverged into Af2-b1 and non-Af2-b1, the latter further diverging into Af2-b2 and -b3 (Fig. 2B).

TABLE 2. Classification of Af2 isolates based on partial sequences[1]

Geographic or ethnic origins	Total no. of isolates	No. of isolates classified as		Reference[2]
		Af2-a	Af2-b	
Africa				
CAR (Bangui)	3	0	3	a
CAR (Pygmy)	2	1	1	b
Ethiopia (Addis Ababa)	3	0	3	a
Kenya (Nairobi)	3	0	3	a
Mauritania (Nouakchott)	6	0	6	a
Morocco (Fes/Ifrane)	6	0	6	a
Niger (Tessaoua)	4	0	4	a
Sudan (Khartoum)	7	0	7	a
Zambia (Lusaka)	2	1	1	a
Asia				
China (Urumqi)	1	0	1	c
India (Varanasi)	9	0	9	a
Saudi Arabia (Riyadh)	8	0	8	a
Turkey (Ankara)	3	0	3	a
Uzbekistan (Tashkent)	1	0	1	d
Europe				
Spain (Badalona/Gitano[3])	1	0	1	e
Italy (North)	1	0	1	f
Total	60	2	58	

[1] Af2 isolates were classified according to nucleotide variations in IG region or VP1 gene (see Fig. 2).
[2] a, Sugimoto et al., 1997; b, Chima et al., 1998; c, Guo et al., 2001; d, Saruwatari et al., 2002; e, Agostini et al., 2001a; f, Pagani et al., 2003.
[3] Immigrants from northwest India.

Both the VT-intergenic sequences and the type-determining sequences within the VP1 gene included a few single-nucleotide polymorphisms (SNPs) useful for classifying Af2 isolates as Af2-a or -b (the subclassification of Af2-b into Af2-b1, -b2, or -b3 was not possible with reported partial sequences) (Fig. 2A). Based on these SNPs, we subclassified 60 Af2 isolates worldwide into Af2-a or -b (Table 2). Thus, two additional Af2-a isolates were detected in the Central African Republic (CAR) and Zambia, but the majority belonged to Af2-b.

Geographic distribution of Af2 subgroups

The domains of Af2 subgroups are shown on a map (Fig. 3) according to the subgroup classification based on the phylogenetic analysis of 51 complete JCV DNA sequences (Fig. 1) supplemented with the subclassification of Af2 based on SNPs (Table 2). The following domains of Af2-a and -b are evident from Figure 3. Af2-a occupied a narrow domain including mainly southern Africa. In contrast, Af2-b occupied a wide domain encompassing almost all of Africa, excluding southern Africa, and neighboring regions of Asia and Europe. These domains partially overlapped each other at their boundaries.

As described above, proto-Af2-b apparently split into three subgroups, Af2-b1–b3. Each of these subgroups also exhibited a distinct domain (Fig. 3). Thus Af2-b1 occupied a domain from central to western Africa (CAR, Niger, Mauritania, and Morocco); Af2-b2 occupied a domain from eastern Africa (Zambia, Tanzania, Kenya, and Ethiopia) and a part of western Asia (Saudia Arabia); and Af2-b3 occupied western, southern, and central-western Asia (Turkey, India, and Uzbekistan) and southeastern Europe (Greece and Macedonia).

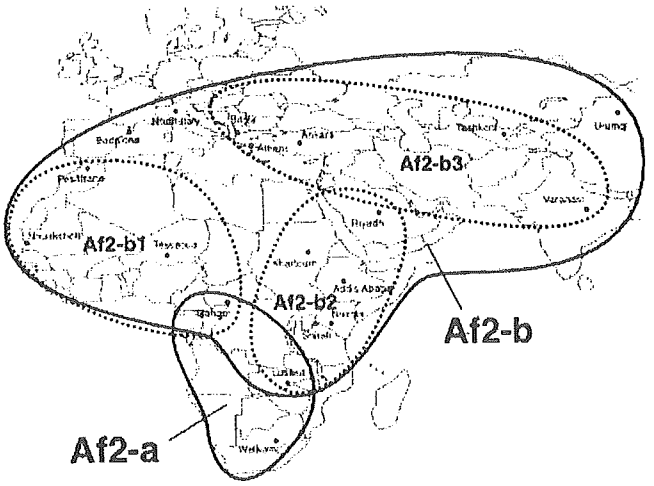

Fig. 3. Map showing geographic domains of Af2 subgroups. Sites of urine collection are indicated by dots. Sites where same subgroup was detected are bounded by lines.

Time scale for divergence of Af2

Assuming that JCV coevolved with human populations (i.e., the first divergence of JCV occurred 100,000 years ago), Hatwell and Sharp (2000) and Sugimoto et al. (2002) estimated the rate of synonymous substitutions to be 4×10^{-7} per synonymous site per year. Using this rate, we attempted to elucidate the time scale of the divergence of types Af2-a and -b. We calculated the average K_s values (synonymous substitutions per synonymous site) between Af2-a and Af2-b1, between Af2-a and Af2-b2, and between Af2-a and Af2-b3 for each of the genes for VP1, VP2, and large T antigen (LTag) (Table 3). The means of these K_s values, calculated with the weight for the length of each gene, were 0.026, 0.023, and 0.024 synonymous substitutions per synonymous site between Af2-a and Af2-b1, between Af2-a and Af2-b2, and between Af2-a and Af2-b3, respectively. According to the rate of synonymous substitutions (4×10^{-7}/site/year) (Hatwell and Sharp, 2000; Sugimoto et al., 2002), we estimated that the divergence between Af2-a and Af2-b1, between Af2-a and Af2-b2, and between Af2-a and Af2-b3 occurred 32,500, 28,800, and 30,000 years ago, respectively.

Similarly, we estimated the date of the split of Af2-b into various subclusters. As described above, we estimated the average K_s values for each of the three genes (VP1, VP2, and LTag) between isolates belonging to Af2-b1 and Af2-b2, between isolates belonging to Af2-b1 and Af2-b3, and between isolates belonging to Af2-b2 and Af2-b3 (Table 3). From the means of these K_s values and the rate of synonymous substitutions (4×10^{-7}/site/year), we estimated that the divergence between Af2-b1 and Af2-b2, between Af2-b1 and Af2-b3, and between Af2-b2 and Af2-b3 occurred 14,000, 15,000, and 10,000 years ago, respectively.

TABLE 3. Synonymous nucleotide substitutions among clusters or subclusters within genotype Af2[1]

Gene	Af2-a vs. Af2-b1	Af2-a vs. Af2-b2	Af2-a vs. Af2-b3	Af2-b1 vs. Af2-b2	Af2-b1 vs. Af2-b3	Af2-b2 vs. Af2-b3
VP1	0.028	0.035	0.033	0.011	0.009	0.011
VP2	0.013	0.011	0.015	0.002	0.006	0.004
LTag	0.031	0.022	0.024	0.015	0.017	0.008
Mean[2]	0.026	0.023	0.024	0.011	0.012	0.008

[1] For three genes, estimated numbers of synonymous substitutions per synonymous site as averages of comparisons among five Af2-a (SO-1–5), five Af2-b1 (CA-9, MA-8, MR-6, and NG-3 and -5), five Af2-b2 (ET-5 and -8, KE-1 and -2, and ZA-2), and five Af2-b3 (GR-17, IN-4, TU-16, UZ-4, and -19).
[2] Calculated with weight for length of each gene.

DISCUSSION

The genetic variations in human mtDNA and the Y chromosome were the basis for a hypothesis that a single out-of-Africa migration to Asia, followed by subsequent migrations across Eurasia, generated European and Asian populations, with further migrations to the Americas and Oceania producing Native American and Oceanic populations (Cavalli-Sforza and Feldman, 2003; Forster, 2004, and references therein). Genetic variations in JCV, however, suggested a significantly different out-of-Africa scenario. The phylogeny of JCV is characterized by the presence of three superclusters (types A, B, and C) that generated various genotypes of JCV in the world (Sugimoto et al., 2002). Type A generated the major European genotypes (EU-a and -b); type B generated the major African (Af2), a minor European (B1-c), most Asian (B1-a, -b, and -d, B2, CY, MY, and SC) and Oceanic genotypes (2E, 8A, and 8B); and type C generated a minor African genotype (Af1) (Yogo et al., 2004, and references therein). Af1 probably represents an ancestral genotype of JCV from which the type-A supercluster was directly generated (Pavesi, 2003). Thus, the phylogenetic relationships among JCV genotypes suggest that the migration "out of Africa" was actually comprised of two major human migrations, one carrying an ancestral genotype belonging to type A, and the other carrying an ancestral genotype belonging to type B.

Nevertheless, regarding human dispersals within Africa, JCV phylogeny and genetic approaches using human mtDNA and the Y chromosome came to similar conclusions, albeit with minor differences. The L0 and L1 haplogroups of human mtDNA as well as the A and B haplogroups of human Y chromosome are thought to represent the earliest African dispersals (Forster, 2004; Jobling and Tyler-Smith, 2003; Cavalli-Sforza and Feldman, 2003, and references therein). Indeed, several clades of haplogroups L0 and L1 were interpreted as the surviving footprints of an early spread across Africa predating the onset of major African reexpansion (see below) (Watson et al., 1997; Salas et al., 2004). Af1, the possible ancestral type of JCV (Pavesi, 2003), may represent the expansion of the earliest Africans.

African reexpansions were studied in some detail using haplogroups of mtDNA. Watson et al. (1997) suggested that early African expansions 60,000–80,000 years ago and late expansions about 52,000 and 19,000 years ago could be traced using distinct mtDNA haplogroups. The Y-chromosome haplogroup E was implicated in African reexpansion (Jobling and Tyler-Smith, 2003, and references therein). According to a phylogenetic tree relating 65 JCV isolates worldwide (Sugimoto et al., 2002), the first split in the type-B supercluster generated Af2 and non-Af2. The former diverged into the Af2 subgroups described in this study, and the latter diverged into B1-c (a minor European genotype) and all major Asian, Oceanian, and Native American genotypes. The first split in type B was estimated to have occurred about 50,000 years ago (Sugimoto et al., 2002). We assume that Af2 subgroups identified in this study represent late human dispersals within Africa with some migrations to neighboring regions of Asia and Europe, whereas the descendants of non-Af2 (i.e., the type-B genotypes excluding Af2) represent a major out-of-Africa expansion to Eurasia, Oceania, and the Americas. We estimated that the division of Af2 into Af2-a and -b occurred 30,000 years ago, and that the division of proto-Af2-b into Af2-b1, -b2, and -b3 occurred 10,000–20,000 years ago.

The most recent widespread demographic shift within the continent was probably the Bantu dispersal, originating in western Africa 3,000–4,000 years ago, spreading both east and south (Salas et al., 2002, and references therein). In the present study, we analyzed samples from Zambia, Tanzania, Kenya, and South Africa where Bantus dispersed, and detected Af2-a and -b2. Chima et al. (1998) detected Af1, rather than Af2, in all four samples from a Bantu group within CAR. The presence of various genotypes of JCV in regions to which Bantus expanded may have been caused by admixture between Bantus and various indigenous populations. Indeed, it was suggested that the southeast Bantu-speakers have a composite origin in the maternal line of descent (Salas et al., 2002).

The Af2-b3 domain encompasses northern India, Turkey, Greece, Macedonia, and Uzbekistan. Most of these countries, excluding Uzbekistan, are included in the eastern domain of Indo-European languages (Ruhlen, 1987). People originally carrying Af2-b3 would have spoken a unique African language, but after migrating out of Africa, they might have intermixed with a population of people speaking an Indo-European language. For sociopolitical reasons, the latter language would have been adopted as the major language in the admixed population. However, when this admixed population migrated into Central Asia (e.g., Uzbekistan), the Indo-European language might have become extinct as a result of adaptation to the Central Asian society.

ACKNOWLEDGMENTS

We are grateful to all urine donors. The helpful suggestions of the reviewers are gratefully acknowledged.

LITERATURE CITED

Agostini HT, Ryschkewitsch CF, Stoner GL. 1996. Genotype profile of human polyomavirus JC excreted in urine of immunocompetent individuals. J Clin Microbiol 34:159–164.

Agostini HT, Ryschkewitsch CF, Brubaker GR, Brubaker GR, Shao J, Stoner GL. 1997a. Five complete genomes of JC virus type 3 from Africans and African Americans. Arch Virol 142: 637–655.

Agostini HT, Yanagihara R, Davis V, Ryschkewitsch CF, Stoner GL. 1997b. Asian genotypes of JC virus in Native Americans and in a Pacific island population: markers of viral evolution and human migration. Proc Natl Acad Sci USA 94:14542–14546.

Agostini HT, Deckhut A, Jobes DV, Girones R, Schlunck G, Prost MG, Frias C, Perez-Trallero E, Ryschkewitsch CF, Stoner GL. 2001a. Genotypes of JC virus in East, Central and Southwest Europe. J Gen Virol 82:1221–1231.

Agostini HT, Jobes DV, Stoner GL. 2001b. Molecular evolution and epidemiology of JC virus. In: Khalili K, Stoner GL, editors. Human polyomaviruses: molecular and clinical perspectives. New York: John Wiley & Sons. p 491–526.

Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. Nat Genet [Suppl] 33:266–275.

Chesters PM, Heritage J, McCance DJ. 1983. Persistence of DNA sequences of BK virus and JC virus in normal human tissues and in diseased tissues. J Infect Dis 147:676–684.

Chima SC, Ryschkewitsch CF, Stoner GL. 1998. Molecular epidemiology of human polyomavirus JC in the Biaka Pygmies and Bantu of Central Africa. Mem Inst Oswaldo Cruz 93:615–623.

Cole CN, Conzen SD. 2001. *Polyomaviridae*: the viruses and their replication. In: Knipe DM, Howley PM, editors. Field's virology, 4th ed. Philadelphia: Lippincott Williams & Wilkins. p 2141–2173.

Dubois V, Moret H, Lafon ME, Brodard V, Icart J, Ruffault A, Guist'hau O, Buffet-Janvresse C, Abbed K, Dussaix E, Ingrand D. 2001. JC virus genotypes in France: molecular epidemiology and potential significance for progressive multifocal leukoencephalopathy. J Infect Dis 183:213–217.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Forster P. 2004. Ice ages and the mitochondrial DNA chronology of human dispersals: a review. Philos Trans R Soc Lond [Biol] 359:255–264.

Frisque RJ, Bream GL, Cannella MT. 1984. Human polyomavirus JC virus genome. J Virol 51:458–469.

Greenberg JH. 1963. The languages of Africa. Bloomington, IN: Indiana University Press.

Guo J, Kitamura T, Ebihara H, Sugimoto C, Kunitake T, Takehisa J, Na YQ, Al-Ahdal MN, Hallin A, Kawabe K, Taguchi F, Yogo Y. 1996. Geographical distribution of the human polyomavirus JC virus type A and B and isolation of a new type from Ghana. J Gen Virol 77:919–927.

Guo J, Sugimoto C, Kitamura T, Ebihara H, Kato A, Guo Z, Liu J, Zheng SP, Wang YL, Na YQ, Suzuki M, Taguchi F, Yogo Y. 1998. Four geographically distinct genotypes of JC virus are prevalent in China and Mongolia: implications for the racial composition of modern China. J Gen Virol 79:2499–2505.

Guo Z, Zheng SP, Sugimoto C, Wang YL, Zheng H-Y, Takasaka T, Kitamura T, Guo J, Yogo Y. 2001. JC virus genotypes in northwestern China: implications for its population history. Anthropol Sci 109:203–212.

Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL. 2001. Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol 18:1189–1203.

Hatwell JN, Sharp PM. 2000. Evolution of human polyomavirus JC. J Gen Virol 81:1191–1200.

Jobes DV, Chima SC, Ryschkewitsch CF, Stoner GL. 1998. Phylogenetic analysis of 22 complete genomes of the human polyomavirus JC virus. J Gen Virol 79:2491–2498.

Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612.

Kato A, Kitamura T, Sugimoto C, Ogawa Y, Nakazato K, Nagashima K, Hall WW, Kawabe K, Yogo Y. 1997. Lack of evidence for the transmission of JC polyomavirus between human populations. Arch Virol 142:875–882.

Kato A, Sugimoto C, Zheng H-Y, Kitamura T, Yogo Y. 2000. Lack of disease-specific amino acid changes in the viral proteins of JC virus isolates from the brain with progressive multifocal leukoencephalopathy. Arch Virol 145:2173–2182.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120.

Kitamura T, Aso Y, Kuniyoshi N, Hara K, Yogo Y. 1990. High incidence of urinary JC virus excretion in nonimmunosuppressed older patients. J Infect Dis 161:1128–1133.

Kitamura T, Kunitake T, Guo J, Tominaga T, Kawabe K, Yogo Y. 1994. Transmission of the human polyomavirus JC virus occurs both within the family and outside the family. J Clin Microbiol 32:2359–2363.

Kitamura T, Sugimoto C, Kato A, Ebihara H, Suzuki M, Taguchi F, Kawabe K, Yogo Y. 1997. Persistent JC virus (JCV) infection is demonstrated by continuous shedding of the same JCV strains. J Clin Microbiol 35:1255–1257.

Kunitake T, Kitamura T, Guo J, Taguchi F, Kawabe K, Yogo Y. 1995. Parent-to-child transmission is relatively common in the spread of the human polyomavirus JC virus. J Clin Microbiol 33:1448–1451.

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC. 2003. Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA 100:171–176.

Padgett BL, Walker DL. 1973. Prevalence of antibodies in human sera against JC virus, an isolate from a case of progressive multifocal leukoencephalopathy. J Infect Dis 127:467–470.

Pagani E, Delbue S, Mancuso R, Borghi E, Tarantini L, Ferrante P. 2003. Molecular analysis of JC virus genotypes circulating among the Italian healthy population. J Neurovirol 9:559–566.

Page RD. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358.

Pavesi A. 2003. African origin of polyomavirus JC and implications for prehistoric human migrations. J Mol Evol 56:564–572.

Ruhlen M. 1987. A guide to the world's languages. Volume 1, classification. Stanford: Stanford University Press.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.

Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, Macaulay V, Carracedo A. 2002. The making of the African mtDNA landscape. Am J Hum Genet 71:1082–1111.

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74:454–465.

Saruwatari L, Sugimoto C, Kitamura T, Ohno N, Sakai E, Shresta P, Hoa BK, Phi PTP, An HPH, Tuyet NTA, Honjo T, Kobayashi N, Zheng H-Y, Takasaka T, Yogo Y. 2002. Asian domains of four major genotypes of JC virus, Af2, B1-b, CY and SC. Arch Virol 147:1–10.

Sugimoto C, Hasegawa M, Kato A, Zheng H-Y, Ebihara H, Taguchi F, Kitamura T, Yogo Y. 2002. Evolution of human polyomavirus JC: implications for the population history of humans. J Mol Evol 54:285–297.

Sugimoto C, Kitamura T, Guo J, Al-Ahdal MN, Shchelkunov SN, Otova B, Ondrejka P, Chollet J-Y, El-Safi S, Ettayebi M, Grèsenguet G, Kocagöz T, Chaiyarasamee S, Thant KZ, Thein S, Moe K, Kobayashi N, Taguchi F, Yogo Y. 1997. Typing of urinary JC virus DNA offers a novel means of tracing human migrations. Proc Natl Acad Sci USA 94:9191–9196.

Suzuki M, Zheng H-Y, Takasaka T, Sugimoto C, Kitamura T, Beutler E, Yogo Y. 2002. Asian genotypes of JC virus in Japanese-Americans suggest familial transmission. J Virol 76: 10074–10078.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.

Tominaga T, Yogo Y, Kitamura T, Aso Y. 1992. Persistence of archetypal JC virus DNA in normal renal tissue derived from tumor-bearing patients. Virology 186:736–741.

Watson E, Forster P, Richards M, Bandelt HJ. 1997. Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61:691–704.

Yogo Y, Sugimoto C. 2001. The archetype concept and regulatory region rearrangement. In: Khalili K, Stoner GL, editors. Human polyomaviruses: molecular and clinical perspectives. New York: John Wiley & Sons. p 127–148.

Yogo Y, Iida T, Taguchi F, Kitamura T, Aso Y. 1991. Typing of human polyomavirus JC virus on the basis of restriction fragment length polymorphisms. J Clin Microbiol 29:2130–2138.

Yogo Y, Sugimoto C, Zheng H-Y, Ikegaya H, Takasaka T, Kitamura T. 2004. JC virus genotyping offers a new paradigm in the study of human populations. Rev Med Virol 14:179–191.

Zheng H-Y, Kitamura T, Takasaka T, Chen Q, Yogo Y. 2004. Unambiguous identification of JC virus strains transmitted from parents to children. Arch Virol 149:261–273.

JC polyomavirus lineages common among Kiribati Islanders: implications for human dispersal in the Pacific

TOMOKAZU TAKASAKA[1], NOBUTAKA OHTA[2], HUAI-YING ZHENG[1,3], HIROSHI IKEGAYA[4], KOICHI SAKURADA[4], TADAICHI KITAMURA[1], YOSHIAKI YOGO[1]*

[1]*Department of Urology, Faculty of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655 Japan*
[2]*Yaizu City Hospital, Shizuoka, Japan*
[3]*Japanese Foundation for AIDS Prevention, Tokyo, Japan*
[4]*National Research Institute of Police Science, Chiba, Japan*

Received 19 July 2005; accepted 10 November 2005

Abstract Although *JC polyomavirus* (JCPyV) genotyping is a promising method in clarifying the peopling of the Pacific, the distribution of JCPyV lineages has been elucidated only in a few Pacific populations. To clarify JCPyV lineages of the Kiribati Islanders, we collected urine samples from 48 Kiribati fishermen (mainly from the Gilbert Islands), who called in at a Japanese port for unloading. From the urine samples, we amplified the 610-bp IG region (VT-intergenic region) of the viral genome using the polymerase chain reaction. We obtained and sequenced IG fragments from 35 samples. From the resultant sequences, together with reference IG sequences, a neighbor-joining phylogenetic tree was constructed to classify the JCPyV isolates into lineages. We detected 2E (one of the major Pacific lineages) in 27 samples (77%), SC (the major Southeast Asian lineage) in seven samples (20%), and EU-a (one of the major European lineages) in one sample (3%) (the rare EU-a isolate was probably recently introduced to Kiribati by Europeans). A phylogenetic analysis based on complete viral DNA sequences revealed that the SC isolates in Kiribati belonged to a new sublineage (named SC-g) within the SC lineage. The present findings agree with the view that the 2E lineage of JCPyV accompanied the Austronesian dispersals in the Pacific. In addition, our findings suggest that ancient Southeast Asians, carrying SC-g, also migrated to the central Pacific islands.

Key words: *JC polyomavirus*, human dispersals, Pacific populations, Austronesians, Kiribati, Micronesia

Introduction

JC polyomavirus [abbreviated JCPyV according to the ICTVdB Index of Viruses (Universal Virus Database of the International Committee on Taxonomy of Viruses, 2005)] causes an opportunistic infection in the central nervous system, known as progressive multifocal leukoencephalopathy (PML) (Padgett et al., 1971). This virus, however, is ubiquitous in human populations, infecting children asymptomatically, and then persisting in the kidney and urinary tract (Ikegaya et al., 2004; Yogo et al., 2004 and references therein; Boldorini et al., 2005). Renal JCPyV is reactivated in most adults, who excrete progeny viruses in the urine. This virus is usually transmitted from parents to children through long-term cohabitation (Kunitake et al., 1995; Suzuki et al., 2002; Zheng et al., 2004a), and rarely transmitted among human populations unless these human populations are closely intermixed (Kato et al., 1997).

JCPyV strains in the world can be classified into more than 10 major lineages, each occupying a unique geographical domain (Sugimoto et al., 1997; Cui et al., 2004; Yogo et al., 2004), suggesting that the evolution of JCPyV occurred in association with human populations. Although all JCPyV lineages potentially induce PML, it has been suggested that one JCPyV lineage (B1-c or 2B) more readily induces PML than the other lineages (Agostini et al., 1998). Nevertheless, it seems unlikely that the geographic distribution of JCPyV lineages was affected by the potential difference in pathogenic capacity among JCPyV lineages, because (1) PML was extremely rare before the epidemic of human immunodeficiency syndrome, and (2) JCPyV variants in PML brains never return to the human population (Yogo and Sugimoto, 2001). JCPyV has evolved almost as fast as the human mtDNA control region that has frequently been used for anthropological studies (Hatwell and Sharp, 2000; Sigurgardottir et al., 2000; Sugimoto et al., 2002a). Thanks to the high rate of evolution of JCPyV, together with its host-linked mode, JCPyV serves as a highly sensitive and reliable marker in tracing the history of human populations (Stoner et al., 2000; Agostini et al., 2001; Yogo et al., 2004). Indeed, JCPyV has been used as a marker to elucidate the origins of various ethnic populations (Cui et al., 2004; Miranda et al., 2004; Yogo et al., 2004 and references therein; Zheng et al., 2004b, 2005a, 2005b; Ikegaya et al., 2005; Takasaka et al., 2005).

* Corresponding author. e-mail: yogo-tky@umin.ac.jp
phone: +81-3-5800-8662; fax: +81-3-5800-8917

Published online 28 January 2006
in J-STAGE (www.jstage.jst.go.jp) DOI: 10.1537/ase.050719

Stoner and his colleagues explored JCPyV lineages in Guam and Papua New Guinea (PNG) (Ryschkewitsch et al., 2000; Jobes et al., 2001). These studies led the authors to discover two major Pacific lineages of JCPyV 2E and 8. In addition, they found that a lineage, 7A (or SC according to our nomenclature), spread across Southeast Asia and South China, also occurs in Guam at a significantly high rate.

To extend the findings noted above, Stoner and his colleagues examined JCPyV lineages in various islander populations of the western Pacific (Yanagihara et al., 2002). The authors collected urine samples from athletes and coaches participating in the XIth South Pacific Games in Guam in 1999. They selected nine JCPyV-positive samples whose donors were derived from Fiji, Kiribati, New Caledonia, PNG, the Solomon Islands, Tonga, Vanuatu, and Wallis and Futuna. They used the selected samples for amplification of the whole viral genome followed by sequencing, and found that lineages 2E and 8 are widely distributed in western Pacific populations. Lineage 8 was divided into two distinct lineages, 8A and 8B, and it was found that 8A was confined to PNG, while 8B was widespread in the Pacific. Based on these and previous findings (Ryschkewitsch et al., 2000; Jobes et al., 2001), the authors suggest three broad human movements to the Pacific: (1) ancient migrations before 40000 years ago accompanied by 8A and 8B, (2) much later migrations beginning 5000 years ago carrying 2E, and (3) relatively recent movements carrying 7A directly from Southeast Asia or South China (Yanagihara et al., 2002). Movements carrying 2E were postulated to correlate with the arrival of Austronesian-language speakers (Yanagihara et al., 2002).

The inference about human dispersals in the Pacific is potentially of interest, but it should be noted that it was based on fragmentary data. Indeed, only one or a few JCPyV isolates were examined for most of the Pacific islands (Yanagihara et al., 2002), although a significantly large number of isolates were examined for a few islands (Ryschkewitsch et al., 2000; Jobes et al., 2001).

Recently, we had an opportunity to study JCPyV isolates of Kiribati men, mainly from the Gilbert Islands (see Figure 1 for the location of Kiribati). We here report the JCPyV lineages that were observed, and discuss their implications for human dispersal patterns of the Pacific.

Materials and Methods

Urine samples

Urine samples were collected from 48 Kiribati fishermen aged 20–36 (average 25.5) years. Sampling was done when these men docked at a Japanese port (the port of Yaizu) for unloading, and with informed consent. We assume that the results obtained from urine samples of males reflects the

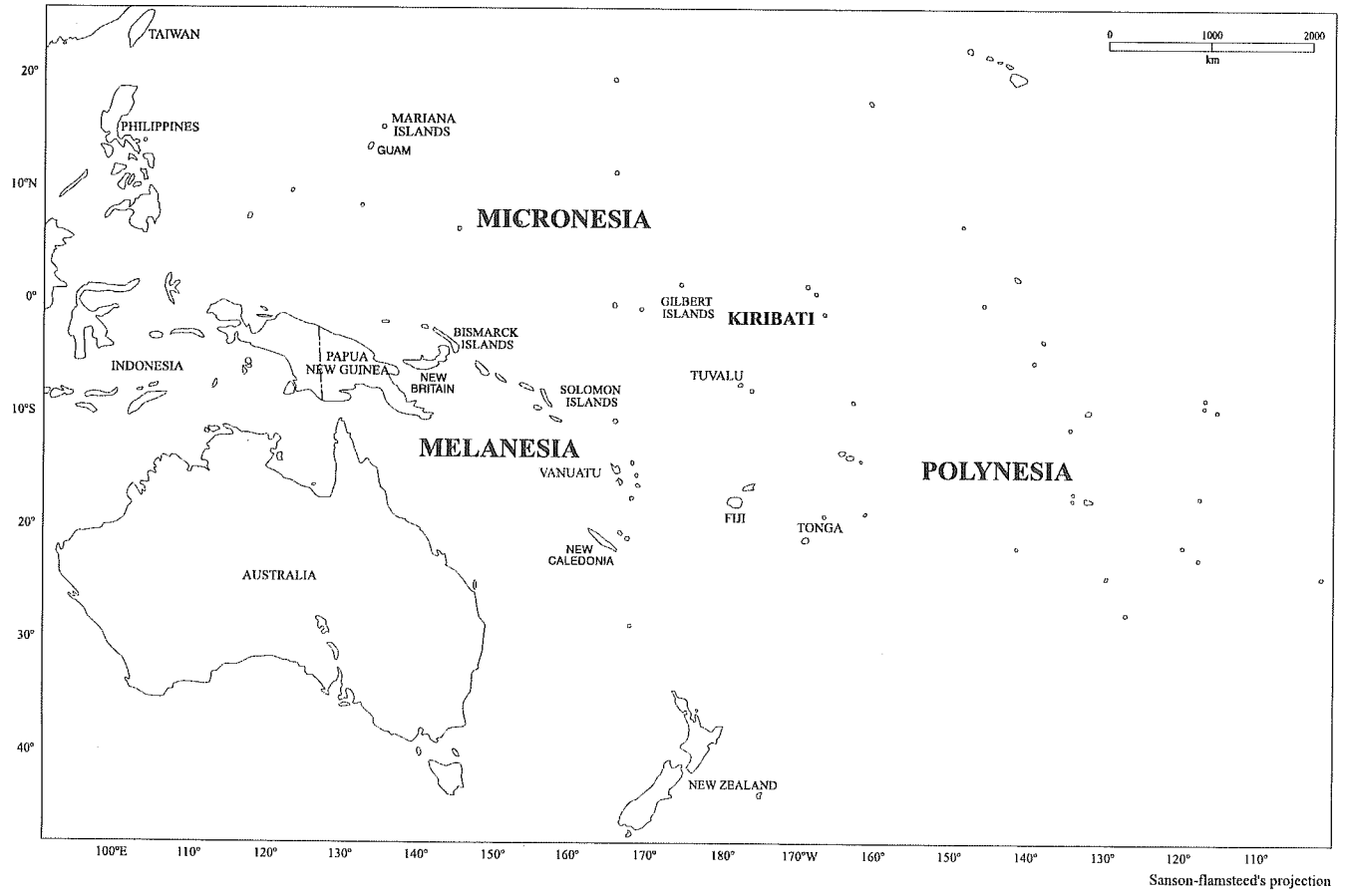

Figure 1. A map of Oceania showing sites at which urine samples were collected in this and previous studies.

general JCPyV profile of the population, as no gender gap has been reported in the transmission, persistence, and shedding of JCPyV (Kitamura et al., 1994, 1997; Kunitake et al., 1995; Agostini et al., 1996). The urine donors were indigenous inhabitants of Abaiang ($n = 4$), Abemama ($n = 2$), Aranuka ($n = 1$), Arorae ($n = 2$), Banaba ($n = 2$), Beru ($n = 2$), Butaritari ($n = 4$), Kuria ($n = 5$), Maiana ($n = 1$), Makin ($n = 3$), Onotoa ($n = 2$), Nikunau ($n = 2$), and Tarawa ($n = 18$) islands. Most of these islands, except for Banaba, belong to the Gilbert Islands. Urine samples were collected in 10 mM EDTA (pH 8.0), and sent to the Department of Urology, Faculty of Medicine, The University of Tokyo, where DNA was extracted as described previously (Kitamura et al., 1990).

PCR

From extracted DNA, the 610-bp IG region (VT-intergenic region) of the viral genome was amplified by PCR using primers P1 and P2 (Kunitake et al., 1995). The IG region encompasses the 3′ terminal regions of both the T-antigen and VP1 genes, and was established to be a region of the JCPyV genome that contains abundant type-determining sites (Ault and Stoner, 1992). The total reaction volume of 50 μl contained 1.25 units of HotStar Taq DNA polymerase (QIAGEN GmbH, Hilden, Germany), 200 μM of each dNTP, 1.5 mM $MgCl_2$, 0.5 μM primers, and a PCR Buffer supplied by the manufacturer. A 2.5-μl volume of sample DNA was added to the PCR mixture. After enzyme activation at 95°C for 15 min, PCR amplification was performed for 50 cycles. The cycle profile was 94°C for 1 min, 55°C for 1 min, and 72°C for 2 min. Both activation and amplification were carried out in a Thermal Sequencer (Asahi Techno Glass Corporation, Tokyo, Japan).

Cloning and sequencing of amplified fragments

The amplified fragments were digested with a combination of *Hin*dIII and *Pst*I, which excised the IG fragment (Kunitake et al., 1995). The digested DNA was ligated to *Hin*dIII- and *Pst*I-digested, alkaline phosphatase-treated pBluescript II SK (+) (Stratagene, La Jolla, USA), and was used to transform competent *Escherichia coli* HB101 cells (Takara Bio Inc., Otsu, Japan). Recombinant clones containing IG sequences were selected by digestion with a combination of *Hin*dIII and *Pst*I followed by agarose gel electrophoresis. For each PCR product, two recombinant clones carrying IG sequences were purified using a QIAprep 8 Turbo Miniprep kit (QIAGEN), and sequenced using the universal T3 and T7 primers and an ABI PRISM 373S DNA sequencer (Applied Biosystems, Foster City, USA).

Cloning and sequencing entire JCPyV genomes

Entire JCPyV DNAs were cloned into pUC19 at the unique *Bam*HI site as described previously (Yogo et al., 1991). The resultant complete JCPyV DNA clones were prepared using a QIAprep 8 Turbo Miniprep kit (QIAGEN). Purified plasmids were used for a cycle sequencing reaction set up using the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences, Buckinghamshire, England). The primers used were a set reported previously (Agostini et al., 1997), excluding JIG-9, JIG-10, JIG-53, and SEC-11, and four additional primers: W-11 [5′-AGTTTTTGAGGGAACAGAGGAG-3′, corresponding to nucleotide (nt) 283 to 310 in the JCPyV (Mad-1) genome (Frisque et al., 1984)], W-1a (5′-CTTCGCCAGCTGT-CACGTAAGGCTTCTG-3′, corresponding to nt 283–310), and the M13 universal forward and reversal primers. Primers were added to a final concentration of 0.25 pmol/μl in a final reaction volume of 20 μl. The cycling conditions were 25 cycles of 30 s at 96°C, 15 s at 50°C, and 60 s at 60°C. The reaction was terminated at 4°C. Cycle sequencing products were purified on Centri-Sep columns (Princeton Separations, Adelphia, NJ, USA). DNA sequencing was performed using an automated sequencer (ABI PRISM 373S DNA sequencer, Applied Biosystems).

Phylogenetic analysis

The noncoding control region of the JCPyV genome was excluded from phylogenetic analysis, as this region is hypervariable especially in JCPyV isolates derived from the brain and the cerebrospinal fluid of PML patients with PML (Yogo and Sugimoto, 2001). DNA sequences were aligned using CLUSTAL W (Thompson et al., 1994) with a gap opening penalty of 15.00 and gap extension penalty of 6.66. To evaluate the phylogenetic relationships among DNA sequences, we used the neighbor-joining (NJ) method (Saitou and Nei, 1987) using the CLUSTAL W program. Divergences were estimated with Kimura's two-parameter method (Kimura, 1980). To assess the confidence of branching patterns of the NJ tree, bootstrap probabilities (BPs) were estimated with 1,000 bootstrap replicates (Felsenstein, 1985) using CLUSTAL W. BPs larger than 70% were considered to be significant (Hillis and Bull, 1993). A phylogenetic tree was visualized using the TREEVIEW program (Page, 1996).

Results

Using a PCR that amplifies the 610-bp IG region, we detected JCPyV DNA from 35 (73%) of the 48 urine samples collected from Kiribati fishermen (donors from which JCPyV DNA was detected are shown in Table 1). We cloned the amplified IG regions, and sequenced two representative clones for each urine sample. The two sequences were usually identical. However, single nucleotide differences were sometimes detected between the sequences derived from the same urine samples, probably because of errors during PCR; in such cases, we sequenced another clone to obtain the consensus sequence. A preliminary phylogenetic analysis of the 35 IG sequences indicated that 34 Kiribati sequences (KB-1 to KB-19 and KB-21 to KB-35) belonged to Asian lineages within Type B, with a single sequence (KB-20) belonging to Type A (European lineage). We constructed a phylogenetic tree from the Kiribati Type-A IG sequence (KB-20) together with reference Type-A IG sequences (Sugimoto et al., 1997, 2002b). According to the resultant tree (not shown), KB-20 was classified as EU-a, a lineage prevalent in Europe (Sugimoto et al., 2002b). The other Kiribati sequences were subjected to a phylogenetic analysis with reference IG sequences representing Asian and Pacific lineages of JCPyV (Kunitake et al., 1995; Kato et al., 1997; Sugimoto et al.,

Table 1. Origins of JCPyV-positive urine samples and detected isolates

Urine donor	Age in years	Island[1]	Isolate	Lineage	Accession no.[2] for:	
					IG sequence	Complete sequence
1	27	Abaiang	KB-1	2E	AB220648	—[3]
2	21	Kuria	KB-2	2E	AB220649	—
3	21	Abemama	KB-3	2E	AB220650	—
4	36	Tarawa	KB-4	SC-g	AB220651	AB220939
5	26	Tarawa	KB-5	2E	AB220652	—
6	31	Maiana	KB-6	2E	AB220653	—
7	26	Tarawa	KB-7	2E	AB220654	—
8	24	Arorae	KB-8	2E	AB220655	—
10	32	Butaritari	KB-9	2E	AB220656	—
11	24	Banaba	KB-10	2E	AB220657	—
12	32	Tarawa	KB-11	2E	AB220658	—
13	34	Kuria	KB-12	2E	AB220659	—
14	25	Abaiang	KB-13	SC-g	AB220660	AB220940
15	25	Kuria	KB-14	SC-g	AB220661	AB220941
20	23	Arorae	KB-15	2E	AB220662	—
21	25	Onotoa	KB-16	2E	AB220663	—
22	23	Beru	KB-17	2E	AB220664	—
23	25	Abaiang	KB-18	2E	AB220665	—
25	22	Tarawa	KB-19	2E	AB220666	—
26	29	Kuria	KB-20	EU	AB220667	—
27	29	Onotoa	KB-21	2E	AB220668	—
28	24	Makin	KB-22	2E	AB220669	—
29	23	Butaritari	KB-23	SC-g	AB220670	AB220942
30	22	Tarawa	KB-24	SC-g	AB220671	AB220943
31	25	Tarawa	KB-25	2E	AB220672	—
34	23	Butaritari	KB-26	2E	AB220673	—
37	29	Nikunau	KB-27	2E	AB220674	—
39	25	Beru	KB-28	2E	AB220675	—
40	28	Tarawa	KB-29	2E	AB220676	—
42	25	Butaritari	KB-30	2E	AB220677	—
44	23	Tarawa	KB-31	SC[4]	AB220678	—
47	30	Tarawa	KB-32	2E	AB220679	—
48	25	Abaiang	KB-33	SC[4]	AB220680	—
49	25	Tarawa	KB-34	2E	AB220681	—
50	21	Tarawa	KB-35	2E	AB220682	—

[1] All but Banaba belonged to the Gilbert Islands.
[2] GenBank/EMBL/DDBJ accession numbers are shown.
[3] Sequences were not determined.
[4] Not subclassified.

1997, 2002b; Chang et al., 1999; Jobes et al., 2001; Miranda et al., 2003, 2004). On the phylogenetic tree obtained (Figure 2), we found seven sequences in the SC cluster and 27 sequences in the 2E cluster (see Table 1).

To subclassify the SC isolates in Kiribati, we established and sequenced five clones containing complete SC genomes from the samples of Kiribati. The resultant complete SC DNA sequences, together with 40 complete SC sequences reported previously (Saruwatari et al., 2002; Sugimoto et al., 2002a: Takasaka et al., 2004), were used to construct an NJ tree. According to the resultant tree (Figure 3), the SC isolates worldwide were classified into several clusters identified previously (SC-a to -f and SC-x) (Saruwatari et al, 2002; Takasaka et al., 2004) and a new cluster designated SC-g. All five sequences from Kiribati, but none from the other geographic regions, were included in the SC-g cluster. The bootstrap probability (73%) for the SC-g cluster was significantly high (Hills and Bull, 1993). SC-c to -e contained isolates only from Myanmar, and SC-f contained isolates worldwide.

Discussion

The geographic distribution of the JCPyV lineages seen in the Pacific islander populations should help elucidate the history of human settlement in the Pacific region. Excluding those introduced into the Pacific during the last 500 years, the JCPyV lineages detected in the Pacific islander populations include 2E, 8A, 8B and a few sublineages (SC-f, SC-x and SC-g) of the SC lineage (Ryschkewitsch et al., 2000; Jobes et al., 2001; Yanagihara et al., 2002; Miranda et al., 2003, 2004; Takasaka et al., 2004; and this study). We summarize the geographic distribution of these 'Pacific' JCPyV lineages. (1) The 2E lineage predominates in the Micronesian islands (i.e. Guam of the Mariana Islands and Kiribati Islands) where a substantial number of JCPyV isolates were analyzed (Ryschkewitsch et al., 2000; Jobes et al., 2001; this study). Furthermore, this lineage was sporadically identified in Melanesia and Polynesia (Yanagihara et al., 2002). In addition, 2E has been identified in some islands of the Philippines (Miranda et al., 2003, 2004; Takasaka et al., 2004).

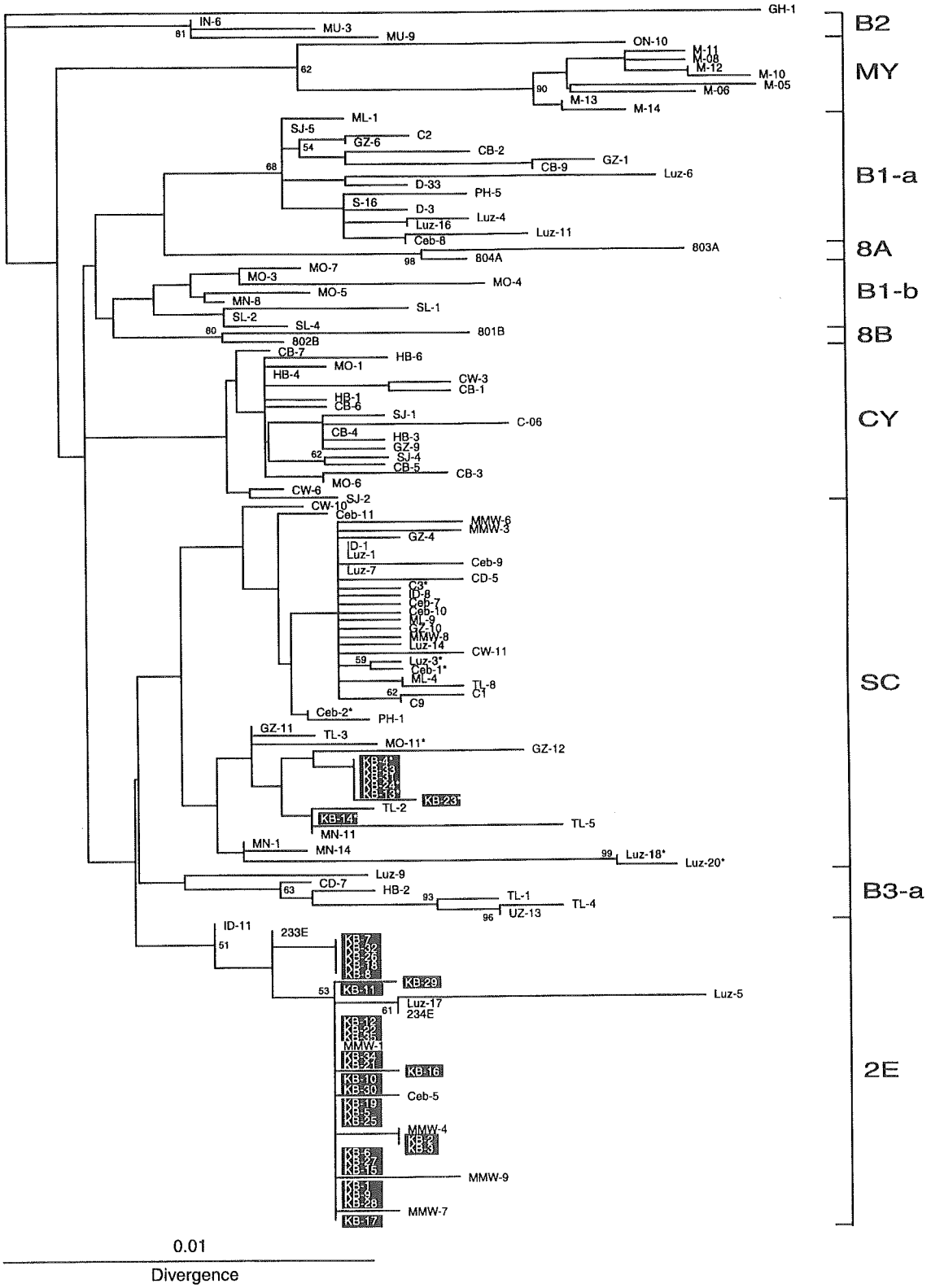

Figure 2. Phylogenetic tree used to classify the detected JCPyV isolates into distinctive lineages. IG sequences in Kiribati together with reference sequences detected previously in Asia and Oceania were used to construct an NJ phylogenetic tree using CLUSTAL W. The phylogenetic tree was visualized using TREEVIEW. The tree was rooted using an isolate (GH-1) belonging to lineage Af1 as the outgroup, as this lineage is probably an ancestral type (Pavesi, 2003). Lineages are indicated on the right of the tree. Asterisks identify isolates (within the SC lineage) analyzed using the whole genome approach (Saruwatari et al., 2002; Takasaka et al., 2004; this study). Symbols for reference sequences were described in references (Kunitake et al., 1995; Kato et al., 1997; Sugimoto et al., 1997; Kitamura et al., 1997; Chang et al., 1999; Jobes et al., 2001); Miranda et al., 2003, 2004. Isolates whose sequences were determined in this study are shown in white on black backgrounds.

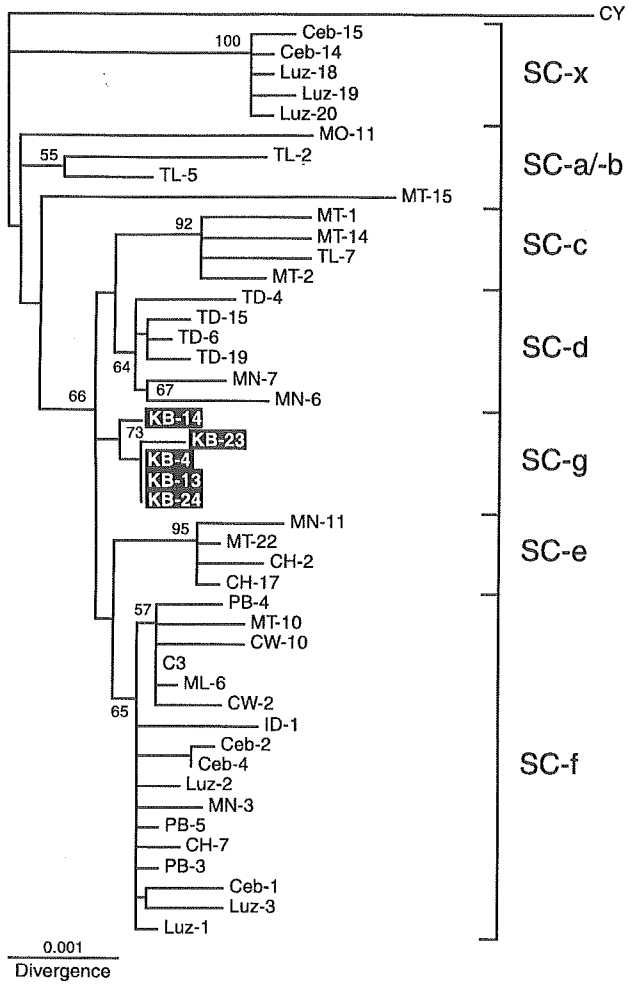

Figure 3. NJ phylogenetic tree relating 65 complete JCPyV (SC) DNA sequences. A phylogenetic tree was constructed from complete sequences, excluding regulatory sequences, using the NJ method. The phylogenetic tree was visualized using the TREEVIEW program. The tree was rooted using an isolate (CY) belonging to lineage CY as the outgroup, as this lineage is a distinct lineage closely related to SC (Sugimoto et al., 2002a). The symbols for sequences, along with their accession numbers, are available in references (Saruwatari et al, 2002; Takasaka et al., 2004). The numbers at nodes in the tree indicate bootstrap probabilities (percent) obtained by 1,000 replicates (only those for major clusters are shown). SC sublineages (SC-a to SC-g and SC-x) are indicated. Isolates whose sequences were determined in this study are shown in white on black backgrounds.

(2) Several isolates with 8A and 8B lineages were first detected in the highlands of PNG and New Britain of the Bismarck Islands (Ryschkewitsch et al., 2000; Jobes et al., 2001), and then 8B isolates were sporadically detected in Melanesia and Polynesia, but 8A isolates were not (Yanagihara et al., 2002). There were no isolates classified as 8A or 8B in island or mainland Southeast Asia. (3) JCPyV isolates belonging to a sublineage (SC-f) of the SC lineage was detected at a higher rate only in Guam (Ryschkewitsch et al., 2000; Jobes et al., 2001). (4) A single isolate belonging to another SC sublineage (SC-x), a sublineage detected among Filipino populations at a low but significant rate (Miranda et al., 2003; Takasaka et al., 2004), was identified in a Hawaiian (Yanagihara et al., 2002). (5) Finally, a new sublineage (SC-g) belonging to the SC lineage was detected at a low but significant rate in the Kiribati islanders (this study).

The above summarized geographic distribution patterns of the Pacific JCPyV lineages together with those of language families in the Pacific (Ruhlen, 1987) suggest that lineage 2E is associated with the dispersal of Austronesian-speaking peoples in the Pacific, while 8A and 8B are associated with that of Papuan-speaking peoples. As genetic diversity among SC isolates is most marked in the southwestern Asian region that includes Myanmar, Thailand, and Southwest China (Saruwatari et al, 2002), it seems reasonable to assume that SC sublineages evolved from an ancestral SC somewhere in mainland Southeast Asia. The present distribution patterns of various SC sublineages in the Pacific probably represent ancient and recent human migrations from Southeast Asia.

There are many genetic studies that have addressed the peopling of the remote Pacific. Two major models have been distinguished by these studies. The first model was proposed by Bellwood (1978, 1991), and posits that Austronesian-speaking and pottery-making farmers, derived from Taiwan or South China, rapidly settled the islands of central and eastern Melanesia and then islands of western Polynesia (e.g. the Solomon Islands) between 1350 and 800 BC. Central Micronesian islands, including the islands of Kiribati, appear to have been settled about 2500–2000 years ago, possibly from Melanesia. Eastern Polynesian islands were the last settled by aceramic populations between 600 and 1250 AD. According to this model, it is assumed that when the proto-Austronesians passed through Melanesia, they did not become very integrated with indigenous Papuan-speaking peoples who had been living in the area for at least 30000 years. The second model proposes that the proto-Polynesians evolved around the center of island Southeast Asia during the late Pleistocene (> 10000 years ago) (Oppenheimer, 1998; Oppenheimer and Richards, 2001). Some geneticists have provided evidence supporting the first model (Cox, 2005; Trejaut et al., 2005), while others have supported the second model (Richards et al., 1998; Capelli et al., 2001; Kayser et al., 2001; Hurles et al., 2002).

To distinguish between the two alternative models explaining the origin and dispersal of Austronesians, we need to identify the homeland of 2E and the extent of intermixture in various Pacific islands between the Austronesian (2E) and Papuan (8A and 8B) JCPyV lineages. The identification of the homeland of 2E requires elucidating the detailed geographic distribution pattern of this lineage in Southeast Asia. We previously detected 2E in modern Filipino populations on Luzon and Cebu Islands at low rates (Miranda et al., 2003; Takasaka et al., 2004) and in an indigenous population (Mamanwa) on Mindanao Island at a higher rate (Miranda et al., 2004). We undoubtedly need to explore this lineage on other Southeast Asian islands, including eastern and western Indonesia, to clarify the overall distribution profile of this lineage on the islands and mainland of Southeast Asia, which will eventually lead to the identification of the homeland of 2E.

At present, a considerable number of JCPyV isolates were examined on only four islands or island groups: PNG,

Guam of the Mariana Islands, New Britain of the Bismarck Islands, and the Gilbert Islands (plus a surrounding island) (Ryschkewitsch et al., 2000; Jobes et al., 2001; this study). 2E was mainly detected on the Gilbert Islands (this study), and 2E and SC-f were equally detected on Guam (Ryschkewitsch et al., 2000; Jobes et al., 2001). There were no isolates with the Papuan JCPyV lineages (8A and 8B) detected on any of these islands. From these observations, we tentatively conclude that an admixture of Austronesian (2E) and Papuan (8A and 8B) JCPyV lineages were rather rare in islander populations of Micronesia. Nevertheless, rarity of admixture remains to be confirmed by future studies performed with large sample sizes on various Pacific islands, especially those in Polynesia.

Acknowledgments

We are grateful to all urine donors. We thank Mrs Yutaka Ito and Keitaro Katayama for help with urine collection. This study was supported in part by grants from the Ministry of Health, Labour and Welfare, Japan.

References

Agostini H.T., Ryschkewitsch C.F., and Stoner G.L. (1996) Genotype profile of human polyomavirus JC excreted in urine of immunocompetent individuals. Journal of Clinical Microbiology, 34: 159–164.

Agostini H.T., Ryschkewitsch C.F., Singer E.J., Baumhefner R.W., and Stoner G.L. (1998) JC virus type 2B is found more frequently in brain tissue of progressive multifocal leukoencephalopathy patients than in urine from controls. Journal of Human Virology, 1: 200–206.

Agostini H.T., Ryschkewitsch C.F., Brubaker G.R., Shao J., and Stoner G.L. (1997) Five complete genomes of JC virus type 3 from Africans and African Americans. Archives of Virology, 142: 637–655.

Agostini H.T., Jobes D.V., and Stoner G.L. (2001) Molecular evolution and epidemiology of JC virus. In: Khalili K. and Stoner G.L. (eds.), Human Polyomaviruses: Molecular and Clinical Perspectives. John Wiley & Sons, New York, pp. 491–526.

Ault G.S. and Stoner G.L. (1992) Two major types of JC virus defined in progressive multifocal leukoencephalopathy brain by early and late coding region DNA sequences. Journal of General Virology, 73: 2669–2678.

Bellwood P. (1978) Man's Conquest of the Pacific. The Prehistory of Southeast Asia and Oceania. Collins, Auckland.

Bellwood P. (1991) The Austronesian dispersal and the origin of languages. Scientific American, 265: 70–75.

Boldorini R., Veggiani C., Barco D., and Monga G. (2005) Kidney and urinary tract polyomavirus infection and distribution: molecular biology investigation of 10 consecutive autopsies. Archives of Pathology and Laboratory Medicine, 129: 69–73.

Capelli C., Wilson J.F., Richards M., Stumpf M.P.H., Gratrix F., Oppenheimer S., Underhill P., Pascali V.L., Ko T.-M., and Goldstein D.B. (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. American Journal of Human Genetics, 68: 432–443.

Chang D., Sugimoto C., Wang M., Tsai R.T., and Yogo Y. (1999) JC virus genotypes in a Taiwan aboriginal tribe (Bunun): implications for its population history. Archives of Virology, 144: 1081–1090.

Cox M.P. (2005) Indonesian mitochondrial DNA and its opposition to a Pleistocene era origin of proto-Polynesians in Island Southeast Asia. Human Biology, 77: 179–188.

Cui X., Wang J.C., Deckhut A., Joseph B.C., Eberwein P., Cubitt C.L., Ryschkewitsch C.F., Agostini H.T., and Stoner G.L. (2004) Chinese strains (Type 7) of JC virus are Afro-Asiatic in origin but are phylogenetically distinct from the Mongolian and Indian strains (Type 2D) and the Korean and Japanese strains (Type 2A). Journal of Molecular Evolution, 58: 568–583.

Felsenstein J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution, 39: 783–791.

Frisque R.J., Bream G.L., and Cannella M.T. (1984) Human polyomavirus JC virus genome. Journal of Virology, 51: 458–469.

Hatwell J.N. and Sharp P.M. (2000) Evolution of human polyomavirus JC. Journal of General Virology, 81: 1191–1200.

Hills D.M. and Bull J.J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology, 42: 182–192.

Hurles M.E., Nicholson J., Bosch E., Renfrew C., Sykes B.C., and Jobling M.A. (2002) Y chromosomal evidence for the origins of Oceanic-speaking peoples. Genetics, 160: 289–303.

Ikegaya H., Iwase H., and Yogo Y. (2004) Detection of identical JC virus DNA sequences in both human kidneys. Archives of Virology, 149: 1215–1220.

Ikegaya H., Zheng H.-Y., Saukko P.J., Varesmaa-Korhonen L., Hovi T., Vesikari T., Suganami H., Takasaka T., Sugimoto C., Ohasi Y., Kitamura T., and Yogo Y. (2005) Genetic diversity of JC virus in the Saami and the Finns: implications for their population history. American Journal of Physical Anthropology, 128: 185–193.

Jobes D.V., Friedlaender J.S., Mgone C.S., Agostini H.T., Koki G., Yanagihara R., Ng T.C.N., Chima S.C., Ryschkewitsch C.F., and Stoner G.L. (2001) New JC virus (JCV) genotypes from Papua New Guinea and Micronesia (Type 8 and Type 2E) and evolutionary analysis of 32 complete JCV genomes. Archives of Virology, 146: 2097–2113.

Kato A., Kitamura T., Sugimoto C., Ogawa Y., Nakazato K., Nagashima K., Hall W.W., Kawabe K., and Yogo Y. (1997) Lack of evidence for the transmission of *JC polyomavirus* between human populations. Archives of Virology, 142: 875–882.

Kayser M., Brauer S., Weiss G., Underhill P.A., Roewer L., Schiefenhovel W., and Stoneking M. (2000) Melanesian origin of Polynesian Y chromosomes. Current Biology, 10: 1237–1246.

Kimura M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution, 16: 111–120.

Kitamura T., Aso Y., Kuniyoshi N., Hara K., and Yogo Y. (1990) High incidence of urinary JC virus excretion in nonimmunosuppressed older patients. The Journal of Infectious Diseases, 161: 1128–1133.

Kitamura T., Kunitake T., Guo J., Tominaga T., Kawabe K., and Yogo Y. (1994) Transmission of the human polyomavirus JC virus occurs both within the family and outside the family. Journal of Clinical Microbiology, 32: 2359–2363.

Kitamura T., Sugimoto C., Kato A., Ebihara H., Suzuki M., Taguchi F., Kawabe K., and Yogo Y. (1997) Persistent JC virus (JCV) infection is demonstrated by continuous shedding of the same JCV strains. Jounal of Clinical Microbiolgy, 35: 1255–1257.

Kunitake T., Kitamura T., Guo J., Taguchi F., Kawabe K., and Yogo Y. (1995) Parent-to-child transmission is relatively common in the spread of the human polyomavirus JC virus. Journal of Clinical Microbiology, 33: 1448–1451.

Miranda J.J., Sugimoto C., Paraguison R., Takasaka T., Zheng H.-Y., and Yogo Y. (2003) Genetic diversity of JC virus in the modern Filipino population: implications for the peopling of

the Philippines. American Journal of Physical Anthropology, 120: 125–132.
Miranda J.J., Takasaka T., Zheng H.-Y., Kitamura T., and Yogo Y. (2004) JC virus genotype profile in the Mamanwa, a Philippine Negrito tribe, and implications for its population history. Anthropological Science, 112: 173–178.
Oppenheimer S. (1998) Eden in the East: The Drowned Continent of Southeast Asia. Weidenfeld and Nicholson, London.
Oppenheimer S. and Richards M. (2001) Fast trains, slow boats, and the ancestry of the Polynesian islanders. Science Progress, 84: 157–181.
Padgett B.L., Walker D.L., ZuRhein G.M., Eckroade R.J., and Dessel B.H. (1971) Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. The Lancet, 297: 1257–1260.
Page R.D.M. (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences, 12: 357–358.
Pavesi A. (2003) African origin of polyomavirus JC and implications for prehistoric human migrations. Journal of Molecular Evolution, 56: 564–572.
Richards M., Oppenheimer S., and Sykes B. (1998) MtDNA suggests Polynesian origins in eastern Indonesia. American Journal of Human Genetics, 63: 1234–1236.
Ruhlen M. (1987) A Guide to the World's Languages, Volume 1: Classification. Stanford University Press, Stanford.
Ryschkewitsch C.F., Friedlaender J.S., Mgone C.S., Jobes D.V., Agostini H.T., Chima S.C., Alpers M.P., Koki G., Yanagihara R., and Stoner G.L. (2000) Human polyomavirus JC variants in Papua New Guinea and Guam reflect ancient population settlement and viral evolution. Microbes and Infection, 2: 987–996.
Saitou N. and Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogentic trees. Molecular Biology and Evolution, 4: 406–425.
Saruwatari L., Zheng H.-Y., Takasaka T., Sugimoto C., Sakai E., Bo Bo, Nwe Nwe Aung, Kitamura T., Yogo Y., and Ohno N. (2002) Peopling of Myanmar as demonstrated by genotyping of urinary JC virus DNA. Anthropological Science, 110: 235–249.
Sigurgardottir S., Helgason A., Gulcher J.R., Stefansson K., and Donnelly P. (2000) The mutation rate in the human mtDNA control region. American Journal of Human Genetics, 66: 1599–1609.
Stoner G.L., Jobes D.V., Fernandez-Cobo M., Agostini H.T., Chima S.C., and Ryschkewitsch C.F. (2000) JC virus as a marker of human migration to the Americas. Microbes and Infection, 2: 1905–1911.
Sugimoto C., Kitamura T., Guo J., Al-Ahdal M.N., Shchelkunov S.N., Otova B., Ondrejka P., Chollet J.-Y., El-Safi S., Ettayebi M., Grèsenguet G., Kocagöz T., Chaiyarasamee S., Thant K.Z., Thein S., Moe K., Kobayashi N., Taguchi F., and Yogo Y. (1997) Typing of urinary JC virus DNA offers a novel means of tracing human migrations. Proceedings of the National Academy of Sciences of the United States of America, 94: 9191–9196.
Sugimoto C., Hasegawa M., Kato A., Zheng H.-Y., Ebihara H., Taguchi F., Kitamura T., and Yogo Y. (2002a) Evolution of human polyomavirus JC: implications for the population history of humans. Journal of Molecular Evolution, 54: 285–297.
Sugimoto C., Hasegawa M., Zheng H.-Y., Demenev V., Sekino Y., Kojima K., Honjo T., Kida H., Hovi T., Vesikari T., Schalken J.A., Tomita K., Mitsunobu Y., Ikegaya H., Kobayashi N., Kitamura T., and Yogo Y. (2002b) JC virus strains indigenous to northeastern Siberians and Canadian Inuits are unique but evolutionally related to those distributed throughout Europe and Mediterranean areas. Journal of Molecular Evololution, 55: 322–335.
Suzuki M., Zheng H.-Y., Takasaka T., Sugimoto C., Kitamura T., Beutler E., and Yogo Y. (2002) Asian genotypes of JC virus in Japanese-Americans suggest familial transmission. Journal of Virology, 76: 10074–10078.
Takasaka T., Miranda J.J., Sugimoto C., Paraguison R., Zheng H.-Y., Kitamura T., and Yogo Y. (2004) Genotypes of JC virus in Southeast Asia and the western Pacific: implications for human migrations from Asia to the Pacific. Anthropological Science, 112: 53–59.
Takasaka T., Kitamura T., Sugimoto C., Guo J., Zheng H.-.Y., and Yogo Y. (2005) Phylogenetic analysis of major African genotype (Af2) of JC virus: implications for origin and dispersals of modern Africans. American Journal of Physical Anthropology DOI: 10.1002/ajpa.20208 (in press).
Thompson J.D., Higgins D.G., and Gibson T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22: 4673–4680.
Trejaut J.A., Kivisild T., Loo J.H., Lee C.L., He C.L., Hsu C.J., Li Z.Y., and Lin M. (2005) Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. PLoS Biology, 3: e247.
Universal Virus Database of the International Committee on Taxonomy of Viruses (ICTVdB) (2005) ICTVdB Index of Viruses, http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/fr-index.htm
Yanagihara R., Nerurkar V.R., Scheirich I., Agostini H.T., Mgone C.S., Cui X., Jobes D.V., Cubitt C.L., Ryschkewitsch C.F., Hrdy D.B., Friedlaender J.S., and Stoner G.L. (2002) JC virus genotypes in the western Pacific suggest Asian mainland relationships and virus association with early population movements. Human Biology, 74: 473–488.
Yogo Y. and Sugimoto C. (2001) The archetype concept and regulatory region rearrangement. In: Khalili K. and Stoner G.L. (eds), Human Polyomaviruses: Molecular and Clinical Perspectives. John Wiley & Sons, New York, pp. 127–148.
Yogo Y., Iida T., Taguchi F., Kitamura T., and Aso Y. (1991) Typing of human *polyomavirus JC* virus on the basis of restriction fragment length polymorphisms. Journal of Clinical Microbiology, 29: 2130–2138.
Yogo Y., Sugimoto C., Zheng H.-Y., Ikegaya H., Takasaka T., and Kitamura T. (2004) JC virus genotyping offers a new paradigm in the study of human populations. Reviews in Medical Virology, 14: 179–191.
Zheng H.-Y., Kitamura T., Takasaka T., Chen Q., and Yogo Y. (2004a) Unambiguous identification of *JC polyomavirus* strains transmitted from parents to children. Archives of Virology, 149: 261–273.
Zheng H.-Y., Zhao P., Suganami H., Ohasi Y., Ikegaya H., Kim J.C., Sugimoto C., Takasaka T., Kitamura T., and Yogo Y. (2004b) Regional distribution of two related Northeast Asian genotypes of JC virus, CY-a and -b: implications for the dispersal of Northeast Asians. Microbes and Infection, 6: 596–603.
Zheng H.-Y., Ikegaya H., Nakajima M., Sakurada K., Takasaka T., Kitamura T., and Yogo Y. (2005a) Two distinct genotypes (MY-x and MX) of JC virus previously identified in Hokkaido Ainu. Anthropological Science, 113: 225–231.
Zheng H.-Y., Kojima K., Ikegaya H., Takasaka T., Kitamura T., and Yogo Y. (2005b) JC virus genotyping suggests a close contact or affinity between Greenland Inuit and other circumarctic populations. Anthropological Science, 113: 291–293.

ANTHROPOLOGICAL SCIENCE

Dispersal of southeastern Asians based on a global phylogenetic analysis of *JC polyomavirus* isolates of genotype SC

LEI SARUWATARI[1], HUAI-YING ZHENG[2,3], HIROSHI IKEGAYA[4], TOMOKAZU TAKASAKA[2], JING GUO[5], TADAICHI KITAMURA[2], YOSHIAKI YOGO[2], NORIKAZU OHNO[1]*

[1]*Department of Anatomy, School of Dentistry, Aichi-Gakuin University, 1-100 Kusumoto-cho, Chikusaku, Nagoya, 464-8650 Japan*
[2]*Department of Urology, Faculty of Medicine, The University of Tokyo, Japan*
[3]*Japanese Foundation for AIDS Prevention, Tokyo, Japan*
[4]*National Research Institute of Police Science, Chiba, Japan*
[5]*Department of Biochemistry, University of Alberta, Canada*

Received 29 June 2005; accepted 5 January 2006

Abstract The SC genotype of the *JC polyomavirus* is the key to understanding the origin and dispersal of southeastern Asians. We performed a phylogenetic analysis based on complete DNA sequences of 24 SC isolates worldwide, including 11 in China, three in Vietnam, two in Malaysia, two in Myanmar, one in Indonesia, two in Mauritius, one in Zambia, one in South Africa, and one in Hawaii, USA. The results suggest that although SC isolates worldwide can be classified into several subgroups, only one (SC-f) has attained a worldwide distribution. This conclusion was confirmed by a single nucleotide polymorphism analysis of 275 reported partial SC DNA sequences worldwide. Based on the present findings, inferences can be made regarding the ancient dispersals of southeastern Asians carrying particular SC subgroups.

Key words: *JC polyomavirus* (JCPyV), complete DNA sequences, phylogenetic analysis, genotype SC, dispersal of southeastern Asians

Introduction

JC polyomavirus (JCPyV) is a member of the *Polyomaviridae* family. Its genome is a single molecule of covalently closed, circular double-stranded DNA of about 5100 bp in length (Cole and Conzen, 2001). Although JCPyV causes progressive multifocal leukoencephalopathy (PML) in immunocompromised patients (Padgett et al., 1971), it is ubiquitous in the human population, infecting children asymptomatically and then persisting in renal tissue throughout life (Yogo et al., 2004). In most adults, renal JCPyV is not latent but replicates to excrete progeny into urine (Kitamura et al., 1990, 1994; Agostini et al., 1996). JCPyV strains around the world can be classified into more than ten major genotypes, with each genotype occupying a unique geographical domain (Sugimoto et al., 1997; Cui et al., 2004; Yogo et al., 2004). Hence, JCPyV genotyping analysis has provided new insights into the origins of various ethnic groups throughout the world (Yogo et al., 2004).

Based on the relationships between JCPyV genotypes and human populations (Yogo et al., 2004), those populations carrying the same genotype are assumed to be related to each other. Furthermore, if a JCPyV genotype is classified into subgroups using an elaborate phylogenetic analysis and the distribution of these subgroups in various human populations is determined, new insights into the relationships among related but ethnically different populations can be gained. Some successful examples of this approach have been reported recently (Zheng et al., 2003, 2004; Ikegaya et al., 2005; Takasaka et al., 2005). In the current study, we focused our attention on an Asian genotype of JCPyV named SC. Sugimoto et al. (1997) found that this genotype occurs mainly in southeastern Asia and southern China, with minor SC isolates in central and northern Asia and southern Africa. Using phylogenetic analysis based on complete viral DNA sequences, Saruwatari et al. (2002b) subclassified many SC isolates in Myanmar, together with a few in other countries, into six subgroups, SC-a to -f. Most subgroups (SC-a to -e) contained isolates in Myanmar, but one (SC-f) included isolates in various regions of southeastern Asia and southern China. Using complete viral DNA sequences, Takasaka et al. (2004) analyzed phylogenetic relationships among JCPyV isolates in the Philippines and in other geographic regions. While a novel SC subgroup (named SC-x) was identified among modern Filipinos based on this analysis, it was also confirmed that most Filipino isolates belong to the SC-f subgroup.

To extend the findings of Saruwatari et al. (2002b) and Takasaka et al. (2004), we performed a profound phylogenetic analysis of many SC isolates recovered at various sites worldwide. We determined the complete sequences of 15 clones containing full-length JCPyV genomes obtained in this and a previous study (Guo et al., 1996). The SC sequences thus obtained, those recently reported but left unclassified into subgroups (Cui et al., 2004; Venter et al.,

* Corresponding author. e-mail: kiwa@dpc.aichi-gakuin.ac.jp
phone: +81-52-751-2561; fax: +81-52-752-5988

Published online 8 March 2006
in J-STAGE (www.jstage.jst.go.jp) DOI: 10.1537/ase.050629