JC virus Type 2: definition of subtypes based on DNA sequence analysis of ten complete genomes

Hansjürgen T. Agostini,† Yukiko Shishido-Hara, Robert W. Baumhefner, Elyse J. Singer, Caroline F. Ryschkewitsch and Gerald L. Stoner

1 Neurotoxicology Section, National Institute of Neurological Disorders and Stroke, Bldg 36, Room 4A-29, National Institutes of Health, Bethesda, MD 20892-4126, USA
2 Neurology Service, VAMC West Los Angeles, Los Angeles, CA 90073, USA

Five major genotypes of JC virus (JCV) have been defined based on nucleotide differences in the VP1 gene of the DNA sequence. These types are probably a result of virus evolution in geographically isolated population groups. One of the first genotypes identified, Type 2, was found to represent strains of Asian origin. In order to further define the spectrum within Type 2 strains, the entire 5.1 kb genome of nine urinary strains of JCV was amplified by PCR with one pair of primers. These urine samples were obtained in the USA (California and New Mexico) from three European Americans, three Native Americans, two African Americans and one Hispanic American. The complete genome of an Asian JCV strain (Tokyo-1) isolated from progressive multifocal leukoencephalopathy (PML) brain in Japan was also sequenced. Here, we report the analysis of these ten DNA sequences and their deduced protein translations. Two phylogenetically distinct subtypes of Type 2 were found, 2A and 2B, which differ from each other by 0.8–1.1% of the coding region sequence. A 215 bp product amplified with primers in the VP1 gene contains enough sequence information to distinguish the major types and subtypes of JCV and is suitable for application in viral epidemiological studies. The investigation of these genomic variations is of special interest because JCV Type 2 strains are found at a significantly higher frequency in brain tissue of patients with PML than would be predicted from their excretion in a control population.

JCV Type 1 is localized in Europe, Type 2 in Asia and Type 3 in Africa (Agostini et al., 1995). Type 4, found in the USA, is closely related to Type 1, but changes to a Type 3-like sequence within the gene of the major capsid protein VP1, possibly as a result of a recombination event (Agostini et al., 1996a, 1996b). A strain tentatively assigned as Type 5 (Agostini et al., 1996b), has been reassigned as a subtype of Type 3 following complete genome analysis (Agostini et al., 1997a). To date, the major JCV genotypes of the world have also been found in the USA population, but it is likely that additional genotypes remain to be discovered in more isolated indigenous populations. Definition of JCV genotypes on the level of the entire genome will be a useful tool in epidemiological studies and may also illuminate other aspects of virus pathobiology.

Although JCV infects 70–90% of the population during late childhood and persists in the kidney and possibly other internal organs, the only known illness caused by JCV is the central demyelinating disease, progressive multifocal leukoencephalopathy.
encephalopathy (PML). An underlying immunodeficiency favours cytotoxic replication of JCV in oligodendrocytes, the myelin-forming cells. PML is usually fatal within 3–9 months (Walker, 1985). About 5% of AIDS patients show neuropathological evidence of PML at autopsy (Kuchelmeister et al., 1993; Berger & Concha, 1995).

JCV was identified in PML brain by electron microscopy in 1965. The prototype JCV strain (Mad1) was isolated from PML brain in 1971 in Madison, WI, USA (Padgett et al., 1971; Walker, 1985). In 1981, a JCV strain known as Tokyo-1 was isolated from PML brain in Japan (Nagashima et al., 1981). The complete DNA sequence of the JCV (Mad1) genome was reported in 1984 (Frisque et al., 1984). In 1988, the complete DNA sequence of another isolate from a PML patient, termed GS/B, which differed slightly in sequence from Mad1, was reported in Germany (Loeber & Dörries, 1988).

The circular genome of JCV can be divided into divergently transcribed early and late coding regions and the non-coding control region containing the origin of DNA replication which regulates their transcription. Following early region expression [large T and small t antigens and the newly described T’ (Trowbridge & Frisque, 1995)] and DNA replication, the late proteins consisting of capsid proteins VP1–VP3 and the agnoprotein are synthesized and virions are assembled in the cell nucleus. In PML brain tissue, the regulatory region to the late side of ori shows unique combinations of deletions and duplications which appear to be derived from an archetypal structure persisting in the urinary tract (Yogo et al., 1990; Ault & Stoner, 1993). These rearrangements are independent of the genotypes assigned from the coding region sequence (Iida et al., 1993; Agostini et al., 1997c).

The information about JCV genotypes is based mainly on the VT intergenic region which includes the 3′ ends of the genes for VP1 and large T antigen (Ault & Stoner, 1992; Iida et al., 1993; Guo et al., 1996), or on type-specific changes within a short VP1 gene fragment located upstream from the VT intergenic region (Agostini et al., 1995, 1996a, 1996b). Whole genome amplification obviates the need to clone directly the low levels of virus found in urine and provides a suitable substrate for direct cycle sequencing of any part of the viral genome. In addition, whole genome amplification followed by direct cycle sequencing provides an averaged sequence at each position, thus eliminating the need for analysis of multiple clones to obtain a sequence representative of the entire virus population. Here, we report amplification of the entire circular viral genome (5–1 kb) of nine urinary strains of JCV in a single PCR by methods previously described (Agostini & Stoner, 1995). We now report the complete genome DNA sequence of nine Type 2 strains and a tenth sequence which may represent a new genotype most closely related to Type 2B. These ten isolates include nine urinary strains and the JCV (Tokyo-1) strain obtained from the brain of a Japanese PML patient (Nagashima et al., 1981, 1982). In order to establish a representative system for the definition of major JCV types and subtypes in clinical samples, sequences obtained were compared with the prototype Type 2 sequence, GS/B, and we show how typing based on the short fragment within the VP1 gene corresponds to phylogenetic relationships based on entire coding region genomes.

### Methods

#### Clinical samples

Urine samples were obtained from individuals in California and New Mexico, USA (three European Americans, two African Americans, three Native Americans and one Hispanic American). The complete JCV genome was amplified from two urine samples collected from the Hispanic American male with multiple sclerosis at different times. Both PCR products were sequenced to confirm the type determination (Type 2A). The carrier of a strain which has not yet been classified, tentatively termed strain X01, was the offspring of an Irish mother and an Italian father. For additional information about the carriers of different strains and their geographical origin see Table 1.

Unfrozen urine (15–50 ml) was sent to the National Institutes of Health in Bethesda, USA. After centrifugation at 4300 r.p.m. for 10 min, the cell pellet was resuspended in lysis buffer containing 50 mM Tris (pH 8.0), 1 mM EDTA, 0.45% NP40 and 0.45% Tween 20, adjusted to pH 8.0, and digested with proteinase K at a final concentration of 0.2 mg/ml. For further details, see previous descriptions by Agostini et al. (1995) and Stoner & Ryschkewitsch (1995).

#### Isolation and cloning of JCV (Tokyo-1).

JCV (Tokyo-1) was isolated from the brain of a 70-year-old Japanese patient, who developed PML during the treatment of hepatoma and pulmonary tuberculosis (Nagashima et al., 1982). The virus was isolated from the autopsy material of the affected part of PML brain (Nagashima et al., 1982). The virus DNAs were extracted directly from the PML brain or from human foetal glial cells (PHFG) at 6 or 10 passages after inoculation. After digestion with EcoRI they were cloned into the EcoRI site of pUC13. More than 50 clones were isolated and characterized. For this study, we selected the pCT-TC clone isolated from the PHFG culture. This clone contained the full-length genome and had an identical restriction pattern to the virus DNA which was directly cloned from the PML brain (Matsuda et al., 1987).

#### PCR

For the initial detection of JCV in clinical samples, PCR primers JLP-15 and JLP-16 were used [JLP-15, 5′ ACAGTTGGCCAGATT-TCCACTACC 3′ (1710–1734); JLP-16, 5′ TAAAGGCTCCCCCTTACGAGAAAA 3′ (1924–1942); numbering based on JCV (Mad1) [Frisque et al., 1984]]. A two-step PCR of 30 s at 94 °C and 1 min at 63 °C for 50 cycles amplified a 215 bp fragment from the VP1 gene, which includes the shorter product of 129 bp used for JCV genotyping in earlier work (Agostini et al., 1995, 1996a, 1996b; Stoner & Ryschkewitsch, 1995). The entire circular genome of all nine urinary JCV strains was amplified in a single PCR as described previously (Agostini & Stoner, 1995; Agostini et al., 1997a). Briefly, after BamHI digestion of the JCV genome, primers overlapping the single BamHI site were used to generate a linear 5·1 kb product. The DNA sequence in the region of the BamHI site itself was determined independently following amplification of a separate short PCR fragment spanning the restriction site.

#### Sequencing of PCR products

Products of three complete genome PCR amplification reactions for each urine sample were combined, gel-purified, and directly cycle sequenced. Primer series JG-1 to JG-12 and SEC-1 to SEC-12 (Agostini et al., 1997a) were end-labelled with [γ-32P]ATP (Amersham). The initial denaturation for 1 min at 95 °C was followed by 30 cycles of 30 s at 95 °C for denaturation and 1 min at 60 °C for primer annealing and 1 min at 72 °C for extension.
65 °C for annealing and elongation. Products were run for 90 or 180 min on a 6% polyacrylamide gel containing 50% (w/v) urea (National Diagnostics). Overlapping sequences from both strands were determined. For additional details, see Agostini et al. (1997a).

Sequencing of JCV (Tokyo-1). Both strands of the entire sequence of JCV (Tokyo-1) were determined by the dideoxy termination method using T7 Sequenase version 2.0 and [35S]dATP (Amersham). Sequencing primers that did not contain repetitive or complementary structures were utilized. Purified double-stranded DNA of the pJCT-TC clone was used as a template in a single-cycle reaction. Sequence reaction products were electrophoresed using gels composed of 6% polyacrylamide, 8 M urea (50%) and 1 × TBE. Sequence compression and deformity due to secondary structure in the DNA template was resolved using formamide gels composed of 6% polyacrylamide, 40% formamide, 7 M urea and 1 × TBE. The sequences of both strands were confirmed to be identical.

Computer analysis. Multiple sequences were analysed with the 8-Unix version of the GCG programs (Genetics Computer Group, Madison, WI, USA) on a Silicon Graphics Indigo R4000 computer. Consensus sequences of subtypes 2A and 2B were generated by the Pretty program. Phylogenetic analysis utilized the UPGMA method (Distances and Phylogenetic analysis) with the GrowTree program of GCG with the UPGMA method as shown in Fig. 1. The 21 complete coding regions of JCV were distinguished. This analysis is based on the entire JCV genome, excluding only the regulatory region to the late side of the origin of replication (ori) (position 1–267, archetypal numbering (Yogo et al., 1990)). Strains of JCV Type 2 subtypes 2A and 2B differ by about 0.8–1.0% as shown in Table 2. The subtype assignment of strain #230 is uncertain; the distance matrix shows it to be equidistant from strains of both the 2A and 2B subtypes (0.7–0.9%), although in the phylogenetic tree it falls into the subtype 2B group (see below). In contrast to the differences between subtypes 2A and 2B (<1%), differences between the other major genotypes (Types 1, 3 and 6) are 1.2–2.3% (Table 3). Variation between strains within a given subtype is usually less than 0.5% (Table 2). The Japanese strain JCV (Tokyo-1) is the prototype for subtype 2A whereas the European Type 2 strain JCV/GS/B is classified as subtype 2B.

Based on the distance matrix calculation for the coding regions of 21 sequences, a phylogenetic tree was generated using the GrowTree program of GCG with the UPGMA method as shown in Fig. 1. The 21 complete coding regions depicted in the phylogram include previously published JCV (Mad1) (Frisque et al., 1984) and GS/B (Loeber & Dorries, 1988) sequences, and the five complete African and African American Type 3 genomes reported previously (Agostini et al., 1997a). In addition, Type 1, Type 4 (#402) and Type 6 sequences are included (Agostini et al., 1998b; H. T. Agostini, C. F. Ryschkewitsch & G. L. Stoner, unpublished results). The two subtypes of Type 2 fall into distinct branches of the tree. Strain #230 is associated with the subtype 2B group, but the distance matrix analysis shows that it differs about equally from both subtypes 2A and 2B, so it may represent a new

Table 1. Origin of JCV Type 2 and related strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Sex</th>
<th>Age (years)</th>
<th>Ethnicity*</th>
<th>Sample</th>
<th>Primary disease</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>#223</td>
<td>M</td>
<td>36</td>
<td>AFAM</td>
<td>Urine</td>
<td>HIV infection</td>
<td>CA, USA</td>
</tr>
<tr>
<td>#224</td>
<td>M</td>
<td>72</td>
<td>HISP</td>
<td>Urine</td>
<td>Multiple sclerosis</td>
<td>CA, USA</td>
</tr>
<tr>
<td>#225</td>
<td>F</td>
<td>20</td>
<td>NAAM</td>
<td>Urine</td>
<td>None</td>
<td>NM, USA</td>
</tr>
<tr>
<td>#226</td>
<td>F</td>
<td>78</td>
<td>NAAM</td>
<td>Urine</td>
<td>None</td>
<td>NM, USA</td>
</tr>
<tr>
<td>#227</td>
<td>M</td>
<td>41</td>
<td>EUAM</td>
<td>Urine</td>
<td>Multiple sclerosis</td>
<td>CA, USA</td>
</tr>
<tr>
<td>#228</td>
<td>M</td>
<td>63</td>
<td>NAAM</td>
<td>Urine</td>
<td>None</td>
<td>NM, USA</td>
</tr>
<tr>
<td>#229</td>
<td>M</td>
<td>48</td>
<td>EUAM</td>
<td>Urine</td>
<td>HIV infection</td>
<td>CA, USA</td>
</tr>
<tr>
<td>#230</td>
<td>M</td>
<td>44</td>
<td>AFAM</td>
<td>Urine</td>
<td>Multiple sclerosis</td>
<td>CA, USA</td>
</tr>
<tr>
<td>Tokyo-1</td>
<td>M</td>
<td>70</td>
<td>Japanese</td>
<td>PML Brain</td>
<td>Hepatoma/tuberculosis</td>
<td>Tokyo, Japan</td>
</tr>
</tbody>
</table>

X01 M 30 EUAM+ Urine Multiple sclerosis CA, USA

* AFAM, African American; HISP, Hispanic; NAAM, Native American; EUAM, European American.
† Subject had Irish mother and Italian father.

Results

Subtypes of JCV Type 2

Using the distance matrix analysis of coding region sequences, two subtypes of JCV Type 2, 2A and 2B, can be distinguished. This analysis is based on the entire JCV genome, excluding only the regulatory region to the late side of the origin of replication (ori) (position 1–267, archetypal numbering (Yogo et al., 1990)). Strains of JCV Type 2 subtypes 2A and 2B differ by about 0.8–1.1% as shown in Table 2. The subtype assignment of strain #230 is uncertain; the distance matrix shows it to be equidistant from strains of both the 2A and 2B subtypes (0.7–0.9%), although in the phylogenetic tree it falls into the subtype 2B group (see below). In contrast to the differences between subtypes 2A and 2B (<1%), differences between the other major genotypes (Types 1, 3 and 6) are 1.2–2.3% (Table 3). Variation between strains within a given subtype is usually less than 0.5% (Table 2). The Japanese strain JCV (Tokyo-1) is the prototype for subtype 2A whereas the European Type 2 strain JCV/GS/B is classified as subtype 2B.
Table 2. Nucleotide differences (%) between JCV Type 2 strains

Comparisons within a subtype are given in italics, whereas comparisons between subtypes 2A and 2B are given in bold.

<table>
<thead>
<tr>
<th>JCV strain</th>
<th>Subtype 2A</th>
<th>Subtype 2B</th>
<th>Subtype 2x*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#229</td>
<td>#225</td>
<td>#226 Tokyo-1</td>
</tr>
<tr>
<td>#228</td>
<td>0.06</td>
<td>0.25</td>
<td>0.45</td>
</tr>
<tr>
<td>#229</td>
<td>0.23</td>
<td>0.23</td>
<td>0.43</td>
</tr>
<tr>
<td>#225</td>
<td>0.02</td>
<td>0.21</td>
<td>0.33</td>
</tr>
<tr>
<td>#226</td>
<td>0.31</td>
<td>0.31</td>
<td>0.45</td>
</tr>
<tr>
<td>Tokyo-1</td>
<td>0.04</td>
<td>0.19</td>
<td>0.23</td>
</tr>
<tr>
<td>#224</td>
<td>0.06</td>
<td>0.21</td>
<td>0.45</td>
</tr>
<tr>
<td>#223</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>#227 GS/B</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Subtype 2A (con): #224, #225, #226 and Tokyo-1.
Subtype 2B (con): #223, #227 and GS/B.
Subtype 2x*: Unclassified Type 2 strain.

Table 3. Nucleotide differences (%) between Type 2 subtypes and other major genotypes

Differences below 1% are given in italics.

<table>
<thead>
<tr>
<th>Type/strain</th>
<th>Subtype 2A (con)*</th>
<th>Subtype 2B (con)†</th>
<th>#230‡</th>
<th>#308§</th>
<th>X01</th>
<th>#601¶</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mad1</td>
<td>2.28</td>
<td>1.96</td>
<td>2.13</td>
<td>2.24</td>
<td>2.05</td>
<td>2.32</td>
</tr>
<tr>
<td>Subtype 2A</td>
<td></td>
<td></td>
<td>0.89</td>
<td></td>
<td></td>
<td>1.11</td>
</tr>
<tr>
<td>Subtype 2B</td>
<td></td>
<td></td>
<td>0.72</td>
<td>1.18</td>
<td>0.81</td>
<td>1.96</td>
</tr>
<tr>
<td>Subtype 2C</td>
<td></td>
<td></td>
<td>1.25</td>
<td>1.20</td>
<td>1.96</td>
<td></td>
</tr>
<tr>
<td>#230</td>
<td></td>
<td></td>
<td></td>
<td>1.54</td>
<td>2.05</td>
<td></td>
</tr>
<tr>
<td>#308</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.35</td>
</tr>
</tbody>
</table>

* Consensus of six subtype 2A strains.
† Consensus of three subtype 2B strains.
‡ A Type 2 strain (this work).
§ A Type 3 strain (Agostini et al., 1997a).
¶ A Type 6 strain (unpublished data).

subtype distinct from both 2A and 2B (Table 2). To date, only this single strain has been observed, and the frequency of these variants in the population is unknown.

Complete genome versus short fragment typing

Complete genome analysis validates the typing assignments based on short DNA fragment amplification (Agostini et al., 1996b). The fragment generated by primers JLP-1 and -4 or JLP-15 and -16 (a 129 or 215 bp fragment within the VP1 gene, respectively) allows differentiation of all major genotypes (Types 1–4 and Type 6), and also subtypes of JCV Types 1 and 2 (Agostini et al., 1996b). Details of Type 1 subtypes have been described (Agostini et al., 1998b). With regard to JCV Type 2, three subtypes 2A, 2B and 2C were initially identified by variations at nucleotide positions 1837 and 1850 (Fig. 2). Subgroups 2A, 2B and 2C are characterized by nucleotides at positions 1837/1850 of T/A, C/G and T/G, respectively. Although the difference between subtypes 2A and 2B is reflected in the phylogenetic tree analysis (Fig. 1), the subtype 2C strains which have been completely sequenced fall into either the 2A group (#228, #229) or the 2B group (#230).

The consensus sequence of subtype 2A, which consists of strains #224, #225, #226 and Tokyo-1, differs from that of subtype 2B (strains #223, #227 and GS/B) at a total of 40 positions within the coding region (0.82%). These sites can be used for subtyping in any region of the genome, and are listed...
Definition of JCV Type 2 subtypes

Fig. 1. Phylogenetic tree of JCV genotypes. The first number in each strain designation stands for the major genotype. Type 2 can be divided into subtypes 2A and 2B. All Native American strains (225, 226 and 228) and a European strain (Tokyo-1) fall into the Type 2A group, while an African American strain (223) and a European strain (GS/B) fall into the Type 2B group. Two of the strains originally designated Type 2C based on the sequence of the 215 bp VP1 gene fragment fall into the Type 2A group (228 and 229), while one falls into the Type 2B group (230). However, strain 230 is clearly distinct from other Type 2B strains and may represent an additional subtype of Type 2. Type 4 (402) is a variant of the Type 1 genotype, but is distinctive in the 215 bp VP1 gene fragment. Strain X01 is a tentative designation pending additional isolates and its classification as a distinct genotype or as a subtype of Type 2. For origins of strains see Table 1. For DNA sequences see Agostini et al. (1997a) and the GenBank database.

Fig. 2. Definition of JCV genotypes in the 215 bp VP1 gene fragment. This sequence allows distinction of all major JCV genotypes and subtypes identified to date. Note that only strain X01 and Type 6 are identical at these sites. Type 2C strains defined in this typing fragment actually fall into the 2A (228 and 229) or 2B groups (230) when the entire genome is analysed (see text and Fig. 1). It seems likely that most strains with T/G at positions 1837/1850 will be variants of the 2A subtype.

Variability within the genome

When Type 1, 2 and 3 consensus sequences are compared, the primary sequence varies most within the capsid proteins, especially VP1, and the 3’ end of the second exon of large T antigen (data not shown). The agnoprotein gene has very few variations among the three major genotypes (Agostini et al., 1997a). This is in contrast to the relative variation between the consensus sequences of JCV Type 2 subtypes 2A and 2B. With four nucleotide differences in the agnoprotein gene (1-9%), the relative variability is higher than that between major genotypes (Type 1 vs 2A, 1-4%; 1 vs 2B, 0-46%; 1 vs 3, 0-46%; 2A vs 3, 0-93%; 2B vs 3, 0-93%). The differences between other genes of the subtype 2A and 2B genomes were lower (VP2, 0-58%; VP3, 0-74%; VP1, 1-23%; T antigen exon 1, 0-41%; T antigen exon 2, 0-72%; small t antigen, 0-19%).

The intergenic region between the 3’ end of the agnoprotein gene and the start of the VP2 gene is a highly conserved region of the JCV genome. Individual mutations are rare and the consensus sequences of Types 1, 2A, 2B and 3 are identical within this area. An exception is the insertion of a single base pair in this element in strain #224 (see below).

Strain X01

The 30-year-old carrier of strain X01 was born in the USA, the child of an Irish mother and an Italian father. Over the course of 8 months, three urine samples were positive for the same JCV strain. A comparison of the coding region of X01 with Type 1 (Mad1), subtype 2A consensus, subtype 2B consensus and Type 3 (308) showed that strain X01 differs from Type 1 by 2-05%, from subtype 2A by 1-33%, from subtype 2B by 0-81% and from Type 3 by 1-54% of the nucleotide sequence (Table 3). Thus, the status of strain X01 is ambiguous. It could be viewed either as an additional European genotype, or as a new subtype within the Type 2 group which is more closely related to subtype 2B strains than to subtype 2A.

Genotype-related protein changes

As shown in Fig. 3(b), the nucleotide differences in JCV subtype 2A and 2B strains result in ten predicted amino acid
Fig. 3. Sequence comparison of JCV Type 2 subtypes. (a) Subtyping positions for JCV Type 2 subtypes 2A and 2B. The consensus sequences of JCV Types 2A and 2B differ at a total of 40 sites within the 4854 bp of the coding region of the JCV genome (excludes only 267 bp of the archetypal regulatory region between ori and the ATG start site of the agnoprotein gene). The A → T change at position 3768 (box) changes the Gln to Leu within the zinc finger motif. (b) Predicted amino acid changes in Type 2A, 2B, strain #230 and strain X01. Only those positions where the consensus sequence of Types 2A and 2B or strains #230 and X01 differ from each other are included. The corresponding amino acids in Types 1 and 3 are listed. Note that position 301 (Leu/Gln) in the large T antigen distinguishes Types 2A and 2B, which follow Type 3 and Type 1, respectively.

...
Definition of JCV Type 2 subtypes

X01, the hydrophilic residue Gln is found at that position, as it is in all Type 1 strains. Another variable site is located in the carboxy terminus of large T antigen at position 653.

Single nucleotide deletions and insertions

The majority of JCV strains analysed do not differ in length from the prototype strain JCV (Mad1) within the overall coding region (4854 bp). However, nucleotide additions or deletions can occur in intergenic regions or in introns which do not affect the reading frame. As shown in Fig. 4, strain #224 (subtype 2A) has an insertion of deoxythymidine in the agnoprotein–VP2 intergenic region. This insertion was confirmed in multiple PCR amplifications. Two other strains closely related to each other and originally termed subtype 2C (#217 and #230) both have a deletion of deoxyadenosine at position 4454 (Mad1 numbering) (Fig. 4). Compression artefacts of sequencing have been eliminated as explanations for these deletions by additional sequencing of a short-range amplified product of the large T antigen intron using a different pair of primers (data not shown).

Regulatory region

Of the nine urinary strains sequenced, six had the expected archetypal regulatory region to the right of ori. Details on minor rearrangements in strains #224, #226 and #230 will be published elsewhere. The regulatory region sequences of strains #223 and #227 were identical to the American consensus sequence (Agostini et al., 1996 b), which differs from the original Asian isolate (Yogo et al., 1990) only at position 217. The usefulness of the regulatory region for subtyping of JCV strains is limited. Previously, deoxyadenosine at position 217 (archetype numbering) (Yogo et al., 1990) seemed to be an exclusive but inconsistent feature of subtype 2A strains, but the unrearranged regulatory region of strain X01 and #230 also show the same nucleotide change. Other point mutations in the regulatory region are listed in Table 4.

Discussion

This analysis of ten complete JCV genomes confirms the existence of the Type 2 group of sequences and allows the assignment of virus subtypes. These results extend those based on RFLP (Yogo et al., 1991) or partial sequence analysis (Ault & Stoner, 1992; Agostini et al., 1995, 1996 b, 1997 a). Here, we have characterized subtypes 2A and 2B, and a possible new subtype represented by strain #228. Two other strains (agp-VP2 intergenic region. This insertion was confirmed in multiple PCR amplifications. Two other strains closely related to each other and originally termed subtype 2C (#217 and #230) both have a deletion of deoxyadenosine at position 4454 (Mad1 numbering) (Fig. 4). Compression artefacts of sequencing have been eliminated as explanations for these deletions by additional sequencing of a short-range amplified product of the large T antigen intron using a different pair of primers (data not shown).

Table 4. Strain specific nucleotide variation in the regulatory region

<table>
<thead>
<tr>
<th>Subtype†</th>
<th>Strain</th>
<th>Nucleotide position*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>92 107 120 159 217 227</td>
</tr>
<tr>
<td>2A</td>
<td>#225</td>
<td>A  T  G  A  C  A  T</td>
</tr>
<tr>
<td>2C</td>
<td>#228</td>
<td>—    —    A  A  —</td>
</tr>
<tr>
<td>2C</td>
<td>#229</td>
<td>—    C    —    A  —</td>
</tr>
<tr>
<td>2B</td>
<td>#230</td>
<td>—    —    —    —    G</td>
</tr>
<tr>
<td>X01</td>
<td>—</td>
<td>—    —    —    —    G</td>
</tr>
</tbody>
</table>

* Position of nucleotides based on strain CY (archetype) numbering (Yogo et al., 1990). Consensus nucleotides are given in the first row and represent the American consensus with A at position 217 (Agostini et al., 1996 b). —, Identity to consensus.

† Genotype assignments based on short typing fragment amplified by primers JLP-15 and -16.

Comparison of the DNA sequence of complete JCV genomes is complicated by the fact that isolates from different PML tissues vary in length due to deletions and duplications occurring in the regulatory region between the origin of DNA replication and the beginning of the agnoprotein gene (Yogo et al., 1990; Ault & Stoner, 1993; Agostini et al., 1997 c). To circumvent this problem, the analysis of sequence similarity has been limited to 4854 bp [positions 277–5130, JCV (Mad1)]
numbering). Little information is lost as the rearranging portion of the regulatory region contains few typing sites except for deoxycytidine at position 133 in African strains (Agostini et al., 1995).

The ability to define JCV genotypes and subtypes is important for several reasons. Firstly, identification of type-specific sites allows design of PCR primers for reliable diagnosis of JCV infection in brain biopsies or cerebrospinal fluids. Placement of primers in constant regions of JCV genomes ensures that all genotypes will be amplified equally. It is particularly important that the 3’ ends of primers do not coincide with variable, type-determining sites (Ault et al., 1994).

Secondly, although the full biological significance of JCV coding region genotypes is not yet understood, there is the possibility that nucleotide changes might alter its disease-causing potential. Indeed, a comparison of the genotype distribution of JCV amplified from the brain of PML patients with that from urine of control individuals revealed that Type 2 strains are more frequent in the brain than in the urinary tract of the control group, suggesting that infection with the Type 2 genotype may be associated with a higher risk of developing PML than infection with JCV strains of other genotypes (Agostini et al., 1997b). The increased risk of developing PML associated with Type 2 infection compared with non-Type 2 strains was estimated to be threefold. Subtype analysis of the infecting strains indicated that the numbers of subtype 2A and 2B strains did not increase equally. In fact, the excess was almost entirely due to an increase in the numbers of subtype 2B strains found in PML brains (Agostini et al., 1998a). This finding points toward an unknown factor which differs between Type 2 subtypes and influences the neurotropism and/or neurovirulence of JCV.

Thirdly, genotyping is important for understanding the evolution of JCV within the human population. The origin of JCV genotypes may reflect coevolution of the virus within different lineages of the human family. Genotypes characteristic of Europe (Type 1), Asia (Type 2), and Africa (Type 3) have been described (Yogo et al., 1991; Ault & Stoner, 1992; Agostini et al., 1996b; Guo et al., 1996; Stoner et al., 1998). A second African genotype, Type 6, has been identified in Africa (Guo et al., 1996) and in an African American (H. T. Agostini, C.F. Ryschkewitsch & G. L. Stoner, unpublished results). It is likely that JCV arrived in the Americas with the ancestors of Native Americans who came across the Bering land bridge from Asia some time between 12,000 and 30,000 years ago. This is supported by the fact that JCV subtype 2A, the most common genotype found in Asia, is also found in a high percentage of modern Native Americans (Agostini et al., 1997d). JCV infection will be useful to follow migration of human populations in more recent times as well, e.g. the migration of Europeans, Africans and Asians to the Americas within the last 500 years. The identification of JCV genotypes appears to provide a quasi-genetic marker which can complement information obtained from the human genome (Cavalli-Sforza et al., 1994), or from maternally inherited mitochondrial DNA (Vigliant et al., 1989).

Based on genotyping in a 129 or 215 bp amplified VP1 gene fragment, the Type 2 sequences were first assigned to subgroups 2A, 2B and 2C (Fig. 2). These subgroups are characterized by nucleotides at positions 1837/1850 of T/A, C/G, and T/G, respectively. Complete genome analysis has established that most of the T/G group (subtype 2C) can be included in subtype 2A on the basis of the complete genome DNA sequence. Two of the three strains in the T/G subgroup (#228 and #229) fell into the subtype 2A group when the entire genome was considered. On the other hand, a third 2C strain, #230, differed by 0-7-0-9% from both subtypes 2A and 2B, and it may possibly represent a separate and distinct subtype. It is noteworthy that all three original subtype 2C strains (#228, #229, #230) have the nucleotide A at position 3768 which translates as Leu-301 in the zinc finger motif of the large T antigen (Agostini et al., 1997a). This amino acid residue is also found in subtype 2A, Type 3 and Type 6, but not in Type 1, subtype 2B or strain X01, all of which have glutamine. The possibility that this mutation in an essential functional region influences the biological activity of T antigen is currently under investigation.

Rules for assigning JCV genotypes have not yet been defined, but some guidelines can be suggested. Type 1 strains, represented by JCV (Mad1), differ from strains of all other genotypes by about 2-0% of the genome (Table 3). Similarly, the African American strain #601, a representative of Type 6, differs from strains of other genotypes by about 2-0% of its DNA sequence. Types 2 and 3, however, are more closely related, differing from each other by only 1-2-1-3% (Table 3). From Table 2, it is evident that different subtypes of Type 2 (2A, 2B, #230) differ from each other by 0-7-1-1%. Strain-to-strain variation within the subtypes 2A or 2B is less than 0-5%. As the major JCV genotype differences probably reflect sequence divergence accompanying the human diaspora, it will be of interest to define the remaining extant types and their relationship to indigenous populations of the world.

We thank deeply Drs Kazuo Nagashima, Michiyuki Matsuda and Kotaro Yasui who isolated and cloned JCV (Tokyo-1). The DNA sequence of JCV (Tokyo-1) was determined by Yukiko Shishido-Hara with their kind support and encouragement, mostly in Japan in the slow virus infection research projects supported by the Japanese Ministry of Health and Welfare and the Japanese Science and Technology Agency. The sequencing was completed at NIH in Bethesda. This work was also supported in part by the Deutsche Forschungsgemeinschaft, Bonn (grant Ag 19/1-1). The support and encouragement of Dr Henry deF. Webster is gratefully acknowledged. Urine samples were obtained from the National Neurological Research Specimen Bank, VAMC, Los Angeles, CA 90073, which is sponsored by NINDS/NIMH, National Multiple Sclerosis Society, Hereditary Disease Foundation, Comprehensive Epilepsy Program, Tourette Syndrome Association, Dystonia Medical Research Foundation, and Veterans Health Service and Research Administration, Department of Veterans Affairs.
References


Received 9 October 1997; Accepted 16 January 1998