JC virus Type 1 has multiple subtypes: three new complete genomes

Hansjürgen T. Agostini,† Caroline F. Ryschkewitsch and Gerald L. Stoner

Neurotoxicology Section, National Institute of Neurological Disorders and Stroke, Bldg 36, Room 4A-29, National Institutes of Health, Bethesda, MD 20892-4126, USA

The complete genomes of three new Type 1 strains of JC virus (JCV) from urine have been analysed. These were subtype 1A, subtype 1B and Type 4 as assigned from a short typing fragment in the VP1 gene. They differ from Mad1 (subtype 1A) by less than 1·0% of the DNA sequence. Based on its complete genome, the JCV Type 4 strain falls into a Type 1 subgroup. Type 4, with several Type 3-like sites in the short typing fragment, is a possible recombinant strain. The consensus of Type 1 DNA sequences is distinguished within the coding region from both Type 2 (strain GS/B) and five Type 3 (African and African American) strains at 64 sites. Most mutations are silent, but at 21 positions amino acid changes occur. Our findings define the subtypes of JCV Type 1 and support the validity of genotyping within the short VP1 fragment.

JC virus (JCV) is found by PCR analysis in the urine of about 40% of the general population over the age of 30 years (Kitamura et al., 1990; Agostini et al., 1996b; Stoner et al., 1996). In severely immunocompromised individuals, JCV causes a neurological infection, progressive multifocal leukoencephalopathy (PML) (Berger & Concha, 1995). Infection of oligodendrocytes and astrocytes leads to relentlessly progressive demyelination, usually with death in 3 to 9 months (Major et al., 1992). The prototype strain, JCV(Mad1), was isolated from PML brain tissue in primary human foetal glial cells in 1971 in Madison, Wisconsin, USA (Padgett et al., 1971). Its complete genome sequence was reported in 1984 (Frisque et al., 1984). In 1988, the sequences of two additional isolates from the brain and kidney of the same PML patient, termed GS/B and GS/K, were reported in Germany (Loeber & Dörries, 1988). The sequences of these JCV strains from brain and kidney differed significantly in the regulatory region, but not in the coding region.

Studies of 105 control individuals from Pennsylvania and Maryland in the USA identified four genotypes of JCV (Agostini et al., 1996b). These types can be distinguished in a 129-bp fragment of the VP1 gene. This fragment, amplified by primers JLP-1 and -4, also contains two sites which identify putative subtypes of Type 1 (subtypes 1A and 1B). In this classification, JCV(Mad1) is a subtype 1A strain (Agostini et al., 1996b). Type 1 is the major genotype among the European/American population (64%). Type 2 (also known as Type B) is an Asian genotype of JCV (Yogo et al., 1991). A third genotype, Type 3, has been identified as African in origin (Agostini et al., 1995, 1997). A fourth type appears to combine a Type 1 sequence and a short segment from a Type 3 (subtype 123 (subtype 1B) and #402 (Type 4), in order to determine whether the sequence of the entire genome correlates with classification in the 129-bp fragment amplified with primer pair JLP-1 and -4.

Urine is an excellent source of the virus, and the complete genome of JCV (5·1 kb) can be amplified in a single PCR reaction in about one-third of the individuals excreting JCV (Agostini & Stoner, 1995). The method obviates the need to clone the virus directly from urine and yields a product suitable for direct cycle sequencing (Agostini et al., 1997). As direct cycle sequencing from the whole genome PCR product provides an averaged DNA sequence, it can eliminate the need for sequencing of multiple clones from the same sample. Detailed study of virus variation in the general population is now feasible (Agostini & Stoner, 1995; Agostini et al., 1997).

For this analysis of JCV Type 1 DNA sequences, urine samples were obtained either fresh (#124) or frozen (#123 and #402) from Caucasian male multiple sclerosis patients aged 52 to 63 years (Stoner et al., 1996). Urinary cell pellets were prepared as described previously and were lysed with proteinase K (Agostini et al., 1995, 1996b). Initially, viral DNA was detected by PCR amplification using primers JLP-1 and -4.
Table 1. Pairwise comparison of Type 1 DNA sequences with Type 2 and Type 3

<table>
<thead>
<tr>
<th>Strain*</th>
<th>#124 (1A)</th>
<th>#123 (1B)</th>
<th>#402 (1C)</th>
<th>GS/B</th>
<th>#308 (2)</th>
<th>#309 (3)</th>
<th>#310 (3)</th>
<th>#311 (3)</th>
<th>#312 (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mad1 (1A)</td>
<td>0.33</td>
<td>0.58</td>
<td>0.86</td>
<td>2.0</td>
<td>2.2</td>
<td>2.5</td>
<td>2.6</td>
<td>2.1</td>
<td>2.3</td>
</tr>
<tr>
<td>#124 (1A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#123 (1B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#402 (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.95</td>
<td>2.1</td>
<td>2.3</td>
<td>2.6</td>
<td>2.2</td>
</tr>
<tr>
<td>GS/B (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.3</td>
<td>1.5</td>
<td>1.6</td>
<td>1.2</td>
</tr>
<tr>
<td>#308 (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.41</td>
<td>0.45</td>
<td>0.62</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>#309 (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.66</td>
<td>0.78</td>
<td>0.54</td>
</tr>
<tr>
<td>#310 (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.86</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>#311 (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.58</td>
<td></td>
</tr>
</tbody>
</table>

* No. in brackets is the genotype designation.

Differences (%) between coding region sequences (4854 bp) were obtained using the GAP program of the GCG Sequence Analysis Software Package. Underlined nos represent % difference between strains designated as subtypes in the PCR fragment amplified with primers JLP-1 and -4. Nos in bold type represent % difference between strains designated as different types in the PCR product amplified with primers JLP-1 and -4.

JCV-positive samples were processed for PCR amplification of the complete 5.1 kb JCV genome as described previously (Agostini & Stoner, 1995). Briefly, the lysed samples were digested with the restriction enzyme BamHI before the PCR reaction. Viral DNA was amplified using primers BAM-1 and BAM-2 with a 5’ overlap at the restriction enzyme site (GGATCC). The reaction was run for 39 cycles with increasing annealing/extension time (6–12 min) at 64 °C and denaturation (30 s) at 94 °C followed by a final extension (10 min) at 72 °C.

PCR products from complete genome amplification using primers BAM-1 and BAM-2 were cloned according to the TA-cloning procedure (Invitrogen). Briefly, after gel purification (Qiagen) 300 ng of the 5.1-kb-long PCR product were incubated with 1 U Tag DNA polymerase (Perkin Elmer Cetus) and dATP in standard Taq buffer for 15 min at 72 °C in order to add a deoxyadenosine overhang at the 3’ end. The DNA was purified by phenol–chloroform extraction and ethanol precipitation and dissolved in TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 7.5). PCR products were ligated overnight at 14 °C into the pCRII vector (Invitrogen). Competent E. coli (INVαF; Invitrogen) were transformed by heat shock at 42 °C. After overnight culture of positive clones in 50 ml LB broth containing ampicillin (100 µg/ml), about 250 µg plasmid was purified following the Qiagen preparation procedure. To identify the JCV DNA insert, plasmids were digested with BamHI and HincII (Agostini & Stoner, 1995).

Purified plasmids were sequenced by direct cycle sequencing (#124 and #402) or by commercial automated sequencing (#123) (Sequetech). Portions of the sequence were checked using short-range PCR products (Agostini et al., 1995, 1996b). In addition, the sequence of strain #124 was checked by direct cycle sequencing of the PCR-amplified complete genome without cloning (Agostini et al., 1997). The primers and methods used for sequencing of the entire genome have been described previously (Agostini et al., 1997). By direct cycle sequencing of the PCR product without cloning, five apparent PCR-induced errors were identified in the cloned sequence of strain #124.

GenBank/EMBL Data Library accession numbers for the Type 1 sequences reported for the first time in this study are as follows: strain #124 (subtype 1A) is AF015526; #123 (subtype 1B) is AF015527; and #402 (Type 4) is AF015528. The accession number of the complete genome of JCV(Mad1) is J02227 (Frisque et al., 1984). The JCV(GS/B) sequence has been previously reported (Loeber & Dörries, 1988). For accession numbers of the five Type 3 strains see Agostini et al. (1997).

Numbering of the JCV coding region and the nonrearranging regulatory region to the early side of ori is that of JCV(Mad1). Numbering of the regulatory region rearranged in PML tissue is that of the unarranged (archetypal) sequence (Yogo et al., 1990).

The sequencing results showed that subtype 1A (#124) was more closely related to the prototype strain Mad1 (also identified as subtype 1A), than it was to the subtype 1B (#123) strain or the Type 4 strain (#402) (Table 1). Coding region
JCV Type 1 genomes

Fig. 1. Differences of the consensus sequence of four JCV Type 1 strains from both strain GS/B (Type 2) and the African consensus sequence (Type 3). (a) Nucleotide sequence differences. The sequence of the Type 1 strains g123, g124, and g402 is reported in this work. Mad1 is the prototype JCV strain (Frisque et al., 1984), and is designated subtype 1A (Ault & Stoner, 1992; Agostini et al., 1996b). Type 2 is represented by strain GS/B (Yogo et al., 1991b; Ault & Stoner, 1992b; Agostini et al., 1996b). A Type 3 consensus is based on strains g308–g312 from three Tanzanians and two African Americans. The complete genomes of these five strains have been reported previously (Agostini et al., 1997). (b) Variable amino acid positions in JCV proteins. Positions are listed where nucleotide changes in one of the consensus sequences or strain GS/B change 21 amino acids in an early or late protein. After the amino acid position number, the amino acid residues at that position are listed in parentheses in the order of Type 1 consensus, Type 2 (strain GS/B) and Type 3 consensus.

Comparison (4854 bp) showed that strain #402 differed by 0±86 and 0±95% from the subtype 1A and subtype 1B strains, respectively. JCV Type 1 strains were found to differ from a Type 2 strain (prototype GS/B) in 2±0–2±2% of their coding region sequence. When compared to five Type 3 strains, the subtype 1A, subtype 1B and Type 4 sequences differ by 2±1–2±6% (Table 1).

In the regulatory regions, all three strains were archetypal with ‘A’ at position 217 as previously described (Agostini et al., 1996b). In strain #402, a mutation of ‘T’ to ‘C’ at position 252 was confirmed by short-range PCR amplification and sequencing.

These results allow the operational definition of JCV genotypes as those sequences differing by 1% or more, whereas subtypes within the type differ from each other by about 0±50–10%. Less than 0±5% sequence divergence over the entire coding region is assumed to represent the usual strain-to-strain variation. Thus, the existence of subtypes 1A and 1B, proposed from the 129-bp VP1 typing fragment, is supported by analysis of the complete genome sequence. Furthermore, Type 4, of which the majority of the sequence is Type 1, appears to be a subtype distinct from both subtype 1A and subtype 1B. Therefore, on the level of the complete genome, Type 4 strains could be classified as subtype 1C.

A study of urine samples from 60 Hungarians showed that 25 (40%) excreted JCV (Stoner et al., 1998). All strains were identified as subtype 1A or subtype 1B. No Type 2, Type 3 or Type 4 strains were found in the Hungarian samples. This confirms that Type 1 is the European genotype of JCV. In the USA, subtype 1B predominates over subtype 1A (42–22%) (Agostini et al., 1996b). In contrast, among the Hungarian strains, subtype 1A predominated (64%), with the remainder (36%) classified as subtype 1B. We have previously speculated that the apparent recombinant strain Type 4 may have originated in the USA when the JCV carried by an African American recombined with a Type 1 virus of European origin (Agostini et al., 1996a). The failure to detect any Type 4 strains in Hungary, in contrast to the 16% found in the USA, tends to support this theory, but additional populations must be studied before it can be concluded that the emergence of Type 4 strains is strictly a New World phenomenon.

These four Type 1 related sequences (#123, #124, #402...
and Mad1) define a consensus sequence except at five positions where the strains are evenly divided between two different nucleotides (positions 1940, 2502, 2712, 3035 and 4286). Comparison of the consensus sequence with a Type 2 sequence (strain GS/B) and the consensus of Type 3 sequences from Africa showed that 64 sites are unique to Type 1 sequences (Fig. 1a). An additional 32 nucleotides are shared only with Type 2, while 25 are shared only with Type 3 strains (not shown). At two positions, all three genotypes display different nucleotides (positions 1786 and 2592).

While most of these nucleotide changes are silent, 21 of them change an amino acid residue (Fig. 1b). The most variable protein is VP1 with seven amino acid changes between the genotypes. An interesting amino acid mutation in the subtype 1C (Type 4) sequence changes the basic lysine residue in subtype 1A and 1B strains to the neutral threonine in Type 4 strains due to a Type 3-like mutation at nucleotide position 1959 (Agostini et al., 1996a). Another non-conservative amino acid change occurs in the zinc finger motif, an important functional region of the large T antigen. At amino acid position 301, Type 1, Type 2 (GS/B) and Type 4 have a hydrophilic residue (glutamine), but in Type 3 strains this residue is changed to the highly hydrophobic leucine. These kinds of amino acid changes could underlie the biological differences between JCV types and subtypes (Agostini et al., 1996b).

A phylogenetic tree based on the sequences compared in Table 1 illustrates the genetic relationships defined there (Fig. 2). Most closely related are the two subtype 1A strains, Mad1 and #124. The subtype 1B strain and Type 4 (#402) are more distantly related. Type 2 (GS/B) falls into a separate group, as do the Type 3 sequences from East Africa. Further definition of Type 2 sequences is in progress. The relationships among five Type 3 strains (three African and two African American) have been described (Agostini et al., 1997).

JCV genotypes are also of interest from an anthropological point of view as they may make it possible to trace the movement of human populations. The predominance of Type 1 in the USA correlates with the European origin of the majority of the population. Similarly, the finding of Type 3 in African Americans, but not in European Americans (Agostini et al., 1997), is consistent with the migration of American blacks from Africa. The significance of Type 2 (Asian) strains and their subtypes in this regard is currently under study. As an infectious disease, JCV strains can move within and between population groups, but the dominance of characteristic types in certain populations after hundreds of years of contact is consistent with JCV transmission early in life, either within the family or within the immediate community (Kitamura et al., 1994; Kunitake et al., 1995).

In summary, our results from complete genome coding region DNA sequence analysis validate the use of JCV genotyping in the short VP1 typing fragment amplified with primers JLP-1 and -4. Strains predicted to be subtypes 1A and 1B differ from each other by only 0.54 and 0.58%, whereas two subtype 1A strains differ by only 0.33%. Type 4, as defined by the VP1 fragment, actually represents a subgroup of Type 1 as previously suggested (Agostini et al., 1996a).

H.T.A. was supported in part by the Deutsche Forschungsgemeinschaft, Bonn (grant Ag 19/1-1). The support and encouragement of Dr Henry def. Webber is gratefully acknowledged. Brain tissue specimens and 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 21 October 1997; Accepted 13 November 1997