Phylogenetic evidence of widespread distribution of genotype 3 JC virus in Africa and identification of a type 7 isolate in an African AIDS patient

Marietjie Venter, Sheilagh B. Smit, Patricia Leman and Robert Swanepoel
Special Pathogen Unit, National Institute for Communicable Diseases, Private Bag X4, Sandringham 2131, South Africa

JC virus (JCV) is the cause of progressive multifocal leukoencephalopathy (PML) in immunocompromised patients. The paucity of reports from Africa has led to the hypothesis that PML is rare because of an absence of virus genotypes associated with the condition. Genotypes 3 and 6 have been identified in East and West Africa but the distribution of types across the rest of Africa is unknown. Full-length sequences of five JCV cerebrospinal fluid samples from PML patients in South Africa are reported here. Three isolates from African AIDS patients grouped with type 3A or 3B, and one with type 7, while one from a Caucasian leukaemia patient grouped with type 2D. Widespread distribution of type 3 on the continent may reflect migration patterns in antiquity, but this is the first report of type 7 in an African individual. Type 2D has only been isolated previously in South Asia, although transmission of this genotype to Europeans who later settled in South Africa is not unlikely.

Analysis of JCV from urine or the CNS has revealed the existence of at least seven genotypes, which appear to have evolved in separate geographic regions conforming to global patterns of human migration (Agostini et al., 2001; Hatwell & Sharp, 2000; Jobes et al., 1998; Stoner et al., 2000). Type 1 is found in Europe and the USA and type 2 in Asia, with 2A and 2C predominating in north-eastern Asia, India and in indigenous Americans (thought to have migrated from Asia to America), 2B in Eurasia, 2D in India and 2E in Guam (Agostini et al., 2001). Types 3 and 6 have been identified in East and West Africa, respectively. Type 4 is closely related to type 1 and has been identified only in the USA, while type 5 is a recombinant between type 2B and type 6 (Hatwell & Sharp, 2000). Type 7 is the dominant strain in southern China where it was isolated from the urine of immunocompromised individuals without PML (Ou et al., 1997), but it has also been identified in urine samples from an Indian hill tribe (Agostini et al., 2001). Type 7 is close to type 2, but is considered a distinct genotype (Hatwell & Sharp, 2000; Jobes et al., 1998).

In the USA, JCV genotype 2, especially 2B, was found most likely to cause PML (Agostini et al., 1997b, 1998) and the apparent lack of PML in AIDS patients in West Africa was ascribed to the absence of type 2 from that region (Chima et al., 1999). In France, most cases of PML were associated with JCV type 1 (66 %) or type 2 (19-7 %). Genotypes 1, 4, 2 and 3 were detected in 52-3, 30-8, 12-3 and 4-6 %, respectively, of urine specimens from France. Infection with JCV genotypes 1 and 2 was found to correlate with PML, while infection with type 4 could represent a lower risk for
PML. All type 3 infections from France were in PML patients from North Africa and it was suggested that PML is rare in Africa only because AIDS patients frequently die before the onset of PML (Dubois et al., 2001).

JCV has a supercoiled circular DNA genome of 5.1 kb divided into three main regions, the non-coding regulatory region containing the origin of replication (ori), the early region that encodes viral regulatory proteins (large T and small t) and the late region, which contains the structural proteins (VP1, VP2, VP3 and the agnoprotein). The regulatory region exhibits extensive variation, with rearrangements, insertions, duplications and deletions occurring in isolates from PML-affected brains relative to archetypal strains from the urinary tract (Agostini et al., 1999c; Jobs et al., 1998; Yogo et al., 1999). This may affect levels of virus transcription and replication, which may in turn alter the cellular host range and be responsible for switching between states of latent and lytic infection (Jensen & Major, 2001). Sequencing of the promoter/enhancer region may be responsible for switching between states of latent and lytic infection (Jensen & Major, 2001). Sequencing of the promoter/enhancer region may thus provide useful supplementary information (Agostini et al., 2001), but has limited value for typing, since all rearranged sequences appear to be unique (Agostini et al., 1999c). Genotyping is based on sequence analysis of either the VP1 region or the full genome with exclusion of the regulatory region that encodes viral regulatory proteins (large T and small t). The regulatory region was used for genotyping as described in Hatwell & Sharp (2000) and the minimum evolution tree is shown in Fig. 1. Regulatory regions of the South African isolates were analysed separately in comparison with the archetypal sequence (Yogo et al., 1990). P distances were calculated with MEGA version 2.1 (Kumar et al., 2000).

In the present study we determined full genome sequences of five JCV isolates from PML cases in South Africa, the first to be reported from this part of the world. For detection of JCV, cerebrospinal fluid samples from four African AIDS patients (SA28_03; SA84_00; SA296_02; SA21_01) and one 77-year-old Caucasian of Afrikaans descent with advanced leukaemia (SA27_03) were heated to 95°C for 10 min and first-round PCR was performed using the forward primer Jig 3 (5′-CATGCTCTTCTGTTACGTTG-3′) (Agostini et al., 1997a) and the reverse primer Jig 3R (5′-TGATT-ACAGCATTTTTGTCTG-3′), while the nested PCR was carried out using primers JLP15 and JLP16 (Newman & Frisque, 1999). Full genome amplification was achieved using the method of Agostini & Stoner (1995), followed by nested PCR of overlapping regions to obtain sufficient amounts for direct cycle sequencing using the primers of Agostini et al. (1997a). Nucleotide sequencing was carried out on both strands using the ABI Prism Big Dye Terminator Cycle sequencing reaction kit v2.0 and analysed on an ABI 377 sequencer (Perkin-Elmer Applied Biosystems). Phylogenetic analysis of the full genome minus the regulatory region was used for genotyping as described in Hatwell & Sharp (2000) in comparison with previously typed sequences (Agostini et al., 2001; Hatwell & Sharp, 2000; Jobs et al., 1998).

Nucleotide alignments of 4846 sites starting at the initiation codon of the agnoprotein gene were generated with CLUSTAL_X 1.64b (Thompson et al., 1997) using the multiple alignment option. The first position of the agnoprotein was counted as nt 275 to reflect the nucleotide positions described for the VP1 typing sites in Stoner et al. (2000). The maximum-parsimony and neighbour-joining methods of PHYLIP version 3.5c (Felsenstein, 1993) and the neighbour-joining and minimum evolution methods of MEGA version 2.1 (Kumar et al., 2000) produced essentially identical trees under 100 bootstrap evaluations. The midpoint rooting option in MEGA was adopted as advocated by Hatwell & Sharp (2000) and the minimum evolution tree is shown in Fig. 1. Regulatory regions of the South African isolates were analysed separately in comparison with the archetypal sequence (Yogo et al., 1990). P distances were calculated with MEGA version 2.1 (Kumar et al., 2000).

In the analysis of the full genome coding region of the five South African JCV isolates, three from African AIDS patients of the Xhosa tribe clustered as subgenotypes of type 3, the fourth isolate from a member of the Zulu tribe clustered with type 7 and the isolate from a Caucasian leukaemia patient clustered with type 2 (Fig. 1). Genotyping was accomplished by comparison with reference sequences from all genotypes identified to date (Agostini et al., 2001). The phylogenetic analysis shown in Fig. 1 included all subtypes, all available type 7 strains and representative sequences of all type 2 and 3 subtypes. In the initial phylogenetic analysis, the South African sequences were aligned with all 261 complete genome sequences in GenBank to identify strains that were the most closely related to the South African strains.

Genotype 3 isolate SA28_03 clustered with subtype 3A and was most closely related to isolate 308 from Tanzania (P distance=0.002) (accession no. U73500.1). Isolates SA84_00 and SA296_02 clustered with subtype 3B, and both had P distances of 0.002 compared with isolate 311, which was isolated from an African-American and the only other full genome sequence of this subgenotype in GenBank (Fig. 1). A number of unique nucleotide substitutions in the South African isolates relative to the other subtype 3 strains was visible in the nucleotide alignment, including nt 1648 and 1816 in VP1 of SA28_03 and nt 1915 and 2146 in VP1 of SA84_00. A deletion was also identified in the non-coding region between VP1 and the large T antigen of SA296_02 (results not shown). Amino acid sequences for the individual proteins were 99–100 % identical to other type 3 isolates, except for the VP1 proteins of SA84_00 and SA28_03, which were 98 % identical. A number of unique amino acid substitutions were visible in isolates SA296_02, SA28_03 and SA84_00 relative to all other JCV isolates (Fig. 2). Subtype 3A was identified in urinary isolates in Tanzania, East Africa, while subtype 3B was found in West Africa where most African-Americans originated; the subtype has been identified in African-Americans and in a Gambian AIDS patient that died of PML in Germany (Agostini et al., 1997d). The South African subtype 3 isolates are the first to be isolated from confirmed PML cases in AIDS patients in Africa.

The fourth South African isolate, SA21_01 from an AIDS patient of the Zulu tribe in the Kwazulu-Natal province, clustered with JCV type 7 (Fig. 1) and was closest to two
isolates from Taiwan (P distance = 0.002) (accession nos AB077873.1 and AB077876.1) (results not shown). Unique nucleotide substitutions in the genome of SA21_01 relative to all other JCV isolates were visible in nt 1756 in VP1 and nt 3483, 3500, 3533 and 3640 in the large T antigen (results not shown). Amino acid changes occurred in the large T antigen in aa 344 (T→S), 363 (V→L) and 396 (H→L) (Fig. 2). Type 7 has not been identified in Africans before.

<table>
<thead>
<tr>
<th>Large T antigen</th>
<th>Small 1 antigen Agnoprotein</th>
<th>VP1</th>
<th>VP2</th>
<th>VP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>111122333 3330445566</td>
<td>66</td>
<td>1</td>
<td>1111111 222333333</td>
<td>1111122 222223333</td>
</tr>
<tr>
<td>2119189024 6690193505</td>
<td>67</td>
<td>281</td>
<td>34556</td>
<td>5601123669 66602334</td>
</tr>
<tr>
<td>9992104134 346873573</td>
<td>70</td>
<td>608</td>
<td>58391</td>
<td>506378485 57901255</td>
</tr>
<tr>
<td>Type 1A</td>
<td>124</td>
<td>VMTGCEQT VHDIESRES NF</td>
<td>VGR</td>
<td>DSRQT</td>
</tr>
</tbody>
</table>

Fig. 2. Predicted amino acid changes in the South African JCV sequences. The alignments were relative to type 1 and representatives of the each genotype shown in Fig. 1 are included in the comparison. Only sites that were variable in at least one sequence are shown. Nucleotide numbers are shown vertically above the sequences.
It remains to be determined how widespread type 7 infections are in this population group.

Isolate SA27_03 from a Caucasian leukaemia patient of Afrikaans descent clustered with JCV type 2D (Fig. 1). The closest related sequences were isolates FA-1a and FA1 from Japan (accession nos AB103402.1, AB103387.1), strain 246D, 247D and 248D from China (accession nos AF363832.1, AF363833.1 and AF363834.1) and isolate MO-3 from Mongolia (AB048551.1), all with P distances of 0.004 to the South African isolate. Although they are relatively rare, subtypes 2A, 2B and 2C have previously been found in Caucasians, constituting <10% of JCV in the urine of European Americans and Germans, while type 2D was found in South India. The boundary region in eastern Europe or eastern Asia where predominance of the type 1/4 clade switches over to type 2 has not yet been identified (Agostini et al., 2001; Jobes et al., 1998). Afrikaners derive mainly from Dutch (the first settlers arrived in 1652), French and German (16th century) and British (19th century) settlers (Hammond et al., 1997), suggesting that JCV infection in the population will probably be similar to that found in western Europe. Although type 2D has not been reported in Europeans before, this may reflect transmission of type 2D from southern Asia to Europeans in antiquity.

All five South African PML isolates had rearranged regulatory regions relative to the archetype. The subtype 7 isolate, SA21_01, was characterized by insertions of up to 11 nt in five positions. Deletions or substitutions occurred in all of the other isolates, with the largest deletion occurring in SA28_03 at nt 36–69 (results not shown).

This is the first report of JCV-associated cases of PML in southern Africa and provides the first evidence that genotype 3 occurs as far south as South Africa and that subtype 3A may be associated with PML. The occurrence of subtype 3 in the region is to be expected: the affected patients were of the Xhosa tribe, believed to have moved south with the other Bantu peoples of South Africa during the great migration of the Niger–Congo linguistic groups in the last three centuries BC. Both East and West African migrants took part in the southern migration, which may explain the presence of subtypes 3A and 3B in South Africa (Oliver, 1991). More surprising is the isolation of JCV type 7 from an African patient and its association with PML. The genotype is known in Asia, but since large numbers of Indians have settled in the KwaZulu-Natal province over the past two centuries, it cannot be excluded that this case resulted from a recent transmission event. The relative occurrence of JCV genotypes in the province is unknown, but Agostini et al. (2001) noted that African subtype 3 isolates were closer to types 2 and 7 than to types 1, 4 or 6 and suggested the existence of an ancient Asian–African link. The subtype 2D isolate identified in the leukaemia patient of Afrikaans descent probably reflects transmission events in Eurasia. A number of unique amino acid substitutions in South African sequences relative to the other isolates are shown in Fig. 2 in the large T antigen, VP1, VP2 and VP3 proteins. Although the biological significance of amino acid changes in these proteins is not yet fully understood, these changes may possibly influence the disease-causing potential, functionality and tropism of the virus and explain differences in strains found in urine, brain tumours and PML (Agostini et al., 1997b; Boldorini et al., 2003; Caldarelli-Stefano et al., 2000). Finally, the present findings confirm that PML-associated South African JCV isolates have rearranged regulatory regions relative to archetypal virus. Further investigations are needed to determine more accurately the incidence of PML in southern Africa and the relative frequencies of JCV genotypes in the region.

Acknowledgements

Our sincere thanks to Dr Lucille Blumberg for liaising with clinicians to obtain clinical information about the PML cases and to the medical staff of various public and private laboratories who have provided us with specimens of suspected cases. The Special Pathogen Unit at the NICD acts as a World Health Organization reference centre for viral haemorrhagic fevers, arbovirus and other emerging diseases for Africa.

References


