Journal of General Virology (1998), **79**, 2491–2498. Printed in Great Britain

Phylogenetic analysis of 22 complete genomes of the human polyomavirus JC virus

David V. Jobes, Sylvester C. Chima, Caroline F. Ryschkewitsch and Gerald L. Stoner

Neurotoxicology Section, National Institute of Neurological Disorders and Stroke, Bldg 36, Room 4A-27, National Institutes of Health, Bethesda, Maryland 20892, USA

The polyomavirus JC (JCV) establishes a persistent infection in the kidneys, and is the virus agent that causes the demyelinating disease progressive multifocal leukoencephalopathy. PCR and DNA sequence analyses of partial JCV genomes have shown that there are at least four main JCV types, each associated with a specific geographical region. Type 1 is of European origin, Type 2 is Asian, Type 3 is found in individuals of African decent and Type 4 is a potential recombinant of Types 1 and 3, and is widely distributed throughout the population of the United States. A comprehensive phylogenetic analysis of 22 complete JCV genomes excluding part of the regulatory region was accomplished using neighbour-joining, UPGMA and maximum parsimony methods. The resulting UPGMA and parsimony phylogenies suggest that the European Type 1 strains diverged from the other types during the evolution of JCV and that each of the other genotypes (and subtypes) falls into well-supported clades. This is the first whole genome approach to phylogeny reconstruction for JCV and represents a significant improvement over earlier studies that were limited to partial JCV sequences and the neighbour-joining method.

Introduction

The polyomavirus JC (JCV), a member of the family *Papovaviridae*, is the causative agent of the central nervous system demyelinating disease progressive multifocal leukoencephalopathy (PML), a condition usually associated with immunocompromised individuals (Berger & Concha, 1995). A site of persistent infection is the kidneys, from which the virus is silently shed in the urine. Upon reactivation it selectively infects oligodendrocytes and astrocytes in the brain (Major *et al.*, 1992; Walker, 1985). Prior to the AIDS epidemic PML cases were uncommon, usually occurring in older patients, patients receiving immunosuppressive agents, or rarely in patients with no identifiable immunodeficiency. PML now affects approximately 5% of AIDS patients and is the AIDS defining illness in 25% of these cases (Berger & Levy, 1993). Up to 40% of immunocompetent Caucasians excrete JCV in the urine as detected by PCR (Agostini *et al.*, 1996), providing an easily accessible source of JCV DNA.

Genotype analysis of JCV from both PML and non-PML patients revealed at least four main types that have evolved in specific geographical regions. JCV Type 1 is of European origin, Type 2 is Asian, Type 3 is found only in Africans and African Americans and Type 4, a putative recombinant of Types 1 and 3, is found throughout the United States. Genotype profiles were elucidated by sequencing a 610 bp fragment, termed the V–T intergenic region (Ault & Stoner, 1992), or by sequence analysis of a short 129 or 215 bp segment of the VP1 coding region, upstream of the V–T region (Agostini *et al.*, 1998*b*, *c*). The evidence suggests that JCV genotyping based on the short VP1 fragments is congruent with that defined by the whole genome (Agostini *et al.*, 1998*b*).

Sequence analysis of the prototype JCV Type 1 genome (JCV Mad-1) from a PML patient revealed a circular double-stranded DNA molecule of 5130 bp divided roughly in half by the early and late region genes (Fig. 1). The early region encodes the viral regulatory proteins, the small t and large T antigens, which share N-terminal sequences, but have different C-terminal sequences. The late region encodes the VP1, VP2 and VP3 capsid proteins. The VP3 coding region actually resides within the VP2 coding region, and the 5′ end of VP1 starts within the 3′ end of VP2/3 (Fig. 1). All JCV types also encode a fourth late protein, the agnoprotein, which may serve to facilitate nuclear localization of the major capsid protein (VP1) or to increase the efficacy of virus transmission from cell to cell (Cole, 1996; Shishido *et al.*, 1997).

While the JCV coding regions are highly conserved in both

Author for correspondence: Gerald Stoner.
Fax +1 301 402 1030. e-mail stoner@helix.nih.gov

0001-5595 © 1998 SGM

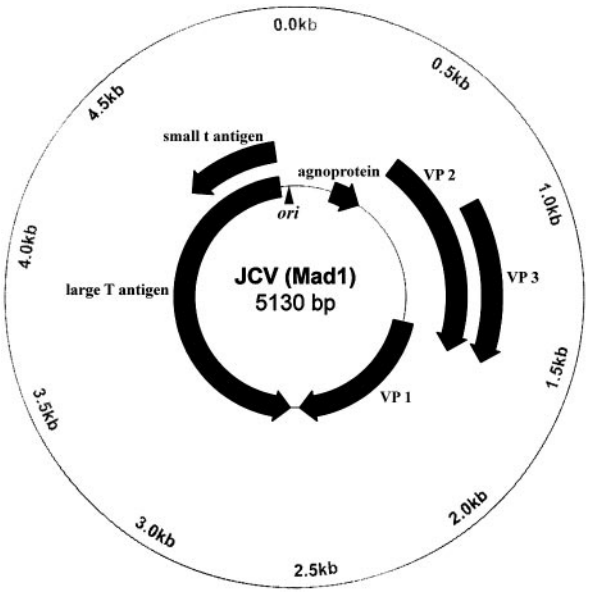

Fig. 1. Schematic diagram showing the gene arrangement of the 5·1 kb double-stranded DNA genome of JCV Mad-1 (Frisque *et al.*, 1984). Early proteins: large T and small t antigens. Late proteins: VP1, VP2, VP3 (capsid proteins) and the agnoprotein. The viral origin of DNA replication (*ori*) is also indicated.

length and sequence, there is extensive strain-to-strain variability in the non-coding regulatory region of PML-derived JCV (designated PML-type). This variation in genome size between different strains is attributed to deletions, duplications and rearrangements in the JCV regulatory region (which includes the viral origin of DNA replication) located between the 5′ end of the agnoprotein gene and the 5′ end of the small t and large T antigen genes (Fig. 1). It has been shown that all the unique rearrangements could be derived directly from the non-rearranged archetypal (kidney) form of the virus (Ault & Stoner, 1993; Agostini *et al.*, 1997*b*). The regulatory region was excluded from this phylogenetic analysis for two reasons. First, elimination of the variable regulatory region allows a direct comparison of both PML and non-PML JCV types because accurate DNA sequence alignment of this region is extremely difficult due to the unique rearrangement patterns observed in PML-type JCV. Second, because there does not appear to be any association between rearrangement patterns and genotypes, the regulatory region would not reflect the true evolution of the JCV types, but would instead merely indicate variation within an individual strain (ontogeny), which is not phylogenetically informative.

The *Papovaviridae* family of viruses is composed of the polyomaviruses (JCV) and papillomaviruses. The polyomaviruses and papillomaviruses are unrelated members of the family that share only superficial characteristics, including a double-stranded DNA genome and a virus capsid composed of 72 pentameric capsomers (Brady & Salzman, 1986; Walker & Frisque, 1986). Traditionally, the polyomaviruses have been classified according to the small genome size (~ 5·1 kb), the circular, supercoiled nature of the double-stranded DNA genome and the arrangement of genes into early and late region on opposite strands within the genome (Brady & Salzman, 1986; Walker & Frisque, 1986). These broad characteristics, however, are not useful for determining taxonomic relationships within the polyomaviruses, as they are shared by all members.

The utility of DNA sequences for elucidating the evolutionary relationships of organisms is well-documented (Cann *et al.*, 1987; Chase *et al.*, 1993; Field, 1988; Woese, 1987; Woese *et al.*, 1990). Previous phylogenetic analysis of JCV using the neighbour-joining method utilized a short, contiguous region located between the 3′ end of the VP1 gene and the 3′ end of the large T antigen, called the V–T intergenic region (Ault & Stoner, 1992; Sugimoto *et al.*, 1997). This 610 bp region also contains a short (approximately 75 bp) non-coding intergenic spacer that separates the VP1 and large T antigen genes. Due to the small size of the V–T intergenic region, the limited amount of sequence difference between closely related strains in the V–T region and the availability of phylogeny programs other than neighbour-joining, we undertook a phylogenetic assessment of 22 complete JCV genomes (minus the regulatory region). We used three separate methods: (1) neighbour-joining, (2) unweighted pair group method using arithmetic averages (UPGMA) and (3) maximum parsimony. This is the first reported whole genome approach determining the taxonomic relationships of all known JCV types and represents a significant improvement over earlier studies employing partial JCV sequences and which were limited to the neighbour-joining method.

Methods

■ **Sample collection information.** Urine samples, the source of JCV DNA, were collected from various individuals and relevant information for each is provided in Table 1.

■ **Phylogenetic analysis.** A multiple sequence alignment of 22 JCV types (Table 1) was performed using the Pileup alignment algorithm using a gap creation penalty of 5 and a gap extension penalty of 1 and saved as a multiple sequence format (MSF) file. A total of 4854 bp was aligned for each strain and two gaps were required to optimize the alignment (4856 total characters). The aligned sequences were then subjected to three separate analyses: (1) neighbour-joining, (2) UPGMA and (3) maximum parsimony (PAUP) using the DISTANCES and GROWTREE programs for generating neighbour-joining and UPGMA trees and PAUPSEARCH and PAUPDISPLAY (part of PAUP 4.0*) (Swofford, 1993) for the parsimony analysis. The neighbour-joining and UPGMA trees were calculated using both the Jukes–Cantor and Kimura two-parameter correction methods (Jukes & Cantor, 1969; Kimura, 1980). The resulting trees were identical regardless of the correction method used. All programs are available in version 9.1 of the Genetics Computer Group (GCG) package (Oxford Molecular Group).

■ **Neighbour-joining and UPGMA.** Of the phylogeny methods, cluster analysis is the simplest and often fastest approach to tree construction. The UPGMA clustering (Sokal & Michener, 1958) method

Table 1. List of 22 JCV types used for phylogenetic analysis separated by genotype and strain name and number

All sequences have been submitted to GenBank and accession numbers are indicated. Abbreviations: PML, progressive multifocal leukoencephalopathy; TB, tuberculosis; MS, multiple sclerosis; HIV, human immunodeficiency virus; SLE, systemic lupus erythematosus; RA, rheumatoid arthritis. Strains Mad-1, GS/B and Tai-3 were not sequenced at the NIH. Analysis of the sequence of the Tokyo-1 strain was initiated in Japan and completed at the NIH.

Genotype	Strain no.	Ethnic group	Sex/age (years)	Disease/sample	Place*	GenBank acc. no.	Reference
Type 1							
1A	Mad-1	Caucasian	M/38	Hodgkin's/PML brain	WI	J02227	Frisque *et al.* (1984)
1A	#124	Caucasian	M/52	MS/urine	CA	AF015526	Agostini *et al.* (1998*b*)
1B	#123	Caucasian	M/55	MS/urine	CA	AF015527	Agostini *et al.* (1998*b*)
Type 2							
2A	Tokyo-1	Japanese	M/70	TB/PML brain	Japan	AF030085	Agostini *et al.* (1998*c*)
2A	#224	Hispanic	M/72	MS/urine	CA	AF015529	Agostini *et al.* (1998*c*)
2A	#225	Native American	F/26	Unknown/urine	NM	AF015530	Agostini *et al.* (1998*c*)
2A	#226	Native American	F/78	Unknown/urine	NM	AF015531	Agostini *et al.* (1998*c*)
2B	GS/B	German	Unknown	Lymphoma/PML brain	Germany	M20322	Loeber & Dörries (1988)
2B	#223	African American	M/36	HIV/urine	CA	AF015532	Agostini *et al.* (1998*c*)
2B	#227	Caucasian	M/41	MS/urine	CA	AF015533	Agostini *et al.* (1998*c*)
2C	#228	Native American	M/63	Unknown/urine	NM	AF015534	Agostini *et al.* (1998*c*)
2C	#229	Caucasian	M/48	HIV/urine	CA	AF015535	Agostini *et al.* (1998*c*)
2C	#230	African American	M/44	MS/urine	CA	AF015536	Agostini *et al.* (1998*c*)
Type 3							
3A	#308	African	Unknown	Unknown/urine	Tanzania	U73500	Agostini *et al.* (1997*a*)
3A	#309	African	Unknown	Unknown/urine	Tanzania	U73500†	Agostini *et al.* (1997*a*)
3A	#310	African	Unknown	Unknown/urine	Tanzania	U73500†	Agostini *et al.* (1997*a*)
3A	#312	African American	M/41	MS/urine	CA	U73502	Agostini *et al.* (1997*a*)
3B	#311	African American	M/59	Control/urine	MD	U73501	Agostini *et al.* (1997*a*)
Type 4	#402	Caucasian	M/63	MS/urine	CA	AF015528	Agostini *et al.* (1996)
Type 5	#501‡	Caucasian	M/30	MS/urine	CA	AF015684	Agostini *et al.* (1998*c*)
Type 6	#601	African American	F/35	SLE/PML brain	NJ	AF015537	Unpublished data
Type 7	Tai-3	Chinese	Unknown	RA/urine	Taiwan	U61771	Ou *et al.* (1997)

* American state or country.
† Described in relation to strain #308 using the same accession number.
‡ Previously designated as strain #X01 by Agostini *et al.* (1998*c*).

calculates the distance (dissimilarity) between taxa, which are termed operational taxonomic units (OTUs). The two OTUs with the smallest distance are clustered together forming a composite OTU which is then treated as a single OTU for further clustering. Each round of clustering generates a larger composite OTU until the last two (a single OTU and the composite OTU) are joined, forming the completed UPGMA tree.

The neighbour-joining method (Saitou & Nei, 1987) is conceptually similar to clustering except that, instead of combining individual taxa, the method starts with a star-like graph (essentially an unrooted, unresolved phylogenetic tree) and then determines the closest OTUs. Unlike UPGMA, which focuses on clustering taxa, neighbour-joining keeps track of nodes of the tree to minimize the total branch lengths at each stage of clustering. The neighbour-joining method is extremely fast and is able to handle very large data sets.

■ **Maximum parsimony.** Due to the size of the data set for the parsimony analysis (22 strains, 4854 bp), a heuristic search using SIMPLE stepwise addition, MULPARS (an option to keep all most parsimonious trees) and tree bisection-reconnection (TBR) branch swapping was performed in PAUPSEARCH. A strict consensus tree of the seven most parsimonious trees was then generated. To test the support of individual clades in the parsimony analysis, 100 bootstrap replicates were performed using the heuristic search option and TBR, offered in the PAUPSEARCH program. Bootstrap analysis (Felsenstein, 1985) is a statistical method that involves randomly sampling a perturbed data set. Each new perturbation creates a new data set that was generated by randomly resampling the original data (sites within a DNA sequence alignment) with replacement, creating a data set that represents some characters that are used more than once, some that are used only once and others that are not used at all. The 100 bootstrap replicates were used as an index of support for each group (clade) and the bootstrap percentage value (the percentage of occurrences of a particular group) was placed above the branch leading to that group (Fig. 2D).

Results

The phylogenetic relationships of JCV, as determined by neighbour-joining, UPGMA, maximum parsimony and boot-

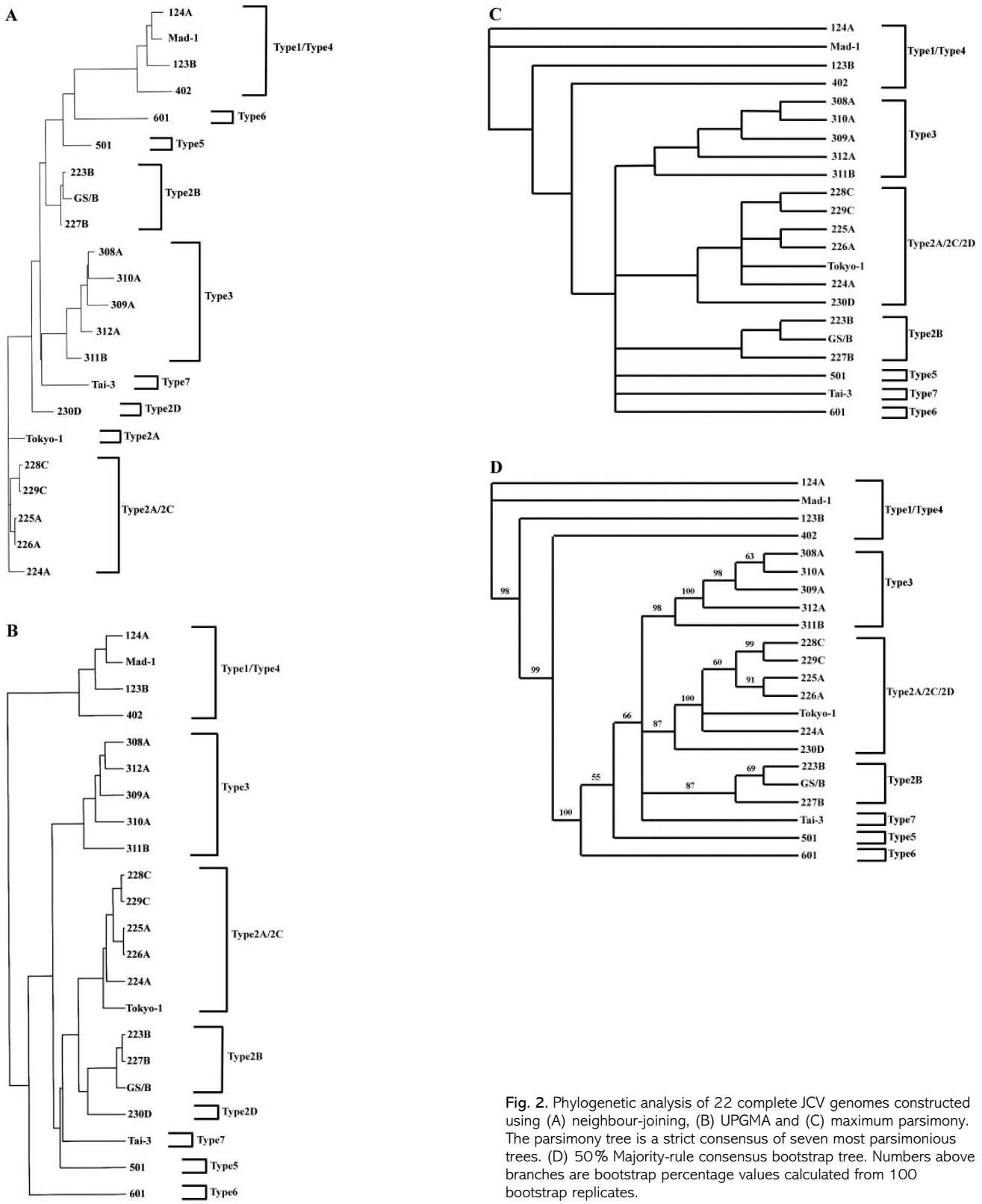

Fig. 2. Phylogenetic analysis of 22 complete JCV genomes constructed using (A) neighbour-joining, (B) UPGMA and (C) maximum parsimony. The parsimony tree is a strict consensus of seven most parsimonious trees. (D) 50% Majority-rule consensus bootstrap tree. Numbers above branches are bootstrap percentage values calculated from 100 bootstrap replicates.

strapping are presented in Fig. 2(A–D). The maximum parsimony approach generated seven most parsimonious trees with a tree length of 440 steps, a consistency index of 0·7864 and a retention index of 0·8547. Of the 4856 total characters in the data set (the two gaps were treated as missing data), 4523 characters were constant, 161 were phylogenetically uninformative and 172 were phylogenetically informative sites.

In all most parsimonious trees, strains #124A and Mad-1 are consistently placed as the most basal group. Strains #123B and #402 are also clustered with #124A and Mad-1, suggesting a close affinity within the Type 1 group. The Tanzanian (#308A, #309A, #310A) and African American (#311B, #312A) strains also form a clade that is present in all three analyses, although in the UPGMA tree, the Tanzanian and African American types are not separated according to group, as they are in the neighbour-joining and parsimony trees (Fig. 2B). Instead, strains #308A and #312A are placed together, with #309A as a sister group. Strains #310A and #311B are then positioned as the basal groups within this Type 3 clade. Strain #311B was originally classified tentatively as a possible Type 5 JCV, but was later reclassified as Type 3, based on the 99·48% sequence similarity to a consensus Tanzanian sequence (Agostini *et al*., 1997*a*). This phylogenetic analysis further supports the reclassification and shows that strain #311B clusters with the Type 3 clade in all phylogenetic trees (Fig. 2 A–D).

Both UPGMA and parsimony methods provide clearly resolved taxonomic relationships of the Type 2 JCV strains. The 228C/229C and the 225A/226A clades are supported using all three approaches, but the placement of Tokyo-1 and #230D varies according to the method employed. UPGMA and parsimony place Tokyo-1 close to the Type 2A/2C clade, whereas the neighbour-joining tree placement is somewhat ambiguous, occupying a lone branch between Type 2A/2C and the rest of the types (Fig. 2A). Interestingly, strain #230D is positioned with the Type 2A/2C clade using parsimony, with the Type 2B clade using UPGMA and is actually placed as the basal member of the group comprised of Type 7, Type 3, Type 2B, Types 5 and 6 and Types 1 and 4 in the neighbour-joining tree. The Type 2B group is placed together as a clade separate from the Type 2A clade with the only difference being the placement of strains #223B with #227B in the UPGMA tree as opposed to GS/B in the neighbour-joining and parsimony trees (Fig. 2A–C). Contrary to the UPGMA and parsimony trees, the neighbour-joining tree actually separates the Type 2A/2C and 2B group by placing the Type 3 clade as an intermediate and placing the Type 2A/2C clade as the most basal assemblage (Fig. 2A).

The remaining JCV types fall into very different arrangements. The single Type 4 sequence (#402) is consistently placed with the Type 1 sequences with neighbour-joining and UPGMA, but in an intermediate position between the Type 1 group and the remaining types using parsimony.

Table 2. Comparison of JCV genotypes elucidated by whole genome analysis (this paper) and V–T intergenic region analysis alone (Sugimoto *et al*., 1997)

In order to make a direct comparison, V–T intergenic region sequences were extracted from the whole genome sequences and compared to the genotypes determined by Sugimoto *et al*. (1997). JCV types not available for comparison are indicated by dashes.

Whole genome genotypes (Type and subtype)	V–T region genotypes (Type and subtype)
Type 1A	Type A subtype EU
Type 1B	Type A subtype EU
Type 2A	Type B subtype MY
Type 2B	Type B subtype B1
Type 2C	Type B subtype MY
Type 2D	Type B subtype B1
Type 3	Type B subtype AF2
Type 4	Type A subtype EU
Type 5	–
Type 6	Type C subtype AF1
Type 7	Type B subtype SC
–	Type B subtype CY
–	Type B subtype AF3
–	Type B subtype B2

Neighbour-joining places Tai-3 (originating in Taiwan and tentatively designated Type 7) as a sister group to the Type 3 clade, while both UPGMA and parsimony suggest a closer association with the Asian Type 2 groups. Strain #501 is the only representative of a putative Type 5 JCV. Of the seven most parsimonious trees, four suggest a very tight association with the only Type 6 sequence analysed (strain #601), located basal to Types 2 and 3, while the remaining three most parsimonious trees place strain #501 with the Type 2B group (not shown). Interestingly, the neighbour-joining tree supports an affinity of the Type 5 sequence with the Type 1 group (Type 6 is also placed in this group). The Type 6 sequence, represented here by strain #601, is placed at various positions in the phylogenetic trees, grouping either with Type 1 (neighbour-joining), or as a strain located basal to the Type 2 and Type 3 clades (Fig. 2A–C).

Fig. 2(D) is a 50% majority-rule consensus bootstrap tree derived from parsimony and shows that all JCV types are strongly supported groups, as indicated by the high bootstrap percentage values ranging from 55–100%.

Table 2 shows a comparison of JCV genotypes based on two different approaches, whole genome analysis and the V–T intergenic region (Sugimoto *et al*., 1997). Although there are several cases where a direct comparison is not possible, both methods have determined that there are multiple genotypes (genetic variants) of JCV and that each can be found in specific ethnic groups (also see Table 1 for ethnic groups).

Discussion

Whole genome versus partial sequences for JCV phylogeny reconstruction

Previous attempts at constructing a molecular phylogeny of JCV genotypes utilized the small V–T intergenic region (610 bp) and the neighbour-joining method (Sugimoto *et al*., 1997). Although it is the most variable area in the JCV genome (not including the regulatory region), the V–T region nevertheless shows little variability between closely related JCV strains and may not provide enough informative sites (Li & Graur, 1991) to construct an accurate phylogenetic tree. Utilizing the whole JCV genome, minus the regulatory region (4854 bp), substantially increases the number of phylogenetically informative sites and more adequately resolves relationships between the JCV genotypes. To test this, a phylogenetic tree (with the same taxa as the whole genome approach) using the 610 bp V–T intergenic region was constructed with the parsimony option of PAUP and with UPGMA and neighbour-joining (data not shown). Parsimony analysis showed that of the 611 total characters (a single gap was required in the alignment), 534 sites were invariant, 36 were phylogenetically uninformative and only 41 sites were informative. In contrast, of the 4856 characters in the whole genome data set, 4523 were invariant between the strains, 161 were uninformative and 172 sites were phylogenetically informative. The whole genome approach, therefore, provides a fourfold increase of informative sites over the V–T region alone. This increase in informative sites translates into a much better resolved phylogeny for JCV. For instance, using only the V–T region, parsimony was unable to resolve any clear relationships between strains 308, 309 and 310 (i.e. the consensus tree showed a polytomous relationship in this clade). Also, the V–T region sequences placed strain Tai-3 with the Type 3 group with all three methods, which, although similar to the neighbour-joining tree constructed from whole genome sequences, is very different from both the UPGMA and parsimony trees generated from whole genomes (Fig. 2A–C). To further complicate matters, the Type 2 strain #224A was assigned an ambiguous and unresolved position in the UPGMA and neighbour-joining trees due to a long branch length. Whole genome sequences, however, consistently placed #224A with the Type 2A/2C group (which includes Type 2D using parsimony), regardless of the method used.

For the maximum parsimony method, which traces the evolution of individual characters (i.e. nucleotides), an increase in the number of informative sites allows a more highly resolved phylogeny and hence more accurately depicts the 'true' phylogeny. Likewise, for the distance methods (UPGMA and neighbour-joining), an enlarged data set potentially contributes more variation to the analysis (up to 2·4% for JCV in this case), which helps differentiate between closely related strains. This is especially critical since only one tree is generated using the UPGMA and neighbour-joining methods, even though there may be other possible arrangements of the JCV types (e.g. if three JCV types are all 1% different, there could be three possible arrangements, but only one is displayed). The small branch lengths (indicating very little sequence difference), poorly resolved relationships among several JCV lineages and low bootstrap percentage values observed in an earlier study (Sugimoto *et al*., 1997), may indicate that there is not enough variation to accurately delimit specific types using only the 610 bp V–T intergenic region.

The problem of single, non-representative phylogenetic trees is circumvented when using maximum parsimony, since all most parsimonious trees are calculated and displayed. The parsimony principle states that simpler hypotheses are preferable to complicated ones (i.e. the phylogenetic tree that requires the fewest number of evolutionary steps to explain the data is preferred). Parsimony treats each nucleotide as a discrete, individual character, meaning that the complete genome of the JCV data set provides 4854 potential characters from which to reconstruct the evolutionary relationships of the different JCV types. However, since most of the sites are invariant (4523), only a small percentage ($172/4854 = 3{\cdot}5\%$) will actually contribute to the phylogenetic analysis. It is important to point out, however, that these informative sites are present throughout the genome and therefore represent nucleotide evolution in JCV on a genome-wide scale better than analysis of partial JCV sequences. In other words, base compositional bias (mutation bias) that may act on one region of the genome would not adversely affect a phylogeny if whole genome sequences were used to create the phylogeny.

The divergence of JCV Type 1 from other JCV genotypes

Based on the UPGMA and parsimony phylogenetic trees, the JCV Type 1 strains appear to have diverged early in the evolution of JCV (Fig. 2B–D), although it is difficult to assign a time to this divergence. These data, however, do not necessarily indicate that JCV Type 1 is an ancestral state of JCV. Type 1 is widely distributed (Europeans and European Americans) and may not have been under the same selective constraints as in some smaller, less widely distributed or genetically distinct populations. It is possible that JCV Type 1 may actually be evolving at a different rate than are other types.

The single Type 4 (#402) sequence is associated with the Type 1 clade using neighbour-joining and UPGMA (Fig. 2A, B). The parsimony tree, however, separates the Type 1 and Type 4 species, with #402 occupying a single lone branch close to, but distinct from, Type 1 (Fig. 2C). This relationship is supported by a high bootstrap value (99%; Fig. 2D). The majority of the Type 4 sequence is nearly identical to Type 1, but does contain parts of the VP1 gene from Type 3. Parsimony is able to discriminate between the fragments that are derived from Type 3 and does show that the unique derivation of the

Type 4 VP1 sequence supports the separation of Type 4 from the Type 1 clade. If future studies show that recombination in JCV is not unique, parsimony may become the method of choice for constructing phylogenetic relationships within this family.

While the three available phylogeny methods do not provide a completely congruent interpretation of the relationship of Type 1 to the other JCV types, it should be stressed that this preliminary study can be bolstered by the inclusion of more whole genome sequences to help bridge the gap between JCV Type 1 strains and the remaining taxa.

Individual clades are strongly supported by all three phylogeny methods

Despite the uncertain placement of JCV Type 1 in the overall phylogeny, all three phylogeny methods do clearly differentiate the types into distinct groups (clades) based on genotype. For instance, #124A, Mad-1, #123B and #402 (the possible Type 1/3 recombinant) are always grouped together, regardless of the method, and are also strongly supported by bootstrap values (Fig. 2D). Likewise, the Type 2A/2C, Type 2B and Type 3 groups are clearly distinct clades and do indeed represent individual genotypes. The inclusion of the Tai-3 strain (a possible Type 7) in the Type 2 clade (parsimony and UPGMA trees) suggests a still present link between the Asiatic Type 2 strains and this strain from Taiwan, a population originating from South China. Until more samples are available, the placement of #501 is still speculative, since the consensus parsimony tree yielded a polytomous branching order with Types 6 and 7. UPGMA and neighbour-joining also provided an unsatisfactory placement of #501.

It is tempting to speculate that the Type 6 strain (#601) branched off from the other African strains and represents an early African JCV type, but until more Type 6 whole genomes from Africa are analysed, it is not clear whether this single Type 6 sequence found in an African American is representative of a true consensus Type 6 genome sequence.

Relevance of molecular phylogenetics to the biology of JCV

JCV causes the chronic demyelinating disease of the central nervous system, PML, which occurs in a background of immunosuppression in humans. PML is usually fatal within 1 year of diagnosis. Rare cases have been reported in patients with no evident immunosuppression and patients occasionally show prolonged survival. Is it possible that the latter may represent less virulent strains of the infective agent? In a study of JCV genotypes in PML patients and urine from controls, Agostini *et al*. (1998*a*) reported that JCV Type 2 and subtype 2B are significantly more frequent in brain tissue from PML patients than in urine of normal controls within the same population. In another study of four cases of PML in West African AIDS patients, it was found that, although there were no significant differences in the pathomorphology of African and non-African PML, PCR analysis showed that two patients had the JCV Type 3 sequence and another had the JCV Type 6 sequence (one sample did not amplify). The African PML patients with the Type 3 sequence showed increased numbers of neoplastoid astrocytes and inflammatory infiltrates while the PML patient with JCV Type 6 displayed more extensive reactive astrocytosis (S. C. Chima, unpublished data). These results tentatively indicate genotype-specific differences in the biology of JCV.

The exploration of strain-to-strain differences among JCV isolates will be facilitated by the establishment of a reliable and consistent method for differentiation of the various JCV genotypes. The comprehensive molecular phylogeny presented here is a step toward a better understanding of relationships between the JCV genotypes and will hopefully aid in further molecular studies of diseases associated with human JCV.

The authors wish to thank Ms Beth Pernerewski for critical reading of the manuscript. We would also like to thank Dr Yukiko Shishido-Hara who determined the complete sequence of JCV (Tokyo-1). Brain tissue specimens and urine samples were obtained from the National Neurological Research 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

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

Agostini, H. T., Ryschkewitsch, C. F., Singer, E. J. & Stoner, G. L. (1997*b*). JC virus regulatory region rearrangements and genotypes in progressive multifocal leukoencephalopathy: two independent aspects of virus variation. *Journal of General Virology* **78**, 659–664.

Agostini, H. T., Ryschkewitsch, C. F., Singer, E. J., Baumhefner, R. W. & Stoner, G. L. (1998*a*). JC virus Type 2B is found more frequently in brain tissue of progressive multifocal leukoencephalopathy (PML) patients than in control urines. *Journal of Human Virology* **1**, 200–206.

Agostini, H. T., Ryschkewitsch, C. F. & Stoner, G. L. (1998*b*). JC virus Type 1 has multiple subtypes: three new complete genomes. *Journal of General Virology* **79**, 801–805.

Agostini, H. T., Shishido-Hara, Y., Baumhefner, R. W., Singer, E. J., Ryschkewitsch, C. F. & Stoner, G. L. (1998*c*). JC virus Type 2: definition of subtypes based on DNA sequence analysis of ten complete genomes. *Journal of General Virology* **79**, 1143–1151.

Ault, G. S. & 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.

Ault, G. S. & Stoner, G. L. (1993). Human polyomavirus JC promoter/enhancer rearrangement patterns from progressive multifocal leukoencephalopathy brain are unique derivatives of a single archetypal structure. *Journal of General Virology* **74**, 1499–1507.

Berger, J. R. & Concha, M. (1995). Progressive multifocal leukoencephalopathy: the evolution of a disease once considered rare. *Journal of Neurovirology* **1**, 5–18.

Berger, J. R. & Levy, R. M. (1993). The neurologic complications of human immunodeficiency virus infection. *Medical Clinics of North America* **77**, 1–23.

Brady, J. N. & Salzman, N. P. (1986). The papovaviruses. General properties of polyoma and SV40. In *The Papovaviridae*, vol. l, *The Polyomaviruses*, pp. 1–26. Edited by N. P. Salzman. New York: Plenum Press.

Cann, R. L., Stoneking, M. & Wilson, A. C. (1987). Mitochondrial DNA and human evolution. *Nature* **325**, 31–36.

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., Duval, R. R., Price, R., Hillis, H. G., Qui, Y.-L., Kron, K. A., Rettig, J. H., Conti, E., Palmer, J. D., Manhart, J. R., Sytsma, K. J., Michaels, H. J., Kress, W. J., Karol, K. G., Clark, W. D., Hedren, M., Gaut, B. S., Jansen, R. K., Kim, K.-J., Wimpee, C. F., Smith, J. F., Furnier, G. R., Strauss, S. H., Xiang, Q.-Y., Plunkett, G. M., Soltis, P. S., Williams, S. E., Gadek, P. A., Quinn, C. J., Eguiarte, L. E., Golenberg, E., Learn, G. H., Graham, S., Barrett, S. C. H., Dayanandan, S. & Albert, V. A. (1993). Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. *Annals of the Missouri Botanical Garden* **80**, 528–580.

Cole, C. N. (1996). Polyomavirinae: the viruses and their replication. In *Fields Virology*, 3rd edn, pp. 1997–2025. Edited by B. N. Fields, D. M. Knipe & P. M. Howley. Philadelphia: Lippincott–Raven.

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

Field, K. (1988). Molecular phylogeny of the animal kingdom. *Science* **239**, 748–753.

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

Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism*, pp. 21–132. Edited by H. N. Munro. New York: Academic Press.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **47**, 471–486.

Li, W.-H. & Graur, D. (1991). *Fundamentals of Molecular Evolution*. Sunderland, MA, USA: Sinauer Associates.

Loeber, G. & Dörries, K. (1988). DNA rearrangement in organ-specific variants of polyomavirus JC strain GS. *Journal of Virology* **62**, 1730–1735.

Major, E. O., Amemiya, K., Tornatore, C. S., Houff, S. A. & Berger, J. R. (1992). Pathogenesis and molecular biology of progressive multifocal leukoencephalopathy, the JC virus-induced demyelinating disease of the human brain. *Clinical Microbiology Reviews* **5**, 49–73.

Ou, W.-C., Tsai, R.-T., Wang, M., Fung, C.-Y., Hseu, T.-H. & Chang, D. (1997). Genomic cloning and sequence analysis of Taiwan-3 human polyomavirus JC virus. *Journal of the Formosan Medical Association* **96**, 511–516.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

Shishido, Y., Nukuzuma, S., Mukaigawa, J., Morikawa, S., Yasui, K. & Nagashima, K. (1997). Assembly of JC virus-like particles in COS7 cells. *Journal of Medical Virology* **51**, 265–272.

Sokal, R. R. & Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* **38**, 1409–1438.

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., Grésenguet, G., Kocagöz, T., Chaiyarasamee, S., Thant, K. Z., 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. *Proceedings of the National Academy of Sciences, USA* **94**, 9191–9196.

Swofford, D. L. (1993). PAUP: phylogenetic analysis using parsimony, version 3.1.1. Champaign, IL, USA: Illinois Natural History Survey.

Walker, D. L. (1985). Progressive multifocal leukoencephalopathy. In *Handbook of Clinical Neurology*, vol. 47, *Demyelinating Diseases*, pp. 503–524. Edited by P. J. Vinken, G. W. Bruyn, H. L. Klawans & J. C. Koetsier. Amsterdam: Elsevier.

Walker, D. L. & Frisque, R. J. (1986). The biology and molecular biology of JC virus. In *The Papovaviridae*, vol. 1, *The Polyomaviruses*, pp. 327–377. Edited by N. P. Salzman. New York: Plenum Press.

Woese, C. R. (1987). Bacterial evolution. *Microbiological Reviews* **51**, 221–271.

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eukarya. *Proceedings of the National Academy of Sciences, USA* **87**, 4576–4579.

Received 30 March 1998; Accepted 10 June 1998