

Utility of JC polyomavirus in tracing the pattern of human migrations dating to prehistoric times

Angelo Pavesi

Correspondence
Angelo Pavesi
angelo.pavesi@unipr.it

Department of Genetics Anthropology & Evolution, University of Parma, Parco Area delle Scienze 11/A, I-43100 Parma, Italy

JC virus (JCV) is a double-stranded DNA polyomavirus co-evolving with humans since the time of their origin in Africa. JCV seems to provide new insights into the history of human populations, as it suggests an expansion of humans from Africa via two distinct migrations, each carrying a different lineage of the virus. A possible alternative to this interpretation could be that the divergence between the two lineages is due to selective pressures favouring adaptation of JCV to different climates, thus making any inference about human history debatable. In the present study, the evolution of JCV was investigated by applying correspondence analysis to a set of 273 fully sequenced strains. The first and more important axis of ordination led to the detection of 61 nt positions as the main determinants of the divergence between the two virus lineages. One lineage includes strains of types 1 and 4, the other strains of types 2, 3, 7 and 8. The distinctiveness of the Caucasian lineage (types 1 and 4), largely diffused in the northern areas of the world, was almost entirely ascribed to synonymous substitutions. The findings provided by the subsequent axes of ordination supported the view of an evolutionary history of JCV characterized by genetic drift and migration, rather than by natural selection. Correspondence analysis was also applied to a set of 156 human mitochondrial genome sequences. A detailed comparison between the substitution patterns in JCV and mitochondria brought to light some relevant advantages of the use of the virus in tracing human migrations.

Received 29 September 2004

Accepted 3 February 2005

INTRODUCTION

The human polyomavirus JC (JCV) is a small DNA virus belonging to the family *Polyomaviridae*. Its genome is a single molecule of circular double-stranded DNA of 5.1 kb (Frisque *et al.*, 1984). Epidemiological studies have shown that JCV infection is widespread in the human population, with a seroprevalence rate ranging from 70 to 90% (Padgett & Walker, 1973). After primary infection, which occurs during childhood without any clinical manifestations, the virus persists lifelong in renal tissue. Reactivation of latent JCV is frequent in adults, as proved by the detection of viral progeny in the urine of a high percentage (20–80%) of healthy individuals over 30 years of age (Kitamura *et al.*, 1990; Chang *et al.*, 2002). The highest percentage of excretors has been found in populations of Asian descent (Agostini *et al.*, 1997a), while the lowest has been observed in Africans (Chima *et al.*, 1998) and Arctic tribes (Sugimoto *et al.*, 2002a).

Sequencing of JCV has revealed the existence of several distinct genotypes. Types 1 and 4, closely interrelated to each other, were found not only in Europe (Agostini *et al.*, 2001a), but also in indigenous populations living in northern Japan, North-East Siberia and northern Canada (Sugimoto *et al.*, 2002a; Yogo *et al.*, 2003). Types 3 and 6 are characteristic of sub-Saharan Africa: type 3 was isolated

in Ethiopia (Sugimoto *et al.*, 2002b), Tanzania (Agostini *et al.*, 1997b) and South Africa (Venter *et al.*, 2004), and type 6 in Ghana (Guo *et al.*, 1996; Kato *et al.*, 2000). Both genotypes were also found in the Biaka Pygmies and Bantus from Central Africa (Chima *et al.*, 1998). Types 2 and 7 show a large geographical distribution (Sugimoto *et al.*, 1997). Type 2 includes several variants, with subtype 2A mainly in the Japanese population and native Americans (excluding Inuits), 2B in Eurasians, 2D in Indians, and 2E in Australians and western Pacific populations (Fernandez-Cobo *et al.*, 2002; Yanagihara *et al.*, 2002; Zheng *et al.*, 2003; Miranda *et al.*, 2004; Takasaka *et al.*, 2004). Subtype 7A was found to be characteristic of southern China and South-East Asia (Saruwatari *et al.*, 2002), while subtype 7B of northern China, Mongolia and Japan (Sugimoto *et al.*, 2002b; Zheng *et al.*, 2004a). A third subtype (7C), spread throughout northern and southern China, has recently been characterized by Cui *et al.* (2004). Finally, type 8 was found in Papua New Guinea and the Pacific Islands (Jobes *et al.*, 2001; Yanagihara *et al.*, 2002).

The ubiquitous distribution of JCV, combined with a transmission mechanism largely within families or populations (Kunitake *et al.*, 1995; Kato *et al.*, 1997; Suzuki *et al.*, 2002; Zheng *et al.*, 2004b), make it an attractive candidate for reconstructing human migrations dating to prehistoric

times. The close relationship of JCV found in native Americans with that in North-East Asia is consistent with the migration of Amerindian ancestors from Asia across the Bering land bridge (Agostini *et al.*, 1997a). Doubts regarding the reliability of JCV as a marker of human evolution (Wooding, 2001) have recently been dispelled by a whole-genome phylogenetic analysis focused on the distinction between slow- and fast-evolving sites (Pavesi, 2003). By this approach, it was proposed that the association of JCV with humans originated in Africa, since type 6 was found to be the putative ancestral genotype. It was also demonstrated how type 6 gave rise to two independent evolutionary lineages: one including types 1 and 4, the other including types 2, 3, 7 and 8 (Pavesi, 2003).

The diffusion in the world of both lineages was elucidated through the analysis of over 1000 sequences of the genomic region of JCV with the highest variation rate (Pavesi, 2004). By using synthetic geographical maps, it was hypothesized that the expansion of *Homo sapiens* from Africa was mediated by two migration waves, each carrying a different virus lineage (Pavesi, 2004). This finding is a valuable one, because it sheds new light on the pattern of human evolution yielded so far by human genes, supporting the hypothesis of one single expansion from Africa into Asia and from there to the other continents (reviewed by Cavalli-Sforza & Feldman, 2003).

The view that the dual exit of JCV from Africa mirrors two migrations on the part of our ancestors is appealing. However, the objection can be raised that the present genetic diversity between the two virus lineages – one (types 1 and 4) mainly diffused in the northern areas of the world and the other (types 2, 3, 7, and 8) in the central and southern areas – is the result of selective pressures favouring adaptation to different climates. In this case, large-scale inferences concerning human evolution should be treated with caution, since a reliable reconstruction of human history is based on phenomena such as genetic drift or migration, and not natural selection (Cavalli-Sforza *et al.*, 1994). A possible response to this objection could be a more subtle analysis of the genome sequence of JCV, with the aim of characterizing the type of nucleotide substitutions causing the deep divergence between the two virus lineages.

In this study, I propose to illustrate an approach to investigate the evolution of JCV based on correspondence analysis (Lebart *et al.*, 1984). The main advantage of this method derives from a mathematically adequate representation of a set of related sequences. It allows not only an elucidation of the evolutionary relationships between sequences, as do the standard phylogenetic methods, but also the identification of those nucleotide positions where systematic changes have occurred in the past. Correspondence analysis was also applied to a large set of complete mitochondrial genomes, whose sequence has been made available by recent studies on global mitochondrial DNA (mtDNA) diversity in humans (Ingman *et al.*, 2000; Mishmar *et al.*, 2003; Ingman & Gyllenstein, 2003). Thanks to the elevated analytical power of

the method, a detailed comparison between the patterns of change underlying the evolution of JCV and human mtDNA is presented.

METHODS

Sequence data. The complete genome sequence of 275 JCV strains was collected from the EMBL database (Release 76). This corpus of data contains 12 isolates from Africa, 44 from North America, 10 from Central America, 15 from South America, 145 from Asia, 29 from Europe and 20 from Oceania. The control region (267 bp) was removed from each sequence, because it frequently reveals large deletions or duplications, especially in JCV strains isolated from the brain of patients with progressive multifocal leukoencephalopathy (Agostini *et al.*, 1997c). The length of the selected sequences was thus comprised between 4791 and 4860 bp. A total of 156 complete mtDNA sequences of humans of diverse ethnic origins was also collected from the EMBL database. This set of data includes 32 sequences from Africa, 9 from the Americas, 37 from Asia, 30 from Europe and 48 from Oceania. The D-loop control region (1120 bp) was removed from each sequence, because it shows a substitution pattern characterized by a high frequency of homoplasy (Ingman & Gyllenstein, 2001). The length of the selected sequences varied from 15 437 to 15 451 bp.

The accession numbers of the JCV sequences under examination are as follows: AB038249–AB038255, AB048545–AB048582, AB074575–AB074591, AB077855–AB077879, AB081005–AB081030, AB081600–AB081618, AB081654, AB092578–AB092587, AB103387, AB103402–AB103423, AB104487, AF004349–AF004350, AF015526–AF015537, AF030085, AF281599–AF281626, AF295731–AF295739, AF300945–AF300967, AF363830–AF363834, AF396422–AF396435, AY121907–AY121915, J02226, U61771 and U73500–U73502. Five sequences (AB038254, AB038255, AF015537, AF030085 and AF004350) derived from patients, died of progressive multifocal leukoencephalopathy. The inclusion of these sequences does not affect the present analysis, since no amino acid substitutions that could be correlated with disease have been detected so far (Kato *et al.*, 2000). The nomenclature system of JCV was in accordance with Agostini *et al.* (2001b). The correlation between Agostini's classification and that developed by Sugimoto *et al.* (2002b) is reported by Cui *et al.* (2004). The accession numbers of the mtDNA sequences are AY195745–AY195792, AF346963–AF347015, AY289051–AY289102, D38112, J01415 and X93334.

Data analysis. The 275 genome sequences of JCV were aligned using the CLUSTAL W program (Thompson *et al.*, 1994). A multiple alignment of 4867 nt positions was obtained. Two sequences were excluded, because of the presence of a large deletion in the VP2 gene (63 bp in the sequence AB103402) or in the small t antigen gene (38 bp in the sequence AB103407). CLUSTAL W was also used to align the 156 mtDNA sequences, yielding a multiple alignment of 15 465 sites. Each alignment was examined to detect variable sites. A total of 1030 variable sites were found in JCV, 944 of them without gaps. The mtDNA sequence showed a total of 1035 variable sites, 994 of them without gaps.

Correspondence analysis was carried out on the JCV sequences formed by nucleotides at the variable sites lacking gaps. Each sequence was converted into a vector consisting of 1s and 0s, depending on whether a given nucleotide is present at a given position or not. For example, the most variable position of JCV (122 A, 10 T, 133 G and 8 C) was represented as a string of four binary characters: for the 122 sequences with A the string is 1000, for the 10 sequences with T the string is 0100, for the 133 sequences with G the string is 0010 and for the 8 sequences with C the string is 0001. Those positions containing only two or three types of nucleotide were converted into a string of two and three

binary characters, respectively. According to these rules, each viral sequence was represented as a vector of 2032 binary characters. This yielded the matrix A (273 rows and 2032 columns) with elements a_{ij} . The matrix P with elements p_{ij} was computed as follows:

$$p_{ij} = a_{ij}/S$$

where S is the sum of all the elements of the matrix A .

The matrix D with elements d_{ij} was computed as follows:

$$d_{ij} = [p_{ij}/(p_i p_j)^{1/2}] - (p_i p_j)^{1/2}$$

where p_i is the sum of the 2032 elements of the row i in the matrix P , and p_j is the sum of the 273 elements of the column j in the matrix P . The matrix D is the matrix of the differences between the binary data observed and the binary data expected in the null hypothesis of a lack of relation between sequences.

The product between the transpose of the matrix D (D^T with 2032 rows and 273 columns) and the matrix D itself (273 rows and 2032 columns) yielded the matrix E (2032 rows and 2032 columns). Eigenvectors and eigenvalues of the matrix E were calculated with the EIGEN subroutine from the statistical language R (Ihaca & Gentleman, 1996; www.r-project.org). Only the first 10 eigenvectors were taken into consideration, yielding the matrix F (2032 rows and 10 columns). The product between the matrix D and the matrix F gave the matrix G (273 rows and 10 columns).

The information carried by the matrix G is crucial, since it provides the position coordinates of the 273 JCV sequences on the first 10 axes of ordination. The percentage variation fraction associated with each axis was calculated as the ratio of the corresponding eigenvalue to the sum of all eigenvalues. The evolutionary relationships between sequences were visualized by the construction of bidimensional plots.

The information carried by the first 10 eigenvectors (matrix F) is also very valuable, since it reveals the nucleotide positions that maximally contribute to the JCV clustering at each axis of ordination. To obtain this information, the absolute values of the 2032 elements of each eigenvector were sorted in increasing order. The highest values correspond to the 'important' positions, whose phylogenetic relevance was evaluated further with the χ^2 test. The eigenvectors were finally used to draw bidimensional plots, in which the variable sites were represented as a heterogeneous cloud of points.

Correspondence analysis was also applied to the 156 human mtDNA

sequences. Each sequence, consisting of 994 variable sites, was converted into a vector of 2000 binary characters, and then subjected to the same processes of calculation described above.

Finally, by using the method of Nei & Gojobori (1986), the rates of synonymous and non-synonymous substitutions in the protein-coding regions of both JCV and mitochondrial genomes were estimated. The degree of similarity between individual amino acid residues was evaluated with the BLOSUM 62 substitution matrix (Henikoff & Henikoff, 1992).

RESULTS

Correspondence analysis of the JCV sequences

The map obtained from the two first axes of ordination shows a subdivision of the 273 JCV sequences into three groups (Fig. 1a). In particular, the projection of points on axis 1, which accounted for 7.8% of the total variation, placed type 6 between the group including types 1 and 4 and that including types 2, 3, 7 and 8. Since the distinctiveness of type 6 is also stressed by the projection of points on axis 2, the main finding yielded by axis 1 is the clear separation of the 52 strains belonging to types 1 and 4.

The identification of the nucleotide substitutions responsible for the divergence of types 1 and 4 requires the examination of a further map, which was obtained from the first two eigenvectors (Fig. 1b). On axis 1 of this map, the set of points with a location similar to that assigned to types 1 and 4 in the previous map includes the corresponding 'important' positions. For example, the point with the highest value of the first eigenvector (0.097) corresponds to a genomic position where all strains of types 1 and 4 show adenine and all the others thymine. Points with a value approaching zero correspond, however, to the large amount of sites having a poor phylogenetic relevance.

By choosing as the threshold a χ^2 value of 1.07 ($P < 0.30$ for 1 d.f.), a total of 61 nt positions with a pattern of change determining the distinctiveness of the virus lineage formed

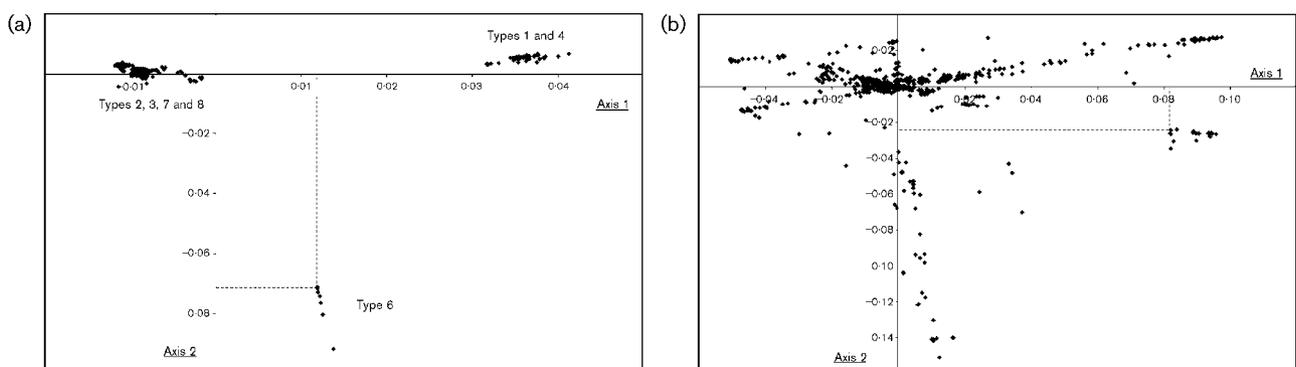


Fig. 1. (a) Plot of axis 1 versus axis 2 of a correspondence analysis of the genome sequence from 273 JCV strains. (b) Plot of axis 1 versus axis 2 of a correspondence analysis of the 944 ungapped variant sites (2032 binary characters) taken from the aligned JCV sequences. Dashed lines exemplify the projection of points on the two axes of the map.

Table 1. List of nucleotide positions causing the divergence of types 1 and 4

First eigenvector* (absolute value)	Nucleotide position†	Genomic region	Nucleotide‡	χ^2 value§
0.0973	5017	Non-coding region	A	0.00
0.0973	4951	t antigen/T antigen	C	0.00
0.0973	3968	T antigen	A	0.00
0.0973	3442	T antigen	G	0.00
0.0973	2978	T antigen	C	0.00
0.0973	2671	T antigen	T	0.00
0.0973	2464	VP1	G	0.00
0.0973	2224	VP1	A	0.00
0.0973	1993	VP1	C	0.00
0.0973	1654	VP1	A	0.00
0.0973	1151	VP2/VP3	C	0.00
0.0973	1050	VP2/VP3	C	0.00
0.0973	1048	VP2/VP3	G	0.00
0.0973	738	VP2	A	0.00
0.0961	5026	Non-coding region	C	0.02
0.0960	4121	T antigen	A	0.00
0.0960	1221	VP2/VP3	A	0.00
0.0959	3446	T antigen	G	0.00
0.0958	3566	T antigen	G	0.00
0.0956	4714	T antigen	G	0.22
0.0956	3836	T antigen	G	0.22
0.0956	3479	T antigen	G	0.22
0.0956	3443	T antigen	A	0.22
0.0956	2386	VP1	T	0.22
0.0956	1251	VP2/VP3	T	0.22
0.0956	1056	VP2/VP3	C	0.22
0.0956	990	VP2/VP3	G	0.22
0.0954	3134	T antigen	C	0.08
0.0946	3710	T antigen	T	0.02
0.0945	2266	VP1	G	0.02
0.0945	993	VP2/VP3	C	0.29
0.0943	3611	T antigen	G	0.29
0.0943	3830	T antigen	T	0.29
0.0937	2320	VP1	G	0.30
0.0936	2455	VP1	A	0.04
0.0935	3809	T antigen	A	0.31
0.0934	2494	VP1	C	0.31
0.0933	2177	VP1	T	0.04
0.0933	3961	T antigen	G	0.37
0.0932	3020	T antigen	T	0.37
0.0920	2726	T antigen	T	0.19
0.0919	434	Agnoprotein	G	0.11
0.0919	1100	VP2/VP3	A	0.19
0.0917	2984	T antigen	C	0.70
0.0910	3053	T antigen	C	0.11
0.0909	3605	T antigen	G	0.11
0.0907	2245	VP1	C	0.94
0.0906	3768	T antigen	T	0.16
0.0904	3178	T antigen	A	0.76
0.0897	2011	VP1	C	0.18
0.0896	4891	T antigen/T antigen	G	0.77
0.0896	4434	Non-coding region	A	0.89
0.0896	2293	VP1	C	0.89

Table 1. cont.

First eigenvector* (absolute value)	Nucleotide position†	Genomic region	Nucleotide‡	χ^2 value§
0.0892	4076	T antigen	G	0.29
0.0892	1771	VP1	C	0.13
0.0889	4975	T antigen/T antigen	A	0.89
0.0886	5039	Non-coding region	C	1.02
0.0882	1359	VP2/VP3	A	0.29
0.0861	3455	T antigen	A	0.30
0.0860	2596	Non-coding region	T	0.46
0.0854	1071	VP2/VP3	G	0.45

*The eigenvector values are sorted in increasing order.

†The position number agrees with the numbering system by Frisque *et al.* (1984).

‡Nucleotide characteristic of types 1 and 4.

§The χ^2 threshold was set at 1.07 ($P < 0.30$ for 1 d.f.).

by types 1 and 4 were found (Table 1). In particular, five sites were localized in the intergenic regions or in the intron of the large T antigen gene. The great majority of sites (48) were characterized by synonymous substitutions. Only eight positions showed a non-synonymous change, most of them determining a conservative substitution (Asn vs His, Ala vs Val, Ala vs Ser, Asp vs Glu, Glu vs Asp, Arg vs Lys and Asn vs Ser). The only point mutation yielding a non-conservative amino acid exchange (the hydrophilic Gln residue vs the hydrophobic Leu residue) was located in the second exon of the large T antigen gene. Interestingly, such a substitution occurs in close proximity to the T antigen zinc finger motif, which is essential for the replication of viral DNA (Swenson *et al.*, 1996).

In the map in Fig. 1(a), the distinctiveness of the ancestral genotype (type 6) was highlighted by the projection of

points on the second axis of ordination, which accounted for 3.8% of the total variation. In the map in Fig. 1(b), the corresponding 'important' positions were recognized as the set of points showing extreme negative values. The χ^2 test demonstrated that the significant divergence of type 6 is due to nucleotide substitutions occurring at 33 sites (Table 2). Again, most of the changes were localized at the third codon position or in the non-coding region. Only three substitutions were found at the second codon position, all determining a conservative amino acid exchange (Asn vs Ser, Lys vs Arg and Phe vs Tyr).

A detailed examination of JCV clustering at the remaining eight axes of ordination provided further phylogenetic information (Table 2). Although the variation fraction associated with each axis progressively decreased (from 3.8 to 1.6%), all but one axis yielded evidence for the divergence

Table 2. Clustering of JCV at the first 10 axes of ordination of correspondence analysis

Axis of ordination*	Genotype diverging†	No. important positions‡	Frequency of nucleotide substitutions			
			Non-coding region	Synonymous	Non-synonymous conservative	Non-synonymous non-conservative
1 (7.81)	1 and 4 (52)	61	5	48	7	1
2 (3.81)	6 (7)	33	4	26	3	0
3 (2.51)	2A (70)	16	0	12	4	0
4 (2.25)	3 (7)	16	3	12	1	0
5 (1.98)	2E (8)	11	2	7	1	1
6 (1.95)	8B (8)	11	7	4	0	0
7 (1.85)	8A (3)	17	4	9	4	0
8 (1.72)	—	—	—	—	—	—
9 (1.70)	4 (12)	12	1	9	2	0
10 (1.61)	7B (40)	11	2	9	0	0

*The numbers in parentheses indicate the percentage fraction of variation associated with each axis.

†The numbers in parentheses indicate the frequency of the strains belonging to a given type or subtype of JCV.

‡The important positions are those responsible for the divergence of the corresponding type or subtype of JCV.

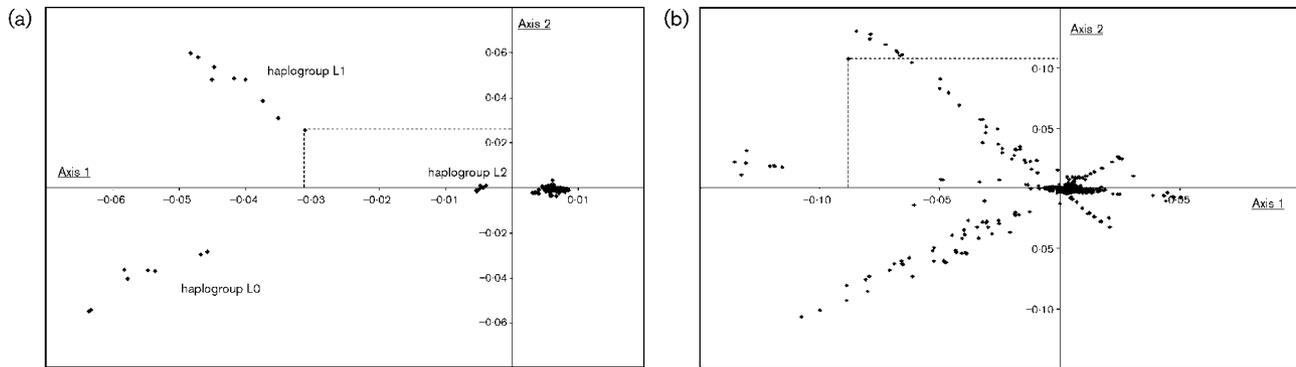


Fig. 2. (a) Plot of axis 1 versus axis 2 of a correspondence analysis of 156 human mtDNA sequences. (b) Plot of axis 1 versus axis 2 of a correspondence analysis of the 994 ungapped variant sites (2000 binary characters) taken from the aligned mtDNA sequences. Dashed lines exemplify the projection of points on the two axes of the map.

of a particular type or subtype of the virus. A complementary analysis of the kind of substitutions occurring at the 'important' positions peculiar to each axis, highlighted a pattern of genetic variation poorly affected by selective pressures (see last four columns in Table 2).

Correspondence analysis of the human mtDNA sequences

The map in Fig. 2(a) was obtained from the first two axes of ordination, which accounted for 4.6 and 3.3%, respectively, of the total variation. It yielded evidence for a grouping of the 156 mtDNA sequences into four clusters. In particular, the projection of points on axis 1 assigned marked negative values (from -0.031 to -0.064) to a fairly heterogeneous set of 17 sequences, thus stressing their separation from the rest. Such sequences, which were isolated from indigenous populations living in sub-Saharan Africa, belong to haplogroups L0 and L1.

A complementary examination of the map drawn with the first two eigenvectors (Fig. 2b) revealed the presence of a number of positions that are crucial for the divergence of haplogroups L0 and L1. Such positions are recognizable as the set of points with an extreme negative value of their coordinate positions on axis 1. By using the χ^2 test, a total of 17 'important' positions were accurately identified (Table 3). Six of them were localized in genes encoding rRNA or tRNA, the others in the protein-coding genes. Nine of the 11 substitutions occurring in the protein-coding region were of the synonymous type. The two non-synonymous changes led to conservative amino acid replacements (Ala vs Thr and Val vs Ile).

The projection of points on the second axis of the map (Fig. 2b) revealed that most of the mtDNA sequences fall into a group with coordinate positions close to zero. The remaining sequences were separated into two groups. The eight sequences of haplogroup L0, with coordinate positions ranging from -0.028 to -0.056 , were placed in the

lower half of axis 2. The nine sequences of haplogroup L1, with coordinate positions comprised between 0.026 and 0.060 , were placed in the upper half of the same axis. The nucleotide variations causing the distinction between haplogroups L0 and L1 were detected by a dual sorting of the second eigenvector: in decreasing order to find substitutions specific for L0, and in increasing order for L1. The variation pattern at such diagnostic sites, 16 for L0 and 19 for L1, is reported in Table 4.

Unlike the analysis of JCV, the examination of the remaining eight axes of ordination provided poor phylogenetic information (Table 4). Only the projection of points on the third axis, which accounted for 2.25% of the total variation, yielded evidence for a clustering of seven mtDNA sequences. Such sequences, which were isolated from autochthonous populations inhabiting West Africa, belong to haplogroup L2. Their distinctiveness was ascribed to a very low number of nucleotide substitutions (Table 4). The other axes of ordination did not provide any clustering of the haplogroups that are peculiar to human populations living outside Africa. At the most, a clustering of a small number of sequences, which belong to the African haplogroups already discriminated by the first three axes of ordination, was found.

Overall, the correspondence analysis of the human mtDNA highlighted the pattern of variation responsible for the divergence of three of four African haplogroups (L0, L1, L2 and L3). Like the analysis of JCV, this pattern appeared to be poorly affected by natural selection, being characterized mainly by nucleotide substitutions of the synonymous type.

Comparison between the patterns of synonymous and non-synonymous substitutions in the protein-coding regions of JCV and human mtDNA

The exclusion of the non-coding region from each viral sequence led to a sequence formed by the six protein-coding

Table 3. List of nucleotide positions causing the divergence of the haplogroups L0 and L1

First eigenvector* (absolute value)	Nucleotide position†	Genomic region	Nucleotide‡	χ^2 value§
0.136	13506	NADH dehydrogenase 5	T	0.00
0.136	8655	ATPase 6	T	0.00
0.136	8468	ATPase 8	T	0.00
0.136	2758	16S rRNA	A	0.00
0.136	825	12S rRNA	A	0.00
0.133	2885	16S rRNA	C	0.06
0.131	10688	NADH dehydrogenase 4L	A	0.01
0.131	10810	NADH dehydrogenase 4	C	0.01
0.131	7146	Cytochrome oxidase 1	G	0.06
0.121	4104	NADH dehydrogenase 1	G	0.26
0.119	13105	NADH dehydrogenase 5	G	0.12
0.119	13650	NADH dehydrogenase 5	T	0.35
0.119	7256	Cytochrome oxidase 1	T	0.35
0.119	3594	NADH dehydrogenase 1	T	0.35
0.119	1018	12S rRNA	A	0.35
0.119	769	12S rRNA	A	0.35
0.116	7521	tRNA Asp	A	0.46

*The values of the eigenvector are sorted in increasing order.

†The number of the positions agree with the numbering system of the Cambridge reference sequence for human mitochondrial DNA (Anderson *et al.*, 1981).

‡Nucleotide characteristic of haplogroups L0 and L1.

§The cut-off value of the χ^2 was set at 1.07 ($P < 0.30$ for 1 d.f.).

genes, having a length of 4572 bp. The exclusion from each mtDNA sequence of the genes encoding rRNAs or tRNAs, as well as of the intergenic regions, yielded a sequence of 11 334 bp long and included the 13 protein-coding genes.

Using the method by Nei & Gojobori (1986), the 273 sequences of JCV were compared to each other. In each comparison, the number of synonymous (S_d) and non-synonymous (N_d) differences was evaluated. At the end of the calculation, a mean value of S_d equal to 47.4, with a standard deviation (SD) of 27.6, was found. The mean value of N_d was 12.2, with a SD of 6.1. The same process of calculation was carried out on the 156 mtDNA sequences.

A mean value of S_d equal to 22.4 (SD = 14.6) and a mean value of N_d equal to 8.2 (SD = 3.7) were obtained.

Although the virus contains a protein-coding region over two times shorter than mtDNA, it exhibited a mean amount of synonymous changes over twice as great. This remarkable difference was investigated further by comparing the mean number of synonymous substitutions per site observed in JCV ($K_s = 0.0516$, SD = 0.0034) with that found in mtDNA ($K_s = 0.0082$, SD = 0.0010).

The trend of the mean number of synonymous substitutions per site (K_s), averaged over a sliding-window region of

Table 4. Clustering of mtDNA at the first three axes of ordination of correspondence analysis

Axis of ordination*	Haplogroup diverging†	No. important positions‡	Frequency of nucleotide substitutions			
			rRNA/tRNA genes	Synonymous	Non-synonymous conservative	Non-synonymous non-conservative
1 (4.63)	L0/L1 (17)	17	6	9	2	0
2 (3.29)	L0 (8)	16	2	12	2	0
2 (3.29)	L1 (9)	19	0	12	6	1
3 (2.25)	L2 (7)	6	2	4	0	0

*The numbers in parentheses indicate the percentage fraction of variation associated with each axis.

†The numbers in parentheses indicate the frequency of a given mtDNA haplogroup.

‡The important positions are those responsible for the divergence of the corresponding haplogroup.

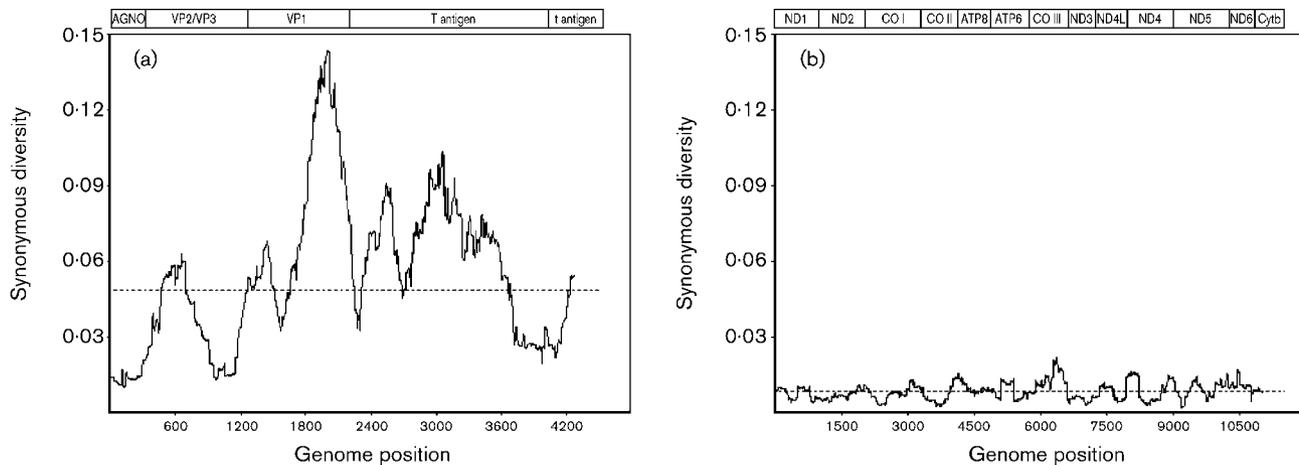


Fig. 3. Trend of the mean number of synonymous substitutions per site (K_s), averaged over a sliding-window region of 100 codons, along the entire coding sequence of 273 JCV strains (a) and 156 human mitochondria (b). The dashed line indicates the mean value of K_s over the entire coding sequence (1524 codons for JCV and 3778 codons for mtDNA). At the top of the figure the organization of the coding region of JCV and mtDNA is presented.

100 codons, was evaluated along the entire coding sequences of both JCV and mitochondrial genomes (Fig. 3). The examination of the two profiles evidenced a considerably higher rate of synonymous substitution in JCV. In particular, it was found that most of the coding region of JCV exhibits a K_s value invariably more elevated with respect to the highest K_s value observed in the mtDNA sequence.

DISCUSSION

In this study, the hypothesis that the pattern of variation of JCV meets with the expectations of the neutral theory of evolution, which is described entirely by random genetic drift, mutation and migration (Kimura, 1983), was verified. Although several tests of neutrality have been developed, they suffer from the limitation that the information content of sequence data are condensed into schematic statistics and, as a consequence, information is partially lost (reviewed by Kreitman, 2000). The results of applying correspondence analysis to JCV suggest that this technique is a valuable tool to extract significant evolutionary trends from sequence data.

The clustering of the 273 genome sequences of JCV along the first axis of ordination of the correspondence analysis (Fig. 1a) resembles that provided by a principal coordinate analysis of about 100 genome sequences (Pavesi, 2003) or over 1000 sequences of the genomic region with the highest variation rate (Pavesi, 2004). Similar to what has been found using standard phylogenetic methods (Cui *et al.*, 2004; Yogo *et al.*, 2004 and references therein), such a clustering confirms an important feature of the evolutionary history of JCV, namely an early emergence of two different lineages from the common root given by the ancestral African genotype (type 6). One lineage includes the 52 strains of

types 1 and 4 (see right side of axis 1), the other includes the 214 strains of types 2, 3, 7 and 8 (see left side of axis 1).

The first point to stress is the geographical distribution of the 52 strains of types 1 and 4. Half of them were found in Europe. One strain was found in Morocco. Six strains were isolated in North America from individuals of European origin. Eight strains were isolated from autochthonous populations inhabiting the northeastern edge of Siberia, such as the Nanais, Koryaks, Chukchis, Luskys and Yukaghirs. Two strains were found in the Canadian Inuits, an indigenous Arctic populace speaking an Eskimo-Aleut language (Ruhlen, 1991). The remaining nine strains were found in Japan: four of them belong to the Ainu, a pre-agricultural native population of great anthropological interest (Bannai *et al.*, 2000).

Since it has been proved that types 1 and 4 arose from type 6 as an independent lineage (Pavesi, 2003), its geographical distribution could reflect a prehistoric migration of humans from Africa into Europe and from there to northern Asia. The hypothesis that types 1 and 4 were acquired by modern humans when they migrated into Europe and came in contact with archaic populations (*Homo neanderthalensis*) seems to be rather unlikely. The transmission of JCV, in fact, requires close and prolonged contact between individuals living in the same ethnic group (Kunitake *et al.*, 1995), as proved by the lack of transmission between populations inhabiting the same geographical area yet only occasionally intermingling with each other (Kato *et al.*, 1997). The geographical distribution of the other lineage of JCV (East Africa, Eurasia, Asia, Americas, Oceania and the Pacific Islands) is compatible with the pattern of migration yielded by human genes (Cavalli-Sforza & Feldman, 2003).

The finding that the divergence of the Caucasian lineage of JCV (types 1 and 4) was accompanied by synonymous, rather than non-synonymous substitutions (Table 1) seems to exclude the hypothesis of a divergence due to selective pressures favouring adaptation to cold climates. The hypothesis of an additional early expansion of humans from Africa to the northern areas of the world (Fig. 4), previously suggested by synthetic maps (Pavesi, 2004) or phylogenetic trees (Yanagihara *et al.*, 2002; Sugimoto *et al.*, 2002a, b; Yogo *et al.*, 2003), seems to be substantiated by the virtual lack of marks of natural selection in the divergence of types 1 and 4. The only adaptation change is probably a non-conservative amino acid replacement (Gln vs Leu) found in the T antigen gene. Besides the Caucasian lineage, this substitution also occurs in five viral strains of subtype 2B, belonging to the alternative lineage yet showing a geographical distribution similar to that of types 1 and 4. Thus, the Gln→Leu change seems to be affected by selection, although its functional significance remains to be determined.

The peopling of the various continents was further elucidated by examining the JCV clustering at the subsequent axes of ordination (Table 2). The peopling of the Americas by populations of Asian ancestry was found to be associated, at the third axis, with the divergence of the subtype 2A. The clustering of JCV at the fifth, sixth and seventh axes can be correlated with three broad migrations playing a role in the peopling of Oceania and the Pacific Islands: an ancient one characterized by subtype 8A, followed by subtype 8B and much later by type 2E (Yanagihara *et al.*, 2002). The distinctiveness of the African types was apparent at the second and fourth axes of ordination. At the tenth axis, the peopling of northern China and Mongolia was found to be associated with the divergence of the subtype 7B. The ninth axis discriminated exclusively the strains of type 4, suggesting that the major migration of humans carrying the Caucasian lineage could be subdivided into more subtle branches.

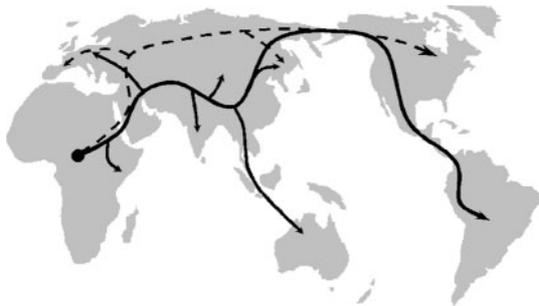


Fig. 4. The two-migration model of the expansion from Africa of *Homo sapiens* suggested by JCV. The migration traced with a solid line is compatible with that yielded by human genes. The migration traced with a dashed line indicates the additional route of expansion suggested by JCV but undetectable with human genes.

Correspondence analysis of the human mtDNA yielded a first axis of ordination which separated the sequences according to the following order: the haplogroups L0, L1, L2 and finally the remaining haplogroups (Fig. 2a). This clustering is similar to the topology of the consensus phylogenetic trees including more or less the same sequences and constructed with the neighbour-joining method (Ingman *et al.*, 2000; Mishmar *et al.*, 2003). Both trees, in fact, show a basal branching pattern where the deepest branch is represented by the haplogroup L0. The following two deepest branches include the haplogroup L1 and L2, respectively.

By comparing the pattern of change peculiar to mtDNA with that of JCV, some relevant differences can be appreciated. The first difference lies in the fact that the first axis of ordination of JCV situates the ancestral type 6 in the middle, between the two lineages arising from it (Fig. 1a). The first axis of mtDNA, on the other hand, places the ancestral haplogroup L0 at the extreme left, since it gave rise to one sole lineage (Fig. 2a).

The second difference stems from the finding that the various axes of ordination are much more informative in JCV than in mtDNA. Indeed, the peopling of the world by humans carrying different types or subtypes of JCV can be correlated with most of the first 10 axes of ordination (Table 2). In the case of mtDNA, the phylogenetic information is limited, however, to the more ancient African haplogroups L0, L1 and L2 (Table 4). The lack of discrimination of the fourth African haplogroup (L3) is consistent with the fact that such a haplogroup usually features, in the consensus trees, with the non-African haplogroups (Ingman *et al.*, 2000; Mishmar *et al.*, 2003).

The third difference depends on the rate of substitution found in the protein-coding region of JCV and mtDNA. Although JCV is known to be a very slowly evolving virus, it shows a mean nucleotide diversity (59.6) double that of mtDNA (30.6). By removing the bias due to the different length of the protein-coding region (4572 bp in JCV and 11 334 bp in mtDNA), it was found that the mean number of synonymous substitutions per site of JCV ($K_s = 0.0516$) is over six times higher than that of mtDNA ($K_s = 0.0082$). The difference between the substitution rates can explain why the number of the 'important' nucleotide positions, which is those positions where systematic changes have occurred in the past, was much higher in JCV with respect to the human mtDNA (see first three axes of ordination in Tables 2 and 4). The greater amount of silent changes in JCV can be appreciated by comparing the trends of the mean synonymous diversity shown in Fig. 3.

The findings reported here support the hypothesis that the human mtDNA, unlike JCV, shows a nucleotide diversity too low to trace the pattern of migrations subsequent to the split between African and non-African populations. Since it is known that mtDNA evolves at a speed 5–10 times higher than the nuclear DNA (Vawter & Brown, 1986), it is

likely that a reconstruction of human history based on the nucleotide sequence of DNA fragments from autosomal or sex-linked loci is an even more difficult task.

It is important to note, however, that improved methods for a large-scale characterization of human genome diversity have provided in the last years valuable information concerning the small nuclear polymorphism or the microsatellite loci. For example, Zhivotovsky *et al.* (2003) studied 377 autosomal microsatellite polymorphisms in 52 world populations and constructed a phylogenetic tree whose two oldest branches include, respectively, hunter-gatherer and farmer populations from sub-Saharan Africa.

Nevertheless, a ubiquitous, usually harmless, symbiote co-evolving with the human host and showing a sufficiently sensitive variation rate could be an alternative approach. A few viruses have been used for inferences about human evolution, such as the hepatitis G virus (Pavesi, 2001), the papillomavirus (Ho *et al.*, 1993; Ong *et al.*, 1993) and the T-cell lymphotropic virus (Miura *et al.*, 1994; Salemi *et al.*, 1999). In the case of the latter two, the main drawback is a transmission mechanism prevalently horizontal. Although the hepatitis G virus does not cause liver disease and is largely transmitted from mother to infant, the finding that it can recombine raises doubts on its ability to trace human history (Worobey & Holmes, 2001). Finally, and most importantly, what we expect from a virus are novel clues on human history, rather than a pure replication of the pattern yielded by human genes.

The JC polyomavirus, exhibiting the unusual feature of a twofold exit from Africa (Pavesi, 2003), could shed new light on the number of migrations leading to the peopling of the various continents. The virtual lack of pathogen power (the virus can cause disease only in 5% of severely immunocompromised patients), the absence of genetic recombination (the unique strain whose sequence suggested recombination has now been discredited due to the inability to repeat the result in the same patient), the strong ethnicity due to a transmission mechanism within the family or in the same community, and the easy detection in individuals due to the high frequency of urinary excretion support the effectiveness of JCV in tracing the history of human populations. The findings reported here, supporting the virtual absence of marks of natural selection in JCV evolution, would encourage further sampling of virus isolates from historical populations, thus providing a more exhaustive picture of our past.

ACKNOWLEDGEMENTS

The author is grateful to Franco Conterio for his support and encouragement. A special thank you goes to Adriano Cosenza for preparing the figures and to Maria Cristina Cignatta for language revision. The critical comments from the anonymous referee and the helpful suggestions from Luca Cavalli-Sforza and Alessio Peracchi are gratefully acknowledged. This work was financed by the MURST (Ministero dell'Università e della Ricerca Scientifica e Tecnologica).

REFERENCES

- Agostini, H. T., Yanagihara, R., Davis, V., Ryschkewitsch, C. F. & Stoner, G. L. (1997a). Asian genotypes of JC virus in Native Americans and in a Pacific Island population: markers of viral evolution and human migration. *Proc Natl Acad Sci U S A* **94**, 14542–14546.
- Agostini, H. T., Ryschkewitsch, C. F., Brubaker, G. R., Shao, J. & Stoner, G. L. (1997b). Five complete genomes of JC virus type 3 from Africans and African Americans. *Arch Virol* **142**, 637–655.
- Agostini, H. T., Ryschkewitsch, C. F., Singer, E. J. & Stoner, G. L. (1997c). JC virus regulatory region rearrangements and genotypes in progressive multifocal leukoencephalopathy: two independent aspects of virus variation. *J Gen Virol* **78**, 659–664.
- Agostini, H. T., Deckhut, A., Jobes, D. V. & 7 other authors (2001a). Genotypes of JC virus in East, Central and Southwest Europe. *J Gen Virol* **82**, 1221–1231.
- Agostini, H. T., Jobes, D. V. & Stoner, G. L. (2001b). Molecular evolution and epidemiology of JC virus. In *Human Polyomavirus: Molecular and Clinical Perspectives*, pp. 491–526. Edited by K. Khalili & G. L. Stoner. New York, NY: Wiley-Liss Inc.
- Anderson, S., Bankier, A. T., Barrell, B. G. & 11 other authors (1981). Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.
- Bannai, M., Ohashi, J., Harihara, S., Takahashi, Y., Juji, T., Omoto, K. & Tokunaga, K. (2000). Analysis of HLA genes and haplotypes in Ainu (from Hokkaido, northern Japan) supports the premise that they descent from Upper Paleolithic populations of East Asia. *Tissue Antigens* **55**, 128–139.
- Cavalli-Sforza, L. L. & Feldman, M. W. (2003). The application of molecular genetic approaches to the study of human evolution. *Nat Genet* **33**, 266–275.
- Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994). Genetic history of world populations. In *The History and Geography of Human Genes*, pp. 60–157. NJ: Princeton University Press.
- Chang, H., Wang, M., Tsai, R. T., Lin, H. S., Huan, J. S., Wang, W. C. & Chang, D. (2002). High incidence of JC virusuria in JC-seropositive older individuals. *J Neurovirol* **8**, 447–451.
- Chima, S. C., Ryschkewitsch, C. F. & Stoner, G. L. (1998). Molecular epidemiology of human polyomavirus JC in the Biaka Pygmies and Bantu of Central Africa. *Mem Inst Oswaldo Cruz* **93**, 615–623.
- Cui, X., Wang, J. C., Deckhut, A., Joseph, B. C., Eberwein, P., Cubitt, C. L., Ryschkewitsch, C. F., Agostini, H. T. & Stoner, G. L. (2004). Chinese strains (Type 7) of JC virus are Afro-Asiatic in origin but are phylogenetically distinct from the Mongolian and Indian strains (Type 2D) and the Korean and Japanese strains (Type 2A). *J Mol Evol* **58**, 568–583.
- Fernandez-Cobo, M., Agostini, H. T., Britez, G., Ryschkewitsch, C. F. & Stoner, G. L. (2002). Strains of JC virus in Amerind-speakers of North America (Salish) and South America (Guarani), Na-Dene-speakers of New Mexico (Navajo), and modern Japanese suggest links through an ancestral Asian population. *Am J Phys Anthropol* **118**, 154–168.
- Frisque, R. J., Bream, G. L. & Cannella, M. T. (1984). Human polyomavirus JC virus genome. *J Virol* **51**, 458–469.
- Guo, J., Kitamura, T., Ebihara, H. & 9 other authors (1996). Geographical distribution of the human polyomavirus JC virus types A and B and isolation of a new type from Ghana. *J Gen Virol* **77**, 919–927.

- Henikoff, S. & Henikoff, J. G. (1992).** Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* **89**, 10915–10919.
- Ho, L., Chan, S. Y., Burk, R. D. & 19 other authors (1993).** The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. *J Virol* **67**, 6413–6423.
- Ihaca, R. & Gentleman, R. (1996).** R: a language for data analysis and graphics. *J Comput Graph Stat* **5**, 299–314.
- Ingman, M. & Gyllensten, U. (2001).** Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. *J Hered* **92**, 454–461.
- Ingman, M. & Gyllensten, U. (2003).** Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res* **13**, 1600–1606.
- Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. (2000).** Mitochondrial genome variation and the origins of modern humans. *Nature* **408**, 708–713.
- Jobes, D. V., Friedlaender, J. S., Mgone, C. S. & 7 other authors (2001).** New JC virus (JCV) genotypes from Papua New Guinea and Micronesia (Type 8 and Type 2E) and evolutionary analysis of 32 complete JCV genomes. *Arch Virol* **146**, 2097–2113.
- Kato, A., Kitamura, T., Sugimoto, C., Ogawa, Y., Nakazato, K., Nagashima, K., Hall, W. W., Kawabe, K. & Yogo, Y. (1997).** Lack of evidence for the transmission of JC polyomavirus between human populations. *Arch Virol* **142**, 875–882.
- Kato, A., Sugimoto, C., Zheng, H. Y., Kitamura, T. & Yogo, Y. (2000).** Lack of disease-specific amino acid changes in the viral proteins of JC virus isolates from the brain with progressive multifocal leukoencephalopathy. *Arch Virol* **145**, 2173–2182.
- Kimura, M. (1983).** *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- Kitamura, T., Aso, Y., Kuniyoshi, N., Hara, K. & Yogo, Y. (1990).** High incidence of urinary JC virus excretion in nonimmunosuppressed older patients. *J Infect Dis* **161**, 1128–1133.
- Kreitman, M. (2000).** Methods to detect selection in populations with applications to the human. *Annu Rev Genomics Hum Genet* **1**, 539–559.
- Kunitake, T., Kitamura, T., Guo, J., Taguchi, F., Kawabe, K. & Yogo, Y. (1995).** Parent-to-child transmission is relatively common in the spread of the human polyomavirus JC. *J Clin Microbiol* **33**, 1448–1451.
- Lebart, L., Morineau, A. & Warwick, K. A. (1984).** *Multivariate Descriptive Statistical Analysis. Correspondence Analysis and Related Techniques for Large Matrices*. New York: Wiley & Sons.
- Miranda, J. J., Takasaka, T., Zheng, H.-Y., Kitamura, T. & Yogo, Y. (2004).** JC virus genotype profile in the Mamanwa, a Philippine Negrito tribe, and implications for its population history. *Anthropol Sci* **112**, 173–178.
- Mishmar, D., Ruiz-Pesini, E., Golik, P. & 10 other authors (2003).** Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A* **100**, 171–176.
- Miura, T., Fukunaga, T., Igarashi, T. & 17 other authors (1994).** Phylogenetic subtypes of human T-lymphotropic virus type I and their relations to the anthropological background. *Proc Natl Acad Sci U S A* **91**, 1124–1127.
- Nei, M. & Gojoberi, T. (1986).** Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. *Mol Biol Evol* **3**, 418–426.
- Ong, C. K., Chan, S. Y., Campo, M. S. & 8 other authors (1993).** Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. *J Virol* **67**, 6424–6431.
- Padgett, B. L. & Walker, D. L. (1973).** Prevalence of antibodies in human sera against JC virus, an isolate from a case of progressive leukoencephalopathy. *J Infect Dis* **127**, 467–470.
- Pavesi, A. (2001).** Origin and evolution of GBV-C/hepatitis G virus and relationships with ancient human migrations. *J Mol Evol* **53**, 104–113.
- Pavesi, A. (2003).** African origin of polyomavirus JC and implications for prehistoric human migrations. *J Mol Evol* **56**, 564–572.
- Pavesi, A. (2004).** Detecting traces of prehistoric human migrations by geographic synthetic maps of polyomavirus JC. *J Mol Evol* **58**, 304–313.
- Ruhlen, M. (1991).** *A Guide to the World's Languages*. CA: Stanford University Press.
- Salemi, M., Vandamme, A. M., Desmyter, J., Casoli, C. & Bertazzoni, U. (1999).** The origin and evolution of human T-cell lymphotropic virus type II (HTLV-II) and the relationship with its replication strategy. *Gene* **234**, 11–21.
- Saruwatari, L., Sugimoto, C., Kitamura, T. & 12 other authors (2002).** Asian domains of four major genotypes of JC virus, Af2, B1-b, CY and SC. *Arch Virol* **147**, 1–10.
- Sugimoto, C., Kitamura, T., Guo, J. & 16 other authors (1997).** Typing of urinary JC virus DNA offers a novel means for tracing human migrations. *Proc Natl Acad Sci U S A* **94**, 9191–9196.
- Sugimoto, C., Hasegawa, M., Zheng, H. Y. & 14 other authors (2002a).** JC virus strains indigenous to northeastern Siberians and Canadian Inuits are unique but evolutionally related to those distributed throughout Europe and Mediterranean areas. *J Mol Evol* **55**, 322–335.
- Sugimoto, C., Hasegawa, M., Kato, A., Zheng, H. Y., Ebihara, H., Taguchi, F., Kitamura, T. & Yogo, Y. (2002b).** Evolution of human polyomavirus JC: implications for the population history of humans. *J Mol Evol* **54**, 285–297.
- Suzuki, M., Zheng, H. Y., Takasaka, T., Sugimoto, C., Kitamura, T., Beutler, E. & Yogo, Y. (2002).** Asian genotypes of JC virus in Japanese-Americans suggest familial transmission. *J Virol* **76**, 10074–10078.
- Swenson, J. J., Trowbridge, P. W. & Frisque, R. J. (1996).** Replication activity of JC virus large T antigen phosphorylation and zinc finger domain mutants. *J Neurovirol* **2**, 78–86.
- Takasaka, T., Miranda, J. J., Sugimoto, C., Paraguison, R., Zheng, H.-Y., Kitamura, T. & Yogo, Y. (2004).** Genotypes of JC virus in Southeast Asia and the western Pacific: implications for human migrations from Asia to the Pacific. *Anthropol Sci* **112**, 53–59.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
- Vawter, L. & Brown, W. M. (1986).** Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. *Science* **234**, 194–196.
- Venter, M., Smit, S. B., Leman, P. & Swanepoel, R. (2004).** Phylogenetic evidence of widespread distribution of genotype 3 JC virus in Africa and identification of a type 7 isolate in an African AIDS patient. *J Gen Virol* **85**, 2215–2219.
- Wooding, S. (2001).** Do human and JC virus show evidence of host-parasite codemography? *Infect Genet Evol* **1**, 3–12.
- Worobey, M. & Holmes, E. C. (2001).** Homologous recombination in GB virus C/hepatitis G virus. *Mol Biol Evol* **18**, 254–261.
- Yanagihara, R., Nerurkar, V. R., Scheirich, I. & 9 other authors (2002).** JC virus in the western Pacific suggest Asian mainland

relationships and virus association with early population movements. *Hum Biol* 74, 473–488.

Yogo, Y., Zheng, H.-Y., Hasegawa, M., Sugimoto, C., Tanaka, S., Honjo, T., Kobayashi, N., Ohta, N. & Kitamura, T. (2003). Phylogenetic analysis of JC virus DNAs detected in Ainu: an attempt to elucidate the origin and diversity of the Ainu. *Anthropol Sci* 111, 19–34.

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

Zheng, H. Y., Sugimoto, C., Hasegawa, M. & 8 other authors (2003). Phylogenetic relationships among JC virus strains in

Japanese/Koreans and native Americans speaking Amerind or Na-Dene. *J Mol Evol* 56, 18–27.

Zheng, H. Y., Zhao, P., Suganami, H. & 7 other authors (2004a). Regional distribution of two related Northeast Asian genotypes of JC virus, CY-a and -b: implications for the dispersal of Northeast Asians. *Microbes Infect* 6, 596–603.

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

Zhivotovsky, L. A., Rosenberg, N. A. & Feldman, M. W. (2003). Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *Am J Hum Genet* 72, 1171–1186.