ELSEVIER

Infection, Genetics and Evolution 1 (2001) 3–12

Infection, Genetics and Evolution

www.elsevier.com/locate/meegid

Do human and JC virus genes show evidence of host–parasite codemography?

Stephen Wooding*

Department of Anthropology, University of Utah, 270 South 1400 East, Salt Lake City, UT 84112-0060, USA

Received 28 October 2000; received in revised form 2 February 2001; accepted 5 February 2001

Abstract

Information about similarities and differences in the demographic history of host and parasite populations is potentially useful for making inferences about a variety of evolutionary processes. However, it is difficult to observe the historical demographic properties of natural populations directly. Here, the extent of demographic similarity in a host and its parasite was examined indirectly by inferring long-term population history from patterns of genetic variation. Nucleotide sequence diversity in human and JC virus (JCV) DNA is consistent with a long-term demographic connection between the two species: both show evidence of large-scale population expansion. However, genetic data also suggest that the two species have different patterns of population substructuring. These similarities and differences have implications for adaptive evolution in JCV that are not evident when the two species are considered separately.

Keywords: JC virus; Host; Parasite

1. Introduction

Host–parasite interactions cause correlations between host and parasite populations that can be used to learn about evolutionary processes (Page and Charleston, 1998). For example, studies of evolutionary "arms races" use correlations between host and parasite adaptations to learn about evolutionary trajectories (Hamilton, 1993), and phylogenetic studies can use correlations in the long-term population history of hosts and parasites to make inferences about evolutionary rates (Ochman et al., 1999).

Models of population dynamics play an important role in the study of host–parasite interactions by describing their demographic consequences (Poulin, 1998; Boots and Sasaki, 1999). Under assumptions about parasite transmission rate, virulence, host reproduction rate and other parameters, population dynamic models make predictions about the properties of host–parasite interactions that can be compared against natural observations (May, 1991). Nonetheless, empirical studies of host–parasite population dynamics are uncommon (Berthier et al., 2000). One problem is that evolutionary processes are difficult to observe directly on short time scales (Anderson and May, 1979).

An alternative to the use of direct observation in studies of population dynamics is the use of molecular genetic data to make inferences. Population genetic analyses can use existing patterns of genetic variation to infer historical processes such as population size, population subdivision and natural selection (Rogers, 1995; Wooding and Ward, 1997; Zanotto et al., 1999). It should be possible to test hypotheses about specific host–parasite relationships by using the tools of population genetics to look for demographic correlations involving host–parasite pairs. I present here a population genetic analysis of the parasite JC virus (JCV) and its obligate host, *Homo sapiens*.

JCV is a small, circular DNA virus that resides with little pathological effect in the urinary tract of 70–90% of adults world wide (Ault and Stoner, 1992). JCV is transmitted horizontally, but it is passed most often from parent to child — an effectively vertical pattern of inheritance — and established infections persist in the host for extended periods (Kunitake et al., 1995).

Evidence that the pathogenicity of JCV is low and that its pattern of transmission is largely vertical have led to the suggestion that the virus may be a good surrogate for human samples in studies of migration and population affiliation (Sugimoto et al., 1997; Guo et al., 1998; Chima et al., 2000; Ryschkewitsch et al., 2000; Hatwell and Sharp, 2000). This suggestion is supported by some empirical findings, which

* Tel.: +1-801-581-4959; fax: 1-801-581-6252.
E-mail address: stephen.wooding@anthro.utah.edu (S. Wooding).

1567-1348/01/$ – see front matter
PII: S1567-1348(01)00002-8

show that JCV and human populations do share certain demographic properties. The relationship between phylogenetic relatedness and geographical distribution is similar in the two species, for example (Sugimoto et al., 1997; Guo et al., 1998). The extent to which patterns of genetic variation in JCV are representative of patterns in humans is still questionable, however. Superficial similarities in phylogeny and geographical distribution might be observed even if deeper differences are present. Factors like isolation by distance could lead to similarities on a broad scale even if factors like population subdivision or natural selection are acting locally.

To investigate the extent to which human and JC virus population histories are similar, I compared DNA sequences from Africa, Asia and Europe with a focus on their implications for demography. Human population history has been studied extensively and genetic diversity in humans has several distinctive features that provide a good basis for comparison. Under the null hypothesis that JCV and human demographies have historically been identical, JCV should show evidence of the same distinctive features as humans. Departures from similarity are potentially informative about the relevance of underlying factors such as natural selection and regional differentiation.

2. Materials and methods

2.1. Data

A sample of 379 nucleotide sequences from the V–T intergenic region of JCV was assembled from Sugimoto et al. (1997). These sequences are from 42 populations on three continents (Fig. 1, and see Fig. 1 in Sugimoto et al., 1997). The sequences are 612 nucleotides long and span portions of two coding regions (capsid VP1 and Large T antigen) and a non-coding segment (intergenic region) of the JCV genome (Ault and Stoner, 1992).

Sequence variation in the sample is composed of both nucleotide substitutions and insertions/deletions. One sequence from Italy (accession number AB004477 in Genbank; type IT-4 in Sugimoto et al., 1997) contains a long sequence gap of unknown origin, so it was removed from all analyses, leaving a total of 378 nucleotide sequences, with 11 from Italy, in this paper.

Sequences were aligned by eye, since the overall differences between them were low, and were then partitioned into two main datasets. Amino acid sequences were inferred for the coding portions of the dataset using the translation code designated by Genbank (standard). DNA sequences were modified by removing four nucleotide positions at which an insertion or deletion was observed, leaving a total of 608 nucleotide positions.

Computer files containing the JCV datasets used in this paper are available from the author.

	Location	Symbol	n
Africa	Accra, Ghana	GH	4
	Addas Abbaba, Ethiopia	ET	8
	Bangui, Central African Republic	CA	11
	Fes/Ifane, Morocco	MR	21
	Khartoum, Sudan	SU	9
	Lusaka, Zambia	ZA	5
	Nairobi, Kenya	KE	8
	Nouakchott, Mauritania	MA	10
	Port Louis, Mauritius	MU	8
	Tessaoua, Niger	NG	8
	Welkom, South Africa	SO	6
Asia	Ankara, Turkey	TU	15
	Beijing, China	CB	10
	Chengdu, China	CD	10
	Chiang Mai, Thailand	TL	11
	Colombo, Sri Lanka	SL	5
	Guangzhou, China	GZ	13
	Harbin, China	HB	6
	Ishikawa, Japan	IK	11
	Jakarta, Indonesia	ID	17
	Masai, Malaysia	ML	14
	Okinawa, Japan	ON	11
	Pamalican Is., Phillipines	PH	8
	Riyadh, Saudi Arabia	SA	20
	Seoul, South Korea	SK	14
	Shenyang/Jinzhou, China	SJ	7
	Taipei, China	TP	9
	Tokyo, Japan	TY	14
	Ulaanbaatar, Mongolia	MO	12
	Varanasi, India	IN	17
	Wuhan, China	CW	10
	Yangon, Myanmar	MN	15
Europe	Athens, Greece	GR	20
	Barcelona, Spain	SP	13
	Budapest, Hungary	HU	17
	Deventer, Netherlands	N	12
	Illertissen, Germany	G	8
	London, United Kingdom	UK	6
	Novosibirsk, Russia	RS	14
	Prague, Czech Republic	CR	18
	Rome, Italy	IT	11*
	Stockholm, Sweden	SW	18

Fig. 1. Sampled locations. Locations, symbols and sample sizes of data assembled from Sugimoto et al. (1997); (* only 11 of 12 sequences from Italy reported by Sugimoto et al. were used in this study. Genbank ID AB004477, type IT-4, was excluded because it has a long gap of unknown origin.

2.2. Analyses

Five analyses that are standard in studies of human genetic diversity were applied to patterns of diversity in JCV. These were chosen to focus on characteristics of human data that are regarded as hallmark sources of evidence about human demographic history.

First, patterns of differentiation among individual DNA sequences and amino acid sequences, and among populations of DNA sequences, were examined using principal components analysis, which gives a summary description of genetic differences (Harpending and Jenkins, 1973). Second, Tajima's *D*-test was applied to the nucleotide sequence data to test the hypothesis that the JCV population has been stationary (i.e. variation is selectively neutral and

population size has been constant) (Tajima, 1989). Third, Rogers' method of moments was applied to the distribution of pairwise differences among sequences (or mismatch distribution) to estimate historical demographic parameters (Rogers, 1995, 1997). Fourth, levels of genetic diversity within continents were compared using p, the per nucleotide difference between sequences (Li, 1997). Fifth, the ratio of synonymous substitutions per synonymous nucleotide position and non-synonymous substitutions per non-synonymous nucleotide position (i.e. K_s/K_a ratios) were calculated using the tools of (Yang, 2000), and analyzed in pairwise comparisons as described by Nei and Kumar (2000).

The results of these analyses were compared with similar analyses of human genetic data in the published literature (Harpending and Rogers, 2000; Rogers, 1995; Cavalli-Sforza et al., 1994; Eller, 1999; Stoneking, 1997; Zhao et al., 2000).

3. Results and discussion

Patterns of genetic diversity in the JC virus have been used as a surrogate for human genetic diversity in several studies of population structure and migration (Sugimoto et al., 1997; Guo et al., 1998). Such studies have relied on an apparent large-scale correlation between JCV and human population structure: on an average, distantly related human populations contain distantly related JCV lineages, and closely related human populations contain closely related JCV lineages (Sugimoto et al., 1997). However, a variety of factors could lead to superficial similarities between human and JC virus populations even if deeper differences are present. If isolation by distance occurs in both populations, for example, then human and JCV populations might show similar patterns of regional differentiation superimposed on differences in population size trends or population subdivision.

Given the rising availability of human genetic data, a more effective approach is to treat the historical relationship between human and JCV as an unknown. Human history is exceptionally well studied and information about humans should provide a good basis for comparison with other species. Evidence for low levels of pathogenicity and largely vertical patterns of transmission in JCV suggest the null hypothesis that human and JCV populations have strongly correlated demographic histories. Inferred differences may be useful for learning about long-term demographic and evolutionary processes in the virus.

Human genetic diversity is characterized by four hallmark trends that are relevant to the investigation of long-term demographic processes in JCV (Harpending and Rogers, 2000). First, phylogenies of human genes tend to be comb-shaped, with most terminal branches being nearly the same length (Cann et al., 1987; Harpending and Rogers, 2000). Second, the distribution of pairwise differences among linked polymorphisms, or mismatch distribution, is a smooth unimodal wave in most human populations (Rogers and Harpending, 1992; Harpending and Rogers, 2000; Wooding and Rogers, 2000; Alonso and Armour, 2001). Third, genetic differences between human populations tend to be correlated with geographical distance, and genetic differences between continents tend to be relatively uniform (Eller, 1999; Harpending and Rogers, 2000; Cavalli-Sforza et al., 1994). Fourth, human populations within Africa tend to be more genetically diverse than populations on other continents (Stoneking, 1997). These patterns are consistent with Late Pleistocene population growth and range expansion originating in Africa (Harpending and Rogers, 2000).

Among surveys of genetic diversity in JCV, the sample of Sugimoto et al. (1997) is most informative in a demographic context. It includes a large number of sample populations distributed over three continents and is comparable in both size and geographical distribution to a number of human genetic datasets (Harpending and Rogers, 2000; Cavalli-Sforza et al., 1994) (Fig. 1). In earlier phylogeographic analyses of the sample, three main trends were identified. Phylogenetic trees relating DNA sequences showed evidence of several clades with relatively limited geographical distributions, including a distinct, basal European clade (designated Eu) (Fig. 1 in Sugimoto et al., 1997). Graphical displays of the data supported the existence of regional differences in lineage composition and relatedness (Sugimoto et al., 1997; Guo et al., 1998). Populations with known histories of human migration and admixture showed evidence of migration and admixture with respect to viral diversity (Kato et al., 1997; Sugimoto et al., 1997).

Earlier findings in JCV are consistent with observations based on human genes (Harpending and Rogers, 2000), but there are some important differences. The tendency of human genetic lineages to group geographically, and the distinctness of migrant lineages in admixed populations are well established (Cavalli-Sforza et al., 1994). However, the presence of a distinct, basal clade of European lineages is unexpected. In human gene genealogies the basal position of lineages from Africa is often interpreted as evidence for an African origin of modern humans (Cann et al., 1987; Horai et al., 1995; Thompson et al., 2000; Cavalli-Sforza et al., 1994). The phylogenetic position of European lineages in JCV is inconsistent with the argument that JCV-infected humans prior to their hypothesized expansion out of Africa.

The distinctness of the Eu clade recognized by Sugimoto et al. (1997) is also unexpected. In a principal components analysis of differences among V–T intergenic region nucleotide sequences based on synonymous nucleotide substitutions only, the first two principal components of variation account for 9.2 and 5.2% of the total variance, respectively, and divide JCV lineages into two distinct clusters (Fig. 2). A large cluster (A) is observed, which contains lineages found predominantly in Asia and Africa, and a smaller cluster (B) is observed, which corresponds to the Eu clade identified previously. Such clusters are not typically observed in principal components maps of large samples of human genetic

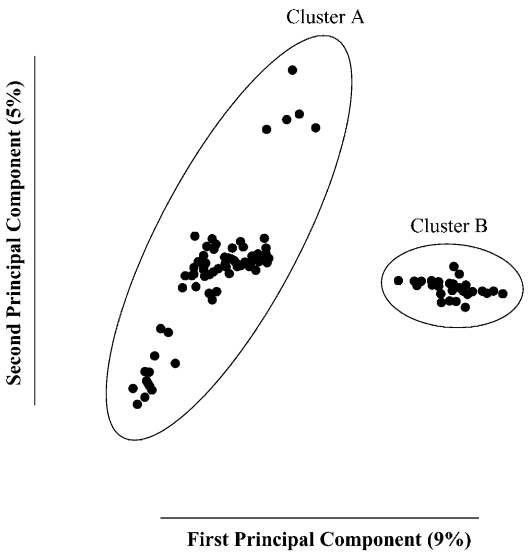

Fig. 2. Principal components map of JCV DNA sequences. Percent values on the axes indicate the proportion of variance accounted for by each component, but otherwise the axes are dimensionless.

diversity, which tend to have more diffuse distributions. Some forms of balancing selection could cause clustering to occur, but such selection would only explain the presence of clusters, not their different geographical distributions (Page and Holmes, 1998).

A principal component analysis of JCV subpopulations based on DNA sequence variation is consistent with the hypothesis of regional subdivision, as well. Like JCV nucleotide sequences, JCV subpopulations are divided into two main clusters defined by geography (Fig. 3). These population clusters are not as distinct as the DNA sequence clusters. Although African and Asian populations are composed mostly of lineages from cluster A and European populations are composed mostly of lineages from cluster B, the geographical division is not strict. Lineage clusters A and B are both found on all three continents.

One notable feature of the principal components map of JCV subpopulations is the similarity of Asian and African subpopulations, which differ from European subpopulations. Analyses of human genetic diversity ordinarily reflect more uniform between-continent differences (Eller, 1999; Ding et al., 2000). Another notable feature is that subpopulations located near the geographical intersection of continents are located near the intersection of continents on the principal components map, as well. Populations from Greece, Turkey, Morocco and Saudi Arabia, for example, fall between populations from Europe and Africa, and Europe and Asia. This pattern is consistent with the argument that isolation by distance may be the source of similarity in the large-scale genetic structure of human and JCV populations.

Levels of diversity within continents differ from observations in humans as well. Measures based on the mean number of nucleotide differences between human genes consistently find more genetic variability in Africa than in Asia and Europe. This trend is generally interpreted as evidence for African origins (Stoneking, 1997). In a comparison of DNA sequences from chromosome 22, for example, Zhao et al. (2000) found π values in Europe, Asia and Africa of 0.0077, 0.0075 and 0.0085, respectively. In JCV, diversity is higher in Europe ($\pi = 0.0185$) than in Asia ($\pi = 0.0171$) or Africa ($\pi = 0.0169$).

The phylogenetic differences between lineage clusters A and B and overall differences in diversity between the European and African/Asian clusters suggest that the JCV subpopulations found in Europe are demographically distinct from JCV populations elsewhere. Geographically

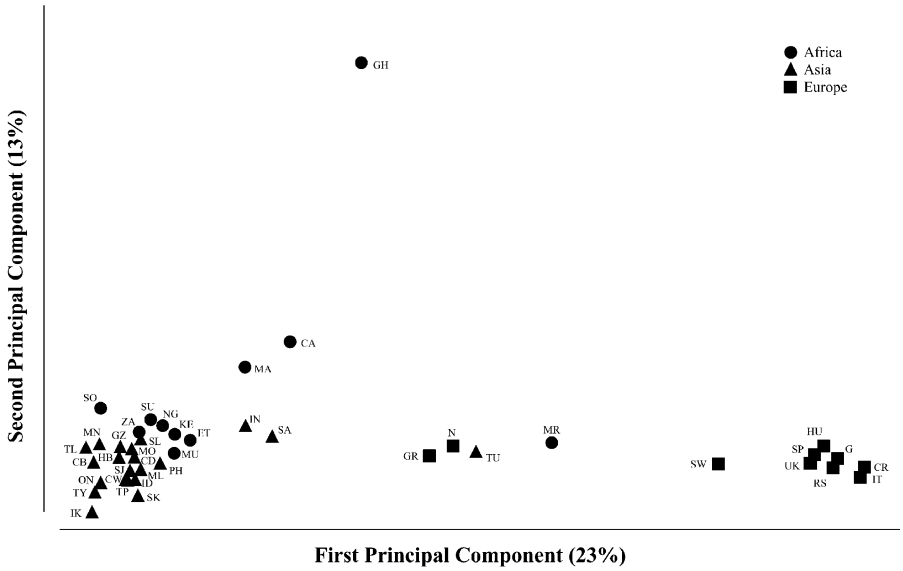

Fig. 3. Principal components map of JCV subpopulations based on DNA sequence variation. Percent values on the axes indicate the proportion of variance accounted for by each component, but otherwise the axes are dimensionless. A key for population symbols is shown in Fig. 1.

defined demographic units may be present. An effective method for comparing the demographic properties of populations is to compare the distribution of pairwise differences among sequences, or mismatch distribution. The mismatch distribution is a well-established tool for studying populations that have been subdivided or have changed in size over time. Whereas single, panmictic populations tend to show moderately rough mismatch distributions, strongly subdivided populations tend to show two modes (Marjoram and Donnelly, 1994). Whereas populations that have expanded exhibit smooth mismatch distributions with a single mode correlated with the time of expansion, populations that have not expanded exhibit rough mismatch distributions (Rogers and Harpending, 1992; Rogers, 1995).

The mismatch distribution in JCV DNA sequences shows evidence of both expansion and subdivision (Fig. 4a). Unlike the mismatch distribution for human mtDNA, which is smooth and unimodal, the mismatch distribution of JCV sequences is smooth, but strongly bimodal, with peaks at 11 and 21 nucleotide differences. Such patterns are unusual, but have been described in other cases population subdivision followed by population expansion (Wooding and Ward, 1997). No methods are available to estimate demographic parameters for subdivided, expanding populations, but the distinctness of JCV clusters A and B indicates that it is appropriate to analyze them separately.

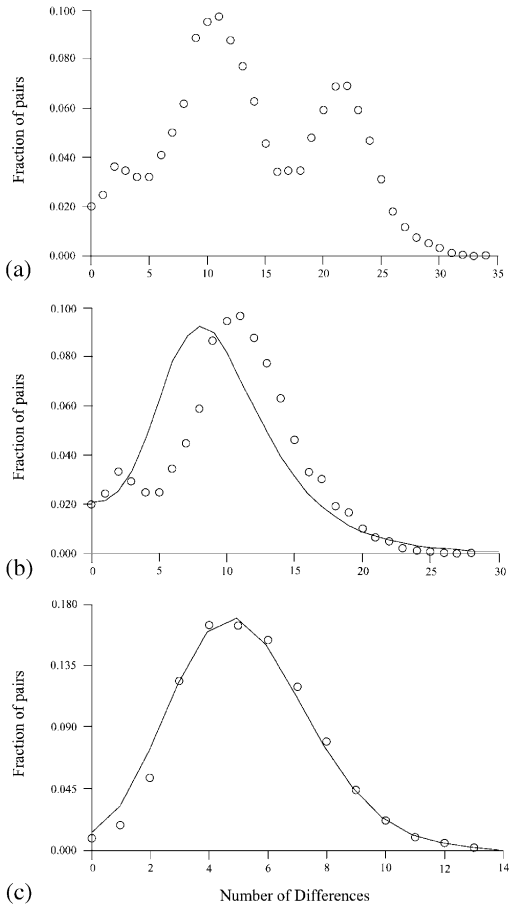

Fig. 4. Mismatch distributions. (a) Mismatch distribution of all JCV nucleotide sequences. Circles represent observed differences; (b) mismatch distribution of JCV nucleotide sequences from cluster A. Circles represent observed differences. Solid line represents the mismatch distribution expected for the parameters estimated for cluster A using Rogers' method of moments ($\hat{\theta}_0 = 3.52$, $\hat{\theta}_1 = 47.32$ and $\hat{\tau} = 6.82$); (c) mismatch distribution of JCV nucleotide sequences from cluster B. Circles represent observed differences. Solid line represents the mismatch distribution expected for the parameters estimated for cluster B using Rogers' method of moments ($\hat{\theta}_0 = 0.01$, $\hat{\theta}_1 = 118.20$ and $\hat{\tau} = 5.35$).

Separate analyses of the lineages in cluster A and B yield results expected under population growth, without subdivision. A test of the frequency spectrum of mutations in each sample using Tajima's *D*-statistic, which compares the total number of variable nucleotide positions in a sample of sequences with the mean pairwise difference, rejects the hypothesis of demographic stationarity for both groups at the 95% confidence level ($D = -1.76$ for cluster A, and -1.94 for cluster B) (Tajima, 1989). The mismatch distributions of the two clusters also match expectations under population expansion when analyzed separately. Both are smooth and unimodal. However, the modal difference within cluster A is nine nucleotides, and the modal difference within cluster B is four nucleotides (Fig. 4b).

The application of Rogers' method of moments yields $\hat{\theta}_0 = 3.52$, $\hat{\theta}_1 = 47.32$ and $\tau = 6.82$ for cluster A and $\hat{\theta}_0 = 0.01$, $\hat{\theta}_1 = 118.20$ and $\tau = 5.35$ for cluster B (Rogers, 1995). The estimate of τ for cluster A seems slightly too low, possibly due to substructuring within the cluster. Under a two-epoch "sudden change" model of population history, which makes the assumption that a population changed instantaneously from an ancient population size to the modern one, $\theta = 2N_0\mu$ provides an estimate of ancient population size N_0) (the number of haploid genomes in the population), where μ is the per nucleotide substitution rate times the number of nucleotides in the sequence under consideration. Similarly, $\tau = 2\mu t$ provides an estimate of the length of time t that has passed since the size change occurred (Rogers, 1995).

Under the synonymous nucleotide substitution rate for primate polyomaviruses proposed by Yasunaga and Miyata (1982) (3.8×10^{-8} synonymous substitutions per site per year), $\blacksquare = 2.3 \times 10^{-5}$ and estimates of τ obtained using the method of moments imply an expansion time of 150,000 years before present for cluster A and 115,000 years before present for cluster B. These expansion times are similar to the expansion time of 66,000–150,000 years before presented for humans inferred from mitochondrial DNA polymorphism (Rogers, 1995). However, the time of expansion for cluster B under this substitution rate predates the permanent occupation of Europe by humans, which occurred after the last glacial maximum around 18,000 years ago (Klein, 1999). Such an early expansion time implies that this substitution rate is too low.The rate of 3.8×10^7 synonymous substitutions per site per year proposed for JCV by Hatwell and Sharp (2000) yields different results. It implies an expansion time of around 14,000 years before present for cluster A

and 11,000 years before present for cluster B. However, this rate rests on the assumption that the earliest divergences in JCV correspond to those found in humans. It is circular to use Hatwell and Sharp's estimate for comparisons between JCV and human diversity here since the substitution rates of the two species were not calculated independently.

The signature of population expansion and subdivision in JCV could be due to a variety of factors. One possibility is that JCV populations have been divided into demes corresponding to isolated human subpopulations. Another possibility is that natural selection is causing between-continent differentiation. Under the latter hypothesis, variation should be present in the phenotypes of JCV variants. These might appear in the form of amino acid substitutions in the V–T intergenic region.

Translations of the DNA sequences reported by Sugimoto et al. (1997) show that amino acid sequence variation is present in both of the coding regions reported. A total of 37 different sequences defined by 27 variable amino acid positions is observed (Fig. 5). A total of 23 of the observed variants occur only once in the sample, with the remaining 14 occurring two or more times.

The structure of a parsimony network generated using the method described by Wooding and Ward (1997) shows some of the structure suggested by the principal components analysis of JCV DNA sequences in Figs. 2 and 3; amino acid sequences in clusters A and B fall into different parts of the network. Europe and Africa each contain fewer amino acid variants than Asia does, and every lineage occurring more than once in the sample is found in Asia. Also, as suggested

Variable Position

Amino Acid Sequence

```
                    0000000000111111111122222222
                    1234567890123456789012345678
JCV_A_1             ALFLVFGGSRVDIQMRMKPTQYHGSASC
JCV_A_2             T...I.......VE...........T..
JCV_A_4             T........F...VE.............
JCV_A_5             T.........K...E.............
JCV_A_6             T.........I..E.............
JCV_A_7             T............EI............
JCV_A_8             T............E............F.
JCV_A_9             T............E.............
JCV_A_10            T............E.........S....
JCV_A_11            T..............K..........F.
JCV_A_12            T..............K...........
JCV_A_13            T....................P.....F.
JCV_A_14            T.........................F.
JCV_A_15            T..........................
JCV_A_19            T...........VE.....A........
JCV_A_20            T...........VE.....N..P.....
JCV_A_21            T...........VE.......F....F.
JCV_A_25            T...........VE..........P.F.
JCV_A_26            T...........VE............F.
JCV_A_27            T...........VE.............
JCV_A_28            T...........VE............Y.
JCV_A_29            T...........VE...........T..
JCV_A_31            T...........VE...R..........
JCV_A_32            T...........VE..T...........
JCV_A_33            T......R.....E.............
JCV_A_34            T....L......VE...........T..
JCV_A_35            T..S........VE.............
JCV_A_36            T.S........GVE.............
JCV_A_37            TP...........E.............
JCV_A_3             T.....D.....VG.K.....FN...F.
JCV_A_16            T...........VE.K.....FN...F.
JCV_A_17            T...........VE.K.....FN.....
JCV_A_18            T...........VE.K..Q..FN...F.
JCV_A_22            T...........VE.......FN...F.
JCV_A_23            T...........VE.......FN...FY
JCV_A_24            T...........VE.......FN.....
JCV_A_30            T...........VE........N...F.
```

Fig. 5. Summary of variable amino acid positions. Variable sites are listed and numbered in the order they occur in the sequence data. Sequence JCV_A_1 is the reference sequence. Below it, letters indicate substitutions and dots indicate identity. Unshaded sequences are from cluster A and shaded sequences are from cluster B.

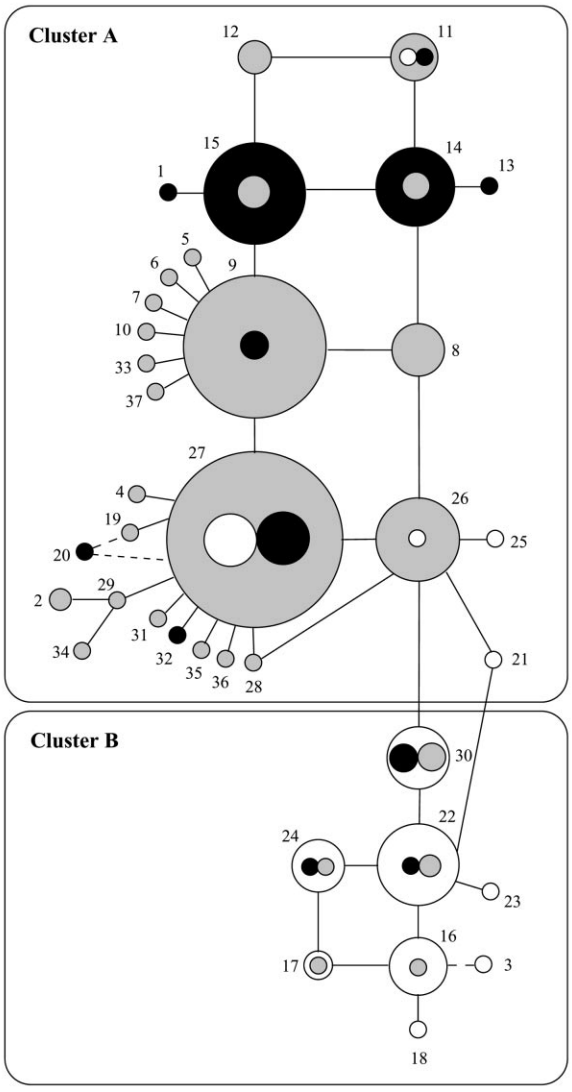

Fig. 6. Parsimony network relating amino acid variants. Node sizes are proportional to lineage frequencies. Within each node, shading indicates the relative abundance of the lineage in each of the three continents: black represents Africa, gray represents Asia and white represents Europe.

by the principal components analysis of JCV subpopulations in Fig. 3, the lineages endemic to Europe are relatively distinct from those found in Africa and Asia, although lineage compositions of the continents overlap (Fig. 6).

The population structure suggested by the relative abundances of amino acid lineages in each region is reflected in a principal components map of JCV subpopulations based on amino acid sequence variation (Fig. 7). In this map, European subpopulations are distinct from African and Asian subpopulations, but African and Asian subpopulations are indistinguishable from each other and form a single nebulous cluster (Fig. 7). The extent of overlap among African and Asian subpopulations to the exclusion of Europeans is remarkable given that the amino acid sequence variation is a subset of DNA sequence variation distinguishing the three continents. The simplest explanation for this pattern is that natural selection has occurred, resulting in convergent evolution.

The pattern of amino acid substitution in the parsimony network is consistent with the hypothesis that convergent evolution has occurred in JCV proteins (Fig. 8). Of the eight amino acid substitutions informative in a parsimony analysis (i.e. substitutions in which at least two variants are observed in two or more samples), five appear to have occurred only once in JCV's history, but three must have occurred more than once. Two (16 R-K and 22 Y-F) can be explained by one convergent amino acid substitution each. One substitution (27 S-F) requires at least six convergent amino acid substitutions to explain. For example, if the 27 S-F substitution occurred once between lineage 11 and 12, then the similarity of lineage 14 to 15 would require a parallel mutation, the similarity of lineage 8 to 9 would require another parallel mutation and so on.

An alternative explanation for the recurrence of 27 S-F substitutions in JCV haplotypes is genetic recombination. Recombination has been regarded in much of the published literature as an insignificant factor in natural JCV populations, but the pattern observed here would be easily explained by the presence of a recombination hot spot. Unfortunately, an analysis using the method of Jakobsen and Easteal (1996), which tests for recombination by identifying cladistic incompatibilities among variable sites, shows that the 27 S-F substitution is the last cladistically informative substitution in the data set. It is impossible to tell whether this position is cladistically incompatible with substitutions both upstream and downstream, which would point to repeated amino acid substitution, or if it only incompatible with substitutions that are upstream, which would point to recombination.

One way to distinguish the hypothesis of multiple substitution from recombination would be to examine the synonymous nucleotide changes in the vicinity of the amino acid substitution. If recombination has occurred, then synonymous substitutions upstream from the 27 S-F substitution should be incompatible with those downstream. Unfortunately, not only is the informative amino acid substitution of interest the last amino acid substitution in the sequence, the non-synonymous change that causes the substitution is the third to last informative nucleotide substitution in the sequence. No convincing pattern is found in the final two informative positions. Comparisons of more JCV DNA sequence data will be required to answer the question of whether the 27 S-F changes have occurred through parallel substitution or recombination.

The bimodality of JCV's mismatch distribution, the basal position of the European clade in a phylogeny, the excessive

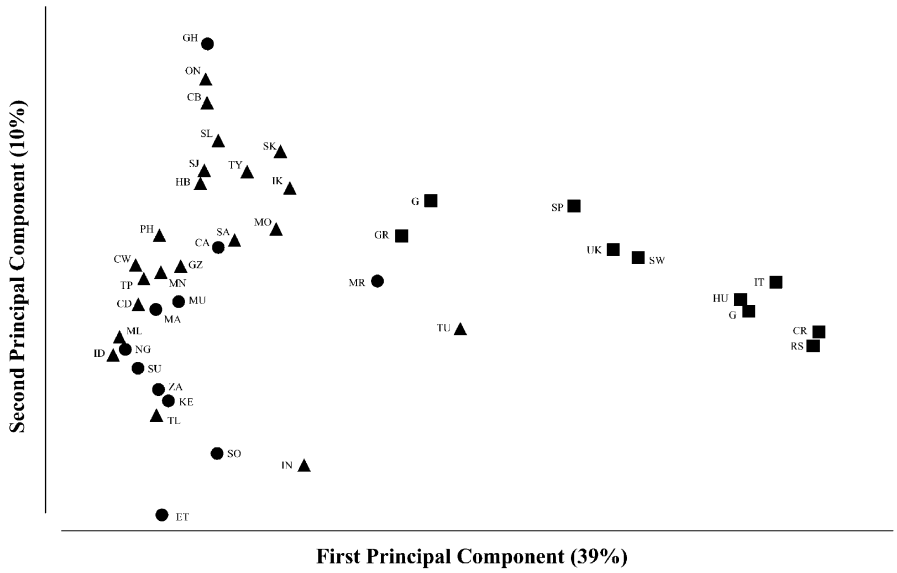

Fig. 7. Principal components map of JCV subpopulations based on amino acid sequence variation. Percent values on the axes indicate the proportion of variance accounted for by each component, but otherwise the axes are dimensionless. A key for population symbols is shown in Fig. 1.

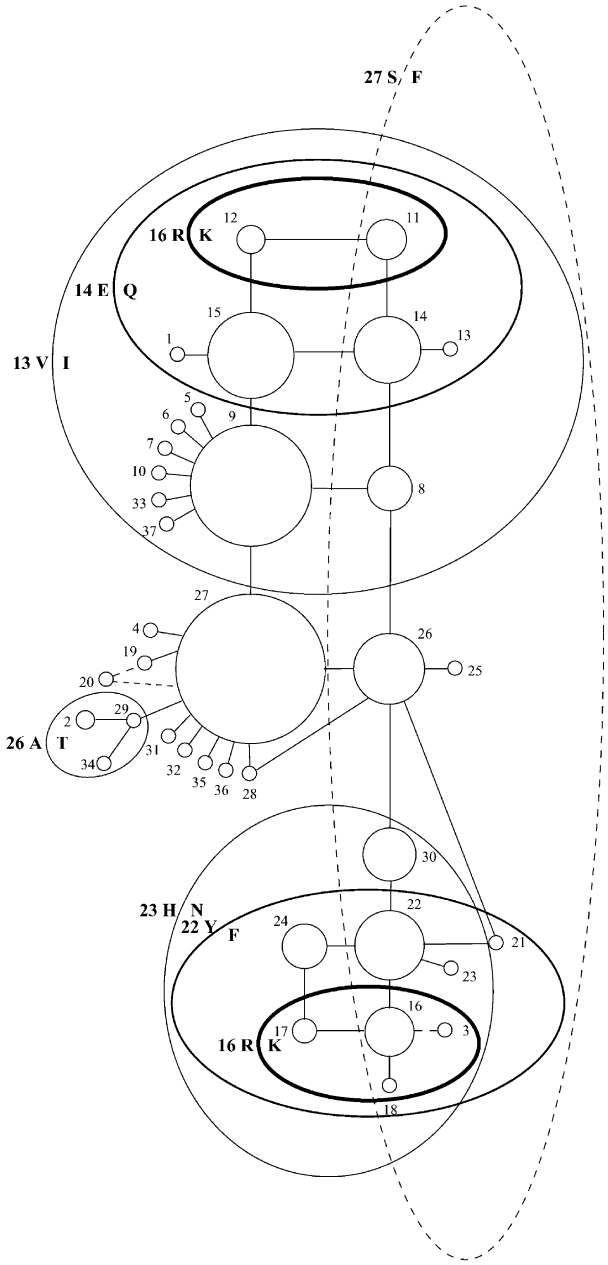

Fig. 8. Parsimony network indicating amino acid substitutions. Contours define the state of amino acid substitutions in which each variant is found in two or more samples (i.e. substitutions informative in a parsimony analysis). The number–letter pairs on each contour indicate the substitution and its alternate states. The number is the substitution number as defined in Fig. 5. The number inside a contour indicates the residue found in lineages inside the contour. The letter outside a contour indicates the residue found in lineages outside the contour. The dashed contour indicates the 27 S-F substitution.

genetic diversity in Europe and the predominantly European origins of cluster B imply that although JCV and humans have a tightly coupled biological relationship, their demographic histories differ. JCV DNA sequences are most likely a poor source of data for studies of human demography outside of the low-resolution applications reported so far. The use of genetic diversity in the JCV parasite to study human populations may be possible at low scales of resolution, but differences between JCV and human demography indicate that inferences about human populations based on diversity in the virus should be treated with caution.

Given the abundant and rapidly increasing base of human genetic data and the depth of detailed investigations into human population history based on genetic, linguistic, morphometric and archaeological evidence (Cavalli-Sforza et al., 1994), an alternative approach to analyzing genetic diversity in JCV is justified. Information about human demographic history should be used as a context for understanding the demographic history of the virus, rather than the other way around. The incorporation of well-supported prior findings about either host or parasite populations into studies of host–parasite coevolution is likely to provide a useful foundation for identifying otherwise unobservable trends of the evolutionary process. In the case of the human–JCV relationship, similarities and differences in patterns of genetic diversity have a variety of implications that are not evident when the two species are considered separately.

The most obvious pattern of diversity in JCV that is placed in context by information about human population history is evidence for population growth. Whereas evidence for a population expansion in JCV in the absence of an expansion in humans would be clear evidence of an epidemiological sweep, evidence for an expansion in both species allows the possibility that they expanded simultaneously. The hypothesis that human and JCV populations expanded simultaneously will be testable if more accurate estimates of nucleotide substitution rates in JCV can be obtained. Information about human history is important for interpreting phylogenetic topologies and diversity levels in JCV populations as well. These lines of evidence suggest that JCV's relationship with humans began in Europe. While not statistically testable, the basal position of a European clade of JCV lineages, and the relatively high level of diversity (π) levels there, are both more consistent with European origins than they are with prior evidence for human origins in Africa, which are supported by analogous patterns in human genes.

Patterns of between-continent divergence in JCV are similar to patterns found in humans in that they both show evidence for isolation by distance. But the distinct difference between European and Africa/Asian JCV subpopulations, and the marked similarity between subpopulations from Asia and Africa, indicate that human population structure alone cannot account for diversity patterns in the virus. The population dynamics of JCV must include the influence of factors such as natural selection or life history traits, such as horizontal transmission, that could enforce between-region similarities and differences independent of those found in humans. High levels of amino acid diversity and evidence for purifying selection imply that natural selection is an important underlying force that may be important in the regional differentiation of viral and human populations.

Further examination of differences between regional subpopulations will be required to define fine-scale trends in the adaptive evolution of JCV. The specific comparison of European and non-European populations should be particularly informative in understanding these trends.

Acknowledgements

Helpful comments and discussion were provided by Henry Harpending, John Hawks, Alan Rogers and four anonymous reviewers. This project was supported partly by a NIH Genome Sciences Training Grant to the University of Utah.

References

Alonso, S., Armour, J.A.L., 2001. A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outside Africa. Proc. Natl. Acad. Sci. U.S.A. 98, 864–869.

Anderson, R.M., May, R., 1979. Population biology of infectious diseases. Part I. Nature 280, 361–367.

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. J. Gen. Virol. 73, 2669–2678.

Berthier, K., Langlais, M., Auger, P., Pontier, D., 2000. Dynamics of a feline virus with two transmission modes within exponentially growing host populations. Proc. R. Soc. London Ser. B 267, 2049–2056.

Boots, M., Sasaki, A., 1999. Small worlds and the evolution of virulence: infection occurs locally and at a distance. Proc. R. Soc. London Ser. B 266, 1933–1938.

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

Cavalli-Sforza, L.L., Menozzi, P., Piazza, A., 1994. The History and Geography of Human Genes. Princeton University Press, Princeton, NJ.

Chima, S.C., Ryschkewitsch, C.F., Fan, K.J., Stoner, G.L., 2000. Polyomavirus JC genotypes in an urban united states population reflect the history of African origin and genetic admixture in modern African Americans. Hum. Biol. 72, 837–850.

Ding, Y., Wooding, S., Harpending, H., Chi, H., Li, H., Fu, Y., Pang, J., Yao, Y., Yu, J.X., Moyzis, R., Zhang, Y., 2000. Population structure and history in East Asia. Proc. Natl. Acad. Sci. U.S.A. 25, 14003–14006.

Eller, E., 1999. Population substructure and isolation by distance in three continental regions. Am. J. Phys. Anthopol. 108, 147–159.

Guo, J., Sugimoto, C., Kitamura, T., Ebihara, H., Kato, A., Guo, Z., Liu, J., Zheng, S.P., Wang, Y.L., Na, Y.Q., Suzuki, M., Taguchi, F., Yogo, Y., 1998. Four geographically distinct genotypes of JC virus are prevalent in China and Mongolia: implications for the racial composition of modern China. J. Gen. Virol. 79, 2499–2505.

Hamilton, W.H., 1993. Haploid dynamical polymorphism in a host with matching parasites: effects of mutation/subdivision, linkage and patterns of selection. J. Heredity 84, 328–338.

Harpending, H.C., Jenkins, T., 1973. Genetic distance among southern African populations. In: Crawford, M., Workman, P. (Eds.), Method and Theory in Anthropological Genetics. University of New Mexico Press, Albuquerque, pp. 177–199.

Harpending, H.C., Rogers, A.R., 2000. Genetic perspectives on human origins and differentiation. Ann. Rev. Genomics Hum. Gen. 1, 361–385.

Hatwell, J.N., Sharp, P.M., 2000. Evolution of human polyomavirus JC. J. Gen. Virol. 81, 1191–1200.

Horai, S., Hayasaka, K., Kondo, R., Tsugane, K., Takahata, N., 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. U.S.A. 92, 532–536.

Jakobsen, I.B., Easteal, S., 1996. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. BioInformatics 12, 291–295.

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.

Klein, R.G., 1999. The Human Career: Human Biological and Cultural Origins. University of Chicago Press, Chicago.

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 virus. J. Clin. Microbiol. 33, 1448–1451.

Li, W.H., 1997. Molecular Evolution. Sinauer Associates.

Marjoram, P., Donnelly, P., 1994. Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. Genetics 136, 673–683.

May, R.M., 1991. The dynamics and genetics of host–parasite associations. In: Toft, C.A., Aeschlimann, A., Bolis, L. (Eds.), Parasite–Host Associations: Coexistence or Conflict? Oxford Science Publications, Oxford, pp. 102–128.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, UK.

Ochman, H., Elwyn, S., Moran, N.A., 1999. Calibrating bacterial evolution. Proc. Natl. Acad. Sci. U.S.A. 96, 12638–12643.

Page, R.D.M., Charleston, M.A., 1998. Trees within trees: phylogeny and historical association. TREE 13, 356–359.

Page, R.D.M., Holmes, E.C., 1998. Molecular Evolution: a Phylogenetic Approach. Blackwell Scientific Publications, London.

Poulin, R., 1998. Evolutionary Ecology of Parasites: from Individuals to Communities. Chapman & Hall, New York.

Rogers, A.R., 1995. Genetic evidence for a Pleistocene population explosion. Evolution 49, 608–615.

Rogers, A.R., 1997. Mismatch version 4.3. Computer program distributed by the author, Department of Anthropology, University of Utah.

Rogers, A.R., Harpending, H.C., 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. E 9, 552–569.

Ryschkewitsch, C.F., Friedlaender, J.S., Mgone, C.S., D.V. Jobes, D., Agostini, H.T., Chima, S.C., Alpers, M.P., Koki, G., Yanagihara, R., Stoner, G.L., 2000. Human polyomavirus JC variants in Papua New Guinea and Guam reflect ancient population settlement and viral evolution. Microbes and Infection 2, pp. 987–996.

Stoneking, M., 1997. Recent African origin of human mitochondrial DNA: review of the evidence and current status of the hypothesis. In: Progress in Population Genetics and Evolution, pp. 1–14.

Sugimoto, C., Kitamura, T., Guo, J., Al-Hadal, M.N., Shchelkunov, S.N., Otova, B., Ondrejka, P., Chollet, J.Y., El-Safi, S., Ettayebi, M., Gresenguet, G., Kocagoz, T., Chaiyarasamee, S., Thant, K.Z., Moe, K., Kobayashi, N., Taguchi, F., Yogo, Y., 1997. Typing of urinary JC virus DNA offers a novel means of tracing human migrations. Proc. Natl. Acad. Sci. U.S.A. 94, 9191–9196.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Thompson, R., Pritchard, J.K., Shen, P., Oefner, P.J., Feldman, M.W., 2000. Recent common ancestry of human Y chromosomes: evidence form DNA sequence data. Proc. Natl. Acad. Sci. U.S.A. 7360–7365.

Wooding, S., Rogers, A., 2000. A Pleistocene population X-plosion? Hum. Biol. 72, 693–695.

Wooding, S., Ward, R., 1997. Phylogeography and Pleistocene evolution in the North American black bear. Mol. Biol. E 14, 1096–1105.

Yang, Z., 2000. Phylogenetic Analysis by Maximum Likelihood (PAML), version 3.0c. Computer program distributed by the author, University College London, London, UK.

Yasunaga, T., Miyata, T., 1982. Evolutionary changes of nucleotide sequences of papova viruses BKV and SV40: they are possibly hybrids. J. Mol. Evol. 19, 72–92.

Zanotto, P.M., Kallas, E.G., de Souza, R.F., Holmes, E.C., 1999. Positive selection in the nef gene of HIV-1. Genetics 153, 1077–1089.

Zhao, Z., Jin, L., Fu, Y., Ramsay, M., Jenkins, T., Leskinen, E., Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Ramos-Onsins, S., Yu, N., Li, W., 2000. Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc. Natl. Acad. Sci. U.S.A. 97, 11354–11358.