A third genotype of the human parvovirus PARV4 in sub-Saharan Africa

Peter Simmonds,1 Jill Douglas,1 Giovanna Bestetti,2 Erika Longhi,2 Spinello Antinori,2 Carlo Parravicini2 and Mario Corbellino2

1Centre for Infectious Diseases, University of Edinburgh, Summerhall, Edinburgh EH9 1OQ, UK
2Infectious Diseases and Tropical Medicine and Pathology Departments, Luigi Sacco Hospital, Milano, Italy

PARV4 is a recently discovered human parvovirus widely distributed in injecting drug users in the USA and Europe, particularly in those co-infected with human immunodeficiency virus (HIV). Like parvovirus B19, PARV4 persists in previously exposed individuals. In bone marrow and lymphoid tissue, PARV4 sequences were detected in two sub-Saharan African study subjects with AIDS but without a reported history of parenteral exposure and who were uninfected with hepatitis C virus. PARV4 variants infecting these subjects were phylogenetically distinct from genotypes 1 and 2 (formerly PARV5) that were reported previously. Analysis of near-complete genome sequences demonstrated that they should be classified as a third (equidistant) PARV4 genotype. The availability of a further near-complete genome sequence of this novel genotype facilitated identification of conserved novel open reading frames embedded in the ORF2 coding sequence; one encoded a putative protein with identifiable homology to SAT proteins of members of the genus Parovirus.

The GenBank/EMBL/DDBJ accession numbers for the NG_OR and CD_BM sequences determined in this study are EU252026, EU254716 and EU874248–EU874250.

A supplementary figure and two supplementary tables are available with the online version of this paper.
Europe and other Western countries. The rapid changes in circulating virus populations and the extremely limited sequence diversity, particularly of PARV4 genotype 1, suggests that PARV4 has recently emerged and spread in parenterally exposed populations in the USA and Europe (Manning et al., 2007; Brown & Simmonds, 2007).

To explore the genetic variability of PARV4 further and extend knowledge of PARV4 variants circulating outside Europe and the USA, we analysed PARV4 variants infecting study subjects with AIDS originating from sub-Saharan Africa. A total of two samples from 13 individuals screened [using 1 μg aliquots of DNA from samples of bone marrow (n=6) or lymph node (n=7)] were positive for PARV4 DNA sequences by PCR, using primers from the ORF1 region as described (Longhi et al., 2007), and a second set of primers from the ORF2 (structural) gene region (see Supplementary Table S1, available in JGV Online). Viral loads were semi-quantified by titration of input DNA using triplicate log10 dilution steps as previously described (Manning et al., 2007). The sample NG_OR contained 620 DNA copies per 10^6 cells, compared with only 7.3 copies per 10^6 cells in the sample CD_BM.

The first patient (NG_OR) was a 38-year-old heterosexual woman from Nigeria with Pneumocystis jiroveci pneumonia, disseminated tuberculosis, pancytopenia and oesophageal candidosis. The patient’s CD4^+ T cell count was 91 cells μl^-1, the HIV-1 viral load was 21 118 copies ml^-1 and serology tests for HCV and hepatitis B virus (HBV) were negative. The second individual (CD_BM) was a 46-year-old heterosexual Congolese man with acute respiratory distress syndrome and bilateral parotid gland enlargement. A diagnosis of Pneumocystis jiroveci pneumonia and cystic lesions of the parotid duct associated with reactive lymphadenopathy and germinal centre depletion was made but the patient died shortly thereafter. His CD4^+ T cell count was 13 cells μl^-1 and the HIV viral load was 18 166 copies ml^-1. Serology for HCV infection was negative. While the first individual had been resident in Italy for the preceding 6 years, the other patient had recently arrived from Africa; in both cases, HIV infections had been acquired in Africa.

Amplified DNA from both regions was sequenced and compared with available corresponding sequences from currently described genotype 1 and 2 strains. Sequences from both sub-Saharan African study subjects grouped separately from those of genotype 1 and 2 to form a third bootstrap-supported group (Fig. 1a). To characterize this new putative genotype of PARV4 further, an almost complete genome sequence of NG_OR was obtained through amplification with a set of eight overlapping nested primer sets spanning the PARV genome (see Supplementary Table S2). Primers used for complete genome sequences failed to amplify PARV4 sequences from the sample CD_BM, a finding potentially attributable to their reduced sensitivity compared with the single-copy detection capability of the primers used for initial screening (Manning et al., 2007). The near-complete genome sequence of NG_OR contained intact ORF1 and ORF2 coding sequences, with no insertions or deletions relative to genotype 1 and 2 sequences. Between ORF1 and ORF2

| Table 1. Comparison of nucleotide and translated amino acid sequences of NG_OR and other PARV4 genotypes. |

<table>
<thead>
<tr>
<th>ORF1 (positions 282–2271)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NG_OR</td>
</tr>
<tr>
<td>2.9</td>
</tr>
<tr>
<td>Genotype 1</td>
</tr>
<tr>
<td>Genotype 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ORF2 (positions 2278–5519)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NG_OR</td>
</tr>
<tr>
<td>1.6</td>
</tr>
<tr>
<td>Genotype 1</td>
</tr>
<tr>
<td>Genotype 2</td>
</tr>
</tbody>
</table>

Mean pairwise nucleotide (below the diagonal) and translated amino acid (above the diagonal) sequence differences (%) in ORF1 (positions 282–2271, numbered based on NC_007018) and ORF2 (positions 2278–5519) between NG_OR and other PARV4 genotypes.
were 107 non-coding nucleotides, compared with 103 and 106 for PARV4 genotypes 1 and 2, respectively. ORF1 and ORF2 sequences in NG_OR were similarly divergent from genotype 1 and 2 sequences (mean nucleotide pairwise distances of 8.0 and 8.3 % for ORF1 and 6.2 and 6.9 % for ORF2) as the genotypes 1 and 2 were from each other (9.1 and 6.6 %; Table 1). The three variants were similarly equidistant when translated amino acid sequences were compared (Table 1). Phylogenetic analysis of near-complete genome sequences confirmed the status of NG_OR as a genotype separate and equally divergent from genotypes 1 and 2 (Fig. 1b).

To investigate whether there was any recombination between NG_OR and other PARV4 genotypes, mean pairwise nucleotide distances between each group were calculated using a sliding window of 300 bases incrementing by 20 bases across the genome (Fig. 2). Each pairwise comparison of genotypes (genotype 1 versus genotype 2, genotype 1 versus NG_OR and genotype 2 versus NG_OR)
showed similar distributions of variability across the genome, providing no evidence for recombination between sequences.

Each genotype comparison showed the same pronounced suppression of variability in ORF2 between positions 2800 and 3500, an occurrence that accounted for the lower overall degree of divergence of ORF2 compared with ORF1 (Table 1). Intriguingly, the region of suppression in variability contained two sequential alternative reading frames (ARFs) of 67 and 86 amino acids [positions 2937–3137 (ARF1) and 3261–3518 (ARF2), numbered according to their position in the prototype PARV4 sequence, NC_007018] that were conserved in all three PARV4 genotypes (Fig. 2). The 5′ ORF encoded a protein of similar length and genome position with identifiable sequence similarity to the small alternatively translated (SAT) protein found in members of the genus Parvovirus [porcine parvovirus (PPV), canine parvovirus (CPV), minute virus of mice (MVM)] (Zadori et al., 2005) (Fig. 2). The homologue of SAT in PARV4 showed 24–29% amino acid sequence similarity to PPV, CPV and MVM, compared with 31–38% within the genus Parvovirus. This contrasts with only 11–18% amino acid sequence identity between PARV4 and parvoviruses in the conventional ORF2 reading frame (i.e. an absence of detectable sequence similarity even after CLUSTAL realignment; data not shown). The 3′ reading frame encoded a protein with no identifiable homology to any other sequence on BLAST searching. A requirement to maintain coding for these two overlapping proteins probably underlies the observed suppression in variability in this genome region.

These findings demonstrate that PARV4 has a wider distribution outside the USA and Europe and that at least one other genotype circulates in sub-Saharan Africa. Although we understand very little about the true exposure frequency of individuals to PARV4 or the nature of the infection it establishes, it is intriguing that PARV4 may also be associated with HIV-1 infection in sub-Saharan Africa in an entirely different epidemiological context from Europe and the USA. The risk factor identified for HIV infection in the study subjects was unprotected heterosexual exposure; neither subject had histories of any intravenous drug use, blood or blood product transfusion, invasive surgery or other identifiable blood-borne route of virus exposure. A lack of parenteral exposure is supported by negative serological tests for HCV in both subjects. Our understanding of the origins and spread of PARV4 would be greatly enhanced by further investigation of differences in the epidemiology, genetic diversity, association with HIV infections and transmission routes of PARV4 between Western countries and Africa which have been highlighted by these initial findings.

The availability of increasing amounts of sequence information is of further value in understanding the nature of PARV4 evolution and the constraints imposed on sequence change. The evidence obtained for the existence of small ARFs in ORF2 justifies further investigation of their potential expression and functional roles to complement ongoing studies of expression of similar small proteins, such as SAT, in other parvoviruses.

References