Journal of General Virology (2016), **97**, 1888–1898 **DOI** 10.1099/jgv.0.000497

Analysis of the full genome of human group C rotaviruses reveals lineage diversification and reassortment

Maria Cristina Medici,[1] Fabio Tummolo,[1] Vito Martella,[2] Maria Cristina Arcangeletti,[1] Flora De Conto,[1] Carlo Chezzi,[1] Enikő Fehér,[3] Szilvia Marton,[3] Adriana Calderaro[1]† and Krisztián Bányai[3]†

Correspondence
Maria Cristina Medici
mariacristina.medici@unipr.it

[1]Unit of Microbiology and Virology, Department of Clinical and Experimental Medicine, University of Parma, Parma, Italy

[2]Department of Veterinary Medicine, University Aldo Moro of Bari, Valenzano, Italy

[3]Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary

Group C rotaviruses (RVC) are enteric pathogens of humans and animals. Whole-genome sequences are available only for few RVCs, leaving gaps in our knowledge about their genetic diversity. We determined the full-length genome sequence of two human RVCs (PR2593/2004 and PR713/2012), detected in Italy from hospital-based surveillance for rotavirus infection in 2004 and 2012. In the 11 RNA genomic segments, the two Italian RVCs segregated within separate intra-genotypic lineages showed variation ranging from 1.9 % (VP6) to 15.9 % (VP3) at the nucleotide level. Comprehensive analysis of human RVC sequences available in the databases allowed us to reveal the existence of at least two major genome configurations, defined as type I and type II. Human RVCs of type I were all associated with the M3 VP3 genotype, including the Italian strain PR2593/2004. Conversely, human RVCs of type II were all associated with the M2 VP3 genotype, including the Italian strain PR713/2012. Reassortant RVC strains between these major genome configurations were identified. Although only a few full-genome sequences of human RVCs, mostly of Asian origin, are available, the analysis of human RVC sequences retrieved from the databases indicates that at least two intra-genotypic RVC lineages circulate in European countries. Gathering more sequence data is necessary to develop a standardized genotype and intra-genotypic lineage classification system useful for epidemiological investigations and avoiding confusion in the literature.

Received 7 March 2016
Accepted 3 May 2016

INTRODUCTION

Rotaviruses (RVs) form a genus of the family *Reoviridae* and are major cause of acute gastroenteritis in humans and animals worldwide. RVs are classified into at least nine antigenically distinct groups or species (A–I) on the basis of a common group antigen, the inner capsid protein (VP6) and of sequence comparison (Matthijnssens *et al.*, 2012; Mihalov-Kovács *et al.*, 2015). Each member of the genus shares the capsid morphology and the unique genome structure with 11 segments of dsRNA, encoding six structural viral proteins (VP1–VP4, VP6 and VP7) and five or six non-structural proteins (NSP1–NSP5 and NSP6) (Estes & Greenberg, 2013).

Species A–C and H RVs are associated with acute gastroenteritis in humans and have remarkably different epidemiologic features and public health importance (Estes & Greenberg, 2013). Species A RVs (RVAs) have been identified mostly in infants and young children (Parashar *et al.*, 2003). Species B (RVBs) and H (RVHs) RVs are known to cause severe diarrhoea, primarily in adults (Rahman *et al.*, 2007; Estes & Greenberg, 2013). Species C RVs (RVCs) have been detected in either sporadic cases or outbreaks of gastroenteritis worldwide, and the overall disease burden and epidemiology of this group of RVs remain to be investigated (Adah *et al.*, 2002; Castello *et al.*, 2009; Medici *et al.*, 2009; Schnagl *et al.*, 2004).

†These authors contributed equally to this work.

The GenBank/EMBL/DDBJ accession numbers for the nucleotide sequences reported in this paper are KT206182–KT206203.

A genotyping classification scheme based on VP4 (P type) and VP7 (G type) has been proposed for RVCs (Jiang *et al.*, 1999; Martella *et al.*, 2007; Marthaler *et al.*, 2013), following the outlines of the full genome. All human RVCs share high conservation within the VP7 and VP4 genes and are classified as G4P[2]. The limited number of genomic sequence data on RVCs hinders the establishment of a proper classification system for complete genomic designation of RVC strains (Mawatari *et al.*, 2014; Moutelíková *et al.*, 2014), similar to the classification system developed for RVAs (Matthijnssens *et al.*, 2008). Implementing a classification system based on the 11 genome segments would be important to better understand the epidemiology of RVC infections (Baek *et al.*, 2013; Marton *et al.*, 2015; Yamamoto *et al.*, 2011; Zhirakovskaia *et al.*, 2016). Available genome sequence information suggests that at least two major human lineages can be distinguished based on backbone gene combinations. One lineage includes strains detected in various areas of the globe including the historic strain Bristol, whilst the other lineage includes a few strains detected during the past decade in Far East Asian countries (Yamamoto *et al.*, 2011).

In the area of Parma, Northern Italy, continual hospital-based surveillance for RV infection has been conducted uninterruptedly over the last 28 years. RVCs were detected sporadically (one case in 2004, one case in 2005 and one case in 2012). Sequence and phylogenetic analyses of the VP4, VP7 and VP6 genes of RVCs detected in 2004 and 2005 revealed that these viruses were highly related to each other (Medici *et al.*, 2009).

In order to obtain data on the genetic make-up of the Italian human RVCs and to understand if RVCs have changed in this period, full-length sequences of the 11 RNA segments were determined for the RVC strain detected in 2012, RVC/Human-wt/ITA/PR713/2012/G4P[2] (PR713/2012), and for the RVC strain detected in 2004, RVC/Human-wt/ITA/PR2593/2004G4P[2] (PR2593/2004). The sequences were analysed genetically and phylogenetically in comparison with sequences retrieved from the databases. The findings of this study contribute to knowledge of the genetic diversity and evolution of human RVCs.

RESULTS

Phylogenetic trees were built using the nucleotide sequences of the coding regions. Lineage and genotype constellations of the two Italian human RVC strains are provided in Table 1, while the phylogenetic trees are shown in Fig. 1.

Strains PR2593/2004 and PR713/2012 represented distinct genetic lineages, with 4.8 % nucleotide and 2.4 % amino acid variation in the VP7 gene, and 3.4 % nucleotide and 2.1 % amino acid sequence variation in the VP4 gene (Table 1). Within the other RNA segments, the rate of sequence variation ranged from 1.9 % (VP6) to 15.9 % (VP3) at nucleotide level, and from 0.3 % (VP6) to 13.7 % (VP3) at amino acid level.

In the phylogenetic tree of VP7, strains PR2593/2004 and PR713/2012 segregated into genotype G4 but formed two discrete intra-genotypic lineages (lineage I and II, respectively) (Fig. 1). Similarly, the VP4 genes of the Italian RVCs clustered into genotype P[2] but into separate intra-genotypic lineages. Likewise, in the VP1-VP2, VP6, and NSP1–NSP5 genes the two RVC strains were characterized as I2-R2-C2-A2-N2-T2-E2-H2, although they segregated into two distinct lineages in each genome segment. Finally, in the VP3 gene, strain PR2593/2004 was characterized as genotype M3, while strain PR713/2012 was designated as genotype M2. High bootstrap values were produced in the 11 phylogenetic trees, with the two major lineages segregating robustly in almost all gene segments.

Table 1. Genomic constellations and nucleotide and amino acid sequence differences (%) of individual gene segments among the two Italian RVCs, PR2593/2004 and PR713/2012

Gene segment	Genomic constellations		PR2593/2004 versus PR713/2012	
	PR2593/2004	PR713/2012	Nucleotide	Amino acid
VP7	G4	G4	4.8	2.4
VP4	P[2]	P[2]	3.4	2.1
VP6	I2	I2	1.9	0.3
VP1	R2	R2	5.7	2.8
VP2	C2	C2	5.1	1.6
VP3	M3	M2	15.9	13.7
NSP1	A2	A2	6.6	4.8
NSP2	N2	N2	4.8	2.9
NSP3	T2	T2	6.9	6
NSP4	E2	E2	4.6	6
NSP5	H2	H2	6.7	8

Strain nomenclature by Matthijnssens *et al.* (2008) is implemented using the genotype designations as recommended by Yamamoto *et al.* (2011).

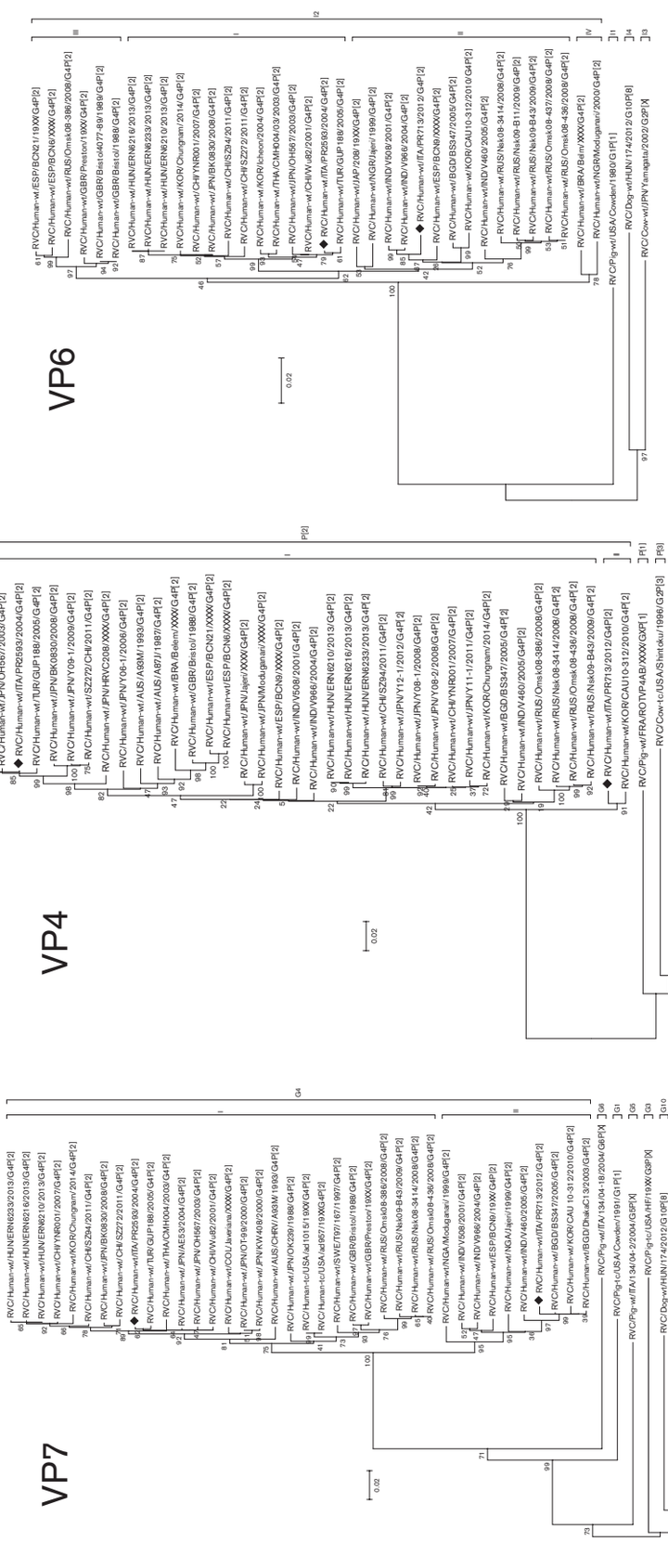
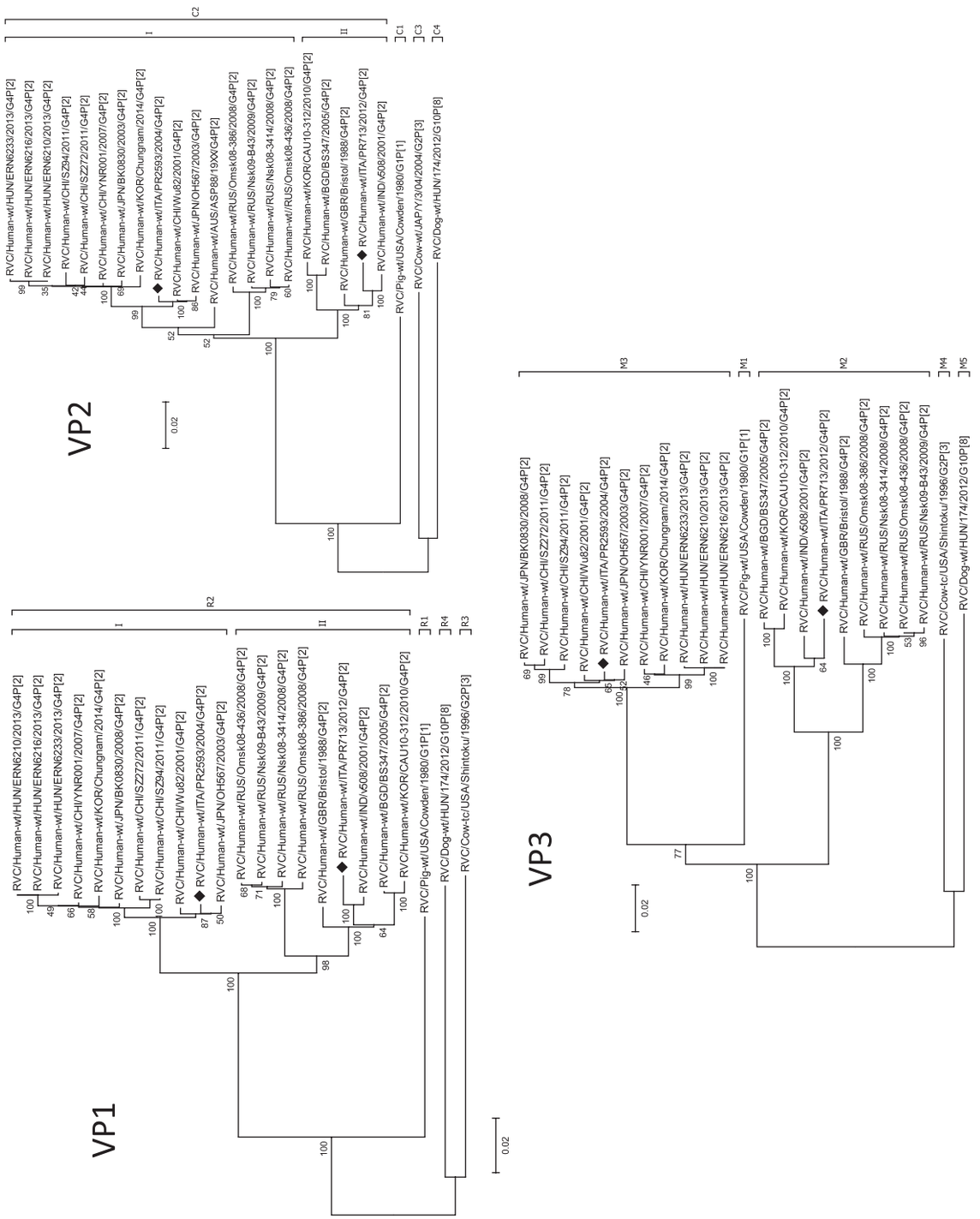


Fig. 1. Phylogenetic dendrograms based on the nucleotide sequences (ORF) of all 11 RVC gene segments. ◆ Italian human RVC strains. Trees were built using the maximum-likelihood method, and bootstrapped with 1000 repetitions. All the full-length human RVC sequences of each individual gene segment available in the NCBI databases were included in the phylogenetic analyses. The VP7 tree only shows a selection of 31 out of 68 human RVC sequences available in databases in order to eliminate redundancy and prepare a readable figure. Bootstrap values are indicated at nodes of branches. Scale bar indicating the number of nucleotide substitutions per site is shown for each tree. Strain designation by Matthijnssens *et al.* (2008).

Fig. 1. (cont.)


VP1
RVC/Human-wt/HUN/ERN6210/2013/G4P[2]
RVC/Human-wt/HUN/ERN6216/2013/G4P[2]
RVC/Human-wt/HUN/ERN6233/2013/G4P[2]
RVC/Human-wt/CHI/SZ94/2011/G4P[2]
RVC/Human-wt/CHI/SZ272/2011/G4P[2]
RVC/Human-wt/KOR/Chungnam/2014/G4P[2]
RVC/Human-wt/JPN/BK0830/2008/G4P[2]
RVC/Human-wt/CHI/SZ272/2011/G4P[2]
RVC/Human-wt/CHI/SZ94/2011/G4P[2]
RVC/Human-wt/CHI/Wu82/2001/G4P[2]
RVC/Human-wt/ITA/PR2593/2004/G4P[2]
RVC/Human-wt/JPN/OH567/2003/G4P[2]
RVC/Human-wt/RUS/Omsk08-436/2008/G4P[2]
RVC/Human-wt/RUS/Nsk09-B43/2009/G4P[2]
RVC/Human-wt/RUS/Nsk08-3414/2008/G4P[2]
RVC/Human-wt/RUS/Omsk08-386/2008/G4P[2]
RVC/Human-wt/GBR/Bristol/1988/G4P[2]
RVC/Human-wt/ITA/PR713/2012/G4P[2]
RVC/Human-wt/IND/k508/2001/G4P[2]
RVC/Human-wt/BGD/BS3447/2005/G4P[2]
RVC/Human-wt/KOR/CAU10-312/2010/G4P[2]
RVC/Pig-wt/USA/Cowden/1980/G1P[1]
RVC/Dog-wt/HUN/174/2012/G10P[8]
RVC/Cow-tc/USA/Shintoku/1996/G2P[3]
I
II
R1
R4
R3
R2
0.02
VP2
RVC/Human-wt/HUN/ERN6233/2013/G4P[2]
RVC/Human-wt/HUN/ERN6216/2013/G4P[2]
RVC/Human-wt/HUN/ERN6210/2013/G4P[2]
RVC/Human-wt/CHI/SZ94/2011/G4P[2]
RVC/Human-wt/CHI/SZ272/2011/G4P[2]
RVC/Human-wt/CHI/YNR001/2007/G4P[2]
RVC/Human-wt/JPN/BK0830/2003/G4P[2]
RVC/Human-wt/KOR/Chungnam/2014/G4P[2]
RVC/Human-wt/ITA/PR2593/2004/G4P[2]
RVC/Human-wt/CHI/Wu82/2001/G4P[2]
RVC/Human-wt/JPN/OH567/2003/G4P[2]
RVC/Human-wt/AUS/ASP88/19XX/G4P[2]
RVC/Human-wt/RUS/Omsk08-386/2008/G4P[2]
RVC/Human-wt/RUS/Nsk09-B43/2009/G4P[2]
RVC/Human-wt/RUS/Nsk08-3414/2008/G4P[2]
RVC/Human-wt//RUS/Omsk08-436/2008/G4P[2]
RVC/Human-wt/KOR/CAU10-312/2010/G4P[2]
RVC/Human-wt/BGD/BS3447/2005/G4P[2]
RVC/Human-wt/GBR/Bristol/1988/G4P[2]
RVC/Human-wt/ITA/PR713/2012/G4P[2]
RVC/Human-wt/IND/k508/2001/G4P[2]
RVC/Pig-wt/USA/Cowden/1980/G1P[1]
RVC/Cow-wt/JAP/Y/3/04/2004/G2P[3]
RVC/Dog-wt/HUN/174/2012/G10P[8]
I
II
C1
C3
C4
C2
0.02
VP3
RVC/Human-wt/JPN/BK0830/2008/G4P[2]
RVC/Human-wt/CHI/SZ272/2011/G4P[2]
RVC/Human-wt/CHI/SZ94/2011/G4P[2]
RVC/Human-wt/CHI/Wu82/2001/G4P[2]
RVC/Human-wt/ITA/PR2593/2004/G4P[2]
RVC/Human-wt/JPN/OH567/2003/G4P[2]
RVC/Human-wt/CHI/YNR001/2007/G4P[2]
RVC/Human-wt/KOR/Chungnam/2014/G4P[2]
RVC/Human-wt/HUN/ERN6233/2013/G4P[2]
RVC/Human-wt/HUN/ERN6210/2013/G4P[2]
RVC/Human-wt/HUN/ERN6216/2013/G4P[2]
RVC/Pig-wt/USA/Cowden/1980/G1P[1]
RVC/Human-wt/BGD/BS3447/2005/G4P[2]
RVC/Human-wt/KOR/CAU10-312/2010/G4P[2]
RVC/Human-wt/IND/k508/2001/G4P[2]
RVC/Human-wt/ITA/PR713/2012/G4P[2]
RVC/Human-wt/GBR/Bristol/1988/G4P[2]
RVC/Human-wt/RUS/Omsk08-386/2008/G4P[2]
RVC/Human-wt/RUS/Nsk08-3414/2008/G4P[2]
RVC/Human-wt/RUS/Omsk08-436/2008/G4P[2]
RVC/Human-wt/RUS/Nsk09-B43/2009/G4P[2]
RVC/Cow-tc/USA/Shintoku/1996/G2P[3]
RVC/Dog-wt/HUN/174/2012/G10P[8]
M3
M1
M2
M4
M5
0.02

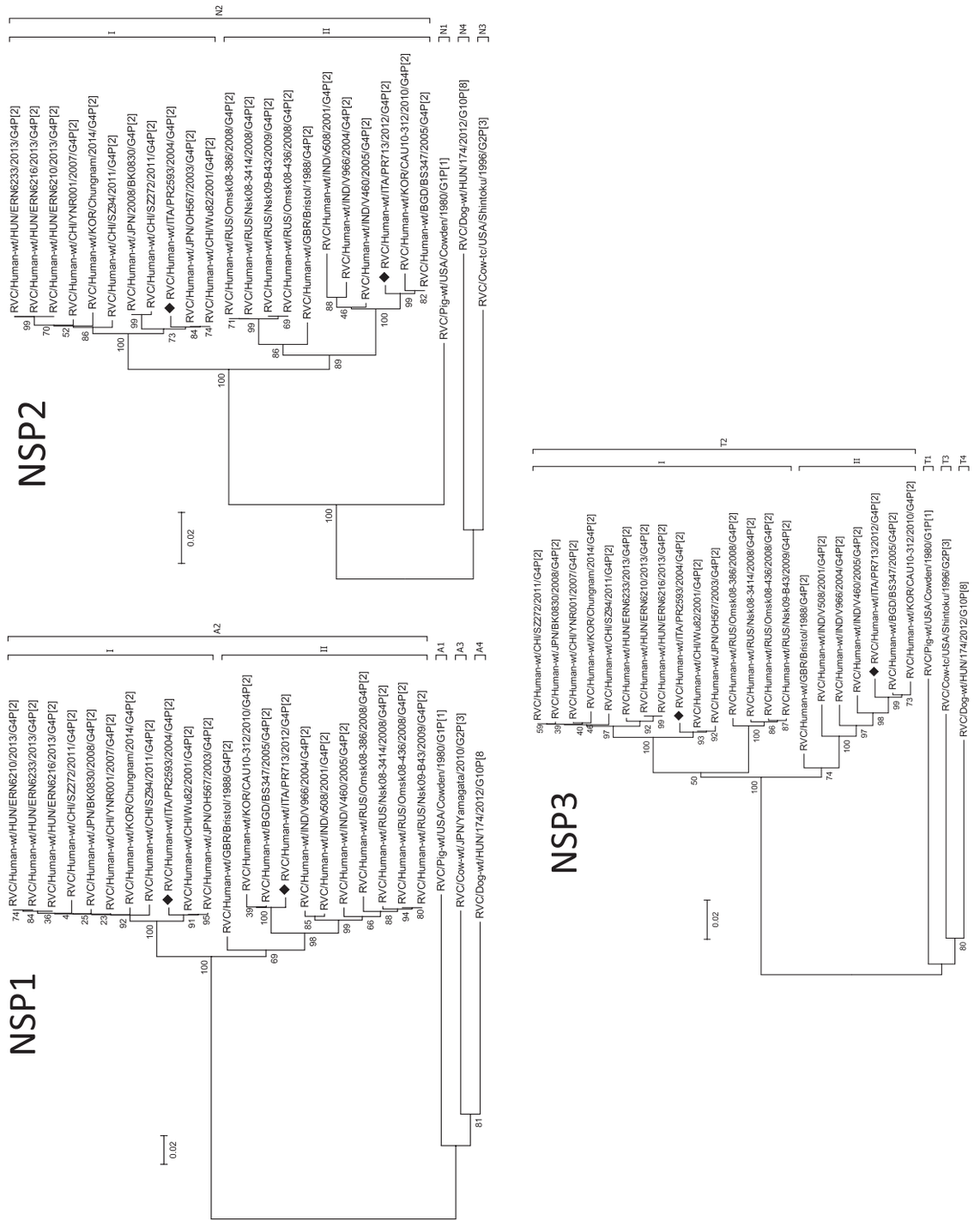

NSP1
NSP2
NSP3
0.02
RVC/Human-wt/HUN/ERN6210/2013/G4P[2]
RVC/Human-wt/HUN/ERN6233/2013/G4P[2]
RVC/Human-wt/HUN/ERN6216/2013/G4P[2]
RVC/Human-wt/CHI/SZ272/2011/G4P[2]
RVC/Human-wt/JPN/BK0830/2008/G4P[2]
RVC/Human-wt/KOR/Chungnam/2014/G4P[2]
RVC/Human-wt/CHI/YNR001/2007/G4P[2]
RVC/Human-wt/CHI/SZ394/2011/G4P[2]
RVC/Human-wt/ITA/PR2593/2004/G4P[2]
RVC/Human-wt/CHI/Wu82/2001/G4P[2]
RVC/Human-wt/JPN/OH567/2003/G4P[2]
RVC/Human-wt/GBR/Bristol/1988/G4P[2]
RVC/Human-wt/KOR/CAU10-312/2010/G4P[2]
RVC/Human-wt/BGD/BS347/2005/G4P[2]
RVC/Human-wt/ITA/PR2713/2012/G4P[2]
RVC/Human-wt/IND/V966/2004/G4P[2]
RVC/Human-wt/IND/V460/2005/G4P[2]
RVC/Human-wt/IND/V508/2001/G4P[2]
RVC/Human-wt/RUS/Nsk08-386/2008/G4P[2]
RVC/Human-wt/RUS/Nsk08-3414/2008/G4P[2]
RVC/Human-wt/RUS/Omsk08-436/2008/G4P[2]
RVC/Human-wt/RUS/Nsk09-B43/2009/G4P[2]
RVC/Pig-wt/USA/Cowden/1980/G1P[1]
RVC/Cow-wt/JPN/Yamagata/2010/G2P[3]
RVC/Dog-wt/HUN/174/2012/G10P[8
RVC/Cow-tc/USA/Shintoku/1996/G2P[3]
A1
A2
A3
A4
I
II
N1
N2
N3
N4
T1
T2
T3
T4

Fig. 1. (cont.)

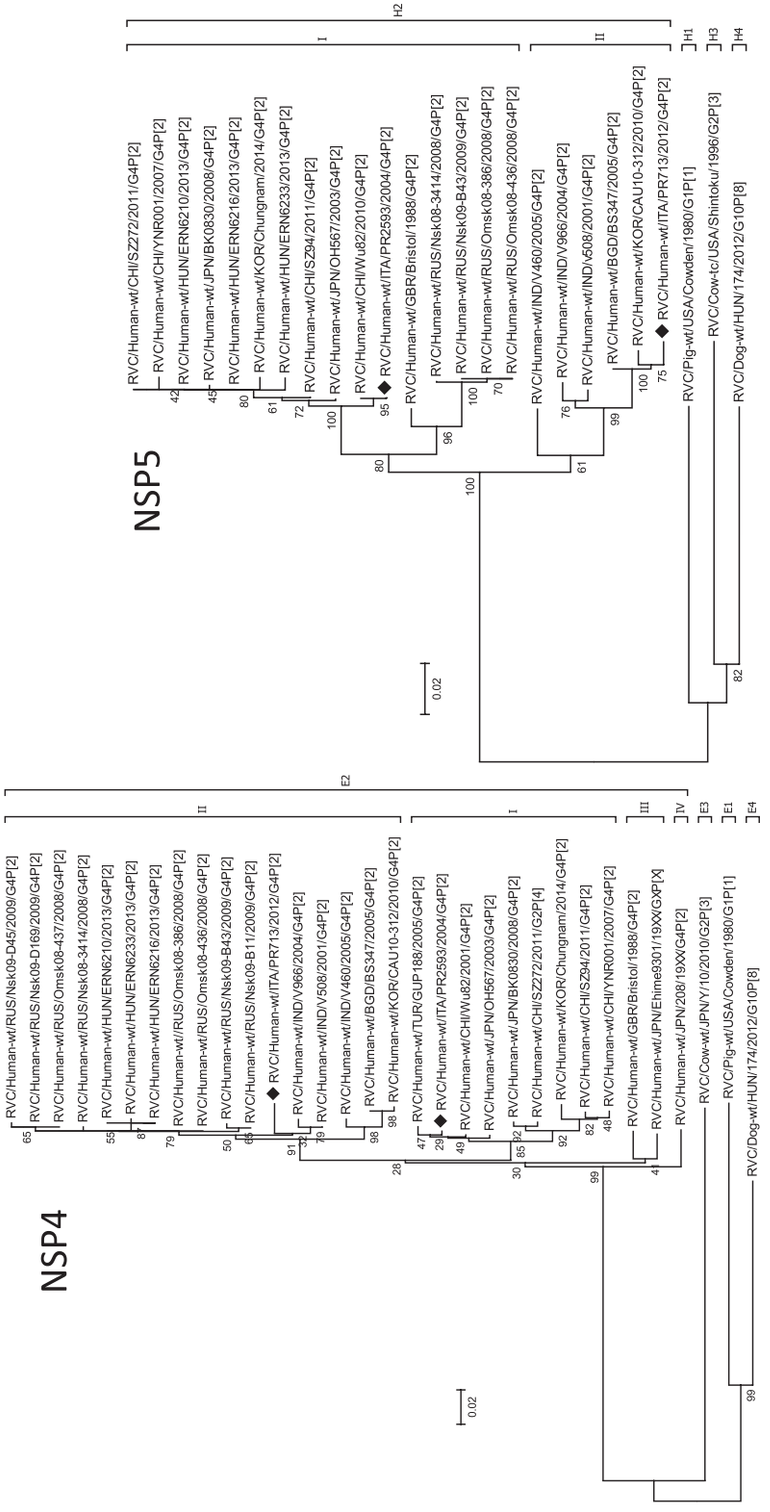

NSP4
RVC/Human-wt/RUS/Nsk09-D45/2009/G4P[2]
RVC/Human-wt/RUS/Nsk09-D169/2009/G4P[2]
RVC/Human-wt/RUS/Omsk08-437/2008/G4P[2]
RVC/Human-wt/RUS/Nsk08-3414/2008/G4P[2]
RVC/Human-wt/HUN/ERN6210/2013/G4P[2]
RVC/Human-wt/HUN/ERN6233/2013/G4P[2]
RVC/Human-wt/HUN/ERN6216/2013/G4P[2]
RVC/Human-wt//RUS/Omsk08-386/2008/G4P[2]
RVC/Human-wt//RUS/Omsk08-436/2008/G4P[2]
RVC/Human-wt/RUS/Nsk09-B43/2009/G4P[2]
RVC/Human-wt/RUS/Nsk09-B11/2009/G4P[2]
◆ RVC/Human-wt/ITA/PR713/2012/G4P[2]
RVC/Human-wt/IND/V966/2004/G4P[2]
RVC/Human-wt/IND/V508/2001/G4P[2]
RVC/Human-wt/IND/V460/2005/G4P[2]
RVC/Human-wt/BGD/BS3447/2005/G4P[2]
RVC/Human-wt/KOR/CAU10-312/2010/G4P[2]
RVC/Human-wt/TUR/GUP188/2005/G4P[2]
◆ RVC/Human-wt/ITA/PR2593/2004/G4P[2]
RVC/Human-wt/CHI/Wu82/2001/G4P[2]
RVC/Human-wt/JPN/OH5567/2003/G4P[2]
RVC/Human-wt/JPN/BK0830/2008/G4P[2]
RVC/Human-wt/CHI/SZ272/2011/G2P[4]
RVC/Human-wt/KOR/Chungnam/2014/G4P[2]
RVC/Human-wt/CHI/SZ94/2011/G4P[2]
RVC/Human-wt/CHI/YNR0001/2007/G4P[2]
RVC/Human-wt/GBR/Bristol/1988/G4P[2]
RVC/Human-wt/JPN/Ehime9301/19XX/GXP[X]
RVC/Human-wt/JPN/208/19XX/G4P[2]
RVC/Cow-wt/JPN/Y110/2010/G2P[3]
RVC/Pig-wt/USA/Cowden/1980/G1P[1]
RVC/Dog-wt/HUN/174/2012/G10P[8]
0.02
II
I
III
IV
E2
E3
E1
E4
NSP5
RVC/Human-wt/CHI/SZ272/2011/G4P[2]
RVC/Human-wt/CHI/YNR0001/2007/G4P[2]
RVC/Human-wt/HUN/ERN6210/2013/G4P[2]
RVC/Human-wt/JPN/BK0830/2008/G4P[2]
RVC/Human-wt/HUN/ERN6216/2013/G4P[2]
RVC/Human-wt/KOR/Chungnam/2014/G4P[2]
RVC/Human-wt/HUN/ERN6233/2013/G4P[2]
RVC/Human-wt/CHI/SZ94/2011/G4P[2]
RVC/Human-wt/JPN/OH5567/2003/G4P[2]
RVC/Human-wt/CHI/Wu82/2010/G4P[2]
◆ RVC/Human-wt/ITA/PR2593/2004/G4P[2]
RVC/Human-wt/GBR/Bristol/1988/G4P[2]
RVC/Human-wt/RUS/Nsk08-3414/2008/G4P[2]
RVC/Human-wt/RUS/Nsk09-B43/2009/G4P[2]
RVC/Human-wt/RUS/Omsk08-386/2008/G4P[2]
RVC/Human-wt/RUS/Omsk08-436/2008/G4P[2]
RVC/Human-wt/IND/V460/2005/G4P[2]
RVC/Human-wt/IND/V966/2004/G4P[2]
RVC/Human-wt/IND/V508/2001/G4P[2]
RVC/Human-wt/BGD/BS3447/2005/G4P[2]
RVC/Human-wt/KOR/CAU10-312/2010/G4P[2]
◆ RVC/Human-wt/ITA/PR713/2012/G4P[2]
RVC/Pig-wt/USA/Cowden/1980/G1P[1]
RVC/Cow-tc/USA/Shintoku/1996/G2P[3]
RVC/Dog-wt/HUN/174/2012/G10P[8]
0.02
I
II
H2
H1
H3
H4

Fig. 1. (cont.)

Table 2. Number of human RVC sequences available in the databases and nucleotide and amino acid sequence identities (%) of individual gene segments among the human RVCs

Gene segment	Human RVC sequences available in the databases		Sequence identities between full-length sequences	
	Total sequences	Full-length sequences	Nucleotide	Amino acid
VP7	91	71	94.8–99.6	95.5–100
VP4	73	41	94.8–99.8	96.1–100
VP6	55	36	96.2–99.6	98.2–100
VP1	20	20	94.0–99.9	96.7–100
VP2	21	21	94.2–99.6	97.5–100
VP3	20	20	83.7–99.4	85.3–100
NSP1	22	22	93.1–99.9	93.1–100
NSP2	22	22	94.2–99.6	95.2–100
NSP3	22	22	92.5–99.9	94.0–100
NSP4	47	29	94.6–99.8	92.0–100
NSP5	22	22	92.2–99.9	90.6–100

Interrogation of the databases was terminated in December 2015.

On the whole, considering also the two RVC strains analysed in this study, a total of 91 VP7, 73 VP4, 55 VP6, 20 VP1, 21 VP2, 20 VP3, 22 NSP1, 22 NSP2, 22 NSP3, 47 NSP4 and 22 NSP5 sequences of human RVC strains, detected over nearly a three-decade period, are available in the databases, but full-length genome sequence is available only for 20 RVC strains. The nucleotide and deduced amino acid sequence identity values of individual gene segments among the human RVCs are shown in Table 2. In all the RNA segments, the human RVC strains showed $\geq$92.2 % nucleotide and $\geq$90.6 % amino acid sequence conservation. The highest sequence conservation was found in the VP6 gene (96.2–99.6 % nucleotide and 98.2–100 % amino acid identity). In contrast, within the VP3 gene, sequence identity was as low as 83.7 % and 85.3 % at nucleotide and amino acid level, respectively.

Throughout the 11 gene segments, the human RVCs had a monophyletic origin and exhibited a high nucleotide difference, 13 % (NSP3) to 31 % (NSP1), with animals RVC strains available in the NCBI databases.

The limited number of sequences deposited in the databases for gene segments other than VP7 and VP4 is insufficient to provide clear insights into the mechanisms driving the evolution of human RVCs, and full genome sequence data are available mostly for Asian RVC strains. By comparison with the human RVC sequences retrieved from the databases, the topology of phylogenetic trees based on the 11 genome segments was substantially similar, allowing the differentiation of at least two major intra-genotypic lineages. Intra-genotypic lineage cut-off values at nucleotide level calculated for individual gene segments ranged from 2.5 % (VP7 and VP4) to 5.5 % (NSP1) (Table 3).

The two human RVC lineages were found to circulate also in other European countries. In particular, the Italian strain PR2593/2004 clustered together with reassortant human RVCs detected in Hungary in 2013 (strains ERN6210, ERN6216, and ERN6233) and with other human RVCs detected in Far East Asian countries (China, strains YNR001 and Wu82; Japan, strains BK0830 and OH567) in the last decade (lineage I). Conversely, strain PR713/2012 segregated closely with human RVCs detected in South Asia (India, strain v508; Bangladesh, strain BS347) in the early 2000s and with a Korean RVC (strain CAU10-312) detected in 2010 (lineage II). On the basis of whole-genome sequences, strain PR2593/2004 showed 98.4–99.3 % nucleotide and 97.7–100 % amino acid identity with three recent Hungarian RVCs, whilst strain PR713/2012, displayed 83.9–97 % nucleotide and 85.9–99.7 % amino acid identity, although it circulated during the same time as the Hungarian RVCs.

Considering the dataset of the 20 human RVCs with full-length genome sequences, the distinction into two lineages was consistently observed for most strains (Table 4). Excluding the three Hungarian RVCs with reassortment affecting the NSP4 gene, reassortment events were found in South Asia RVCs, such as strains BS347 (Bangladesh) and v508 (India) with all segments of lineage II except the VP4 (lineage I) gene. Reassortment events were observed also for Russian RVCs that showed a mixed I/II (strains Omsk-08-436; Nsk-08-3414; Nsk-09-B43) or I/II/III (strain Omsk-08-386) genetic backbone. Likewise, the historic strain Bristol detected in the United Kingdom in 1988 showed a mixed lineage I/II/III genetic backbone.

DISCUSSION

Molecular analysis of the whole genome allowed designation of the genotype constellation of two Italian human

Table 3. Intra-genotypic lineage nucleotide cut-off for individual RVC gene segments

Gene segments	Lineage cut-off (%)
VP7	2.5
VP4	2.5
VP6	2.5
VP1	5.1
VP2	4.3
VP3	–
NSP1	5.5
NSP2	3.8
NSP3	4.1
NSP4	3.8
NSP5	4.7

For VP3 gene, no intra-genotypic lineage was observed.

RVC strains as G4P[2]-I2-R2-C2-M2/M3-A2-N2-T2-E2-H2. This genome backbone is strictly conserved across all the human RVCs characterized thus far. This has led to the hypothesis that human RVCs, unlike human RVAs, may have undergone a recent bottleneck event in their evolution (Martella *et al.*, 2010). Alternatively, RVCs would be less prone to inter-species transmission than RVAs, thus limiting the extent of their genetic/antigenic diversification in the human host (Bányai *et al.*, 2006; Martella *et al.*, 2010). Also, in this study, evidence for zoonotic transmission of RVC strains was not observed, by comparing human and animal RVC sequences available in the NCBI databases. However, it should be considered that there is a limited number of sequences in the databases to confirm this hypothesis and that the RVC sequences released in databases are not representative of global RVC isolates. A possible zoonotic role of animal RVCs has been anecdotally suggested by both virological and serological investigations (Gabbay *et al.*, 2008; Iturriza-Gómara *et al.*, 2004), and therefore the possibility that animal RVC strains have acted as donors of genetic material for human RVCs can not be ruled out (Martella *et al.*, 2010).

Comprehensive analysis of the whole data set, i.e. 415 gene sequences of human RVCs available in the databases, allowed us to obtain hints at the evolution of these viruses and to characterize in detail the Italian RVC strains. Lineage distinction and nomenclature were only intended for the limited purposes of this study, as a consensual, largely adopted classification should necessarily rely on the analysis of a larger data set. Only 20 complete genome sequences were available at the time of database interrogation. Sequence analysis of RVC strains for which a complete full-genome sequence is available, revealed the existence of at least two major genome configurations, herewith defined as type I and type II. Human RVCs of type I had genome segments of lineage I and were all associated with the M3 VP3 genotype, including the Italian strain PR2593/2004. Conversely, human RVCs of type II had genes of lineage II and were all associated with the M2 VP3 genotype, including the Italian strain PR713/2012. With some exceptions, due to inter-lineage reassortment in the VP4, VP6 and NSP4 (Baek *et al.*, 2013; Marton *et al.*, 2015; Yamamoto *et al.*, 2011), the type I and type II backbones in human RVCs appeared well conserved, and the two Italian RVC strains were found to be representative members of the two major genome configurations. The origin of VP3 genetic diversity in human RVCs is unclear. RVCs with the M3 genotype have been circulating for at least 15 years, after their first identification in China in 2001 (Wang *et al.*, 2007). Whether this VP3 genotype was derived from a past inter-species reassortment event from a yet unidentified animal host remains to be investigated. Our findings, together with the data gathered recently in Eastern Europe (Marton *et al.*, 2015), indicate that the appearance in Europe of human RVCs with the M3 genotype dates back to 2004, at least.

Intriguingly, in our analysis, the prototype human RVC strain Bristol, detected in 1988, displayed a puzzling genome configuration, with three genes of the type I, six genes (including the M2 VP3) of the type II and two genes of type III lineage. Therefore, the RVC genome configurations currently circulating worldwide have genes derived from the reference strain Bristol. This clearly suggests that the genome configuration of human RVCs has evolved over the years due to mechanisms of genetic drift, intermingled with intra-typic reassortment, although, due to the limited genome sequence data, it is impossible to determine the exact temporal order of these events. It is possible that this simple scenario depicted in our analysis will be challenged and revised, as more sequence data will be available for human RVCs. Nonetheless, the observed genetic heterogeneity reflects the results of processes of selection and diversification acting on a global scale. Stringent calculations with more sequence data will be necessary to develop a standardized intra-genotypic classification system.

As full-genome sequence data are available only for a few RVC strains, it is possible that the RVC genetic variability is currently underestimated. Whole-genome sequencing has become widely adopted in the research of RVA, where it offers a platform to assess more in-depth RVA diversity and evolution (Bányai *et al.*, 2010; Matthijnssens *et al.*, 2011). Likewise, accumulation of whole-genome sequence data for RVCs will provide new insights into the mechanisms driving the evolution and affecting the epidemiology of this neglected group of RVs in humans.

METHODS

Samples and laboratory diagnosis. The RVC strain PR2593/2004 was detected in the stool of a 2-year-old child in 2004, while strain PR713/2012 was identified from a 21-year-old male individual in 2012. The patients were admitted with acute gastroenteritis at the University Hospital of Parma, Italy, during the RV season (December–April).

RVCs were detected from 10 % (w/w) phosphate-buffered saline faecal suspension by electron microscopy and by silver staining after PAGE of viral dsRNA extracted by phenol–chloroform (Medici *et al.*, 2004) for

Table 4. Intra-genotypic lineages of individual segments of RVC strains with full-length genome sequence

Strain	Country	Year	Genes lineage designation										
			VP7	VP4	VP6	VP1	VP2	VP3*	NSP1	NSP2	NSP3	NSP4	NSP5
RVC/Human-wt/CHI/Wu82/2001/G4P[2]	China	2001	**I**	**I**	**I**	**I**	**I**	M3	**I**	**I**	**I**	**I**	**I**
RVC/Human-wt/JPN/OH567/2003/G4P[2]	Japan	2003	**I**	**I**	**I**	**I**	**I**	M3	**I**	**I**	**I**	**I**	**I**
RVC/Human-wt/ITA/PR2593/2004/G4P[2]	**Italy**	**2004**	**I**	**I**	**I**	**I**	**I**	M3	**I**	**I**	**I**	**I**	**I**
RVC/Human-wt/CHI/YNR001/2007/G4P[2]	China	2007	**I**	**I**	**I**	**I**	**I**	M3	**I**	**I**	**I**	**I**	**I**
RVC/Human-wt/JPN/BK0830/2008/G4P[2]	Japan	2008	**I**	**I**	**I**	**I**	**I**	M3	**I**	**I**	**I**	**I**	**I**
RVC/Human-wt/CHI/SZ272/2011/G4P[2]	China	2011	**I**	**I**	**I**	**I**	**I**	M3	**I**	**I**	**I**	**I**	**I**
RVC/Human-wt/CHI/SZ94/2011/G4P[2]	China	2011	**I**	**I**	**I**	**I**	**I**	M3	**I**	**I**	**I**	**I**	**I**
RVC/Human-wt/HUN/ERN6233/2013/G4P[2]	Hungary	2013	**I**	**I**	**I**	**I**	**I**	M3	**I**	**I**	**I**	II	**I**
RVC/Human-wt/HUN/ERN6210/2013/G4P[2]	Hungary	2013	**I**	**I**	**I**	**I**	**I**	M3	**I**	**I**	**I**	II	**I**
RVC/Human-wt/HUN/ERN6216/2013/G4P[2]	Hungary	2013	**I**	**I**	**I**	**I**	**I**	M3	**I**	**I**	**I**	II	**I**
RVC/Human-wt/KOR/Chungnam/2014/G4P[2]	South Korea	2014	**I**	**I**	**I**	**I**	**I**	M3	**I**	**I**	**I**	**I**	**I**
RVC/Human-wt/IND/v508/2001/G4P[2]	India	2001	II	I	II	II	II	M2	II	II	II	II	II
RVC/Human-wt/BGD/BS347/2005/G4P[2]	Bangladesh	2005	II	I	II	II	II	M2	II	II	II	II	II
RVC/Human-wt/KOR/CAU10-312/2010/G4P[2]	Korea	2010	II	II	II	II	II	M2	II	II	II	II	II
RVC/Human-wt/ITA/PR713/2012/G4P[2]	**Italy**	**2012**	II	II	II	II	II	M2	II	II	II	II	II
RVC/Human-wt/RUS/Omsk-08-386/2008/G4P[2]	Russia	2008	I	I	III	II	I	M2	II	II	I	II	I
RVC/Human-wt/RUS/Omsk-08-436/2008/G4P[2]	Russia	2008	I	I	II	II	I	M2	II	II	I	II	I
RVC/Human-wt/RUS/Nsk-08-3414/2008/G4P[2]	Russia	2008	I	I	II	II	I	M2	II	II	I	II	I
RVC/Human-wt/RUS/Nsk-09-B43/2009/G4P[2]	Russia	2009	I	I	II	II	I	M2	II	II	I	II	I
RVC/Human-wt/GBR/Bristol/1988/G4P[2]	United Kingdom	1988	I	I	III	II	II	M2	II	II	II	III	I

The Italian RVCs in this study are shown in bold type. All the strains were characterized as G4-P[2]-I2-R2-C2-M2/M3-A2-N2-T2-E2-H2.
*In VP3, the genotype is reported. Strain designation by Matthijnssens *et al.* (2011).

identification of the typical migration pattern (4–3–2–2 pattern). Stool specimens were stored at −80 °C until used.

Whole-genome sequencing. Viral RNA was extracted (Viral RNA Mini kit, QIAGEN) and then denatured at 97 °C for 5 min. Reverse transcription (RT) was performed using AMV reverse transcriptase (Promega) with random hexamer, tailed by a common PCR primer sequence (Djikeng *et al.*, 2008). PCR was carried out with Taq DNA polymerase (Thermo Scientific), applying the following thermal profile: initial denaturation (95 °C for 3 min) was followed by 40 cycles of amplification (95 °C for 30 s, 48 °C for 30 s, 72 °C for 2 min) and terminated at 72 °C for 8 min. Approximately 30 to 40 ng cDNA obtained by the random RT-PCR were subjected to enzymatic fragmentation and adapter ligation using the reagents supplied in the NEBNext Fast DNA Fragmentation & Library Prep Set for Ion Torrent kit (New England Biolabs) according to the manufacturer's instruction (Maan *et al.*, 2007). Barcoded adapters were retrieved from the Ion Xpress™ Barcode Adapters kit (Life Technologies). After size selection, emulsion PCR was carried out on a OneTouch v2 instrument according to the manufacturer's protocol (Life Technologies). Templated beads were enriched on an Ion

One Touch ES machine. The sequencing protocol recommended for the Ion PGM™ Sequencing 200 Kit v2 on a 316 v2 chip was strictly followed. Raw sequence data were processed using the CLC Genomics Workbench version 7 (http://www.clcbio.com/). In brief, quality trimming (quality limit=0.05) was adjusted to include reads with lengths of 40 nt and 180 nt. Trimmed sequence reads were mapped onto reference human RVC sequences downloaded from the GenBank. Relevant mapping parameters included (i) the minimum overlap with reference sequences (set to 20 nt) and (ii) the minimum identity along the overlap (80 %). To improve the mapping results, the following gap penalties were applied for the data set: mismatch cost=2, insertion cost=3, deletion cost=3. After visual inspection of the sequence alignments and remapping onto the obtained consensus sequence, a single consensus sequence was finalized for each genome segment for both strains. For strains PR713/2012 and PR2593/2004, respectively, 10121 of 54326 and 36170 of 62262 trimmed reads mapped onto the assembled genome sequences, resulting in sufficient average coverage for both strains (range for PR713/2012: 168X, NSP2 and 1016X, NSP5; range for PR2593/2004: 84X, VP6 and 595X, NSP5).

Phylogenetic analysis. Multiple sequence alignments and phylogenetic tree constructions were performed with MEGA6 (Tamura *et al.*, 2013), applying the maximum-likelihood method. The reliability of the phylogenetic trees was assessed by bootstrap re-sampling over 1000 replicates. Reference sequences were retrieved from the NCBI databases for all those humans RVC strains whose full genomes were available for comparison.

Genotype designation was assigned to all gene segments of RVC on the basis of previous literature (Jiang *et al.*, 1999; Martella *et al.*, 2007; Matthijnssens *et al.*, 2008; Baek *et al.*, 2013) and lineage numbers were allocated according to Yamamoto *et al.* (2011). Nucleotide intragenotypic cut-off values for each gene segment were calculated with Geneious v6.2 software package (Kearse *et al.*, 2012).

GenBank accession numbers (VP7, VP4, VP6, VP1–VP3, NSP1–NSP5) for each individual genomic segment of the PR2593/2004 and PR713/2012 strains are KT206182–KT206192 and KT206193–KT206203, respectively.

ACKNOWLEDGEMENTS

This study was partly supported by project 'Epidemiologia molecolare e studio dei meccanismi evolutivi di norovirus e rotavirus umani' granted by the University of Parma, Italy (Fondi di Ateneo 2014). K. B. was supported by the Momentum program and the Hungarian Scientific Research Program (OTKA, T100727).

REFERENCES

Adah, M. I ., Wade, A., Oseto, M., Kuzuya, M. & Taniguchi, K. (2002). Detection of human group C rotaviruses in Nigeria and sequence analysis of their genes encoding VP4, VP6, and VP7 proteins. *J Med Virol* **66**, 269–275.

Baek, I. H., Than, V. T., Kim, H., Lim, I. & Kim, W. (2013). Full genomic characterization of a group C rotavirus isolated from a child in South Korea. *J Med Virol* **85**, 1478–1484.

Bányai, K., Jiang, B., Bogdán, A., Horváth, B., Jakab, F., Meleg, E., Martella, V., Magyari, L., Melegh, B. & other authors (2006). Prevalence and molecular characterization of human group C rotaviruses in Hungary. *J Clin Virol* **37**, 317–322.

Bányai, K., Papp, H., Dandár, E., Molnár, P., Mihály, I., Van Ranst, M., Martella, V. & Matthijnssens, J. (2010). Whole-genome sequencing and phylogenetic analysis of a zoonotic human G8P[14] rotavirus strain. *Infect Genet Evol* **10**, 1140–1144.

Castello, A. A., Argüelles, M. H., Rota, R. P., Humphrey, C. D., Olthoff, A., Gentsch, J. R., Glass, R. I., Glikmann, G. & Jiang, B. (2009). Detection and characterization of group C rotavirus in Buenos Aires, Argentina, 1997–2003. *J Med Virol* **81**, 1109–1116.

Djikeng, A., Halpin, R., Kuzmickas, R., Depasse, J., Feldblyum, J., Sengamalay, N., Afonso, C., Zhang, X., Anderson, N. G. & other authors (2008). Viral genome sequencing by random priming methods. *BMC Genomics* **9**, 5.

Estes, M. K. & Greenberg, H. B. (2013). Rotaviruses. In *Fields Virology*, 6th edn, pp. 1347–1401. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott Williams & Wilkins.

Gabbay, Y. B., Borges, A. A., Oliveira, D. S., Linhares, A. C., Mascarenhas, J. D., Barardi, C. R., Simões, C. M., Wang, Y., Glass, R. I. & Jiang, B. (2008). Evidence for zoonotic transmission of group C rotaviruses among children in Belém, Brazil. *J Med Virol* **80**, 1666–1674.

Iturriza-Gómara, M., Clarke, I., Desselberger, U., Brown, D., Thomas, D. & Gray, J. (2004). Seroepidemiology of group C rotavirus infection in England and Wales. *Eur J Epidemiol* **19**, 589–595.

Jiang, B., Gentsch, J. R., Tsunemitsu, H., Saif, L. J. & Glass, R. I. (1999). Sequence analysis of the gene encoding VP4 of a bovine group C rotavirus: molecular evidence for a new P genotype. *Virus Genes* **19**, 85–88.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S. & other authors (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649.

Maan, S., Rao, S., Maan, N. S., Anthony, S. J., Attoui, H., Samuel, A. R. & Mertens, P. P. (2007). Rapid cDNA synthesis and sequencing techniques for the genetic study of bluetongue and other dsRNA viruses. *J Virol Methods* **143**, 132–139.

Martella, V., Bányai, K., Lorusso, E., Decaro, N., Bellacicco, A., Desario, C., Corrente, M., Greco, G., Moschidou, P. & other authors (2007). Genetic heterogeneity in the VP7 of group C rotaviruses. *Virology* **367**, 358–366.

Martella, V., Bányai, K., Matthijnssens, J., Buonavoglia, C. & Ciarlet, M. (2010). Zoonotic aspects of rotaviruses. *Vet Microbiol* **140**, 246–255.

Marthaler, D., Rossow, K., Culhane, M., Collins, J., Goyal, S., Ciarlet, M. & Matthijnssens, J. (2013). Identification, phylogenetic analysis and classification of porcine group C rotavirus VP7 sequences from the United States and Canada. *Virology* **446**, 189–198.

Marton, S., Deák, J., Dóró, R., Csata, T., Farkas, S. L., Martella, V. & Bányai, K. (2015). Reassortant human group C rotaviruses in Hungary. *Infect Genet Evol* **34**, 410–414.

Matthijnssens, J., Ciarlet, M., Heiman, E., Arijs, I., Delbeke, T., McDonald, S. M., Palombo, E. A., Iturriza-Gómara, M., Maes, P. & other authors (2008). Full genome-based classification of rotaviruses reveals a common origin between human Wa-like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. *J Virol* **82**, 3204–3219.

Matthijnssens, J., Ciarlet, M., McDonald, S. M., Attoui, H., Bányai, K., Brister, J. R., Buesa, J., Esona, M. D., Estes, M. K. & other authors (2011). Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). *Arch Virol* **156**, 1397–1413.

Matthijnssens, J., Otto, P. H., Ciarlet, M., Desselberger, U., Van Ranst, M. & Johne, R. (2012). VP6-sequence-based cutoff values as a criterion for rotavirus species demarcation. *Arch Virol* **157**, 1177–1182.

Mawatari, T., Hirano, K., Tsunemitsu, H. & Suzuki, T. (2014). Whole-genome analysis of bovine rotavirus species C isolates obtained in Yamagata, Japan, 2003–2010. *J Gen Virol* **95**, 1117–1125.

Medici, M. C., Martinelli, M., Arcangeletti, M. C., Pinardi, F., De Conto, F., Dodi, I., Virdis, R., Abelli, L. A., Aloisi, A. & other authors

(2004). Epidemiological aspects of human rotavirus infection in children hospitalized with acute gastroenteritis in an area of northern Italy. *Acta Biomed* **75**, 100–106.

Medici, M. C., Abelli, L. A., Martinelli, M., Martella, V., Dettori, G. & Chezzi, C. (2009). Molecular characterization of group C rotaviruses detected in children in Italy. *Acta Biomed* **44**, 62–65.

Mihalov-Kovács, E., Gellért, Á., Marton, S., Farkas, S. L., Fehér, E., Oldal, M., Jakab, F., Martella, V. & Bányai, K. (2015). Candidate new rotavirus species in sheltered dogs, Hungary. *Emerg Infect Dis* **21**, 660–663.

Moutelíková, R., Prodělalová, J. & Dufková, L. (2014). Prevalence study and phylogenetic analysis of group C porcine rotavirus in the Czech Republic revealed a high level of VP6 gene heterogeneity within porcine cluster I1. *Arch Virol* **159**, 1163–1167.

Parashar, U. D., Hummelman, E. G., Bresee, J. S., Miller, M. A. & Glass, R. I. (2003). Global illness and deaths caused by rotavirus disease in children. *Emerg Infect Dis* **9**, 565–572.

Rahman, M., Hassan, Z. M., Zafrul, H., Saiada, F., Banik, S., Faruque, A. S., Delbeke, T., Matthijnssens, J., Van Ranst, M. & other authors (2007). Sequence analysis and evolution of group B rotaviruses. *Virus Res* **125**, 219–225.

Schnagl, R. D., Boniface, K., Cardwell, P., McCarthy, D., Ondracek, C., Coulson, B., Erlich, J. & Morey, F. (2004). Incidence of group C human rotavirus in central Australia and sequence variation of the VP7 and VP4 genes. *J Clin Microbiol* **42**, 2127–2133.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729.

Wang, Y. H., Zhou, X., Kobayashi, N., Zhu, Z. R., Liu, M. Q. & Yang, Z. Q. (2007). Amplification of entire fragment of group C rotavirus genes and the sequence analysis on VP6 and VP7 genes in the field strain Wu-82 isolated in Wuhan area. *Chin J Zoonoses* **23**, 1167–1171.

Yamamoto, D., Ghosh, S., Kuzuya, M., Wang, Y. H., Zhou, X., Chawla-Sarkar, M., Paul, S. K., Ishino, M. & Kobayashi, N. (2011). Whole-genome characterization of human group C rotaviruses: identification of two lineages in the VP3 gene. *J Gen Virol* **92**, 361–369.

Zhirakovskaia, E., Tikunov, A., Klemesheva, V., Loginovskikh, N., Netesov, S. & Tikunova, N. (2016). First genetic characterization of rotavirus C in Russia. *Infect Genet Evol* **39**, 1–8.