Tracing the origin and co-phylogeny of the caliciviruses

Graham J. Etherington,1 Susan M. Ring,1 Michael A. Charleston,2,3 Jo Dicks,3 Vic J. Rayward-Smith4 and Ian N. Roberts1

Correspondence
Ian N. Roberts
Ian.Roberts@bbsrc.ac.uk

1Department of Food Safety Science, Institute of Food Research, Norwich Research Park, Colney Lane, Norwich NR4 7UA, UK
2Evolutionary Biology Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK
3Computational Biology Group, John Innes Centre, Norwich Research Park, Colney Lane, Norwich NR4 7HA, UK
4School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK

Received 26 October 2005
Accepted 22 December 2005

Caliciviruses infect a wide range of mammalian hosts and include the genus Norovirus, the major cause of food-borne viral gastroenteritis in humans. Using publicly available sequence data and phylogenetic analysis tools, the origins and virus–host co-phylogeny of these viruses were investigated. Here, evidence is presented in support of host switching by caliciviruses, but showing that zoonotic transfer does not appear to have occurred in the history of these viruses. The age or demography of the caliciviruses cannot yet be estimated with any firm degree of support, but further studies of this family, as new dated sequences become available, could provide key information of importance to human health and in understanding the emergence of food-borne disease.

INTRODUCTION

Caliciviruses are a family of viruses that comprise the genera Vesivirus, Lagovirus, Norovirus and Sapovirus. They have a positive-sense, single-stranded RNA genome of around 7.5 kb, usually containing three open reading frames (ORFs). ORF1 encodes the non-structural proteins, ORF2 encodes the capsid protein and ORF3 encodes a minor structural protein. In some caliciviruses, ORF1 and ORF2 are fused to form a contiguous polyprotein (Clarke & Lambden, 1997). Caliciviruses infect a broad host range of animals that includes reptiles, cetaceans, cattle, lagomorphs (rabbits and hares), pigs, cats, skunks, chimpanzees and humans (van Regenmortel et al., 2000). Human caliciviruses in the genera Norovirus and Sapovirus are the major cause of viral gastroenteritis in humans (Koopmans & Duizer, 2004), with symptoms that may include nausea, diarrhoea, vomiting, abdominal cramping, fever and general malaise (Hardy & Estes, 1996; van Regenmortel et al., 2000). Norovirus was found to be the most common cause of intestinal disease in a UK population cohort component (where the study population included those with a similar opportunity for exposure and excluded those without that opportunity) and usually infected between 20 and 160 people per month per 100 000 population (combined individual and outbreak cases) (Food Standards Agency, 2000). The same study also estimated that 1 % of children will have suffered from the effects of norovirus gastroenteritis by their first birthday (Food Standards Agency, 2000). In the USA, noroviruses are the most common cause of gastroenteritis, with an estimated 23 million cases annually (Centers for Disease Control and Prevention, 2003).

Complete genomic sequences of caliciviruses are available from the GenBank database (Benson et al., 2002) for several mammalian hosts. These host species include cow, hare, rabbit, human, cat, sea lion and pig. Partial genome sequences are available for various other host species, but derive from varying parts of the genome and so cannot be aligned for use in phylogenetic studies. A much larger number of sequences from the polymerase and capsid protein genes are also available and have been used extensively for phylogenetic analysis (Berke et al., 1997; Green et al., 2000), resulting in the definition of a number of genogroups (Berke et al., 1997). However, the co-phylogeny of these viruses with their hosts and their demographic growth have yet to be investigated in any detail. Demographic histories of viruses vary greatly (Yusmin et al., 2001; Twiddy et al., 2003). With the availability of dated sequence information and the
introduction of coalescent methods (Kingman, 1982; see Methods), it is now possible to estimate (i) nucleotide substitution rate, (ii) the date of the most recent common ancestor (MRCA) and (iii) the population size over time (demographic growth) for a dataset of interest (Jenkins et al., 2002). Estimates of all three of these parameters are reliant on the existence of a molecular clock of evolution.

Zoonotic transfer of viruses is a constant threat, both from the food chain (Slifko et al., 2000) and via direct transfer (Meslin et al., 2000). Recent high-profile examples include a variety of RNA viruses, notably Influenza A virus (Hong Kong chicken flu HSN1) (Yuen et al., 1998), Ebola and Marburg viruses (Schou & Hansen, 2000) and severe acute respiratory syndrome coronavirus (Stavrinides & Guttman, 2004). In this paper, we have extended analysis of the calicivirus genome and sequence evolution by relating it to host phylogeny. We have also examined the origin of the noroviruses and comment on the nature of cross-species and cross-population transfer in the history of food-borne virus evolution and with regard to the potential future relationship of these viruses with our own species.

METHODS

The calicivirus sequences used in this study were obtained from public domain databases. All accession numbers and links to software are available at http://vir.sgmjournals.org/supplemental/index.shtml.

Co-phylogeny analysis. Calicivirus genomic sequences for seven mammalian hosts were obtained from GenBank. The sequences were aligned using the CLUSTAL_X software (Thompson et al., 1997) and a phylogenetic tree was obtained using the DNAML option of the PHYLIP software package (Felsenstein, 1989; using the model as described by Felsenstein & Churchill, 1996). Although the molecular clock model, which is required by the co-phylogeny software we will introduce below, is not the optimal model for this dataset, the resulting topology of the phylogeny is broadly similar to that obtained by the no-clock model. The basic topology of the host tree was obtained from the Tree of Life Web Project (Maddison et al., 2001). The dates of the splits in the tree were obtained from literature sources (Kumar & Hedges, 1998; Bininda-Emonds et al., 1999; Lui et al., 2001; Murphy et al., 2001).

The TreeMap software (http://www.it.usyd.edu.au/~mcharles) for host–parasite co-evolution (Charleston & Page, 1998) was used to estimate the co-phylogeny of the virus and host trees. All potential optimal solutions were examined and the probability of each co-phylogeny having arisen by chance was calculated using the TreeMap randomization test. We also tested the significance of the number of co-divergent events (CEs) and non-co-divergent events (NCEs) observed in the optimal solutions. One thousand sets of random ‘parasite’ trees were generated separately for each test. For the first set, the maximum number of CEs and for the second, the minimum number of NCEs was calculated when analysed with our fixed host tree. The observed values of CEs and NCEs from the optimal solutions were then compared with the corresponding distribution obtained from the randomly generated trees. These tests gave a further indication of whether our optimal trees could have arisen by chance alone.

The analysis outlined above was carried out on complete calicivirus genome sequences. In addition, for each of our species, we took a subset of the genomic sequence, the capsid sequence, and analysed them using a similar strategy. Finally, we compared the two sets of results. We did this in order to compare our whole-genome results with those of rapidly evolving sequences, possibly under high selective pressure, where the evolutionary picture might be different.

Tree dating analysis. Viral strains associated with an outbreak will sometimes contain the date of the outbreak in the annotation of their sequences. This extra information may be used in the phylogenetic analysis process. A phylogenetic tree containing sequences with known isolation dates (dated tips) has two scales: the timescale of the tree measured in years and the changes in branch lengths measured in number of substitutions per site. If we assume a single rate of evolution across the tree, we are left with a linear relationship between time and substitution rate. Consequently, we can use the time between dated tips to calibrate the clock of the entire tree and hence estimate the MRCA and substitution rate.

The TipDate program (Rambaut, 2000) was used to carry out this process. We used TipDate to try to estimate substitution rates and the age of the calicivirus and norovirus MRCA using dated capsid sequences. Several norovirus capsid sequences, isolated from humans, are associated with known isolation dates. These capsid sequences were used to examine the age of the norovirus and calicivirus trees. Two datasets were assembled.

Dataset 1 consisted of 30 norovirus capsid sequences from dated outbreaks. Although more sequences were available, we restricted the size of the dataset to make analyses manageable. However, in order to ensure that our estimates of the substitution rate and MRCA were representative of the noroviruses as a whole, we chose the subset of sequences that represented the broadest available range of dated viruses.

Dataset 2 comprised 21 dated norovirus capsid sequences, which were known to form a closely related clade within a larger dataset, plus capsid sequences from each of the complete calicivirus genomes used in the earlier co-phylogeny mapping analysis. It was necessary to form this additional dataset in order to examine both noroviruses as an isolated genus (dataset 1) and the caliciviruses as a whole family (this dataset). For those complete genomes where no isolation date was given, the submission year to GenBank was used in its place. Although potential differences between the submission date and the date of outbreak may bias our results, the use of submission dates would be expected to provide a reasonable indication of whether the caliciviruses as a whole have an ancient or a recent origin.

For each of our two datasets, the sequences were aligned using CLUSTAL_X (Thompson et al., 1997). Maximum-likelihood trees were estimated using the DNAML option in PHYLIP (Felsenstein, 1989; using the model as described by Felsenstein & Churchill, 1996) (see Supplementary material available in JGV Online). The capsid sequences in dataset 1 were rooted using the outgroup Vesicular exanthema of swine virus (VESV), whilst those in dataset 2 were rooted using Hepatitis E virus. VSV was chosen for dataset 1 as it occurs in the same family as the noroviruses, but in a different genus. Hepatitis E virus was chosen as the outgroup for dataset 2 as it is a closely related virus, but is classified outside the family Caliciviridae. Hence, both outgroups were similar enough to align with their respective datasets, but divergent enough to be grouped, taxonomically, outside all of the clades found in their respective datasets.

Using TipDate and the phylogenetic trees obtained above as input, three models – single rate (SR), single rate dated tips (SRDT) and different rate (DR) – were tested. SR, the simplest model, is the single rate (molecular clock) model where the sequences are considered to be contemporaneous. SRDT is the single rate model that takes into account non-contemporaneous sequences. The DR model, the most complex, permits each branch of the tree to have a different rate of substitution. The outputs of the models differ in that the SRDT model...
estimates the date of the MRCA, the substitution rate and the likelihood of the tree under the model, whilst SR and DR output the likelihood of the tree only. Additionally, for the SRDT model, 95% confidence intervals (CI) for the substitution rate and MRCA were found by finding the parameter estimates that gave a log likelihood score of 1+92 less than the maximum value (Jenkins et al., 2002). Each model used the HKY85 model of nucleotide substitution (Hasegawa et al., 1985), as this contains an intermediate number of parameters, ensuring that the analysis would not be significantly under- or over-parameterized. Both homogeneous (SRDT Homo) and heterogeneous (SRDT Hetero) rate variation between sites were examined for all three models, where the latter was assumed to have a gamma distribution with four rate categories and an alpha shape parameter of 0.5 (as used by Chare et al., 2003). The maximum log likelihood (ln L) of the tree under each model was noted and a likelihood ratio test (LRT) was carried out between SR and DR and between SRDT and DR for both the homogeneous and the heterogeneous rate cases.

RESULTS

Co-phylogeny analysis

Using TreeMap, the ‘tanglegram’, a graphical representation of the relationship between host and virus phylogenies, was taken and examined for incongruence. As can be seen from Fig. 1, the host and virus trees for the caliciviruses do not match exactly, indicating that the evolutionary origins of some of the viruses do not lie with their current hosts. Identical results were obtained using the capsid sequence dataset (results not shown), indicating a similar evolutionary process within this rapidly evolving region to that acting over the whole genome. TreeMap searches for all potential host–virus trees (termed a ‘jungle’; Charleston, 1998) and then calculates the costs of each tree. Jungles enable swift achievement of optimal solutions under general weighting schemes, including minimization of the total number of events and maximization of co-speciation events. According to the model, two aspects of event costs are of primary importance. Firstly, the cost of co-divergence is less than that of any other kind. Secondly, three event types (excluding co-divergence) are permitted: duplication, lineage loss and host switch. Without knowing the relative likelihoods of these events, it is impossible to assign costs to them; we therefore simply minimize their total number. This is equivalent to assigning unit costs, but we offer no justification for such an assignment and must be governed by practicality. TreeMap returns all possible maps that could be optimal under an arbitrary set of event costs and here we chose those that are most parsimonious. From the complete genome dataset, no such assumptions were made. Six good solutions were found, of which two had equally low costs of 7. These two optimal solutions were examined further using the TreeMap randomization test and both were found to be significant at the 5% level ($P=0.01 \pm 0.007$ for the first and $P=0.02 \pm 0.014$ for the second). The first of these two solutions is shown in Fig. 2. The only difference between this solution and the second least-cost solution is the direction of the host switch between Lagomorpha (rabbits and hares) and Carnivora. We tested the two solutions for the number of both CEs and NCEs. Both were significant at the 5% level, with $P=0.015 \pm 0.00384$ for co-divergences and $P=0.013 \pm 0.00358$ for non-co-divergences.

Fig. 1. Apposing phylogenies showing host (left), virus (right) and deduced relationships (centre) were produced using TreeMap (Charleston & Page, 1998). The viral genomes used were: Rabbit hemorrhagic disease virus (RabbitHDV), San Miguel sea lion virus (SMSLV), norovirus (human calicivirus GII; not shown), Norwalk virus (HumanNLV), Hawaii calicivirus (not shown), European brown hare syndrome virus (BrownHareSV), Feline calicivirus (FelineCV), Bovine calicivirus strain NB (BovineCV), Porcine enteric calicivirus (PorcineCV) and Vesicular exanthema of swine virus (SwineVESV). Although three human viruses were included in the original analysis, it was found that they grouped closely together; thus, in order to gain more overall clarity in the tree, only one human virus was included in the analysis shown. Accession numbers are provided in Supplementary material (available in JGV Online).
Consequently, we were confident that the co-phylogeny estimated by TreeMap had not occurred by chance alone.

The results indicated an early host switch between *Carnivora* and *Lagomorpha* to have occurred together with a later host switch from sea lion (*San Miguel sea lion virus, SMSLV*) to pig lineages (VESV). The analysis also indicated that Porcine enteric calicivirus, Bovine calicivirus, *European brown hare syndrome virus*, *Rabbit hemorrhagic disease virus* and the human caliciviruses show co-divergence with their hosts. Regarding the timing of the host-switching events, we could draw two conclusions. Firstly, if the two ultrametric trees of the reconstruction are rooted together using the host tree as a timescale, the results indicate that the ancient host switch between the *Lagomorpha* and the *Carnivora* occurred approximately 60 million years ago during the Palaeocene epoch, a period characterized as the starting point of a dramatic evolutionary radiation by mammals occupying the vacant ecological niches left by the recently extinct dinosaurs (Alroy, 1999). It is tempting, therefore, to speculate that an ancestral *Carnivora* virus infected an early lineage of *Lagomorpha* via newly created links in the Palaeocene food chain. However, if we were to speculate that the age of the virus tree was insignificant when compared with that of the host tree, then our results would clearly suggest a co-phylogeny that includes co-divergence but not co-speciation. This would mean that, although the viral phylogeny resembles that of the host phylogeny, it probably did not occur through co-evolution. Generally speaking, a given virus would be expected to be able to jump to a more closely related host (e.g. rabbit to hare) more easily than to a more distantly related host (e.g. rabbit to human). Therefore, as the viruses jump from closely related host to closely related host, they mirror the phylogeny of their chosen hosts. This ‘preferential host switching’ has been suggested in primate lentiviruses (Charleston & Robertson, 2002) and would provide an explanation for the incongruence towards the base of the co-phylogeny (Fig. 1). This incongruence could be the result of viruses switching to slightly more distantly related hosts and then infecting more closely related hosts at a later stage. Equally, incongruence could have been caused by viruses switching to and from intermediate hosts not included in the analysis. In the following analyses, we attempted to find evidence in favour of one or other of these two hypotheses (co-divergence or co-speciation).

**Tree dating analysis**

In the analysis of dataset 1, 30 noroviruses were used to estimate a substitution rate and the age of the MRCA. Of the two SRDT models (see Methods), SRDT Hetero gave a significantly better fit to the data than SRDT Homo, with estimates of 1678 years for the MRCA and 0.002707 for the number of substitutions per site per year (Table 1a). However, although the SRDT Hetero model provided a
Table 1. TipDate analysis of (a) norovirus evolution (dataset 1) and (b) calicivirus evolution (dataset 2) for six models

<table>
<thead>
<tr>
<th>Model*</th>
<th>MRCA‡</th>
<th>Substitution rate§</th>
<th>Log likelihood (ln L)</th>
<th>Parameters</th>
<th>Degrees of freedom</th>
<th>2Δ§</th>
<th>P value¶</th>
<th>Accept/reject this model when compared with DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Dataset 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR Homo</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>SRDT Homo</td>
<td>2169 (1468, ∞)</td>
<td>0-000462 (0-0, 0-00698)</td>
<td>−15613</td>
<td>57</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SR Homo</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>DR Hetero</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>SRDT Hetero</td>
<td>1678 (2133, ∞)</td>
<td>0-002707 (0-0, 0-004072)</td>
<td>−14361</td>
<td>35</td>
<td>27</td>
<td>210</td>
<td>&lt;0-01</td>
<td>Reject</td>
</tr>
<tr>
<td>SR Hetero</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>(b) Dataset 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR Homo</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>SRDT Homo</td>
<td>1574 (1024, ∞)</td>
<td>0-000905 (0-0, 0-001401)</td>
<td>−18786</td>
<td>28</td>
<td>25</td>
<td>178</td>
<td>&lt;0-01</td>
<td>Reject</td>
</tr>
<tr>
<td>SR Homo</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>DR Hetero</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>SRDT Hetero</td>
<td>5127 (3831, ∞)</td>
<td>0-002688 (0-0, 0-004333)</td>
<td>−18047</td>
<td>33</td>
<td>25</td>
<td>224</td>
<td>&lt;0-01</td>
<td>Reject</td>
</tr>
<tr>
<td>SR Hetero</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

*Different rate model (DR), single rate model (SR) and single rate dated tips model (SRDT), with each analysed under rate homogeneity (Homo) and rate heterogeneity (Hetero).
‡Estimated years to MRCA (lower/upper limit at 95% CI) (∞ denotes an infinite age and is associated with a lower confidence limit for the substitution rate of 0). While no estimate of years to MRCA can be accepted with confidence, all are consistent with emergence within the last 5000 years.
§Estimated number of substitutions per site per year (lower/upper limit at 95% CI).
¶Difference in the number of degrees of freedom (d.f.) between the model under analysis and DR (d.f.) and the P value obtained from the LRT.

significantly better fit to the data than the SR Hetero model, the DR Hetero model provided the best overall fit. This is reflected in a lower confidence limit of 0 for the substitution rate and the corresponding upper limit of ∞ for the MRCA in the SRDT Hetero model. In the analysis of dataset 2, the capsid sequences of the calicivirus genomes used in the co-phylogeny analysis were added to a known clade of 21 noroviruses. Again, of the two SRDT models, SRDT Hetero provided the better fit to the data, with estimates of 5127 years for the MRCA and 0-002688 for the number of substitutions per site per year (Table 1b). However, as for dataset 1, the DR Hetero model provided the best overall fit.

The acceptance of the DR Hetero model goes some way towards explaining the topology of the norovirus and calicivirus trees (see Supplementary material available in JGV Online), where the tips of the best DNA ML trees are not consistent with an SRDT model (i.e. the tips of many of the earlier isolated sequences are further away from the root of the tree than those sampled more recently). Consequently, although the MRCA and substitution rates obtained by the SRDT Hetero models are consistent with analogous estimates for other RNA viruses (Drake & Holland, 1999; Jenkins et al., 2002; Holmes, 2003) and in this light we are inclined to believe that they represent something close to the real values, they cannot be relied upon, as the confidence intervals clearly indicate.

If we assume that the MRCA estimated by the SRDT Hetero model is approximately correct at 5000 years and the sequence data used in this analysis span a series of outbreaks covering a 14 year period, we are looking at perhaps only 0·3 % of the length of the tree. Ideally, a much wider range of sequence isolation dates would have been preferred, especially from closely related strains. Although they are not available at present, it may be possible to obtain these sequences and to carry out a more thorough analysis in future years. In addition, it would be interesting to examine a number of dated complete-genome sequences, rather than capsid sequences. The norovirus capsid sequences are likely to have evolved faster than non-structural proteins, thus influencing the results obtained. In conclusion, although it is possible that the norovirus and calicivirus sequences analysed have arisen within the last 5000 years, we have not been able to provide conclusive evidence for this age and hence are unable to estimate dates of the host switches discovered in the previous analysis.
DISCUSSION

A co-phylogeny analysis of seven host species and their respective caliciviruses showed strong evidence for two host switches in the evolution of these species. While we were fairly confident of the phylogeny of the host species, covering a 90 million year period, we were less certain regarding the evolution of the viral sequences. Due to a number of factors, including different rates of evolution in different lineages and the small range of dated sequences compared with the likely age of the tree, it was not possible to obtain accurate estimates of the viral tree age and the accompanying substitution rate. Hence, it is impossible to determine at present whether the co-phylogeny of the caliciviruses with their mammalian hosts is due to ancient co-evolution or recent preferential host switching and co-divergence. However, if the evolution of the caliciviruses is found in future to have followed that of most other RNA viruses, then the latter is most certainly the case.

In the absence of a valid substitution rate, it is also difficult to draw conclusions concerning the demographic growth of viruses compared with their hosts and thus to make predictions concerning food safety. In future years, it will become possible to obtain dated sequences from caliciviruses that cover a greater time period, providing better information on the substitution rate and MRCA. As this occurs, it will be interesting to examine the demographic growth of the human-related noroviruses to examine what factors influence the spread of the viruses, e.g. did the growth of the viruses increase in line with human population growth or perhaps emerge with the first large-scale movements of the human race around the world? The type of demographic growth observed, e.g. exponential, logistic or piece-wise logistic (Pybus & Rambaut, 2002), will provide valuable information for food-safety strategies.

Throughout our analysis, we could find no evidence of zoonotic behaviour in caliciviruses. In all of the evolutionary co-phylogeny reconstructions examined, none showed any suggestion of non-human-to-human host switches. Of course, we cannot rule out the possibility that the inclusion of more virus sequences from more closely related hosts may indicate the potential for transfer amongst, for example, primate lineages. However, it is unlikely that the more distant relatives investigated in the present study pose a significant threat to humans.

In contrast, this might not be the case for some of the other hosts in our analysis. In the co-phylogeny (Fig. 1), a virus representing an early ancestor of SMSLV jumped from sea lion to pig, evolving into VESV. It has been established for some time that SMSLV is very similar to VESV in many respects and that the respective viral particles are morphologically indistinguishable (Clarke & Lambden, 1997). Furthermore, experimental research (Berry et al., 1990) has shown that pigs are readily infected by SMSLV, resulting in a transmissible vesicular disease. Indeed, it has been proposed that recent epidemics of VESV in pigs may have resulted from the feeding of marine mammal (seal) meat and fish to pigs as a protein supplement during America’s Great Depression (1929–1941) (Smith et al., 1998). Co-phylogeny results are in agreement with the conclusion of these authors that there must be continuing concern that a VESV-like disease could reappear in the USA due to the large number of marine mammals on the west coast potentially acting as a marine reservoir for caliciviruses (Smith et al., 1998).

In summary, identifying the date for the MRCA of emergent norovirus strains promises to be useful for understanding the circumstances surrounding their origin and the rate at which their populations may be expected to expand and diverge in the future. This will become increasingly feasible as more dated genomic sequences are added to the databases. It will also be interesting to determine whether other enteric viruses follow similar patterns of evolution and demographic growth.

ACKNOWLEDGEMENTS

This work was supported by grants from the Biotechnology and Biological Sciences Research Council (to I. N. R., J. D. and V. J. R.-S.), a BBSRC studentship (to G. J. E.) and a Royal Society Research Fellowship (to M. A. C.). The authors thank Edward C. Holmes, George Lomonossoff, Oliver Pybus and an anonymous reviewer for their valuable advice.

REFERENCES


Evolutionary history of food-borne viruses


