

Recombination Patterns in Aphthoviruses Mirror Those Found in Other Picornaviruses[∇]

Livio Heath,¹ Eric van der Walt,¹ Arvind Varsani,² and Darren P. Martin^{3*}

Department of Molecular and Cell Biology, Faculty of Science, University of Cape Town, Rondebosch, 7701, South Africa¹; Electron Microscopy Unit, University of Cape Town, Rondebosch, 7701, South Africa²; and Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Observatory 7925, South Africa³

Received 30 May 2006/Accepted 4 September 2006

Foot-and-mouth disease virus (FMDV) is thought to evolve largely through genetic drift driven by the inherently error-prone nature of its RNA polymerase. There is, however, increasing evidence that recombination is an important mechanism in the evolution of these and other related picornaviruses. Here, we use an extensive set of recombination detection methods to identify 86 unique potential recombination events among 125 publicly available FMDV complete genome sequences. The large number of events detected between members of different serotypes suggests that horizontal flow of sequences among the serotypes is relatively common and does not incur severe fitness costs. Interestingly, the distribution of recombination breakpoints was found to be largely nonrandom. Whereas there are clear breakpoint cold spots within the structural genes, two statistically significant hot spots precisely separate these from the nonstructural genes. Very similar breakpoint distributions were found for other picornavirus species in the genera *Enterovirus* and *Teschovirus*. Our results suggest that genome regions encoding the structural proteins of both FMDV and other picornaviruses are functionally interchangeable modules, supporting recent proposals that the structural and nonstructural coding regions of the picornaviruses are evolving largely independently of one another.

Foot-and-mouth disease is a highly contagious viral disease of cloven-hoofed animals and remains a major animal health concern affecting agricultural economics worldwide. The causal agent, *Foot-and-mouth disease virus* (FMDV), is the type species of the genus *Aphthovirus*, which constitutes one of nine genera within the family *Picornaviridae* (5). Picornaviruses are small, nonenveloped icosahedral viruses with positive-sense single-stranded RNA genomes. Their single open reading frame (ORF) encodes a polyprotein that is cleaved during a cascade of proteolytic events to yield mature viral proteins (1). The spatial arrangement of the viral genes is somewhat polarized, with the regions encoding the structural proteins (1A, 1B, 1C, and 1D) clustered at the 5' end of the ORF. The majority of the nonstructural proteins (2A, 2B, 2C, 3A, 3B, 3C, and 3D) are encoded in the 3' half of the ORF, with the only exception being the leader protein (L). In FMDV, L is a papain-like protease responsible for the inhibition of cap-dependent translation through cleavage of the eIF4G translation initiation factor (3).

The extensive genetic and antigenic diversity observed in RNA viruses such as FMDV is generally attributed to the error-prone nature of their replication machinery (4). FMDV is thought to be evolving largely through genetic drift but with positive selection contributing substantially to the fixation of mutations, particularly in the capsid coding regions (8, 20, 21). However, incongruence between the

inferred phylogenies of individual subgenomic regions suggests that recombination may also play a significant role in FMDV evolution (7, 32, 34). A number of specific FMDV recombination events have been described thus far. Whereas a very few events have involved exchanges of genome sequences encoding parts of the capsid-coding region (P1), all of the rest seem to have involved genome regions encoding the nonstructural proteins (11, 12, 32). In this study, we sought first to identify the set of unambiguously unique recombination events detectable in publicly available FMDV full genome sequences and secondly to determine the distribution of these events across the FMDV genome. We provide evidence that the P1 regions of both FMDV and picornaviruses in general may be functionally interchangeable modules that facilitate their promiscuous recombinational exchange within different picornavirus species.

MATERIALS AND METHODS

The complete genome sequences of 125 FMDVs were downloaded from <http://virology.wisc.edu/acp> in May 2005 and aligned using ClustalX (31; <http://darwin.uvigo.es/rdp/heath2006.zip>). The FMDV species alignment described by Palmenberg et al. (26) was used to guide a manual editing process. Amino acid sequence alignments and known biological features were taken into account to preserve the biological relevance of the alignment.

Phylogenetic-compatibility analysis was performed using the program TreeOrderScan in the Simmonic 2005 ver. 1.4 package (28, 29). This software generates "optimally ordered" rooted neighbor-joining trees (100 bootstrap replicates; all branches with <70% support are collapsed) for successive fragments along an alignment (300 nucleotides [nt] in length at 100-nt intervals). The taxon orders within each tree are then compared with those generated for all the other 300-nt sequence fragments examined along the alignment. A compatibility matrix is constructed by counting the minimum number of tree topology alterations required to convert the taxon ordering in each tree to that of every other tree. Sequences were assigned to predefined groups (based on their serotypes) in order to compute the numbers of phy-

* Corresponding author. Mailing address: Institute of Infectious Disease and Molecular Medicine, Observatory, Cape Town, 7925, South Africa. Phone: 27-21-406 6366. Fax: 27-21-689 1528. E-mail: Darren@science.uct.ac.za.

[∇] Published ahead of print on 13 September 2006.

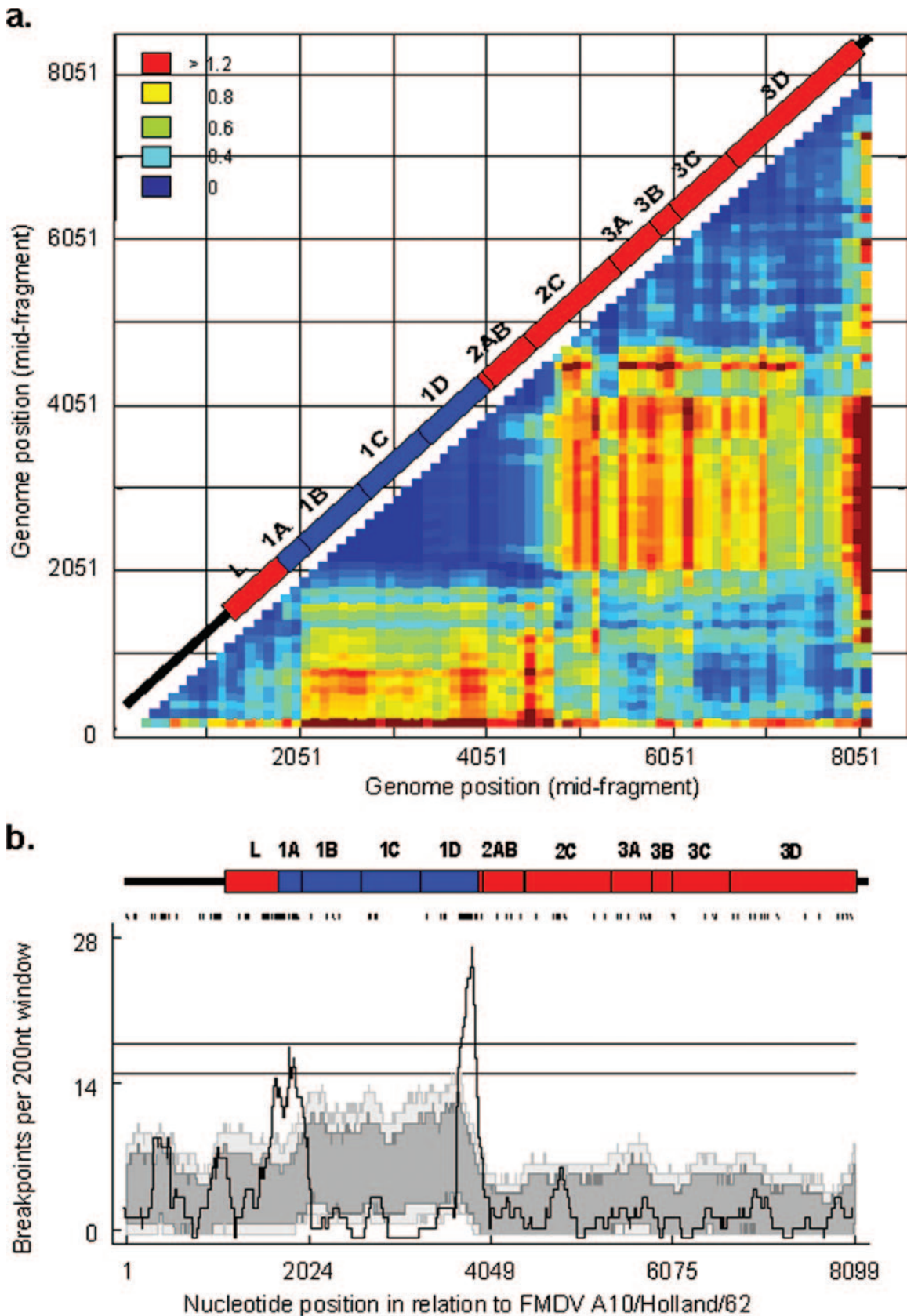


FIG. 1. (a) Phylogenetic compatibility matrix of 125 FMDVs. To generate this plot, a 300-nucleotide window was moved across the alignment with 100-nucleotide steps, and at each position, a neighbor-joining tree was constructed. The compatibility of two trees is measured by the number

logenetic violations between groups of sequences represented in different trees.

To investigate the extent of recombination within the data set, the aligned sequences were examined using the Recombination Detection Program (RDP) (16), GENECONV (25), BOOTSCAN (17), MAXIMUM CHI SQUARE (22), CHIMAERA (19), and SISTER SCAN (6) recombination detection methods as implemented in RDP3 (19), available from <http://darwin.uvigo.es/rdp/rdp.html> (for full details of program settings, see <http://darwin.uvigo.es/rdp/heath2006.zip>). The breakpoint positions and recombinant sequence(s) inferred for every detected potential recombination event (PRE) were manually checked and adjusted where necessary using the extensive phylogenetic and recombination signal analysis features available in RDP3.

Once a set of unique PREs was identified (<http://darwin.uvigo.es/rdp/heath2006.zip>), a breakpoint map containing the positions of all clearly identifiable breakpoints was compiled. A breakpoint density plot was then constructed from this map by moving a 200-nt window 1 nucleotide at a time along the length of the map. At each window position, all identified breakpoints falling within the window were counted, and the number was plotted at the central window position. Significant clustering of breakpoint positions within each window was tested by permutation. Globally significant breakpoint clusters were identified as those windows within the breakpoint density plot that contained more breakpoint positions than the maximum found in more than 95% of 1,000 permuted breakpoint density plots. Locally significant breakpoint clusters were identified as those windows at a particular location within the breakpoint density plot that contained more breakpoint positions than more than 99% of the windows at the same location in the 1,000 permuted breakpoint density plots (for a detailed description of the permutation test, see <http://darwin.uvigo.es/rdp/heath2006.zip>).

RESULTS AND DISCUSSION

The major antigenic determinant of FMDV is 1D. Since the serotypic relatedness of different isolates is reflected in the phylogeny of the 1D gene region, it has traditionally been used in the phylogenetic analysis of FMDV sequences. However, radically incongruent tree topologies between the structural and nonstructural coding regions of FMDV isolates has suggested that the 1D phylogeny may not appropriately reflect the evolutionary histories of different FMDV isolates (7, 32, 34). To address this issue, we analyzed the phylogenetic compatibility of successively generated sequence fragments across the complete genome sequences of 125 FMDV isolates. Phylogenetic-compatibility analysis indicated extensive incongruence between different regions of the genome (Fig. 1). Whereas the phylogenies of the 1A, 1B, and 1C gene regions are consistent with that of 1D, the phylogeny of the remainder of the genome is largely irreconcilable with that of the P1 region. Phylogenetic-incompatibility values among different parts of the 3' portion of the genome are only slightly higher than between different regions of P1, but the phylogeny of the 3' half of the genome is also substantially incompatible with those of the 5' non-translated region and the L gene. The incongruence of different parts of the FMDV genome suggests that intertypic

recombination may have played an important role in the evolution of FMDV.

To detect evidence of individual recombination events within the FMDV alignment, it was examined using a set of six recombination detection methods implemented in RDP3 (19). RDP3 uses a mixture of statistical and phylogenetic methods to both identify evidence of probable recombination events within individual sequences and identify a minimal subset of unique events detectable within an entire alignment. Importantly, the program specifically avoids overestimating the amount of recombination in an alignment by identifying multiple descendants of individual recombination events. Eighty-six unique PREs among FMDV isolates were detected in this way (<http://darwin.uvigo.es/rdp/heath2006.zip>). Included in these 86 PREs are all 8 that have been previously described (2, 32, 34). Twenty-six of the 38 PREs for which two sequences closely related to the presumed recombinant's parental sequences were identified are between members of different serotypes. This implies that gene flow among serotypes is relatively common and that it does not necessarily incur severe fitness costs. Importantly, the results of this analysis are in close agreement with the phylogenetic-compatibility analysis in that the distribution of observed breakpoints appears to be nonrandom (Fig. 1). Whereas there are locally significant breakpoint cold spots in the 1B, 1C, and 1D genes of the P1 region (Fig. 1), the entire P1 region is bounded by two globally significant ($P < 0.01$) breakpoint hot spots.

The results of these FMDV breakpoint distribution and phylogenetic-compatibility analyses mirror those recently reported for the enteroviruses, another picornavirus genus. Phylogenies of the structural and nonstructural genes of enteroviruses are apparently also largely incompatible with one another (14, 24, 29). This suggested to us that a common mechanism may be influencing intertype recombination occurring among both aphthoviruses and enteroviruses and that evidence of the same mechanism might also be detectable in other picornavirus genera. To test this hypothesis, complete genome sequences of species A, B, and C enteroviruses (29) ($n = 28, 61,$ and $51,$ respectively) and teschoviruses (27) ($n = 29$) were subjected to the same set of analyses as that used on the FMDV alignment. There are insufficient full genome sequences available for members of the other picornavirus genera for analysis of these to have yielded any meaningful results. We detected evidence of 32, 119, 58, and 46 independent PREs in the enterovirus species A, B, and C and teschovirus alignments, respectively (<http://darwin.uvigo.es/rdp/heath2006.zip>).

Consistent with previous reports, significant recombination hot spots were detected at the boundaries of the P1 regions of all three enterovirus species (24, 29). These re-

of phylogenetic violations that are required to match the ordering of taxa within the tree. Obtained values are color coded, with phylogenetically compatible regions shown in blue. The FMDV genome diagram superimposed on the diagonal axis was drawn to scale relative to FMDV A10/Holland/62 (AY539751). (b) Potential recombination breakpoint hot spots within the FMDV alignment. Detectable breakpoint positions are indicated by small vertical lines at the top of the graph. A 200-nucleotide window was moved along the alignment 1 nucleotide at a time, and the breakpoints detected within the window region were counted and plotted (solid line). The upper and lower broken lines, respectively, indicate 99% and 95% confidence thresholds for globally significant breakpoint clusters. The light- and dark-gray areas, respectively, indicate local 99% and 95% breakpoint-clustering thresholds, taking into account local regional differences in sequence diversity that influence the abilities of different methods to detect recombination breakpoints (for a detailed description of the permutation test, see <http://darwin.uvigo.es/rdp/heath2006.zip>).

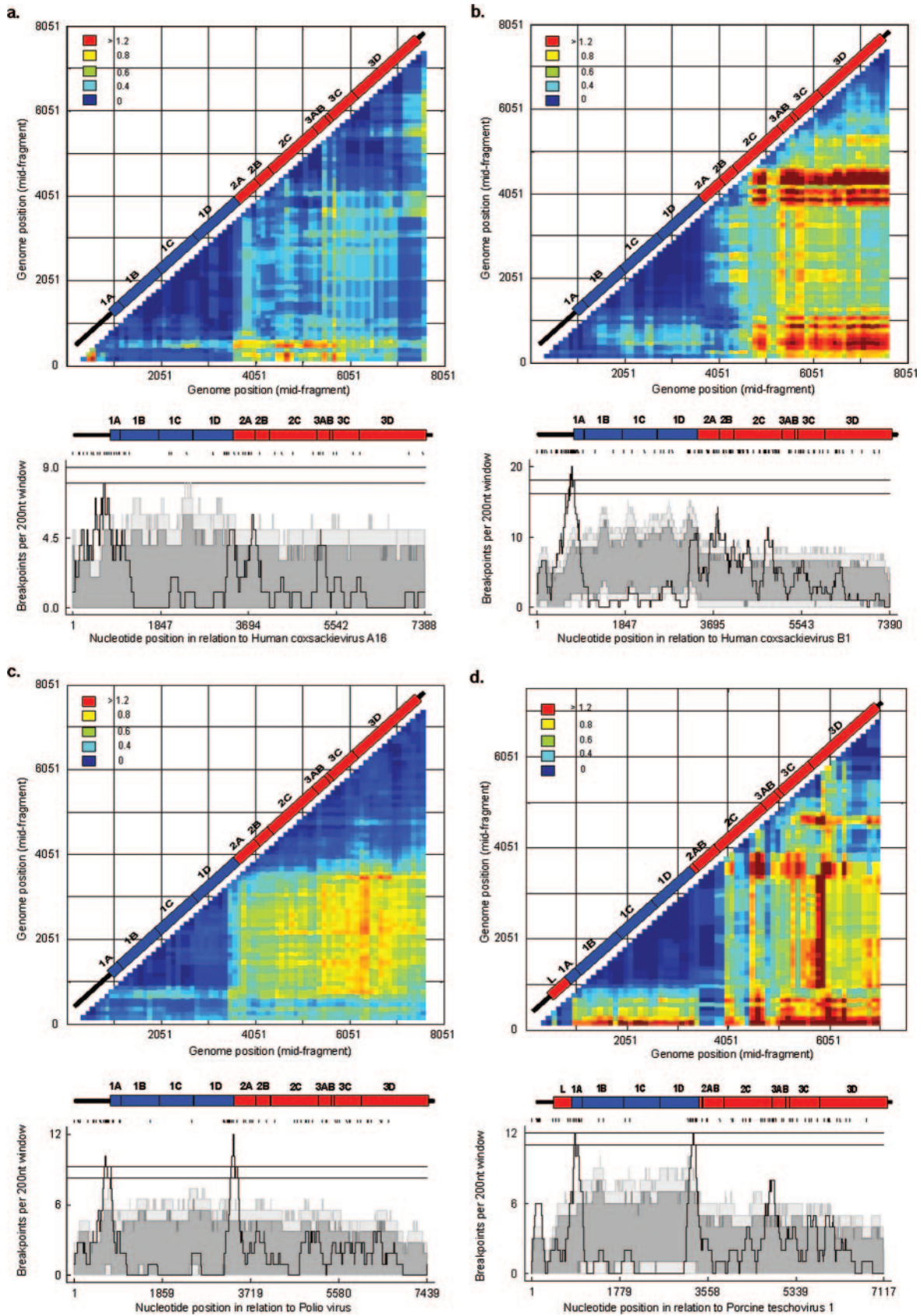


FIG. 2. Phylogenetic-compatibility matrices and breakpoint density plots of enteroviruses and teschoviruses constructed as outlined for Fig. 1. (a) Enterovirus species A. (b) Enterovirus species B. (c) Enterovirus species C. (d) Teschoviruses. The superimposed genome diagrams were drawn

combination hot spots are most clearly defined in the poliovirus and nonpoliovirus species C enteroviruses (Fig. 2c), for which two globally significant breakpoint hot spots are evident at or close to the boundaries of the P1 region. For the species A and B enteroviruses, globally significant hot spots are located close to the 5' boundary of the P1 region (Fig. 2a and b). Although there are breakpoint clusters close to the 3' P1 boundary in both species A and B enterovirus genomes, the significance of these clusters is not statistically supported. A locally significant breakpoint hot spot is, however, apparent in both the enterovirus A and B genomes at the boundary between the 2A and 2B ORFs (Fig. 2a and b). It is perhaps also noteworthy that a nonsignificant breakpoint cluster between the 2A and 2B ORFs is also evident in the enterovirus C genomes (Fig. 2c). The globally significant ($P = 0.01$) breakpoint hot spots at or close to the P1 boundaries in the teschoviruses closely resemble those observed in the FMDV and enterovirus C genomes (Fig. 2d). As with FMDV, the teschovirus 5' untranslated region also contains a locally significant breakpoint cluster. The other locally significant teschovirus breakpoint cluster at the 2C-3AB ORF interface also seems to correspond with breakpoint clusters at or near the 2C-3AB ORF interface in the A, B, and C enteroviruses, with the clusters in enterovirus Bs being locally significant. Importantly, all of the recombination hot spots identified by the breakpoint-clustering analyses coincide very closely with the boundaries of genomic regions with maximum phylogenetic conflict indicated by the phylogenetic-compatibility analyses.

Our study supports recent proposals that the constant generation of variants possessing different combinations of structural and nonstructural proteins is an important feature of enterovirus evolution (13, 15, 24, 29). Moreover, our work suggests that this reassortment of structural- and nonstructural-protein coding regions has played a similarly significant role in the evolution of other picornavirus genera. Recombination in picornaviruses may be analogous to component swapping or pseudorecombination that occurs among viruses with multipartite genomes. Such "pseudopseudorecombination" may be strongly facilitated by the partitioning of structural and nonstructural genes that is common in picornavirus genomes. Clustering of structural-protein genes, for example, will massively increase the probability that single recombination events will transfer them as an intact genome module.

Although it is likely that there is a common mechanism influencing the recombination patterns we have observed in FMDV and other picornaviruses, it is not immediately obvious what that mechanism might be. The recombination hot spots described here could be facilitated by conserved biochemical and/or folded nucleotide structural features at, for example,

the interfaces between the structural- and nonstructural-protein coding sequences. There does not, however, appear to be any correlation between the positions of enterovirus genomic-RNA secondary structures and favored recombination breakpoint positions detected in these viruses (29).

It is also possible that recombination occurs throughout these genomes at roughly the same rate but that recombinants containing breakpoints outside certain well-defined genome regions (such as at the interfaces between the structural and nonstructural genes) are generally less fit than parental viruses. Evaluation of the replicative abilities of recombinant FMDV and poliovirus genomes provides some experimental support for the notion that many newly formed recombinant picornaviruses, if not the vast majority, are substantially less viable than either of their parents (23, 30, 33). We and others have previously demonstrated that selection seems to favor the survival of recombinants in which the regions of sequence inherited from different sources either work well together or do not encode proteins that interact extensively with one another (10, 18). The viability of a newly formed recombinant genome is apparently inversely correlated with both the complexity of interactions between genome regions inherited from its different parents and the relatedness of those parents to one another (10, 18). It is conceivable that picornavirus recombinants with complexly interacting ORFs inherited from divergent parents will be removed by purifying selection. If this is true, what should remain is evidence of a set of recombination events that accurately anticipate protein-protein, protein-RNA, and RNA-RNA interactions between viral components. The apparent paucity of recombination breakpoints within P1 observed in all five picornavirus data sets that we examined may, for example, indicate that breakpoints occurring in this region have a high probability of disrupting complex interactions between the various structural proteins that play crucial roles in the stability and maturation of picornavirus virions (9).

The results presented here should be of considerable use in future analyses of picornavirus evolution. Partitioning picornavirus genomes at the recombination hot spots we have identified should both increase the power of and decrease the rate of false-positive inferences in population-genetic and selection studies. Our RDP project files (<http://darwin.uvigo.es/rdp/heath2006.zip>) can also be directly used with the program RDP3 to obtain mostly recombination-free picornavirus data sets that should be of great use in such studies. Besides this, the files, when loaded into RDP3, are essentially a highly interactive database of probable aphthovirus, enterovirus, and teschovirus recombination. These project files will enable detailed analyses of how, when, and where any of the many individual PREs detected in this study may have occurred.

to scale relative to sequences NC001430, NC001472, NC002058, and AJ011380. The upper and lower broken lines in each breakpoint density graph, respectively, indicate 99% and 95% confidence thresholds for globally significant breakpoint clusters. The light- and dark-gray areas, respectively, indicate local 99% and 95% breakpoint-clustering thresholds, taking into account regional differences in sequence diversity that influence the abilities of different methods to detect recombination breakpoints (<http://darwin.uvigo.es/rdp/heath2006.zip>). It should be noted that apart from a different color scale, the phylogenetic-compatibility analyses presented in panels a, b, and c are a repeat of those, using the same datasets, previously described by Simmonds and Welch (29) and are included here for comparative purposes.

ACKNOWLEDGMENTS

We thank Peter Simmonds for helpful comments on the manuscript.

We thank the Harry Oppenheimer Trust, South African National Bioinformatics Network, and National Research Foundation for funding this work. D.P.M. is supported by the Sydney Brenner Fellowship.

REFERENCES

1. Belsham, G. J. 2005. Translation and replication of FMDV RNA. *Curr. Top. Microbiol. Immunol.* **288**:43–70.
2. Carrillo, C., E. R. Tulman, G. Delhon, Z. Lu, A. Carreno, A. Vagnozzi, G. F. Kutish, and D. L. Rock. 2005. Comparative genomics of foot-and-mouth disease virus. *J. Virol.* **79**:6487–6504.
3. Devaney, M. A., V. N. Vakharia, R. E. Lloyd, E. Ehrenfeld, and M. J. Grubbs. 1998. Leader protein of foot-and-mouth disease virus is required for cleavage of the p220 component of the cap binding protein complex. *J. Virol.* **62**:4407–4409.
4. Domingo, E. 2000. Viruses at the edge of adaptation. *Virology* **270**:251–253.
5. Fauquet, C. M., M. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball. 2005. Virus taxonomy. Eighth report of the International Committee on Taxonomy of Viruses. Academic Press, London, United Kingdom.
6. Gibbs, M. J., J. S. Armstrong, and A. J. Gibbs. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**:573–582.
7. Haydon, D. T., A. D. Bastos, and P. Awadalla. 2004. Low linkage disequilibrium indicative of recombination in foot-and-mouth disease virus gene sequence alignments. *J. Gen. Virol.* **85**:1095–1100.
8. Haydon, D. T., A. D. Bastos, N. J. Knowles, and A. R. Samuel. 2001. Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. *Genetics* **157**:7–15.
9. Jackson, T., A. M. G. King, D. I. Stuart, and E. Fry. 2003. Structure and receptor binding. *Virus Res.* **91**:33–46.
10. Jain, R., M. C. Rivera, and J. A. Lake. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. USA* **96**:3801–3806.
11. King, A. M., D. McCahon, K. Saunders, J. W. Newman, and W. R. Slade. 1985. Multiple sites of recombination within the RNA genome of foot-and-mouth disease virus. *Virus Res.* **3**:373–384.
12. King, A. M., W. R. Slade, J. W. Newman, and D. McCahon. 1980. Temperature-sensitive mutants of foot-and-mouth disease virus with altered structural polypeptides. II. Comparison of recombination and biochemical maps. *J. Virol.* **34**:67–72.
13. Lukashev, A. N. 2005. Role of recombination in evolution of enteroviruses. *Rev. Med. Virol.* **15**:157–167.
14. Lukashev, A. N., V. A. Lashkevich, O. E. Ivanova, G. A. Koroleva, A. E. Hinkkanen, and J. Itonen. 2003. Recombination in circulating enteroviruses. *J. Virol.* **77**:10423–10431.
15. Lukashev, A. N., V. A. Lashkevich, O. E. Ivanova, G. A. Koroleva, A. E. Hinkkanen, and J. Itonen. 2005. Recombination in circulating Human enterovirus B: independent evolution of structural and non-structural genome regions. *J. Gen. Virol.* **86**:3281–3290.
16. Martin, D., and E. Rybicki. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**:562–563.
17. Martin, D. P., D. Posada, K. A. Crandall, and C. Williamson. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res. Hum. Retrovir.* **21**:98–102.
18. Martin, D. P., E. van der Walt, D. Posada, and E. P. Rybicki. 2005. The evolutionary value of recombination is constrained by genome modularity. *PLOS Genet.* **1**:475–479.
19. Martin, D. P., C. Williamson, and D. Posada. 2005. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**:260–262.
20. Martinez, M. A., J. Hernandez, M. E. Piccone, E. L. Palma, E. Domingo, N. Knowles, and M. G. Mateu. 1991. Two mechanisms of antigenic diversification of foot-and-mouth disease virus. *Virology* **184**:695–706.
21. Mateu, M. G., M. A. Martinez, L. Capucci, D. Andreu, E. Giralt, F. Sobrino, E. Brocchi, and E. Domingo. 1990. A single amino acid substitution affects multiple overlapping epitopes in the major antigenic site of foot-and-mouth disease virus of serotype C. *J. Gen. Virol.* **71**:629–637.
22. Maynard Smith, J. 1992. Analysing the mosaic structure of genes. *J. Mol. Evol.* **34**:126–129.
23. Murdin, A. D., H. H. Lu, M. G. Murray, and E. Wimmer. 1992. Poliovirus antigenic hybrids simultaneously expressing antigenic determinants from all three serotypes. *J. Gen. Virol.* **73**:607–611.
24. Oberste, M. S., S. Penaranda, and M. A. Pallansch. 2004. RNA recombination plays a major role in genomic change during circulation of coxsackie B viruses. *J. Virol.* **78**:2948–2955.
25. Padidam, M., S. Sawyer, and C. M. Fauquet. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**:218–225.
26. Palmenberg, A. C., and J. Y. Sgro. 2002. Alignments and comparative profiles of picornavirus genera, 149–155. *In* B. L. Semler and E. Wimmer (ed.), *The molecular biology of picornaviruses*. ASM Press, Washington, D.C.
27. Simmonds, P., A. Tuplin, and D. J. Evans. 2004. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: implications for virus evolution and host persistence. *RNA* **10**:1337–1351.
28. Simmonds, P., and S. Midgley. 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. *J. Virol.* **79**:15467–15476.
29. Simmonds, P., and J. Welch. 2006. Frequency and dynamics of recombination within different species of human enteroviruses. *J. Virol.* **80**:483–493.
30. Teterina, N. L., A. E. Gorbalenya, D. Egger, K. Bienz, M. S. Rinaudo, and E. Ehrenfeld. 2006. Testing the modularity of the N-terminal amphipathic helix conserved in picornavirus 2C proteins and hepatitis C NSSA protein. *Virology* **344**:453–467.
31. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**:4876–4882.
32. Tosh, C., D. Hemadri, and A. Sanyal. 2002. Evidence of recombination in the capsid-coding region of type A foot-and-mouth disease virus. *J. Gen. Virol.* **83**:2455–2460.
33. van Rensburg, H. G., T. M. Henry, and P. W. Mason. 2004. Studies of genetically defined chimeras of a European type A virus and a South African Territories type 2 virus reveal growth determinants for foot-and-mouth disease virus. *J. Gen. Virol.* **85**:61–68.
34. van Rensburg, H. G., D. Haydon, F. Joubert, A. D. Bastos, L. Heath, and L. Nel. 2002. Genetic heterogeneity in the foot-and-mouth disease virus Leader and 3C proteinases. *Gene* **289**:19–29.