Persistence of Multiple Genetic Lineages within Intrahost Populations of Ross River Virus[▽]

Wen J. Liu,¹ Michelle F. Rourke,¹ Edward C. Holmes,^{2,3} and John G. Aaskov^{1,4*}

Australian Army Malaria Institute, Brisbane, Australia¹; Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, Pennsylvania 16802²; Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892³; and Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia⁴

Received 17 December 2010/Accepted 14 March 2011

We examined the structure and extent of genetic diversity in intrahost populations of Ross River virus (RRV) in samples from six human patients, focusing on the nonstructural (nsP3) and structural (E2) protein genes. Strikingly, although the samples were collected from contrasting ecological settings 3,000 kilometers apart in Australia, we observed multiple viral lineages in four of the six individuals, which is indicative of widespread mixed infections. In addition, a comparison with previously published RRV sequences revealed that these distinct lineages have been in circulation for at least 5 years, and we were able to document their long-term persistence over extensive geographical distances.

Old world alphaviruses (e.g., Chikungunya, O'nyong-nyong, and Ross River) cause fever, rash, and arthritis in humans and are associated with infrequent, but major, outbreaks of disease (5). Ross River virus (RRV) infects more than 20 vertebrate hosts, causing clinical disease in humans and horses, and has been recovered from more than 20 mosquito species (18). Since the ability to generate genetic diversity (20) enables RNA viruses to rapidly exploit new ecological niches, including new host species, and may be a key determinant of virulence (9, 14), it is clearly of central importance to determine the extent and structure of genetic diversity in human pathogens such as RRV. In particular, it is important to determine what proportion of intrahost genetic variation is created *de novo* by mutation and what might be due to mixed infections of individual hosts with phylogenetically diverse lineages. To date, measures of genetic diversity in alphaviruses, including RRV, have relied on comparisons of population consensus sequences, often of isolates and occasionally from pools of mosquitoes that may contain more than one infected insect (16). There is only one report of an attempt to quantify genetic diversity in alphaviruses within individual hosts (26), and that study was constrained by the technology of the time (T1 nuclease fingerprinting) (6) and by the use of isolates that had been passaged *in vitro* rather than taken directly from host tissues.

In this study, we compared patterns of genetic diversity in both structural (E2) and nonstructural (nsP3) protein genes in intrahost populations of RRV in serum. The samples were taken from epidemic polyarthritis patients from various ecological settings thousands of kilometers apart in eastern and western Australia (Table 1). The nsP3 protein forms part of the complex of alphaviral nonstructural proteins that replicate

the viral genome (21). However, the C-terminal region of this protein is hypervariable, even within serotypes of alphaviruses (1, 13). The E2 protein attaches to receptors on host cells and contains most of the epitopes that are involved in neutralization by antibodies (12). While nsP3 and the other nonstructural proteins are translated directly from the viral genome, the structural proteins, including E2, are translated from subgenomic 26S RNA that is synthesized in infected cells (21).

Viral RNA was copied, amplified, cloned, and sequenced as described previously (2), except that RNA was extracted from isolates using a QIAamp viral minikit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and RRV oligonucleotide primers (1, 10) were substituted for those used previously with dengue virus (DENV). This approach utilized Expand polymerase (Roche), which is a mixture of *Taq* polymerase and the high-fidelity *Tgo* polymerase. Plasmids were linearized with 10 U of *Sac*II (New England BioLabs) overnight at 37°C before being purified (MinElute PCR purification kit; Qiagen), and inserts were sequenced at the Australian Genome Research Facility. The E2 and nsP3 genes of RRV were sequenced from the intrahost populations in the order PW7, PW14, SN39, PW11, SN11, and SN85 (PW7, PW14, etc., denote individual patients).

The nucleotide sequences of the majority of clones derived from both the E2 and nsP3 genes differed from their respective consensus sequences, with pairwise diversity values ranging from 0.08 to 0.72% (Table 2). As RRV PW7I, which was isolated from serum by culture on C6/36 *Aedes albopictus* cells, was consistently less diverse than the population in the serum from which it was derived (PW7S), no further analyses were performed with the isolates. Genetic diversity was significantly greater in nsP3 than in E2 from three of the six individuals, as was overall nucleotide diversity ($P < 0.01$, χ^2 test). The overall error frequencies in these data were 17×10^{-4} and 27×10^{-4} mutations per site for E2 and nsP3 genes, respectively (Table 2). Although these estimates inevitably contain some amplification-induced errors (3, 19), particularly those mutations that were observed only once in each sample, they are similar to

* Corresponding author. Mailing address: Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia. Phone: 61 7 3138 6193. Fax: 61 7 3138 6030. E-mail: j.aaskov@qut.edu.au.

[▽] Published ahead of print on 23 March 2011.

TABLE 1. Characteristics of localities from which strains of Ross River virus were recovered from patients in 2009 and used in this study

Patient	Locality	Approximate population	Geography
SN11	Nerang, central east coast, Australia	25,000	Urban
SN39	Yamba, central east coast, Australia	5,000	Coastal
SN85	Middlemount, central eastern Australia	5,000	Rural, inland
PW7	Esperance, southwest Australia	10,000	Coastal
PW11	Broome, northwest Australia	10,000	Coastal
PW14	Karratha, northwest Australia	10,000	Coastal

those obtained from intrahost populations of other mosquito-borne RNA viruses that cause acute infections, such as the dengue and West Nile viruses (2, 8, 25). Finally, although the values of the ratio of nonsynonymous-to-synonymous nucleotide substitutions per site (d_N/d_S) varied between both genes and individuals (Table 2), there was no evidence of positive selection in any instance. This was expected, given the short duration of viremia in RRV patients (less than 10 days [17]), and hence the limited time for allele fixation to occur, such that we were considering transient mutations rather than fixed substitutions. Indeed, that the intrahost d_N/d_S values (mean of 0.78) were consistently higher than those observed at the epidemiological scale (mean of 0.25 [10]) is consistent with the presence of a high frequency of transient deleterious mutations that have yet to be removed by purifying selection, as has been observed with other studies of intrahost viral diversity (7).

Highly polymorphic sites were distributed throughout the nsP3 and E2 genes/proteins (Fig. 1). All nsP3 genes contained the recently described duplicated RNA element in the 3' region (1, 10), which suggests that this change has successfully penetrated all members of RRV lineage 3. An nsP3 clone from

FIG. 1. Distribution of variable sites across the E2 (a) and nsP3 (b) nucleotide and amino acid sequences of Ross River virus.

TABLE 2. Measures of intrahost genetic diversity of Ross River virus from epidemic polyarthritides patients

Strain ^a	Gene/protein	No. of clones sequenced	No. of clones with consensus sequence	Total no. of variable nt sites	Total no. of variable aa sites	Total no. of nt changes ^b	Total no. of aa changes	Mean pairwise divergence (%)	d_N/d_S
PW7S (serum)	E2	16	1	48	35	50	37	0.49	1.43
	nsP3	19	3	32	21	37	25	0.23	0.97
PW7I (isolate)	E2	22	12	12	9	13	10	0.09	1.34
	nsP3	23	11	19	11	19	11	0.10	0.61
PW11	E2	19	3	40	28	49	35	0.40	0.97
	nsP3	22	1	55	36	150	49	0.60	0.61
PW14	E2	22	14	10	7	11	8	0.08	1.03
	nsP3	22	5	38	21	40	22	0.21	0.51
SN11 ^c	E2	0	0	0	0	0	0	0	0
	nsP3	22	2	62	27	161	49	0.72	0.30
SN39	E2	21	9	31	14	33	15	0.25	0.40
	nsP3	21	3	49	33	49	33	0.29	0.80
SN85	E2	20	6	53	28	70	39	0.53	0.53
	nsP3	24	0	112	67	143	84	0.70	0.67
Total E2 serum						213			
Total nsP3 serum						580			

^a Refer to Table 1 for details of patients from whom viruses were derived.

^b Values for nsP3 and for E2 in boldface type are significantly different ($P < 0.05$, chi-square test).

^c There was insufficient serum for testing from patient SN11.

FIG. 2. Phylogenetic relationships between E2 (a) and nsP3 (b) genes from intrahost samples of Ross River virus from patients from urban and rural settings in Australia. Note the presence of two clearly distinct lineages. RRV (consensus) sequences generated from previous studies are in red. Individuals that harbor each of the phylogenetically distinct lineages are marked with an asterisk. The phylogenetic position of strain QML1 is also indicated. Both trees are mid-point rooted for purposes of clarity only, and all horizontal branch lengths are drawn to a scale of nucleotide substitutions per site. Bootstrap support values (>70%) are shown for key nodes.

SN85 had a nucleotide change that converted codon 546 from an OPAL stop codon (UGA) to arginine (CGA), confirming an earlier, unpublished observation that this site was polymorphic in RRV recovered from an epidemic polyarthritis patient. Previous studies of a related arthrogenic alphavirus, O'nyong-nyong, identified a similar polymorphism in the nsP3 gene but attributed it to passage *in vitro* in vertebrate cells (11). Given that all previous assignments of an OPAL codon at this position have been based on consensus sequences and rarely in-

involved viruses studied in tissues from natural hosts, it is possible that all alphaviruses may be polymorphic at this site and that the ratio of OPAL-to-arginine codons varies with the host and or host tissue. With respect to E2, several single-nucleotide polymorphisms gave rise to significant changes. Specifically, that at nucleotide (nt) 706 in a clone from SN85 gave rise to a stop codon, and that at nt 790 in clones from PW11 and SN85 generated a threonine-alanine substitution at E2-264, which would have disrupted a glycosylation motif at this position (1).

One of the four most polymorphic sites in E2 (Fig. 1), that at E2-275, occurred in a region of significant E2-E1 intraspikes contact in alphaviruses (23). All amino acid substitutions in known serological epitopes (24) were conservative. The lack of variation in serological epitopes in the E2 protein, which is a major target for neutralizing antibodies, bodes well for the efficacy of an RRV vaccine currently in clinical trials.

The most striking observation from this study was that both the E2 and nsP3 genes contained at least two phylogenetically distinct and well-supported lineages that were found in different patients (Fig. 2). Specifically, a maximum likelihood phylogenetic tree estimated using the PAUP* program (22) (parameter values were determined using ModelTest [15] and are available from the authors on request) revealed that although all the sampled viruses could be classified within lineage 3 of RRV (10), each of three of the individuals with the E2 gene (PW11, SN39, and SN85) and two of the individuals with the nsP3 gene (SN11 and SN85) had two divergent viral lineages, indicative of mixed infections. Although individual SN85 was the only one to harbor phylogenetically distinct lineages in both E2 and nsP3, comparison of the E2 and nsP3 phylogenies was complicated by the relatively small number of samples from each patient and the lack of phylogenetic resolution in parts of the nsP3 tree. Importantly, however, because each lineage is defined by multiple mutations, their presence cannot be explained by experimental error, such as that occurring with RT-PCR and sequencing. Equally striking was that isolate QML1, from a sample taken from an individual in eastern Australia in 2004 (10), fell within the bounds of genetic diversity of the viruses sampled here in 2009, particularly that from individual SN85. Such a phylogenetic pattern indicates that multiple viral lineages have cocirculated (beneath the consensus) within RRV-infected humans for at least 5 years (i.e., they must have diverged prior to the isolation of QML1 in 2004) and perhaps for considerably longer. In addition, these lineages have spread extensively across Australia, as shown by the presence of the two distinct lineages in patients PW11 and SN39, from whom samples were taken almost 3,000 kilometers apart.

While there is growing evidence for mixed infections with RNA viruses that cause acute, self-limiting infections, such as eastern equine encephalomyelitis virus (26) and influenza (4), we know of no other case in which multiple viral populations have been transmitted among individuals for the extended time period documented here. In addition, although the fitness correlates, if any, of these distinct viral lineages are unknown, their persistence in time and space indicates that the population bottlenecks that occur at interhost virus transmission may not be especially narrow.

We thank the staff from Pathwest, Sullivan and Nicolaides Pathology and Queensland Medical Laboratories for providing the sera used in this study.

This study was supported by grants from the National Health and Medical Research Council of Australia and the Cook Estate. E.C.H.

was funded, in part, by grant R01GM80533-04 from the National Institutes of Health.

REFERENCES

1. Aaskov, J. G., A. Jones, W. Choi, K. Lowry, and E. E. Stewart. 2011. Lineage replacement accompanying duplication and rapid fixation of an RNA element in the nsP3 gene in a species of alphavirus. *Virology* **410**:353–359.
2. Aaskov, J. G., K. Buzacott, H. M. Thu, K. Lowry, and E. C. Holmes. 2006. Long-term transmission of defective RNA viruses in humans and *Aedes* mosquitoes. *Science* **311**:236–238.
3. Bracho, M. A., A. Moya, and E. Barrio. 1998. Contribution of *Taq* polymerase-induced errors to the estimation of RNA virus diversity. *J. Gen. Virol.* **79**:2921–2928.
4. Ghedin, E., et al. 2009. Mixed infection and the genesis of influenza diversity. *J. Virol.* **83**:8832–8841.
5. Griffin, D. E. 2007. Alphaviruses, p. 1023–1067. In D. M. Knipe et al. (ed.), *Fields virology*, 5th ed. Lippincott Williams and Wilkins, Philadelphia, PA.
6. Holland, J. J., E. A. Grabau, C. L. Jones, and B. L. Semler. 1979. Evolution of multiple genome mutations during long-term persistent infection by vesicular stomatitis virus. *Cell* **16**:495–504.
7. Holmes, E. C. 2009. The evolution and emergence of RNA viruses. Oxford Series in Ecology and Evolution. Oxford University Press, Oxford, United Kingdom.
8. Jerzak, G., K. A. Bernard, L. D. Kramer, and G. D. Ebel. 2005. Genetic variation in West Nile virus from naturally infected mosquitoes and birds suggests quasispecies structure and strong purifying selection. *J. Gen. Virol.* **86**:2175–2183.
9. Jerzak, G. V., K. Bernard, L. D. Kramer, P. Y. Shi, and G. D. Ebel. 2007. The West Nile virus mutant spectrum is host-dependant and a determinant of mortality in mice. *Virology* **360**:469–476.
10. Jones, A., K. Lowry, J. Aaskov, E. C. Holmes, and A. Kitchen. 2010. Molecular evolutionary dynamics of Ross River virus and implications for vaccine efficiency. *J. Gen. Virol.* **91**:182–188.
11. Lanciotti, R. S., et al. 1998. Emergence of epidemic O'nyong-nyong fever in Uganda after a 35-year absence: genetic characterization of the virus. *Virology* **252**:258–268.
12. Mukhopadhyay, S., et al. 2006. Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses. *Structure* **14**:63–73.
13. Oberste, M. S., M. D. Parker, and J. F. Smith. 1996. Complete sequence of Venezuelan equine encephalitis virus subtype 1E reveals conserved and hypervariable domains within the C terminus of nsP3. *Virology* **219**:314–320.
14. Pfeiffer, J. K., and K. Kirkegaard. 2005. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog.* **1**:e11.
15. Posada, D., and K. A. Crandell. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
16. Powers, A. M., et al. 2001. Evolutionary relationships and systematics of the alphaviruses. *J. Virol.* **75**:10118–10131.
17. Rosen, L., D. J. Gubler, and P. H. Benett. 1981. Epidemic polyarthritis (Ross River) virus infection in the Cook Islands. *Am. J. Trop. Med. Hyg.* **30**:1294–1302.
18. Russell, R. C. 2002. Ross River virus: ecology and distribution. *Annu. Rev. Entomol.* **47**:1–31.
19. Smith, D. B., J. McAllister, C. Casino, and P. Simmonds. 1997. Virus “quasispecies”: making a mountain out of a molehill? *J. Gen. Virol.* **78**:1511–1519.
20. Steinhauer, D. A., E. Domingo, and J. J. Holland. 1992. Lack of evidence for proofreading mechanisms associated with an RNA polymerase. *Gene* **122**:281–288.
21. Strauss, J. H., and E. G. Strauss. 1994. The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* **58**:491–562.
22. Swofford, D. L. 2003. PAUP*: phylogenetic analysis using parsimony (and other methods), version 4. Sinauer Associates, Sunderland, MA.
23. Voss, J. E., et al. 2010. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* **468**:709–712.
24. Vratil, D. C., C. A. Fernon, L. Dalgarno, and R. C. Weir. 1988. Location of a major antigenic site involved in Ross River virus neutralization. *Virology* **162**:346–353.
25. Wang, W.-K., S.-R. Lin, C.-C. King, and S.-C. Chang. 2002. Dengue type 3 virus in plasma is a population of closely related genomes: quasispecies. *J. Virol.* **76**:4662–4665.
26. Weaver, S. C., et al. 1993. Diversity within natural populations of eastern equine encephalomyelitis virus. *Virology* **195**:700–709.