Virus movement maintains local virus population diversity

Jamie C. Snyder*†, Blake Wiedenheft*†, Matthew Lavin‡, Francisco F. Roberto§, Josh Spuhler*†, Alice C. Ortmann*‡, Trevor Douglas*¶, and Mark Young*†‡

*Thermal Biology Institute and Departments of Microbiology, †Plant Sciences and Plant Pathology, and §Chemistry and Biochemistry, Montana State University, Bozeman, MT 59717; and ‡Idaho National Laboratory, Idaho Falls, ID 83415

Communicated by George Bruening, University of California, Davis, CA, October 5, 2007 (received for review January 4, 2007)

Viruses are the largest reservoir of genetic material on the planet, yet little is known about the population dynamics of any virus within its natural environment. Over a 2-year period, we monitored the diversity of two archaeal viruses found in hot springs within Yellowstone National Park (YNP). Both temporal phylogeny and neutral biodiversity models reveal that virus diversity in these local environments is not being maintained by mutation but rather by high rates of immigration from a globally distributed metacommunity. These results indicate that geographically isolated hot springs are readily able to exchange viruses. The importance of virus movement is supported by the detection of virus particles in air samples collected over YNP hot springs and by their detection in metacommunity sequencing projects conducted in the Sargasso Sea. Rapid rates of virus movement are not expected to be unique to these archaeal viruses but rather a common feature among virus metacommunities. The finding that virus immigration rather than mutation can dominate community structure has significant implications for understanding virus circulation and the role that viruses play in ecology and evolution by providing a reservoir of mobile genetic material.

Results

Hot springs studied covered a large area with an average distance of >30 km between them. We found that each hot spring maintained its high-temperature and acidic nature during the 2-year monitoring period (Fig. 1). Each hot spring had its own unique water chemistry, which did not change appreciably over the monitoring period, and received little input from surface water. Additionally, DNA sequencing of 16S rRNA genes revealed that each hot spring supported its own unique collection of seven to nine major archaeal genera, with Sulfolobus being the only genus common to all three sites (Fig. 2).

SVIR sequences were amplified in all three sites from 39 of the 40 total sampling events. Sequencing of 2,165 independent clones revealed 722 unique sequences that formed 12 well supported clades (Fig. 3A). In this study, a clade is representative of a distinct virus species, phylogenetically distinct from other species, but containing a low level of variation. Five of these clades (A–C, I, and K) were detected in all three hot springs. The remaining seven clades (D–H, J, and L) were only detected in one of the three hot springs (Fig. 3B). SVIR-like sequences were amplified from all three monitor sites, but only from 21 of the 40 total sampling events. Sequencing of 779 independent SVIR clones revealed 320 unique sequences that formed six distinct clades (Fig. 4A). The three largest clades (A–C) were represented in all three hot springs, whereas three smaller clades (D–F) were only found in a single hot spring (Fig. 4B). All four previously sequenced SVIR isolates from thermal features on three different continents nested within the YNP clades, whereas the SVIR sequences from Iceland formed clades independent of the Yellowstone clades. Sequence accumulation curves for both SVIR and SVSV are not


The authors declare no conflict of interest.

3To whom correspondence may be addressed. E-mail: tldouglas@chemistry.montana.edu or myoung@montana.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0709445104DC1.

© 2007 by The National Academy of Sciences of the USA
being the only common, but not the dominant, resident in each of the three hot springs. A, *Sulfolobus*

### Fig. 1. General water chemistry of each hot spring monitor site. Although not entirely static, the overall geochemical signature of each spring was distinct and relatively stable over the 2-year sampling course. Pie charts illustrate the distinct geochemical signature of each hot spring. (A) CHMS. (B) RCMS. (C) RHMS. Average values are reported in the table.

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>pH</th>
<th>Cl^-</th>
<th>SO_4^{2-}</th>
<th>DOC</th>
<th>NH_3</th>
<th>P</th>
<th>Al</th>
<th>Sb</th>
<th>As</th>
<th>B</th>
<th>Ca</th>
<th>Fe</th>
<th>Pb</th>
<th>Mn</th>
<th>Mo</th>
<th>K</th>
<th>Si</th>
<th>Na</th>
<th>Zn</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHMS</td>
<td>79.2</td>
<td>2.2</td>
<td>1.8</td>
<td>10.2</td>
<td>2.5</td>
<td>14.0</td>
<td>0.07</td>
<td>20.0</td>
<td>0</td>
<td>0.01</td>
<td>1.4</td>
<td>0</td>
<td>1.47</td>
<td>0.01</td>
<td>0.02</td>
<td>0</td>
<td>9</td>
<td>108</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>RCMS</td>
<td>79.3</td>
<td>3</td>
<td>8.7</td>
<td>5.3</td>
<td>5.3</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0</td>
<td>0.04</td>
<td>1.5</td>
<td>12.2</td>
<td>0</td>
<td>0.08</td>
<td>0</td>
<td>5.2</td>
<td>24.9</td>
<td>11.4</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>RHMS</td>
<td>189.7</td>
<td>0.11</td>
<td>125</td>
<td>2.5</td>
<td>0.86</td>
<td>0.86</td>
<td>1.9</td>
<td>0.125</td>
<td>1.0</td>
<td>7.0</td>
<td>5.0</td>
<td>2.7</td>
<td>0</td>
<td>0.03</td>
<td>0.18</td>
<td>0.65</td>
<td>106.5</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

asymptotic, indicating that more SSV and SIRV local diversity remains to be discovered [see supporting information (SI) Fig. 8]. The majority of both SIRV and SSV clades were detected during the first year of sampling, indicating that most of the dominant SIRV and SSV clades were observed. In addition, the observation of a diminishing and limited number of viral clades indicates that the differential viral clade detection was not a sampling artifact.

The SSV or SIRV clades detected in a particular hot spring could be rapidly replaced (Figs. 3 and 4). For example, the dominant SIRV clade A in the Crater Hills Monitor Site (CHMS) hot spring was completely replaced within 167 days by clades B, H, J, and I, and T, only to reappear as the dominant clade 55 days later. SSV populations follow a similar trend. In the Ragged Hills Monitor Site (RHMS) hot spring, the dominant SSV clade A was replaced within 66 days by clades B and C. There was no clear relationship between the appearance of a clade and its relative abundance in the clone library.

The rank abundance of all subsets of SIRV and SSV sequences were fitted against multiple distribution models. The best fit model was consistently Etienne’s sampling formula (EtSF) (Fig. 5). In contrast to other studies (14), power law functions (i.e., Zipf and Mandelbrot) fit less well than EtSF. Unlike traditional models of rank abundance (e.g., the lognormal), the EtSF (15), a genealogical modification of Hubbell’s zero-sum multinomial (16), uses biologically realistic parameters [e.g., migration and speciation as used by Hubbell (16)] for analyzing virus diversity within these hot springs. Central to this theory is the interplay of virus speciation (virus clade formation) and migration during the generation of metacommunity diversity. Where immigration is negligible, the number of new virus clades produced by the virus community per generation, $\Theta$, is a function dominated by mutation. When the probability of resident extinction being replaced by a migrant ($m$) becomes high, immigration becomes the main cause of local virus diversity. With EtSF, the parameter reflecting the biodiversity-generating capacity of the SIRV and SSV metacommunities, $\Theta$, was estimated at 575 new lineages per metacommunity generation, and an immigration rate of $m$ was estimated at 1.0. A model estimating only $\Theta$, however, shows a better fit, strongly suggesting that these local YNP hot springs are not dispersal-limited (i.e., $m = 1.0$).

Independent analyses were conducted to examine the decay in community composition and phylogenetic distances through time. In both analyses (Figs. 6 and 7), we detected a weak, but significant, pattern with regard to SIRV and SSV viral clade turnover through time. For both analyses, three models were fit to the data: (i) comparing among hot springs including days between sampling, (ii) comparing within hot springs including days between sampling, and (iii) including only days between sampling. The model including just days between sampling events provided a better fit than the other two models, which included the sampling location as a proxy for water geochemistry and archaeal host composition. For the community composition analysis, this finding suggests that water geochemistry and the archaeal community explain little, if any, of the change in viral community composition and phylogenetic distances through time. In both analyses (Figs. 6 and 7), we detect a weak, but significant, pattern with regard to SIRV and SSV viral clade turnover through time. For both analyses, three models were fit to the data: (i) comparing among hot springs including days between sampling, (ii) comparing within hot springs including days between sampling, and (iii) including only days between sampling. The model including just days between sampling events provided a better fit than the other two models, which included the sampling location as a proxy for water geochemistry and archaeal host composition. The decay in community phylogenetic distance through time is weaker than the decay of community composition distance, which is consistent with our observations of SIRV and SSV clade turnover.

### Fig. 2. Comparison of the archaeal community detected at each of the hot springs. (A) CHMS. (B) RHMS. (C) RCMS. Pie charts illustrate the average relative abundance of each genus detected by using 16S rRNA gene analysis. The community composition was relatively stable over the 2-year time course, with *Sulfolobus* being the only common, but not the dominant, resident in each of the three hot springs. A, *Acidianus*; C, *Caldococcus*; D, *Desulfurococcus*; M, *Metallosphaera*; P, *Pyrodictium*; T, *Thermocladium*; V, *Vulcanisaeta*; S, *Stygiolobus*; O, other.
We investigated possible mechanisms for virus movement to and from YNP hot springs. Two obvious movement mechanisms are of viruses through the air or by long-distance underground water movement between hot springs. We speculate that transport through the subsurface waters is unlikely because of the high temperatures of these waters in YNP (>350°C at 100 m) and the unlikely prospect of direct subsurface hydrological connection on an intercontinental scale. In contrast, air transport of microbes is known to occur on a global scale, and virus movement on a local scale is well established (i.e., movement of respiratory viruses through the air and independent of any virus vector species). We were able to detect viral DNA sequences in the air column above a hot spring by using amplification-detection methods with viral-specific primers. These findings suggest that virus transport through the air is a possible mechanism for rapid virus movement between hot springs.

Discussion

Our findings demonstrate that virus movement, not mutation, is the dominant factor controlling the change in viral community
structure. These findings were only possible because of our temporal sampling approach at the local scale. The restricted number of SIRV and SSV clades detected early on during this study suggests globally limited diversity. The high immigration rates estimated with EtSF indicate that the local diversity may closely represent global diversity. Our results strongly suggest that geographically isolated hot springs are readily able to exchange viruses perhaps on a global scale. This finding is in stark contrast to the host of SIRV and SSV, *Sulfolobus*, which has been shown to be geographically limited in its dispersal (13).

We do not believe that the high rate of virus turnover at the three sampled hot springs represents an artifact of inadequate sampling for several reasons. First, only a limited number of viral clades were detected, and all main clades were detected during the first year of sampling. Second, we observed a nonrandom pattern of decay in community composition and phylogenetic diversity through time (Figs. 6 and 7), which is an expectation of neutral biodiversity theory (16).

It is unlikely that the observed viral diversity is being maintained solely by balancing selection of viral clades already present in each hot springs. The shape of the sequence rank abundance curve, with its long singleton sequence tail, is not consistent with balancing selection because there are too many rare types and a high abundance of a few common types (Fig. 5). The maintenance of this type of viral sequence distribution requires the continual input of viral sequences from outside the hot springs. Other types of selection also are unlikely. Each hot spring has a distinct geochemical environment and archaeal community composition that remains relatively steady over time. Unaccounted environmental parameters responsible for dynamic virus population shifts are marginalized by the observation of the same dominant SSV and SIRV clades appearing in all three hot springs despite dramatic differences in their water chemistry and archaeal host composition (Figs. 3B and 4B). If new viral clades were being selected from the low abundance population already present within each individual hot spring, we would expect to observe the appearance of viral clades unique to each hot spring owing to unique selection pressures operating in these three different hot springs. In general, this finding is not what we observe. Instead, we observe common viral clades appearing and disappearing from each hot spring in a stochastic environment.

**Fig. 5.** Rank abundance distribution of 2,944 sequences in 1,042 sequence types for SIRV and SSV sampled at all three hot springs during the ~2-year study. Distributions and their AIC values are shown, with EtSF being the best fit. The abundances of the five most abundant genotypes are reported next to the corresponding point.

**Fig. 6.** Community composition of SIRV and SSV as a function of time. Pairwise comparisons of community composition (Jaccard’s distance) for SIRV (open circles) and SSV (bullets) are shown. The models including within-hot-springs (black lower lines) and among-hot-springs (upper red lines) comparisons along with days between sampling are illustrated for SIRV and SSV. The center lines depict the model, including only days between sampling. The 95% confidence intervals for the mantel correlation of each of the two models were essentially identical.

**Fig. 7.** Phylogenetic relatedness of SIRV and SSV as a function of time. The phylogenetic distance (net-relatedness index) between samples compared with days between sampling are shown for both SIRV (open circles) and SSV (filled circles). The upper (red) and lower (black) lines represent the models, including both days between sampling events and among-hot-springs or within-hot-springs comparisons. The model including only days between sampling events is shown by the center line. The 95% confidence intervals for the mantel correlation of each of the two models were essentially identical.

**Table:**

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>EtSF (G)</td>
<td>1229.93</td>
</tr>
<tr>
<td>EtSF (G, m)</td>
<td>1231.93</td>
</tr>
<tr>
<td>Zipf</td>
<td>2698.80</td>
</tr>
<tr>
<td>Mandelbrot</td>
<td>2900.60</td>
</tr>
<tr>
<td>Lognormal</td>
<td>4950.47</td>
</tr>
<tr>
<td>Broken stick</td>
<td>8647.29</td>
</tr>
<tr>
<td>Preemption</td>
<td>9839.20</td>
</tr>
</tbody>
</table>

**Model AIC estimates**

- **Theta:** 575 ± 10
- **m:** 1.0 ± 1.71E-12

**870 singletons**

**Log abundance (sequences)**

- **y = 1.0480x + 4.2, adj R² = 0.1242 (p = 2.2e-16)**
- **y = 0.0193x + 56.8, adj R² = 0.0038 (p = 0.1824)**

**Snyder et al.**

*PNAS* November 27, 2007 vol. 104 no. 48 19105

**MICROBIOLOGY**
fashion. The traditional explanations for virus biodiversity, including temporal shifts in host availability (12) and habitat heterogeneity (17), do not adequately explain the temporal clade shifts that we observe in both thermophilic virus populations.

Selection for a high viral mutation rate is discounted by the lack of a strong positive relationship between community phylogenetic relatedness and time (Fig. 7). Selection would render samples taken from surviving lineages as highly related (18). If mutation were primarily driving clade dynamics within a local hot spring, we would expect to see a terminal branch ramify through time. We do not observe this pattern. Instead, we observe entire clades disappearing and being replaced (Figs. 3B and 4B). Using mutation rates reported for dsDNA viruses of approximately one misincorporation per $10^{2}$-$10^{9}$ bp, one cannot achieve the observed local sequence diversity by mutation alone.

Our analysis reveals that three small YNP hot springs harbor a highly diverse local virus community that is not dispersal-limited and probably part of a virus metacommunity that is global in extent. High immigration rates and the potentially large size of the local virus community were predicted from the phylogenetic analysis. We observed no geographic structure, $870$ terminals sampled once (after biasing against sampling this class), and resident clades being replaced by immigrant clades. Our observed concordance between phylogeny and neutral modeling indicates extensive virus exchange among these hot springs and potentially between globally distributed thermal features. This finding also is revealed by the close phylogenetic relatedness of globally distributed SSV isolates (Fig. 4A). With immigration rates estimated at the maximum (a resident death has a 100% probability of being replaced by an immigrant), a local viral community should be a fair representation of the global metacommunity (14), which would be consistent with other studies of viral diversity (12). Each hot spring can be viewed as a small island incapable of representing the entire viral metacommunity during a single sampling event. The hot springs viral metacommunity must be extremely large while distributed among a highly fragmented island-like hot spring landscape. Immigration rates of the magnitude we are suggesting may be common among thermal viruses and may account for the detection of Sulfolobus virus types in anomalous settings such as the Sargasso Sea (19).

The impact of virus immigration on virus population dynamics has not been recognized previously because traditional approaches have typically involved culture-dependent methods and limited resampling of the same environment over an extended time period. Our results demonstrate how culture-independent approaches, combined with spatial and temporal sampling, can identify rapid virus movement as a key determinant in maintaining local virus diversity. If our observation of virus movement proves to be general, then understanding the ecology and evolutionary causes of high immigration rates is essential.

Materials and Methods

Hot Spring Monitoring Sites. Three high-temperature (~80°C) acidic (pH ~3.0) sites were selected in YNP (see SI Fig. 9): CHMS (44°39.31′N, 110°28.95′W), Rabbit Creek Monitor Site (RCMS) (44°31.287′N, 110°48.647′W), and RHMS (44°43.653′N, 110°42.862′W).

Sample Collection and DNA Isolation. Total environmental DNA was extracted from 500-ml samples collected from each hot spring, at each time point, by using the UltraCleanWater DNA Extraction Kit (Mo Bio Laboratories).

Water Chemistry. Hot spring water samples were collected and preserved according to Energy Laboratories. Each sample was analyzed for the reported inorganics (chloride and sulfate), organics (dissolved organic carbon), nutrients (ammonia, total nitrogen, and phosphorus), and metals (aluminum, antimony, arsenic, barium, boron, cadmium, calcium, copper, iron, lead, magnesium, manganese, mercury, molybdenum, nickel, potassium, selenium, silicon, sodium, and zinc).

PCR Amplification, Cloning, and Sequencing of Archaeal and Viral Sequences. Archaeal 16S rRNA gene sequences were amplified from total environmental DNA by using archaeal-specific primers 20F (TTCGGGTGTTAGCTGTCCGG) or 348Fs (TC-CAGGCCCTACGGG), paired with the universal primer U14906R (GACGGGCCGGTGTGTTGTRCA) (20). SSV and SIRV sequences were PCR-amplified by using primers identified by comparative virus genomic studies (see SI Fig. 10). Universal SSV primers UnivSSV 3F (CAATCGCCATAAGGTCAGGGTCAGGG) and UnivSSV 4R [CGTTTA(C/T)TACTATAACGTA] were designed by using the complete genome sequences of four SSV isolates (10). These primers amplify ~256 bp of the largest conserved ORF whose function is yet unknown. SIRV primers DBPF (GATAATGACAAATTGGCAAAAAGG) and DBPR (GCTTAATATTATTTAACCTAGATATCC) amplify the entire 405 bp of the coat protein gene. These primers were designed by using the complete genome sequences of two Icelandic isolates (SIRV1 and SIRV2) (21) and partial genome sequences from two YNP isolates (data not shown). PCR products were cloned into pCR2.1 by using TOPO-TA cloning protocols (Invitrogen) and sequenced by using Big Dye Termination protocols on an ABI 3700 automated capillary sequencer (Applied Biosystems).

Phylogenetic Analysis. DNA sequences were manually edited by using Sequerencer 4.2.2 (Gene Codes). Sequence alignments were performed by using CLUSTALX (22). The alignments were subsequently analyzed by PAUP version 4.0b10 (23), and a maximum parsimony analysis was conducted. Bootstrap values were obtained by resampling the data 10,000 times, and a bootstrap 50% majority rule consensus tree was produced. An independent maximum likelihood analysis was conducted by using MrBayes (24). Clade credibility values were derived by using 2 million permutations sampled every 10,000 generations. To reduce the effects of possible sequencing errors, only sequences that contained more than one base change and did not code for stop codons when translated in the appropriate translational reading frame were used in subsequent phylogenetic and rank abundance analysis.

Dissimilarity Index, Rank Abundance Modeling, and Estimations of Migration and Metacommunity. Modeling of sequence abundance curves was performed with the Radfit program in the Vegan package (25), which is written for the R statistical environment (26). Abundances also were fit to the zero-sum multinomial distribution (16) by using Bayesian methods in R code (27) and EeSF (28) by using the program Tetame (15). Temporal (Euclidean) and community (Jaccard’s) distances were generated with the base package in R, as well as with Vegan. Community phylogenetic distances were generated with Phylom (29) by using the net-relatedness index. Regression analyses and mantel tests were performed in R.

Detection of Viral Sequences in the Air Column. An all-glass impinger air sampler [SKC model 224-PCXRS AirChek portable sampling pump, serial no. 584623; SKC Glass Midget Impinger, 25 ml (standard nozzle); and spill-resistant SKC Glass Midget Impinger, 25 ml (fritted nozzle), trap operating at 3 liters/min for 100 min] was used to sample the air column above the CHMS hot spring. A double 0.8/0.45-μm filter (Millipore) was placed in line on the air sampler intake to remove cells, and resulting filtered air passed through 25 ml of
sterile NMN buffer [20 mM NaPO₄ (pH 5.5), 100 mM NaCl, and 1 mM MgCl₂]. After sampling, the buffer was removed, and 1 μl was used as a template for amplification with GenomiPhi (Amersham Biosciences), followed by PCR amplification by using primers specific for SSV: SIRV, Sulfolobus turreted icosahedral virus [(STIV) C651f (GCAATTAGCT-GCGAACATCA) and C1457r (CCGGGTAAGTTGGTTTGAGA) amplifying a portion of C557, a structural protein in STIV], SSV, or SIRV.

This work was supported by National Science Foundation Grant MCB 01322156.

Figure 6.6. A sequence accumulation curve for the SSV-like viruses detected over time in all three monitor sites. This indicates the total diversity of SSV-like sequences in YNP was not captured.

Figure 6.10. A species accumulation curve for the SIRV-like viruses detected over time in all three monitor sites. This indicates the total diversity of SSV-like sequences in YNP was not captured.
...AATAACCGGTAGTCCAACTGGATACGTAAGTTA------------------

*RC7#4 and 9Y-NL are two partially sequenced isolates from YNP.*