doi:10.1111/j.1558-5646.2010.01040.x

GEOGRAPHIC DIFFERENCES IN SEXUAL REASSORTMENT IN RNA PHAGE

Kara J. O'Keefe,[1,2] Olin K. Silander,[3] Helen McCreery,[4] Daniel M. Weinreich,[5] Kevin M. Wright,[6] Lin Chao,[7] Scott V. Edwards,[2] Susanna K. Remold,[8] and Paul E. Turner[1,9]

[1]Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520-8106

[2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

[3]Core Program Computational and Systems Biology, Biozentrum, University of Basel, Basel, Switzerland

[4]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

[5]Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912

[6]Department of Biology, Duke University, Durham, North Carolina 27708

[7]Section of Ecology, Evolution and Behavior, University of California, San Diego, La Jolla, California 92093

[8]Department of Biology, University of Louisville, Louisville, Kentucky 40292

[9]E-mail: paul.turner@yale.edu

Received November 5, 2008
Accepted April 20, 2010

The genetic structure of natural bacteriophage populations is poorly understood. Recent metagenomic studies suggest that phage biogeography is characterized by frequent migration. Using virus samples mostly isolated in Southern California, we recently showed that very little population structure exists in segmented RNA phage of the *Cystoviridae* family due to frequent segment reassortment (sexual genetic mixis) between unrelated virus individuals. Here we use a larger genetic dataset to examine the structure of *Cystoviridae* phage isolated from three geographic locations in Southern New England. We document extensive natural variation in the physical sizes of RNA genome segments for these viruses. In addition, consistent with earlier findings, our phylogenetic analyses and calculations of linkage disequilibrium (LD) show no evidence of within-segment recombination in wild populations. However, in contrast to the prior study, our analysis finds that reassortment of segments between individual phage plays a lesser role among cystoviruses sampled in New England, suggesting that the evolutionary importance of genetic mixis in *Cystoviridae* phage may vary according to geography. We discuss possible explanations for these conflicting results across the studies, such as differing local ecology and its impact on phage growth, and geographic differences in selection against hybrid phage genotypes.

KEY WORDS: **Biogeography, genetic variation, hybridization, phylogenetics, population structure, sex.**

Bacteriophage (phage) are perhaps the most abundant and diverse biological entities on Earth (Fuhrman 1999; Breitbart et al. 2002). Details of their ecology reveal that phage exist in all habitable environments, and that phage play fundamental roles in the functioning of natural ecosystems (Abedon 2008). However, relatively little is known of the population biology and genetics of phage in natural habitats; most of the existing studies address community-level questions, such as the impact of phage on the dynamics of their host bacteria (Wommack and Colwell 2000; Suttle 2005). Additionally, the majority of recent studies have only investigated marine and hot-spring ecosystems (Angly et al. 2006; Snyder et al. 2007; Desnues et al. 2008). Because phage almost certainly constitute the majority of Earth's microbes, it becomes vital to better understand the population biology of phage

in other natural environments, and to elucidate whether geography is prominent in dictating their population structure.

We have recently initiated population biology studies of natural phage isolates from the family *Cystoviridae* (Silander et al. 2005). Cystoviruses characteristically feature a ~13 kb double-stranded RNA genome split up into three separate segments (referred to as L, M, and S, for "large," "medium," and "small," respectively). The genome is enclosed within a polyhedral nucleocapsid surrounded by a lipid outer envelope (Butcher et al. 1997; Mindich 1999; Laurila et al. 2002). The best-known cystovirus is phage ϕ6, isolated over 30 years ago in Nebraska, USA on *Pseudomonas*-infested bean straw (Vidaver et al. 1973) (A.K. Vidaver, pers. comm.). In the laboratory, phage ϕ6 is typically cultured on the plant pathogenic bacterium *Pseudomonas syringae* pathovar *phaseolicola* (*P. phaseolicola*), isolated contemporaneously from bean straw. Phage ϕ6 causes lytic (lethal) infection of bacterial cells. Although it is uncertain which Pseudomonads or other host bacteria cystoviruses primarily infect in the wild, these phage can be readily isolated by sampling from clover, as well as green beans and other vegetables (Mindich et al. 1999; Silander et al. 2005).

One interesting feature of cystoviruses is their ability to undergo segment reassortment, a form of sex (genetic mixis) whereby multiple virus genotypes that co-infect the same cell can exchange segments to create hybrid progeny (Turner et al. 1999; Mindich 2004). It is well known that RNA viruses typically experience elevated mutation rates, owing to the lack of error correction during RNA replication. But reassortment is an additional process that can provide the genetic variation which fuels evolution via natural selection in populations of segmented RNA viruses; reassortment is a genetic-exchange mechanism recognized as important in the evolution of pathogenic viruses such as influenza viruses, and hantaviruses (Rizvanov et al. 2004; Xu et al. 2004). For these reasons, phage ϕ6 provides a useful empirical model for examining questions relating to the evolution of genetic exchange (sex) in viruses and other biological populations (Chao 1992; Turner 2003).

Because virologists could easily observe and manipulate reassortment in the laboratory, the process was long assumed to be important in natural populations of segmented pathogenic and nonpathogenic viruses (Pressing and Reanney 1984; Chao 1988). However, this assumption was only recently confirmed for cystoviruses. Our recent studies of natural phage isolates indicated that cystoviruses were surprisingly prevalent and diverse in the terrestrial systems studied; 13 of 32 collected plants yielded cystoviruses, and a single plant often showed many distinct variants that were frequently only distantly related (Silander et al. 2005). We used partial sequence data from the phage isolates to examine their genetic relatedness. Phylogenetic analyses showed that the relatedness among isolates as inferred from gene trees of individual segments depended strongly on which segment was used to construct the phylogeny. These results suggested that mixis occurs frequently in natural populations of cystoviruses and that variation due to reassortment rivals (or even exceeds) that due to spontaneous mutation (Silander et al. 2005). In addition, consistent with laboratory observations, we found no evidence of measurable homologous recombination (i.e., mixis due to breaking and joining of chromosome segments) within short sequences (up to 413 bases long) of the three genome segments (Silander et al. 2005).

The current study extends our work on virus biogeography by further examining population structure of phage in the family *Cystoviridae*. We again hypothesize that segment reassortment in these terrestrial phage occurs frequently enough to produce heterogeneity in their genetic relatedness, but that homologous recombination does not. In contrast to the earlier study, which sampled primarily from a single site in California, the current study focuses on viruses collected at three geographic locations in the Southern New England region of Northeastern USA. As predicted, we find no evidence for homologous recombination across large stretches of two genome segments. But in contrast to our prior results, we show that reassortment seems to occur less frequently among cystoviruses sampled from these New England locations relative to cystoviruses sampled in Southern California. Specifically, in the California sample, reassortment has erased all evidence of shared evolutionary history between individual segments, whereas in the New England sample evidence of a shared evolutionary history remains, despite similar levels of divergence for both populations. This suggests that the evolutionary importance of mixis may vary according to geography. We conclude by discussing possible factors underlying these differences.

Materials and Methods

CULTURE CONDITIONS

Phage were cultured on *P. phaseolicola* (American Type Culture Collection # 21781), the typical laboratory host of cystovirus ϕ6. Culture medium was LC: Luria-Bertani medium at pH 7.5. All phage were isolated via plaque purification. A plaque forms when a lytic virus particle infects a cell, releasing progeny that subsequently infect neighboring cells; this iterated process results in a visible plaque in a lawn containing superabundant cells, and a typical phage ϕ6 plaque contains ~10^6 to 10^8 particles descended from the originating phage (Dennehy et al. 2007). To form plaques, a phage suspension was mixed with 200 μl of overnight (stationary phase) *P. phaseolicola* culture in 3 mL of soft (0.7%) agar at 42°C. The mixture was overlaid onto solid LC agar (1.5% agar) in a Petri dish and incubated for 24 h at 25°C (Turner and Chao 1998).

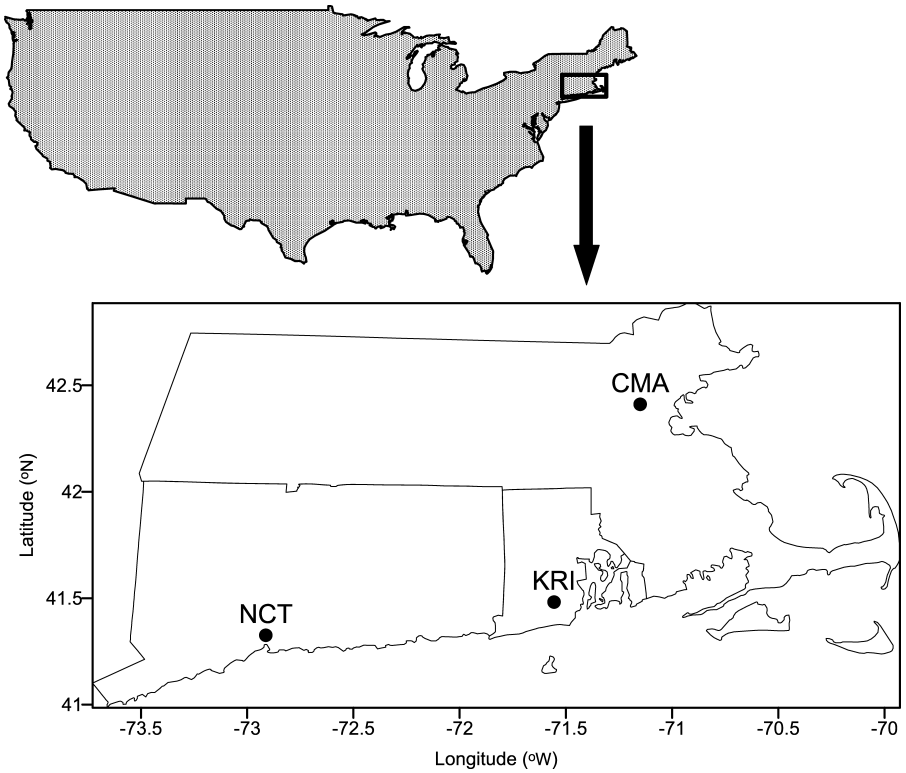

Figure 1. Collection sites in Southern New England, USA: New Haven, Connecticut (NCT), West Kingston, Rhode Island (KRI), and Cambridge, Massachusetts (CMA).

COLLECTIONS AND PHAGE ISOLATION

Collected plants were white clover (*Trifolium repens*), red clover (*T. pratense*), and green bean (*Phaseolus vulgaris*). One hundred two plants were collected from three locations in Southern New England, USA (Fig. 1). Plant collection followed a nested design; at each of the three locations, we identified three or four sublocations containing clover patches. Sublocations were separated by roughly 0.5–1 km. At each sublocation, where plant distributions allowed, one white clover plant and one red clover plant were collected every 10 m along a transect. Collection sites in New Haven, Connecticut (NCT: latitude 41.32539, longitude −72.91244) were weedy edges of grass lawns in two city parks (clover), and a community garden and a farmer's field (bean). Collections in West Kingston, Rhode Island (KRI; 41.48170, −71.55613) were weedy edges of a farmer's field, grassy edges of a bike path, and weedy areas between a road and a pond. Collections in Cambridge, Massachusetts (CMA; 42.41112, −71.15010) were grassy road medians and grassy edges of a bike path. Leaves, stems, and flowers were picked with sterile tweezers and placed in separate clean plastic bags; sealed bags were stored at 4°C for up to two days. Plant material was then crushed with a mortar and pestle and incubated overnight in LC broth at 25°C and 120 rpm shaking. For the first 17 plants collected in NCT, the incubation included addition of *P. phaseolicola* that had been cultured for 5 h in LC broth. The other 26 NCT plants were incubated with and without *P. phaseolicola* to determine whether added bacteria affected phage recovery. Because bacteria did not affect virus recovery (see Results), we chose to incubate the remainder of the collection (CMA and KRI) without *P. phaseolicola*. In all cases, the resulting culture was passed through a 0.2-μm filter to remove bacteria and larger debris and stored at −20°C in 40% glycerol. The stored filtrate was plated at multiple concentrations with *P. phaseolicola*. Up to 10 plaques were isolated per plant, and then re-isolated by plating again with bacteria (double plaque purification). Plaque isolates were then cut from agar using a sterile spatula, and stored at −20°C in 40% glycerol; note that each stored plaque contained the descendants of a single phage in the overnight culture made from a single plant. For the 26 NCT collections that were incubated using two different methods, we isolated only phage that experienced incubation without added bacteria.

REVERSE TRANSCRIPTION, PCR, AND SEQUENCING

To isolate phage RNA the plaque suspension was plated at high titer (up to ∼10–15,000 phage particles per plate) with *P. phaseolicola* and incubated at 25°C for 24 h. The soft agar layer was then harvested. Agar and cells were removed by centrifugation (1440 g for 25 min), and phage were concentrated by ultracentrifugation (107,000 g for 3 h). Phage genomes were isolated using the QIAGEN QIAamp Viral RNA Mini Kit (QIAGEN,

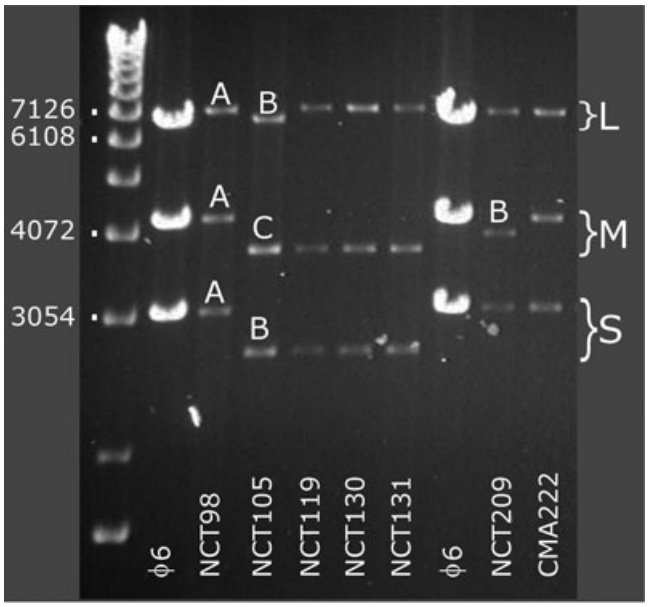

Figure 2. Variation in genome segment size was observed among cystoviruses. This gel fragment contains purified tri-partite genomes of eight cystoviruses. The left-most lane in the gel contains the Invitrogen 1Kb ladder, where the lengths (in base pairs) of selected bands are labeled on the left. The contents of the other columns are marked at the bottom of the gel. Each viral genome consists of three segments, L, M, and S, as labeled on the right. Examples of each observed size class are lettered: two classes of the L and S segments (A and B) and three classes the M segment (A–C).

Germantown, MD), following the manufacturer's spin protocol, which extracts viral nucleic acids (DNA or RNA).

Cystoviruses were then identified by their characteristic tripartite genomes: three visible bands on an agarose gel stained with ethidium bromide or SYBR Safe (Invitrogen, Carlsbad, CA). The tripartite genomes persisted after treatment with DNAse I (Invitrogen), which indicated that they were composed of RNA rather than DNA. Electrophoretic mobility (or "size class") of genome segments was scored on an agarose gel run at 90V for 4 h (Fig. 2). Because cystovirus RNA segments are double-stranded (except for extremely short single-stranded regions at the ends), determination of size-classes was not confounded by differences in RNA secondary structure. The lengths of RNA segments were estimated by comparisons to a DNA ladder (molecular weight standard), where estimated base number was multiplied by a constant (i.e., 1.04) to account for differences between RNA and DNA in molecular weight-to-charge ratio.

Cystovirus genomes were reverse transcribed with Superscript II Reverse Transcriptase (Invitrogen) and random hexamer primers. Two rounds of PCR and sequencing were performed. In the first round, PCR amplification was attempted using each of 19 primer pairs for all cystovirus isolates (66 clones total; see

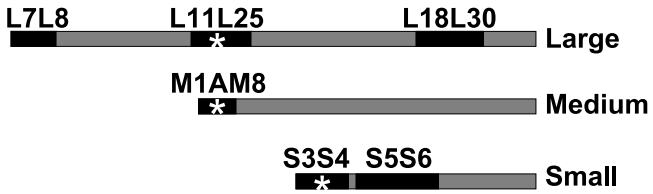

Figure 3. The six loci (PCR amplicons) chosen for sequencing and phylogenetic analysis span about a third of the approximately 13-kb cystovirus genome. Loci marked with asterisks were previously sequenced for California cystovirus isolates (Silander et al. 2005). Locus lengths are roughly to scale.

Results). Reactions used Invitrogen Taq DNA Polymerase (recombinant) and primers that were previously designed for phage ϕ6 (provided by C. Burch, University of North Carolina, Chapel Hill, NC). A "touchdown" protocol was used: for the first 10 temperature cycles, the annealing temperature dropped 1°C per cycle from 60°C to 50°C, and the following 20 cycles had an annealing temperature of 50°C. This first round of PCR was cleaned with the QIAquick PCR Purification Kit or Gel Extraction Kit (QIAGEN, Germantown, MD), or with Exosapit (US Biological, Swampscott, MA). Phage isolates that responded poorly to PCR were excluded from further analysis. The remaining 23 phage were sequenced at six loci where PCR amplification was most successful; below we refer to a sequenced locus using the abbreviation of its associated primer pair. These six loci covered roughly a third of the genome (Fig. 3). The sequenced phage were isolated from 11 individual plants, with between one and four phage per plant. Sequencing was carried out using Applied Biosystems (ABI, Carlsbad, CA) BigDye and the same primers as used during amplification. Sequences were cleaned by ethanol precipitation and read by an ABI 3100 or 3730 sequencer.

For the second round of PCR and sequencing, new primers (sequences available upon request) were designed based on conserved regions of the sequences obtained in the first round.

SEQUENCE CURATION

Sequences were manually corrected using ABI Seqgen and aligned by eye and using ClustalW (Thompson et al. 1994) within the program BioEdit 7.0.4.1 (Hall 2007). All sequences used in this study have been deposited in Genbank (L segment, HM041210-HM041219, HM046256-HM046277; M segment, HM046278-HM046289, and S segment, HM046290-HM046309).

PHYLOGENETIC ANALYSIS

For each locus separately and for concatenations of loci on the L and S genome segments, Bayesian phylogenies were reconstructed in MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003), using the default settings with the model of nucleotide substitution

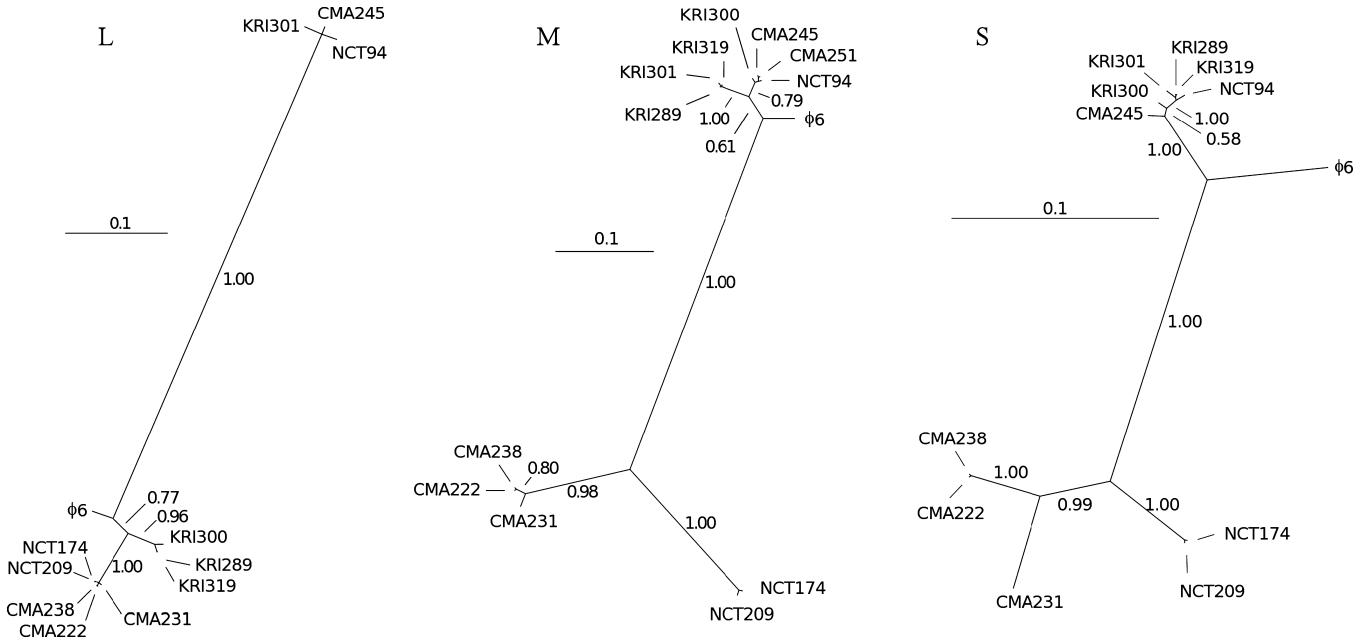

Figure 4. 50% majority-rule consensus phylogenies for genome segments L, M, and S. Highly similar sequences from the same plant are represented by a single sample and sequence data from different loci on the same genome segment were concatenated. The three-letter prefix on sample names indicates geographic origin (CMA: Cambridge, MA; KRI: Kingston, RI; NCT: New Haven, CT). Confidence is indicated by posterior probabilities on each branch. Some posterior probabilities near crowded tips are not presented. For the L segment, the three missing posteriors are all 1.00. For the M segment, the branch below clade CMA245-CMA251 has posterior probability 0.95; the branch below CMA245-CMA251-NCT94 has posterior probability 0.62. For the S segment, the branch below KRI289-KRI319-KRI301 has posterior probability 0.66.

indicated by the AIC criterion (usually the general time reversible model with a gamma distribution of rate variation and a proportion of invariable sites) as implemented by MrModeltest 2.2 (Posada and Crandall 1998). MrBayes was run for 1 million generations, including a burn-in of 300,000 generations, and checked for convergence. If the run had not adequately converged, the analysis was re-run for 2 million generations with a burn-in of 600,000 generations, after which all runs showed convergence. Phylogenies displayed here are 50% majority-rule consensus trees of the New England isolates (plus phage ϕ6) based on the six sequenced loci (Fig. 4), and of all known cystoviruses that align with our isolates at one locus per genome segment (see Fig. S1). Unless indicated otherwise, hereafter "reconstructed phylogenies" refer to the phylogenies of the New England isolates.

We mapped size class onto the L- and M-segment phylogenies for the sequenced cystoviruses. To examine the origin of size variation of the L segment, we tested whether phage ϕ6 and clones 289, 300, and 319 formed a monophyletic group. We constructed a constraint phylogeny that was unresolved except for one partition separating these four clones from all the others. We used PAUP* 4.0b10 (Swofford 2002) to filter the phylogenies sampled by MrBayes (after the burn-in period) during phylogeny reconstruction and to determine how many of these phylogenies were consistent with the constraint phylogeny. The fraction of consistent phylogenies out of total (post-burnin) phylogenies represents the posterior probability of monophyly of the four clones.

RECOMBINATION AND REASSORTMENT

Recombination and reassortment are processes that cause different regions of the genome to be independently inherited. As a result, the evolutionary histories and patterns of relationship of different genome regions may become decoupled, causing phylogenies constructed from these different sequences to disagree. We thus tested for recombination and reassortment among naturally isolated cystoviruses using two phylogenetic approaches to determine whether evolutionary histories and patterns of relationship differed significantly for pairs of loci on the same segment (indicating recombination), and for pairs of genome segments (indicating reassortment). Both approaches accounted for uncertainty in phylogeny estimation by using the distribution of trees visited by MrBayes.

In the first approach, for each pair of loci on the same genome segment (three pairs on L and one pair on S) we asked whether the data (sequence alignments for the two loci) were better described by a single phylogeny or by two different phylogenies. We defined three models: (i) the two loci were constrained to the same model of nucleotide substitution (along with parameters for state frequencies, rate matrix, gamma shape, and proportion of invariable

sites), branch lengths, and topology; (ii) only the branch lengths were estimated separately for the two loci; and (iii) both branch lengths and topology were estimated separately for the loci. The fit of each model (along with the phylogenies and patterns of nucleotide substitution) to the data (the alignment) was indicated by the harmonic mean of the estimated marginal likelihood of the phylogenies sampled after the burn-in. We used the ratio of harmonic means (Bayes factor) taken from two different model analyses, to determine the preferred model. If Ln[Bayes factor] > 1.0, the data were inferred to support one model significantly better than the other (Raftery 1996). Comparisons were also made between models (i) through (iii) where the model of nucleotide substitution and other parameters were always allowed to vary between the loci; results did not differ qualitatively. Importantly, Bayes factors are proportional to the marginal likelihood, and not simply the maximum likelihood value, and thus can sometimes favor less parameter rich models (Nylander et al. 2004). If the data best supported the model with different topologies for each locus (model iii), we concluded that these two loci had experienced recombination and evolved independently. If the model with different topologies did not perform significantly better than one or both of the models having a single topology, we concluded that the two loci had followed the same evolutionary history and experienced little or no recombination. To seek evidence of reassortment between segments, we concatenated sequences from the same genome segment and performed an identical Bayes factor analysis between segments.

The second approach allowed detection of partial similarity between phylogenies. Here we asked how the distance between two phylogenetic topologies compared to the distance expected by chance alone. We began with one phylogeny per segment, constructed from concatenated sequences. We used two measures of phylogeny-to-phylogeny distance: the number of different resolved quartets, and the number of shared resolved quartets. (A quartet is the pattern of relationship between four taxa on a phylogeny, ignoring the other taxa on the phylogeny. Quartets, rather than triplets, were used because trees were unrooted.) Quartet count comparisons occurred as follows. For each segment, we randomly chose 500 phylogenies from 10,000 phylogenies visited by MrBayes after the burn-in, such that the frequency of a topology among the 500 phylogenies roughly represents its posterior probability. We used the program Component 2.0 (Page 1993) to enumerate the different and shared resolved quartets between each of the 500 phylogenies from one genome segment, and each of the 500 phylogenies of a second segment (25,000 pairs). We then established the distribution of phylogeny-to-phylogeny distances expected by chance by comparing the 500 phylogenies of one segment to 500 random coalescent trees generated in PAUP* 4.0b10. We compared the distribution of phylogeny-to-phylogeny distance scores from each pair of genome segments with the distributions of scores expected by chance (Fig. 5). If distance scores fell within the expected distribution, we concluded that reassortment had scrambled the history of one genome segment relative to the other.

LINKAGE DISEQUILIBRIUM

We also tested for reassortment among segments by calculating two measures of linkage disequilibrium (LD), r^2 and D', which measure the magnitude of statistical associations between loci. If segments reassort frequently associations between segments are randomized, consequently reducing LD. But if reassortment occurs only over long timescales, significant LD should exist

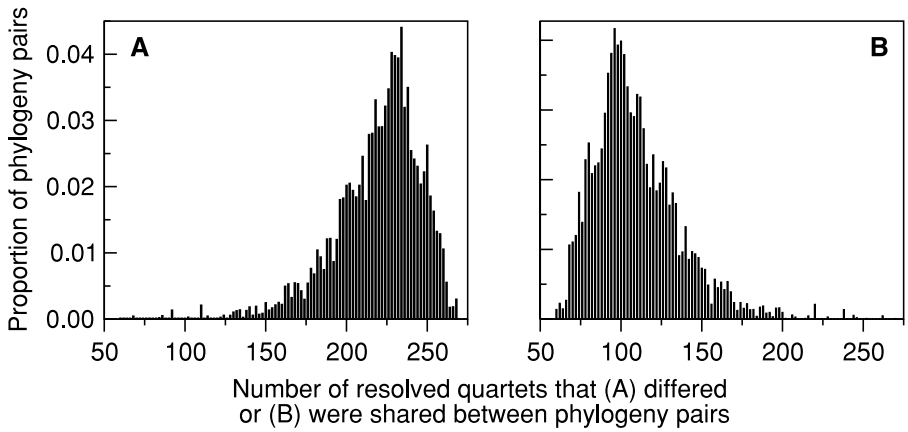

Figure 5. Phylogeny-to-phylogeny distance comparisons show that phylogenies are more similar than expected by chance. In the plot on the left (A) arrows mark the median distance (median number of quartets that differed) between 500 of the most probable phylogenies made from each genome segment. The bars show the distances between pairs of randomized phylogenies, in other words, the proportion of pairs of randomly generated phylogenies that differed by the indicated number of quartets. Similarly, on the right (B), arrows mark the median number of shared quartets for pairs of phylogenies made from each segment. Bars show the proportion of pairs of randomly generated phylogenies that shared the indicated number of quartets.

between segments. r^2 and D′ were measured between all pairs of polymorphic sites (nucleotides) and an average value was obtained for all nucleotides between each pair of segments. One set of segments was then randomized, and the average r^2 or D′ was recalculated. This process was repeated 10,000 times to test whether the observed average LD statistic was significantly greater than expected for a random (unlinked) collection of segments (Silander et al. 2005). For this analysis, we used sequence data from one locus on each segment: S3S4, M1AM8, and L11L25.

Results

CYSTOVIRUS PREVALENCE

Of 102 plants collected, 15 plants yielded cystovirus clones, which formed visible plaques on the permissive host *P. phaseolicola*. Estimated prevalence among plants was 0.15, which was significantly lower than estimated prevalence of 0.41 in Southern California (Fisher's exact test, $P = 0.005$). In the current study, a total of 66 cystovirus clones were isolated (Table 1). The 15 plants that yielded clones were distributed roughly equally across the three geographic locations in New England: 5, 4, and 6 plants obtained at sites in New Haven, CT (NCT), Kingston, RI (KRI), and Cambridge, MA (CMA), respectively. These 15 plants included all of the species targeted for collection: 6 *T. repens*, 6 *T. pratense*, and 3 *P. vulgaris* plants. Two-by-three Fisher's exact tests showed no difference in cystovirus prevalence between plant species ($P > 0.5$) or location within New England ($P = 0.38$). Eleven other plants yielded plaques formed by unknown DNA phage, but these isolates were not analyzed in our study.

The first 17 plants from NCT were incubated with *P. phaseolicola*; only two of these 17 plants yielded cystoviruses, a value slightly lower than the overall recovery rate. For the remaining 26 NCT plants, which were incubated with and without added bacteria, the same three plants consistently yielded cystoviruses; thus, we observed no difference in cystovirus recovery based on isolation method.

DIVERSITY: SIZE OF GENOME SEGMENTS

All three cystovirus genome segments exhibited variation in electrophoretic mobility, which we interpreted as variation in segment length. Two size classes for the L segment, three size classes for the M segment, and two size classes for the S segment were visibly distinguishable (Table 1). Size differences between classes were roughly 200–400 nucleotides, whereas within classes, there was no detectable size polymorphism. The intermediate size class of the M segment ("B" in Table 1) was unique to one *T. repens* plant and one *T. pratense* plant, which were growing a few centimeters apart at location NCT. Size class did not correspond to location; e.g., at locations NCT and CMA we observed size variation for each segment. At KRI, variation was observed for the L segment. Within-plant size variation was found only once, in plant #299 at location KRI.

Five of the 12 potential size class combinations (such as A/A/A, A/A/B, etc. in Table 1) were observed. The most frequent

Table 1. Genetic variation per plant among the 66 cystovirus clones. Each viral genome segment was classified by size: two classes, "A" (larger) and "B" (smaller), for the L and S segments; and three classes, "A" through "C", for the M segment. Most plants harbored only one phage variant by this classification, although one plant yielded two variants.

Plant number	Number of clones isolated	Clone name	Size class			Location
			Segment L	Segment M	Segment S	
		ϕ6	B	A	A	New Haven, CT
14	3	NCT119, NCT130–131	A	C	B	
59	2	NCT94, NCT98	A	A	A	
68	2	NCT104–105	B	C	B	
175	1	NCT174	A	B	A	
153	4	NCT209–212	A	B	A	
219	4	CMA220–223	A	A	A	Cambridge, MA
230	2	CMA231–232	A	A	A	
233	5	CMA238–242	A	A	A	
235	6	CMA245–250	A	A	A	
236	6	CM A251–256	B	C	B	
237	6	CMA257–262	A	C	B	
264	6	KRI272–277	A	A	A	Kingston, RI
288	10	KRI289–298	B	A	A	
299	8	KRI300–307	A and B	A	A	
310	1	KRI319	B	A	A	

combination consisted of the large size class for all three segments (A/A/A), and we noted that the canonical cystovirus phage $\phi 6$ was not a member of this group due to its relatively small L segment. We found statistically significant disequilibrium between the M and S segments (two-way exact test, $P = 0.001$) but not between the L and M segments ($P = 0.59$) nor L and S segments ($P = 0.54$). Thus, based on the frequencies of size classes for the M and S segments, the frequencies of combinations of M- and S-segment size classes were more skewed than expected. For this test, we ignored virus clones that could have been recently derived from the same ancestor, in the plant or in laboratory culture, as discussed further below.

DIVERSITY: RNA SEQUENCES

None of the 66 naturally isolated virus clones was amplified by all primer pairs, and no primer pair worked universally. We observed that a subset of 23 clones responded to the largest number of primers; thus, these 23 viruses were used to examine RNA sequence diversity. For these 23 clones and six primer pairs, we obtained 125 of the 138 possible sequences, with at least one sequence on each segment for all but one clone (CMA251). Sequence alignments revealed high levels of polymorphism (Table 2). Most variation occurred at synonymous sites, suggesting strong selection against amino acid changes. Five small indels (~9 bases or smaller) were found at the M1AM8 and S5S6 loci. When sequences were compared with those from other known cystoviruses, we found two large indels within the locus L7L8, which is near the 5′ end of the L segment. A 337-base insertion was shared between CMA245/248 and KRI301/305 but no other sequences. (This pattern of variation does not correspond to the variation in segment mobility reported immediately above.) The insertion contained all but the first six bases of a 336-base potential open reading frame. Twelve substitutions in the insert separate CMA245/248 and KRI301/305, 11 of which are synonymous. The 12th encodes either asparagine or aspartate, two amino acids of similar polarity. This pattern of substitution suggests strong selection against amino acid changes and therefore that this insertion codes for a protein. The putative ORF does not have significant similarity to any protein currently in Genbank (protein-protein BLAST, E $\geq$ 5.8). It is located only 11 bases away from a previously described indel that contains an ORF encoding for protein "PA" (Mindich et al. 1999). None of our isolates contain the 203-base insertion found by Mindich et al. (1999) on the L segment.

The 23 sequenced clones came from only 11 individual plants. Most viral sequences that were isolated from the same plant differed by only one or two point mutations (less than 0.1% of sites). The exception was plant #299, which contained two very divergent phage genotypes each represented by a pair of closely related sequences (KRI300 and KRI306, and KRI301 and KRI305). The divergent genotypes differed at 1–21% of nucleotide sites, depending on the locus.

Because our isolation process (see Methods) required that viruses grew in vitro, nearly identical genotypes isolated from the same plant sample could be the descendants of a recent common ancestor, either in the plant or in laboratory culture. Therefore in the phylogenetic analysis (see below), we conservatively included only one clone from groups of viruses that were observed to be highly similar in sequence and which were isolated from the same plant. By this logic, we chose one clone from each of the following nine strain groups: NCT94 and NCT98; NCT209, NCT210, NCT211, and NCT212; CMA222 and CMA223; CMA231 and CMA232; CMA238 and CMA239; CMA245 and CMA248; KRI289 and KRI290; KRI300 and KRI306; and KRI301 and

Table 2. **Statistics were calculated in the program DNAsp 4.10 (Rozas et al. 2003) for sequences at each locus. Nearly identical sequences found in the same plant were excluded. Sequence length excludes indels and all other sites that were missing for at least one sample. π is the average pairwise number of nucleotide differences per site (Nei 1987). Because more than two bases were occasionally found at a site, we report θ_η, an estimate of $2N\mu$ (twice the haploid population size times the mutation rate) that takes into account all mutations rather than just the number of segregating sites. Tajima's D was calculated using θ_η. None of the D values are statistically different from zero, which indicates no evidence for historical changes in population size or selective sweeps. The ratio K_a/K_s (the number of nonsynonymous substitutions per nonsynonymous site divided by the number of synonymous and silent substitutions per synonymous or silent site) is much less than 1.0, which indicates strong purifying selection against amino acid changes (Nei and Gojobori 1986; Ohta 1995; Makalowski and Boguski 1998).**

Locus	No. sequences	Sequence length (bases)	Proportion coding	Proportion polymorphic	π	$\theta\eta$	D	Ka/Ks
L7L8	10	544	0.53	0.25	0.09	0.09	0.06	0.04
L11L25	11	702	1.00	0.18	0.08	0.07	0.76	0.00
L18L30	11	846	1.00	0.22	0.10	0.08	0.73	0.01
M1AM8	12	475	0.27	0.25	0.12	0.10	0.91	0.07
S3S4	11	708	0.72	0.17	0.07	0.06	1.35	0.16
S5S6	9	969	0.97	0.19	0.09	0.08	0.59	0.13

KRI305. This process reduced the number of analyzed clones to 11 genotypes, plus an additional phage (CMA251) that yielded sequence only at locus M1AM8.

PHYLOGENETIC ANALYSES

Unrooted phylogenies for all three genome segments were characterized by long internal branches connecting relatively tight clusters of close relatives (Fig. 4; see Fig. S1), indicating that there were widely divergent segment lineages circulating. Isolates did not cluster strictly by geographical location either within New England (Fig. 4) or across New England and California (Fig. S1). New England phylogenies made from different loci on the same genome segment (segments L and S only) were concordant with the trees made from concatenated sequences.

RECOMBINATION

We tested for recombination using Bayes factors to determine whether two loci on the same segment or two genome segments were best described by the same or different phylogenetic topologies. Table 3 lists the model comparisons for all within-segment pairs of loci; all were better described by a single topology, rather than by two different topologies. Of the three models tested, the model that allowed topologies to differ for the two loci (model iii in Table 3) was never preferred. This observation suggests that the isolates have not experienced homologous recombination, which would have allowed two loci on the same segment to follow independent evolutionary paths. The comparison between loci at opposite ends of the longest genome segment, L7L8 and L18L30, was potentially most meaningful as these loci are the most physically distant loci within one segment in the entire genome; recombination should have been more likely between these two distantly separated loci than between any other pair. Although this comparison was ambiguous regarding which model provided the best fit, the model allowing separate topologies (iii) was never the best model. Thus, even this comparison yielded no evidence for recombination.

REASSORTMENT

We again used Bayes factors to test for reassortment between genome segments. In contrast to the within-segment results, pairs of segments were significantly better described by two different phylogenetic topologies (Table 3, bottom). For every segment pair, the best model was the one that allowed different topologies for each segment (model iii in Table 3). This result indicated that segment reassortment has played at least a small role in the evolutionary history of the isolates.

Despite the strong evidence for reassortment, we noted that some similarity existed between segment phylogenies. For example, the group containing isolates NCT174, NCT209, CMA222, CMA238, and CMA231 was found with high support in the phylogenies of all three genome segments (Fig. 4). We quantified this observation by measuring the distance (in quartet counts) between phylogenies of the three genome segments, and comparing those distances to the expected distance between randomly generated phylogenies (Fig. 5). Although the phylogenies of different segments were not identical, we observed significantly more shared quartets and significantly fewer different quartets, compared to randomly constructed phylogenies (i.e., observed values resided in the extreme tails of the distributions; Fig. 5). These data suggested that although there is reassortment between segments, it has not erased all evidence of shared evolutionary history between segments.

Table 3. **Pairwise comparisons of Bayes factors for loci on the same genome segment yielded no evidence for homologous recombination within genome segments (top). Comparisons between genome segments (concatenating all available loci on each segment) indicated that reassortment between genome segments does occur (bottom). Table entries show the relative performance of three models (with the better model on the left of the "greater than" symbol): a model with branch lengths and tree topology fixed between the two loci ("i"), a model with branch lengths free to vary between the loci ("ii"), and a model with both branch lengths and topology free to vary between the loci ("iii"). For this analysis, all sets of sequences from a single segment were concatenated to construct the phylogenies. The relative ranking of model iii indicates whether separate topologies are preferred for the two loci or segments. For no within-segment pair of loci (top) was the model with differing topologies (iii) preferred. In contrast, iii was preferred in every between-segment comparison (bottom). Symbols * and ** indicate a significant difference in Bayes factors of at least 1.0 and at least 2.5 respectively. The headings "Test 1" and "Test 2" refer to comparisons performed with all other model parameters fixed between the pair of sequences (Test 1), and with all other parameters free to vary between the pair of sequences (Test 2).**

Locus 1	Locus 2	Test 1	Test 2
L7L8	L11L25	i>ii>iii * *	i>ii>iii ** **
L7L8	L18L30	ii>iii>i * **	i>iii>ii **
L11L25	L18L30	i>iii>ii **	i>ii>iii **
S3S4	S5S6	i>ii>iii ** **	i>iii>ii ** *
Segment 1	Segment 2	Test 1	Test 2
L	M	iii>ii>i ** **	iii>ii>i ** **
L	S	iii>ii>i ** **	iii>ii>i ** **
M	S	iii>ii>i ** **	iii>ii>i ** **

LINKAGE DISEQUILIBRIUM

LD between segments was measured using two statistics, D′ and r^2. The metric D′ measures the amount of LD relative to the maximum possible LD, given the observed amount of polymorphism in a sample. The related metric r^2 is a measure of how often a specific allele at one locus is associated with a specific allele at a second locus. For both LD metrics, there were highly significant levels of LD between all segments (see Fig. S2). We calculated significance by randomizing the associations between segments and recalculating each metric. For D′, the *P*-values were 0.006, 0.001, and < 0.0001 for the S-M, S-L, and M-L segment associations, respectively. For r^2, *P*-values were <0.0001, 0.041, and 0.018 for the SM, SL, and ML segment associations, respectively (Fig. S1). This analysis was thus consistent with the above analysis of phylogenetic distances using quartet counts.

Discussion

The current study extends our work on virus biogeography (Silander et al. 2005), focusing on segmented RNA phage in the family *Cystoviridae*, which were collected from natural samples of microbes associated with leguminous plants. These efforts complement existing studies of population structure in enteric bacteria (e.g., *Escherichia coli*) and extremophiles (Breitbart et al. 2003; Snyder et al. 2007; Andersson and Banfield 2008). Two major results are evident from our work on phage population structure. First, these studies have revealed that natural populations of cystoviruses contain extensive genetic variation, with frequent changes occurring in both genome size and nucleotide sequence. Second, in the current study we find that although reassortment has occurred multiple times within the New England populations (as was also found in the previous study; Silander et al. 2005), it has not erased all traces of shared evolutionary history between segments within the New England populations. This second point contrasts markedly with the previous study, which found no evidence of shared evolutionary history between segments.

CYSTOVIRUS PREVALENCE

Using samples of wild clover and bean plants from three locations in New England, we observed that cystoviruses could be recovered from 15 out of 102 plant samples. This result indicates the minimum prevalence of cystoviruses (perhaps only an ecologically similar group, or ecotype, of these viruses), because the phage we recovered were limited to strains capable of plaque formation on *P. phaseolicola*, a permissive laboratory host of phage ϕ6 (Duffy et al. 2007). Thus, we have not determined what fraction of total cystoviruses our collections represent. Previous work showed that some cystoviruses that fall within the genetic diversity of phage that we collected (i.e., phage ϕ8, ϕ12, and ϕ13) are not capable of infecting *P. phaseolicola* (Mindich et al. 1999). However, natural variation in host use among cystoviruses has not been studied extensively (but see Duffy et al. 2007 for a related in vitro study). In the future, culture-independent molecular detection techniques could be harnessed to determine the total prevalence of cystoviruses in natural phage communities. (We discuss the effects of culture bias on sample diversity below.)

We observed no significant difference in cystovirus prevalence among the three collection sites in New England. However, cystovirus prevalence estimated in the current study was roughly one-third of that found by Silander et al. (2005), who sampled mostly white clover plants growing on lawns in La Jolla, CA. These differing recovery rates are unlikely to stem from differences in methodology, because highly similar protocols of phage isolation spanned the two studies. However, we noted that Silander et al. enriched for phage by culturing plant samples in LC medium containing *P. phaseolicola*, whereas the current study generally enriched using bacteria-free medium. We examined this methodological difference using 26 plants from location NCT that were incubated with and without *P. phaseolicola*; these results showed that the different enrichment techniques did not alter estimates of cystovirus prevalence.

We speculate that the differences in cystovirus prevalence between sampled locations in New England and Southern California may result from differences in abiotic factors. In New England, the winter is colder and the growing season is shorter than in Southern California. Similarly, the growth of phytopathogenic bacteria is limited by the shorter growing season. Cystoviruses seem to preferentially use these bacteria as hosts; thus, the phage may be reduced in density during cold times of the year, which could in turn result in lower growing-season densities, such as we found in our sampling. The lower prevalence of phage at locations in New England may explain why genetic variation was low within individual plants, and why segment reassortment seems to occur less often in our New England populations than among strains isolated in Southern California (see further discussion below).

To our knowledge, no other studies have provided data on phage prevalence in association with wild plants. In the original paper reporting isolation of phage ϕ6 from infested bean straw (Vidaver et al. 1973), only a single virus was described. A more recent study describing isolation of cystoviruses from market vegetables similarly focused on detailed strain characterization, rather than questions relating to virus biogeography and population biology (Mindich et al. 1999). Additional sampling from distinct geographic locations, plant species, and terrestrial environments would be needed to obtain a more complete understanding of cystovirus populations and communities. Plant pathogenic Pseudomonads appear to be the preferred hosts for cystoviruses; these bacteria (especially *P. syringae* pathovars) are extremely widespread and important constituents of the phyllosphere (Hirano and Upper 1990) and constitute some of the most

significant agricultural pathogens. But aside from the *Cystoviridae*, no other plant-associated phage have been studied at the population level. More research is warranted, and we speculate that terrestrial plant communities may harbor a rich and diverse variety of phage, as observed in phage studies conducted using samples from marine and lake environments (Bergh et al. 1989; Angly et al. 2006; Desnues et al. 2008), soil (Ashelford et al. 1999), and the human gut (Breitbart et al. 2003).

CYSTOVIRUS DIVERSITY

Genome segment size

One surprising aspect of cystovirus diversity revealed by our study is the extensive natural variation in genome-segment size. Segment size varied often between phage isolated from a single plant (plant #299) or from plants of the same species growing only meters apart. Such size variation is perhaps less surprising in organisms with larger genome sizes, such as eukaryotes, bacteria, and DNA phage (e.g., phage T4) whose genomes are on the order of hundreds of kilobases and can contain large noncoding regions. But RNA phage such as cystoviruses are generally of small genome size, which is believed to cause added genetic constraint.

Furthermore, our data show that although some phage may have two segments that are nearly identical, the third segment may differ by a very large fraction of synonymous substitutions, or in some cases, large-scale indels. For example, isolates KRI301 and KRI319, which have nearly identical M and S segments, contain L segments that differ (minimally) by a 337bp indel (more extensive changes may lie on unsequnced portions of the L segment). This indel is on the order of 5% of the entire L segment, and we would expect that it could have significant effects on the efficiency of genome packaging, which could cause selection against reassortants. At first examination, this does not appear to be so.

Three of our naturally isolated clones as well as phage $\phi 6$ appear to have a deletion in the L segment. The low probability that these viruses are monophyletic suggests that the L segment has changed in size multiple times during the evolutionary history of the small number of clones in our small sample. Furthermore, these size changes are not restricted to the L segment: the three size classes recovered for the M segment indicate at least two separate evolutionary events. Intriguing questions concerning genomic constraint, function, and origins are raised by the indels and genome size variation discovered by Mindich et al. (1999) and in the current study: interestingly, an unrelated indel was found by Mindich et al. (1999) that occurs at nearly the same location in the L segment as the 337 bp indel that occurs in our isolates.

Genome size variation has also been observed in other closely related viruses. Bovine rotavirus, a dsRNA virus containing 11 segments per particle, exhibits segment-size variation even during a single outbreak within a herd of hosts (Rodger and Holmes 1979). Natural variation in segment size has also been shown for ssRNA influenza viruses, where indels on the order of 20 amino acids have been associated with host adaptation (Matrosovich et al. 1999). However, the frequent acquisition of novel reading frames and their ability to function independently of other genomic regions (i.e., other segments) has not, to our knowledge, been previously described in segmented RNA viruses.

Sequence variation

Among the 23 cystoviruses that we sequenced, nucleotide diversity was very high, although dN/dS values were low for all loci we examined (Table 2). This extreme nucleotide diversity along with low observed amino acid diversity is a signature of strong purifying selection on amino acid sequence. Thus, we envision cystovirus populations experience high mutation rates (as typical for RNA viruses) that continuously introduce new genetic variation, but that selection continuously removes those variants with altered amino acid sequences. Silander et al. (2005) similarly found high overall sequence diversity for cystoviruses isolated in Southern California. Levels of nucleotide diversity in other phage populations have not been well characterized, providing no basis for comparison. However, our estimate for pairwise nucleotide diversity (π) in cystoviruses lies near the upper end of the range reported for well-studied RNA viruses that infect eukaryotes; e.g., in the highly variable retrovirus HIV-1, subtypes A and D exhibited $\pi \approx 0.16$ in a 12-year study in Uganda (Herbeck et al. 2007).

Potential biases

We note that three methodological details have likely affected our estimates of diversity. The first is culture bias. All phage were isolated using the bacterial strain used in the original isolation of phage $\phi 6$ in Nebraska, USA (Vidaver et al. 1973). However, the resulting culture bias likely underestimates levels of genetic diversity in nature. Despite the fact that we have used a single bacterial host, which most likely selects for more similar phage, we found very high levels of genetic diversity. We expect that if we used an assortment of different bacterial hosts, we would discover even more cystovirus diversity.

Second, from among the cystoviruses that we cultured, we further excluded phage isolates for which we could not generate PCR amplicons and sequence data. This likely excluded some divergent phage. Despite this, we found that the phage we sequenced were highly divergent on a genetic level, containing large indels at multiple loci across different segments.

Finally, it is possible that we have biased our quantitative estimates of diversity (π and θ) upward, as we excluded all phage that were highly similar in genetic sequence and which were isolated from the same plant. The exclusion of these highly similar sequences was important because these could have arisen during the isolation process from a single progenitor phage. Although the exclusion of these sequences leads to potential bias, we believe

that their inclusion would have caused an even stronger biased impression that higher levels of within-plant clonality existed at the time of sampling.

RECOMBINATION AND REASSORTMENT

We used two phylogenetic methods to detect whether recombination or reassortment have been important in the evolution of cystoviruses in the wild. Phylogenetic relationships among the cystovirus isolates were examined separately using data from each of the three segments. In the absence of reassortment or recombination, we would expect the phylogenies to be congruent due to shared evolutionary histories. As in the previous study, we found no evidence of recombination within any of the segments. In contrast, using a method based on Bayes factors, we found significant phylogenetic incongruence between genome segments, which confirmed that some reassortment has occurred in natural populations. However, the second method of quartet-count comparisons detected significant similarities between the segment phylogenies. Therefore, the evolutionary histories of the segments were not completely independent. We concluded that reassortment occurs among the cystovirus populations we sampled in New England, but not frequently enough to erase all evidence of shared evolutionary history.

The Bayes-factors method we used to detect recombination and reassortment is sensitive to small topological differences between phylogenies. One differently placed partition, if well supported, would be enough to yield a significant difference between phylogenies. In our case, significant phylogenetic topological differences between segments may have been driven primarily by the placement of two samples, KRI300 and KRI301, which appeared to be different genetic mixtures of two lineages. These two isolates were found in the same plant and most likely descended from a single co-infection (and reassortment) event sometime in the recent past.

Our assessment of reassortment (and recombination) is likely to underestimate the level of structure in natural cystovirus populations. As discussed above, the diversity in our sample of genetic sequences is likely a subset of existing diversity in nature, due to our culture- and PCR-based methods. We expect that highly divergent phage (such as would have been missed by our culturing and PCR) will reassort (or recombine) less often than genetically similar phage, which would imply more population structure. Thus our finding of significant population structure in New England is likely to be conservative.

The significant similarity in the evolutionary histories between segments is surprising only because it contrasts with the findings of Silander et al. (2005) for the same virus family. In this earlier work, we performed similar phylogeny-to-phylogeny comparisons using isolates mostly from La Jolla, California and found that segment phylogenies were no more similar to each other than to random phylogenies. This result suggested that extraordinary levels of reassortment had effectively scrambled the phylogenies of these previously isolated phage. In contrast, the current study presents quartet-count comparisons that show more similarities and fewer differences than expected by chance suggesting that reassortment has played a lesser role for the New England cystovirus isolates.

The comparison across studies is necessarily indirect. In both studies, to infer population structure, we have used null models that derive from randomizations of the datasets themselves. Because the two datasets are different, it is not possible for us to directly compare the two datasets. However, we note that in the previous study, no evidence of population structure was found (i.e., there was no significant departure from the null model of random mating and complete reassortment), whereas in the current study, we have found strong evidence of population structure (strong departure from the null model). This finding is despite the present study containing fewer isolates and therefore less power to detect population structure. We speculate that one explanation for the difference in findings between the two studies is that cystovirus reassortment rates may depend on geographic location. If geographic variation in reassortment rates exists, this variation may occur for at least two reasons. First, opportunities for co-infection and/or reassortment may be rarer in one location. As we report above, we found that only 15 out of 102 plants sampled in New England yielded cystoviruses and that many fewer plants harbored genetically diverse cystovirus populations. If plants (and hosts) are colonized only rarely, co-infection of the same bacterial cell on an individual plant seems likely only for closely related virus genotypes. Because co-infection is required for reassortment to occur, low cystovirus prevalence therefore could result in low rates of reassortment between all but close relatives. By contrast, about one in two plants sampled in Southern California yielded cystoviruses and the within-plant diversity was relatively high: 10 out of 14 phage isolated from individual clovers had at least one segment that was only distantly related to the orthologous segments from the other phage (Silander et al. 2005). Second, reassortment rates may vary if natural selection operating in one geographic location tends to eliminate hybrid cystoviruses because they are of low fitness, thus obscuring past reassortment events. The breaking-apart of co-adapted gene complexes (here, sets of viral segments) is a potential cost of genetic exchange (sex). Co-adapted gene complexes (epistatically evolved segments) may occur in cystoviruses, but experimental evidence for this phenomenon is lacking.

In our previous study, we suggested that selection against hybrids could have played a role among the Southern California isolates (Silander et al. 2005). In that study, the segment pairs L/M and L/S were found to be in linkage equilibrium (LE), whereas the segment pair S/M was in LD. The presence of LE between two

pairs of segments provided the control that allowed us to infer frequent reassortment at the genomic level. Therefore, any observed LD among segments was interpreted as potentially due to selection. By contrast, in the current study the results of our phylogenetic analyses were corroborated by significant LD among all segments pairs. In this case, the absence of a control (i.e., LE pair) prevents us from definitively concluding by the same rationale that negative selection against hybrids helps explain our observations.

Through the current study, we have gained further insight into the nature of a genetic "population" for cystoviruses. In what a human observer would consider to be a single geographic location, we often observed closely related cystoviruses, but also widely differing genetic types. We constructed concatenated trees using virus isolates drawn from our previous and current studies to examine geographic structure; in particular, we sought to determine whether CA isolates clustered with New England isolates from various locations. As these concatenated trees showed, cystoviruses in the two sampled locales do not exist in genetic isolation despite their vast geographic separation. Rather, the trees confirmed long-range mixis, combined with some degree of geographic clustering.

Differences in reassortment rates between cystovirus (or other viral) populations would imply that populations may vary significantly in the amount of standing genetic variation they harbor, and thus in their capacity for adaptation. Variation in these processes across different viral populations and at different geographic scales is an area toward which future research should be directed.

CONCLUDING REMARKS

Our results address a common but unspoken assumption. Evolutionary biologists (several of the present authors included) have expended great effort in explaining when and why genetic exchange (sex) may be advantageous in viruses and other biological systems (Chao 1992; Chao et al. 1997; Turner 2003; Turner et al. 2009). These investigations assume that the costs and benefits of sexual processes are fairly constant across different populations of a particular microbe. The current results serve as a reminder that the extent of genetic exchange may vary considerably even in a system where sex is mechanistically possible, frequently observed under controlled laboratory conditions, and common in some populations. Just as well-studied eukaryote species are characterized by a wide array of mating patterns and sexual behavior, we should expect similar or even greater variability from viruses.

ACKNOWLEDGMENTS

We thank S. Abedon, R. Froissart, P. Lewis, J. Wertz, B. O'Meara (with his instructive website), E. Simms's laboratory group and the Turner laboratory group for helpful discussions and manuscript comments, M.B. Decker for assistance with some figure graphics, and C. Burch for providing PCR and sequencing primers. This work was supported by a National Science Foundation Microbial Biology Postdoctoral Fellowship DBI-0400834 to KJO, and a National Science Foundation grant DEB-0452163 to PET.

LITERATURE CITED

Abedon, S. T. 2008. Bacteriophage ecology: population growth, evolution, and impact of bacterial viruses. Cambridge Univ. Press, Cambridge, U.K.

Andersson, A. F., and J. F. Banfield. 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. Science 320:1047–1050.

Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. A. Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley, H. Liu, et al. 2006. The marine viromes of four oceanic regions. PLoS Biol. 4:e368. doi:10.1371/journal.pbio.0040368.

Ashelford, K. E., M. J. Day, M. J. Bailey, A. K. Lilley, and J. C. Fry. 1999. In situ population dynamics of bacterial viruses in a terrestrial environment. Appl. Environ. Microbiol. 65:169–174.

Bergh, O., K. Y. Borsheim, G. Bratbak, and M. Heldal. 1989. High abundance of viruses found in aquatic environments. Nature 340:467–468.

Breitbart, M., I. Hewson, B. Felts, J. M. Mahaffy, J. Nulton, P. Salamon, and F. Rohwer. 2003. Metagenomic analyses of an uncultured viral community from human feces. J. Bacteriol. 185:6220–6223.

Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer. 2002. Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. USA 99:14250–14255.

Butcher, S. J., T. Dokland, P. M. Ojala, D. H. Bamford, and S. D. Fuller. 1997. Intermediates in the assembly pathway of the double-stranded RNA virus phi 6. EMBO J. 16:4477–4487.

Chao, L. 1988. Evolution of sex in RNA viruses. J. Theor. Biol. 133:99–112.

———. 1992. Evolution of sex in RNA viruses. Trends Ecol. Evol. 7:147–151.

Chao, L., T. T. Tran, and T. T. Tran. 1997. The advantage of sex in the RNA virus phi 6. Genetics 147:953–959.

Dennehy, J. J., S. T. Abedon, and P. E. Turner. 2007. Host density impacts relative fitness of bacteriophage phi 6 genotypes in structured habitats. Evolution 61:2516–2527.

Desnues, C., B. Rodriguez-Brito, S. Rayhawk, S. Kelley, T. Tran, M. Haynes, H. Liu, M. Furlan, L. Wegley, B. Chau, et al. 2008. Biodiversity and biogeography of phages in modern stromatolites and thrombolites. Nature 452:340–343.

Duffy, S., C. L. Burch, and P. E. Turner. 2007. Evolution of host specificity drives reproductive isolation among RNA viruses. Evolution 61:2614–2622.

Fuhrman, J. A. 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548.

Hall, T. A. 2007. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Ibis Therapeutics, Carlsbad, CA.

Herbeck, J. T., F. Lyagoba, S. W. Moore, N. Shindo, B. Biryahwaho, P. Kaleebu, and J. I. Mullins. 2007. Prevalence and genetic diversity of HIV type 1 subtypes A and D in women attending antenatal clinics in Uganda. Aids Res. Hum. Retroviruses 23:755–760.

Hirano, S. S., and C. D. Upper. 1990. Population biology and epidemiology of *Pseudomonas-syringae*. Annu. Rev. Phytopathol. 28:155–177.

Laurila, M. R. L., E. V. Makeyev, and D. H. Bamford. 2002. Bacteriophage phi 6 RNA-dependent RNA polymerase – molecular details of initiating nucleic acid synthesis without primer. J. Biol. Chem. 277:17117–17124.

Makalowski, W., and M. S. Boguski. 1998. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. Proc. Natl. Acad. Sci. USA 95:9407–9412.

Matrosovich, M., N. Zhou, Y. Kawaoka, and R. Webster. 1999. The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. J. Virol. 73:1146–1155.

Mindich, L. 1999. Reverse genetics of dsRNA bacteriophage Phi 6. Adv. Virus Res. 53:341–353.

———. 2004. Packaging, replication and recombination of the segmented genomes of bacteriophage Phi 6 and its relatives. Virus Res. 101:83–92.

Mindich, L., X. Y. Qiao, J. A. Qiao, S. Onodera, M. Romantschuk, and D. Hoogstraten. 1999. Isolation of additional bacteriophages with genomes of segmented double-stranded RNA. J. Bacteriol. 181:4505–4508.

Nei, M. 1987. Molecular evolutionary genetics. Columbia Univ. Press, New York, NY.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3:418–426.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53:47–67.

Ohta, T. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. 40:56–63.

Page, R. D. M. 1993. User's manual for COMPONENT, Version 2.0. The Natural History Museum, London.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.

Pressing, J., and D. C. Reanney. 1984. Divided genomes and intrinsic noise. J. Mol. Evol. 20:135–146.

Raftery, A. E. 1996. Hypothesis testing and model selection. Pp. 163–188 *in* W. R. Gilks, D. J. Spiegelhalter, and S. Richardson, eds. Markov Chain Monte Carlo in practice. Chapman and Hall, London.

Rizvanov, A. A., S. F. Khaiboullina, and S. St Jeor. 2004. Development of reassortant viruses between pathogenic hantavirus strains. Virology 327:225–232.

Rodger, S. M., and I. H. Holmes. 1979. Comparison of the genomes of simian, bovine, and human rotaviruses by gel-electrophoresis and detection of genomic variation among bovine isolates. J. Virol. 30:839–846.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497.

Silander, O. K., D. M. Weinreich, K. M. Wright, K. J. O'Keefe, C. U. Rang, P. E. Turner, and L. Chao. 2005. Widespread genetic exchange among terrestrial bacteriophages. Proc. Natl. Acad. Sci. USA 102:19009–19014.

Snyder, J. C., B. Wiedenheft, M. Lavin, F. F. Roberto, J. Spuhler, A. C. Ortmann, T. Douglas, and M. Young. 2007. Virus movement maintains local virus population diversity. Proc. Natl. Acad. Sci. USA 104:19102–19107.

Suttle, C. A. 2005. Viruses in the sea. Nature 437:356–361.

Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal-W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Turner, P. E. 2003. Searching for the advantages of virus sex. Orig. Life Evol. Biosph. 33:95–108.

Turner, P. E., C. L. Burch, K. A. Hanley, and L. Chao. 1999. Hybrid frequencies confirm limit to coinfection in the RNA bacteriophage phi 6. J. Virol. 73:2420–2424.

Turner, P. E., and L. Chao. 1998. Sex and the evolution of intrahost competition in RNA virus phi 6. Genetics 150:523–532.

Turner, P. E., R. McBride, and C. W. Zeyl. 2009. Sexual exploits in experimental evolution. Pp. 479–521 *in* T. Garland Jr. and M. R. Rose, eds. Experimental evolution: concepts, methods and applications of selection experiments. Univ. of California Press, Berkeley, CA.

Vidaver, A. K., R. K. Koski, and J. L. Van Etten. 1973. Bacteriophage Phi-6: a lipid-containing virus of *Pseudomonas phaseolicola*. J. Virol. 11:799–805.

Wommack, K. E., and R. R. Colwell. 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64:69–114.

Xu, X. Y., S. E. Lindstrom, M. W. Shaw, C. B. Smith, H. E. Hall, B. A. Mungall, K. Subbarao, N. J. Cox, and A. Klimov. 2004. Reassortment and evolution of current human influenza A and B viruses. Virus. Res. 103:55–60.

Associate Editor: M. Travisano

Supporting Information

The following supporting information is available for this article:

Figure S1. 50% majority-rule consensus phylogenies for the New England isolates and Southern California isolates described in Silander 2005, based on the sequence shared between the two studies.

Figure S2. Measures of linkage disequilibrium between segments.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.