Viruses are integral to life in the ocean, contributing to the disease and mortality of their hosts, catalyzing evolution by mediating gene exchange, and influencing the partitioning of nutrients among trophic levels (1). The genetic material of viruses may consist of single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA) or ssRNA or dsRNA, depending on the virus. Most studies of marine viroplankton over the past two decades have focused on DNA-containing viruses, which appear to be predominantly bacteriophages (2). Our knowledge about the RNA viruses in the marine viroplankton is much more limited, but the data available suggest that virtually all of them infect eukaryotic organisms, most likely protists (3). These RNA viruses were often assumed to be a minor component of the viroplankton, but recent data suggest that, at least at the one location sampled, they were as abundant in seawater as DNA viruses (4).

At present, our knowledge of the RNA viruses that infect the marine protistan plankton is limited to what we have learned from 13 isolates and the results of a few molecular surveys (3). The RNA viruses that have been isolated so far infect some of the major taxa of marine protists, including diatoms (5–9), dinoflagellates (10), raphidophytes (11), prasinophytes (12), and thaustochytrids (13). Phylogenies of viruses in the order Picornavirales based on alignments of the RNA-dependent RNA polymerase (RdRp) sequences are congruent with the established taxonomic assignments by the International Committee on Taxonomy of Viruses (ICTV) (14–16), and thus the RdRp is a useful molecular marker for the investigation of the diversity of viruses in the order Picornavirales (picornavirids). Cultivation-independent surveys targeting the RdRp of picornavirids in samples from temperate waters (17) and subtropical waters (18) have revealed a level of genetic diversity that is poorly represented by the limited number of existing cultures (19).
Metagenomic methods can provide a more comprehensive view of the genetic diversity of RNA viruses than single-gene surveys. This approach has been used to investigate RNA virus diversity in a variety of habitats, including reclaimed water (20), untreated wastewater (21, 22), hot springs (23), a freshwater lake (24), and various marine habitats (4, 25, 26). Application of this method to RNA viruses harvested from coastal waters of British Columbia suggested that most were predominantly positive-sense, single-stranded RNA (+ssRNA) viruses that are distantly related to established taxa (25). Very few dsRNA viral sequences were identified, and no sequences from RNA phage, negative-sense, single-stranded RNA (−ssRNA) viruses, or retroviruses were detected (25). These data offered a first glimpse into the genomic diversity of the natural viral viroplankton in seawater but were limited to a single location in temperate coastal waters. To broaden our understanding of viral viral ecology in the ocean, we estimated the abundance of RNA viruses relative to that of DNA viruses and the diversity of viral RNA viral communities harvested from coastal tropical waters. Because of their small genome sizes, RNA viruses cannot be accurately quantified in seawater even with RNA-specific stains (27, 28); therefore, we used an indirect approach in which the values corresponding to the relative masses of total viral RNA and DNA were divided by estimates of the mass of nucleic acid per RNA or DNA virion in the sample to determine relative abundances. Our data, from a tropical coastal site sampled on two occasions, indicated that the abundance of RNA viruses could at times exceed that of DNA viruses in seawater (4). An initial metagenomic analysis of RNA virus diversity in those samples suggested that, just as they did in one study in temperate waters (25), +ssRNA viruses in the order Picornavirales dominated in tropical coastal waters. The goal of our previous report was to estimate the relative abundances of RNA and DNA viruses. In this report, we expand our analysis of the two Kāne‘ohe Bay viromes. Specifically, we present data on the buoyant density distributions of marine picornavirids, provide a more detailed analysis of the RNA virus metagenomes—including the reconstruction of six genomes—and provide estimates of the diversity of marine picornavirids using a number of different approaches.

RESULTS AND DISCUSSION
General description of the metagenomes. A general characterization of the two metagenomes analyzed in this study can be found elsewhere (4), but the salient features are summarized here to provide a context for the new analyses that are the focus of this paper. Pyrosequencing of the two libraries prepared from samples collected in 2009 and 2010 from coastal O‘ahu resulted in a combined total of 249,941 high-quality reads and approximately 89 Mbp of sequence. The majority of sequences in each library assembled into contigs (69% and 78% in the 2009 and 2010 libraries, respectively). Approximately 54% of the total sequences were most similar to those of RNA viruses, 4% appeared to derive from cells, and 42% could not be assigned. Of the sequences that were identified as viral, >97% were most similar to those of +ssRNA viruses (95% specifically to members of the order Picornavirales). The remaining 3% were most similar to those of dsRNA viruses, with the majority having similarity to Micromonas pusilla reovirus (MpRV), the sole member of the genus Mimmoreovirus.

Buoyant densities of marine picornavirad-like viruses. Buoyant density gradients are frequently used to purify viruses from natural assemblages for analysis (29–31), but there are few reports describing the density distribution of uncultivated assemblages of DNA-containing viruses (32–34) and none for the RNA-containing viruses. This information is useful for optimizing purification strategies (29). Although picornavirads dominated our metagenomic libraries, the libraries were prepared from pooled fractions representing a specific density range (1.38 to 1.53 g ml⁻¹), which was chosen conservatively based only on the distribution of total RNA. To better understand the density distribution of marine picornavirads, we analyzed each fraction separately for the entire gradient (≥1.2 to ≥1.6 g ml⁻¹), amplifying by PCR first with degenerate primers targeting two subclades of marine picornavirads and then with primers designed to specifically target the RdRp genes of a putative high-buoyant density phylotype and a putative low-buoyant density phylotype. Amplification with the degenerate primers resulted in variable amplicon yields that depended on which fraction of a CsCl buoyant density gradient was assayed. The patterns were similar for the two subclades assayed, with both showing a peak in the 1.45 g ml⁻¹ fraction (Fig. 1). The distribution was somewhat broader for subclade 1, the primers for which were also found to capture a broader range of phylogenetic diversity (Culley and Steward [18]). Clone libraries prepared from subclade 1 RdRp amplicons derived from one of the lower-density fractions (1.38 g ml⁻¹) and one of the higher-density fractions (1.49 g ml⁻¹) on either side of the main amplification peak revealed that some sequences were present in both libraries but that others were detected only in one library or the other (see Fig. S1 in the supplemental material). Reverse transcription–quantitative PCR (RT–qPCR) using primers designed to target one of the sequences found only in the lower-density library (phytotype A) or one of the sequences found only in the higher-density library (phytotype B) revealed target distributions consistent with the clone library results (Fig. 1). Specifically, the concentration of phytotype A was much higher than that of phytotype B in the 1.38 g ml⁻¹ fraction and the concentration of phytotype B was higher than that of phytotype A in the 1.49 g ml⁻¹ fraction. In both cases, however, the distribution of target was bimodal, with a local maximum in or near the density fraction from which the sequence derived but an overall maximum occurring in the 1.43 g ml⁻¹ fraction (Fig. 1).

The reason for the bimodal peaks observed for both of the specific phytotypes assayed is unknown. The positions of the clearly separated minor peaks make sense, considering the criteria used to choose the targets, but the presence of major peaks for both targets in the same intermediate density fraction is curious. The pattern does not appear to be a result of nonspecific amplification, since melting curves of the amplicons indicated a single narrow peak at the same melting temperature for all fractions for a given primer set. Differences in total RNA levels among the fractions could have influenced the efficiency of the RT reactions (35), but the offset of the RT–qPCR peaks and the total RNA peak suggest this cannot be the sole explanation for the observed distributions. Two alternative explanations are that (i) identical target sequences are found in viruses that differ in their levels of buoyant density or (ii) many of the viruses of each phytotype were aggregated with each other, with other viruses, or with some other material that altered their equilibrium buoyant density, a phenomenon that has been observed previously (34).

Regardless of the details of the phytotype-specific distributions, the amplification data suggest that the buoyant density distribution of picornavirads was primarily in the density range
from 1.35 to 1.5 g ml⁻¹. The distribution was similar to that of total RNA, the most notable exception being the second peak in total RNA at high densities (ca. 1.6 g ml⁻¹), which is not accompanied by an increasing signal for picornavirads (Fig. 1). This suggests that the RNA in the denser fractions is qualitatively different from that in the primary RNA peak and may not be of viral origin.

**RdRp viral diversity in the metagenome.** A search for all likely RdRp sequences (i.e., those that contained, at a minimum, the same two of seven conserved motifs) returned 531 (517 picornavirad-like and 14 reovirid-like) sequences in the 2009 library and 300 (292 picornavirad-like and 8 reovirid-like) in the 2010 library. A subset of these sequences that were longer and contained four of the seven conserved RdRp motifs were analyzed phylogenetically. These longer sequences (51 picornavirad-like and 3 reovirid-like sequences in 2009; 21 picornavirad-like and 3 reovirid-like sequences in 2010) formed large clusters with other environmental RdRp sequences, although some of these clusters were not well supported (maximum-likelihood support values < 80) and most were related only distantly to RdRp sequences from cultivated representatives (Fig. 1). This suggests that the RNA in the denser fractions is qualitatively different from that in the primary RNA peak and may not be of viral origin.

Extrapolation from the frequency distribution of unique RdRp amino acid sequences using the mean Chao 1 estimator suggested that there were on the order of 600 to 1,000 phylotypes (95% confidence interval [CI] from 500 to 1,500) if extrapolating from the shorter sequences (2009 and 2010 samples) and around 400 sequences (95% CI from 200 to 1,000) if extrapolating from the longer sequences (2009 sample only). To put the sequence diversity into a taxonomic context, we clustered the sequences to a conservatively defined species level (≥68% amino acid [aa] identity) (18), which resulted in a minimum of 39 and 21 different clusters in the 2009 and 2010 libraries, respectively, with five of those appearing in both libraries. Extrapolation using the Chao 1 estimator for the 2009 sample resulted in an estimated species-level richness of 300 (95% CI from 100 to 900). This analysis could not be done for the 2010 sample, or for any of the reovirus-like sequences, because all of the sequences retrieved were unique in those instances.

None of the RdRp nucleotide sequences that we obtained from Kane‘ohe Bay were identical to sequences previously derived from waters of coastal British Columbia. Nor were there identical nucleotide sequences shared between the 2009 and 2010 samples. However, three of the nucleotide sequences from the 2009 sample in this study were identical to sequences from a prior sampling of...
Ka‘e’ohe Bay in 2006. After translation, five 2009 RdRp phylotypes were identical to phylotypes from the 2010 sample.

The RdRp amino acid sequences in our libraries spanned the distance represented in previous targeted gene surveys (17, 18) and include some new, deeply branching groups. Most sequences did not cluster near sequences from the few cultivated representatives. Since viral RdRp sequences tend to cluster based on host phylogeny (18), the phylogenetic distances among the RdRp sequences suggests that there are a great many protists from diverse clades in seawater that are being lysed by RNA viruses at any given time.

**Assembly and analysis of picornavirad-like genomes.** Six complete or near-complete genomes were assembled from the metagenomic libraries (Fig. 4). These contigs (KB2009_con55, KB2009_con15, KB2009_con28, KB2009_con74, and KB2009_con88 and KB2010_con16) ranged in size from 8,330 to 9,465 bp (mean, 9,008 bp) and had GC contents ranging from 36.4% to 46.8% (mean, 41.5%). Primer pairs designed to uniquely
amplify a region within the RdRp gene of each assembled genome resulted in amplicons of the expected size and predicted sequences when the original RNA extract was used as the template (data not shown). Each genome contained either one or two large open reading frames (ORFs) that encoded polyproteins ranging in size from 906 to 2,827 amino acids (Fig. 4 and Table 1). These domains were similar to conserved domains of the nonstructural and structural proteins of known ssRNA viruses in the order Picornavirales.

Phylogenetic analysis of the full-length RdRp genes from these assemblies and evidence from comparative genome organization analyses suggested that four of the assembled genomes (KB2009_con28, KB2009_con74, KB2009_con55, and KB2010_con16) were most similar to those of known diatom-infesting viruses, including the three classified members of the genus Bacillarnavirus (maximum-likelihood support value = 99) (Fig. 5). The RdRp of another near-complete genome (KB2009_con15) was most closely affiliated (maximum-likelihood support value = 100) with members of the family Dicistroviridae. One other genome (KB2009_con88) was more divergent and did not form any well-supported clades with any other viruses in the analysis.

Two of the assembled genomes (KB2009_con55 and KB2010_con16) were nearly identical (99.8%). The 2009 genome was 361 bp shorter than the 2010 genome, missing 231 bp on the 5’ end and 130 bp on the 3’ end, presumably a result of incomplete sequence coverage. The assembly of these genomes in which nearly all (19 of 22) differences were synonymous substitutions provides some confidence in the assemblies and suggests that the genomes derive from functionally equivalent strains. Whether these phylotypes coexist, or whether one replaced the other over time, we cannot discern from the present data. A high degree of genome sequence conservation over time (96% to 98% nucleic acid identity) was also observed among DNA-containing viruses over a 10-year period in coastal California (36). Although the collection dates for our samples were only 10 months apart, the data

FIG 3  Reovirid RdRp tree. An unrooted ML tree based on an alignment of sequences containing regions 1 to 4 conserved in the RdRp of reovirids (48) from this study and sequences from representative viruses from the established genera in the family Reoviridae is shown. All sequences above and below the horizontal line are in the subfamilies Spinareovirinae and Sedoreovirinae, respectively. The specific sequences in the alignment and the full names of the virus species represented by abbreviations in the figure are listed in Table S1 in the supplemental material. The scale bar is the equivalent of 0.3 substitutions per site. The statistical support values shown are percentages greater than 80 as calculated by the aLRT method.
suggest that there could be similar genome stability among planktonic RNA viruses.

Each of the six assembled genomes contained a putative non-structural gene with the highly conserved motifs of a nucleotide triphosphate (NTP)-binding domain and another gene with significant sequence similarity to the catalytic center of a family of +ssRNA virus RdRps. We located a region with sequence similarity to a family of viral 3C cysteine proteases in only two of the genomes (KB2009_con55 and KB2010_con16). The syntenous regions of the other genomes were of similar sizes but had no significant similarity to a known protease in this region to map its location.

Of the nine genomes that have been assembled from marine RNA metagenomes—six from this study and three from a previous study (18)—all but one are related to the genomes of viruses in the order Picornavirales, which is consistent with the relatively high frequency of picornavirus-like RdRps sequences in the libraries. The assembled genomes from this study share several characteristics with picornaviruses. These include a monopartite or bipartite genome, the helicase-protease-replicase nonstructural gene order, and a poly(A) tail (37). The nonstructural gene cassette is closest to the 5' end and is followed by a gene block of structural genes. This configuration is similar to the gene order of viruses in the genus Bacillarnavirus and families Dicistroviridae and Marnaviridae, all taxa within the Picornavirales.

The number of sequences recruiting to various contigs provides some clues about the composition of the RNA viral community (Table 1). Based on the predicted size and the average cover-
TABLE 1 Summary information on assembled genomes

<table>
<thead>
<tr>
<th>Genome name</th>
<th>Sample</th>
<th>Genom size (bp)</th>
<th>% GC</th>
<th>Avg coverage (no. of reads)</th>
<th>ORF 1 size (bp)</th>
<th>ORF 1 size (aa)</th>
<th>ORF 2 size (bp)</th>
<th>ORF 2 size (aa)</th>
<th>5' UTR size (bp)</th>
<th>IGR size (bp)</th>
<th>3' UTR size (bp)</th>
<th>% UTR</th>
<th>Poly(A) tail</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>KB2010_con16</td>
<td>2010</td>
<td>9,465</td>
<td>38.9</td>
<td>529</td>
<td>5,388</td>
<td>1,796</td>
<td>2,718</td>
<td>906</td>
<td>773</td>
<td>296</td>
<td>289</td>
<td>14.3</td>
<td>Y</td>
<td>13</td>
</tr>
<tr>
<td>KB2009_con55</td>
<td>2009</td>
<td>9,104</td>
<td>38.7</td>
<td>115</td>
<td>5,388</td>
<td>1,796</td>
<td>2,718</td>
<td>906</td>
<td>773</td>
<td>296</td>
<td>289</td>
<td>14.9</td>
<td>Y</td>
<td>2</td>
</tr>
<tr>
<td>KB2009_con15</td>
<td>2009</td>
<td>9,387</td>
<td>46.8</td>
<td>169</td>
<td>8,481</td>
<td>2,827</td>
<td>NA</td>
<td>NA</td>
<td>585</td>
<td>NA</td>
<td>298</td>
<td>9.4</td>
<td>Y</td>
<td>3</td>
</tr>
<tr>
<td>KB2009_con28</td>
<td>2009</td>
<td>9,264</td>
<td>45.3</td>
<td>259</td>
<td>8,268</td>
<td>2,756</td>
<td>NA</td>
<td>NA</td>
<td>904</td>
<td>NA</td>
<td>161</td>
<td>11.5</td>
<td>ND</td>
<td>5</td>
</tr>
<tr>
<td>KB2009_con74</td>
<td>2009</td>
<td>8,500</td>
<td>42.6</td>
<td>131</td>
<td>5,415</td>
<td>1,805</td>
<td>2,718</td>
<td>906</td>
<td>61</td>
<td>259</td>
<td>47</td>
<td>4.3</td>
<td>ND</td>
<td>2</td>
</tr>
<tr>
<td>KB2009_con88</td>
<td>2009</td>
<td>8,330</td>
<td>36.4</td>
<td>81</td>
<td>8,127</td>
<td>2,709</td>
<td>NA</td>
<td>NA</td>
<td>28</td>
<td>NA</td>
<td>138</td>
<td>2.0</td>
<td>Y</td>
<td>1</td>
</tr>
</tbody>
</table>

*Estimated minimum genome size, % GC content, average coverage, the sizes of open reading frames (ORFs) and untranslated regions (UTRs), and whether a poly(A) tail was evident are presented for each of the six genomes assembled from the metagenomic libraries. IGR, intergenic region; NA, not available; ND, not detected; Y, yes.

*Listed in this column are the percentages of total bp assigned to a genome calculated by multiplying the predicted size of the genome (in bp) by the average genome coverage value and dividing by the total no. of bp generated in the library (×100).

age for the assembled genomes and the total number of nucleotides sequenced in the library, the recruitment data imply that the community structures of the RNA viruses in these samples differ. For example, on the basis of this type of analysis, we estimate that KB2010_con16 represents 13% of the total RNA viruses (Table 1). However, these estimates of dominance based on recruitment have considerable uncertainty because of the possibility of bias in the production of the metagenomic library as discussed below.

**Comparison of RNA viral metagenomes.** In reciprocal BLAST analyses of the two libraries, 76% of the 2009 sequences had significant similarity (E value ≤ 10^-5) to sequences from the 2010
library and 84% of the 2010 sequences had significant similarity to those from 2009. An intercomparison of these Käne’ohe Bay metagenomes and two other aquatic metagenomes targeting RNA viruses, one from coastal British Columbia (25) and one from an artificial lake in Maryland (24), revealed that sequences most similar to +ssRNA viral genomes outnumbered those matching dsRNA viral genomes in all cases (Fig. 6). Furthermore, no am-bisense, –ssRNA, or retroviral sequences were detected in any of the libraries.

Most of the +ssRNA in the marine samples, but not those in the freshwater sample, were assigned to the order Picornavirales. Among the picornavirad-like sequences, those with similarity to sequences of dicistrovirids and JP-B (a putative viral genome of unknown affiliation assembled from a marine sample) were identified in all four of the metagenomes. Sequences related to the diatom-infecting bacillarnaviruses were common in the Käne’ohe Bay libraries but were not detected in the coastal British Columbia or freshwater libraries. The relative high representation of bacillarnavirus-like sequences in the libraries from Käne’ohe Bay is consistent with the importance of diatoms in this system, blooms of which often dominate the eukaryotic phytoplankton community (38).

Also notable was the detection, only in the freshwater library, of sequences most similar to those of viruses that infect land plants (e.g., tombusvirids, sobemovirids, and members of the Virgaviridae) or insects (iflavirids). The detection of sequences similar to those of plant and insect viruses in the lake, but not the sea, might be attributable in part to viruses entering this shallow retention basin in terrestrial runoff (24). However, many of the novel virus sequences recovered may derive from uncharacterized viruses that infect benthic or planktonic freshwater organisms.

The predominance of +ssRNA virus-like sequences in all of the libraries analyzed may reflect a higher relative abundance of these types of RNA viruses in aquatic environments, but we cannot yet rule out biases from the steps involved in library preparation. The amplification method we used in this study was found to introduce little intragenomic bias when used to sequence individual viruses having a variety of genome configurations (39), but there is no data on the relative efficiencies with which ssRNA versus dsRNA genomes are recovered in a mixture of the two. A hypothesis that has yet to be tested is that the apparent low representation of dsRNA viruses in all of the libraries reflects interference from the reannealing of complementary strands during reverse transcription. Even if this were found to be true, it does not explain the dearth of sequences similar to those of other viruses having single-stranded RNA genomes (negative sense, ambisense, retroviral). Ultimately, quantitative assays of the major viral groups identified by metagenomic analyses will be needed to confirm the apparent dominance of picornavirads among marine RNA viruses.
Conclusion. The six genomes assembled as part of this study represent a significant increase in the number of marine picornavirid genome sequences available and will be useful for designing future experiments to test the ecological contributions of these viruses.

The apparent dominance of picornavirid-like sequences in our samples from coastal tropical waters is consistent with a previous metagenomic analysis of RNA viruses in coastal temperate waters (25), despite the use of very different methods in the two studies for each of three major steps: viral harvesting (ultrafiltration versus flocculation), amplification (sequence-independent single-primer amplification [SISPA] versus linker ligation), and sequencing (454 versus Sanger). This suggests that, as documented for metagenomes of DNA viruses (31, 40), the results are robust with respect to at least some methodological biases and that picornavirids may generally dominate the pool of RNA viruses in the ocean (but perhaps not in freshwater). However, quantitative analyses of additional marine habitats (e.g., polar, open ocean) and a thorough evaluation of potential biases during reverse transcription are needed to either bolster or banish this incipient paradigm.

MATERIALS AND METHODS

A description of the study site and methodological details concerning sample collection, processing, and construction of the metagenome was provided in a companion study (4). Summaries of those elements are provided here, as a convenience to the reader, along with more-detailed descriptions of the analyses specific to this paper.

Sample collection and processing. Surface seawater (<0.5-m depth) was collected on 1 August 2009 (35 liters) and 3 June 2010 (40 liters) in polycarbonate carboys from a pier in Kāne‘ohe Bay, Hawai‘i (21°25′46.80″N, 157°47′31.51″W), a reef-protected, subtropical embayment located on the windward side of O‘ahu. Samples were filtered (Sterivex; Millipore) (0.22 μm pore size) to remove cells and larger particles, and then viruses in the filtrate were concentrated by chemical flocculation and filtration (41) followed by centrifugal ultrafiltration (Amicon 15; Millipore) (30 kDa). Viruses from concentrated, virus-enriched samples were purified using sequential step and continuous CsCl buoyant density gradients (29). Fractions of approximately 0.5 ml were collected from the continuous gradient (22 to 23 fractions per gradient). After buffer exchange into TE (10 mM Tris, 1 mM EDTA, pH 8), nucleic acids were extracted from a portion of each density fraction (QiAamp MinElute Viral Spin kit; Qiagen). Each nucleic acid extract was treated with DNase to remove any copurified DNA, and the RNA concentration was determined by fluorometry (4).

Amplification, cloning, and sequencing of RdRp genes. Reverse-transcription PCR (RT-PCR) with two sets of degenerate primers targeting marine picorna-like virus subclades 1 and 2 (Mplsc1 and Mplsc2; see Table S2 in the supplemental material) was performed with RNA template from each buoyant density fraction according to the protocol described by Culley and Steward (18). The resulting endpoint PCR products were separated on a 1% agarose gel and visualized on a digital gel documentation system. The intensity of PCR amplification in each fraction was measured using Molecular Imaging Software (Kodak). Amplified products from the 1.38 g ml⁻¹ and the 1.49 g ml⁻¹ buoyant density fractions were excised from the gel and purified separately (MinElute Gel Extraction kit; Qiagen). The ends of the purified products were repaired (PCRTerminator End Repair kit; Lucigen) and ligated into the pSMART-HCKan vector (Lucigen). Ligated vector was transformed into Ecloni 10G Supreme cells (Lucigen) via electroporation using the supplier’s recommended conditions. Clones were screened for insertions by PCR amplification, and products in the correct size range were purified and sequenced by Sanger sequencing with fluorescent dye terminators (Applied Biosystems).

Quantification of RdRp phylotypes in the buoyant density gradient. Reverse transcription quantitative PCR (RT-qPCR) was used to determine the abundances of two RNA virus phylotypes identified in the 2009 sample. Primers were designed to target regions of the RdRp unique to each phylotype (see Table S2 in the supplemental material). Reactions were performed with RNA template from each of the RNA viral buoyant density fractions from the 2009 sample. Reaction mixtures for cDNA synthesis (Superscript III; Invitrogen Corporation) consisted of 5 μl of the extracted, DNase-treated RNA template, a 0.2 mM concentration of each deoxynucleoside triphosphate, and a 0.5 μM concentration of reverse primer. Samples were denatured at 65°C for 5 min and cooled on ice and then supplemented with 1× First-Strand Buffer (Invitrogen Corporation), 5 mM dithiothreitol, 40 U RNase (RNaseOUT; Invitrogen Corporation), and 200 U reverse transcriptase (Superscript III; Invitrogen Corporation) to obtain a final reaction mixture volume of 20 μl. The reaction mixtures were brought to 55°C for 60 min and to 70°C for 15 min as a final denaturation step and were then supplemented with 1 μl RNase H (Invitrogen Corporation) and incubated for 20 min at 37°C. The qPCR amplification was performed on a 7300 real-time PCR system (Applied Biosystems) with Power SYBR green PCR Mastermix (Applied Biosystems). The reaction mixtures contained 12.5 μl SYBR green PCR master mix (Applied Biosystems), a 0.2 μM concentration of each primer, and 2 μl of sample cDNA template, with a final volume of 25 μl. For each primer set, reactions were replicated two times, and each set of reaction mixtures contained duplicate samples, standards, and negative controls. Standards consisted of 10-fold serial dilutions (3 × 10⁻⁶ to 3 × 10⁶ molecules per reaction) of target molecule that had been cloned, amplified using appropriate primers, purified by agarose gel electrophoresis, and extracted with a MinElute Gel Purification kit (Qiagen). Quantities of DNA were determined fluorometrically with a Quant-IT DNA Assay kit (Invitrogen Corporation). The thermal cycling protocol consisted of a denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 s, annealing at the primer-specific temperature (see Table S2) for 30 s, extension at 72°C for 35 s, and a final extension at 72°C for 14 min and 25 s. The specificity of each primer set was verified by an analysis of the DNA melting curve of the amplification products over the course of the reaction as well as independent experiments in which the amplification efficiency of the primers was determined in samples spiked with various amounts of nontarget template.

Metagenome construction. For each sampling date, RNA purified from density fractions in the range from 1.38 to 1.49 g ml⁻¹ was pooled and subjected to a random-priming-mediated sequence-independent single-primer amplification (RP-SISPA) reaction (39, 42). After size selection (500 to 1,000 bp) on an agarose gel, RP-SISPA products were sequenced with a GS FLX Titanium platform (Roche Diagnostics Corporation). Sequences were run through a quality-control pipeline (43) to remove short and low-quality reads and presumed artificial replicates, and the ends were trimmed to remove any remaining primer sequences. Rarefaction curves for each RNA viral metagenome were generated in META-VIR (44) using equal samplings of the unassembled, quality-controlled sequences and a clustering percentage of 75.

Identification and analysis of RdRp phylotypes in the metagenome. A hidden Markov model was built and used to identify RdRp-like sequences in each of the metagenomes using HHMNER v 3.0 (45). Markov models were first produced based on amino acid alignments of conserved RdRp regions (46) of representative picornavirid-like viruses and conserved RdRp regions (47) of representative reovirids (see Table S1 in the supplemental material). Sequences fitting the criteria of the models were then subjected to a more detailed analysis using HMMER v 3.0 (45). A profile HMM was first produced based on amino acid alignments of conserved RdRp regions (46) of representative picornavirid-like viruses and conserved RdRp regions (47) of representative reovirids (see Table S1 in the supplemental material). Sequences fitting the criteria of the models were then subjected to a more detailed analysis using HMMER v 3.0 (45). A profile HMM was produced for each phylotype (see Table S2 in the supplemental material). Reactions were performed with RNA template from each of the RNA viral buoyant density fractions from the 2009 sample. Reaction mixtures for cDNA synthesis (Superscript III; Invitrogen Corporation) consisted of 5 μl of the extracted, DNase-treated RNA template, a 0.2 mM concentration of each deoxynucleoside triphosphate, and a 0.5 μM concentration of reverse primer. Samples were denatured at 65°C for 5 min and cooled on ice and then supplemented with 1× First-Strand Buffer (Invitrogen Corporation), 5 mM dithiothreitol, 40 U RNase (RNaseOUT; Invitrogen Corporation), and 200 U reverse transcriptase (Superscript III; Invitrogen Corporation) to obtain a final reaction mixture volume of 20 μl. The reaction mixtures were brought to 55°C for 60 min and to 70°C for 15 min as a final denaturation step and were then supplemented with 1 μl RNase H (Invitrogen Corporation) and incubated for 20 min at 37°C. The qPCR amplification was performed on a 7300 real-time PCR system (Applied Biosystems) with Power SYBR green PCR Mastermix (Applied Biosystems). The reaction mixtures contained 12.5 μl SYBR green PCR master mix (Applied Biosystems), a 0.2 μM concentration of each primer, and 2 μl of sample cDNA template, with a final volume of 25 μl. For each primer set, reactions were replicated two times, and each set of reaction mixtures contained duplicate samples, standards, and negative controls. Standards consisted of 10-fold serial dilutions (3 × 10⁻⁶ to 3 × 10⁶ molecules per reaction) of target molecule that had been cloned, amplified using appropriate primers, purified by agarose gel electrophoresis, and extracted with a MinElute Gel Purification kit (Qiagen). Quantities of DNA were determined fluorometrically with a Quant-IT DNA Assay kit (Invitrogen Corporation). The thermal cycling protocol consisted of a denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 s, annealing at the primer-specific temperature (see Table S2) for 30 s, extension at 72°C for 35 s, and a final extension at 72°C for 14 min and 25 s. The specificity of each primer set was verified by an analysis of the DNA melting curve of the amplification products over the course of the reaction as well as independent experiments in which the amplification efficiency of the primers was determined in samples spiked with various amounts of nontarget template.
regions 1 to 4 conserved in the RdRP of reovirids (48). Translated picornavirid-like RdRP sequences were clustered at the putative species level based on a conservative phylogenetic distance criterion as previously described (18). In essence, the greatest distance between any two officially classified strains of picornavirids belonging to the same species (Human Rhinovirus 2 and Human Rhinovirus 89) was taken as the species threshold. Any sequences whose distance from one another was lesser than the distance represented by that threshold were considered members of the same species and those whose distance from one another was greater than the distance represented by that threshold were considered to be different species. The Chao 1 estimator (49) of total phylotype or species richness was calculated with EstimateS (50). In light of the large uncertainty in the estimator, values were rounded to the nearest hundred.

Maximum-likelihood trees were constructed with PhyML (51) from protein sequences aligned with MAFFT (52) using the auto function.

**Sequence assembly and library comparison.** Sequences were assembled using CLC Genomics Workbench version 5.0 and the following parameters: global alignment, a minimum contig length of 200, mismatch, insertion, and deletion costs set to 3, length fraction set to 0.5, similarity threshold set to 0.8, automatic word value set to 20, and bubble size set to 50. Assembly statistics were presented elsewhere (4). Contigs and singletons from the assembled libraries were classified with MEGAN (53) based on blastx (54) searches of the NCBI nonredundant nucleotide database where hits with an E value \( \leq 10^{-5} \) were considered significant. We used this conservative, but frequently used, cutoff value to further reduce the likelihood of the misclassification of a sequence. All of the individual reads comprising a given contig inherited the taxonomic assignment given to the contig.

We compared the KB libraries by BLAST analysis (blastx; \( E \) value, \( \leq 10^{-5} \) cutoff) where the unassembled reads from one library were used to query the second library and vice versa.

**Analysis of assembled genomes.** Open reading frames were identified with the heuristic approach for gene prediction described by Besemer and Borodovsky (55). Searches of the Conserved Domain Database (CDD)-NCBI database (56) were conducted with the translated ORFs from each genome (Table 1). Searches of a database of experimentally verified internal ribosomal entry site (IRES) structures (57) were conducted with the untranslated regions (UTR) of the six genomes.

**Accession numbers.** Metagenomic data referred to in this paper are available at the CAMERA website (http://camera.calit2.net/) under accession number CAM_PROJ_BROADPHAGE and sample names CAM_SMPL_000815 (1 August 2009) and CAM_SMPL_000824 (3 June 2010). The accession numbers of the MPL phytophysotypes described in this research are listed in Table S1 in the supplemental material.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01210-14/-/DCSupplemental.

Figure S1, EPS file, 0.7 MB.
Figure S2, EPS file, 2.3 MB.
Table S1, PDF file, 0.4 MB.
Table S2, PDF file, 0.1 MB.

**ACKNOWLEDGMENTS**

We thank Gordon O. Walker for assistance with sampling and Shulei Sun and the Broad Institute for assistance with the processing and sequencing of the viromes.

This research was funded in part by the Gordon and Betty Moore Foundation through grant GBMF1799 to the Broad Institute. Samples G4008 (KB2009) and G8294 (KB2010) were sequenced at the Broad Institute. Bioinformatics analyses were supported in part through the use of the COBRE/INBRE Bioinformatics core, supported jointly by NIH Grants from the National Institute of General Medical Sciences (P20GM103516 and P20GM103466). This work was supported by NSF grants to G.F.S. and A.I.C. (OCE 08-26650) and the Center for Microbial Oceanography Research and Education (EF 04-24599).

**REFERENCES**
