Bacteriophages are considered to be the most abundant biological entities on the planet. The *Siphoviridae* are the most commonly encountered tailed phages and contain double-stranded DNA with an average genome size of ~50 kb. This paper describes the isolation from four different activated sludge plants of the phage RRH1, which is polyvalent, lysing five *Rhodococcus* species. It has a capsid diameter of only ~43 nm. Whole-genome sequencing of RRH1 revealed a novel circularly permuted DNA sequence (14,270 bp) carrying 20 putative open reading frames. The genome has a modular arrangement, as reported for those of most *Siphoviridae* phages, but appears to encode only structural proteins and carry a single lysis gene. All genes are transcribed in the same direction. RRH1 has the smallest genome yet of any described functional *Siphoviridae* phage. We demonstrate that lytic phage can be recovered from transforming naked DNA into its host bacterium, thus making it a potentially useful model for studying gene function in phages.

**MATERIALS AND METHODS**

**Bacterial strains and media.** The bacterial strains used in this study are listed in a study by Petrovski et al. (31). All were grown in peptone yeast extract calcium (PYCa) broth or agar (31). All chemicals unless otherwise noted were obtained from Sigma, Australia.

**Phage purification, host range determination, and single-step growth curves.** Phage recovery and purification were performed using *R. rhodochrous* (Rrho39), *R. equi* (Requ10), and *R. erythropolis* (Rery29) as described by Petrovski et al. (32). Ten rounds of phage dilution and single plaque isolation were performed to ensure the final phage suspension resulted from a single virion, before undertaking any further characterization studies.

**Electron microscopy.** RRH1 virions (with and without added T2 phage) were allowed to absorb to Formvar coated 200-mesh copper grids for 5 min. These were washed twice for 1 min in double-distilled water (ddH2O) and negatively stained with 2% (wt/vol) uranyl acetate for 2 min. Excess liquid was absorbed by filter paper and the grids allowed to air dry before being examined under a JEOL JEM-100cx transmission electron microscope (TEM) at an accelerating voltage of 100 kV. Electron micrographs were calibrated using catalase crystals (Electron Microscopy Sciences, Hatfield, PA), and the known *Escherichia coli* from transformed naked phage DNA demonstrated here, offer an exciting new model for better understanding phage biology.
coli Myoviridae phage T2. RRH1 particle sizes were determined by measuring 40 symmetrical phage capsid vertices and 10 tail lengths.

**DNA isolation and sequencing.** RRH1 phages were purified with NaCl-polyethylene glycol (PEG) 8000, and DNA was isolated by using SDS-protease K as described previously (31). Four independent isolates of RRH1 from different geographical locations were sequenced using a Roche GS FLX genome sequencer and Titanium chemistry by Genoseq (UCLA, Los Angeles, CA). The pyrosequencing reads were assembled separately using the gsAssembler (Roche Applied Science, Indianapolis, IN). The resulting single contigs from each isolate had a minimum of 50-times sequence read coverage.

**Genome annotation.** The genome of RRH1 was annotated using the Integrative Services for Genomic Analysis (http://isga.cgb.indiana.edu) (21) interface to the Etaggis (29) software package, following by manual inspection of all gene predictions.

Putative open reading frames (ORFs) longer than 90 bases were predicted using Glimmer3 (13) with the iterative process described by Delcher et al. (12) to enhance predictive accuracy. All predicted start codons were inspected manually for the presence of putative ribosomal binding sites and adjusted as required.

Sequence similarity searches were performed using BLAST X against a nonredundant database, including data sourced from NCBI, Swiss-Prot, and PDB databases by using a significance value of 1e−04. The BLAST X results were used as input for the BLAST-Extend-Repraze algorithm (http://sourceforge.net/projects/bert/) to identify potential frame shifting or point mutations. Protein domain searches were performed using hmmpfam (http://hmmer.janelia.org/) against the PFAM (5) and TIGRFAM HMM (17) databases to identify protein family or domain matches. Each ORF was also checked manually using the conserved domain database (CDD) (14). Possible transmembrane domains were recognized using DAS (dense alignment surface method) transmembrane prediction (http://www.sbc.su.se/~miklos/DAS/) (11). Identified ORFs were also screened for the presence of lipoprotein motifs with the predicted protein sequence (6). Any possible tRNAs and tmRNA genes were screened by using tRNAscan-SE (24, 37).

**Mass spectroscopy.** To identify phage structural proteins, purified virions (~10^-8 PFU) were precipitated from concentrated stocks by using ZnCl₂ (36) to remove contaminating PEG. The pellet was reduced using 100 mM dithiothreitol (DTT) and heat denatured (100°C for 5 min). Samples were loaded into a 12% SDS-polyacrylamide gel electrophoresis gel prior to staining with Coomassie brilliant blue. The entire lane containing all visible protein bands was excised from the gel to create a protein pool. The excised gel was trypsin digested (40) followed by analysis using electrospray ionization (ESI)-time of flight mass spectrometry (TOF-MS) by the Mass Spectroscopy and Proteomic Facility at La Trobe University.

**Electroporation of phage DNA into Rhodococcus erythropolis.** Electrotropic R. erythropolis (Rery29) cells were prepared as described by Sekizaki et al. (38) and stored at ~70°C until required. Aliquots (100 μl) were thawed on ice prior to electroporation using 100 to 500 ng of RRH1 genomic DNA at 2.5 kV/cm. The transformed cells were incubated at 30°C for 2 h in 1 ml PYCa medium, plated onto lawn plates of R. erythropolis (Requ10), and incubated for 2 days prior to being examined for plaques.

**Nucleotide sequence accession number.** The nucleotide sequence for RRH1 has been deposited in GenBank under accession number JN116822.

**RESULTS AND DISCUSSION**

**Isolation and general features of RRH1 phage.** Four phage isolates were obtained using three Rhodococcus strains as hosts (two lysed R. rhodochrous [Rrho39], one lysed R. equi [Requ10], and one lysed R. erythropolis [Rery29]). These phage isolates were obtained from samples of activated sludge from three wastewater treatment plants in Victoria, Australia (Daylesford, Werribee, and Ballarat), and from one in Queensland, Australia (Nambour). Sequencing revealed that all four isolates had identical genomes (see below), suggesting that this phage RRH1 may be widely present in Australian activated sludge communities.

Electron microscopy revealed that RRH1 has a noncontractile tail (~81 ± 1.9 nm) and an isometric capsid (~43 ± 1.2 nm) (Fig. 1A and B). Its morphology is characteristic of Siphoviridae phages, with a capsid among the smallest ever reported for a member of this family.

The average burst size was calculated to be 45 ± 5 particles per infective center, with a latency period of ~2 h in PYCa broth at 30°C. RRH1 phage displays a broad host range within the genus Rhodococcus. From the bacterial strains tested in our collection, RRH1 can propagate lytically on R. equi (Requ10), R. equi (Requ28), R. equi (Requ10), R. erythropolis (Rery29), R. globerus (Rglo35), R. rhodinii (Rrho46), and R. rhodochrous (Rrho39).

**Genome sequencing of RRH1.** These four RRH1 phages were each sequenced using Roche/454 pyrosequencing (see Materials and Methods). The data generated over 10,000 reads for each, with a minimum of 50-times sequence coverage, and identical DNA sequences emerged. The assembled sequences of all four phages were consistent with a circularly permuted, double-stranded DNA (dsDNA) genome, a proposition confirmed by PvuII restriction digest profiles (data not shown). Consequently, only that of RRH1 phage isolated on R. rhodochrous from the Daylesford treatment plant is detailed here. Its genome consists of 14,270 bp with a G+C content of 68.3 mol% that falls within the range 67 to 73 mol% G+C for Rhodococcus species (16). A comparison of the RRH1 genome with those of other Siphoviridae family members revealed that this phage has the smallest genome described to date. The most similar phage at the DNA level is a putative prophage sequence found in the complete genome of Nocardia farcinica IFM 10152 (23), but their similarity is low (5%). The remainder of the RRH1 genome shares no identity with any other sequenced genome.

The RRH1 genome carries 20 putative genes larger than 90 nucleotides and contains no tRNAs (Table 1). The ORFs are numbered consecutively from orf1 to orf20. All 20 genes are located on the same strand and transcribed in the same direction (Fig. 2). Only 12 ORF products share significant homology with other protein sequences in the GenBank database, and only 6 could be annotated functionally. The RRH1 phage genome appears to be modularly organized (as is typical for the Siphoviridae), consisting of gene clusters involved in DNA packaging, cell lysis, and head and tail morphogenesis (Fig. 2). Each is discussed below.

**DNA packaging module.** The putative large terminase appears to be encoded by orf2, as it shares significant similarity with a large terminase found in a prophage of Corynebacterium diphtheriae, as well as containing the motif pfam03354 characteristic of large terminases. This gene product is essential for packaging DNA into the phage head during phage replication (34). Typically, the large
terminase functions as a complex with a small terminase. The latter is responsible for determining specificity of DNA binding (9), while the large terminase performs DNA cleavage after the packing into the prohead (15). Such terminase genes are typically transcribed together in an operon-like structure, in which the latter is responsible for determining specificity of DNA binding (8), and this arrangement is also the case with RRH1.

Structural and lysin modules. The portal protein appears to be encoded by orf4, based on the presence of the distinctive conserved motif pfam04860. The portal protein gene is typically the first gene in the structural protein module of Siphoviridae phages (8), and this arrangement is also the case with RRH1. Phage portal protein genes are typically followed by genes encoding capsid and tail proteins (8). This genomic structure is not maintained in RRH1, in which a putative lysin gene (orf5) is found downstream of orf4. While the predicted amino acid sequence of Orf5 is most closely related to that of a hypothetical protein in a putative prophage of Rhodococcus erythropolis, it also shares identity with the putative lysin proteins of the Mycobacterium phage Bethlehem and other related Mycobacterium phages (19). In addition, Orf5 contains the pfam01471 motif, composed of three alpha helices and other related motifs, which are associated with peptidoglycan binding and bacterial cell wall degradation (7). On the basis of these data, we conclude that orf5 encodes a lysin that is an unusual location within the gene module encoding phage structural proteins.

TABLE 1 Summary of genes carried by RRH1

<table>
<thead>
<tr>
<th>ORF</th>
<th>Genome coordinates</th>
<th>Protein function (Pfam)</th>
<th>Match (%) identity</th>
<th>Protein size in kDa (no. of TM domains)</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>219–599</td>
<td>Putative small terminase</td>
<td></td>
<td>13.3 (1)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>655–1965</td>
<td>Large terminase (pfam03354)</td>
<td>Terminase protein, Corynebacterium diphtheriae (44)</td>
<td>46.5 (0)</td>
<td>7e^-85</td>
</tr>
<tr>
<td>3</td>
<td>1995–2327</td>
<td>Unknown</td>
<td>Hypothetical protein, Gordonia neofaciens (58)</td>
<td>11.9 (0)</td>
<td>3e^-107</td>
</tr>
<tr>
<td>4</td>
<td>2471–3589</td>
<td>Portal protein (pfam04860)</td>
<td>Hypothetical protein, Corynebacterium diphtheriae (52)</td>
<td>40.5 (1)</td>
<td>7e^-54</td>
</tr>
<tr>
<td>5</td>
<td>3586–4320</td>
<td>Lysin (pfam01471)</td>
<td>Hypothetical protein, Rhodococcus erythropolis (52)</td>
<td>26.7 (0)</td>
<td>7e^-61</td>
</tr>
<tr>
<td>6</td>
<td>4308–4442</td>
<td>Unknown</td>
<td></td>
<td>4.5 (1)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4439–6424</td>
<td>Prohead protease (pfam04586)</td>
<td>Hypothetical protein, Gordonia neofaciens (41)</td>
<td>68.4 (4)</td>
<td>7e^-113</td>
</tr>
<tr>
<td>8</td>
<td>6428–6784</td>
<td>Unknown</td>
<td>Hypothetical protein, Gordonia neofaciens (44)</td>
<td>12.3 (0)</td>
<td>5e^-14</td>
</tr>
<tr>
<td>9</td>
<td>6793–7125</td>
<td>Putative structural protein</td>
<td>Hypothetical protein, Gordonia neofaciens (60)</td>
<td>15.0 (0)</td>
<td>2e^-41</td>
</tr>
<tr>
<td>10</td>
<td>7140–7571</td>
<td>Unknown</td>
<td>Hypothetical protein, Gordonia neofaciens (60)</td>
<td>11.8 (0)</td>
<td>6e^-14</td>
</tr>
<tr>
<td>11</td>
<td>7884–8024</td>
<td>Putative structural protein</td>
<td>Hypothetical protein, Gordonia neofaciens (36)</td>
<td>5.3 (0)</td>
<td>3e^-105</td>
</tr>
<tr>
<td>12</td>
<td>8023–8324</td>
<td>Unknown</td>
<td>Hypothetical protein, Corynebacterium diphtheriae (38)</td>
<td>10.9 (1)</td>
<td>2e^-75</td>
</tr>
<tr>
<td>13</td>
<td>8442–10040</td>
<td>Putative tape measure protein</td>
<td>Hypothetical protein, Corynebacterium diphtheriae (32)</td>
<td>53.3 (6)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>10037–11458</td>
<td>Unknown</td>
<td>Hypothetical protein, Corynebacterium diphtheriae (32)</td>
<td>52.3 (2)</td>
<td>1e^-52</td>
</tr>
<tr>
<td>15</td>
<td>11472–12083</td>
<td>Putative structural protein</td>
<td></td>
<td>21.5 (0)</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>12087–12329</td>
<td>Unknown</td>
<td></td>
<td>9.0 (1)</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>12510–13157</td>
<td>Unknown</td>
<td></td>
<td>23.4 (0)</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>13348–13467</td>
<td>Unknown</td>
<td></td>
<td>4.2 (1)</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>13445–13912</td>
<td>Putative structural protein</td>
<td></td>
<td>16.0 (1)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>13909–14184</td>
<td>HNH endonuclease (pfam01844)</td>
<td>HNH endonuclease, Corynebacterium diphtheriae (62)</td>
<td>10.0 (0)</td>
<td>3e^-24</td>
</tr>
</tbody>
</table>

*ORFs were numbered consecutively.

*Predicted function is based on amino acid identity, conserved motifs, (ESI-)TOF-MS, and gene location within functional modules.

*The most closely related gene (only if named) and the name of the organism. The percent identity was based on the best match when a BlastP analysis was performed.

*The predicted protein size and the predicted transmembrane (TM) domains were determined using DAS (11).

*The probability of obtaining a match by chance as determined by BLAST analysis. Only values less than 10^-20 were considered significant.
The N-terminal region of Orf7 contains a pfam04586 motif, characteristic of prohead proteases. The predicted size of Orf7 is 68.4 kDa, but from the SDS-PAGE data (Fig. 3), we were unable to recognize a protein of this size. These data suggest that Orf7 is a fusion protein product involving the prohead protease and the major capsid protein, as reported before for the Lactococcus phage c2 (25). This hypothesis is supported by the observation that one of the most abundant RRH1 phage proteins (and the likely main capsid protein) is only \( \frac{101}{11} \) 40 kDa (Fig. 3).

SDS-PAGE analysis revealed the presence of a major structural protein of approximately 15 kDa. In silico analysis of the 20 putative genes of RRH1 revealed only two putative protein products that match such a size (Orf10 and Orf19). Whole-phage mass spectroscopy revealed 55% peptide coverage of Orf19 but no coverage for Orf10, suggesting that Orf19 is the major structural tail protein. No putative ribosomal slippage sequence in the structural protein genes could be detected, although the sequence has been observed in other Rhodococcus phages (32, 41). The genes orf8 to orf18 have no predicted function, with the sequences of Orf11 and Orf16 to Orf19 sharing no similarity to any other protein. Mass spectroscopy analysis revealed sequences of both the major structural proteins (Orf7 and Orf19) and identified Orf11, Orf13, and Orf15 as minor structural proteins. The largest putative gene in the RRH1 phage genome is orf13. This gene is predicted to encode the tape measure protein, since its closest sequence homologues are those of other phage tape measure proteins.

orf20 encodes a putative endonuclease, based on sequence similarity with a putative HNH endonuclease found in a prophage of...
C. diphtheriae and the presence of the conserved motif pamt01844. Surprisingly, the RRH1 phage genome does not appear to carry any DNA replication genes, such as those encoding helicases, primases, or DNA polymerases. The absence of such gene products suggests that this phage relies on its host to provide all the required genome replication machinery.

It is interesting to note that the RRH1 phage displays a broad host range and relatively small burst size (i.e., phage particles produced per cell infected) for such a small phage. It may be that its limited genome size allows it to function only as a generalist of low efficiency. This hypothesis would suggest that most genes present in larger phage genomes are not essential for lytic replication but are instead important for achieving replication efficiency through host specialization and better use of host resources. Phages like RRH1 would be expected to have a competitive advantage over more specialized phages in environments where there is a high diversity of Rhodococcus spp., since they would be able to exploit the available host resource more fully but be at a competitive disadvantage in environments with a limited range of strains and where host utilization efficiency is paramount. Such a model may explain why both broad- and narrow-host-range phages for the same host can be isolated.

**RRH1 has the smallest known Siphoviridae phage genome.** The genome size of RRH1 is unusually small (14.2 kb). To our knowledge, no functional Siphoviridae phage with a smaller genome has been identified. A search of the NCBI database for complete dsDNA virus genomes using the keywords “bacteriophage” or “phage” identified 537 complete phage genome sequences as of June 2011. The next smallest of the Siphoviridae genomes deposited belongs to the Enterococcus faecalis phage EFRM31, at 16.95 kb (26, 27).

Can RRH1 be used as a model to study phage gene function? RRH1 is genetically very simple, suggesting that it may be of value in aiding our understanding of the Siphoviridae. With the aim of assessing RRH1 phage as a model system, a DNA transformation protocol was developed. Electroporation of naked phage DNA into electrocompliant R. erythropolis cells (Rery29) resulted in viable plaques (~10^7/µg DNA) developing on lawn plates of the host organism. Furthermore, the calculated DNA packaging efficiency for RRH1 is at the lower end for a Siphoviridae phage (33), suggesting that it may be possible to add additional genes to the RRH1 genome. This outcome raises the feasibility of being able to apply standard in vitro molecular biology techniques to study gene function in this phage. Work exploring this attractive possibility is currently in progress.

**Conclusion.** Phages are considered to represent the most genetically diverse biological entities on the planet (18). How little we still know of phage genomics is emphasized by the RRH1 phage described in this paper. Despite its small genome size and limited number of genes, most of its gene products cannot be assigned a function based on sequence similarities. New and improved model systems are needed to better elucidate how phages reproduce and exploit their hosts. The minimal genome size of phage RRH1 combined with its suitability for gene manipulation and other standard in vitro genetic techniques suggests that it may be an excellent candidate for this task.

**ACKNOWLEDGMENTS**

The research was supported by the Australian Research Council Linkage Grant (LP0774913) in collaboration with Melbourne Water (David Gregory) and South East Water (Graham Short), who are thanked for their financial support. Z. A. Dyson was the recipient of an Australian Postgraduate Award Ph.D. scholarship. S. Petrovski was funded by the ARC linkage grant and La Trobe University.

We thank Carolyn Bell and Judy Oliver for providing bacteriophage T2.

**REFERENCES**