Virus Genes (2019) 55:257–265
https://doi.org/10.1007/s11262-018-1625-5

Isolation and characterization of bacteriophage NTR1 infectious for *Nocardia transvalensis* and other *Nocardia* species

Stephanie Taylor[1] · Teagan L. Brown[2] · Joseph Tucci[2] · Peter Lock[3] · Robert J. Seviour[1] · Steve Petrovski[1]

Received: 1 October 2018 / Accepted: 11 December 2018 / Published online: 17 December 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

We describe here the isolation and characterization of the bacteriophage, NTR1 from activated sludge. This phage is lytic for *Nocardia transvalensis, Nocardia brasiliensis* and *Nocardia farcinica*. NTR1 phage has a genome sequence of 65,275 bp in length, and its closest match is to the *Skermania piniformis* phage SPI1 sharing over 36% of its genome. The phage belongs to the *Siphoviridae* family, possessing a long non-contractile tail and icosahedral head. Annotation of the genome reveals 97 putative open reading frames arranged in the characteristic modular organization of *Siphoviridae* phages and contains a single tRNA-Met gene.

Keywords Bacteriophage · *Nocardia* phage · Genome sequence · Phage therapy · Siphoviridae

Bacterial genetic diversity is driven by DNA exchange within microbial communities involving several processes of horizontal gene transfer. One of these, transduction, allows bacteriophages (or phages) to transfer DNA from a donor bacterium to a suitable recipient [1–3]. Phages are believed to have the highest genetic diversity of any known biological entity, and current estimates suggest there are more than 10^{31} phages on Earth, with a predicted 'species' estimate of ca. 100 million [3–7]. With the advent of next generation DNA sequencing, phage genomic data are much more readily generated, and more than 5000 phage genomes have now been sequenced [2, 8–10]. However, comparatively little information is available on their ecological importance, particularly those infective for environmentally important bacteria [11].

The genus *Nocardia* embraces an ecologically ubiquitous group of bacteria, some of which have been implicated in environmental problems including stabilization of foams in wastewater treatment plants because of their high cell surface hydrophobicities [12–14]. Furthermore, some *Nocardia* species are considered to be opportunistic human pathogens[15, 16]. To date, fifty-five *Nocardia* species are known to infect humans [22]. Some of the more prominent include *N. abscessus, N. asteroides, N. brasiliensis, N. farcinica, N. nova* and *N. transvalensis* [17–22]. Referred to collectively as nocardiosis, these infections are more common in immunocompromised patients and are often acquired through infections of the respiratory tract or through skin fissures. These infections then persist as comorbidities, with disease disseminating to the brain, lungs and kidneys [16, 23, 24]. *N. farcinica* has been described as one of the most frequently encountered human pathogens [18], but misdiagnosis of *Nocardia* infections is common mainly because their clinical manifestations are shared with those of other illnesses including tuberculosis, pneumonia, carcinoma and chronic granulomatous disease [19, 25].

Furthermore, the presence of a preexisting disease state within patients can lead to inappropriate or late initiation of treatment, problems thought to contribute to a mortality rate of more than 30% [18, 19, 26]. Antibiotic treatment options are available over long periods (often 6–12 months), but resistance to currently available antibiotics is increasing [4, 19, 26]. Therefore, developing highly targeted alternative forms of treatment is attractive. Of these, phage therapy, exploiting lytic phages targeting pathogenic

Edited by Andrew Millard.

✉ Steve Petrovski
steve.petrovski@latrobe.edu.au

[1] Department of Physiology, Anatomy & Microbiology, La Trobe University, Bundoora, VIC, Australia

[2] Department of Pharmacy & Applied Science, La Trobe Institute for Molecular Science, La Trobe University, Bendigo, VIC, Australia

[3] La Trobe Institute for Molecular Sciences, La Trobe University, Bundoora, VIC, Australia

Nocardia species, is an attractive approach. Phages infective for *Nocardia* species characterized previously include phage NBR1 infective for *N. brasiliensis* and the broad host range phage GTE2 that targets *Gordonia, Rhodococcus* and *Nocardia* [7, 14, 27]. Additional genetic information on other phages lytic for *Nocardia* species may also provide further insight into their evolution and biodiversity.

Activated sludge mixed liquor samples were collected from several wastewater treatment plants in Victoria, Australia, and screened for the presence of phages able to lyse *N. transvalensis* using phage enrichment and plaque plating [27]. Single plaques present on lawn plates of *N. transvalensis* strain CON40 were purified by multiple rounds of single plaque isolation, and the phage eventually obtained is named NTR1. This is the first phage reported as lytic for *N. transvalensis*, and we report its characterization here.

Phage NTR1 belongs to the *Siphoviridae* family in possessing the characteristic long, non-contractile tail (292 ± 9 nm) and an isometric capsid head (60 ± 4 nm) (Fig. 1). This phage was screened against 14 *Nocardia* strains held in our culture collection (Table 1), and plaques were produced on three strains including *N. transvalensis, N. brasiliensis* and *N. farcinica*. These three strains have all been described as pathogens causing Nocardiosis [19].

The genome sequence of phage NTR1 was determined using Illumina technology, generating more than a 100-fold sequence coverage. This phage possesses a dsDNA genome 65,275 bp in size with a G + C content of 68.2 mol%. At the DNA level, it appears to be highly novel, and its closest homologue is phage SPI1, which infects *Skermania piniformis*, although this shared similarity covers only 36% of the genome and 68% sequence homology [28]. Identification of open reading frames was conducted applying a suite of bioinformatics tools and revealed 97 putative open reading frames together with one tRNA-Met (Table 2) [29, 30]. The genome has a typical *Siphoviridae* modular organization, consisting in order of phage DNA packaging genes, structural protein genes, followed by cell lysis and a DNA replication module. The NTR1 DNA sequence was deposited in GenBank with the accession number MF477236.

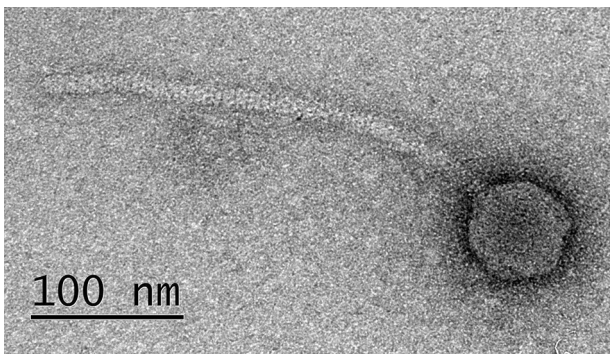

Fig. 1 Electron micrograph of NTR1. Scale bar 100 nm

Table 1 *Nocardia* strains used in this study

Strain reference	Organism	Additional reference
CON4	*Nocardia asteroides*	Q 131
CON12	*Nocardia asteroides*	
CON14	*Nocardia otitidiscaviarum*	AMMRL 19.11
CON15	*Nocardia otitidiscaviarum*	AMMRL 19.12
CON17	*Nocardia carnea*	Queensland State University
CON20	*Nocardia asteroides*	Rockhampton hospital
CON23	*Nocardia asteroides*	DSMZ 43757
CON25	*Nocardia otitidiscaviarum*	DSMZ 43242
CON30	*Nocardia carnea*	DSMZ 43397
CON40	*Nocardia transvalensis*	DSMZ 43405
CON42	*Nocardia brasiliensis*	DSMZ 43758
CON43	*Nocardia brevicatena*	DSMZ 43024
CON47	*Nocardia nova*	ATCC 33726
CON70	*Nocardia farcinica*	

Phage DNA packaging processes generally requires two packaging proteins, which are known as the small and large terminase subunits [31]. The small subunit is essential in determining the specificity of DNA binding, but we were unable to identify such a gene in NTR1 based on shared homology. A large terminase subunit gene was identified by the presence of the conserved motif Pfam 03237, which occurs exclusively in large terminase subunits. In terms of shared homology, this gene is similar to another large terminase gene in a *S. piniformis* phage SPI1 [28]. This subunit is responsible for the cleavage of the phage DNA after packaging into the capsid prohead [32]. Typically, the small terminase gene is located upstream of the large subunit gene [31]. Two genes upstream of the large terminase (*orf1* and *orf2*) are of unknown function and either or both potentially may encode the small terminase subunit. A phylogenetic analysis of the large terminase sequence showed it clustered with *Mycobacterium* phages Roshbush, Qyrzula, PG1 and Cooper all of which have circularly permuted genome ends and belong to cluster B suggesting NTR1 has the same genetic configuration [33].

Tailed phages typically have a modular genome organization [31]. Generally the genes encoding structural proteins are located adjacent to those encoding the terminase subunits [31]. Most genes located between *orf4–orf60* encode putative proteins in other phages are of unknown function, although possession of conserved motif can provide a putative function [7]. These clustered genes in NTR1 appear to encode proteins with a high level of amino

Table 2 Summary of ORFs and gene products in NTR1 phage genome

Gene[a]	Coordinates	Length (bp)	Predicted function (Pfam)[b]	Significant match (% identity)[c]	E-value[d]
1	3…848	846	Nucleoid occlusion protein. Related to ParB partitioning (cd16396)	Hypothetical protein [*Mycobacterium abscessus*] (52)	1e−44
2	845…1720	501	Hypothetical protein	Hypothetical protein [*Selenomas ruminantium*] (52)	4e−51
terL	1723…3408	1685	Phage terminase, large subunit Terminase like family (pfam03237)	Putative large terminase subunit [*Skermania* phage SPI1] (61)	0.0
4	(3904…3401)	503	Hypothetical protein	Hypothetical protein [*Nocardia otitidiscaviarum*] (39)	3e−29
5	(4190…3960)	231	Hypothetical protein	–	–
6	(4587…4345)	243	Hypothetical protein	Hypothetical protein [*Nocardia grenadensis*] (57)	7e−11
7	(4771…4613)	159	Hypothetical protein	–	–
8	(5190…4771)	420	Hypothetical protein	Hypothetical protein [*Nocardia*] (34)	6e−09
9	(5714…5190)	525	Hypothetical protein	Hypothetical protein [*Nocardia brasiliensis*] (41)	9e−09
10	(6282…5848)	535	Hypothetical protein	Hypothetical protein [*Mycobacterium* phage Astraea] (38)	8e−06
11	(6752…6282)	471	Hypothetical protein	Hypothetical protein [*Rhodococcusfascians*] (49)	1e−22
12	(7336…6860)	477	Hypothetical protein	Hypothetical protein [*Streptomyces* sp.] (67)	2e−60
13	(7751…7338)	414	Hypothetical protein	Hypothetical protein [*Gordonia* phage Yvonnetastic] (48)	3e−34
14	8022…8219	198	Hypothetical protein	–	–
15	8347…8574	228	Hypothetical protein	–	
16	8737…8967	231	Hypothetical protein	–	
17	9014…9310	297	Hypothetical protein	Hypothetical protein [*Nocardia thailandica*] (87)	3e−49
18	9390…9710	321	Hypothetical protein		
19	9872…10186	315	Hypothetical protein	Hypothetical protein [*Rhodococcus* phage Finch] (54)	1e−13
20	10326…10508	183	Hypothetical protein		
21	(10784…10605)	180	Hypothetical protein	Hypothetical protein [*Skermania* phage SPI1] (62)	4e−07
22	(11173…10781)	393	Hypothetical protein	Hypothetical protein SPI1_5 [*Skermania* phage SPI1] (43)	7e−22
23	(11785…11390)	396	Hypothetical protein	–	–
24	(12814…11861)	954	Hypothetical protein	Hypothetical protein [*Mycobacterium abscessus*] (39)	3e−65
25	(13133…12924)	210	Hypothetical protein	Hypothetical protein [*Bacillus* phage Bastille] (31)	9e−05
26	(14178…13183)	996	Phage/plasmid like protein Domain of unknown function (DUF932)	DUF945 domain-containing protein [*Mycobacteriumperegrinum*] (52)	1e−97
27	14317…14445	129	Hypothetical protein		
28	14524…14790	267	Hypothetical protein		
29	14855…15067	213	HicA toxin superfamily (COG1724)	Hypothetical protein [*Mycobacterium* sp. E1747] (45)	0.001
30	15069…15416	348	Protein of unknown function (pfam07098)	Hypothetical protein SPI1_10 [*Skermania* phage SPI1] (55)	2e−27
31	15554…17374	1821	Hypothetical protein—putative portal	Hypothetical protein SPI1_11 [*Skermania* phage SPI1] (68)	0.0
32	17371…19584	2214	Hypothetical protein (pfam04233)—putative capsid morphogenesis	Hypothetical protein SPI1_12 [*Skermania* phage SPI1] (57)	0.0

Table 2 (continued)

Gene[a]	Coordinates	Length (bp)	Predicted function (Pfam)[b]	Significant match (% identity)[c]	E-value[d]
33	19620…19784	165	Hypothetical protein	Hypothetical protein SPI1_13 [*Skermania* phage SPI1] (58)	6e−04
34	19879…21915	2037	Putative major capsid	Hypothetical protein SPI1_14 [*Skermania* phage SPI1] (61)	0.0
35	21998…22789	792	Hypothetical protein	Hypothetical protein [*Nocardia testacea*] (58)	4e−84
36	22801…23064	264	Hypothetical protein	Hypothetical protein [*Mycobacterium* phage ShiVal] (53)	2e−16
37	23128…23400	273	Hypothetical protein	Hypothetical protein SPI1_18 [*Skermania* phage SPI1] (68)	5e−18
38	23402…23608	207	Hypothetical protein	Hypothetical protein SPI1_19 [*Skermania* phage SPI1] (69)	6e−08
39	23614…24012	399	Hypothetical protein	Hypothetical protein SPI1_20 [*Skermania* phage SPI1] (53)	1e−25
40	24132…25256	1124	Putative major tail	Hypothetical protein SPI1_21 [*Skermania* phage SPI1] (81)	8e−156
tRNA-Met	25347…25420		tRNA		
41	25824…26600	747	Hypothetical protein—head to tail connector	Hypothetical protein [*Rhodococcus ruber*] (57)	2e−90
42	26600…27133	534	Hypothetical protein—tail completion protein	Hypothetical protein [*Rhodococcus ruber*] (62)	3e−65
43	27149…27487	339	Hypothetical protein—tail assembly	Hypothetical protein [*Mycobacterium* phage Vincenzo] (57)	1e−13
44	27480…27899	420	Hypothetical protein—tail assembly	Hypothetical protein SPI1_29 [*Skermania* phage SPI1] (43)	4e−29
45	27935…28372	437	Hypothetical protein	Hypothetical protein SPI1_30 [*Skermania* phage SPI1] (44)	1e−25
46	28407…33992	5586	Tape measure protein Phage related protein (COG5412)	Hypothetical protein [*Nocardia miyunensis*] (51)	0.0
47	34002…35258	1257	Hypothetical protein—tail assembly	Hypothetical protein [*Rhodococcus ruber*] (64)	0.0
48	35255…35434	180	Hypothetical protein	–	–
49	35419…35805	387	Hypothetical protein—tail assembly	Hypothetical protein [*Rhodococcus ruber*] (66)	1e−52
50	35837…37528	1692	Hypothetical protein—tail assembly	Hypothetical protein [*Rhodococcus rhodochrous*] (54)	0.0
51	37525…38193	669	Hypothetical protein—minor tail	Hypothetical protein [*Rhodococcus ruber*] (44)	8e−50
52	38193…38813	621	Hypothetical protein	Hypothetical protein [*Rhodococcus ruber*] (55)	1e−79
53	38810…39400	591	Hypothetical protein	Hypothetical protein [*Rhodococcus ruber*] (57)	
54	39400…40002	603	Hypothetical protein	Hypothetical protein SPI1 38 [*Skermania* Phage SPI1] (45)	1e−45
55	40023…40562	540	Hypothetical protein	Hypothetical protein SPI1 46 [*Skermania* Phage SPI1] (34)	6e−34
56	40538…40753	216	Hypothetical protein	Hypothetical protein [*Mycobacterium* phage Apizium] (35)	3e−04
57	40773…41138	366	Hypothetical protein	Hypothetical protein [*Hoysella altamirensis*] (43)	9e−19
58	41129…41431	303	Hypothetical protein	Hypothetical protein [*Deinococcus pimensis*] (36)	4e−04
59	41442…42356	915	Hypothetical protein	Hypothetical protein PBI_ASTRAEA_111 [*Mycobacterium* phage Astraea] (41)	4e−64

Table 2 (continued)

Gene[a]	Coordinates	Length (bp)	Predicted function (Pfam)[b]	Significant match (% identity)[c]	E-value[d]
60	42357…42767	411	Hypothetical protein	Hypothetical protein [*Nocardia farcinica*] (50)	4e−26
61	42764…43474	711	Hypothetical protein	Hypothetical protein [*Nocardia farcinica*] (50)	2e−54
62	43501…43968	468	Hypothetical protein	Hypothetical protein [*Nocardia farcinica*] (50)	3e−54
63	(44153…43965)	189	Hypothetical protein	gp45 [*Mycobacterium* phage Pipefish] (54)	1e−15
64	(44874…44242)	633	Hypothetical protein	Hypothetical protein SPI1_25 [*Skermania* phage SPI1] (50)	8e−44
65	(45522…44947)	576	Hypothetical protein	Hypothetical protein [*Nocardia otitidis-caviarum*] (48)	2e−55
66	(46519…45772)	204	Hypothetical protein	–	–
67	45815…46660	846	Putative lysine (pfam13539)	Putative lysine [*Skermania* phage SPI1] (81)	3e−91
68	46657…46926	270	Hypothetical protein	Hypothetical protein [*Nocardia arizonsis*] (88)	7e−11
69	46930…47181	252	Hypothetical protein	Hypothetical protein [*Skermania* phage SPI1] (49)	4e−14
70	(47708…47178)	531	ruvC RuvC Holliday Junction resolvasome (COG0817)	Hypothetical protein SPI1_52 [*Skermania* phage SPI1] (46)	1e−46
71	(48520…48035)	486	Hypothetical protein	Hypothetical protein SPI1_53 [*Skermania* phage SPI1] (48)	1e−39
72	(49631…48525)	1107	Hypothetical protein	Hypothetical protein SPI1_54 [*Skermania* phage SPI1] (49)	2e−105
73	(49900…49631)	270	Hypothetical protein	Hypothetical protein SPI1_56 [*Skermania* phage SPI1] (42)	3e−06
74	(50262…49897)	366	Hypothetical protein (TIGR03789)	Hypothetical protein [*Nocardia jiangxien-sis*] (74)	9e−37
75	(50512…50285)	228	Hypothetical protein	–	–
76	(50509…50736)	228	Hypothetical protein	–	–
77	(52361…50733)	1629	Helicase Superfamily II DNA or RNA helicase (COG1061)	Putative helicase [*Skermania* phage SPI1] (65)	0.0
78	(52442…52645)	204	Hypothetical protein	–	–
79	(53517…52648)	870	Hypothetical protein	HTH DNA binding protein [*Mycobacte-rium* phage Bipper] (44)	1e−70
80	(56022…53599)	2424	AAA domain (pfam13481) Bifunctional DNA primase/polymerase (pfam09250) RecA family ATPase (COG3598)	gp52 [*Mycobacterium* phage Phaedrus] (48)	0.0
81	(56007…56333)	327	Hypothetical protein (cd00569)	Hypothetical protein SPI 60 [*Skermain* phage SPI1] (44)	2e−09
82	(58147…56327)	1821	DNA polymerase DNA polymerase family A (pfam00476)	DNA polymerase [*Nocardia* phage NBR1] (99)	0.0
83	(58159…58545)	387	Hypothetical protein (Pfam11753)	DUF3310 domain containing protein [*Nocardia ignorata*] (44)	2e−15
84	(58542…58733)	192	Hypothetical protein	–	–
85	(58730…58966)	237	Hypothetical protein	Hypothetical protein [*Nocardia* phage NBR1] (56)	2e−15
86	(58963…59319)	357	Hypothetical protein	Hypothetical protein 54 [*Brevibacterium* phage LuckyBarnes] (51)	8e−17
87	59858…60610	753	Hypothetical protein	–	–
88	60985…61464	480	Hypothetical protein	–	–
89	61464…61637	174	Hypothetical protein	–	–

Table 2 (continued)

Gene[a]	Coordinates	Length (bp)	Predicted function (Pfam)[b]	Significant match (% identity)[c]	E-value[d]
90	61624…61773	150	Hypothetical protein	–	–
91	62028…62195	168	Hypothetical protein	–	–
92	62188…62880	693	Hypothetical protein	Hypothetical protein [*Nocardia pseudobrasiliensis*] (70)	1e−43
93	62877…63215	339	Hypothetical protein	–	–
94	63215…63670	456	Hypothetical protein	–	–
95	63667…64293	627	Metallophosphatase domain (cd07390)	Hypothetical protein [*Nocardia flavorosea*] (58)	1e−74
96	64286…64447	162	Hypothetical protein	–	–
97	64748…65248	501	Hypothetical protein	Hypothetical protein [*Mycobacterium abscessus*] (34)	5e−15

[a]Genes numbered consecutively

[b]Predicted function is based on amino acid sequence identity, conserved motifs and gene locations within functional modules. Data output was generated via a suite of bioinformatics tools followed by confirmation and visualization through Geneious

[c]The most closely related gene (if named) and the name of the organism. The percentage identity (%) was based on the highest match when a BlastP analysis was performed

[d]The probability of obtaining a match by chance, as determined by BLAST analysis

acid similarity to those found at the same location in the *S. piniformis* phage SPI1 [28]. They include genes encoding the major tail protein (Orf40), which shares 81% similarity with the hypothetical protein 21 of the *Skermania* SPI1 phage, and the major capsid protein (Orf34), which shares 61% similarity with the hypothetical protein 14 of the *Skermania* SPI1 phage (Table 2). Small clusters of genes were seen throughout the NTR1 phage genome that are also similar to those described in genomes of *Nocardia* and *Rhodococcus* host species, which is not surprising as phage NTR1 can infect *Nocardia* species which are related to *Rhodococcus* species, sharing 24 hypothetical proteins (Table 2) [34]. Whether the NTR1 phage host range extends to *Rhodococcus* species is not known.

The NTR1 phage gene *orf29* is predicted to encode a protein of the HicA toxin superfamily, which is generally part of a two-component system, the *hicAB* cassette, a type II toxin-antitoxin (TA) system [35]. These type II TA systems are the best documented and most abundant systems among bacteria, and consist of a toxin and antitoxin [36]. Despite identifying a putative HicA gene, the gene encoding its component HicB could not be seen in the NTR1 phage genome. It is unclear which of the proteins is the toxin and which is the antitoxin, but in most bacteria, the antitoxin encoding gene (*hicA*) is located upstream of its corresponding toxin encoding gene [35]. However, recent studies have described HicA as the toxin and HicB as the antitoxin [37, 38]. The genes upstream and downstream of *orf29* (putative *hicA*) share no sequence similarity to previously described *hicB* genes. Interestingly no *hicA* or *hicB* genes have been reported previously in *Nocardia* species or its phages.

The location of the genes encoding the phage structural proteins in NTR1 phage was within the region *orf31-orf65*. The major capsid proteins most probably encoded by *orf34*, the major tail by *orf40* and the tape measure protein by *orf46*, have been identified in this phage, based on their sequence homologies, gene lengths and possession of the diagnostic conserved motif COG5412 associated with phage tape measure proteins. A cluster of genes encoding tail morphogenesis and assembly proteins (*orf41-orf44, orf47, orf49-orf51*) were also identified. This domain (COG5412) is of particular interest because of its association with mobilomes such as phages and transposons.

The remaining genes within this cluster encode proteins of unknown function. However, based on their location it is not unreasonable to assume they encode other components or proteins involved in the structural assembly of the phage. This pattern of gene organization is commonly seen in genomes of phages within the order *Caudovirales* [39]. The *Siphoviridae* phage genomes with their modular architecture generally have the gene encoding the main tail protein separate from the tape measure protein by two ORFs, where the second ORF is transcribed *via* a ribosomal slippage event [40]. Within the phage NTR1 genome, the genes encoding the main tail protein and the tape measure protein are separated by five ORFs and a tRNA and lack the ribosomal slippage sequence. It seems likely that the additional genes within this region encode proteins performing a similar function, given the widely different clustering patterns of genes that are responsible for a related function in other phage genomes [2, 7, 13, 14, 28].

The genes between *orf52* and *orf66* encode hypothetical proteins of unknown function, but share sequence

similarities with both bacterial and phage proteins. The DNA replication/maintenance module of NTR1 appears to be encoded by genes *orf67* to *orf83*. Orf70 is a RuvC Holliday junction resolvase, playing a key role in DNA recombination [41]. In Gram-negative bacteria, resolution of Holliday junction recombination intermediates is said to be achieved by the RuvC endonuclease working in conjunction with the RuvAB complex, thereby facilitating both branch migration and resolution [42, 43]. Interestingly, Gram-positive species such as *N. transvalensis* often lack RuvC. Instead, they utilize RecU, an unrelated junction resolution enzyme that can function as part of the RuvAB complex [44]. The *ruvC*-related genes have been detected in phages and prophages, although they appear more distantly related to the wild-type RuvC protein from *E. coli* (19–21% identity) [14, 27, 28, 45, 46]. In the NTR1 genome, the Holliday junction resolvase is substantially homologous to a hypothetical protein in the *Skermania* phage SPI1 genome (46% identity).

Analysis of the NTR1 phage genome suggests that complex recombination events have contributed to its evolution. Holliday junction resolvases act on branched DNA Holliday junctions to resolve their structures during recombination events. Thus, the presence of such a resolvase may be responsible for a series of recombination events within the genome of NTR1 [28]. Orf77 contains the superfamily II DNA helicase motif (COG1061) and shares 65% identity with a putative helicase in the *Skermania* phage SPI1 (Table 2).

The adjacent gene *orf79* encodes a hypothetical protein that shares 44% amino acid sequence identity with the helix turn helix (HTH) DNA binding protein in *Mycobacterium* phage Bipper [47], consistent with it having a regulatory role in NTR1 phage. The next gene *orf80* encodes three domains in different regions of the encoded protein. The first, motif pfam13481, suggests the expressed protein is a DNA repair protein [48], followed by an N-terminal bifunctional primase/polymerase motif (pfam09250) and a RecA-family ATPase, associated with DNA replication, recombination and repair [49]. This protein Orf80 is thought to have multiple functions based on possession of these conserved motifs. The RecA-family ATPase motif suggests its product behaves as a DNA-dependent ATPase, involved in promoting autodigestion of phage repressors, thereby causing a derepression of genes involved in the SOS response. This autodigestive process has an impact on the regulation of DNA repair and prophage induction in response to DNA damage [50, 51]. The Orf92 predicted amino acid sequence shares 70% identity with a hypothetical protein in *Nocardia pseudobrasiliensis* (Table 2), while Orf95 contains a metallophosphatase (MPP) conserved protein domain (cd07390) found in the bacterium *Nocardia flavorosea*, where it is thought to play a role in DNA repair, since both genes are in close proximity to a Holliday junction resolvase gene [52].

While a substantial portion of the NTR1 genome encodes proteins of unknown function, the information presented here provides a foundation for further studies into *Nocardia* phage genomics. The genetic data clearly indicate a close evolutionary relationship between the NTR1 genome and the *S. piniformis* phage SPI1 genome. Additional *Nocardia* phage sequence data may provide further insights into the evolutionary relationships between these phages. Finally, based on the data reported here, NTR1 would not be considered a suitable candidate for Nocardiosis phage therapy because it possesses a putative toxin gene. However, phage endolysin treatment rather than that using the entire virion could be used instead, as described previously [53].

In this study, we report the genomic characterization of a phage that can propagate in three different *Nocardia* species, all of which have clinical relevance. This phage, NTR1, belongs to the family *Siphoviridae* and appears to have a complex evolutionary history.

Acknowledgements We would like to acknowledge the LIMS Bio-Imaging Facility for use of the Electron Microscope and the La Trobe University Genomics Platform.

Author contributions JT, RJS and SP conceived the study as part of a larger project. SP commenced the project and ST and TLB performed the sequencing and annotation. PL generated the transmission electron microscope image. ST and SP performed the DNA sequence analysis and wrote the manuscript. All authors approved the manuscript.

Funding The Petrovski laboratory receives supported from the La Trobe University, Research Focus Area (RFA) securing food, water and the environment.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Research involving human and animal participants This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. CarlosC, GhislainF, SandraC-C, Marie-LiseD, HaraldB (2003) Phage as agents of lateral gene transfer. Curr Opin Microbiol6:417–424
2. PetrovskiS, SeviourRJ, TillettD (2013) Characterization and whole genome sequences of the *Rhodococcus* bacteriophages RGL3 and RER2. Arch Virol158:601–609
3. RohwerF (2003) Global phage diversity. Cell113:141
4. Brown-ElliottBA, BrownJM, ConvillePS, WallaceRJ (2006) Clinical and laboratory features of the *Nocardia* spp. based on current molecular taxonomy. Clin Microbiol Rev19:259–282
5. CasjensSR (2005) Comparative genomics and evolution of the tailed-bacteriophages. Curr Opin Microbiol8:451–458

6. HendrixRW, SmithMC, BurnsRN, FordME, HatfullGF (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world'sa phage. Proc Natl Acad Sci96:2192–2197
7. PetrovskiS, SeviourRJ, TillettD (2014) Genome sequence of the *Nocardia* bacteriophage NBR1. Arch Virol159:167–173
8. DesiereF, McShanWM, vanSinderenD, FerrettiJJ, BrüssowH (2001) Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic streptococci: evolutionary implications for prophage-host interactions. Virology288:325–341
9. HendrixRW (2003) Bacteriophage genomics. Curr Opin Microbiol6:506–511
10. PopeWH, MavrichTN, GarlenaRA, Guerrero-BustamanteCA, Jacobs-SeraD, MontgomeryMT, RussellDA, WarnerMH, HatfullGF (2017) Bacteriophages of *Gordonia*spp. display a spectrum of diversity genetic relationships. mBio8:e01069–e01017
11. Jurczak-KurekA, GasiorT, Najman-FalenczykB, BlochS, DydeckaA, TopkaG, NecelA, Jakubowska-DeredasM, NarajczykM, RichertM, MieszkowskaA, WrobelB, WegrzynG, WegrzynA (2016) Biodiversity of bacteriophages: morphological and biologicial properties of a large group of phages isolated from urban sewage. Sci Rep6:34338
12. De losReyesF (2010) Foaming. In: Microbial ecology of activated sludge. IWA, London, pp 215–258
13. PetrovskiS, DysonZA, QuillES, McIlroySJ, TillettD, SeviourRJ (2011) An examination of the mechanisms for stable foam formation in activated sludge systems. Water Res45:2146–2154
14. PetrovskiS, SeviourRJ, TillettD (2011) Characterization of the genome of the polyvalent lytic bacteriophage GTE2, which has potential for biocontrol of *Gordonia*-, *Rhodococcus*-, and *Nocardia*-stabilized foams in activated sludge plants. Appl Environ Microbiol77:3923–3929
15. BarkaEA, VatsaP, SanchezL, Gaveau-VaillantN, JacquardC, KlenkH-P, ClémentC, OuhdouchY, vanWezelGP (2016) Taxonomy, physiology, and natural products of actinobacteria. Microbiol Mol Biol Rev80:1–43
16. XiaoM, PangL, ChenSC, FanX, ZhangL, LiH-X, HouX, ChengJ-W, KongF, ZhaoY-P (2016) Accurate identification of common pathogenic *Nocardia* species: evaluation of a multilocus sequence analysis platform and matrix-assisted laser desorption ionization-time of flight mass spectrometry. PLoS ONE11:e0147487
17. EkramiA, KhosraviAD, ZadehARS, HashemzadehM (2014) *Nocardia* co-infection in patients with pulmonary tuberculosis. Jundishapur J Microbiol7:e12495
18. GlupczynskiY, BerhinC, JanssensM, WautersG (2006) Determination of antimicrobial susceptibility patterns of *Nocardia* spp. from clinical specimens by Etest. Clin Microbiol Infect12:905–912
19. KandiV (2015) Human *Nocardia* infections: a review of pulmonary nocardiosis. Cureus7:e304–e304
20. McNerneyR (1999) TB: the return of the phage. A review of fifty years of mycobacteriophage research. Int J Tuberc Lung Dis3:179–184
21. TorresO, DomingoP, PericasR, BoironP, MontielJ, VazquezG (2000) Infection caused by *Nocardia farcinica*: case report and review. Eur J Clin Microbiol Infect Dis19:205–212
22. ConvillePS, Brown-ElliottBA, SmithT, ZelaznyAM (2017) The complexities of *Nocardia* taxonomy and identification. J Clin Microbiol56:e01419–e01417
23. AnagnostouT, ArvanitisM, KourkoumpetisTK, DesalermosA, CarneiroHA, MylonakisE (2014) Nocardiosis of the central nervous system: experience from a general hospital and review of 84 cases from the literature. Medicine93:19–32
24. CândidaA, NunoR-P, AntónioS, FernandoM (2015) *Nocardia* infections among immunomodulated inflammatory bowel disease patients: a review. World J Gastroenterol21:6491–6498
25. ChenJ, ZhouH, XuP, ZhangP, MaS, ZhouJ (2014) Clinical and radiographic characteristics of pulmonary nocardiosis: clues to earlier diagnosis. PLoS ONE9:e90724
26. CastellanaG, GrimaldiA, CastellanaM, FarinaC, CastellanaG (2016) Pulmonary nocardiosis in chronic obstructive pulmonary disease: a new clinical challenge. Respir Med Case Rep18:14–21
27. PetrovskiS, SeviourRJ, TillettD (2011) Genome sequence and characterization of the *Tsukamurella* bacteriophage TPA2. Appl Environ Microbiol77:1389–1398
28. DysonZ, TucciJ, SeviourR, PetrovskiS (2016) Isolation and characterization of bacteriophage SPI1, which infects the activated-sludge-foaming bacterium *Skermania piniformis*. Arch Virol161:149–158
29. LopesA, TavaresP, PetitMA, GueroisR, Zinn-JustinS (2014) Automated classification of tailed bacteriophages according to their neck organization. BMC Genom15:1027
30. DelcherAL, BratkeKA, PowersEC, SalzbergSL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics23:673–679
31. RaoVB, FeissM (2008) The bacteriophage DNA packaging motor. Annu Rev Genet42:647–681
32. FujisawaH, MoritaM (1997) Phage DNA packaging. Genes Cells2:537–545
33. HatfullGF (2012) The secret lives of mycobacteriophage. Adv Virus Res82:179–288
34. GaoB, GuptaRS (2012) Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. Microbiol Mol Biol Rev76:66–112
35. MakarovaKS, GrishinNV, KooninEV (2006) The HicAB cassette, a putative novel, RNA-targeting toxin-antitoxin system in archaea and bacteria. Bioinformatics22:2581–2584
36. ParkSJ, SonWS, LeeB-J (2013) Structural overview of toxin–antitoxin systems in infectious bacteria: a target for developing antimicrobial agents. Biochim Biophys Acta (BBA)-Proteins Proteom1834:1155–1167
37. JørgensenMG, PandeyDP, JaskolskaM, GerdesK (2009) HicA of *Escherichia coli* defines a novel family of translation-independent mRNA interferases in bacteria and archaea. J Bacteriol191:1191–1199
38. TurnbullKJ, GerdesK (2017) HicA toxin of *Escherichia coli* derepresses hicAB transcription to selectively produce HicB antitoxin. Mol Microbiol104:781–792
39. VeeslerD, CambillauC (2011) A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiol Mol Biol Rev75:423–433
40. XuJ, HendrixRW, DudaRL (2004) Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol Cell16:11–21
41. AravindL, KooninEV (2000) SAP—a putative DNA-binding motif involved in chromosomal organization. Trends Biochem Sci25:112–114
42. SharplesGJ (2001) The X philes: structure-specific endonucleases that resolve Holliday junctions. Mol Microbiol39:823–834
43. vanGoolAJ, HajibagheriNM, StasiakA, WestSC (1999) Assembly of the *Escherichia coli* RuvABC resolvasome directs the orientation of Holliday junction resolution. Genes Dev13:1861–1870
44. McGregorN, AyoraS, SedelnikovaS, CarrascoB, AlonsoJC, ThawP, RaffertyJ (2005) The structure of *Bacillus subtilis* RecU Holliday junction resolvase and its role in substrate selection and sequence-specific cleavage. Structure13:1341–1351
45. BidnenkoE, EhrlichSD, ChopinMC (1998) *Lactococcus lactis* phage operon coding for an endonuclease homologous to RuvC. Mol Microbiol28:823–834
46. SaitoA, IwasakiH, AriyoshiM, MorikawaK, ShinagawaH (1995) Identification of four acidic amino acids that constitute the

catalytic center of the RuvC Holliday junction resolvase. Proc Natl Acad Sci92:7470–7474
47. HatfullGF (2016) Complete genome sequences of 61 mycobacteriophages. Genome Announc4:e00389–e00316
48. AmmelburgM, FrickeyT, LupasAN (2006) Classification of AAA + proteins. J Struct Biol156:2–11
49. SanchezH, KidaneD, ReedP, CurtisFA, CozarMC, GraumannPL, SharplesGJ, AlonsoJC (2005) The RuvAB branch migration translocase and RecU Holliday junction resolvase are required for double-stranded DNA break repair in *Bacillus subtilis*. Genetics171:873–883
50. LippsG, WeinzierlAO, vonSchevenG, CramerP (2004) Structure of a bifunctional DNA primase-polymerase. Nat Struct Mol Biol11:157–162
51. SelbitschkaW, ArnoldW, PrieferUB, RottschäferT, SchmidtM, SimonR, PühlerA (1991) Characterization of recA genes and recA mutants of *Rhizobiummeliloti* and *Rhizobiumleguminosarum* biovar *viciae*. Mol Gen Genet229:86–95
52. ShinDH, YokotaH, KimR, KimS-H (2002) Crystal structure of conserved hypothetical protein Aq1575 from *Aquifex aeolicus*. Proc Natl Acad Sci99:7980–7985
53. NobregaFL, CostaAR, KluskensLD, AzeredoJ (2015) Revisiting phage therapy: new applications for old resources. Trends Microbiol23:185–191

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.