Journal of Microbiology (2014) Vol. 52, No. 7, pp. 609–618
DOI 10.1007/s12275-014-3380-1

Genotyping, Morphology and Molecular Characteristics of a Lytic Phage of *Neisseria* Strain Obtained from Infected Human Dental Plaque

Ahmed N Aljarbou[1*] and Mohamad Aljofan[1,2]

[1]*Department of Pharmaceutics, College of Pharmacy, Qassim University, Saudi Arabia*
[2]*Department of Microbiology, School of Medicine and Health Sciences, Monash University, 3600, Australia*

(Received Jul 17, 2013 / Revised Mar 3, 2014 / Accepted Mar 12, 2014)

The lytic bacteriaphage (phage) A2 was isolated from human dental plaques along with its bacterial host. The virus was found to have an icosahedron-shaped head (60±3 nm), a sheathed and rigid long tail (~175 nm) and was categorized into the family *Siphoviridae* of the order *Caudovirales*, which are dsDNA viral family, characterised by their ability to infect bacteria and are nonenveloped with a noncontractile tail. The isolated phage contained a linear dsDNA genome having 31,703 base pairs of unique sequence, which were sorted into three contigs and 12 single sequences. A latent period of 25 minutes and burst size of 24±2 particles was determined for the virus. Bioinformatics approaches were used to identify ORFs in the genome. A phylogenetic analysis confirmed the species inter-relationship and its placement in the family.

***Keywords*:** *Neisseria* sp., *Siphoviridae,* lytic phage A2, human dental plaques, contigs, phylogenetic relationships

Introduction

Human oral cavity harbours various microbial communities with important ramification to human health and disease. A recent study of the bacterial biography of the human digestive tract found that the human mouth harbours the highest phylogenetic diversity (Ursell *et al*., 2012), including prokaryotic microbes such as bacteria and archaea and eukaryotes such as fungi and protista (Macarthur and Jacques, 2003). To date, nearly 700 different microbial species have been detected from the human mouth swabs (Paster *et al.,* 2001; Aas *et al*., 2005). Viruses and other bacteriophages have significant influence on the overall microbial abundance with effects on gene transfer processes in various microbial environments of the hosts. However, very little is known about their presence or impact on the oral bacterial community (Bachrach *et al*., 2003). There is a limited number of reports on the presence, genetic and human health impacts of oral bacterial community and there have been a few attempts made to isolate lytic phages from the human mouth (Hitch *et al*., 2004).

The human mouth's microbial community is responsible for periodontal diseases. The antibiotic treatments for mild and severe periodontal diseases may be inappropriate as it affects the whole microbial community and contributes to the development of drug-resistant bacteria. The use of bacteriophages may be an alternative method of controlling pathogenic bacteria since both are natural, non-toxic and specific in nature. Recent approval by the US Food and Drug Administration (FDA) for using *Listeria* phages as food additives (US Food and Drug Administration, 2006) has sparked new interests in employing phages as bactericidal agents in controlling health of human mouth.

The current study aims to identify and characterize the newly isolated bacteriophage infecting an oral *Neisseria* strain from human mouth (described in Aljarbou *et al*., 2012). The bacteriophage A2 was discovered to be a lytic virus, which is the first report of any lytic phage strain infecting *Neisseria* species. There have been reports of *Neisseria* isolations from various sources, e.g., *Neisseria meningitides* and *Neisseria gonorrhoeae* (Chanishvili *et al*., 2001), but no associated lytic phage virus has been detected for any *Neisseria* spp. The *Neisseria* is a genus of Gram-negative bacteria that contains some pathogenic species having major impact on human health including the gonococcus (*N. gonorrhoeae*) and the meningococcus pathogens (*N. meningitidis*). The genus also contains many non-pathogenic species including *Neisseria lactamica.* Therefore, the characterization of *Neisseria* isolated from human mouth area including the isolation of a bacteriophage will give an insight into the microbial environment.

Materials and Methods

Bacterial isolation and media and culture conditions

Toothpicks and dental floss (Johnson and Johnson, REACH) were used to collect materials attached to the dental plaque and between the teeth in the human mouth. These were first collected from three volunteers for bacterial host isolation. Each sample was mixed with 1 ml of PBS buffer to dissolve dental plaque materials. Ten microliter from each mixed sample was added to separate tubes of 10 ml of LB and BHIB. These tubes were incubated aerobically over night at 37°C with shaking at 150 rpm in an orbital shaker. Serial

*For correspondence. E-mail: aaljarbou@yahoo.com

dilutions from overnight growth were streaked out onto blood, LB and BHI agar plates and incubated at 37°C for 24 and 48 h. Colonies were selected according to their color, morphological features and size. Isolated colonies were picked and plated again to ensure purity of the isolated bacteria. Next, 20 µl from each mixed sample was streaked out directly onto blood and LB agar plates to grow bacteria that might not grow well in the broth media. These were incubated for 24 and 48 h. Colonies were then selected and purified as described above.

Detecting lytic phages

LB and BHIB agar were used as a bottom agar, while LB and BHIB containing 0.35% agarose were used as the soft top agar, which was prepared and kept molten in a water bath at 50°C. Soft tops were supplemented with horse blood to various percentages (2%, 3%, and 5% v/v) for bacteria that grew better in the presence of blood. For infection, 100 µl of filtered phage sample was added to 300 µl of the host cell culture that had been grown overnight. The virus particles were allowed to adsorb onto the host cells for 15 min at room temperature, then the infected cells were added to 3 ml of the molten soft top agar in universal tubes and mixed well before being poured onto the bottom agar. This was left to set for a few min; the plates were then inverted and incubated at 37°C. After 24 to 48 h, they were checked for the appearance of plaques.

Purification of phage DNA

The isolated OIB strain was grown overnight in LB, 10 µl of this culture was transferred to 10 ml of LB and incubated at 37°C in an orbital shaker at 150 rpm to an OD_{695} of approximately 0.2. Three hundred microliter of this culture was added to 100 µl of virus stock (~10^6 plaque-forming unit (pfu)/ml). The virus particles were allowed to adsorb onto the host cells for 15 min at room temperature, then the infected cells were added to 3 ml of the molten soft top agar in universal tubes and mixed well before being poured onto the bottom agar. This was left to set for a few min; the plates were then inverted and incubated at 37°C. Fifteen soft top agar plates that had completely lysed were scraped and collected into a 250 ml Sorvall tube. Forty milliliter of SM buffer was added to the collection tube and mixed well, then incubated overnight at 10°C to allow the virus particles to diffuse from the soft top agar into the buffer. The tube was centrifuged in a Beckman centrifuge at 250 g for 25 min, and then the supernatant was transferred to a fresh tube. The supernatant was filtered with a 0.45 µm and then a 0.22 µm Millipore filter to ensure the removal of agar and cell debris. The filtered A2 virus particles were mixed with 1/8 volume polyethylene glycol (PEG) 6000 solution [2.5 M NaCl, 20% (w/v) PEG 6000] and incubated on ice for 30 min. Sample was then centrifuged at 16,000 × *g* for 10 min and the virus pellet re-suspended in 0.5 ml of 10 mM Tris pH 7.5, 10 mM $MgCl_2$, 100 mM NaCl. Nucleic acids were digested by adding 10 U of DNase I and 10 µg/ml RNase A and incubated for 30 min at 37°C. Then it was extracted using an equal volume of phenol: chloroform as described (Sambrook *et al*., 2001).

DNA sequencing

Extracted viral genomes were cut using restriction endnuclease (New England Biolabs, USA). DNA fragments of 0.5 to 2 kb in size were ligated to the pGEM-T Easy vector, and then introduced into competent cells of *E. coli* JM109. DNA clones were sequenced by AGOWA (Germany). Primer walking and PCR sets accomplished gaps between the contigs.

Analysis of sequence

The viral genomic DNA sequences were assembled using the Lasergene SeqMan version 7.0 program (DNAStar, USA). Homology and Annotation of the Viral Genomes of the nucleic acid sequences were first analysed using the online Basic Local Alignment Search Tool (BLAST). BLASTN was first used to check if there was any known match to gene homology in the GenBank, via comparing each nucleotide query sequence against the nucleotide sequence database. Next, TBLASTX was used to uncover the sequence identity, and Open Reading Frame (ORF) Finders at http://www.ncbi.nlm.nih.gov/projects/gorf/ and GeneMark at http://opal.biology.gatech.edu/ GeneMark/index.html were used to predict and analyse the genes present in the sequences. The predicted ORFs were compared to the GenBank databases using the BLASTP program. Statistics for each of the ORFs were calculated using the ProtParam program (http://www.expasy.ch/tools/protparam.html). GC content in the virus contigs was calculated using the online base composition tools at http://atmolbiol-tools.ca/Jie_Zheng.

Pulse Field Gel Electrophoresis (PFGE)

A 2% low melting point agarose (Seaplaque® CTG agarose, USA) in 0.5 × TBE was prepared. Forty microliter of sample was added into an eppendorf tube, followed by the addition of 40 µl of agarose, which was mixed gently by pipetting to avoid the presence of bubbles. The mixed sample was transferred into a plug mould and left on the bench for 2 h. One milliliter of lysis buffer was added to the eppendorf tube, followed by the agarose plug, pushing the agarose out from the back. This was incubated overnight in a water bath at 55°C.

Gels were run for 18 h at 14°C using a CHEF DR-III system (Bio-Rad, USA) at 6.0 V/cm with initial and final pulse times of 5 sec and 13 sec, respectively. Following electrophoresis, the gel was stained for 30 min in 0.5 × TBE containing Gel Red (Biotium, USA). The gel image was captured using a G:BOX gel documentation system (Syngene, UK).

One step growth curve

The latent period and burst size were determined by the one step growth curve method, as described by Ellis and Delbrück (1939). The cells were infected with phage at 0.1 multiplicity of infection (MOI), to ensure that each cell was infected with only one phage. After infection the culture media was diluted at 100-fold to avoid multiple cycles of growth and lysis inhibition. Basically, a phage stock was added to an overnight cell culture containing 10 mM of $MgSO_4$ and incubated on ice for 10 min. The infected cells were centrifuged at 10,000 × g for 5 min at 4°C. The supernatant was

removed and infected cells were washed twice with PBS buffer, and then pelleted. They were then resuspended in fresh LB broth containing 10 mM $MgSO_4$ to aid adsorption and subsequently incubated in a water bath at 37°C. Samples were taken every 5 min up to 95 min, and were immediately titred by the plaque assay method. The one step growth experiment was repeated three times to observe if there was any variation in the results.

Electron microscopy

Preparations of virus particles were visualised by transmission electron microscopy at the Electron Microscopy facility of the University of Leicester. Approximately 3 µl of virus sample was placed on fresh glow-discharged Pioloform-coated grids and fixed in glutaraldehyde vapor for 2 min. Excess solution was removed from the grid using filter paper. The grid surface was washed with H_2O, then left to dry at room temperature for about 3–5 min and negatively stained with 1% (v/v) uranyl acetate and viewed on a JEOL 1220 microscope (Pagaling *et al.*, 2007).

Analysis of structural proteins

The virus particles were prepared from soft-top agar plates and concentrated by CsCl gradient method. The concentrated viral particles were mixed with 2× SDS loading buffer and heated for 5 min at 95°C and resolved by 12% of SDS-PAGE. Bands of interest were excised from the gel and an in-gel trypsin digest performed (Speicher *et al.*, 2000).

The bands were destained using 200 mM ammonium bicarbonate/20% acetonitrile, followed by reduction (10 mM dithiothreitol, Melford Laboratories Ltd., UK), alkylation (100 mM iodoacetamide, Sigma, UK) and enzymatic digestion (sequencing grade modified porcine trypsin, Promega, UK) using an automated digest robot (Multiprobe II Plus EX, Perkin Elmer, UK). LC-MS/MS was carried out upon each sample using a 4000 Q-Trap mass spectrometer (Applied Biosystems, UK).

Fragmentation spectra generated by LC-MS/MS were searched using the MASCOT search tool (Matrix Science Ltd., UK) against the nucleotide contig sequences supply using appropriate parameters. The criteria for protein identification were based on the manufacturer's definitions (Matrix Science Ltd). Candidate peptides with probability-based Mowse scores exceeding threshold ($P<0.05$), are considered significant or indicating an extensive homology.

Results

Detecting lytic phages in the human mouth

Over one hundred different bacterial colony types were isolated from human mouth. Different types of 0.35% w/v agar, soft-top, LB and, brain heart infusion were used. Soft-tops were supplemented with horse blood to various percentages (2%, 3%, and 5% v/v) for bacteria that grew better in the presence of blood. The dental plaque containing saliva from volunteers was mixed, filtered and used to infect overnight cultures of the isolated bacteria; these infected cultures were plated on soft tops, resulting in many instances in the appearance of plaques. A single plaque was then used to re-infect the same host culture; however, this procedure produced plaques only at the first and second times of propagation, after which plaques did not appear. The basis for loss of infectivity or development of host resistance upon serial passage is not known, but is a fairly common observation (Campbell *et al.*, 1985). One putative plaque-forming virus was an exception and continued to re-plaque on soft-top agar through multiple rounds of propagation, its host, the oral isolated bacterium (OIB), turned out to have plaque morphologies due to a lytic phage and was called A2 virus. This lytic phage was isolated from saliva following the infection experiment.

Identification and taxonomic classification of the OIB strain

To identify the OIB strain, 16S rRNA gene was amplified by PCR. Sequencing of the 1365 bp of the amplification product

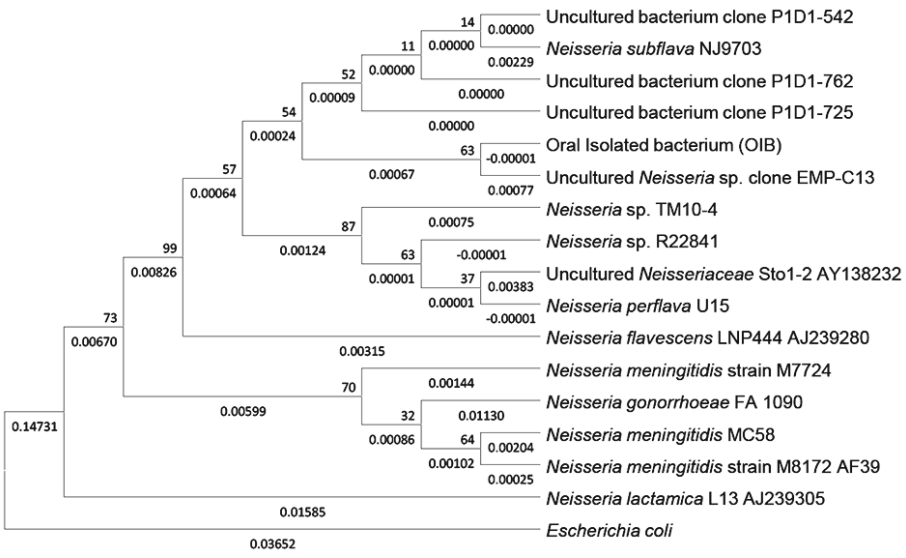

Fig. 1. A phylogenetic tree showing evolutionary relationships of the OIB strain to 17 taxa. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 2.5×10^{-1} is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown above the branches (Felsenstein *et al.*, 1985) (next to the branches). The evolutionary distances were computed using the Jukes-Cantor method (Jukes and Cantor, 1969) and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 1312 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 (Tamura *et al.*, 2007).

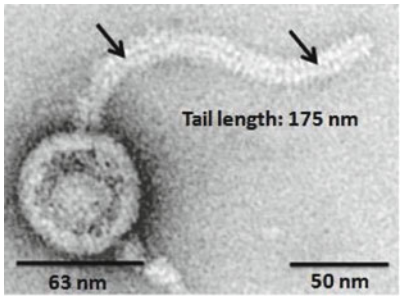

Fig. 2. Electron micrograph of A2 virus. The black arrows show the inner tube, running from the head to the end of the tail; the figure shows different size and length of capsid and tail that were detected in one plaque which was caused by the A2 virus.

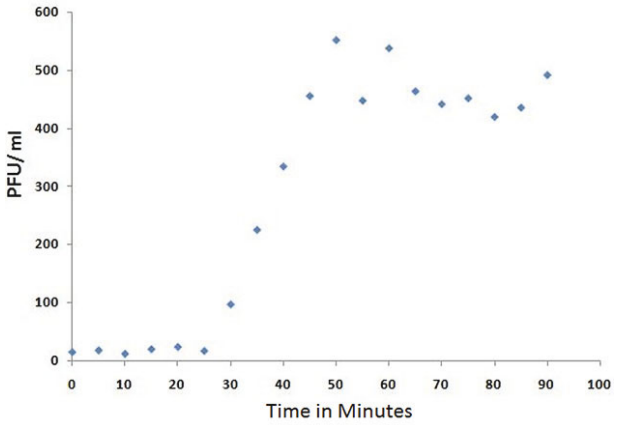

Fig. 3. Single step growth A2 with the OIB strain. The figure shows the latent period and the burst size, which were observed within 90 min. The latent period started at zero time and ended at 25 min of incubation.

and its subsequent analysis using BLASTN revealed hits to different types of *Neisseria* spp. The closest match with 99% identity was to an uncultured *Neisseria* sp. clone EMP_C13 (accession number EU794238), which originated from bovine fecal sample. The closest match to a cultured bacterium was 99% identical to *Neisseria subflava* NJ9703 (accession number AF479578). The *Neisseria perflava* and *N. subflava* were amplified in the human upper respiratory tract. To confirm the phylogenetic relationships between the OIB strain and diverse *Neisseria* strains, a phylogenetic tree was created (Fig. 1), based on the 16S rRNA gene sequence. *N. gonorrhoeae* FA 1090 was included in the analysis because the A2 genome had high sequence homology to the putative prophage in this strain (Piekarowicz *et al*., 2007).

Based on these phylogenetic analyses, the closest match of the OIB strain is *N. subflava* and *N. perflava.* Moreover, the bootstrap confidence on the *N. perflava* lineage is low at 57%. The 73% bootstrap value made it possible that the most closely related stain to the isolate is any of the *Neisseria* spp.; except *N. lactamica* L 13AJ239305, which appeared grouped alone in the phylogenetic analyses. Consistent to this relationship, the *Neisseria* species is known to frequently exchange chromosomal genes, which confounded their exact phylogenetic relationships in closely related species (Smith *et al*., 1999).

Transmission electron microscopy analysis of the viral particles

The A2 virus, which possess typical lytic behavior, as its titer was found to be increased by the plaque assay method and the lysis formed by the virus was very clear and not as turbid as usually obtained in this phage. The size of the plaques on the agar soft-top varied from 1 mm to 1.5 mm (data not shown). The plaques caused by A2 virus were visible within 3 h of incubation on the soft-tops at 37°C.

The A2 virus particles were isolated from plaques, stained with uranyl acetate and visualized using transmission electron microscope. According to ICTV (http://www.ncbi.nlm.nih.gov/ICTV), the A2 virus can be classified into order *Caudovirales*, family *Siphoviridae*, based on their morphology having an icosahedron-shaped head, sheathed-rigid tail with no base plates or tail fibres (Fig. 2). The head size is 60±3 nm in diameter. It has a thick non-contractile tail 175 nm long, covered with sheath striations. No collar or, connection neck is detectable. An inner tube was visible running from the head to the end of the tail (Fig. 2).

Single-step growth curve for the A2 virus

The latent period or eclipse and the burst size were determined for this virus using the single-step growth curve as described by Ellis and Delbrück (1939). A very low multiplicity of Infection (MOI) was applied to ensure host cells were infected with only one virus. Based on single-step growth, as shown in Fig. 3, the phage demonstrated typical lytic phage characteristics with the bacterial host, i.e. a latent period lasting 25 min, followed by an increase in the number of virus particles. The latent period was observed to be 5 min less than the *E. coli* phage (Ellis and Delbrück, 1939) and 7 min longer than phage T4 which grows on *E. coli*

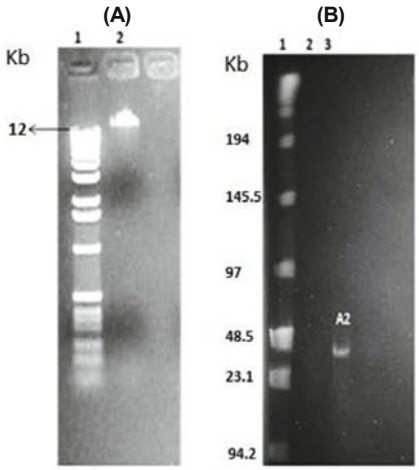

Fig. 4. The A2 virus genome. (A) Gel electrophoresis of the A2 genome: Lanes: 1, 1 kb marker; 2, The A2 viral genome was precipitated with PEG 6000 and extracted with equal volume of phenol: chloroform. 5 µl was run on 1% agarose gel electrophoresis. The virus band was set closely above the 12 kb marker; (B) Pulsed field gel electrophoresis of the A2 genome.

CTGCCTTTACTATCTATCAGTACTTGTTGAATATCAGTATTTACTAGTGTGGTCTCAG
CC**TGATTAGGGTTAGCC-TGATTAGGGTTAGCC-TGATTAGGGTTAGCC-TGA**
TTAGGGTTAGCCTGATTAGGGTTAGCC-TGATTAGGGTTAGCC-TGATTAGGGT
TAGCC-GATTAGGGTTAGCCTGATTAGGCTCATCAAAAACAATGTAATCTGTTGT
TCCGTTCCCATTCTTGCGGACTTGC

Fig. 5. The diagram represents the direct repeats of the Open Reading Frame 2 (ORF 2). The replication gene that was identified in ORF 2 is placed in contig A, and that it has eight sets of direct repeats (bold). The size of each of the direct repeats is 15 bp. The orders of assembled contigs were based on ORF and TBLASTX analysis.

(Hadas *et al*., 1997). The average burst size of the three experiments was calculated as 24±2 virus particles per bacterial cell.

Genome characterization, type, and size

It was confirmed that the A2 virus had a linear dsDNA genome following digestion with restriction endonucleases. The A2 viral genome appeared to sit immediately above the 12 kb marker (Fig. 4A). However, analysis obtained from PFGE showed that the size of the A2 viral genome, band 3, is just under the 48.5 and above 23.1 kb marker (Fig. 4B).

Cloning, sequencing and assembling the A2 viral nucleic acids

Fifty clones originated from *Sau*3A1, thirty-six from *Mfe*I and 10 from a *Hae*III digests were sequenced. Assembly of the sequences into contigs was done and ten small contigs were formed, with a few single sequences that did not overlap. The orders of assembled contigs were based on ORF and TBLASTX analysis. It was found that most of the contigs significantly matched prophage in both the *Neisseria gonorrhoeae* FA 1090 and *N. gonorrhoeae* NCCP11945 genomes (Table 1). Using PCR and primer walking, all the gaps were filled except for two, which resulted in the formation of three contigs: A, B and C (Fig. 5). The multiple ORFs in the contigs had the same direction and order as the genes on the

Table 1. A2 ORFs, gene products and, functional assignments

Contig	ORF	MW(kDa)	aa	Putative RBS	E value	Identities	Predicted function (domain)	Significant matches (accession number)
A	1	10.05	37		3e-11	88%	Unknown	Putative phage associated protein (PPAP)/ *N. gonorrhoeae* FA 1090 (YP_207633)
	2	40.8	218	AGGA (-1)	1e-15	35%	Phage replication protein	*Bacillus cereus G9842* (NP_833429)
	3	13.5	56				No match	
	4	53.7	169	AAG (-8)	8e-25	42%	Replicative DNA helicase (pfam00772)	*N. gonorrhoeae FA 1090* (YP_207635)
B	5	64	262		1e-142	93%	Terminase (pfam03237)	*N. gonorrhoeae* FA 1090 (YP_207645)
	6	174	715	AGG (-7)	0.0	89%	Unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207646)
	7	74.8	311	GGAG (-6)	4e-112	63%	Prohead protease (pfam04586)	*N. gonorrhoeae* FA 1090 (YP_207647)
	8	57	229	GGAG (-7)	7e-93	91%	Unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207648)
	9	55.8	227		7e-93	91%	Recombinational DNA repair protein (pfam03837)	*Actinobacillus pleuropneumoniae* (ZP_0013513)
	10	85.5	345	GGA (-6)	9e-119	90%	Unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207648)
	11	26.4	109	AGG (-10)	1e-28	67%	Unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207649)
	12	43.5	176		6e-64	67%	Unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207650)
	13	39.3	159		2e-32	63%	Unknown (COG3425)	PPAP / *N. gonorrhoeae* FA 1090 (YP_207651)
	14	34.6	142	AGG (-8)	4e-70	86%	Unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207652)
	15	63.4	257	AAGGA (-6)	1e-118	82%	Unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207653)
	16	26.8	109	AGGA (-7)	3e-46	87%	Unknown (PRK06975)	PPAP / *N. gonorrhoeae* FA 1090 (YP_207654)
	17	21.7	88		1e-41	89%	Unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207655)
	18	49.2	199	GGA (-9)	1e-90	79%	Unknown(CHP2217)	PPAP / *N. gonorrhoeae* FA 1090 (YP_207656)
	19	71.4	288	GAG (-9)	1e-119	75%	Unknown (BR0599)	PPAP / *N. gonorrhoeae* FA 1090 (YP_207657)
	20	36.5	147	GGT (-7)	3e-58	72%	Unknown (PRK10838)	PPAP / *N. gonorrhoeae* FA 1090 (YP_207658)
	21	13.5	548	GAGGT (-5)	0.0	79%	Unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207660)
	22	257.3	1059		3e-166	74%	Unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207660)
C	23	12.5	74		4e-04	39%	Unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207860)
	24	12.6	50				No match	
	25	28	114	AGGAGG (-4)	9e-28	53%	unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207669)
	26	13.7	57	AGGTGGTGG(-2)	8e-05	55%	unknown	PPAP / *N. gonorrhoeae* NCCP11945 (YP_002002040)
	27	38.8	157	AAG (-8)	4e-62	72%	Putative ATP binding protein (pfam08291)	*N. meningitidis* FAM18 (YP_974952)
	28	25.2	103	GGA (-0)	8e-23	47%	unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207664)
	29	40.5	161	GGAA (-1)	3e-42	53%	unknown	PPAP / *N. gonorrhoeae* FA 1090 (YP_207663)
	30	24.1	296	AAGG (-4)	3e-18	88%	Tail length tape measure protein	*N. gonorrhoeae* FA 1090 (YP_207672)
	31	214.6	884		0.0	78%	Tail length tape measure protein	*N. gonorrhoeae* FA 1090 (YP_207672)
	32	62.5	131		1e-25	56%	Putative Integrase (cd00397)	*N. gonorrhoeae* FA 1090 (YP_207674)

prophage in the *N. gonorrhoeae FA 1090* genome.

Sequence analysis of A2 virus

All three contigs had significant matches to other viral genes in the GenBank database. Table 1 shows the genes detected using GenMark (www.exon.gatech.edu) and ORFs, followed by the structure and function of the predicted genes. ORFs were often preceded by a sequence displaying similarity to the consensus ribosome-binding site (RBS), TAAGGAGGT (Shine and Dalgarno, 1974, 1975). The ORF and TBLASTX analyses showed the strongest matches to a putative phage in the *N. gonorrhoeae FA 1090* gene, while other matches were to a phage of *Bacillus cereus G9842,* a recombinant DNA repair protein belonging to *Actinobacillus pleuropneumoniae* and a putative ATP binding protein of *N. meningitidis* FAM18. Significant similarities were also found using BLASTN analysis of the three contigs of the A2 virus in different parts of the virus genome. As shown in Table 1; contig B, has four parts of the virus genome matched to *N. gonorrhoeae* FA 1090 (accession number CP001050). In contig C, only one part of the virus genome matched to the same bacterium and no matches to contig A. A total of 32 ORF genes were identified, out of which 17 were matched to phage proteins of unknown function and 13 to known genes or domains, while only two of the 32 had no significant matches to the GenBank database (Table 2 and Fig. 6). In total of 161 clones that were sequenced, 14 single sequences did not overlap with the three contigs that had been developed for the purpose while five single sequences were matched to four putative phage associated proteins and one to a putative replicative DNA helicase of *N. gonorrhoeae* FA 1090. The other single sequences were individually matched to a phage terminase small subunit of *E. coli* RS218, a single-stranded binding protein of *N. meningitidis* 053442, a putative integron gene cassette protein of an uncultured bacterium, an inner membrane insertion protein of *Serratia proteamaculans* 568, a DNA-damage-inducible protein of *Haemophilus influenzae* PittHH and a cloning vector. The remaining three sequences had no matches.

The estimated size of the A2 virus genome is 35 to 40 kb, based on the PFGE result. However, only 24.424 kb is accounted for by the total nucleotides obtained from the three contigs. If the 14 single sequences were to belong to the A2 virus and not to other prophages detected using electron microscopy, then only 3-9 kb might be missing from the total viral genome. As the 14 single sequences contain 7279 nucleotides plus the total number of nucleotides in the three contigs (24424), the total is 31703 nucleotides. Thus, the total number of all the sequenced nucleotides is close to the virus genomic size obtained by the PFGE. One conclusion which may be drawn is that one or both of the two gaps may have genes that cannot be cloned as a result of being lethal to *E. coli*; for example, the lysin gene (Bourguet *et al*., 2012). It is known that five successive genes forming a cluster are usually found in viruses of Gram-negative bacteria: these are endolysin, holin, antiholin and Rz/Rz1 equivalents (Wang *et al*., 2006). Another possible conclusion is that there are large gaps that cannot be filled by PCR.

The GC content of the A2 virus

The GC content can be used as an indicator of replication direction in many prokaryotes. It is known that the lowest GC region of the whole genome indicates the origins of replication and the highest GC region indicates the termination (Grigoriev *et al*., 1998, 1999). The contig A of the A2 virus had a lower GC content than contigs B and C. The ORFs of the A contig showed significant similarity to replication genes which may be the beginning of the A2 virus genome (Fig. 6A). The ORF analyses indicated that the structural order of the A2 virus gene is similar to that of a prophage in the

Table 2. **Protein analysis of A2 virion**

kDa	No. of different peptides observed	Score	Function	Nearest match to the GenBank
128.2	41	1397	DNA-directed RNA polymerase beta chain	*N. meningitidis FAM18*
124	9	398	Ribonuclease E	*N. meningitidis Z2491*
107.6	25	968	Pyruvate dehydrogenase E1 component	*N. meningitidis FAM18*
98.3	37	1527	Elongation factor EF-2	*N. meningitidis Z2491*
94.8	26	983	Putative isocitrate dehydrogenase	*N. meningitidis FAM18*
85.1	15	510	Putative chaperone protein	*N. meningitidis FAM18*
76.8	14	565	Aspartyl-tRNA synthetase	*N. meningitidis Z2491*
71	15	654	30S ribosomal protein S1	*N. meningitidis MC58*
65.3	19		Chaperonin, 60 kDa	*N. meningitides MC58*
60.4	6	339	Inositol-5-monophosphate dehydrogenase	*N. meningitidis Z2491*
57	7	326	Putative phosphate acetyltransferase	*N. meningitidis MC58*
52.6	3	143	Putative phage associated protein	*N. gonorrhoeae FA1090*
45.5	9	672	Translation elongation factor Tu	*N. meningitidis MC58*
43	10	575	Citrate synthase	*N. meningitidis FAM18*
39.4	11	263	Outer membrane protein	*N. meningitidis* MC58
35.5	8	471	Alcohol dehydrogenase	*N. meningitidis Z2491*

aa, Number of amino acids
MW, Molecular weight
RBS, Putative 5' upstream ribosome binding sequence (RBS) TAAGGAGGT.

N. gonorrhoeae A1090 genome. The ORF arrangement is divided into modules that are common to both bacterial and archaeal viruses. The genes relating to the establishment of cell infection are in the early region and those relating to DNA synthesis are in the middle region, while genes relating to virus assembly and cell lysis are located in the late region (Hendrix *et al.*, 1999; Brussow and Desiere, 2001). The GC content of Contig A was found to be 47.1%. The contigs B and C were found to have higher values, at 50.6 and 49.7%, respectively. The GC content of these contigs is lower than that of four prophages (54–57%), but similar to one prophage (49%) found in the *Neisseria gonorrhoeae* genome. These percentages are comparable to the average GC content (52.5%) of the *N. gonorrhoeae* genomic DNA (Piekarowicz *et al.*, 2007).

Bioinformatics analysis of gene products

All of the predicted ORFs were searched for similarity against the GenBank databases using BLASTP analysis. The ORFs 1, 6, 8, 10–23, 25, 26, 28 and 29 showed significant matches to hypothetical proteins and putative phage-associated proteins with unknown functions, all of which were matched to prophage in the *N. gonorrhoeae* FA 1090 genome (Table 1 and Fig. 6 IV). The ORFs 3 and 24 showed no match or similarity against the databases. ORF 2 in contig A was matched to a phage replication protein that is found in early gene that plays an important role in replication of viral genomes. Careful analysis showed that this gene has a 15-base pair region that is sequentially repeated eight times (Figs. 5 and Fig. 6IV). The presence of a series of repeats has been identified in many phage genes; for example, it was found that all lactococcal phages have a series of repeats in the replication genes. It has been proved that these series of repeats are the origins of phage replication (Ostergaard *et al.*, 2001).

Based on the length of the repeat, it could be either a cohesive end or a short exact direct repeat. In the case of cohesive ends, the two ends of the phage genomic DNA are single stranded identical in length and complementary to each other. After infection, the two ends anneal to each other to form a circular genomic DNA that serves as a tool for DNA replication mechanism. The length of the cohesive ends varies between viruses and commonly is between 7 to 19 nucleotides long, and can be at either 5′- or 3′- protruding strands (Hershey and Burgi, 1965; Ellis and Dean, 1985). Other viruses with linear dsDNA genomes require short exact direct repeats at the end of the genome to maintain genome integrity following DNA replication. Each round of replication would ordinarily result in loss of sequences at the 5′ ends. These repeats allow concatemers formation and resolution with endonucleases to maintain genome integrity.

The function of terminal short direct repeats in linear dsDNA genome is to maintain genome integrity. Either by intra-molecular base paring to form a closed circular DNA molecule, which is then replicated by a rolling circle mechanism, as explained by the '*cos*' sequences of lambdoid phages or to facilitate concatemerization of multiple copies of the linear genomes as exemplified by phage T4. Thus, the number of the 8 repeats found at the end of linear genome of the A2 virus could mark the end and the beginning of the genome. The 8 repeated sets were also searched against the GenBank databases using BLASTN, BLASTP, and BLASTX. Only a significant similarity was detected using BLASTP, which was over the total length of the repeated sets (39 aa)

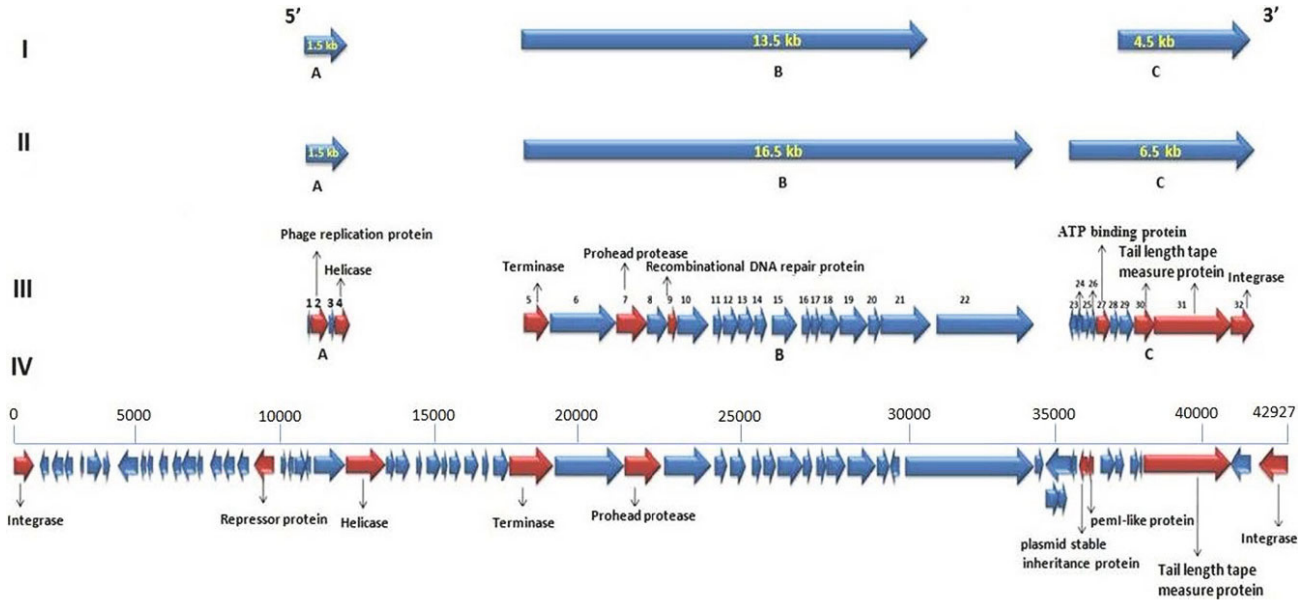

Fig. 6. Genome annotation of the A2 virus. Representation of the direction and order of the three contigs A, B and C based on the ORF matches to a prophage in the *N. gonorrhoeae FA 1090* genome. Part I shows the three contigs obtained using the restrictions endonucleases *Hae*III, *Sau*3AI and *Mfe*I. Part II shows the extension was mad in contigs B and C using *Dra*I and *Eag*I restriction endonucleases. Part III shows the genes order in the three contigs and then these were compared to the genes order of the prophage in the *N. gonorrhoeae* FA 1090 part IV. The known structural genes are represented by red arrows and the unknown putative phage-associated protein genes by blue ones.

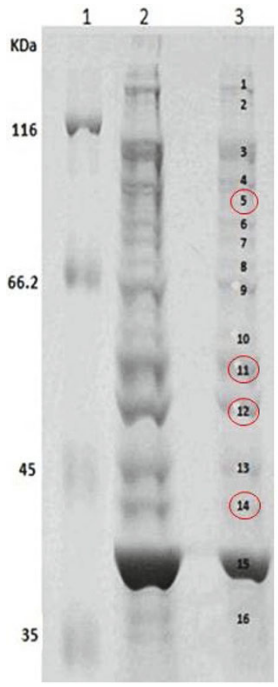

Fig. 7. SDS-PAGE analysis of A2 virion proteins. Lane 1 is the protein molecular weight marker unstained standard. Molecular mass in kDa of the marker is shown on the left. Lane 2 (band 1–16) contains less dilution of the structural protein samples than lane 3, of which the circled bands are directly related to the translated genomic sequence. Gel was stained with Brilliant Blue G-Colloidal and destained as indicated in the 'Materials and Methods'.

with E-value of 2e-07 and identity of 62% to a hypothetical protein of *Plasmodium falciparum* 3D7 (accession number XP_001350291).

The ORF 4 matched a putative replicative DNA helicase; most of the 12 known hexameric helicases have roles in DNA replication, recombination and transcription (Patel and Picha, 2000). The ORF 5 was matched to a terminase gene, but it was not clear whether it was a large or small subunit. Terminase genes are responsible for ATP-dependent packing of concatameric DNA in phage capsids. The small subunit possesses DNA recognition specificity, while the large subunit has catalytic activity (Fujisawa and Morita, 1997). The ORFs 7, 9, 13, 16, 18, 19, and 20 were matched to a variety of domains; see Table 1 for more details. The ORF 27 was matched to a putative ATP binding protein while ORFs 30 and 31 were matched to a tail-length tape-measure protein, which has an important role in the assembly and determination of length of the phage tail (Katsura, 1987; Pedersen *et al*., 2000). Finally, ORF 32 was matched to a putative integrase. It is an enzyme that may be divided into the serine and the tyrosine recombinase families, which is based on their mode of catalysis. Integrase play a variety of important roles such as mediating the recombination site between two DNA recognition sequences and the attachment site in the phage and the bacteria (Groth and Calos, 2004).

Structural proteins

The structural proteins were analysed by SDS-PAGE (Fig. 7) and the predicted protein mass (kDa), the number of observed peptides, the matched score, the nearest similarity and the origins of the protein and fragments, which are based on the GenBank database search are shown on Table 2. The viral capsid showed three major protein bands as well as many minor bands. Six bands (9, 10, 11, 12, 13 and 14) were cut and sequenced. Comparisons of the sequences against the GenBank databases and the contigs of the A2 viral genome showed clear matches with the bacterial host proteins from the NCBI nr database, while only three bands (11, 12, and 14) of the six sequenced bands, matched ORFs 6, 8 and 10 of the viral genome (Table 2). It was considered promising that three of the six sequenced protein bands matched the corresponding ORF genes, so the remainder of the ten protein bands are being sequenced. Interestingly, the annotation analyses of the ten sequenced protein bands matched significantly to bacterial proteins rather than the A2 virus capsid proteins (data not shown). No corresponding ORF genes were detected, except one peptide of band 5, which matched ORF 22.

The peptides of the four bands (5, 11, 12, and 14) that had corresponding ORFs are listed in Table 2. Bands 11 and 12 could be the major structural genes which are involved in the capsid and tail morphogenesis, based on the site order of the corresponding ORF genes, 8 and 10 (Fig. 6). The ORF 8 is set next to a prohead protease gene, which indicates that ORFs 8 and 10 are probably the major structural genes. These ORFs were matched to putative phage-associated proteins of unknown function in the *N. gonorrhoeae* A1090 genome. The band 11 had five peptides (K.SAQANGEPL NK.G, K.AITNINVGNQR.A, K.SGVTPTPTAVVSAGAGK.I, R.LPAYVQGVGNLLQVR.T, and K.GFTQPTSFTTGLQTY DLSAPSQK.L) that were matched to two corresponding ORFs. Four of the five peptides were matched to ORF 8 and one to ORF 10. It is also notable that the molecular mass of band 11 and ORF 8 were identical by being 57 kDa each. This supports the contention that band 11 corresponds to ORF 8 and not ORF 10. However, band 12 had ten peptides (K.TPLSQGFISR.V, K.LIIGNGGAPLIK.L, K.AITNINVG NQR.A, R.GGVINHEMVER.N, R.LSPDTIYVNAR.D, K.L NVDVNNTANIK.A, K.SGVTPTPTAVVSAGAGK.I, R.Q EYYQIEWPLR.T, R.LPAYVQGVGNLLQVR.T, K.GFTQP TSFTTGLQTYDLSAPSQK.L) matching three corresponding ORFs. The remaining ten peptides were matched to ORF 6, four to ORF 8 and five to ORF 10. It was found that the molecular mass of ORFs 6, 8 and 10 were 174.13 kDa, 57 kDa and 85.59 kDa, respectively. The only similarity was between ORF 8 and band 12, whose molecular weights were 52.6 kDa. It is not clear whether one of the protein bands was split into two. However, the similarity in molecular weight suggests that the split did not occur. It was also found that four peptides in bands 11 and 12 that matched to ORF 8 were the same peptides with translated nucleotides. Therefore, despite having other similarities, the sequence homology of the protein bands 11 and 12 were considerably different. A GenBank database search showed that band 11 had a significant match to putative phosphate acetyltransferase of *N. meningitidis MC58.* In contrast, a sequence of band 12 had a significant match to a putative phage-associated protein of *N. gonorrhoeae* FA 1090. In the case of bands 5 and 14, only one peptide, K.ALRNIIVQAR.Y and R.SLCELLL, from each match related to ORFs of 22 and 10, respectively. The bands 5 and 14 exhibited significant matches in the database to putative isocitrate dehydrogenase and citrate synthase of *N. meningitidis* FAM18. Therefore, these bands were

concluded to play an important role in the capsid structure of the A2 virus.

Discussion

Isolating lytic phages from the human mouth using the culture-based method is possible, but it requires much patience and careful examination of the plaques formed by lysis of the bacterial host. Knowledge of the bacterial effect on the soft-top agar is also required, as many bacteria can form plaques due to their antibacterial activity (Hitch *et al*., 2004). The isolated virus, A2, and its host (OIB) are unknown, and they have not been characterised previously. The virus was characterised by genotypic and phenotypic tests in order to identify its taxonomy. The clones were sequenced for the A2 virus, which fell into three contigs and 14 single sequences. Much effort was made to complete the genome sequence of the A2 virus; however, while a genotypic extension was obtained, the virus genome remained incomplete.

The A2 virus behaved like a lytic virus with the latent period of 25 min and the burst size of 24±2 particles determined by a single-step growth curve. It is also confirmed that the A2 virus is the first lytic virus that infects one of the *Neisseria* spp. This proves that detecting lytic phages from the human mouth for other pathogenic bacteria is possible and they could potentially be used as a second treatment option for bacteria that resist to a variety of antibiotics.

Intriguingly, there were significant matches to the host proteins, which were not unexpected, since several reasons could explain such matches and that they might not be considered to be due to contamination impending from the host proteins. It is likely that the A2 virus may have a lipid associated with its capsid. In the case of band 15 (Fig. 7), it seems that the major band protein, in accordance to the databases, a match was available to an outer membrane protein of *N. meningitidis* MC58; but no corresponding ORFs were detected. As indicated above, this protein may have significant roles in the A2 virus capsid, and similar protein may also be present in the cell membrane of the host.

The features of other bands, which significantly matched bacterial proteins rather than viral proteins (or even had no corresponding ORFs), could be the virus proteins which is supported by the observations that the bands 4, 5 and 11 are significantly matching host proteins in the GenBank databases, but they still have no corresponding genes. Thus, this could be extended to other bands that had significant matches to host proteins but not to the ORFs. Secondly, the virus genome has not been completely sequenced, thus some corresponding ORFs would be missing. Again, as indicated, similarities of the viral and the host proteins as well as contamination from the host proteins could be possible. Moreover, the GenBank databases may have much poorer coverage of viral than bacterial proteins and thus, the matches were significantly more likely to be made to bacterial genome rather than viruses. More importantly, the completion of the analysis of these proteins would be greatly improved if the viral genome was available intact to completely sequence, which is a priority that will be undertaken in the near future. In conclusion, the A2 virus is the first isolated bacteriophage that infects *Neisseria* spp., which belongs to the bacteriophage family *Siphoviridae*. The suitability of the identified bacteriophage as a potential phage therapy agent against *Neisseria* will also be assessed. This will potentially contribute to the treatment of antibiotic resistant bacterium.

Acknowledgements

M.A. is supported by National Health and Medical Research Council of Australia (Peter Doherty Biomedical Fellowship #APP1037092). The authors of the manuscripts would like to acknowledge Dr Shaun Heaphy at the Department of Infection, Immunity and Inflammation, Leicester Medical School, University of Leicester for allowing us to avail research facilities to conduct this study.

References

Aas, J.A., Paster, B.J., Stokes, L.N., Olsen, I., and Dewhirst, F.E. 2005. Defining the normal bacterial flora of the oral cavity. *J. Clin. Microbiol.* **43**, 5721–5732.

Aljarbou, A.N., de Luca, A., and Aljofan, M. 2012. Isolation of a new *Neisseria* phage from the oral cavity of healthy humans. *Antivirals Antiretrovirals* **1**, 416 doi:10.4172/416.

Bachrach, G., Leizerovici-Zigmond, M., Zlotkin, A., Naor, R., and Steinberg, D. 2003. Bacteriophage isolation from human saliva. *Lett. Appl. Microbiol.* **36**, 50–53.

Bourguet, F.A., Souza, B.E., Hinz, A.K., Coleman, M.A., and Jackson, P.J. 2012. Characterization of a novel lytic protein encoded by the *Bacillus cereus* E33L gene *ampD* as a *Bacillus anthracis* antimicrobial protein. *Appl. Environ. Microbiol.* **78**, 3025–3027.

Brussow, H. and Desiere, F. 2001. Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. *Mol. Microbiol.* **39**, 213–222.

Campbell, L.A., Short, H.B., Young, F.E., and Clark, V.L. 1985. Autoplaquing in *Neisseria gonorrhoeae*. *J. Bacteriol.* **164**, 461–465.

Chanishvili, N., Chanishvili, T., Tediashvili, M., and Barrow, P.A. 2001. Phages and their application against drug-resistant bacteria. *J. Chem. Technol. Biotechnol.* **02**, 68–2575.

Ellis, D.M. and Dean, D.H. 1985. Nucleotide sequence of the cohesive single-stranded ends of *Bacillus subtilis* temperate bacteriophage phi 105. *J. Virol.* **55**, 513–515.

Ellis, E.L. and Delbrück, M. 1939. The growth of bacteriophage. *J. Gen. Physiol.* **22**, 365–384.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783–791.

Fujisawa, H. and Morita, M. 1997. Phage DNA packaging. *Genes Cells* **2**, 537–545.

Grigoriev, A. 1998. Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res.* **26**, 2286–2290.

Grigoriev, A. 1999. Strand-specific compositional asymmetries in double-stranded DNA viruses. *Virus Res.* **60**, 1–19.

Groth, A.C. and Calos, M.P. 2004. Phage integrases: biology and applications. *J. Mol. Biol.* **335**, 667–678.

Hadas, H., Einav, M., Fishov, I., and Zaritsky, A. 1997. Bacteriophage T4 development depends on the physiology of its host *Escherichia coli*. *Microbiology* **143**, 179–185.

Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E., and Hatfull, G.F. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl. Acad. Sci. USA* **96**, 2192–2197.

Hershey, A.D. and Burgi, E. 1965. Complementary structure of interacting sites at the ends of lambda DNA molecules. *Proc. Natl. Acad. Sci. USA* **53**, 325–330.

Hitch, G., Pratten, J., and Taylor, P.W. 2004. Isolation of bacteriophages from the oral cavity. *Lett. Appl. Microbiol.* **39**, 215–219.

Jukes, T.H. and Cantor, C.R. 1969. Evolution of protein molecules. *In* Munro, H.N. (ed.), Mammalian Protein Metabolism, pp. 21–132, Academic Press, New York, N.Y., USA.

Katsura, I. 1987. Determination of bacteriophage lambda tail length by a protein ruler. *Nature* **327**, 73–75.

Macarthur, D.J. and Jacques, N.A. 2003. Proteome analysis of oral pathogens. *J. Dent. Res.* **82**, 870–876.

Ostergaard, S., Brondsted, L., and Vogensen, F.K. 2001. Identification of a replication protein and repeats essential for DNA replication of the temperate lactococcal bacteriophage TP901-1. *Appl. Environ. Microbiol.* **67**, 774–781.

Pagaling, E., Haigh, R.D., Grant, W.D., Cowan, D.A., Jones, B.E., Ma, Y., Ventosa, A., and Heaph, S. 2007. Sequence analysis of an Archaeal virus isolated from a hypersaline lake in Inner Mongolia, China. *BMC Genomics* **8**, 410.

Paster, B.J., Boches, S.K., Galvin, J.L., Ericson, R.E., Lau, C.N., Levanos, V.A., Sahasrabudhe, A., and Dewhirst, F.E. 2001. Bacterial diversity in human subgingival plaque. *J. Bacteriol.* **183**, 3770–3783.

Patel, S.S. and Picha, K.M. 2000. Structure and function of hexameric helicases. *Annu. Rev. Biochem.* **69**, 651–697.

Pedersen, M., Ostergaard, S., Bresciani, J., and Vogensen, F.K. 2000. Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901-1 involved in tail length determination and baseplate assembly. *Virology* **276**, 315–328.

Piekarowicz, A., Klyz, A., Majchrzak, M., Adamczyk-Poplawska, M., Maugel, T.K., and Stein, D.C. 2007. Characterization of the dsDNA prophage sequences in the genome of *Neisseria gonorrhoeae* and visualization of productive bacteriophage. *BMC Microbiol.* **7**, 66.

Saitou, N. and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 2001. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., USA.

Shine, J. and Dalgarno, L. 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346.

Shine, J. and Dalgarno, L. 1975. Terminal-sequence analysis of bacterial ribosomal RNA. *Eur. J. Biochem.* **57**, 221–230.

Smith, N.H., Holmes, E.C., Donovan, G.M., Carpenter, G.A., Spratt, B.G., and Spratt, B.G. 1999. Networks and groups within the genus *Neisseria*: analysis of *argF*, *recA*, *rho*, and 16S rRNA sequences from human *Neisseria* species. *Mol. Biol. Evol.* **16**, 773–783.

Speicher, K.D., Kolbas, O., Harper, S., and Speicher, D.W. 2000. Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. *J. Biomol. Tech.* **11**. 74-86.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599.

Ursell, L.K., Clemente, J.C., Rideout, J.R., Gevers, D., Caporaso, J.G., and Knight, R. 2012. The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites. *J. Allergy Clin. Immunol.* **129**, 1204–1208.

US Food and Drug Administration. 2006. Food additives permitted for direct addition to food for human consumption; bacteriophage preparation. *Fed. Regist.* **71**, 47729–47732.

Wang, J., Hu, B., Xu, M., Yan, Q., Liu, S., Zhu, X., Sun, Z., Reed, E., Ding, L., Gong, J., Li, Q.Q., and Hu, J. 2006. Use of bacteriophage in the treatment of experimental animal bacteremia from imipenem-resistant *Pseudomonas aeruginosa*. *Int. J. Mol. Med.* **17**, 309–317.