Isolation, characterization, and complete genome analysis of P1312, a thermostable bacteriophage that infects Thermobifida fusca

Jatuporn Cheepudom †, Cheng-Cheng Lee †, Bingfu Cai and Menghsiao Meng *

Graduate Institute of Biotechnology, National Chung Hsing University, Taichung, Taiwan

Thermobifida fusca is a moderately thermophilic and cellulolytic actinobacterium. It is of particular interest due to its ability to not only produce a variety of biotechnologically relevant enzymes but also serve as an alternative host for metabolic engineering for the production of valuable chemicals from lignocellulosic agricultural wastes. No bacteriophage that infects T. fusca has been reported, despite its potential impacts on the utilization of T. fusca. In this study, an extremely thermostable bacteriophage P1312 that infects T. fusca was isolated from manure compost. Electron microscopy showed that P1312 has an icosahedral head and a long flexible non-contractile tail, a characteristic of the family Siphoviridae. P1312 has a double-stranded DNA genome of 60,284 bp with 93 potential ORFs. Thirty-one ORFs encode proteins having putative biological functions. The genes involved in phage particle formation cluster together in a region of approximately 16 kb, followed by a segment containing genes presumably for DNA degradation/modification and cell wall disruption. The genes required for DNA replication and transcriptional control are dispersed within the rest of the genome. Phylogenetic analysis of large terminase subunit suggests that P1312 is a headful packaging phage containing a chromosome with circularly permuted direct terminal repeats.

Keywords: bacteriophage, Thermobifida fusca, lignocellulosic agricultural wastes, large terminase subunit, amidase, phage purification, endolysins, actinobacteriophage

Introduction

Thermobifida fusca, a moderately thermophilic soil actinobacterium, is known for its ability to produce a battery of cellulolytic enzymes (Maki et al., 2009; Adav et al., 2010; Gomez del Pulgar and Saadeddin, 2014). Sequence analyses of the genome of T. fusca YX strain suggests that it produces nine cellulases (including endocellulase, exocellulase, and cellobiosidase), at least five hemicellulose hydrolysis-related enzymes and many other glycoside hydrolases (Lykidis et al., 2007). Recent studies demonstrated that T. fusca may also secret lignin degradation-promoting enzymes. For example, a copper-containing polyphenol oxidase exhibits an activity for oxidation of phenolic lignin related compounds, and this activity boosts the digestion function of cellulase/xylanase toward sugarcane bagasse (Chen et al., 2013). Besides cellulolytic enzymes, T. fusca also produces a variety of oxidoreductases, such as heme-containing peroxidase (van Blooise et al., 2010) and catalase (Lončar and Fraaije, 2015), which are potentially useful in industries for detoxification and decolorization.
Cellulosic biomass is a low cost, abundant and renewable source for biofuels. To utilize it, great efforts have been devoted to engineer *Saccharomyces cerevisiae* so that the genetically modified yeasts can ferment xylose, cellulose, and cellobiose, and cellulosic oligosaccharides for bioethanol production (Katahira et al., 2006; Ha et al., 2011). Nonetheless, directly using the recalcitrant lignocellulose for biofuel production is still a challenging task. *T. fusca* has potential to be an alternative host for metabolic engineering to transform the sugars embedded in lignocellulose into biofuels and green chemicals. For example, an engineered *T. fusca* strain was able to convert untreated plant biomass to 1-propanol after an exogenous gene of *T. fusca* lignocellulose into biofuels and green chemicals. For example, metabolic engineering to transform the sugars embedded in *T. fusca* can be fully modified for the purpose of producing valuable commodities from various cellulosic agricultural wastes.

Bacteriophages have contributed to the development of a variety of molecular tools for biotechnology. For example, the origin of replication from f1 and λ phages was used in the construction of plasmids for single-stranded DNA production (Reece, 2004) and routine recombinant DNA operation, respectively (Boyd and Sherratt, 1995). In addition, recombinases from bacteriophages such as phage λ, P1, and PY54 have been widely used to modify prokaryotic species and to create transgenic animals and plants (Sauer and Henderson, 1988; Nafissi and Slavcev, 2014). On the other hand, bacteriophages have potentials to foul industries that use bacteria to produce fermented products or bioactive molecules. No detailed reports of bacteriophages that infect *T. fusca* are present in the literature. Thus, we set out to isolate *T. fusca*-infecting phages with long-term goals to understand the bacterium-phage interactions and look for useful genetic elements for the development of molecular tools specific for *T. fusca*. A thermostable tailed bacteriophage of the *Siphoviridae* family was then isolated in this study.

Tailed bacteriophages constitute the *Caudovirales* order. Phages with contractile tails are further subdivided into the *Myoviridae*, while those with short and long non-contractile tails are into *Podoviridae* and *Siphoviridae*, respectively. Tails are critical to the infection because they contain proteins required for the specific recognition of the hosts and trigger DNA release from the heads. The genomes of tailed bacteriophages are composed of several functional modules. For *Siphoviridae*, the modular arrangement as follows: packaging, head morphogenesis, tail morphogenesis, lysis, recombination, lytic/lysogenic control, excision, and DNA replication modules is seen repeatedly in numerous temperate phages of Gram-positive low GC-content bacteria (Stevens et al., 2011), whereas lambdaid phages that infect Gram-negative bacteria have a genome consisting of modules in order of packaging, head morphogenesis, tail morphogenesis, recombination, lytic/lysogenic control, DNA replication, and lysis (Campbell, 1994). Characteristics and genome analysis of this newly discovered *T. fusca*-infecting phage are addressed herein.

**Materials and Methods**

**Bacterial Strain**

*T. fusca* NTU22 strain (Chen et al., 2013) was routinely cultivated by transferring 10⁸ spores into 50 ml CYC medium (30 g sucrose, 6 g Casamino acids, 3 g NaNO₃, 2 g yeast extract, 1 g K₂HPO₄, 0.5 g KCl, 0.5 g MgSO₄·7H₂O, 0.01 g FeSO₄·7H₂O, 1 l distilled water, pH 8.0) and incubated aerobically at 50°C with 200 rpm shaking.

**Phage Isolation**

Compost collected from eight different sites in the suburbs of Taichung city, Taiwan, was the source of phage screening. One gram of the compost was added into 10 luria-Bertani medium (LB) and incubated aerobically at 50°C overnight. One ml aliquot of the sample was irradiated (120 mJ/cm², 30 s) using a CL-1000 UV crosslinker (UVP, California, USA) before being mixed with 0.1 ml *T. fusca* spore (10⁶ cfu/ml) and 15 ml CYC media with 0.7% agar. The warm mixture was immediately poured onto a solidified CYC agar plate and incubated at 50°C for 2 days. One of the plaques formed in the bacterial lawn was transferred into 50 ml 1-day-old culture of *T. fusca* and further incubated at 50°C for 1 day. The bacterial lysate was centrifuged, and the supernatant was sterilized using a 0.45-µm filter and stored at 4°C. The isolated bacteriophage capable of infecting *T. fusca* was named P1312 hereafter.

**Phage Titer**

Phage titer was estimated by plaque assay method. Briefly, 1 ml phage sample at the 10⁻⁷–10⁹ dilution was mixed with 10² spore of *T. fusca* in 15 ml pre-warmed (50°C) CYC soft agar medium, and the mixture was poured evenly onto a CYC solid agar medium. The solidified plates were then incubated at 50°C for 2 days, and the number of plaques formed in the lawn of *T. fusca* was counted to determine the phage titer.

**Phage Stability, Adsorption, and Burst Size**

To determine the structural stability of P1312, the purified phage sample was diluted in water and incubated at temperatures ranging from 60–95°C for the indicated periods of time (10–45 min), and the residual infectivity was determined by the plaque assay. The burst size of P1312 was determined by one-step growth method (Lin et al., 2012) with slight modifications. Briefly, P1312 was added into an overnight culture of *T. fusca* in 20 mM phosphate buffer (pH 8.0) to give a multiplicity of infection (MOI) of 1.0 according to the initial spore number of *T. fusca*. The mixture was incubated at room temperature for 10 min to allow the phage particles to attach to the host. The supernatant, after centrifugation at 12,000 rpm for 5 min, was collected to determine the number of unattached phage particles. The percentage of adsorption was calculated using the formula [(initial titer–titer in supernatant)/initial titer] x 100%. The pelleted cells were washed two more times with CYC medium and resuspended in 50 ml of pre-warmed CYC and incubated at 50°C. Aliquots of 1 ml culture were taken at intervals and the phage titers in the clarified supernatant were then determined as aforementioned. The burst size (Bₚ) of P1312 was calculated as...
B_p = P_t/P_o where P_t is the phage titer at the plateau phase and P_o is the initial infective titer, which was estimated on the basis of the plaque-forming units arising from the initially washed cells.

**Phage Purification**

Hundred milliliter broth of *T. fusca*, which had been cultivated for 1 day, was inoculated with 1 ml of P1312 stock (10^9 pfu/ml) and the cultivation was continued for another 2 days at 50°C, 200 rpm. The clarified broth, after centrifugation at 12,000 rpm, 4°C, for 10 min, was adjusted to contain 1.5 M NaCl and 10% (w/v) polyethylene glycol (PEG) 8000. The mixture was placed on ice for at least 1 h and subjected to centrifugation at 12,000 rpm, 4°C, for 15 min to precipitate the phage particles. The pellet was resuspended in TBS buffer (50 mM Tris-HCl [pH 8.0], 150 mM NaCl) that additionally contained 0.25 U/ml DNase I and 10 µg/ml RNase A (Takara Bio). After incubation at 30°C for 1 h, the supernatant was passed through a filter with 0.45 µm pore size and loaded onto a Tricorn 10/600 column packed with Sephacryl S-500HR (Amersham Biosciences). The chromatography was performed with TBS buffer at a flow rate of 0.7 ml/min.

**Transmission Electron Microscopy (TEM)**

The purified phage particles on a carbon-coated copper grid were negatively stained with 2% uranyl acetate (pH 4.0) for 10 min and observed at an accelerating voltage of 120 kV under a transmission electron microscope (Jeol LEM-1400).

**Genome Analysis of Phage P1312**

Genomic DNA was obtained from the purified phage P1312 after phenol/chloroform extraction and ethanol precipitation. Whole genome sequencing of P1312 was performed using Illumina Miseq (Tri-1 Biotech, Inc). The count of reads was 115,622 with the average length of 175 bases per read. The sequence data could be assembled into a circular single contig of 60,284 base pairs using the *de novo* assembly algorithm of CLC Genomics Workbench (Qiagen). The nucleotide sequence of the genome of P1312 is deposited at GenBank under accession number KT021004. Identification of potential open reading frames (ORFs) within the phage genome was performed by using Bacterial Annotation System (van Domselaar et al., 2003) as well as Glimmer/RBSfinder program (Delcher et al., 1999), and the functions of proteins encode by the ORFs were predicted based on BLASTp program and conserved domain search (http://www.ncbi.nlm.nih.gov/). The probable replication origin of P1312 was predicted by the GenSkew program (http://genskew.csb.univie.ac.at/).

**Phylogenetic Analysis**

Protein sequences of terminase large subunit from a variety of bacteriophages were retrieved from the biological database, National Center for Biotechnology Information (NCBI). Multiple sequence alignments of those terminase sequences were performed by using the ClustalW program with default parameters in MEGA 6.0 version (Tamura et al., 2013). Phylogenetic tree was built by the neighbor-joining method and phylogenies were determined by bootstrap analysis of 10,000 replicates in MEGA 6.0 version.

**Identification of Proteins Associated with P1312 Virions**

Proteins of P1312, purified via PEG precipitation and gel filtration chromatography, were identified by tandem mass spectrometry using an Applied Biosystems QStar LC-MS/MS spectrometer (Life Technologies Corp., Carlsbad, USA). The obtained spectrometry information was analyzed with Mascot software (Matrix Science Ltd., London, UK) using the NCBI non-redundant database and the specific database created in this study based on the predicted ORFs of phage P1312 (Table 1). The important parameter settings for Mascot analysis were as follows: mass values, monoisotopic; protein mass, unrestricted; peptide mass tolerance, ±0.5 Dalton; fragment mass tolerance, ±0.5 Dalton; and maximal missed cleavages, 2.

**Results and Discussion**

**Isolation and Purification of Phage P1312**

Compost collected from several suburban farms was tested for the presence of phages that infect *T. fusca* according to the method described in Materials and Methods. A chicken manure sample was found to contain phages that could grow in *T. fusca* and resulted in lytic plaques in the lawn of the bacterium (Figure 1). After plaque purification and phage propagation, the virions were purified by PEG precipitation and gel filtration using a Sephacryl S-500HR column as described in Materials and Methods. The phage particles could be obtained from the fractions corresponding to the void volume of the chromatography (Figure 2). According to the TEM photos and the genome organization (described below), we believe that the purified sample contained only one type of phage. If there were more than two bacteriophages, more genes for phage specific proteins such as tape measurement protein and large subunit of terminase would be expected.

**Infectivity of P1312**

Since *T. fusca* is a moderately thermophile, it was interesting to know the thermal stability of P1312. The phage particles were incubated at 60, 70, 80, 90, and 95°C. Aliquots were withdrawn periodically during the incubation and the residual infectivity was determined by the plaque assay. The infectivity of P1312 remained intact after the incubations at 90°C for 45 min, indicating that P1312 is extremely thermostable (Figure 3). Nonetheless, 5 min incubation at 95°C was able to inactivate P1312 effectively. For one-step growth experiments, P1312 was added into the culture of *T. fusca* at an MOI of 1. After 10 min incubation, which resulted in 69% adsorption of the phage particles to the host, the infected host cells were transferred into a fresh CYC medium and the phage particles released into the medium during the following cultivation were determined by the plaque assay. There was a latent period of ~45 min before the rapid increase of the phage titer, and lysis was complete by ~120 min (Figure 4). The growth kinetics of P1312 suggests that the average burst size is 57 pfu per infected cell.
**TABLE 1 | Predicted open reading frames (ORFs) of P1312 and predicted database matches.**

| ORF | Start | End  | Protein kDa (Start) | Predicted functions | BLASTP (best match) | Ident (%) | LC-M/M^2 |
|-----|-------|------|---------------------|---------------------|---------------------|-----------|-----------|
| 001 | 243   | 980  | 27.8 (V)            | Phage head morphogenesis protein | WP_017602198/ Nocardiap lucentensis | 133/249   | 38        |
| 002 | 987   | 1475 | 18.3 (V)            | Hypothetical protein           | WP_030728966/ Streptomyces | 80/131    |           |
| 003 | 1462  | 2805 | 50.6 (M)            | Phage terminase large subunit  | WP_027740749/ Streptomyces | 289/435   | 38        |
| 004 | 5101  | 2795 | 83.4 (V)            | Phage portal protein           | WP_040692131/ Nocardiap lucentensis | 299/405   | 14        |
| 005 | 4301  | 5080 | 29.0 (L)            | Hypothetical protein           | WP_040692134/ Nocardiap lucentensis | 45/61     |           |
| 006 | 5774  | 5085 | 25.2 (V)            | Capsid                          | WP_040692203/ Nocardiap lucentensis | 225/283   | 18        |
| 007 | 5121  | 6005 | 31.3 (V)            | gp27                             | WP_040692137/ Nocardiap lucentensis | 34/68     |           |
| 008 | 6036  | 6461 | 15.5 (V)            | Phage protein                    | WP_040692204/ Nocardiap lucentensis | 84/138    | 63        |
| 009 | 6644  | 6363 | 10.1 (V)            | gp27                             | WP_040692134/ Nocardiap lucentensis | 29/91     |           |
| 010 | 6954  | 6376 | 21.7 (V)            | No match                         | WP_017602206/ Nocardiap lucentensis | 80/120    | 66        |
| 011 | 6487  | 6963 | 18.2 (M)            | Hypothetical protein            | WP_040692137/ Nocardiap lucentensis | 38/94     |           |
| 012 | 6978  | 7340 | 13.3 (M)            | Phage head-tail adaptor          | WP_040692140/ Nocardiap lucentensis | 49/70     |           |
| 013 | 7340  | 7732 | 14.8 (V)            | gp27                             | WP_030714734/ phiSASD1 phage | 389/41    |           |
| 014 | 8508  | 7699 | 29.7 (V)            | No match                         | WP_017602209/ Nocardiap lucentensis | 45/62     |           |
| 015 | 8512  | 8703 | 7.1 (V)             | Phage tail protein               | WP_017602211/ Nocardiap lucentensis | 19/53     |           |
| 016 | 8703  | 9098 | 14.3 (M)            | Hypothetical protein            | WP_040692145/ Nocardiap lucentensis | 67/116    |           |
| 017 | 9119  | 9406 | 11.2 (V)            | Phage tail protein               | WP_017602211/ Nocardiap lucentensis | 48/99     |           |

(Continued)
| ORF | Start | End   | Protein kDa (Start)* | Predicted functions                          | BLASTP³ (best match)       | Ident (%)² | LC-M/M¹ |
|-----|-------|-------|----------------------|---------------------------------------------|----------------------------|------------|---------|
| 019 | 9411  | 9587  | 5.9 (V)              | gp35.1                                      | YP_003714742 phoSASD1 phage | 37/96      |         |
| 020 | 9418  | 14391 | 177.1 (V)            | Tail tape measure protein                   | WP_040703000 Nocardiosiopsis salina | 156/362    | 27      |
| 021 | 14388 | 16361 | 71.7 (V)             | Phage tail protein                         | WP_017602213 Nocardiosiopsis lucentensis | 358/672    |         |
| 022 | 16358 | 19648 | 120.6 (M)            | Phosphodiesterase                          | WP_026128263 Nocardiosiopsis lucentensis | 520/933    |         |
| 023 | 19670 | 20212 | 19.3 (M)             | Hypothetical protein                       | WP_040692409 Nocardiosiopsis lucentensis | 71/180     | 23      |
| 024 | 20237 | 20497 | 9.6 (M)              | No match                                   | WP_017946204 Streptomyces sp. CNS615 | 53/68      |         |
| 025 | 20494 | 21270 | 28.4 (M)             | O-methyltransferase                        | WP_042170120 Streptomyces sp. NIBRC 110036 | 41/124     |         |
| 026 | 21359 | 21267 | 3.2 (M)              | No match                                   | WP_043471037 Kitasatospora sp. MBT66 | 47/109     |         |
| 027 | 21428 | 22423 | 37.2 (V)             | Glycosyl/glycerophosphate transferases      | WP_045995834 Streptomyces sp. NRRL S-104 | 35/120     |         |
| 028 | 22494 | 23063 | 20.1 (L)             | Hypothetical protein                       | WP_017602347 Nocardiosiopsis lucentensis | 147/213    |         |
| 029 | 23346 | 23068 | 10.5 (M)             | Hypothetical protein                       | WP_042170120 Streptomyces sp. NIBRC 110036 | 41/124     |         |
| 030 | 23167 | 23517 | 11.9 (V)             | No match                                   | WP_017946204 Streptomyces sp. CNS615 | 53/68      |         |
| 031 | 23530 | 23408 | 4.9 (M)              | No match                                   | WP_043471037 Kitasatospora sp. MBT66 | 47/109     |         |
| 032 | 23850 | 23530 | 12.0 (M)             | No match                                   | WP_043471037 Kitasatospora sp. MBT66 | 47/109     |         |
| 033 | 23822 | 24076 | 8.5 (V)              | No match                                   | WP_043471037 Kitasatospora sp. MBT66 | 47/109     |         |
| 034 | 24019 | 23870 | 5.9 (M)              | No match                                   | WP_043471037 Kitasatospora sp. MBT66 | 47/109     |         |
| 035 | 24099 | 24554 | 13.5 (L)             | No match                                   | WP_043471037 Kitasatospora sp. MBT66 | 47/109     |         |
| 036 | 24608 | 24126 | 17.6 (L)             | Hypothetical protein                       | WP_017602347 Nocardiosiopsis lucentensis | 58/114     |         |
| 037 | 24451 | 24597 | 5.4 (V)              | No match                                   | WP_043471037 Kitasatospora sp. MBT66 | 47/109     |         |
| 038 | 24655 | 24882 | 8.6 (M)              | Transcriptional regulator                  | WP_017602347 Nocardiosiopsis lucentensis | 58/114     |         |
| 039 | 25042 | 25404 | 13.5 (V)             | Protein involved in exopolysaccharide biosynthesis | WP_043471037 Kitasatospora sp. MBT66 | 47/109     |         |
| 040 | 25530 | 25973 | 16.3 (L)             | Sensory protein kinase CreG                | WP_017602347 Nocardiosiopsis lucentensis | 173/266    |         |
| 041 | 25970 | 26845 | 30.8 (M)             | N-acetylmuromyl-L-alanine amidase/Lysin     | WP_017602347 Nocardiosiopsis lucentensis | 173/266    |         |
| 042 | 27063 | 27344 | 9.9 (L)              | Hypothetical protein (Similar to Tfu_2915) | WP_017602347 Nocardiosiopsis lucentensis | 45/73      |         |

(Continued)
| ORF | Start | End  | Protein kDa (Start)a | Predicted functions                            | BLASTPb(best match) | Ident (%)c | LC-M/Md |
|-----|-------|------|-----------------------|-----------------------------------------------|----------------------|------------|---------|
| 043 | 27313 | 28419| 38.9 (L)              |                                               |                      |            |         |
| 044 | 28350 | 27490| 32.0 (V)              | Transcriptional regulator                     | WP_018223844         | 54/131     |         |
|     |       |      |                       | (Helix-turn-helix family protein)              | Salinispora pacifica | (41)       |         |
| 045 | 28819 | 28406| 14.5 (M)              |                                               |                      |            |         |
| 046 | 29336 | 28905| 16.5 (M)              |                                               |                      |            |         |
| 047 | 30751 | 29333| 50.3 (V)              | Hypothetical protein                          | ADU28745             | 88/289     |         |
|     |       |      |                       |                                               | Nitrosococcus watsonii C-113 | (30)       |         |
| 048 | 29466 | 29365| 3.7 (L)               |                                               |                      |            |         |
| 049 | 29438 | 30544| 42.5 (M)              | Transposase, IS605 orfB                        | WP_016188297         | 384/368    | 29      |
|     |       |      |                       |                                               | Thermobifida fusca   | (99)       |         |
| 050 | 30724 | 31998| 44.1 (L)              |                                               |                      |            |         |
| 051 | 31767 | 30778| 34.9 (M)              | Hypothetical protein                          | WP_037974524         | 40/107     | 14      |
|     |       |      |                       |                                               | Synergistes jonesii  | (37)       |         |
| 052 | 33166 | 31952| 45.3 (M)              | Transcriptional regulator                     | WP_017615773         | 223/399    |         |
|     |       |      |                       | (Helix-turn-helix XRE-family like proteins)    | Nocardiopsis salina  | (56)       |         |
| 053 | 34868 | 33456| 50.7 (V)              | Replicative DNA helicase                      | WP_035111511         | 174/451    | 31      |
|     |       |      |                       |                                               | Corynebacterium freiburgense | (39)       |         |
| 054 | 34798 | 35892| 37.3 (V)              |                                               |                      |            |         |
| 055 | 36066 | 34912| 42.3 (V)              | Hypothetical protein                          | WP_040793245         | 49/130     |         |
|     |       |      |                       |                                               | Nocardia paucivorans | (38)       |         |
| 056 | 37130 | 36525| 22.0 (M)              | Hypothetical protein                          | WP_040721854         | 37/87      |         |
|     |       |      |                       |                                               | Streptomonospora alba| (43)       |         |
| 057 | 36735 | 37016| 9.8 (V)               |                                               |                      |            |         |
| 058 | 37102 | 38487| 45.7 (V)              |                                               |                      |            |         |
| 059 | 37779 | 37183| 22.1 (M)              |                                               |                      |            |         |
| 060 | 38460 | 37840| 23.7 (M)              | Hypothetical protein                          | WP_017541564         | 59/168     |         |
|     |       |      |                       |                                               | Nocardiopsis halophila| (35)       |         |
| 061 | 38523 | 40232| 64.3 (L)              |                                               |                      |            |         |
| 062 | 40349 | 38715| 55.7 (V)              | Hypothetical protein                          | WP_013475335         | 89/160     |         |
|     |       |      |                       |                                               | Micromonospora sp. L5| (56)       |         |
| 063 | 42526 | 40484| 72.8 (V)              | DNA segregation ATPase FtsK'SpoII             | WP_017602285         | 253/481    |         |
|     |       |      |                       |                                               | Nocardiopsis lucentensis| (53)       |         |
| 064 | 44174 | 42486| 58.0 (V)              | Hypothetical protein                          | WP_017602419         | 96/293     |         |
|     |       |      |                       |                                               | Nocardiopsis lucentensis| (33)       |         |
| 065 | 43517 | 43747| 8.7 (M)               |                                               |                      |            |         |
| 066 | 44513 | 44346| 6.2 (V)               |                                               |                      |            |         |
| 067 | 45035 | 44577| 16.5 (V)              | ABC-ATPase                                     | WP_037918697         | 51/119     |         |
|     |       |      |                       |                                               | Streptomyces yeoshonensis | (43)       |         |
| 068 | 45020 | 46795| 67.3 (L)              | Intergrase/Recombinase                         | YP_005087274         | 155/501    |         |
|     |       |      |                       |                                               | Rhodococcus phage REQ1 | (31)       |         |
| 069 | 47265 | 46846| 15.3 (M)              |                                               |                      |            |         |
| 070 | 47870 | 47286| 22.4 (V)              | Hypothetical protein                          | WP_036322164         | 44/113     | 51      |
|     |       |      |                       |                                               | Microbispora sp. ATCC PTA-5024 | (39)       |         |
| 071 | 47971 | 48198| 8.6 (M)               | Transcriptional regulator                     | WP_0120322242        | 33/63      |         |
|     |       |      |                       | (Helix-turn-helix XRE-family like proteins)    | Pelotomaculum thermopropionicum | (52)       |         |

(Continued)
General Features of P1312

The morphology of P1312 was visualized using a transmission electron microscope (Figure 5). P1312 has a head in hexagonal outline, plausibly icosahedral, and a long flexible non-contractile tail, suggesting that P1312 belongs to the family Siphoviridae. The head is approximately 56 nm in diameter, and the tail is approximately 250 nm long and 11 nm wide. The complete nucleotide sequence of the genome of P1312 was determined with Illumina Miseq and assembled into a circular genome of 60,284 base pairs using the assembly algorithm of CLC Genomics Workbench. The GC content of the phage genome is 65.9%, close to 67.5% of the host’s chromosome (Lykidis et al., 2007). The possible ORFs in the genome were predicted by Bacterial Annotation System and Glimmer/RBSfinder program. Integrating the computing results indicated the presence of 93 ORFs, arranged on both strands. Among them, 36 ORFs start with methionine, while 44 and 13 ORFs start with valine and leucine, respectively. The biggest ORF encodes a protein containing 1657 amino acid residues; by contrast, the smallest one encodes a peptide of 30 amino acids. The general information of the proteins encoded by the ORFs regarding their lengths, putative functions, and homologs found by BLASTp are listed in Table 1. The phage head morphogenesis protein-encoding ORF was arbitrarily assigned as the first ORF, followed by terminase genes, so that the presented gene order is similar to that of many bacteriophages such as Salmonella bacteriophage ES18.

### Table 1

| ORF | Start | End | Protein kDa (Start)a | Predicted functions | BLASTPb(best match) | Ident (%)c | LC-M/Md |
|-----|-------|-----|----------------------|---------------------|---------------------|-----------|--------|
| 072 | 48520 | 48948 | 16.5 (M) | No match | | | |
| 073 | 49402 | 48866 | 18.2 (L) | No match | | | |
| 074 | 49094 | 49405 | 11.2 (V) | No match | | | |
| 075 | 49402 | 50229 | 30.7 (M) | No match | | | |
| 076 | 49935 | 49747 | 6.6 (M) | No match | | | |
| 077 | 50446 | 51360 | 33.9 (V) | Exodeoxyribonuclease VIII | WP_017972488 | 156/282 | |
| 078 | 51357 | 52118 | 28.2 (V) | Hypothetical protein | WP_045740910 | 117/172 | |
| 079 | 52187 | 52909 | 27.5 (V) | DNA polymerase III subunit epsilon (DnaQ-like exoribonuclease) | WP_026118477 | 121/234 | |
| 080 | 52936 | 53097 | 5.9 (V) | No match | | | |
| 081 | 53097 | 53723 | 22.6 (M) | No match | | | |
| 082 | 53720 | 54208 | 17.9 (V) | Holliday junction resolvase | WP_033299115 | 102/157 | |
| 083 | 54205 | 54645 | 16.3 (V) | WhiB family transcriptional regulator | WP_011595908 | 50/107 | |
| 084 | 54648 | 55439 | 30.3 (M) | Hypothetical protein | WP_017602245 | | |
| 085 | 55657 | 55343 | 11.0 (M) | No match | | | |
| 086 | 55382 | 56044 | 24.7 (M) | Hypothetical protein | WP_043984960 | 80/215 | |
| 087 | 56029 | 57438 | 51.9 (V) | Hypothetical protein | WP_017602245 | 135/271 | |
| 088 | 56919 | 56311 | 21.1 (L) | No match | | | |
| 089 | 57435 | 58103 | 24.4 (M) | Hypothetical protein | WP_030282501 | 30/83 | |
| 090 | 58214 | 58504 | 14.5 (V) | No match | | | |
| 091 | 58573 | 58908 | 12.3 (M) | Hypothetical protein | WP_037865813 | 61/118 | |
| 092 | 58872 | 59621 | 9.3 (M) | No match | | | |
| 093 | 59887 | 59571 | 10.5 (V) | No match | | | |

---

a The first translated amino acid is shown in parentheses.
b The matched homologs in Streptomyces phage phiSASD1 are also included although they are not the best hits.
c The percentage identity is calculated based on the number of identical amino acid residues (numerator) over the number of compared residues (denominator) and shown in parentheses.
d The sequence coverage (%) determined by mass spectrometry of the protein is indicated.
Cheepudom et al.

Thermobifida fusca-infecting phage

FIGURE 1 | Plaques of P1312 in the lawn of T. fusca NTU22.

FIGURE 2 | Purification of P1312 by gel filtration chromatography. The PEG-precipitated P1312 was loaded into a 50 ml Sephacryl S-500HR column. The proteins were eluted with TBS buffer and monitored with a UV detector. AU is referred to absorbance units.

(Casjens et al., 2005) and Enterococcus faecalis bacteriophage φEf11 (Stevens et al., 2011). Thirty-one ORFs encode proteins that contain acknowledged functional domains or motifs, while 23 ORFs encode hypothetical proteins that have homologs found in other phages, prophages or bacteria, despite the lack of putative conserved domains. The rest of the predicted ORFs encode proteins sharing no significant identity with proteins in the databases; therefore, some of them may represent false positives.

The purified phage particles were subjected to protein identification by LC-MS/MS. This method identified 24 out of the 93 predicted proteins, including many phage structural proteins and enzymes such as N-acetylmuramoyl-L-alanine amidase and replicative DNA helicase (Table 1). Nine identified proteins (encoded by ORF6, 10, 14, 50, 75, 76, 85, 88, and 90) actually do not have significant homologs in the databases; therefore, they may represent novel proteins. It is noteworthy that either DNase I or RNase A, which were used in the pretreatment prior to the size exclusion chromatography, was not detected. The absence of the nucleases in the phage preparation suggests that the chromatographic method could effectively separate the soluble proteins from the virions.

Provisional Functions of P1312 Proteins

Genome organization of P1312 is illustrated in Figure 6. ORFs are displayed on both DNA strands with many of them overlapping. The cluster encompassing ORF1–21 is responsible for phage particle formation. ORF1 produces the head morphogenesis protein. The product of ORF3 contains an ATP-binding cassette transporter nucleotide-binding domain and shares similarity to the large subunit of phage terminases. ORF4 is supposed to produce the phage portal protein, the junction between the phage head and tail proteins, through which DNA passes during packaging and injection. ORF7 encodes
a phage capsid protein. The product of ORF8 belongs to a family composed of proteins from a variety of bacteriophages. The region from ORF12 to ORF21 probably is involved in tail formation. In details, ORF12 directs the synthesis of the phage head–tail adaptor, while ORF13, 15, 16, 17, 18, and 21 likely encode the phage tail components. ORF20, the largest gene in the genome, encodes the tail tape measure protein. For lambdoid phages, the size of the tape measurement protein corresponds to the tail length by a fairly constant of 0.15 nm per amino acid residues (Katsura, 1990). The average tail length of P1312 is approximately 250 nm, accordant to the size of the phage tail tape measure protein. Many of the proteins encoded by the assumed head and tail gene modules share similarities to the proteins encoded by Streptomyces phage phiSASD1 (Wang et al., 2010) (Table 1). The comparison of the modules between P1312 and phiSASD1 is shown in Figure 7. Despite the resemblance, a couple of differences in the gene order are noticed. First, the directions of portal protein gene are opposite. Second, the N terminus of gp38 of phiSASD1 shares a small similar region with a tail protein of P1312 (product of ORF21), while the C terminus has a region similar to the CBM_4_9 domain present in the phosphodiesterase of P1312 (product of ORF22). Apart from the head and tail gene modules, the other modules share less similarity between P1312 and phiSASD1.

The segment from ORF22 to ORF49 is notable for producing enzymes for a variety of functions such as DNA degradation/modification and cell wall disruption. ORF22 encodes a protein of 1096 amino acid residues. This protein is predicted to contain an N-terminal signal peptide, a metallophosphatase (MPP) domain and a domain related to carbohydrate-binding module (CBM) family 4_9 at the C-terminus. Proteins in the MPP superfamily are functionally diverse; the members include Mre11/SbcD-like exonuclease, Dbr1-like RNA lariat debranching enzymes, YkE-like phosphodiesterases, purple acid phosphatases (PAPs), YbbF-like UDP-2,3-diacylglycerol hydrolases and acid sphingomyelinases according to the conserved domain database, NCBI. By the N-terminal signal peptide and the C-terminal CBM, the product of ORF22 may be secreted through the cell membrane and anchors within the cell wall architecture. The MPP domain is thought to have a DNA degradation activity that may prevent the superinfection of the host by other bacteriophages. The cell wall of T. fusca mainly contains peptidoglycan and polyglycerolphosphate-lipoteichoic acid (Rahman et al., 2009); the latter is covalently linked to the outer leaflet of the cytoplasmic membrane. Therefore, the alternative function of the MPP domain may be involved in the hydrolysis of the cell wall by hydrolyzing the phosphodiester bonds in the teichoic acid polymers. ORF25 directs the synthesis of a putative class I adenylmethionine-dependent methyltransferase. Presumably, this enzyme has an activity for nucleic acid modification. ORF27 encodes a protein similar in amino acid sequence to glycosyl/glycerophosphate transferase, a protein responsible for the polymerization of the main chain of the cell wall-associated teichoic acid. The product of ORF27, as being a phage protein, is thus assumed to interfere with, rather than promote, the host's function of teichoic acid synthesis.

The protein encoded by ORF41 is a putative N-acetylmuramoyl-L-alanine amidase, presumably able to disintegrate the peptidoglycan of the host cell by cleaving the amide bond between N-acetylmuramoyl and L-alanine. It is noteworthy that this phage protein does not contain a secretory signal sequence, consistent with a long-term observation that the endolysin produced by double stranded DNA bacteriophages requires a small membrane protein, known as a holin, to permeabilize the membrane for its access to the peptidoglycan (Young et al., 2000). Interestingly, we found that this putative amidase was associated with the phage particles according to the results of tandem mass spectrometry. This association may reflect an artifact caused by an unspecific interaction of the protein to the virion. Alternatively, it may implicate an involvement of the amidase in the infection step by assisting the phage to inject its DNA into the host cell. Actually, tail-associated peptidoglycan-degrading enzymes involved in localized cell wall degradation have been evidenced in a number of bacteriophages such as Tuc2009 (Kenny et al., 2004) and φ29 (Xiang et al., 2008). ORF402 encodes a 93-amino acid polypeptide with two transmembrane helices predicted by the TMPred program (Hofmann and Stoffel, 1993). It is tempting to assume that this polypeptide serves as the holin for the phage amidase to pass through the membrane. ORF49 encodes a transposase belonging to IS605 family, which is also present in the genome of T. fusca and many other bacteria. Presumably, phage P1312 acquired this gene from its bacterial host.

Genes responsible for DNA replication and transcriptional regulation are situated sporadically in the remaining half of the genome. Presumably, ORF53 directs the synthesis of a DNA-like replicative helicase that unwinds the DNA duplex at the replication fork. ORF63 encodes a protein whose C-terminus shares similarity to DNA translocase ftsK of Actinoplanes, suggesting the involvement of the protein in DNA segregation. The protein encoded by ORF68 contains a putative serine recombinase domain that is usually found associated with pflm00239 in putative integrases/recombinases of mobile genetic elements of diverse bacteria and phages. Therefore, this protein may catalyze the integration and excision of the phage genome and from the host chromosome, respectively. It is noteworthy that P1312 behaved as a lytic bacteriophage in this study; therefore, the culture condition may disfavor the integration
function of the protein probably through suppressing the protein expression. The protein encoded by ORF77 is related to the PDDEXK superfamily, and appears to be an exonuclease VIII. The product of ORF79 is a DnaQ-like exonuclease, belonging to DEDDh 3'-5' exonuclease family. Probably, it acts as a proofreading subunit (epsilon) of polymerase III for the DNA replication of P1312. ORF 82 encodes a protein of RusA superfamily that can resolve Holliday junction intermediates by its endonuclease activity. BLASTp analysis also suggests the presence of five transcriptional regulators, encoded by ORF38, 44, 52, 71, and 83. P1312 may use them to control the gene expressions, in terms of transcriptional timing and level, for its successful proliferation. The probable replication origin of P1312 was predicted by the GenSkew program, an application for computing and plotting nucleotide skew data. The resulting GC-skew plot (not shown) suggests that the region around the nucleotide 38440 (close to ORF60) could be the initiation site for replication.
FIGURE 7 | Comparison of the head and tail gene modules between P-1312 and phiSASD1. The percentage identity between the compared proteins (see details in Table 1) is shown within the gray shadow.
FIGURE 8 | Neighbor-joining phylogenetic tree of large terminase subunit amino acid sequences. The sequences, except that from P1312, were classified as described previously (Casjens et al., 2005), and bootstrap analysis was performed with 1000 repetitions. The node of phylogenetic tree shows the bootstrap confidence values above 50%.
DNA Packaging Strategy

All known tailed-bacteriophage virions contain a single linear dsDNA chromosome, because the passage of the portal protein is not wide enough to allow two parallel dsDNAs to be threaded simultaneously into the head during packaging and out of the virion during injection. As a member of Siphoviridae, P1312 probably contains a linear genome, rather than a circular one as determined by whole genome sequencing using Illumina Miseq. Six types of termini of the chromosomes in tailed-bacteriophage virions have been studied (Casjens and Gilcrease, 2009); they are (1) single-stranded cohesive ends, (2) circularly permuted direct terminal repeats, (3) short, non-permuted, direct terminal repeats, (4) long, non-permuted, direct terminal repeats, (5) terminal host DNA sequences, and (6) covalently bound terminal proteins. These different types of ends reflect varied DNA replication strategies and depend on terminase actions during DNA packaging. A previous bioinformatic analysis indicated that the different functional classes of phage-encoded terminases can usually be predicted from the amino acid sequence of large terminase subunits (Casjens et al., 2005). According to the determined nucleotide sequence, the termini of type 5 and 6 were precluded from the genome of P1312. To have a clue about the most probable packaging strategy employed by P1312, the large terminase subunit of P1312 was compared, in amino acid sequence, to those of bacteriophages with known chromosomal termini. Based on the resulting phylogenetic tree (Figure 8), P1312 is classified into a clan that uses the P22-like headful strategy for DNA packaging. Accordingly, the genomic termini of P1312 are possibly circularly permuted direct terminal repeats. To support this proposition, the phage genome was digested with a variety of restriction enzymes, and the digested products were heated at 80°C for 15 min, followed by a fast or slow cooling process. The samples were then analyzed by agarose gel electrophoresis (Figure 9). If the genome has single-stranded cohesive ends, the two restriction fragments, which have a single cohesive terminus, will join together and appear as a larger fragment in the slow chilled sample. That the restriction patterns of P1312 did not alter in response to the different cooling processes excludes the possibility of the presence of cohesive ends. For headful packaging phages that contain terminally redundant and circularly permuted chromosomes, the restriction pattern may consist of all the fragments expected from a circular genome plus a submolar pac fragment as in the case of P22 (Casjens and Gilcrease, 2009). However, if the headful terminase makes imprecise series initiation cleavage, as is the case for sf6 and ES18, no visible pac fragment will be detected in the restriction pattern. Instead, the terminal fragments will show as blur backgrounds between bands due to their variable lengths (Casjens and Gilcrease, 2009). The restriction patterns created by the various restriction enzymes in Figure 9 are consistent with the predicted results based on a circular P1312 genome. No pac fragment was observed in the electrophoresis gel in this study; however, blur backgrounds were actually present. This result supports the proposition that P1312 is a P22-like headful packaging phage.

Concluding Remarks

Bacteriophage P1312 is the first well-described phage to infect T. fusca. The tolerance to high temperature indicates that P1312 has adapted to the thermophilic stage of composting as T. fusca. In general, P1312 has a genome organization similar to Siphoviridae of Gram-positive bacteria; however, a notable difference is the direction of the portal protein-encoding gene, which renders the genome of P1312 unique. Nocardiopsis lucentensis, an actinomyces species in saline soil habitats (Yassin et al., 1993), is the number 1 ranked organism that provides best-matched homologs with the proteins encoded by the putative ORFs of P1312. This fact suggests the presence of prophages in N. lucentensis or implies a possibly evolutionary origin of P1312. P1312 may represent a valuable resource for understanding the molecular interaction between T. fusca and its co-evolutionary bacteriophages.

Acknowledgments

This work was supported by grants MOST 103-2313-B-005-032-MY3 from the Ministry of Science and Technology, Taiwan, ROC. We would also like to express our gratitude to Professor Michael Bagdasarian, Michigan State University, for his help in the preparation of the manuscript.
References

Adav, S. S., Ng, C. S., Arulmani, M., and Sze, S. K. (2010). Quantitative iTRAQ secretome analysis of cellulolytic Thermobifida fusca. J. Proteome Res. 9, 3016–3024. doi: 10.1021/pr100174z

Boyd, A. C., and Sherratt, D. J. (1995). The pCL plasmid: versatile cloning vectors based on the bacteriophage lambda origin of replication. Gene 153, 57–62. doi: 10.1016/0378-1119(94)00788-T

Campbell, A. (1994). Comparative molecular biology of lambdoid phages. Annu. Rev. Microbiol. 48, 193–222. doi: 10.1146/annurev.mi.48.100194.001205

Casjens, S. R., and Gilcrease, E. B. (2009). Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. Methods Mol. Biol. 502, 91–111. doi: 10.1007/978-1-60327-565-1_7

Casjens, S. R., Gilcrease, E. B., Winn-Stampley, D. A., Schicklimaier, P., Schmieger, H., Pedulla, M. L., et al. (2005). The generalized transducing Salmonella bacteriophage ES18: complete genome sequence and DNA packaging strategy. J. Bacteriol. 187, 1091–1104. doi: 10.1128/JB.187.3.1091-1104.2005

Chen, C. Y., Hsieh, Z. S., Cheepudom, J., Yang, C. H., and Meng, M. (2013). A 24.7-kDa copper-containing oxidase, secreted by Thermobifida fusca, significantly increasing the xylanase/cellulase-catalyzed hydrolysis of sugarcane bagasse. Appl. Microbiol. Biotechnol. 97, 8977–8986. doi: 10.1007/s00253-013-4727-y

Delcher, A. L., Harmon, D., Kasif, S., White, O., and Salzberg, S. L. (1999). Thermobifida fusca Katsura, I. (1990). Mechanism of length determination in bacteriophage lambda. Trends Microbiol. 26, 1–18. doi: 10.1016/0968-0004(90)90004-D

Kenny, J. G., McGrath, S., Fitzgerald, G. F., and van Sinderen, D. (2004). Hofmann, K., and Stoffel, W. (1993). TMbase—a database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler 374, 166.

Hofmann, K., and Stoffel, W. (1993). TMbase - a database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler 374, 166.

Katahira, S., Mizuike, A., Fukuda, H., and Kondo, A. (2006). Ethanol fermentation from lignocellulosic hydrolysate by a recombinant xylose- and cellobiose-glycosaccharide- assimilating yeast strain. Appl. Microbiol. Biotechnol. 72, 1136–1143. doi: 10.1007/s00253-006-0402-x

Katsura, I. (1990). Mechanism of length determination in bacteriophage lambda tails. Adv. Biophys. 26, 1–18. doi: 10.1016/0065-227X(90)90004-D

Kenny, J. G., McGrath, S., Fitzgerald, G. F., and van Sinderen, D. (2004). Bacteriophage Tuc2089 encodes a tail-associated cell wall-degrading activity. J. Bacteriol. 186, 3480–3491. doi: 10.1128/JB.186.11.3480-3491.2004

Lin, Y. R., Chiu, C. W., Chang, F. Y., and Lin, C. S. (2012). Characterization of a new phage, termed Ef1 and its comparison with all available phage and predicted prophage genomes. FEMS Microbiol. Lett. 317, 9–26. doi: 10.1111/j.1574-6968.2010.02203.x

Reece, R. J. (2004). Analysis of Genes and Genomes. Hoboken, NJ: John Wiley & Sons.

Nafissi, N., and Slavec, B. (2014). Bacteriophage recombination systems and biotechnical applications. Appl. Microbiol. Biotechnol. 98, 2841–2851. doi: 10.1007/s00253-014-5512-2

Rahman, O., Pfitzenmaier, M., Pester, O., Morath, S., Cummings, S. P., Hartung, T., et al. (2009). Macroamphiphilic components of thermostable actinomycetes: identification of lipoteichoic acid in Thermobifida fusca. J. Bacteriol. 191, 152–160. doi: 10.1128/JB.01105-08

Cheepudom et al. (2015) Thermobifida fusca-infecting phage. J. Bacteriol. 189, 2477–2486. doi: 10.1128/JB.01899-06

Maki, M., Leung, K. T., and Qin, W. (2009). The prospects of cellulase producing bacteria for the bioconversion of lignocellulosic biomass. Int. J. Biol. Sci. 5, 500–516. doi: 10.7150/ijbs.5.500

Copyright © 2015 Cheepudom, Lee, Cai and Meng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.