The Genomic Sequence and Annotation of Bacteriophage HK239

Alice Ann Wright
Western Kentucky, aliwright84@gmail.com

Follow this and additional works at: http://digitalcommons.wku.edu/theses
Part of the Biotechnology Commons, and the Genetics Commons

Recommended Citation
http://digitalcommons.wku.edu/theses/208

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
A Thesis
Presented to
The Faculty of the Department of Biology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirement for the Degree
Masters of Biology

By
Alice Ann Wright
December 2010
THE GENOMIC SEQUENCE AND ANNOTATION OF BACTERIOPHAGE HK239

Date Recommended 8/16/2010

Rodney King

Director of Thesis
Sigrid Jacobshagen

Claire Rinehart

Dean, Graduate Studies and Research Date
Acknowledgements

This research project was made possible by the support of several individuals. Dr. Gail Christie (Department of Microbiology and Immunology, Virginia Commonwealth University) generated the library necessary to sequence the genome. Dr. Kinchel Doerner assisted in the use of TransTerm. Dr. Andersland helped obtain pictures and measurements of the HK239 viral particles. Dr. Steve Cresawn (Department of Biology, James Madison University) provided the Phamerator analysis. The Biotechnology Center provided the equipment and support for sequencing the genome. Amanda Nolan and Sarah Schrader cloned the HK106 cI gene. The research was funded by WKU Graduate Studies and KY-INBRE NCRR Grant P20 RR16481. Advice and support provided by my committee; Dr. Sigrid Jacobshagen and Dr. Claire Rinehart; and my advisor, Dr. Rodney King, have helped in completing this project.
Table of Contents

Introduction .............................................................................................................3
Materials and Methods .......................................................................................13
Results ..................................................................................................................24
Discussion .............................................................................................................41
Bibliography .........................................................................................................45
Appendix I .............................................................................................................49
<table>
<thead>
<tr>
<th></th>
<th>Table Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Primers used</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>Strains used</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>Genes</td>
<td>28</td>
</tr>
<tr>
<td>4</td>
<td>Terminators</td>
<td>32</td>
</tr>
<tr>
<td>5</td>
<td>Additional genetic elements</td>
<td>32</td>
</tr>
</tbody>
</table>
List of Figures

1. Lytic and lysogenic lifestyles ................................................................. 4
2. Genomic Organization ........................................................................... 6
3. pSMART HCAmp vector ........................................................................ 14
4. Restriction analysis .............................................................................. 25
5. HK239 genome ..................................................................................... 27
6. Verification of ends and cos site ............................................................ 33
7. HK239 and HK106 immunity regions ................................................... 35
8. Phamerator analysis of HK239 and HK022 ........................................... 36
9. Confirmation of HK106 lysogens ............................................................ 37
10. Alignment of HK106 and HK239 cI ...................................................... 38
11. Electron micrograph of HK239 ............................................................. 39
12. Head and tail measurements of HK239 ................................................ 40
Bacteriophages are viruses that infect bacteria and they are the most numerous biological entities on Earth. Temperate phage can adopt two different lifestyles. In the lytic lifestyle, a phage injects its genome into the host and a controlled developmental program ensues. The phage DNA is replicated, phage genes are expressed and new viral particles are assembled. Ultimately, the host cell lyses and the phage particles are released into the environment. In the lysogenic lifestyle, a phage integrates its genome into the host chromosome, creating a prophage. The cell containing the prophage is known as a lysogen. Most prophage genes are not expressed. However, those that are encode a wide variety of functions. One function is exclusion, or the prevention of a different phage type from successfully infecting the lysogenic cell. Most exclusion systems are limited to a specific phage. Bacteriophage HK239 is unique in that it has a wide range of exclusion including Lambda, P1vir, P2, HK022, and T4rII. To learn more about HK239, the genome was sequenced and annotated. The genome is 41,538 bp in length and there are 71 open reading frames. It has a genomic organization similar to other lambda phage and is most closely related to bacteriophage HK022. No additional genes that share homology with known exclusion functions were identified through the sequence analysis of the HK239 genome. It is possible that an open reading frame for which no database matches were found may indeed encode an exclusion function.
**Introduction**

Bacteriophage, or “phage,” are viruses that infect bacteria. Phage are ubiquitous and they are the most numerous biological entities on Earth – one study estimated that there are one to ten million phage per milliliter of seawater [1]. They are relatively simple in genetic organization and have smaller genomes compared to bacteria. This relative simplicity, combined with the ease and rapidity at which large numbers can be generated, has made them a model to better understand molecular processes such as gene expression. In addition, they have served as a tool for moving genetic material between hosts. Many early studies used temperate phage because of their ability to adopt two different lifestyles: lytic and lysogenic. The roots of molecular biology can be traced to a rich array of experiments that were done to understand the elegant genetic switch between these two different lifestyles.

**Lytic Lifestyle**

In the lytic lifestyle the bacteriophage replicates at the expense of the host bacterium (Figure 1). The phage attaches to the host bacterium via a protein on the host’s surface and injects its genome into the host. Phage genes are usually transcribed with the host encoded RNA polymerase. This gene expression occurs in regulated cascades, allowing the lytic cycle to proceed in a very ordered fashion. The genes required for replication of the phage genome are expressed early in infection. The head and tail genes necessary for the formation of the phage particle are expressed later. Late in infection, the phage genome is packaged into the head and the tail is attached, resulting in a complete phage particle. The host cell is then lysed and the new phage particles are released to begin the cycle again [2].
Lytic and lysogenic lifestyles. On the left is the lytic lifestyle and on the right is the lysogenic lifestyle. Phage genomic DNA is in red and the bacterial chromosome in black. In general, the lytic cycle is completed within 45 minutes and approximately 100 viral particles are generated [9].

**Lysogenic Lifestyle**

Like the lytic lifestyle, the phage first attaches to the host bacterium and injects its genome. However, instead of replicating, the phage genome integrates into the host genome, creating a prophage. This is achieved by recombination at *att*, or attachment sites, within the phage and host genomes and is mediated by the phage-encoded integrase protein. A bacterial cell that carries a prophage is called a lysogen (Figure 1). The phage
can exist in this state for many generations of the host until it is induced to enter the lytic life cycle. This can occur spontaneously or, in some cases, upon damage to the host by an external stimulus, such as UV exposure. Upon induction the prophage will excise from the host and begin to replicate [2].

**Organization of the Genome**

Lambdoid phage, a subset of temperate phage, have a distinct genetic organization [3, 4]. Genes are grouped according to function [3] and are expressed in successive cascades during lytic infection (Figure 2, Ref. 2). The initiation of the expression cascade is regulated at the immunity region. This is also the location at which the decision between lysis and lysogeny occurs [2]. In the right operon are the genes for replication of the genome, late antitermination, lysis of the host bacterium, and formation and assembly of the phage particle. In the left operon are the genes for early antitermination, recombination, and other functions that may be specific to that phage [3]. Expression of most genes in both operons is controlled by transcription termination signals, which are read through at certain points during the lytic life cycle [2].

*The Immunity Region*

The immunity region is central to the life cycle of the phage as it contains the genes and regulatory elements necessary for entry into either the lytic or lysogenic life cycle. It is also the point at which the left and right operons diverge [5]. In the establishment of lysogeny, three genes are essential: *cI* (the repressor, a DNA binding protein), *cII*, and *cIII*. The cII and cIII gene products are necessary to initiate repressor synthesis from *P_{RE}*, or promoter for repressor establishment. cII functions as a transcriptional activator and cIII protects cII from degradation by host proteases [6].
After its initial synthesis, the repressor is able to regulate its own expression from $P_{RM}$, the promoter for repressor maintenance [2]. The repressor prevents transcription from the left and right promoters, $P_L$ and $P_R$. It does so by binding to specific sites called operators ($O_L$ and $O_R$) [7]. Cro (another DNA binding protein), on the other hand, promotes the lytic lifecycle [8]. It competes with the CI protein to bind at the operator sites and permit expression of the downstream genes necessary for completion of the lytic life cycle [2].

**Figure 2.** Schematic illustrating the genomic organization of Lambdoid phage [3, 4]. Arrows above the map indicate leftward and rightward transcripts [9].

**Integration**

Integration of the phage genome into the host genome is essential for establishment of lysogeny. Within the phage genome, near the integrase, is an $attP$ site.
Similarly, on the host genome there is an \( \text{attB} \) site. The integrase protein catalyzes recombination between these two sites, allowing for the creation of the prophage [9].

**Antitermination**

Gene expression occurs in regulated cascades, made possible by the presence of promoters and terminators at key sites in the genome. The terminators are thought to prevent inappropriate gene expression during lysogeny. When a phage enters the lytic life cycle, expression of genes downstream of the terminators is required for successful phage replication. A phage encoded protein, called N, promotes transcription antitermination. It recognizes \( \text{nut} \) (N utilization) sites, comprised of \( \text{BoxA} \) and \( \text{BoxB} \) regions, on the nascent RNA and modifies RNA polymerase in such a way that it can read through terminators and transcribe downstream genes [2]. There is a second phage encoded antiterminator protein, called Q, that is expressed later in the phage lytic cycle. Q recognizes \( \text{qut} \) (Q utilization sites) and allows for read through of transcription terminators and of downstream genes [9].

**Replication**

In lambda, genes O and P are responsible for replication of the phage genome. These genes are located directly downstream of \( \text{cro} \) and are expressed early in the lytic life cycle [2].

**The Head and Tail Genes**

The region at the beginning of the genome encodes proteins for viral particle (Figure 2) formation and DNA packaging. The head portal protein and the major head subunit precursor are part of the head protein gene cluster and are involved in forming the capsid [4]. The head maturation protease cleaves the major head subunit precursor,
allowing for expansion and strengthening of the head [4], and then cleaves itself [10]. The terminase is responsible for linearizing the phage DNA and then helps package it into the head [9]. The tail is assembled from tail protein subunits [9].

Lysis

Lysis of the host bacterium is the last stage of the lytic life cycle. At least three genes are usually involved in this process: holin, lysin, and Rz. During lysis, holin is responsible for generating holes in the host membrane. These holes are large enough to permit passage of the enzymes that actually lyse the cells. Lysis is achieved by attacking the peptide or glycosidic bonds in the host cell wall [11].

Importance of Lysogens

Most prophage genes are not expressed. However, those that are expressed have a wide variety of important functions. Phage CTXφ is an example of a medically important phage. It carries the genes that encode cholera toxin, which are expressed in lysogenic strains of Vibrio cholera [12]. Phage are not only important from a medical standpoint. For example, marine cyanophage are known to carry genes involved in photosynthesis, potentially contributing to the metabolism of the host bacterium [13]. This and similar discoveries have led to a rethinking of the roles played by viruses in marine ecosystems and ecology on a broader scale.

In addition to these examples, the presence of a prophage can confer other advantages to the host bacterium. Protection against infection by other phage is a good example. There are two means by which phage are unable to infect a lysogen. The first is homoimmunity. In this case, phage are unable to successfully infect a lysogenic cell of the same immunity type [2]. This is mediated by the prophage repressor which
recognizes the operator binding sites of the infecting phage. If the prophage repressor is able to bind, it will shut down the expression of genes necessary for lytic growth of the infecting phage. The second means of protection is exclusion. Exclusion can be achieved through a broad range of mechanisms which ultimately achieve the same goal: preventing a different type of phage from successfully infecting the lysogen. Some well documented examples are presented below.

**Phage Exclusion Mechanisms**

*P22 SieA*

One of the exclusion mechanisms encoded by P22 involves the product of the *sieA* gene. The *sieA* gene product is believed to exclude phage at the level of injection by preventing entry of phage genomic DNA. *SieA*, like all exclusion genes, is expressed by the prophage. The gene’s expression is constitutive, however the quantity of *SieA* protein is likely regulated by its high percentage of low usage codons [14].

*φ80 cor*

Like the P22 *sieA* gene, the exclusion function encoded by φ80 *cor* also works at the level of injection [15]. The Cor protein has an N-terminal transmembrane helix that allows it to interact with the FhuA protein on the surface of the host cell [16]. FhuA, which normally allows for ferrichrome uptake, also functions as a phage receptor. Any phage that uses this receptor to attach and inject its genome into the host would be inhibited by *cor*-containing lysogens [15]. This function allows φ80, and other phage containing the *cor* gene, to exclude HK022, T1, and N15 [15].
HK022 nun

Not all exclusion mechanisms work at the level of infection. In HK022, the nun gene encodes a transcription terminator whose function blocks lambda phage growth \[17, 18\]. Nun protein recognizes the lambda nutL and nutR sites \[19\]. Binding of Nun at the nut sites prevents N protein from recognizing the same sites. While preventing N from functioning as an antiterminator, Nun also terminates transcription, thus halting the lytic life cycle. In addition to transcription termination, Nun may also prevent N gene translation. There is recent evidence that Nun has a secondary exclusion function that blocks an RNAseIII processing event necessary for N translation \[20\].

e14 lit

e14 is a defective prophage in Escherichia coli K-12 \[21, 22\] that encodes a T4 exclusion function. A protease called Lit (late inhibitor of T4, Ref 23) cleaves EF-Tu, causing all translation to cease \[24\]. This protease is activated by a small peptide called gol (“growth on lit,” Ref. 25). Gol is cleaved from the major head protein of T4 during formation of the phage head \[26\]. Gol binds to EF-Tu which then creates a substrate for Lit \[27\]. Cleaving EF-Tu effectively prevents successful infection by T4 \[28\].

Lambda rex

The Lambda rex system encodes two genes, rexA and rexB \[29\] that are responsible for T4rII exclusion \[30\]. T4rII replication in a Lambda lysogen triggers the rex system \[28\]. RexB is an ion channel \[28, 31\] that is activated by RexA \[28\]. rexB is expressed from its own promoter, pLIT, and the protein is found in larger quantities in the cell than RexA. This ratio changes upon T4rII infection as RexA levels increase,
resulting in activated RexB [28]. This activation causes a loss of membrane potential that kills the host cell and stops the spread of T4rII [28].

**Phage Genomics**

As sequencing technologies have improved, more genomes of organisms have been sequenced, including phage. More than 600 have been sequenced [32]. Phage genomes are generally easier to sequence because of their small size [33]. However, sequencing the genome is only the first step. The genome must also be annotated. Multiple computer based methods have been developed for identifying open reading frames (ORFs). ORFs are the easiest to identify due to the conserved sequences for start and stop codons [34]. Sequencing and annotating phage genomes presents valuable information about the phage and their evolutionary relationships with other phage. Phage appear to be mosaics of each other and this can complicate our ability to establish ancestral relationships [33]. This can be accomplished via genome comparison tools such as dotplots or BLASTs at the protein or nucleotide level to compare individual genes [34]. The availability of the genomic sequence also allows the researcher to ask certain questions about the phage that he or she might not have been able to ask before. For example, relatedness and evolutionary history among organisms/viruses can be more fully explored with a genomic sequence. Also, the functions of genes, whose presence might otherwise have gone unnoticed, can be more fully explored.

**Bacteriophage HK239**

Bacteriophage HK239 is a lambdoid phage and thus shares a similar genetic organization as other members of this group. It was isolated in the early 1970s from cow dung in Hong Kong by Dhillon and Dhillon [35]. It is unique in that it has a wide range
of exclusion [36]. HK239 lysogens were reported to exclude \( \lambda \), T4rII, P1\textit{vir}, P2, and HK022 [35]. Previous work by Wright \textit{et. al.} had attempted to explain how this wide range of exclusion is achieved [37]. This research was done with the only known phage stock available: a lytic mutant. Since lysogens could not be generated to conduct genetic experiments, we decided to attempt to identify the exclusion gene(s) by cloning pieces of HK239 DNA into a plasmid vector and screening cells transformed with these plasmid clones for phage resistance. A clone containing a \( \phi 80 \) \textit{cor} homolog was successfully isolated based on its ability to exclude phage HK022. The specificity of the exclusion (only HK022 growth was prevented) suggested that there were other HK239 genes that encode exclusion functions [37].

The goal of this research was to sequence and annotate the genome of bacteriophage HK239. It was expected that this project would provide more information about the exclusion phenotype of bacteriophage HK239 and insight into why it is a lytic mutant. It was also expected that the genomic sequence would yield some information about the evolutionary relatedness of HK239 and other lambdoid phage.
Materials and Methods

Preparation of Genomic DNA

Cultures of *Escherichia coli* strain LE392 were grown overnight at 37°C in TB (1% tryptone and 0.5% NaCl). The next day the bacterial cells were pelleted and resuspended in half the volume of 10 mM MgSO$_4$. A stock of bacteriophage HK239 was prepared by plating dilutions on LE392. Serial dilutions of the HK239 stock were mixed with 50 µL of an overnight culture of LE392. Following a 10 minute incubation at 37°C, 3 mL of molten (55°C) TB top agar (1% tryptone, 0.5% NaCl, and 0.75% agar) was added to each phage/bacterial mixture, vortexed, and poured onto prewarmed LB agar plates. After cooling, the plates were inverted and incubated overnight at 37°C. The following day plaques were picked with Pastuer pipettes and the plugs were stored in 200 µL TMG (10mM Tris HCl pH 7.4, 10 mM MgSO$_4$ x 7H$_2$O, and 0.1% gelatin) [9].

To 150 µL of overnight LE392 cells, an agar plug of HK239 was added. Fifty mL of LB (1% tryptone, 0.5% yeast extract, and 0.5% NaCl) were added and the cultures were grown at 37°C until clearing was seen. Two hundred µL of chloroform were added to the cultures and vortexed well. The cultures were then centrifuged at 1400 xg for 10 min and the supernatant was recovered. Genomic DNA was extracted from the phage particles using the Qiagen ® lambda maxi kit (titers of $5 \times 10^9$ to $3 \times 10^{10}$ are required for this kit; cat. no. 12562).

Generation of Library

HK239 genomic DNA was sent to Dr. Gail Christie (Department of Microbiology and Immunology, Virginia Commonwealth University) who sheared the DNA with a Hydroshear machine into ~1.1 kb fragments. The fragments were blunt-end ligated into a
pSMART® HCAmp vector between primers SL1 and SR2 (Figure 3). The library was transformed into 10G Elite competent cells from Lucigen as follows: 1 µL of the HK239 library was added to 25 µL competent cells. As a control, an equal volume of pUC19 was used in place of the library. Cells were transferred to a 1 mm electroporation cuvette. The cells were pulsed once at 2.5 kV, 25 µF, and 200 ohm with a time constant of 4.62. To the cells, 975 µL of SOC (2% tryptone, 0.5% yeast extract, 0.05% NaCl, 250 mM KCl, 10 mM MgSO₄, and 20 mM glucose) was added immediately and the cells were transferred to 15 mL tubes. The cells recovered for 1 hr at 37°C. Fifty µL of electroporated cells were plated on each of 10 LB (25 µg/mL) ampicillin plates for the library. Plates were incubated overnight at 37°C. The resulting colonies were picked and grown overnight in 5 mL TB (25 µg/mL) ampicillin at 37°C. The cultures were pelleted and resuspended in 2.5 mL 10 mM MgSO₄.

Figure 3. pSMART® HCAmp vector, accession number AF399742 [41]. The primers, SL1 and SR2, flank the insertion site. The Ap-R gene allows for selection of ampicillin resistant colonies.
Sequencing

Plasmids were isolated from ampicillin resistant clones using the QIAwell® 8 Ultra Plasmid Kit (cat. no. 16152). The purified plasmid DNA was analyzed on 1% agarose gels stained with Ethidium Bromide. Plasmids containing inserts were identified by their altered migration relative to the vector control. Sequencing reactions were performed using SL1 and SR2 primers (see Table 1 and Figure 3). A typical sequencing reaction contained the following: 100 ng DNA, 1 µL primer, 4 µL BigDye Terminator 3.0 (later, during additional sequencing, the reaction was altered to 2 uL BigDye Terminator 3.0 and 2 µL buffer to conserve the BigDye Terminator; ABI cat. no. 4336917), and npH2O to 10 µL total volume. Thermocycler conditions were as follows: 25 cycles of 96°C for 30s, 60°C for 30s, and 72°C for 4 min followed by a 10°C hold. Reactions were cleaned using either the DyeEx™ 2.0 Spin kit from Qiagen (cat. no. 63204) or the Sigma-Spin post-reaction clean-up columns (cat. no. S5059-70EA). The samples were dried down using a centrivap. Samples were resuspended in 15 µL hi-di formamide and loaded onto an ABI3130 for analysis.

Assembly

Sequencing data were analyzed using VectorNTI and Geneious software. Poor sequence was removed from the ends of each read. Homologies to other phage sequences were identified using nucleotide BLAST. Then the data were organized in an Excel spreadsheet according to overlapping homologies to other phage identified by nucleotide BLAST. Sequences were aligned and assembled based on observed overlap. Sequence gaps were resolved by sequencing directly from phage DNA by primer walking.
Regions covered by only one sequence read were re-sequenced and both strands were covered.

**Verification of the assembled contig by restriction analysis**

To verify the assembled genome, three digests were used: MfeI (NEB cat. no. R0589S, 10 U/µL), HindIII (NEB cat. no. R0104S, 20 U/µL), and an AhdI (NEB cat. no. R0584S, 5 U/µL) with NcoI (NEB cat. no. R0193S, 10 U/µL) double digest. The reactions were set up as follows: 1-3 µg of DNA, 1 µL NEB buffer (buffer 4 for MfeI and AhdI with NcoI doubled digest and buffer 2 for HindIII), 1 µL of BSA (AhdI/NcoI only), 1 µL of enzyme, and npH2O to 10 µL total. The digests were incubated for 3hr at 37°C and the enzyme was heat-killed by incubating for 20 min at 65°C. The digests were analyzed on 1% agarose gel and stained with ethidium bromide.

**Annotation**

Open reading frames were identified using two programs: Viral Genome Organizer [38] and GeneMark [39]. All annotated open reading frames were analyzed using nucleotide and protein BLAST. Protein and nucleotide e-values provided in BLAST were used to determine how likely the gene encoded the same function as the homologs identified in the search. The lower the e-value, the more similar the homologs were in sequence. Open reading frames that had high e-values (close to or above one) or had no hits in the database received a number in place of a name. TransTerm was used to identify rho-independent transcription terminators [40]. Additional genomic elements were identified based on homology in other phage.
Table 1. Primers used

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR2</td>
<td>GGTCAGGTATGATTTAAATGGTCAGT</td>
<td>sequence from inserts</td>
</tr>
<tr>
<td>SL1</td>
<td>GCAGTCCAGTTACGCTGGAGTC</td>
<td>sequence from inserts</td>
</tr>
<tr>
<td>S1</td>
<td>CGGATATCGCTGAATATATCGGGT</td>
<td>primer walking</td>
</tr>
<tr>
<td>S2a</td>
<td>CAGCCGCAACACAGTCGAAATT</td>
<td>primer walking</td>
</tr>
<tr>
<td>S2b</td>
<td>CTTTGCCAACCGGCAGCCTTGGAGCT</td>
<td>primer walking</td>
</tr>
<tr>
<td>S3a</td>
<td>CCGTTTCGCATCTGAGTT</td>
<td>primer walking</td>
</tr>
<tr>
<td>S3b</td>
<td>GTCAACCGGCCTCAAATTTGA</td>
<td>primer walking</td>
</tr>
<tr>
<td>S4a</td>
<td>AAAAGGCCACGTCGCTCTCCTGGAGCT</td>
<td>primer walking</td>
</tr>
<tr>
<td>S4b</td>
<td>CGGTGAAACACGCACTCGCAG</td>
<td>primer walking</td>
</tr>
<tr>
<td>S5a</td>
<td>GAAACGGATGGGAATCACCGGT</td>
<td>primer walking</td>
</tr>
<tr>
<td>S5b</td>
<td>GGCCGTGCTTATTACTGCTGCT</td>
<td>primer walking</td>
</tr>
<tr>
<td>S67a</td>
<td>GCCGATTCGAGCTTCAAATAA</td>
<td>primer walking</td>
</tr>
<tr>
<td>S67b</td>
<td>CGGAAATTCATCAGCTACAGCTGACC</td>
<td>primer walking</td>
</tr>
<tr>
<td>S8a</td>
<td>CATCCATCGAGACGACGATCCGT</td>
<td>primer walking</td>
</tr>
<tr>
<td>S8b</td>
<td>CGAAACATATGTGCGCAGCTCCTGCTCT</td>
<td>primer walking</td>
</tr>
<tr>
<td>S9a</td>
<td>GAAAGACCAGCAGCTGCGAGATTC</td>
<td>primer walking</td>
</tr>
<tr>
<td>S9b</td>
<td>GGGTCTATTGTTGGTGGGCTCTCTAAA</td>
<td>primer walking</td>
</tr>
<tr>
<td>S10a</td>
<td>TGCGCTGAGCCTCTATCAGTC</td>
<td>primer walking</td>
</tr>
<tr>
<td>S10b</td>
<td>TGCCGGTATCGCTCTCCTGATTTAA</td>
<td>primer walking</td>
</tr>
<tr>
<td>S11a</td>
<td>GGTGTTGTGCTTCCGCAATGCTATA</td>
<td>primer walking</td>
</tr>
<tr>
<td>S11b</td>
<td>TGCAGCCACGACAGCTGAGT</td>
<td>primer walking</td>
</tr>
<tr>
<td>S12a</td>
<td>GCAAAAGAGGACAGAGACAGACGAG</td>
<td>primer walking</td>
</tr>
<tr>
<td>S12b</td>
<td>CCATCCCTCGTTCTCAGTCTGCTA</td>
<td>primer walking</td>
</tr>
<tr>
<td>S14a</td>
<td>TTCTCATGTCAAGCAGGGA</td>
<td>primer walking</td>
</tr>
<tr>
<td>S14b</td>
<td>GATGGTTTTCCATGCGTGCTTGC</td>
<td>primer walking</td>
</tr>
<tr>
<td>S16a</td>
<td>CGTCAGCGAGGCTTCTGATG</td>
<td>primer walking</td>
</tr>
<tr>
<td>S16b</td>
<td>GTTAAGCCGCGCTGATGACGCTC</td>
<td>primer walking</td>
</tr>
<tr>
<td>S18</td>
<td>CCAATACGTTTAAATCTTCTCTGCGA</td>
<td>primer walking</td>
</tr>
<tr>
<td>S2:1</td>
<td>TCGATCCCCAGACAGCGCAACACAA</td>
<td>primer walking</td>
</tr>
<tr>
<td>S2:8a</td>
<td>CGTCACTCTCCTCAGCTCGCC</td>
<td>primer walking</td>
</tr>
<tr>
<td>S2:8b</td>
<td>GGGTCAATTGGTGCTGTAGGCTCTAAAA</td>
<td>primer walking</td>
</tr>
<tr>
<td>S2:11a</td>
<td>AGCATGTTGCTCTGCAGCGTCG</td>
<td>primer walking</td>
</tr>
<tr>
<td>S2:12b</td>
<td>CATCCTCTCGCTTCTGTATGCCTA</td>
<td>primer walking</td>
</tr>
<tr>
<td>S2:13b</td>
<td>GGCCTGATCTGCTGTCGCTT</td>
<td>primer walking</td>
</tr>
<tr>
<td>S2:14a</td>
<td>TCAAGCCGAGAATGTTCTCGC</td>
<td>primer walking</td>
</tr>
<tr>
<td>S2:14b</td>
<td>TCAACCCACCTGAGTCTCAGCA</td>
<td>primer walking</td>
</tr>
<tr>
<td>S2:17a</td>
<td>CGCCAGCATATCGAGGAACG</td>
<td>primer walking</td>
</tr>
<tr>
<td>S2:17b</td>
<td>CGCTCTGTTATCTGCATCTCGT</td>
<td>primer walking</td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence</td>
<td>Purpose</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>395</td>
<td>CAGTTCAAGGAGGATGCCG</td>
<td>primer walking</td>
</tr>
<tr>
<td>396</td>
<td>GTCATTCTGTGCTGTTC</td>
<td>primer walking</td>
</tr>
<tr>
<td>397</td>
<td>TAATCCTACAAACAAAG</td>
<td>primer walking</td>
</tr>
<tr>
<td>398</td>
<td>ACTGGTTTCCTGTTCCTCA</td>
<td>primer walking</td>
</tr>
<tr>
<td>399</td>
<td>GAACGCTGACGACTGT</td>
<td>primer walking</td>
</tr>
<tr>
<td>400</td>
<td>CTTCTCAGTATGCGTTG</td>
<td>primer walking</td>
</tr>
<tr>
<td>401</td>
<td>TGCGTACCAAACATAAAAC</td>
<td>primer walking</td>
</tr>
<tr>
<td>402</td>
<td>TACCATAAAATAGTACGAGT</td>
<td>primer walking</td>
</tr>
<tr>
<td>403</td>
<td>GCATAGCAAGATGGGTA</td>
<td>primer walking</td>
</tr>
<tr>
<td>404</td>
<td>GCCTCTATCCAGTCGTTG</td>
<td>primer walking</td>
</tr>
<tr>
<td>405</td>
<td>GTGTAATACTTCTGACT</td>
<td>primer walking</td>
</tr>
<tr>
<td>406</td>
<td>CATTCTGGCTGGAGGTTGA</td>
<td>primer walking</td>
</tr>
<tr>
<td>407</td>
<td>GTCAGCAACAAATCTGAT</td>
<td>primer walking</td>
</tr>
<tr>
<td>408</td>
<td>TATCTGTTCCCTCTGACCA</td>
<td>primer walking</td>
</tr>
<tr>
<td>409</td>
<td>GTATGAGACAGAAGTCAGC</td>
<td>primer walking</td>
</tr>
<tr>
<td>410</td>
<td>ACTACAGTAACGGACTGC</td>
<td>primer walking</td>
</tr>
<tr>
<td>411</td>
<td>GATCAGTTCTGCTAGGGT</td>
<td>primer walking</td>
</tr>
<tr>
<td>412</td>
<td>GAAGAGTGCCGATATGGGC</td>
<td>primer walking</td>
</tr>
<tr>
<td>413</td>
<td>CTTCAGAACTGAGTTCT AG</td>
<td>primer walking</td>
</tr>
<tr>
<td>414</td>
<td>GATATCATTCCAGAGGACG</td>
<td>primer walking</td>
</tr>
<tr>
<td>415</td>
<td>GTCTGATATTGTCATGCCA</td>
<td>determining left end</td>
</tr>
<tr>
<td>416</td>
<td>GATATCATTCCAGACCCAG</td>
<td>determining right end</td>
</tr>
<tr>
<td>417</td>
<td>CATACGTAACGGACTGC</td>
<td>primer walking</td>
</tr>
<tr>
<td>418</td>
<td>GACCGATAGAAGAAGTC</td>
<td>primer walking</td>
</tr>
<tr>
<td>419</td>
<td>GATGCCAGCAAAATCTGAT</td>
<td>primer walking</td>
</tr>
<tr>
<td>420</td>
<td>GTAGTGCCTCCTGCTATG</td>
<td>primer walking</td>
</tr>
<tr>
<td>421</td>
<td>GAGAAATGGGTAAGCGA</td>
<td>primer walking</td>
</tr>
<tr>
<td>422</td>
<td>GTCTGATATTGTCATGCCA</td>
<td>determining right end</td>
</tr>
<tr>
<td>423</td>
<td>GTAAACGCGTGATACTAGA</td>
<td>determining left end</td>
</tr>
<tr>
<td>424</td>
<td>CTTGCAGGTAGATTTA</td>
<td>primer walking</td>
</tr>
<tr>
<td>425</td>
<td>CTTAGAAGTAGTATGAG</td>
<td>primer walking</td>
</tr>
<tr>
<td>426</td>
<td>GTTCACCTTTGGTTATTGC</td>
<td>primer walking</td>
</tr>
<tr>
<td>427</td>
<td>GTATTTATGTCACACCG</td>
<td>primer walking</td>
</tr>
<tr>
<td>428</td>
<td>GTTGTGGGAAAGTTATC</td>
<td>primer walking</td>
</tr>
<tr>
<td>429</td>
<td>GATCCCACTGCTAGAGAG</td>
<td>primer walking</td>
</tr>
<tr>
<td>430</td>
<td>CTATGTTTCTAGTGGTTGATC</td>
<td>primer walking</td>
</tr>
<tr>
<td>431</td>
<td>GCAGGGGTGTATGTTGTTG</td>
<td>primer walking</td>
</tr>
<tr>
<td>432</td>
<td>GAGGTATATGACAAACCCAG</td>
<td>primer walking</td>
</tr>
<tr>
<td>433</td>
<td>CATACGCACTTTCTATG</td>
<td>primer walking</td>
</tr>
</tbody>
</table>
Table 1 continued

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>436</td>
<td>CTTTCCAAGCACGGATA</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>437</td>
<td>GCACCCCGTATTAACGATG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>438</td>
<td>CTGGTGGGCAAGGCTGAAGTC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>439</td>
<td>GTTGTCCGTTTCGTCGATC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>440</td>
<td>GTGCCGGGAGTTGACTTGG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>441</td>
<td>GACCCGAAAAGTGGCGAT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>442</td>
<td>GTCCCTGTCGTCTTCTCTCA</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>443</td>
<td>CATCAATTTGACTGTAAT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>444</td>
<td>CAGTCTTTCCAGCTCGCT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>445</td>
<td>TATCCAGTTCGCTCGGCTG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>446</td>
<td>GCATAATCCTTTACTACATC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>447</td>
<td>CGATTCGACGCTTACCAT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>448</td>
<td>CTGGACAATACGTCTCGGT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>449</td>
<td>TAATCCCTACAACCAAAG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>450</td>
<td>CCATATTTGGATTTCGAG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>451</td>
<td>GTAAAGCTTATGCCGACG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>452</td>
<td>GAAACCCTTGTTCGACG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>453</td>
<td>CATGACAACGTACAATG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>454</td>
<td>CATAAGCATTTAGCCCTACAT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>455</td>
<td>TATCCAGCTCAGCTCGT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>456</td>
<td>CTTACCAGACATCAGTCTC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>457</td>
<td>GCCCTAATACAGCGGCG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>458</td>
<td>GTACTCATGCTATGCGG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>459</td>
<td>GAAACCCTTGTTCATG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>460</td>
<td>CATGACAACGTACAATG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>461</td>
<td>CCATATTTGGATTTCGAG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>462</td>
<td>ATTTCCATAATACAGAAGC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>463</td>
<td>CTTGAACGAATCACCAGCTCA</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>464</td>
<td>ATTTCCAAATAATCTG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>465</td>
<td>CATGACAACGTACAATG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>466</td>
<td>GACCCGAAAAGTGGCGAT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>467</td>
<td>GTTGTCCGTTTCGTCGCT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>468</td>
<td>GTATAGTCAGCAAGTAGC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>469</td>
<td>CATCAGTCCGATTAGGCAG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>470</td>
<td>GTGTGGAAGTCTGTCCAGGA</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>471</td>
<td>GTGTGGAAGTCTGTCCAGGA</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>472</td>
<td>GTGTGGAAGTCTGTCCAGGA</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>473</td>
<td>GTGTGGAAGTCTGTCCAGGA</td>
<td>re-sequencing</td>
</tr>
</tbody>
</table>
Table 1 continued

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>474</td>
<td>GAAAGCTCTTCGAGTTATC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>475</td>
<td>CTGGAAGTGTGTGTTTAC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>476</td>
<td>CAAAATCGGCCATTGATGGC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>477</td>
<td>GCCAGCGTTTCGATGGTA</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>478</td>
<td>GAGGCTGCTTAATGGCTA</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>479</td>
<td>CGGATTCGGAATGGCTGC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>480</td>
<td>CAGCGAAGCGTTTGATAAG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>481</td>
<td>GATTCCACITCTGAGACG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>482</td>
<td>GCAGGAATACATCAGGAC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>483</td>
<td>GAGGGTTACCTGACTTAA</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>484</td>
<td>CAGTTCACTTTACCTGAAAT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>485</td>
<td>CTGAATATCCACGCCAAAT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>486</td>
<td>CTGAGCCGGGTCACTGGGC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>487</td>
<td>CAGGGCGTTCGCAGAGCG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>488</td>
<td>GTTATCCGCCTTCCAATC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>489</td>
<td>CGATGCGCTGTTATGATAT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>490</td>
<td>CATTGCTCCGTGTATTACAC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>491</td>
<td>CGGTTGTCACGGAGCCAT</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>492</td>
<td>CGTATCGATCGTCTTCTG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>493</td>
<td>GATGTAGCCCGATGAACAC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>495</td>
<td>CAAAACGCAGGAGTGAACAC</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>496</td>
<td>GAGTGATGTAATACCTG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>497</td>
<td>CGGCAGATGAAGGTTGATG</td>
<td>re-sequencing</td>
</tr>
<tr>
<td>498</td>
<td>GTTGTAGGGCGTTAGGGAAG</td>
<td>re-sequencing</td>
</tr>
</tbody>
</table>

Verification of the genomic ends

The ends of the genome were predicted based on their homology to bacteriophage HK022. Two primers for each end were designed to sequence in the direction of the predicted ends. In addition, an aliquot of the genomic DNA was treated as follows: 100 ng genomic DNA, 1 µL 10x buffer, 1 µL ligase (NEB cat. no. M0202S, 400 U/µL), and water to 10 µL; 14°C overnight. To identify the cos sites, the ligated genomic DNA was used as a template in a PCR reaction with primers 422 and 423. The PCR product was
sequenced with primers 416 and 417. Sequencing reactions were carried out as before. Sequences were aligned using Geneious [41].

**Bioinformatics analysis**

Genome wide comparisons were made using two types of soft-ware. Dotplot analysis was done in Geneious [41] with the HK239 and HK106 (accession number EF120461) immunity regions and the HK239 and HK022 genomes. A Phamerator analysis was performed on the HK239 and HK022 (accession number AF069308) genomes [42].

**Generating an HK106 lysogen and screening for homoimmunity**

Serial dilutions of HK106 were made in TMG. Five µL of each phage dilution were spotted onto a lawn of *E. coli* strain RK898 (MG1655). After allowing the spots to dry, the plates were inverted and incubated overnight at 37ºC. The next day cells from the center of a spot were streaked for isolation on an LB agar plate and grown overnight at 37ºC. After overnight growth, 4 potential lysogens were purified by streaking for isolation on LB agar plates. Isolated colonies were used to inoculate 5 mL of TB for overnight cultures. The next day, the cells were pelleted and re-suspended in 2.5 mL of 10 mM MgSO₄.

Suspected HK106 lysogens were verified by PCR. The reactions were set up as follows: 1 µL of cells, 1 µL primer 469, 1 µL primer 424, 8.3 µL PCR mix (0.6 mM dNTPs, 3X buffer B, and 8.25 mM MgCl₂), and 0.25 µL taq polymerase (FB60050, 5 U/µL). In place of cells, HK106 lysate was used as a positive control. Cycle conditions were as follows: 94ºC 3 min; 25 cycles of 94ºC 1 min, 55ºC 1 min, and 72ºC 1 min; and a 4ºC hold. Samples were analyzed on a 1% agarose gel stained with Ethidium Bromide.
Cultures positive for containing an HK106 lysogen were stored at -80°C in a mixture of 800 µL lysate and 200 µL of 80% glycerol.

**Visualization of HK239 and head and tail measurements**

High titer \((10^{12} \text{ phage/mL})\) liquid lysates of HK239 were prepared as follows. Two cultures were prepared by inoculating 150 mL of LB with 1 mL of *E. coli* strain RK898 overnight culture. The cultures were incubated at 37°C until growth was visible (approximate OD\(_{600}\) = 0.2-0.3). Then 50 µL of a HK239 phage stock was added to the culture. The culture was incubated at 37°C for 6 hours, shaking, in a baffled flask. The cells and debris were pelleted by centrifugation for 10 min at 10,000 xg at 4°C. The supernatant was treated by adding 288 µL of DNase and 57.6 µL of RNase A at a concentration of 1 unit/mL each. The lysate was incubated at room temperature for 30 min. Solid NaCl was added to a concentration of 1M and the lysate was incubated on ice for 1 hour. PEG 8000 was added to a concentration of 10% weight per volume. After the PEG 8000 dissolved, the lysate was transferred to eight 50 mL centrifuge tubes (Oak Ridge centrifuge tubes, PPCO) and incubated on ice for an hour. The lysate was centrifuged at 11,000 xg at 4°C for 10 minutes. The supernatant was discarded and the tubes were inverted and allowed to air dry for five minutes. The phage pellets were resuspended in 4 mL of TMG total. The phage suspension was chloroform extracted twice with an equal volume of chloroform. The aqueous layer was collected and stored in a 15 mL conical tube at 4°C.

For visualization on the TEM, 15 µL of phage lysate were mixed with 15 µL of 1% uranyl acetate. A grid was set on the phage mixture for 30 s and then transferred to water for 30 s. The remaining liquid was wicked off and the phage were examined under
the TEM at 60V. Scale bars were determined using a ruler grid. Tail and head measurements were made using Auto-Montage software from Syncroscopy.

Table 2. *E. coli* strains used.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>LE392</td>
<td><em>supE44 supF58 hsdR514 galK2 galT22 metB1 trpR55 lacY1</em> [9]</td>
</tr>
<tr>
<td>RK898</td>
<td>MG1655; wild-type</td>
</tr>
<tr>
<td>10G Elite</td>
<td>F- <em>mcrA D(mrr-hsdRMS-mcrBC) ϕ80dlacZΔM15 ΔlacX74 endA1 recA1araD139 Δ(ara, leu)7697 galU galK rpsL nupG λ- tonA</em> [43]</td>
</tr>
<tr>
<td>RK1212</td>
<td>HK106 lysogen of RK898</td>
</tr>
</tbody>
</table>
Results

Genomic sequence of HK239 and verification through restriction analysis

The data generated from shotgun sequencing of phage HK239 was assembled into nine contigs. Primer walking was used to close the remaining gaps and generate a single contig. In addition, areas with low coverage were re-sequenced to ensure the quality of the final genomic sequence. The entire sequence was determined on both strands. Through this analysis, we have shown that the HK239 genome is 41,538 bp in length. The assembly was verified by restriction analysis (Figure 4). Lanes 3 and 4 contain HK239 genomic DNA digested with MfeI. The expected fragments were 20,517 bp, 4,823 bp, 4,301 bp, 4,166 bp, 2,868 bp, 2,160 bp, 1,293 bp, 931 bp, and 509 bp. These size estimates correspond to fragments seen on the gel. However, the bands at 4,823 bp and 2,868 bp are faint. These fragments contain the physical ends of the genome which means that they possess cos sites. It is likely that during the digestion these cos sites are annealing to one another, creating a larger band. Although it is difficult to see due to smearing, there may be a band of 7,691 bp, which would be the size of the annealed fragments. Lanes 5 and 6 contain HK239 genomic DNA digested with HindIII. The expected bands were 13,477 bp, 10,163 bp, 6,422 bp, 5,036 bp, 3,810 bp, 1,341 bp, 711 bp, and 606 bp. This corresponds to what is seen on the gel except that the band at 1,341 bp is faint and the one at 606 bp is not visible. Like the MfeI digest, these fragments include the cohesive ends which probably annealed to one another during the digest. This is supported by a band visible at 1,947 bp, the expected size for the annealed fragments. The correspondence between the expected bands and the fragments visible on the gel confirm that the assembly of the genome is correct.
Figure 4. Restriction digests of HK239 genomic DNA. Lanes 1 and 8 contain a Lambda HindIII ladder. Lanes 2 and 7 contain a mid-range ladder (Fisher). Lanes 3 and 4 contain HK239 genomic DNA digested with MfeI, heat inactivated and untreated respectively. Lanes 5 and 6 contain HK239 genomic DNA digested with HindIII, heat inactivated and untreated respectively. The numbers in white correspond to the molecular weight markers.
Annotations

The HK239 genome was fully annotated using Viral Genome Organizer, GeneMark, and TransTerm (Figure 5). Seventy-one open reading frames were annotated based on a comparison of the output from Viral Genome Organizer and GeneMark programs. Not all of these open reading frames could be assigned a function based on protein and nucleotide BLAST analyses so they were assigned an arbitrary name (see Table 3). Twelve rho-independent transcription terminators were identified using the TransTerm program (Table 4). Additional elements such as promoters and protein binding sites were identified based on homology to other phages in the database (Table 5).

**Figure 5 (next page).** The annotated HK239 genome [41]. The black line represents the genomic sequence and the red arrows represent predicted open reading frames. The blue arrows represent promoters and the orange arrows represent terminators. The gray arrows indicate DNA and RNA binding sites and the cos site.
Table 3. Features of Bacteriophage HK239 genes and their homologies to other phage (if any) with the corresponding e-values from nucleotide and protein BLASTs.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Strand</th>
<th>Left End</th>
<th>Right End</th>
<th>Length</th>
<th>Nucleotide e-value</th>
<th>Protein e-value</th>
<th>Homology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminase small subunit</td>
<td>(+)</td>
<td>50</td>
<td>535</td>
<td>485</td>
<td>0</td>
<td>3.00 x 10^{-88}</td>
<td>HK022, HK97</td>
</tr>
<tr>
<td>Terminase large subunit</td>
<td>(+)</td>
<td>542</td>
<td>2056</td>
<td>1514</td>
<td>0</td>
<td>0</td>
<td>HK022, HK97</td>
</tr>
<tr>
<td>Head portal protein</td>
<td>(+)</td>
<td>2056</td>
<td>3330</td>
<td>1274</td>
<td>0</td>
<td>0</td>
<td>HK022, HK97</td>
</tr>
<tr>
<td>Head maturation protease</td>
<td>(+)</td>
<td>3348</td>
<td>4025</td>
<td>677</td>
<td>0</td>
<td>4.00 x 10^{-129}</td>
<td>HK022, HK97</td>
</tr>
<tr>
<td>Major head subunit precursor</td>
<td>(+)</td>
<td>4028</td>
<td>5185</td>
<td>1157</td>
<td>0</td>
<td>0</td>
<td>HK022, HK97</td>
</tr>
<tr>
<td>ORF1</td>
<td>(+)</td>
<td>5219</td>
<td>5545</td>
<td>326</td>
<td>3.00 x 10^{-168}</td>
<td>6.00 x 10^{-57}</td>
<td>HK022, HK97</td>
</tr>
<tr>
<td>Head-tail adaptor protein 1</td>
<td>(+)</td>
<td>5545</td>
<td>5781</td>
<td>236</td>
<td>2.00 x 10^{-118}</td>
<td>7.00 x 10^{-38}</td>
<td>HK022</td>
</tr>
<tr>
<td>ORF2</td>
<td>(+)</td>
<td>5778</td>
<td>5975</td>
<td>197</td>
<td>4.00 x 10^{-95}</td>
<td>4.00 x 10^{-28}</td>
<td>HK022</td>
</tr>
<tr>
<td>Head-tail adaptor protein 2</td>
<td>(+)</td>
<td>5977</td>
<td>6309</td>
<td>332</td>
<td>3.00 x 10^{-173}</td>
<td>4.00 x 10^{-58}</td>
<td>HK022</td>
</tr>
<tr>
<td>ORF3</td>
<td>(+)</td>
<td>6302</td>
<td>6841</td>
<td>539</td>
<td>1.00 x 10^{-99}</td>
<td>0</td>
<td>HK022</td>
</tr>
<tr>
<td>ORF4</td>
<td>(+)</td>
<td>6838</td>
<td>7203</td>
<td>365</td>
<td>0</td>
<td>9.00 x 10^{-65}</td>
<td>HK022</td>
</tr>
<tr>
<td>Major tail subunit</td>
<td>(+)</td>
<td>7258</td>
<td>7758</td>
<td>500</td>
<td>0</td>
<td>4.00 x 10^{-92}</td>
<td>HK022</td>
</tr>
<tr>
<td>ORF5</td>
<td>(+)</td>
<td>7797</td>
<td>8282</td>
<td>485</td>
<td>0</td>
<td>6.00 x 10^{-88}</td>
<td>HK022</td>
</tr>
<tr>
<td>ORF6</td>
<td>(+)</td>
<td>8288</td>
<td>8455</td>
<td>168</td>
<td>8.00 x 10^{-82}</td>
<td>2.00 x 10^{-25}</td>
<td>HK022</td>
</tr>
<tr>
<td>Tail length tape measure</td>
<td>(+)</td>
<td>8478</td>
<td>10898</td>
<td>2420</td>
<td>0</td>
<td>0</td>
<td>HK022</td>
</tr>
<tr>
<td>protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>minor tail protein</td>
<td>(+)</td>
<td>10898</td>
<td>11236</td>
<td>338</td>
<td>1.00 x 10^{-176}</td>
<td>2.00 x 10^{-60}</td>
<td>HK022</td>
</tr>
<tr>
<td>minor tail protein L</td>
<td>(+)</td>
<td>11233</td>
<td>11988</td>
<td>755</td>
<td>0</td>
<td>5.00 x 10^{-147}</td>
<td>HK022</td>
</tr>
<tr>
<td>tail assembly protein 1</td>
<td>(+)</td>
<td>11990</td>
<td>12700</td>
<td>710</td>
<td>0</td>
<td>2.00 x 10^{-135}</td>
<td>mostly HK022</td>
</tr>
<tr>
<td>ORF7</td>
<td>(+)</td>
<td>12730</td>
<td>13071</td>
<td>341</td>
<td>N/A</td>
<td>7.00 x 10^{-12}</td>
<td>Y. pestis phage lipoprotein</td>
</tr>
<tr>
<td>Gene</td>
<td>Strand</td>
<td>Left end</td>
<td>Right End</td>
<td>Length</td>
<td>Nucleotide e value</td>
<td>Protein e value</td>
<td>Homology</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>--------</td>
<td>--------------------</td>
<td>----------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>tail assembly protein 2 (+)</td>
<td>13115</td>
<td>13723</td>
<td>608</td>
<td>1.00 x 10^{-164}</td>
<td>3.00 x 10^{-78}</td>
<td>HK022 partial</td>
<td></td>
</tr>
<tr>
<td>host specificity protein (+)</td>
<td>13776</td>
<td>17231</td>
<td>3455</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>HK97 partial</td>
</tr>
<tr>
<td>ORF8 (+)</td>
<td>17233</td>
<td>17535</td>
<td>302</td>
<td>6.00 x 10^{-110}</td>
<td>1.00 x 10^{-47}</td>
<td>HK022</td>
<td></td>
</tr>
<tr>
<td>ORF9 (+)</td>
<td>17535</td>
<td>18209</td>
<td>674</td>
<td>2.00 x 10^{-122}</td>
<td>1.00 x 10^{-96}</td>
<td>HK022 partial</td>
<td></td>
</tr>
<tr>
<td>cor (-)</td>
<td>18288</td>
<td>18572</td>
<td>284</td>
<td>3.00 x 10^{-133}</td>
<td>9.00 x 10^{-45}</td>
<td>phi80</td>
<td></td>
</tr>
<tr>
<td>tail fiber (+)</td>
<td>18615</td>
<td>19988</td>
<td>1373</td>
<td>1.00 x 10^{-162}</td>
<td>4.00 x 10^{-77}</td>
<td>HK97, beginning only, partial to tail fiber</td>
<td></td>
</tr>
<tr>
<td>msgA (-)</td>
<td>20035</td>
<td>20307</td>
<td>272</td>
<td>N/A</td>
<td>3.00 x 10^{-32}</td>
<td>virulence protein</td>
<td></td>
</tr>
<tr>
<td>ORF10 (+)</td>
<td>20372</td>
<td>20626</td>
<td>254</td>
<td>N/A</td>
<td>6.00 x 10^{-27}</td>
<td>Enterobacter cancerogenus hypothetical</td>
<td></td>
</tr>
<tr>
<td>DNA pol V subunit (+)</td>
<td>20626</td>
<td>21015</td>
<td>389</td>
<td>N/A</td>
<td>7.00 x 10^{-50}</td>
<td>DNA polymerase subunit V</td>
<td></td>
</tr>
<tr>
<td>Integrase (+)</td>
<td>21329</td>
<td>22507</td>
<td>1178</td>
<td>0</td>
<td>0</td>
<td>stx2 converting phage 1717</td>
<td></td>
</tr>
<tr>
<td>AlpA (-)</td>
<td>22488</td>
<td>22679</td>
<td>192</td>
<td>4.00 x 10^{-95}</td>
<td>2.00 x 10^{-95}</td>
<td>E. coli</td>
<td></td>
</tr>
<tr>
<td>ORF11 (-)</td>
<td>22710</td>
<td>22877</td>
<td>167</td>
<td>2.00 x 10^{-78}</td>
<td>7.00 x 10^{-25}</td>
<td>Lambda</td>
<td></td>
</tr>
<tr>
<td>ORF12 (-)</td>
<td>22978</td>
<td>23817</td>
<td>839</td>
<td>N/A</td>
<td>N/A</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>ORF13 (-)</td>
<td>23817</td>
<td>24017</td>
<td>201</td>
<td>N/A</td>
<td>N/A</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>ORF14 (-)</td>
<td>24010</td>
<td>24324</td>
<td>315</td>
<td>N/A</td>
<td>N/A</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>ORF15 (-)</td>
<td>24321</td>
<td>25055</td>
<td>734</td>
<td>0</td>
<td>4.00 x 10^{-110}</td>
<td>partial HK620,</td>
<td></td>
</tr>
<tr>
<td>ORF16 (-)</td>
<td>25052</td>
<td>25534</td>
<td>482</td>
<td>2.00 x 10^{-151}</td>
<td>1.00 x 10^{-61}</td>
<td>HK620</td>
<td></td>
</tr>
<tr>
<td>ORF17 (-)</td>
<td>25531</td>
<td>25695</td>
<td>164</td>
<td>7.00 x 10^{-77}</td>
<td>8.00 x 10^{-23}</td>
<td>HK97</td>
<td></td>
</tr>
<tr>
<td>abc2 (-)</td>
<td>25706</td>
<td>25999</td>
<td>293</td>
<td>3.00 x 10^{-138}</td>
<td>2.00 x 10^{-49}</td>
<td>HK97, HK022</td>
<td></td>
</tr>
<tr>
<td>ssDNA binding protein (-)</td>
<td>26013</td>
<td>26528</td>
<td>515</td>
<td>0</td>
<td>8.00 x 10^{-79}</td>
<td>CP-1639 partial</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Strand</td>
<td>Left End</td>
<td>Right End</td>
<td>Length</td>
<td>Nucleotide e value</td>
<td>Protein e value</td>
<td>Homology</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>--------</td>
<td>--------------------</td>
<td>----------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>ds break repair protein</td>
<td>(-)</td>
<td>26529</td>
<td>27236</td>
<td>707</td>
<td>0</td>
<td>4.00 x 10^{-134}</td>
<td>Sf6</td>
</tr>
<tr>
<td>ORF18</td>
<td>(-)</td>
<td>27336</td>
<td>27494</td>
<td>158</td>
<td>7.00 x 10^{-77}</td>
<td>1.00 x 10^{-22}</td>
<td>Sf6</td>
</tr>
<tr>
<td>kil protein</td>
<td>(-)</td>
<td>27491</td>
<td>27643</td>
<td>152</td>
<td>7.00 x 10^{-62}</td>
<td>6.00 x 10^{-20}</td>
<td>Sf6, HK97 kil protein</td>
</tr>
<tr>
<td>cIII</td>
<td>(-)</td>
<td>27628</td>
<td>27759</td>
<td>131</td>
<td>6.00 x 10^{-62}</td>
<td>2.00 x 10^{-16}</td>
<td>Sf6</td>
</tr>
<tr>
<td>ORF19</td>
<td>(-)</td>
<td>27784</td>
<td>28752</td>
<td>968</td>
<td>0</td>
<td>5.00 x 10^{-167}</td>
<td>Sf6</td>
</tr>
<tr>
<td>ORF20</td>
<td>(-)</td>
<td>28920</td>
<td>29162</td>
<td>242</td>
<td>N/A</td>
<td>N/A</td>
<td>None</td>
</tr>
<tr>
<td>ORF21</td>
<td>(-)</td>
<td>29227</td>
<td>29455</td>
<td>383</td>
<td>0</td>
<td>8.00 x 10^{-69}</td>
<td>Stx2 converting phage 1</td>
</tr>
<tr>
<td>N</td>
<td>(-)</td>
<td>29455</td>
<td>29838</td>
<td>383</td>
<td>0</td>
<td>8.00 x 10^{-69}</td>
<td>Stx2 converting phage 1</td>
</tr>
<tr>
<td>ORF22</td>
<td>(-)</td>
<td>30409</td>
<td>30714</td>
<td>383</td>
<td>0</td>
<td>8.00 x 10^{-32}</td>
<td>None, kilA like protein</td>
</tr>
<tr>
<td>ORF23</td>
<td>(-)</td>
<td>30754</td>
<td>31095</td>
<td>341</td>
<td>N/A</td>
<td>N/A</td>
<td>None, kilA like protein</td>
</tr>
<tr>
<td>cI</td>
<td>(-)</td>
<td>31127</td>
<td>31840</td>
<td>713</td>
<td>0</td>
<td>4.00 x 10^{-135}</td>
<td>HK97</td>
</tr>
<tr>
<td>cro</td>
<td>(+)</td>
<td>31941</td>
<td>32141</td>
<td>200</td>
<td>2.00 x 10^{-88}</td>
<td>1.00 x 10^{-30}</td>
<td>Lambda</td>
</tr>
<tr>
<td>cII</td>
<td>(+)</td>
<td>32279</td>
<td>32575</td>
<td>296</td>
<td>3.00 x 10^{-143}</td>
<td>7.00 x 10^{-49}</td>
<td>HK022, HK97</td>
</tr>
<tr>
<td>ORF24</td>
<td>(+)</td>
<td>32608</td>
<td>32769</td>
<td>161</td>
<td>1.00 x 10^{-78}</td>
<td>3.00 x 10^{-22}</td>
<td>Sf6</td>
</tr>
<tr>
<td>O</td>
<td>(+)</td>
<td>32756</td>
<td>33577</td>
<td>821</td>
<td>0</td>
<td>3.00 x 10^{-160}</td>
<td>HK97</td>
</tr>
<tr>
<td>P</td>
<td>(+)</td>
<td>33574</td>
<td>34950</td>
<td>1376</td>
<td>0</td>
<td>0</td>
<td>HK97</td>
</tr>
<tr>
<td>ninB</td>
<td>(+)</td>
<td>35024</td>
<td>35464</td>
<td>440</td>
<td>0</td>
<td>7.00 x 10^{-79}</td>
<td>HK97, ninB</td>
</tr>
<tr>
<td>DNA N-6-adenine methyltransferase</td>
<td>(+)</td>
<td>35431</td>
<td>35988</td>
<td>557</td>
<td>0</td>
<td>4.00 x 10^{-103}</td>
<td>HK97, most/E. coli</td>
</tr>
<tr>
<td>ninE</td>
<td>(+)</td>
<td>35985</td>
<td>36167</td>
<td>183</td>
<td>4.00 x 10^{-85}</td>
<td>1.00 x 10^{-26}</td>
<td>HK620</td>
</tr>
<tr>
<td>ninF</td>
<td>(+)</td>
<td>36164</td>
<td>36334</td>
<td>171</td>
<td>4.00 x 10^{-80}</td>
<td>1.00 x 10^{-23}</td>
<td>HK620</td>
</tr>
<tr>
<td>Gene</td>
<td>Strand</td>
<td>Left end</td>
<td>Right End</td>
<td>Length</td>
<td>Nucleotide e value</td>
<td>Protein e value</td>
<td>Homology</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>----------</td>
<td>-----------</td>
<td>--------</td>
<td>--------------------</td>
<td>----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>roi</td>
<td>(+)</td>
<td>36327</td>
<td>37049</td>
<td>722</td>
<td>0</td>
<td>9.00 x 10^{-136}</td>
<td>HK620</td>
</tr>
<tr>
<td>ORF25</td>
<td>(+)</td>
<td>37049</td>
<td>37339</td>
<td>290</td>
<td>1.00 x 10^{-126}</td>
<td>5.00 x 10^{-48}</td>
<td>HK620</td>
</tr>
<tr>
<td>RusA</td>
<td>(+)</td>
<td>37336</td>
<td>37698</td>
<td>362</td>
<td>2.00 x 10^{-176}</td>
<td>5.00 x 10^{-62}</td>
<td>HK620</td>
</tr>
<tr>
<td>ninH</td>
<td>(+)</td>
<td>37695</td>
<td>37883</td>
<td>188</td>
<td>2.00 x 10^{-88}</td>
<td>3.00 x 10^{-28}</td>
<td>HK620</td>
</tr>
<tr>
<td>Q</td>
<td>(+)</td>
<td>37880</td>
<td>38503</td>
<td>623</td>
<td>0</td>
<td>1.00 x 10^{-118}</td>
<td>HK620</td>
</tr>
<tr>
<td>holin</td>
<td>(+)</td>
<td>38937</td>
<td>39620</td>
<td>683</td>
<td>1.00 x 10^{-166}</td>
<td>2.00 x 10^{-53}</td>
<td>HK620</td>
</tr>
<tr>
<td>lysin</td>
<td>(+)</td>
<td>39244</td>
<td>39720</td>
<td>476</td>
<td>0</td>
<td>1.00 x 10^{-88}</td>
<td>HK022</td>
</tr>
<tr>
<td>Rz</td>
<td>(+)</td>
<td>39717</td>
<td>40154</td>
<td>437</td>
<td>0</td>
<td>5.00 x 10^{-75}</td>
<td>HK97</td>
</tr>
<tr>
<td>Rz1</td>
<td>(+)</td>
<td>39916</td>
<td>40101</td>
<td>186</td>
<td>9.00 x 10^{-87}</td>
<td>3.00 x 10^{-26}</td>
<td>HK022</td>
</tr>
<tr>
<td>ORF26</td>
<td>(+)</td>
<td>40191</td>
<td>40466</td>
<td>275</td>
<td>N/A</td>
<td>N/A</td>
<td>None</td>
</tr>
<tr>
<td>HNH endonuclease</td>
<td>(+)</td>
<td>40846</td>
<td>41199</td>
<td>353</td>
<td>0</td>
<td>2.00 x 10^{-53}</td>
<td>HK022, mostly</td>
</tr>
<tr>
<td>ORF27</td>
<td>(+)</td>
<td>41199</td>
<td>41402</td>
<td>203</td>
<td>4.00 x 10^{-100}</td>
<td>N/A</td>
<td>HK022, HK97</td>
</tr>
</tbody>
</table>
Table 4. Predicted Rho-independent transcription terminators in the HK239 genome.

<table>
<thead>
<tr>
<th>DNA Sequence</th>
<th>Strand</th>
<th>Left end</th>
<th>Right end</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCCCGTAATAACGGGGCTTAATTTTT</td>
<td>(+)</td>
<td>3309</td>
<td>3335</td>
<td>27</td>
</tr>
<tr>
<td>GGGCGGGGAAACCCGCCCTTT</td>
<td>(+)</td>
<td>5192</td>
<td>5213</td>
<td>21</td>
</tr>
<tr>
<td>CCGGCCCTTGAGCCGGTTTTTT</td>
<td>(+)</td>
<td>7220</td>
<td>7241</td>
<td>22</td>
</tr>
<tr>
<td>AGGGCGGCAACGCCCTTAATTAATCAGGATT</td>
<td>(+)</td>
<td>7763</td>
<td>7792</td>
<td>29</td>
</tr>
<tr>
<td>CCCGCTTCGGCAGGGTTTTTTTT</td>
<td>(+)</td>
<td>13083</td>
<td>13103</td>
<td>20</td>
</tr>
<tr>
<td>GCCACCTTCGGGTGCTTTTTTTAT</td>
<td>(+)</td>
<td>13743</td>
<td>13766</td>
<td>23</td>
</tr>
<tr>
<td>CCGCCTGATGGCCGTTTTCTTTTT</td>
<td>(-)</td>
<td>22695</td>
<td>22717</td>
<td>22</td>
</tr>
<tr>
<td>ACTCGCTACGGCGAGTTTTTTGTTTT</td>
<td>(-)</td>
<td>22891</td>
<td>22914</td>
<td>23</td>
</tr>
<tr>
<td>TGCCCTCCAGTGAGGGCGATTTTTT</td>
<td>(-)</td>
<td>28765</td>
<td>28793</td>
<td>28</td>
</tr>
<tr>
<td>CCCGCCACAGAGCCGGTTTTTTCTTT</td>
<td>(-)</td>
<td>30302</td>
<td>30327</td>
<td>25</td>
</tr>
<tr>
<td>AGGCCCTGCTGTAATCGCAGGCTTTTTTAT</td>
<td>(+)</td>
<td>34976</td>
<td>35007</td>
<td>31</td>
</tr>
<tr>
<td>GCCCTGATTAATAATCTCGGGGCTTTTTTGCGTTTTT</td>
<td>(+)</td>
<td>38611</td>
<td>38645</td>
<td>34</td>
</tr>
</tbody>
</table>

Table 5. Additional genomic elements predicted by homology to other phages in the database.

<table>
<thead>
<tr>
<th>Element name</th>
<th>Strand</th>
<th>Left end</th>
<th>Right End</th>
<th>Length</th>
<th>Homology</th>
</tr>
</thead>
<tbody>
<tr>
<td>boxB left</td>
<td>(-)</td>
<td>30004</td>
<td>30018</td>
<td>14</td>
<td>HK106</td>
</tr>
<tr>
<td>boxA left</td>
<td>(-)</td>
<td>30062</td>
<td>30070</td>
<td>8</td>
<td>HK106</td>
</tr>
<tr>
<td>O_L1</td>
<td>(-)</td>
<td>30089</td>
<td>30105</td>
<td>16</td>
<td>HK106</td>
</tr>
<tr>
<td>P_L</td>
<td>(-)</td>
<td>30087</td>
<td>30115</td>
<td>28</td>
<td>HK106</td>
</tr>
<tr>
<td>O_L2</td>
<td>(-)</td>
<td>30113</td>
<td>30129</td>
<td>16</td>
<td>HK106</td>
</tr>
<tr>
<td>O_L3</td>
<td>(-)</td>
<td>30133</td>
<td>30149</td>
<td>16</td>
<td>HK106</td>
</tr>
<tr>
<td>P_RM</td>
<td>(-)</td>
<td>31847</td>
<td>31875</td>
<td>28</td>
<td>HK106</td>
</tr>
<tr>
<td>O_R3</td>
<td>(+)</td>
<td>31851</td>
<td>31867</td>
<td>16</td>
<td>HK106</td>
</tr>
<tr>
<td>O_R2</td>
<td>(+)</td>
<td>31874</td>
<td>31890</td>
<td>16</td>
<td>HK106</td>
</tr>
<tr>
<td>O_R1</td>
<td>(+)</td>
<td>31898</td>
<td>31914</td>
<td>16</td>
<td>HK106</td>
</tr>
<tr>
<td>P_R</td>
<td>(+)</td>
<td>31888</td>
<td>31916</td>
<td>28</td>
<td>HK106</td>
</tr>
<tr>
<td>boxA right</td>
<td>(+)</td>
<td>32149</td>
<td>32157</td>
<td>8</td>
<td>HK106</td>
</tr>
<tr>
<td>boxB right</td>
<td>(+)</td>
<td>32180</td>
<td>32194</td>
<td>14</td>
<td>HK106</td>
</tr>
<tr>
<td>P_RE</td>
<td>(+)</td>
<td>32270</td>
<td>32297</td>
<td>27</td>
<td>HK97</td>
</tr>
</tbody>
</table>
Verification of the genomic ends

Bacteriophage genomes are linear when they are packaged into the viral particle, meaning that they do have physical ends. The ends of HK239 were predicted based on homology to bacteriophage HK022. The ends were confirmed by designing two primers sets for each end and sequencing off the ends of HK239 genomic DNA. Sequencing analysis showed good quality sequence until the end of the genome was reached, at which point the sequence abruptly stopped. In some instances, the ends are cohesive and would not be detected by sequencing off the end of a linear genome. This is true in the case of a 3’ overhang. To address this, a PCR product, amplified across the predicted ends of ligated HK239 genomic DNA, was sequenced. This revealed additional bases not seen in the original genomic sequence. These additional bases represent the cos site, which is 10 bp in length (Figure 6).

Figure 6. Verification of the ends of HK239 and identification of the cos site. Sequences from top to bottom: consensus sequence, sequence from left end, sequence from right end, sequence of PCR products generated from ligated genomic DNA, and HK239 genomic sequence. The highlighted sequence on the PCR product and the HK239 genomic sequence represents the cos site [41].
Bioinformatics Analysis

Once the genome of HK239 had been completed, it was compared to other phage genomes. It is known that lambdoid phages are genetic mosaics of one another [34], and this also holds true for HK239. Among the information revealed by HK239’s mosaicism, two pieces were particularly important: HK239’s potential homoimmunity to HK106 (accession number EF120461) and HK239’s high degree of homology to HK022 (accession number AF069308).

The HK239 and HK106 immunity regions are highly homologous (Figure 7A). The N, cl, cro, and cII genes, the operator binding sites, the nut sites, and the pR, pL, pRM, and pRE promoters are all very similar to each other with only a few mismatches. Only the genes directly downstream of cl, ORFs 22 and 23 in HK239 and hicA, hicB, and a hypothetical protein in HK106, show no homology to each other. This can be seen in the dotplot (Figure 7A), a nucleotide by nucleotide comparison of the immunity regions. The diagonal lines indicate areas of homology. These data led to further experiments to explore the potential homoimmunity of the two phage, which is discussed in the next section.

Genome wide comparisons led to the discovery of a high degree of homology between HK239 and HK022. This can be seen in Figure 8, a phamerator analysis of the HK239 and HK022 genomes. Like a dotplot, this analysis also shows homology between nucleotide sequences; however the phamerator shows additional details such as the relative location of genes within the sequence. The homology is represented by purple shading and the white indicates no homology. From the phamerator analysis, it can be
seen that the head and tail genes share a high degree of homology. There is also some homology between the host specificity genes and the lysis genes as well.

A.

Figure 7. A. Dot plot of HK239 and HK106 immunity regions (HK106 accession number: EF120461) [41]. Regions of homology are indicated by the diagonal line. B. Immunity test: Plating of HK239 on RK898 (left) and an HK106 lysogen (right). A 5 μL aliquot of each serial dilution ranging from $10^{-1}$ to $10^{-12}$ were spotted onto laws of 898 (left) and an HK106 lysogen (right). The order of the dilutions started from the top left of the plate and moved right.
Figure 8. Phamerator output of HK022 (accession number AF069308) and HK239 genomes [42]. The purple shading and other colors indicate homology. The absence of shading indicates no homology.

**Generation of an HK106 lysogen**

Since the sequence analysis showed that bacteriophages HK106 and HK239 shared significant similarity in their immunity regions (Tables 3 and 5, Figure 8), the potential for homoimmunity was explored. We knew that HK106 could form lysogens whereas HK239 could only be grown lytically. The immunity region contains the genetic information that controls the lysis/lysogeny decision.

We took advantage of the similarity of the HK239 and HK106 immunity regions to investigate the reason for the HK239 clear plaque phenotype. An HK106 lysogen was generated and confirmed by PCR with HK106 specific primers (Figure 9). Three out of the 4 colonies obtained were positive for an HK106 prophage. Dilutions of HK239 were spotted on a lawn of an HK106 lysogen. Clearing was seen only at the lowest dilution, suggesting that HK239 is homoimmune to HK106 or alternatively, HK106 lysogens can
exclude HK239 (Figure 7C). Of the genes and genetic elements in the immunity region, 
cI was studied first because there are a number of sequence differences between HK106 
and HK239 (Figure 10A). However, in the protein alignment (Figure 10B), there are 
only three amino acid differences, one of which is a non-conservative change: from 
glycine (HK106) to aspartic acid (HK239). It is not known if these differences affect the 
functionality of the repressor and contribute to the clear plaque phenotype. The HK106 
cI gene was cloned into an expression vector (a pBAD18 plasmid) and it was 
demonstrated that HK239 cannot form plaques on strains that carry this plasmid. This 
result confirms that HK239 and HK106 are hommoimmune (data not shown).

![Image of gel electrophoresis](image_url)

**Figure 9.** Confirmation of HK106 lysogen. Purified suspected lysogens were screened 
for the presence of the HK106 prophage by PCR with primers 424 and 469. Lane 1 
contains a 100 bp ladder. Lanes 2-4, show an expected band of 287 bp. Lane 5 did not 
contain any product, indicating the colony did not contain a HK106 prophage. Lane 6 is 
a positive control amplified from an HK106 lysate. Lane 7 is a negative control. The 
numbers in white indicate the size of the marker bands flanking the product.
Figure 10 (previous page). Nucleotide (A) and protein (B) alignments of the HK106 and HK239 cI genes [40]. The green shading indicates homology. Gaps indicate differences in the sequences.

Head and tail measurements

HK239 particles were visualized on the TEM by negative staining (Figure 11). Sixty-one phage were measured using the Syncroscopy software. The head and tail measurements were 49.7 nm and 133.2 nm respectively (Figure 12).

Figure 11. Electron micrographs of Bacteriophage HK239. The phage were deposited on formvar coated grids and stained with 1% uranyl acetate. The scale bar is approximately 100 nm.
Figure 12. Head and tail measurements of 61 HK239 bacteriophage. The average head length, as indicated by the black line, is 49.7 nm. The average tail length, as indicated by the black line, is 133.2 nm.
Discussion

The sequencing and annotation of bacteriophage HK239 has yielded new insights. Many of the genes have been identified based upon matches in the database. Some open reading frames that have been assigned a function are described below.

The HK239 immunity region

When annotating the genome, it became apparent that the HK239 immunity region is highly homologous to regions of the HK106 immunity region (Figure 7A). In the HK239 and HK106 immunity regions (Figure 7B), the N (an antiterminator), cI, cII, and cro genes are highly homologous. The right operator binding sites are identical and the left have only a couple mismatches. The boxA and boxB sites in both operons are identical. The left and right operon promoters, PL and PR, are identical. The promoters that drive repressor synthesis, PRM and PRE, contain only a few mismatches. The significance of these differences is unknown. To discover if mutations in the left operator binding sites or cI were responsible for the clear plaque phenotype, HK239 was plated on a lawn of an HK106 lysogen (Figure 7C). Only at the lowest dilution did HK239 plate on the lysogen and that was probably due to the fact that the bacteria were overwhelmed with infecting phage resulting in killing. These data present two possibilities: 1) the repressor of HK106 recognizes the operator binding sites of HK239 or 2) HK106 is capable of excluding HK239. It is known from sequence alignments (Figure 10) that there are only three amino acid differences between the two proteins. One of these is a non-conservative change: glycine (HK106) to aspartic acid (HK239). However, it is not known if these differences are responsible for the clear plaque phenotype. The HK106 cI gene was cloned into an expression vector and it was
demonstrated that HK239 cannot form plaques on cells that contain the construct. This confirms that the two phage are homooimmune and suggests that it is the HK239 repressor that is defective or not expressed (data not shown).

**Genes expressed in a prophage**

In addition to the repressor, a limited set of genes can be expressed in a prophage. An example of this includes genes encoding exclusion functions. One known exclusion gene in HK239 is the φ80 *cor* homolog [37]. In φ80, the *cor* gene product excludes HK022 and other phage that inject their genome by attaching to the FhuA receptor [15]. In addition to exclusion mechanisms, virulence factors are also expressed from prophage. The annotation of HK239 revealed a homolog of *msgA* from *Salmonella typhimurium*. In *S. typhimurium*, this gene encodes a factor that enhances the organism’s ability to survive within macrophage [44]. Although there is strong homology, it is not known if this gene is expressed or has any role in pathogenesis in *E. coli*.

No additional exclusion genes were identified in sequencing the HK239 genome. This does not mean that they are not present, but suggests that a gene with a currently unknown function may encode a novel exclusion mechanism. Candidates for this include the genes directly downstream of *cI*. Since HK106 and HK239 are homooimmune, it may be possible to generate a recombinant of HK239 that is capable of forming lysogens. With a lysogen, it would be possible to examine the broad exclusion phenotype of HK239 and test candidate exclusion genes by deleting them and looking for a loss of the phenotype.
Relatedness to other phage

It has been observed among phage families, particularly lambdoid phage, that phages are genetic mosaics of each other [45]. This is true for HK239 and this mosaicism is apparent in Figure 8. The head and tail genes, Q, and the lysis genes are all highly homologous to one another as indicated by the purple shading. In fact, at many of these regions of homology, the homology ends at the end of the reading frame, the host specificity gene being an obvious exception (Figure 8). It has been proposed that most of this homology occurs due to recombination at gene boundaries because recombination within genes is deleterious to the phage [46]. The head and tail genes are more often than not transferred as one unit because the gene products must be able to interact with one another [46]. This is true in the case of HK239 and HK022. The head and tail genes are nearly identical and this similarity is reflected in the TEM measurements. HK239 has a tail 133.2 nm in length and a head 49.7 nm in diameter on average. The head and tail measurement for HK022 are 135 nm and 55 nm respectively [4]. The physical dimensions of these phages (head diameter and tail length) are consistent with other members of the family *Siphoviridae*. Members of this family are characterized by a double stranded DNA genome, icosahedral heads, and non-contractile tails [47].

The mosaicism in HK239 indicates that it has a shared ancestry with many other phages (Table 13). However, it has the highest degree of homology to HK022. This is evident in figure 8 where more than half the HK239 genome shows homology (as seen in the purple shading) to HK022. Also, more than a third of the open reading frames, 26 out of 71, have partial or full homology to HK022. No other phage genome in the database shows this degree of homology.
Conclusions

The genomic sequence of HK239 has revealed that it has a similar genetic make-up to other lambdoid phage, it shares homology in the immunity region with HK106, and is mostly closely related to HK022. HK106 readily forms lysogens but HK239 cannot. Our experiments show that a HK239 cannot grow on cells that express the HK106 cI confirming that the two phage are homoimmune. It may be possible to generate an HK239 recombinant that can form lysogens. Such a lysogen would allow closer examination of HK239’s exclusion phenotype. The combination of the availability of lysogen and the genomic sequence will make it possible to delete candidate exclusion genes and look for the inability to exclude specific phages. While the genomic sequence has not directly answered questions about HK239’s wide range of exclusion, it has made possible other avenues for exploring the exclusion range.
Bibliography


42. **Steve Cresawn.** James Madison University. Personal communication.


Appendix I. The HK239 GenBank file.

LOCUS     HK239_genome 41538 bp DNA linear 14-MAY-2010
DEFINITION Bacteriophage_HK239.
ACCESSION
VERSION
KEYWORDS
SOURCE      Bacteriophage_HK239
ORGANISM  Bacteriophage_HK239
          Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
          Lambda-like viruses.
REFERENCE   1 (bases 1 to 41538)
          Wright, A.A., King, R.A. and Christie, G.E.
          Submitted (14-MAY-2010) Department of Biology, Western Kentucky
          University, 1906 College Heights Blvd, Bowling Green, KY 42101, USA
FEATURES             Location/Qualifiers
  source          1..41538
                  /organism="Bacteriophage_HK239"
                  /mol_type="genomic DNA"
  gene            50..535
                  /gene="terminase small subunit"
  CDS             50..535
                  /gene="terminase small subunit"
                  /codon_start=1
                  /transl_table=11
                  /product="terminase small subunit"
                  /translation="MADKRIRSDSSAAAVQAMKNAAVDTIDPPSHAGLEKKAEPFWHDN
IRSKALDSWTPADLLAAVELANNNQLYITVLRLKDLRKEERIRGEERDEGLIKDLRK
QIVELQRTILAQRRLQIHSATNGESRDQKKRNQNDRARNTKNEHQDQDDNL
IAFPKHG"
  gene            542..2056
                  /gene="terminase large subunit"
  CDS             542..2056
                  /gene="terminase large subunit"
                  /codon_start=1
                  /transl_table=11
                  /product="terminase large subunit"
AMKSVGDGIEETIRDVENWLRSVGLTKSQAVGLIARFKSAIRSESEGDGNEAQINA
LLQSIKSFPSNLGK"

<table>
<thead>
<tr>
<th>Type</th>
<th>Start</th>
<th>End</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene</td>
<td>4028</td>
<td>5185</td>
<td>major head subunit precursor</td>
</tr>
<tr>
<td>CDS</td>
<td>4028</td>
<td>5185</td>
<td>major head subunit precursor</td>
</tr>
<tr>
<td>terminator</td>
<td>5192</td>
<td>5213</td>
<td>terminator</td>
</tr>
<tr>
<td>gene</td>
<td>5219</td>
<td>5545</td>
<td>ORF1</td>
</tr>
<tr>
<td>CDS</td>
<td>5219</td>
<td>5545</td>
<td>ORF1</td>
</tr>
</tbody>
</table>

/translation="MSELALIQKAIIEESQQKMTQLFDAAQKAEIESTGQVSKQLQSDLMKV
QEELTKSGTRLDFLEQKLSGAENGPEEKKSFSERAASEEELIKSWDGKQGTFGAKTF
NKSLGSDADSAGSLIQPMIPGILIIMPGLRRLTIRDLAOGRTSSNALEYVREEVFT
NNADVVEAEKALKPESDITFSKQTANVKTIAHVQASRQVMDDAPMLQSYINNR
LMYGLALKKEEGQLNGDGTGDNLGLNKVATAYDTSLNATGDTRADIAAIYQ
VTESEFASGILNPRDWHNILKDNENGYIFGGPQAFTSINMWGLPVVPTKAQQ
AAGTFTVGGFDMASQVWDRMDATVEVSREDRNFKNMLTILCEERLALAHY
RPTAIKGTSSGS"

/standard_name="terminator"

<table>
<thead>
<tr>
<th>Type</th>
<th>Start</th>
<th>End</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene</td>
<td>5545</td>
<td>5781</td>
<td>head-tail adaptor protein 1</td>
</tr>
<tr>
<td>CDS</td>
<td>5545</td>
<td>5781</td>
<td>head-tail adaptor protein 1</td>
</tr>
</tbody>
</table>

/translation="MAIDVLDVISLFLFKQQIEFEEDDRDELITLYAQAADFYMCRWCDEPAWKVAADIPAAVKGAVVLFADMFEHRTAQSEVQLEYENAAERMMFIHRNWRGKAEEEGS"

<table>
<thead>
<tr>
<th>Type</th>
<th>Start</th>
<th>End</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene</td>
<td>5545</td>
<td>5781</td>
<td>head-tail adaptor protein 1</td>
</tr>
<tr>
<td>CDS</td>
<td>5545</td>
<td>5781</td>
<td>head-tail adaptor protein 1</td>
</tr>
</tbody>
</table>
/translation="MEPGRFRHRVKILTFTSRDPGSQPVESWTGGNPVPAEVKRSGREQLSGGAEYAT квартирные рывердеккергквкнсгк"  
gene  5778..5975  
/CDS  5778..5975  
/gene="ORF2"  
/NOTE="Similar to Bacteriophage HK022 gp8"  
/codon_start=1  
/transl_table=11  
/product="unknown"

/translation="MNEQQLVNEIAVIREQITAQREQTEISAESN MALCDVIIQSLAGE LDEEAQQTYLSGKPRG"  
gene  5977..6309  
/CDS  5977..6309  
/gene="head-tail adaptor protein 2"  
/codon_start=1  
/transl_table=11  
/product="head-tail adaptor protein 2"

/translation="MQAGKLRHRITLQEPVKVQNTTGAVINTWRDVATVRAEVSPLSA REFIAAAQASQGEITTRIVIRYRAGVTRKHRILFRGAVYNIHGVLPDPKSGREYLT LPCSEGVNDG"  
gene  6302..6841  
/CDS  6302..6841  
/gene="ORF3"  
/NOTE="Similar to Bacteriophage HK022 gp10"  
codon_start=1  
/transl_table=11  
/product="unknown"

/translation="MADSVEVSLTGLESLLGKMEAVSEVTRNKAGRFA LRKAANIIIRDRA RSNASRVDDPLTKEIAKH NIVASFSSKQF RRTGDRLAFRVGVMG GARQ YANTKAN VRKGRA GTKYKTSGDKGNPGGTDWY WRFLEFGTEHTSARPILRPM NG VND VINVFSTEMGKAIDRAIRLAMKKGT TA"  
gene  6838..7203  
/gene="ORF4"
CDS 6838..7203
/gene="ORF4"
/note="Similar to Bacteriophage HK022 gp11"
/codon_start=1
/transl_table=11
/product="unknown"
/translation="MIAPIFSVCASSPEVTDLLGSNPVRIYPFGIQQDDNVVYPYVVWQNITGSPENYIAQRPDADFTQLQVDAYADTVDEVIAVATALRDAIEPHAHITRWGGQERSPETKRYRYSFDVDWIVTR"

terminator 7220..7241
/standard_name="terminator"

gene 7258..7758
/gene="major tail subunit"
CDS 7258..7758
/gene="major tail subunit"
/codon_start=1
/transl_table=11
/product="major tail subunit"
/translation="MSVLTQGTQLFVLVKGVSEVECITAFSPGSNPAQPQIEDTCLSERFDRSYKRGLRTPGTASLTLNADPKNTSHIMLYNLSISDDEQDLTFAIGWSDGTASPTAAENGAEGGAVDGLVLPDSRTWFVFVKGYVSDFPFDFSAVTVSTSTASIQRSGAVWVVKVT""

terminator 7764..785
/standard_name="terminator"

gene 7797..8282
/gene="ORF5"
CDS 7797..8282
/gene="ORF5"
/note="Similar to Bacteriophage HK022 gp13"
/codon_start=1
/transl_table=11
/product="unknown"
/translation="MKLTLDAKESGAFTGRPEKEITWTQTGDKKATATVYVVRPMGYHTATSDVLAFGKVDGVAAGRIAASICDEHGKPIFTPADITGEADPERGALDGGLTVALLAIPQEVNLGKTSSAPKTNSGASSSTSEASEAEQSPKRGKSSHISKSPNSQSTGNTGTV""

gene 8288..>8458
/gene="ORF6"
CDS  8288..8458
/gene="ORF6"
/note="Similar to Bacteriophage HK022 gp14"
/codon_start=1
/transl_table=11
/product="unknown"

/translation="MLRVEWGAGLVSSMIANVRDPKRPPFNPTDFTLHFTKVKAADGPI SLEEARASWT"

/gene  8478..10898
/gene="tail length tape measure protein"
CDS  8478..10898
/gene="tail length tape measure protein"
/codon_start=1
/transl_table=11
/product="tail length tape measure protein"

/translation="MASKSLGTLTIDLIAKVGGFVSGLSQAERASQKWRKQVKEDAAAA AAAMTGFAVAGAAIGAGVAGVAGYNLLLKTTSRQITEDESREDWAKSLNMSTQSLAW QYAAEKAAGVSGQDAMIFKDVGDVAKILDAVLNKGSEAVGALDGLSAKLAG ESPDKQQLLAISDALEKVSKNAEKTILESGLNDLSKMLPLLNGSEKLRQYMDAA KKFGVAPDDADIEKLVRVNALFEDMETQVNGVKIELAAGLASSVDLSGLQKSIGD MGDVFKDPAVIQGLTLDVGVDLAVLVRVGAEAGKLDQYKGGSSVGLNAS IPEIERRIKNLNADLDDKGLASFNRIQMIDVSGLKEAERELQQRLARFLRNSQSTLP KIKLPEPVKTYTLGAGGETNGKPQKNTSGQKLDSAFSAERSYMRIELIDTTGK KTAVVTEQQKLQFDIADGKILQGLNETQKKRLASLAQEVDRLNAVKKANEANAK VAAFVANLQEQNENARADLDGVIDQAGGLDKQRERLRLRELSIERSYLDQQRDLQ KQYQSGDISQTQVDRETQALKDAQAERLGQEDYYSQIDALQSDWVTGARDQL ADWVDDSTNYATLADAMKSLSGISSNIVDMLNGKNKASWKDVGVVLKIEQ VMVNMMIANAASSISLFAGAAASSASSGTAIQSYGALQFNAKGGVYSSADLS QYSNSVSSPTLAFAKGAGLMGEAGPEAIMPLTRAADGSLGVRAMGISGLTPG GSSAPQVSQIDGNGNTQTQASGGYEQFREVSFVDRYRELIGRDLSPGGA VWNLAKGGR"

/gene  10898..11236
/gene="minor tail protein"
CDS  10898..11236
/gene="minor tail protein"
/codon_start=1
/transl_table=11
WQEESQKYGSVFIFIQPYDGVIDTVFTQSQLEMVGLDISDGITKIEFRHQGGLFYRINSR1K1V1ER1C1

| terminator | 13083..13103 |
| standard_name | "terminator" |
| gene | 13115..13723 |
| /gene | "tail assembly protein 2" |
| CDS | 13115..13723 |
| /gene | "tail assembly protein 2" |
| /codon_start | 1 |
| /transl_table | 11 |
| /product | "tail assembly protein 2" |

/translation="MKEVMTTIQLLGVGLGKTFGRTHQR1ARTGEAAIALSKTLPGFESFMISSKRRGLTFAVFKGKRNIAADEMPFSEGDDVVRIMPVIIGSKRAGLLQITILGAVLITAATGPGIGAFAAAAGGLTGFAAATGASLVGLGGVIQLLSPQPSGIAKKSADNRAASYAFGGVTNTAAQYVPVPLLYGKRRIGGAISAGIYVEDQQ"

| terminator | 13743..13766 |
| standard_name | "terminator" |
| gene | 13776..17231 |
| /gene | "host specificity protein" |
| CDS | 13776..17231 |
| /gene | "host specificity protein" |
| /codon_start | 1 |
| /transl_table | 11 |
| /product | "host specificity protein" |

/translation="MATDKVLKGRKGS5SSRTPTEQPDQDLQSVAKAKILVALGEGFAGQLTGDITYLDTGALENADGQNSFSGVTFWEFRAGTFAQKYIQGIPG1NTEISVGETVSSATAWTRTFTNTQLSARFRLKWSLFPQEDDDGDLVGSVNYAIDLQTDGTWQTTLNTSVDKGTTGYSIERSH1DLPQAGSTWTIRLKR1TSDANSAK1GDTMMLQSFTEVIDAKLRYPNTHLYVEFDQQSGQ1SCEPRGVR1VPD1Y1PETR1YSGTW1TAFKQDAWT1DNP1WA1FYDV1SDRF1G1HL1TAANIDKW1TLYQV1AYC1DQMVPD1KG1NGT1EPRY1CN1YQ1QR1DAY1TV1LRF1FA1IFR1GMT1YWG1GDQ1VALADM1PRD1DV1SY1TRAN1V1G1GR1FTY1SS1TK1SY1T1ALV1SW1D1PNG1N1AY1AD1AME1PVFEQALVARYG1FN1Q1LEM1TA1GCTR1QSEAR1KGR1W1G1LT1N1K1R1D1V1V1S1F1VD1V1LD1G1N1I1P1P1GY1IA1VA1DELL1SG1KV1MG1GR1SA1V1N1G1R1V1KL1D1R1V1A1DA1AP1G1D1RL1L1N1L1PS1G1AS1Q1SRT1QAV1N1G1E1ST1VT1T1AY1SETP1Q1A1E1V1W1VES1DE1L1AQ1Y1R1V1V1S1V1SD1NN1D1G1T1F1S1IT1GA1W1HDP1D1KY1AC1T1G1A1I1D1Q1R1PV1SP1VP1PG1N1Q1S1P1AN1IV1S1S1F1V1V1Q1N1S1V1ET1M1R1V1SW1D1QA1Q1N1AI1AY1EA1Q1W1RR1N1D1GN1W1V1N1P1R1T1T1S1F1D1V1PG1I1Y1A1GR1LY1V1R1V1RAIN1A1E1ISS1GW1GY1SEE1K1T1LT1K1GV1GN1P1P1PV1G1F1A1SE1V1V1F1G1E1L1N1W1G1F1P1A1N1T1D1TL1K1TE1I
QYSLTGSEDADAILSDVPQRPKYYQMQMLAGGQFWYRAQLVDRGTGNEGYTD
WVRGQASIDVSDITDIVEDIKESDTFKELIESAVDSNEKIAGMADDIRQNAADL
EQQALAIKENADGLAQAEVKIDEISVMDGMTGKVKNSSIAVQNSLAQVTSRRSQ
TATNAGNSASIDRIDIADTSAVARALVTLDAAGGNVSNATDLTLTETLADFTQ
ASATKINSLTVTNGQTAAINQTAAVADVNGNLSAMYNNKVGVSSNGQYYAA
GMGIGVENTPSGMQSQVIIFLADRFAVTTMVGGTVTLFPVIQNGQAIIRDVTVG
DGTISNAKIGNYIQSNNYVAGVSGWKLDSGTFENYGSTAGEGAMKQTNTISVK
DNNVLRVQFRGLTGPF

gene  17233..17535
  /gene="ORF8"
CDS  17233..17535
  /gene="ORF8"
  /note="Similar to Bacteriophage HK022 gp24"
  /codon_start=1
  /transl_table=11
  /product="unknown"

/translation="MAYGIQTWDSAGKPNNYGIKPVSVVGRIGQQAAGQNSGSWSFTVPS
GMKVFALSLEDGGSVSIVAGRNTIVTAASSVGLGNYPSKCEVVVFMEKA"

gene  17535..18209
  /gene="ORF9"
CDS  17535..18209
  /gene="ORF9"
  /note="Similar to Bacteriophage HK022 gp25"
  /codon_start=1
  /transl_table=11
  /product="unknown"

/translation="MAEFGAMILMDNGNPVTFQSTPFCLYGKYTFNSSANGSSOQVAQ
NIALNADYPVMVFIKTTNTAQPPTPVMYSRNGGNVYAVGVPYNSFTLTHSYVFA
IFPQIPLKWGLAIWDASGKLVTNERSVLSDLQVTGTPGANGGINIDQTLGSWSA
VAPAQLGQTIIVNNSTKPTIYITINAYSSCRFDGANTRINAGGTSTGAGSPPGGLT
TGISLTAINTAAYD"

gene  complement(18288..18572)
  /gene="cor"
CDS  complement(18288..18572)
  /gene="cor"
  /codon_start=1
  /transl_table=11
/product="cor"

/translation="MFKRSGLFAGCFKCIERTYPAPAERRISSPVLRLVVLFTHTVNLNRSFLTANDCSAAANWFLFLNYASASCKHDSACHIADNQFSFHDSDIQRNSVS"
gene 18615..19988
/gene="tail fiber"
CDS 18615..19988
/gene="tail fiber"
/codon_start=1
/transl_table=11
/product="tail fiber"

/translation="MLYNTGTIAINGNTATGKGTNWTPASQVRAGQTIIVMSNPVQLFQISSVNSATMTVTPAASPALSQKYAILVSDISVDGLAQAMSQLIKEYDENIAGWETFATTSANQTITVTINGTTTIPGIGKLAQKGSNGAVTVADGGTGATTAEGSRLNGLGNSATRNVTGAAGTVAGDDSFEKIAKLGAATKTDTGEGNVLITGSFGVGSKILPVISDIWDKQGSRFCNVTPATLGPGIYGSGIRLDNIGSGATYAAQQSFALIFSGKVIQFMGMTDGVDTGWMKIYHTGNTTRASDGTLKAASPIVRLFGNGKCQLNDESEGCTVTRLATGKYLVEGCEGLNSDAWGGIDGGFDIPTDRNKPWILWDYEVNADGVLLIQTYHRTFPSAPAYARNERDGINDGEPIIDIPSDFVSFVSRVEMPANSIWNQKQKAIEEAAKSAEEVQ"
gene complement(20035..20307)
/gene="msgA"
CDS complement(20035..20307)
/gene="msgA"
/codon_start=1
/transl_table=11
/product="msgA"

/translation="MVMFVELVYDKRNVEGLKGAREIILAELKKRVHQIFPDAEVKVKPMQANGLNSDASKSDREKLNRMLEEMFEESDMWLVALVEFTVLQNGV"
gene 20372..20626
/gene="ORF10"
CDS 20372..20626
/gene="ORF10"
/codon_start=1
/transl_table=11
/product="unknown"
/translation="MRAILMPRKSIDIHAAFLASIEQNQKGYLCLNTNKFINKLREKSWHFSQADANTWIERYQPDFADKTTNGSQRNYWILRNMGRTVF"
gene 20626..21015
CDS 20626..21015
/gene="DNA pol V subunit"
/codon_start=1
/transl_table=11
/product="DNA pol V subunit"

/translation="MGFPSPATDYVEQRISLDERIITRPAAATYFMRAGATHYREGILNGALLVVDASMSPCDGSLLVCTDSGEFRIKRYRTHPRPHLENLENGKRSSLLLKDEVSDTSRPVFGVITYIINDARSGEFDYPLK"
gene 21329..22507
CDS 21329..22507
/gene="integrase"
/codon_start=1
/transl_table=11
/product="integrase"

/translation="MAISDTKLRTIYGKPYSGPQEVADADGLSVRISPKGVIQFQYRYRWHGKPNRLGLGRYPSLSDLKDRQITADLRNLYFSGTDPRTYFEEKVENSMITVAQCLDYWFDNYVSTTLREKTKALRYRSTVMKRMHDAPNPASSTVQKWVVLTEEERNPRRAQVRQLSRLSAISWCRRQLIDSCAIMSIQPRDFSGRSAEVDVRVLSYHELAKIWLAIERSRASSTSNKLHLHQLMLWGLARILSELRLAKKTEFDLLENVWTVPKEHSMKMGNRVRIPFQIKPFLEKAMTTYNDVLFPGEDIKPIIIAAANRFVRNIRGGMDLGYWRTHDFRRXLVRLSEMVNPHEVTERMLGHELGGIMSVYNKHDWIEAQRKAYELHADKLFWHIRSISD"
gene complement(22488..22679)
CDS complement(22488..22679)
/gene="AlpA1"
/codon_start=1
/transl_table=11
/product="AlpA1"

/translation="MTDTSLIPKEKEVNNKLGVSSRQTIWNYTKRHGFKPKVRTHPKSYLREAPEGWILNGGVNQKCS"
terminator complement(22695..22717)
  /standard_name="terminator"
gene complement(22710..22877)
  /gene="ORF11"
CDS complement(22710..22877)
  /gene="ORF11"
  /note="Similar to Bacteriophage Lambda gp35"
  /codon_start=1
  /transl_table=11
  /product="unknown"

/translation="MHFRVTGEWNGEPNRVIEAENINDCYDHMIWAQIAHADVNTIRIEELKEHQAA"
terminator complement(22891..22914)
  /standard_name="terminator"
gene complement(22978..23817)
  /gene="ORF12"
CDS complement(22978..23817)
  /gene="ORF12"
  /codon_start=1
  /transl_table=11
  /product="unknown"

/translation="MELSRLDKIKEINDDWLMDDCEGNKNCDDVVKEMARIALASLEAEPVAVLLSGGGAKNNVSFDSSGNAYADPLREVTPLYTAPPAPVSIPAAMEIDDDDFS
AFEHGKAVGVNARASMQLAEPVSNSDELPLDYLQGHKDGLEWAAQLAEANHPQTGDWLYDDQIDLARAIKGPDMPTQGGNPVIPDGPLWSICEQMPVIGELNW
RTSFPLLVTCEIGVIPAYYGVFSVNGDRHYGFMEKSLKYGDDNGNHPQTNEYGLISNVTHWMLPPEPPQEVNRG"
gene complement(23817..24017)
  /gene="ORF13"
CDS complement(23817..24017)
  /gene="ORF13"
  /codon_start=1
  /transl_table=11
  /product="unknown"

/translation="MANSLLETCNWQILRAELARNPNMAMTLQKLDIMIEHSVRAAIE
SHRVDWFSEAERKAIVN"
gene complement(24010..24324)
CDS complement(24010..24324) /gene="ORF14"
/codon_start=1
/transl_table=11
/product="unknown"

/gene="ORF14"
/translation="MSTITREWŁQQAINDYESVRDELPFGŁDYQGNILAALRIALASLEA AEKRIAELAKLDSADKLQDSAHRGLQHGFSFGQTDNQAGFEECLSLAYGARGK DNG"

gene complement(24321..25055) /gene="ORF15"
/CDS complement(24321..25055) /gene="ORF15"
/codon_start=1
/transl_table=11
/product="unknown"

/gene="ORF15"
/translation="MSQIDYQLRAKAÉKATCGEWSLEYGDGRFDGDALIHREAAGYI PICRIEGAHPESSFDEDFQMEQQANAEFIAASNPATVLALLDERERNQQYIKRRD QENEMIALTVVAKLRVELEGKDSKIANTLAERDALREGEMGDARHNSTRAADITY FQOLVECEIPAGSGVEYVDDMREKLEAEKRIAELREVVLPQCYSMLHARVD FDEPYHTEMVYRQHVLEALHNAGINVTEACKGEAS"

gene complement(25052..25534) /gene="ORF16"
/CDS complement(25052..25534) /gene="ORF16"
>Note="Similar to Bacteriophage HK620 hkaH"
/codon_start=1
/transl_table=11
/product="unknown"

/gene="ORF16"
/translation="MKQMSLIEDMGFLKGCIPRLKVENETNAEYLVRKFGGELESKLETA LRECRASEITIDNLAEKCAALDAKKEKFAVECAATKIAAYLKSGRQDFSLNTTA TDAFLAVRAQGVEMYADNLDNGADDAERGGFDYAVRFLRSEASGVRLFADQ LRKGGSQ"

gene complement(25531..25695) /gene="ORF17"
/CDS complement(25531..25695) /gene="ORF17"
/note="Similar to Bacteriophage HK97 gp38"
/codon_start=1
/transl_table=11
/product="unknown"

/translation="MRGLAYNPGILPAEMIIRQRVKPMPSEELKLKNRVFSVNDNKYLNA
MWRSRGKK"
gene complement(25706..25999)
    /gene="abc2"
CDS complement(25706..25999)
    /gene="abc2"
    /codon_start=1
    /transl_table=11
    /product="abc2"

/translation="MPAPLYGADDARRCSCNSVSEVLKFRKNYDLIMSLPQETKDEKEFR
HCIWLAEKERERIYQTSIPFRKATYTHFPEIDPRLRNYRSRYGASND"
gene complement(26013..26528)
    /gene="ssDNA binding protein"
CDS complement(26013..26528)
    /gene="ssDNA binding protein"
    /codon_start=1
    /transl_table=11
    /product="ssDNA binding protein"

/translation="MASRGVNKVIIIIGHLGDPEIRYSPSGTAFANITVATSEQWRDKQTG
EQKEQTEWHVVMGSKLAELASEYLRKGEVYLEGKLRTWRKQDQGQDRFTT
EVIVGVGTMQMLGKQGGEQSSPQRNNQQRQQQSPQQGNHSEPPMDFDD
DIPFAPVTLPFPRHAHAI"
gene complement(26529..27236)
    /gene="ds break repair protein"
CDS complement(26529..27236)
    /gene="ds break repair protein"
    /codon_start=1
    /transl_table=11
    /product="ds break repair protein"

/translation="MDLNKFDEPPFCPEDIEWRIQQSGKTRDGKVVAMVLAYVTNRAIM
KRDVCGKAGWRNEYRDIPNNGVCEGISIKIDSEWVTKWAANETQVEAVKG
GRSGAMKRAAVQWGIGRYLKLEEGFAQTSLDKKHGWHRAKLKDGTGFYW"
LPPSLPGWAIPASDNKPSPENTNQKSPSVDCEQILKDFSDYASTETDKKKLIERYQ
RDWQLMAGNEEAQAKCVCVQVMNIRVNELKQA"
gene complement(27336..27494)  
/gene="ORF18"
CDS complement(27336..27494)  
/gene="ORF18"
  /note="Similar to Bacteriophage Sf6 gene 30"
  /codon_start=1
  /transl_table=11
  /product="unknown"
  /translation="MSFTDNWSDEEFIRQMKEILOGDIHVTCHHSEGEGVTETHVHAESSSLVSP"
gene complement(27491..27643)  
/gene="kil protein"
CDS complement(27491..27643)  
/gene="kil protein"
  /codon_start=1
  /transl_table=11
  /product="kil protein"
  /translation="MRNEIAINHQLRERRAQKNKAIIARFIGDSDKMWFEANKAMKSAINIPWYRRK"
gene complement(27628..27759)  
/gene="cIII"
CDS complement(27628..27759)  
/gene="cIII"
  /codon_start=1
  /transl_table=11
  /product="cIII"
  /translation="MIYAIAGGARMGAFLQNSLLERITRKLRLDGRWKRVEVLLCAMK"
gene complement(27784..28752)  
/gene="ORF19"
CDS complement(27784..28752)  
/gene="ORF19"
  /note="Similar to Bacteriophage Sf6 gene 33"
  /codon_start=1
  /transl_table=11
  /product="unknown"
/translation="MSEVTDLVVIEKANAMIVFQSADQIEILQKVEREVMSFVPDITTAK
GRKEIASLAYKVAQTKYLDGLGKDLVAELKEIPKLIDANRKTVRDRLDELKAK
ARQPLTDYEEEQARIKAEEEAKAAAALAKQIESDHEIAILMDREFDQRQREEARL
KAQEKEKREHERLLKREAAEEKARAEAKAIAEIAAARREAEEKAAAAERAEERE
RIEAEQRAAQREAEAAERAEREKQAIEAERRKQAQEAAERRDDAEAEKEQARIAE
EKRIKEEERRAKDKAHRKEVNNKLADLKVAGASEVAKNIITAVKGEVFATKITY"
gene complement(28920..29162)
  /gene="ORF20"
CDS complement(28920..29162)
  /gene="ORF20"
  /codon_start=1
  /transl_table=11
  /product="unknown"

/translation="MRVKTMGASPXSQRGIQGTNLNTEKGMWGKEDVTEQAVKAAVE
HLMIKDQKYAYEKTGKWLISHQLVDKLPEEAD"
gene complement(29227..29451)
  /gene="ORF21"
CDS complement(29227..29451)
  /gene="ORF21"
  /codon_start=1
  /transl_table=11
  /product="unknown"

/translation="MNQTYIPSCLRNLPKQKAKPRKQAIAKDAEAVIDQAIQLLREELRSG
KLEGMMMPYQRGYLSAIKSVLEKSEL"
gene complement(29455..29838)
  /gene="N"
CDS complement(29455..29838)
  /gene="N"
  /note="Similar to Bacteriophage HK106 N"
  /codon_start=1
  /transl_table=11
  /product="N"

/translation="MTRRTQFKGNSRSRRRERLKAALANGVLAEEAISSEVLHRIPTLS"
RAQIQAKGTHETPERIEDAKPIKFMAQDVQQKEYRNLERAAIVYANEFGHKQPETGVCLPNVAIYAAGYRKSQQLTAR"
  misc_feature  30004..30018
    /standard_name="BoxB"
    /note="Similar to Bacteriophage HK106 boxB"
misc_feature  30062..30070
    /standard_name="BoxA"
    /note="Similar to Bacteriophage HK106 boxA"
promoter complement(30087..30115)
    /standard_name="pL"
    /note="Similar to Bacteriophage HK106 pL"
misc_feature  30089..30105
    /standard_name="oL1"
    /note="Similar to Bacteriophage HK106 oL1"
misc_feature  30113..30129
    /standard_name="oL2"
    /note="Similar to Bacteriophage HK106 oL2"
misc_feature  30133..30149
    /standard_name="oL3"
    /note="Similar to Bacteriophage HK106 oL3"
terminator complement(30302..30327)
    /standard_name="terminator"
gene complement(30409..30714)
    /gene="ORF22"
CDS complement(30409..30714)
    /gene="ORF22"
    /codon_start=1
    /transl_table=11
    /product="unknown"

/translation="MTNRGRVPHTHFSMLNELTFLVAPLEQAGYTLPEKMPDISEGRVFSQWLRDNRGVEPFTFPYNEYPDGRTRVFLYPNEYCRFQTILQRSVAASVR"
gene complement(30754..31095)
    /gene="ORF23"
CDS complement(30754..31095)
    /gene="ORF23"
    /codon_start=1
    /transl_table=11
    /product="unknown"
/translation="MTQFQLALIAREVDGEVIHLRTKDGYNATAMCKSAGKLLADYTRLKTTQDFDELSRDGMPISELIQSFKGRAENQGTWVPDIAINLAQWLSPKFAVQVSRWVREWMSG"
gene complement(31127..31840) /gene="cI"
CDS complement(31127..31840) /gene="cI"
/translation="Similar to Bacteriophage HK106 cI"/codon_start=1/transl_table=1/product="cI"

/translation="MSAKKKPLTQEQLDARRLKIAYEKKKNELGQLQESVADKMGMGQSGVGALFNDINALNAYNAALLAKILNVSEEFPSIAREIYEMYEAVSMQPSLRRSEYEPVFVQAGMFRPLRTFTKGDKEKWVSTTKASDSAFWLEVEGNMTAPGKPSFDMLILVDPEQAVEPGDFCIARLGGDEFTKIKLRDSEGQVFQQLPNPQYPMIPCNESCVSVGVKVIASQWPEETFG"
promoter complement(31847..31857) /standard_name="pRM"/note="Similar to Bacteriophage HK106 pRM"
misc_feature 31851..31867 /standard_name="oR3"/note="Similar to Bacteriophage HK106 oR3"
misc_feature 31874..31890 /standard_name="oR2"/note="Similar to Bacteriophage HK106 oR2"
promoter 31888..31916 /standard_name="pR"/note="Similar to Bacteriophage HK106 pR"
misc_feature 31898..31914 /standard_name="oR1"/note="Similar to Bacteriophage HK106 oR1"
gene 31941..32141 /gene="cro"
CDS 31941..32141 /gene="cro"/note="Similar to Bacteriophage HK106 cro"/codon_start=1/transl_table=1
/product="cro"

/translation="MEQRITLKDYAIRFGQTKTAKDGLVYQSAINKAIHAGRKLIFLTINAD
GSVYAAEIKPFPNSNKTTA"

misc_feature 32149..32157
  /standard_name="boxA"
  /note="Similar to Bacteriophage HK106 boxA"

misc_feature 32180..32194
  /standard_name="boxB"
  /note="Similar to Bacteriophage HK106 boxB"

promoter complement(32270..32297)
  /standard_name="pRE"
  /note="Similar to Bacteriophage HK106 pRE"

gene 32279..32575
  /gene="cII"
  /codon_start=1
  /transl_table=11
  /product="cII"

/translation="MTQASYSKPTQREIDRAETDLLINLSTLTQRGLAKMIGCHESKISRT
DWRFIASVLCAFGMASDISPISRAFKYALDGITKKKSPAATEDSEQIDMQF"

gene 32608..32769
  /gene="ORF24"

CDS 32608..32769
  /gene="ORF24"
  /note="Similar to Sf6 gene 42"
  /codon_start=1
  /transl_table=11
  /product="unknown"

/translation="MTKRRKQYQEKEIRHPDSPEGLVVAAANNRAFAERLVGVVRLAK
AGVKHGRR"

gene 32756..33577
  /gene="O"

CDS 32756..33577
  /gene="O"
  /codon_start=1
/transl_table=11
/product="O"

/translation="MGVVKLANDYRPQLEVVEHRVADTEDGFMRVANEITDSLLMADLT VQQMKVMAIRMKTGQFKNPMDFRTLNTQIAMTGIHHTHVCAAKRQLIERKFLI ADGVKIGVKNVQSQWISQDSTSLLAHTANKTLAKSANGYKPSQLNTKDNIQKTIN TNTPLPNNGGGQKVQPKRKAERIDYESFLNAYTVEGDRPLHAVAVNEKRKR RLKKIIPLKTPNVDGFRAYVRAFVHQAKFYFGDNDTGWADFDYLLREDSLTI GVREGKFADRGIA"
gene 33574..34950
/gene="P"
CDS 33574..34950
/codon_start=1
/transl_table=11
/product="P"

/translation="MKQDIEASVIGGLLIGGLTPTASDVLATLEPEAFSIPLYRKAFEVIRK QARNRNIDLALMVACGEEEEHTSILMTSKNCPSANLKYGAMYGVAHNYHRL VLEIMDEMPQGTIDASSQAMDLEVKRLSAIRKPRDEVKPVRLGEIITDYTDL DRRLRNGEESDTLKGTIEELDAITGGMNAEDLVIIAARPDMGKTELALKIAEGV ASRVPGDVRVRGLIFSMSMALQIAERSIANAGRMSVSLRNPAAMDDEGWARVANGMSQLADLDVWVVDASRLSVEIRSAERHKQENPNSLIMVDYLGLIEKPK KADRNDLAIASISGSKAMKDLKTPVSLSDVKEKRNKRPTNADLRDSGS IEQDADSIIMLYREAYDENSASSAAPFAEIIVTKRFGSLGTVVQRFCNGHFVACD QDEARQICTSSNAPAARGRYYAQGADV"
terminator 34976..35007
/standard_name="terminator"
gene 35024..35464
/gene="ninB"
CDS 35024..35464
/codon_start=1
/transl_table=11
/product="ninB"

/translation="MKKLTFEIRSPAQQAIHAVQQILPDPTKIPvVTTiQERNRSLDQNR KLWACLGDVSRQVWHRGRWLSDEWKKCVFTAALKQQDVVPNLAGNGFVIG QSTSRMVNEFÆELLIQAFGTERGVWSDEARLAEWKKARWGDRRAA"
gene 35431..35988
/gene="DNA N-6-adenine-methyltransferase"
CDS 35431..35988
/gene="DNA N-6-adenine-methyltransferase"
/codon_start=1
/transl_table=11
/product="DNA N-6-adenine-methyltransferase"
/translation="MESAMGRSGCMTIKSNTPAHDKDCWQTPLWLFDALDIEFGFCLDS AASDKNALCAHWLTEADDALNSEWVSAGAIWNNPPYSNIRPWEKAAEQCIQQ RQTVVMLVPEDMSVGWSKALESVDEVRITDGRINFIEPSTGLEKKGSML LIWPFRIPRRIPFRMTTASALMAIGGARRAA"
/gene 35985..36167
/gene="ninE"
CDS 35985..36167
/gene="ninE"
/codon_start=1
/transl_table=11
/product="ninE"
/translation="MRRQRRSITDIICENCKYLPTKRSRNKRKPIPKESDKVFNYTAHLW DIRWLRRERARKTR"
/gene 36164..36334
/gene="ninF"
CDS 36164..36334
/gene="ninF"
/codon_start=1
/transl_table=11
/product="ninF"
/translation="MIDPNRSYEQQSVERALTCANCGQKLHVLEVHV ButtCEHCAELMSDP NSSMYEEDDE"
/gene 36327..37049
/gene="roi"
CDS 36327..37049
/gene="roi"
/codon_start=1
/transl_table=11
/product="roi"
/translation="MNELINGNAIKMTSIEIAELVGRHDKNVKRTIETLAKNGVIRIPQIEV"
SERINNLGFNVQYEHYVFEGEQGKRDSIVVVAQLSPEFTARLVDVDRWDRLED2399
NIPKTLPEALRLAALAEQKMQLENQLAIAAAPKVFEADRVGEASGILIGNFAKVV
GIGPNKLFAWMRDHIILASGSRRNVPMQEYMDRHYFTVKEATAVNTNHGIIQISF
TTK1TGRGQQWLTRKLLDNGMLKVTGEAA"

gene 37049..37339
/gene="ORF25"

CDS 37049..37339
/gene="ORF25"

/note="Similar to Bacteriophage HK97 gp66"
/codon_start=1
/transl_table=11
/product="unknown"

/translation="MANLRK1ARGRECQVR1YGV1ENGNPETTVL1AHYRMA11CGTGMK
PDDILIGAWAC1SACHDEIDRTRHTNNLDNKDAKLYHLEGVIRTQAILLK1EG1K1P"

gene 37336..37698
/gene="RusA"

CDS 37336..37698
/gene="RusA"

/codon_start=1
/transl_table=11
/product="RusA"

/translation="MNEYQ1FVL1PPSVNTYWRMRGSQYYI1SAK1QK1KYR1DVQ1QIIRQL
KL1D1FTKSRLRIK1VIAVPDSRR1D1LNIKLGL1DSLIHAGFAEDDEQFD1DIRVIRG
VK1PG1RL1IGIK1TELENA"

gene 37695..37883
/gene="ninH"

CDS 37695..37883
/gene="ninH"

/codon_start=1
/transl_table=11
/product="ninH"

/translation="MNATIQ1T1PELL1QTR1GN1QTEVARMLSCARGTVLKYNRDSKGERH
IVNGVLMV1KVKGKR1GRP"

gene 37880..38503
/gene="Q"

CDS 37880..38503
/gene="Q"
/codon_start=1
/transl_table=11
/product="Q"

/translational="MRLESVAKFHSPKSPMMSDSPRATASDLSGTDVMAAMGMAQSAGFGMAAFCGKHELSQNDKQKAINYLMQFAHKVSGKYRGVAKLEGNTKAKVQLVLATFAYADYCRSAATPGARCRDCHGTGRAVDIAKTERWGVVVEKECGRCKGVGYSRMPASATYRAVTMLIPNLQTQPTWSRTVKPLYDALVVQCHKEESIADNILNAVTR"
terminator 38611..38645
   /standard_name="terminator"
gene 38937..39260
   /gene="holin"
CDS 38937..39260
   /gene="holin"
   /codon_start=1
   /transl_table=11
   /product="holin"

/translational="MKMPEKNDLLAAILAAKEQGIGAILAFAMAYLRGRYNGGAFKTKTVIDATMCAIIAWFIRDLDFAGLSSNLAYITSVFIGYIGTDSIGSLIKRFAAKKGVEDGGNQ"
gene 39244..39720
   /gene="lysin"
CDS 39244..39720
   /gene="lysin"
   /codon_start=1
   /transl_table=11
   /product="lysin"

/translational="MVEINNNQRKAFLDMLAWSEGTDNGRQKTRNHGYDVIVGGELFTDFSADHPKIVTLKLKSTAAGRYQLLSRWDDAYRKLQLKDFTSPKSKQDAVALQQIKERGALPMIDRGDIRQADRCNSIWASLPGAGYGQFEHKADNLIAKFKTEAGGTVREIEV"
gene 39717..40154
   /gene="Rz"
CDS 39717..40154
   /gene="Rz"
   /codon_start=1
   /transl_table=11
   /product="Rz"
/product="Rz"

/translation="MSRVTAIISALVICIIVCLSNAVHNHYRDNAITYKAQRDKATYIIADM QKRQRDVAELDARYTKELADANATIESLRAVSAGRRKLQVAATCAKSTTTGAS SMGDGESPRLTADAELNYYRLRSGIDRITAQVNYLQYEYIRTQCLK"

gene 39916..40101
/CDS 39916..40101
/gene="Rz1"
/codon_start=1
/transl_table=11
/product="Rz1"

/translation="MLTRLSKVSVLMFLLGVSACKSPPPVQSQRPEPAWAMEKAQDLQ QMLNSIITVSEVESTG"

gene 40191..40466
/CDS 40191..40466
/gene="ORF26"
/codon_start=1
/transl_table=11
/product="unknown"

/translation="MTKLYHRISTFLSGCWAFITSISFAIFSFGSTACSLRGLWRAISALAP KFLPEKAVWRIVERMCSESVREKINVGRHRPTGALCSPLL"

gene 40846..41199
/CDS 40846..41199
/gene="HNH endonuclease"
/codon_start=1
/transl_table=11
/product="HNH endonuclease"

/translation="METGMNkelrvyGsRwDKARLRFQHQhpCVMCEEQGRITPATV VDHIVPHKLKDALKSGNPLAISKAQLLFWSKENWQPCLKAHHDSTKQRMEKSG AVGCDANGYPLDPASHWST"

gene 41199..41402
/CDS 41199..41402
/gene="ORF27"
/product="ORF27"
/codon_start=1
/transl_table=11
/product="unknown"

/translation="MKDLIEYRDGKVQLAIDGVEMKRVTSTIQSHTVGEDVPTLTVSGHVWSEYKGKDHKLEQVDKHSAC"

/standard_name="cos site"

BASE COUNT 10667 a 10205 c 10617 g 10049 t

ORIGIN

1 gttgagttt aacgctaac ccgat**** ttgtagcata tggcgagata
61 acaaatccgt tccgacagt cggcgccgc ggttcaggg ctagaaaatg cagctgtgga
121 cccactcgat cccgcgtccc atgcaggttt ggagaaaaaa gcgagccag cagctgtgga
181 caatcagcag acgtagcag tcagcagcg cccctccttc caaagcttat cggccaacca ggttcaggg
241 gtgtagttt atagcagctt atatcaacct tttaagcag aaaaagagc cagctgtgga
301 caaatcctcag cccctccttc cggccaacca ggttcaggg cccctccttc caaagcttat cggccaacca ggttcaggg
361 gttgagttt aacgctaac ccgat**** ttgtagcata tggcgagata
421 acaaatccgt tccgacagt cggcgccgc ggttcaggg ctagaaaatg cagctgtgga
481 cccactcgat cccgcgtccc atgcaggttt ggagaaaaaa gcgagccag cagctgtgga
541 caatcagcag acgtagcag tcagcagcg cccctccttc caaagcttat cggccaacca ggttcaggg
601 gtgtagttt atagcagctt atatcaacct tttaagcag aaaaagagc cagctgtgga
661 gttgagttt aacgctaac ccgat**** ttgtagcata tggcgagata
721 acaaatccgt tccgacagt cggcgccgc ggttcaggg ctagaaaatg cagctgtgga
781 cccactcgat cccgcgtccc atgcaggttt ggagaaaaaa gcgagccag cagctgtgga
841 gtgtagttt atagcagctt atatcaacct tttaagcag aaaaagagc cagctgtgga
901 caatcagcag acgtagcag tcagcagcg cccctccttc caaagcttat cggccaacca ggttcaggg
961 gttgagttt aacgctaac ccgat**** ttgtagcata tggcgagata
1021 acaaatccgt tccgacagt cggcgccgc ggttcaggg ctagaaaatg cagctgtgga
1081 caatcagcag acgtagcag tcagcagcg cccctccttc caaagcttat cggccaacca ggttcaggg
1141 cccactcgat cccgcgtccc atgcaggttt ggagaaaaaa gcgagccag cagctgtgga
1201 gttgagttt aacgctaac ccgat**** ttgtagcata tggcgagata
1261 acaaatccgt tccgacagt cggcgccgc ggttcaggg ctagaaaatg cagctgtgga
1321 cccactcgat cccgcgtccc atgcaggttt ggagaaaaaa gcgagccag cagctgtgga
1381 gttgagttt aacgctaac ccgat**** ttgtagcata tggcgagata
1441 acaaatccgt tccgacagt cggcgccgc ggttcaggg ctagaaaatg cagctgtgga
1501 cccactcgat cccgcgtccc atgcaggttt ggagaaaaaa gcgagccag cagctgtgga
1561 gttgagttt aacgctaac ccgat**** ttgtagcata tggcgagata
1621 acaaatccgt tccgacagt cggcgccgc ggttcaggg ctagaaaatg cagctgtgga
1681 cccactcgat cccgcgtccc atgcaggttt ggagaaaaaa gcgagccag cagctgtgga
1741 gttgagttt aacgctaac ccgat**** ttgtagcata tggcgagata
4201 cgatctggaa cagaaactgg catcggcgc tgagaatcct ggtgagaaga aatccttctc
4261 tgaacgggct gctgaagagc ttattaagtc atgggacggt aaacagggca cctttggcgc
4321 taaaacgttt aacaagtcac taggcagtga cgctgattcg gctggctcac tgatccagcc
4381 atgcagatc ccaggcatca tcatgccagg cctgcgccgt cttaccattc gtgatctgct
4441 gggctcaggg cgcacttcca gtaacgctct ggaatatgtg cgtgaagagg tgtttaccaa
4501 taacgccgac gtggtggcag agaaagcact gaaggcagaa tcctcagcga
4561 gca ctgggtgcag gcatcacgtc aggtgatgga
4621 tgatgcgcca atgcttcagt cctacattaa caaccgcctc atgtacggcc tggcactgaa
4681 ggaagagggc caagctctga aacgccagaac gttctagggat aaccttgaag ggtctacaca
4741 agttgcaacc gcctgccgct gtccacaggc gacaccccgct gacatcttctgct
4801 tcctctctct gttcttcgct gctctctgta ctaataaggc cagaggttcg ggttgctcat
4861 cgtgagactt tggcttagcg cctgctgagcgt tggctctgcct gaaatcttca gaaagcccct
4921 ctgctgacagc gccttcctct cttgccggtg ggtgctgact ggtgctggtt cttgctcact
4981 gcctgctggtg gccgcgcgca gcctggtgct gaaatcttca gaaagcccct
5041 ggtctgctgct ccctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
5101 ctgctgacagc gccttcctct cttgccggtg ggtgctgact ggtgctggtt cttgctcact
5161 gggacacctt tcctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
5221 cttgctgctg ggtgctgact ggtgctggtt cttgctcact
5281 gggacacctt tcctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
5341 ggtctgctgct ccctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
5401 ggcctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
5461 gggacacctt tcctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
5521 agggggtcc cggctgaggt aaagaggatc agcggcagag agcagctttc aggcggccgc
5581 gggacacctt tcctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
5641 cggctgaggt aaagaggatc agcggcagag agcagctttc aggcggccgc
5701 cggctgaggt aaagaggatc agcggcagag agcagctttc aggcggccgc
5761 gtaataaggc ccctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
5821 gtaataaggc ccctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
5881 gggacacctt tcctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
5941 gggacacctt tcctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
6001 cggctgaggt aaagaggatc agcggcagag agcagctttc aggcggccgc
6061 gggacacctt tcctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
6121 gggacacctt tcctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
6181 gggacacctt tcctctctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
6241 ctgctgacagc gccttcctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
6301 ctgctgacagc gccttcctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
6361 ctgctgacagc gccttcctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
6421 ctgctgacagc gccttcctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
6481 ctgctgacagc gccttcctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
6541 ctgctgacagc gccttcctct cttgctgctg ggtgctgact ggtgctggtt cttgctcact
9001 ccatcctgga gagtctggtg aatgaagcgt ccaagagctg cccagcttg gacaagacc
gt881 gaagaccag cagaagcgt cccagcttg gacaagacc
gt882 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt883 tctctgagc gctgcttggt gatttctggtg aatgaagcgt ccaagagctg cccagcttg
gt884 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt885 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt886 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt887 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt888 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt889 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8810 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8811 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8812 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8813 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8814 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8815 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8816 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8817 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8818 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8819 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8820 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8821 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8822 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8823 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8824 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8825 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8826 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8827 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8828 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8829 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8830 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8831 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8832 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8833 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8834 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8835 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8836 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8837 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8838 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8839 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8840 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8841 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8842 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8843 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8844 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8845 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8846 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8847 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8848 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
gt8849 ccccaagctg gacagagtt tcgagagtctg ggaagacag
gt8850 gatatcagc ggtgagaatg gattggcttg gacagagtt tcgagagtctg ggaagacag
ccagccaaat ctatctggtg gcagggtgag gagtataaag cgtggccatg ccagattgag
11401
gggatcgaag cttcaacctc gggaagtagt gcccagccaa aactttcggt tgctaacctc
11461
gacagttcga tcaccgcgct ttgcctggct tacgacgata tgctgcaggc gaaggtgacg
11521
11581
11641
gctgaaggtt ttctacatcg atgcaaagag cagtgaaaac
11701
11761
11821
11881
caccatcag ccgccgac gctgaaggtt ttctacatcg atgcaaagag cagtgaaaac
11941
12001
12061
12121
12181
12241
cacatactg ccgccgac gctgaaggtt ttctacatcg atgcaaagag cagtgaaaac
12301
caccatcag ccgccgac gctgaaggtt ttctacatcg atgcaaagag cagtgaaaac
12361
gcaggggtg gccgggggag gggatttacg ctaccatggg atggtaaagc
12421
12481
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
12541
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
12601
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
12661
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
12721
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
12781
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
12841
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
12901
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
12961
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13021
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13081
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13141
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13201
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13261
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13321
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13381
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13441
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13501
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13561
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13621
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13681
cgccgagc cttcttgggt tggcctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
13741
tgctgcttg gccgggggag gggatttacg ctaccatggg atggtaaagc
gcccgaaggg cgccagctca agttcccgaa ctcctaccga acagcctgat gatctgcaat ctgtagcgaa ggcaaaaatc ctc tgggcgc ttggggaagg tgagtttgca gggcagctea
23401 gtctttgtgt cctgcaagat agtccagcg taactcatca ctattactta caggttcggc
tctgacagtaa ccacgcaacc ggtonggcgggtaaactctct cgcagccggtg ggcacacgct
ttaaagccgcttcatcttc gcacgggggtc ggccataacg tggccactat cggagtgcac
ttggttcccc tccagtatggg gggggtccggtg ggcacatagt gttggagctt ggttccagcgc
tgcgtatgccta gcgtctctct ctcgctttct gcacactcttc tccgtttccg ggtttctgtc
23461 gtctttgtgt cctgcaagat agtccagcg taactcatca ctattactta caggttcggc
tctgacagtaa ccacgcaacc ggtonggcgggtaaactctct cgcagccggtg ggcacacgct
ttaaagccgcttcatcttc gcacgggggtc ggccataacg tggccactat cggagtgcac
ttggttcccc tccagtatggg gggggtccggtg ggcacatagt gttggagctt ggttccagcgc
tgcgtatgccta gcgtctctct ctcgctttct gcacactcttc tccgtttccg ggtttctgtc
23521 gttctctct cgcagccggtg ggcacacgct
ttaaagccgcttcatcttc gcacgggggtc ggccataacg tggccactat cggagtgcac
ttggttcccc tccagtatggg gggggtccggtg ggcacatagt gttggagctt ggttccagcgc
tgcgtatgccta gcgtctctct ctcgctttct gcacactcttc tccgtttccg ggtttctgtc
23581 gttctctct cgcagccggtg ggcacacgct
ttaaagccgcttcatcttc gcacgggggtc ggccataacg tggccactat cggagtgcac
ttggttcccc tccagtatggg gggggtccggtg ggcacatagt gttggagctt ggttccagcgc
tgcgtatgccta gcgtctctct ctcgctttct gcacactcttc tccgtttccg ggtttctgtc
23641 gttctctct cgcagccggtg ggcacacgct
ttaaagccgcttcatcttc gcacgggggtc ggccataacg tggccactat cggagtgcac
ttggttcccc tccagtatggg gggggtccggtg ggcacatagt gttggagctt ggttccagcgc
tgcgtatgccta gcgtctctct ctcgctttct gcacactcttc tccgtttccg ggtttctgtc
23701 caacgattca cgcagccatct tttgcaacac gtcacagtgtt ttaattccttc gcacacttc
23761 catcaaaaac gtaactcatg tattattttt ccaatcgac gggggtccggtg ggcacatagt gttggagctt ggttccagcgc
tgcgtatgccta gcgtctctct ctcgctttct gcacactcttc tccgtttccg ggtttctgtc
23821 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
23881 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
23941 atattctgtt ttcagcgcctt aatttggattt ggtttctgtt gttttctgtt gttttctgtt
23901 caacagcgaac gacatctgtc ccttggcttt gcctctcttt ctcgcacatg ggtttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24001 caacagcgaac gacatctgtc ccttggcttt gcctctcttt ctcgcacatg ggtttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24061 caacagcgaac gacatctgtc ccttggcttt gcctctcttt ctcgcacatg ggtttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24121 caacagcgaac gacatctgtc ccttggcttt gcctctcttt ctcgcacatg ggtttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24181 caacagcgaac gacatctgtc ccttggcttt gcctctcttt ctcgcacatg ggtttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24241 caacagcgaac gacatctgtc ccttggcttt gcctctcttt ctcgcacatg ggtttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24301 caacagcgaac gacatctgtc ccttggcttt gcctctcttt ctcgcacatg ggtttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24361 caacagcgaac gacatctgtc ccttggcttt gcctctcttt ctcgcacatg ggtttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24421 caacagcgaac gacatctgtc ccttggcttt gcctctcttt ctcgcacatg ggtttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24481 caacagcgaac gacatctgtc ccttggcttt gcctctcttt ctcgcacatg ggtttctgtt gttttctgtt
ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24541 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24601 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24661 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24721 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24781 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24841 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24901 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
24961 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25021 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25081 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25141 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25201 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25261 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25321 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25381 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25441 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25501 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25561 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25621 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25681 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25741 ttcctcggtt ccttgcgagtt cgtttctgtt gttttctgtt gttttctgtt gttttctgtt
25801 ggaattgagt ctggtaaatt cgctcgccgct ctcttttcct tgcagaacccat atacaaatgc
25861 gaacttccttc ttcatttcct gcggcctgct gtaggcagat tatltaatttc tagttttttc
25921 tgaattttaca caacatactc gatagcggagt tggcggaca ccggcgcgcc gctacgcag
25981 cttatgggac cggctggata atcggttcgt gtcagcctgc gtcatttttc gttttttcct
tgctgcaat atccaattgc atcagcgctg ctgagggag ggctgctgag
26041 aaggggagag ttactggtgc cttctttctc gcacccctt gcggtaatc tgtgctgcgt
26091 gactgctcag cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26141 gacgttcagag cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26191 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26241 gatggagagt cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26291 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26341 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26391 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26441 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26491 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26541 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26591 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26641 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26691 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26741 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26791 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26841 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26891 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
26941 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27001 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27051 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27101 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27151 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27201 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27251 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27301 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27351 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27401 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27451 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27501 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27551 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27601 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27651 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27701 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27751 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27801 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27851 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27901 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
27951 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
28001 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
28051 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
tgctgcaat atacaattgc atcagcgctg ctgagggag ggctgctgag
28101 atatactgcgg cggcgtttacct ctggtacagc atctggttc tttttgttgc ttttttttgt
ccttgcctgc gcccttgctc gccctgctgc ctggctttctc ttca
gcttctc ctttttaagc gttcttcatg
cctcgttgccttt tcctgctccg ctttgagtct tgcctcttct ctttggcggt caaattcgcg
atccatcaaa atagctattt catggtcaga ctcaatttgc tttgcgagag cttcagctgc
tgccttagct tcttcttcgg ctttaatccg cgcctgttct tcctcataat cagtaagagg
ctggcgtgcc ttggctttca gctcatcaag gcgatcacgc actgtcttgc ggttggcatc
aattagcttt ggaatctttc tcatgtagc aacaaggtct tggcaagac cagcatgagac
ttttttcgatg ggcagatcct tctcatcttt caaatctaat caaatctttag ggtgagggc
cgctttcaaat caagatgtat cgcagctttcc catttggtatt
ctctcgttgg aatgaataca cagcttgaga ttgattatat ccccccccag ggtgtttgtaa
30241 ttagat ccagcagcag cgcagtcatg tctttgtaat cggtagtattag ttcttttatt ttcttttatt taatttttttag gttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
33001 tcagttttgc gctgcaccaag gcagacattat tgactgtaaa ttacctcatg ctgtatgcgt
33061 gaaaatcgcga tggaaacagg tggatttcga tggacagtgc cagggacagct taaccatagc
ttaaaccagct aataaaacat tgaactaatgc gtagaagatgc gttacctatc ccaagatgct
33071 tchcctccc taacttacg cagggatgact cagatgcgtg
ttcataatgc gtaggttgaact tgggaagtcg caggttgtaa cggatgatta
tgggaaagtgcc gcctacgcgt cgggttggag cggatggtttgc ggcgaacagcat ggcgtttgatc
cggcatgtgc caggttttgtc ggggaaagtgc gcctatgcgtg
t黄金fts tggagatggc gtaattgctc cctcgtgcc tcagctctgc
tctccagttc gacagcaattgc ggcgcagcat gcagcgtttt gcgcagttcgc
tctccccct gcgcagctc tggagatggt gacagcatcgc
tcatatctgc ggcatgccat gcagctgtc tggagatggc
tcgggaaagtc ggggaaagtgc gcctatgcgtg
tgggaaagtgcc gcctacgcgt cgggttggag cggatggtttgc ggcgaacagcat ggcgtttgatc
cggcatgtgc caggttttgtc ggggaaagtgc gcctatgcgtg
t黄金fts tggagatggc gtaattgctc cctcgtgcc tcagctctgc
tctccagttc gacagcaattgc ggcgcagcat gcagcgtttt gcgcagttcgc
tcatatctgc ggcatgccat gcagctgtc tggagatggc
tcgggaaagtc ggggaaagtgc gcctatgcgtg