12-2010

The Genomic Sequence and Annotation of Bacteriophage HK239

Alice Ann Wright
Western Kentucky, aliwright84@gmail.com

THE GENOMIC SEQUENCE AND ANNOTATION OF BACTERIOPHAGE HK239

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

List of Tables

List of Figures

THE GENOMIC SEQUENCE AND ANNOTATION OF BACTERIOPHAGE HK239

Alice Ann Wright December 2010 90 pages

Directed by: Rodney King, Sigrid Jacobshagen, and Claire Rinehart

Department of Biology Western Kentucky University

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, P1*vir*, 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].

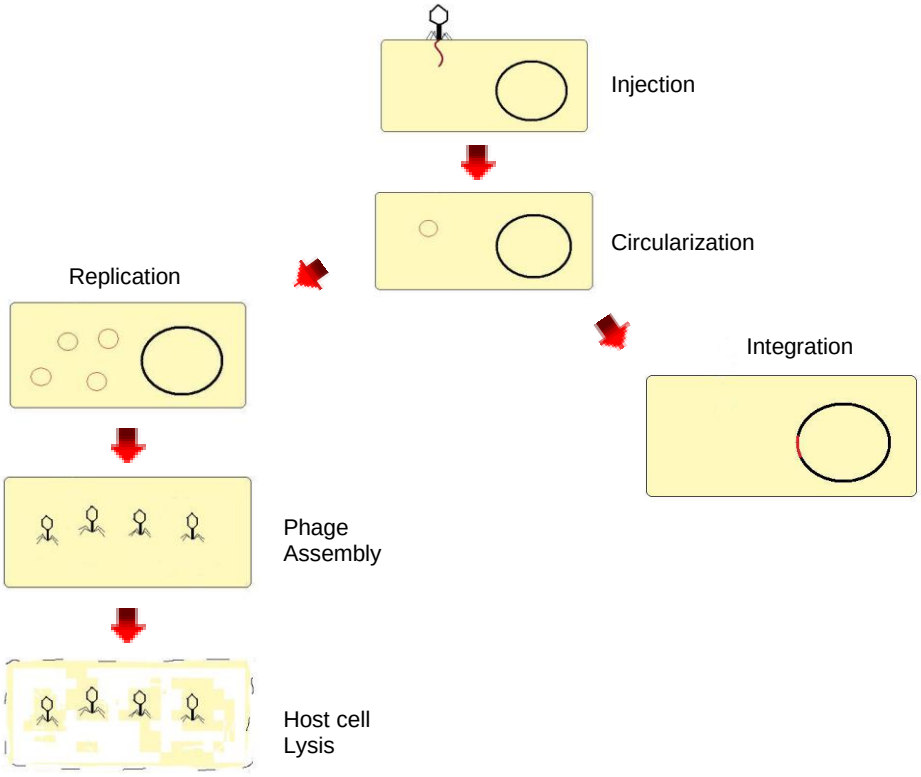


Figure 1. Lytic and lysogenic lifestyles. On the left is the lytic lifestyle and on the right is the lysogenic lifestyle. Phage genomic DNA is in red and the bacterial chromosome in black. In general, the lytic cycle is completed within 45 minutes and approximately 100 viral particles are generated [9].

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].

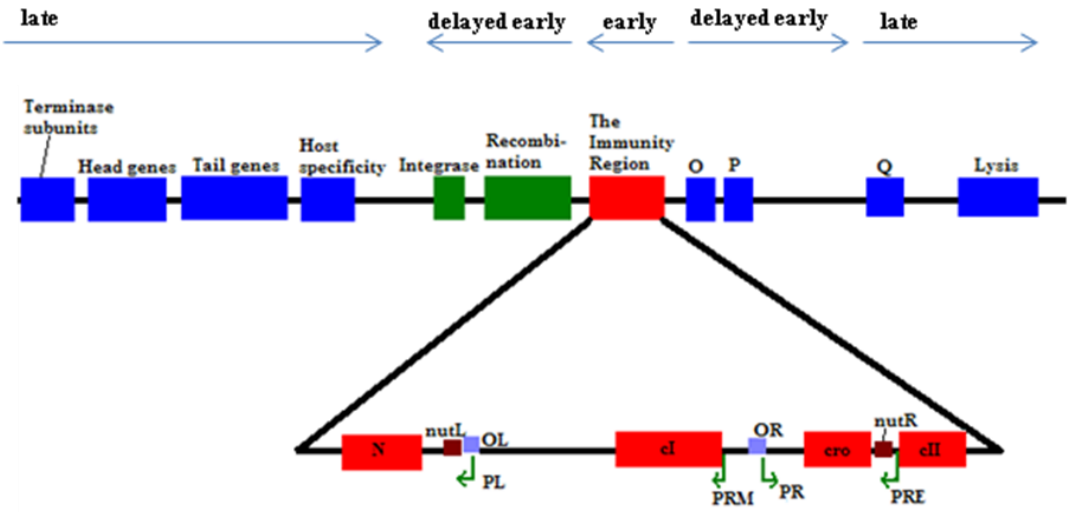


Figure 2. Schematic illustrating the genomic organization of Lambdoid phage [3, 4]. Arrows above the map indicate leftward and rightward transcripts [9].

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 *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 *nut* (N utitlization) sites, comprised of *BoxA* and *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 *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 *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-teminal 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 λ, T4rII, P1*vir*, P2, and HK022 [35]. Previous work by Wright *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 φ80 *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 $7H_2O$, 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 5×10^9 to 3×10^{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 uL 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 $MgSO_4$, 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 $MgSO_4$.

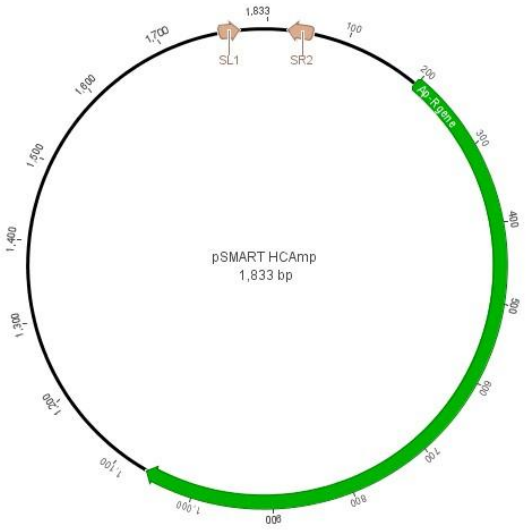


Figure 3. pSMART® HCAmp vector, accession number AF399742 [41]. The primers, SL1 and SR2, flank the insertion site. The Ap-R gene allows for selection of ampicillin resistant colonies.

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

(Table 1). 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

Primer	Sequence	Purpose
SR2	GGTCAGGTATGATTTAAATGGTCAGT	sequence from inserts
SL1	GCAGTCCAGTTACGCTGGAGTC	sequence from inserts
S1	CGGATATCGCTGAAATTATCGGTG	primer walking
S2a	CAGCGCCACACAGTCGAAATT	primer walking
S2b	CTTTGCCAACGGCGCCAAGT	primer walking
S3a	CCGTTCGTCGCATCTCGTTG	primer walking
S3b	GGTCAACGGCGTCAAAATTGA	primer walking
S4a	AAAAGGCCAGTCGCCTCTGGAGCT	primer walking
S4b	CGGTGAAAACAACGAACTCTCG	primer walking
S5a	TGAACGATGGCGATCACCGT	primer walking
S5b	GGCCGTGCTTATTACTGCTGCT	primer walking
S67a	GCCGATTCGACGCTTACCATAA	primer walking
S67b	CGGAAATTCAGTACAGCCTGACC	primer walking
S8a	CATCCATCGAGACAGAGATTTCGT	primer walking
S8b	CGAACAATTATGTTGCCGGCTCT	primer walking
S9a	GAAAGACCAGCTGCCGGAGT	primer walking
S9b	GGGTCATTTGGTGTGGGTTCTAAA	primer walking
S10a	TGCGCTGAGCCTCTATCCAGTC	primer walking
S10b	TGCCGTTATCGTCTCCGTATTTAA	primer walking
S11a	GGTTGTGCTTCCGCAATGCTATA	primer walking
S11b	TGCAGCACGAAGCATCTGATG	primer walking
S12a	GCAAAAGAGGCAGCAGAACGAG	primer walking
S12b	CCATCCTTCGTTTCGTATGCGTA	primer walking
S14a	TTCTCATGTTCAAGCCGGGA	primer walking
S14b	GATGGTTTCATGCGCGTTGC	primer walking
S16a	CGTCACGGGGCTTTCTGATG	primer walking
S16b	GTTAAGCCGCTGTATGACGCTC	primer walking
S18	CCAAATACGTTAATCTTCTCGCGA	primer walking
S2:1	TCGATCCCAGACAGCCACCAAC	primer walking
S2:8a	CGTCAATCTTCACCTCGGCC	primer walking
S2:8b	GGGTCATTTGGTGTGGGTTCTAAA	primer walking
S2:11a	AGCATGTTGCATCGCGTCGA	primer walking
S2:12b	CCATCCTTCGTTTCGTATGCGTA	primer walking
S2:13b	GCCTGAATCTGCGCTCTGCTT	primer walking
S2:14a	TCAAGCCGGGATGTTCTCGC	primer walking
S2:14b	TCAACCCACCTGGTCACGCA	primer walking
S2:17a	CGCCAGCATATCGAGGAACG	primer walking
S2:17b	CGCTCTGGTTATCTGCATCATCGT	primer walking

Table 1 continued

Primer	Sequence	Purpose
395	CAGTTCAGGAAGGATGCCG	primer walking
396	GTCATTCTGGTCTGTTTC	primer walking
397	TAATCCCTACAACCAAAG	primer walking
398	ACTGGTTCCTGTTTCTCA	primer walking
399	GAACGCTGACGAACTGAT	primer walking
400	CTTCTCGGTAATGCGTTG	primer walking
401	TGCGTACCAAACTAAAATC	primer walking
402	TACCATAAATAGTACGCAGT	primer walking
403	GCATAGCAAGATGGGTA	primer walking
404	GCCTCTATCCAGTCGTGT	primer walking
405	GTGTAATACTTCTGAACT	primer walking
406	CATTCTGGCTTGAGGTTGA	primer walking
407	GTCACGAACAAATCTGAT	primer walking
408	TATCTGTTCCCTCTGACCA	primer walking
409	GTATGAGCAGAGTAACCG	primer walking
410	ACTACAGTAACGGACTGC	primer walking
412	GATCAGTTCGTCAGCGTT	primer walking
413	GAAGAGTCCGATATGTGGC	primer walking
414	CTTTGAACTGAGTTCTGCG	primer walking
415	GATATCATTCAGGACGAGC	primer walking
416	GATCTGATATTGTCATGCCA	determining left end
417	CAAACTCGAACAGGTAGAC	determining right end
418	GAAAGCAATAGAAGAAGC	primer walking
419	GATGCCAGCAAAAGTGATC	primer walking
420	GTAGTGCGTCCTGCTAATG	primer walking
421	GAGAAATGGGTAAGCACA	primer walking
422	GTCTATCCAGTTCTCCCACAC	determining right end
423	GTAAAACGGTGATATAGAG	determining left end
424	CTTGCGGTGATAGATTTA	primer walking
425	CTTAGAAGTGAGTATGAG	primer walking
426	GTTCACTTTGGTTATTGC	primer walking
427	GTATTTATGTCAACACCG	primer walking
428	GTTGTGGGGAAAGTTATC	primer walking
429	GATCCCATGCAATGAGAG	primer walking
430	CTATGTTTAGTGAGTTGTATC	primer walking
431	GCAGGGGTGTATTGTTTG	primer walking
432	GAGGTATATGACAAACCGAG	primer walking
433	CATACGCACTTTTCTATG	primer walking

Table 1 continued

Primer	Sequence	Purpose
436	CTTTCCCAAGCACGGATA	re-sequencing
437	GCACCCCGTATTAACGATG	re-sequencing
438	CTGGTGGGCAAGGCTGAAGTC	re-sequencing
439	GTTGTCCGTTCGTCGCATC	re-sequencing
440	CTGCTGAGGGGAGATTCG	re-sequencing
441	GACCCGAAAAGTGGCGAT	re-sequencing
442	GTCCCTGTCGTCTTCCTCA	re-sequencing
443	CAGGACGACAACGTGGTC	re-sequencing
444	CGGTAGAGTAGATTGGGA	re-sequencing
445	CATCAATTTGACTGTAAT	re-sequencing
446	CAGTCTTTCCAGCTCGCT	re-sequencing
447	TATCCAGTTCGCTCGGCTG	re-sequencing
448	GCATAATCCTTACTACATC	re-sequencing
449	CGATTCGACGCTTACCAT	re-sequencing
450	CTGGACAATACGTCTGCGT	re-sequencing
451	GTAACGCTTATGCCGACG	re-sequencing
452	CTCCTCCCTACGCTGTTAC	re-sequencing
453	TAATCCCTACAACCAAAG	re-sequencing
454	CGCCATAAATACAGCGGC	re-sequencing
455	GAAACCCTTGTTCATGGC	re-sequencing
456	CATGACAACGTACAATGA	re-sequencing
457	CATAAGCATTGCCACTATC	re-sequencing
458	CCATATTTGGATTTCGAG	re-sequencing
459	CTTCAGCGATTATGCGTC	re-sequencing
460	GATGAGCCATTCTGCCTG	re-sequencing
461	CTTGAACGAATCACCCGTA	re-sequencing
462	GAATCACCAATAAATCTG	re-sequencing
463	ATTTCCAATAATCAGAAC	re-sequencing
464	CTAAAGGTACTCACGAAAC	re-sequencing
465	GCCTGCGGGACTATTGC	re-sequencing
466	GATCTCTCACCTACCAAAC	re-sequencing
467	GTACTCATGGTTATATGT	re-sequencing
468	GTATAGTCAGCAAGTAGC	re-sequencing
469	CATCTCGTAGATTTCTCTG	re-sequencing
470	CTATCACCGCAAGGGATA	re-sequencing
471	CATCAGTCCGATTAGCAG	re-sequencing
472	GTGTGGAAGTCTGTCACCGA	re-sequencing
473	GTTGCCAGAACGTCGCTG	re-sequencing

Table 1 continued

Primer	Sequence	Purpose
474	GAAAGCCTTCGAGGTTATC	re-sequencing
475	CTGGAAGTGTGTGTTTAC	re-sequencing
476	CAAATCGGCATTGATGGC	re-sequencing
477	GCCAGCGTTTCGATGGTA	re-sequencing
478	GAGGCTGCTTAATGGCTA	re-sequencing
479	CGGATTCGGAATGGCTGC	re-sequencing
480	CAGCGAAGCGTTTGATAAG	re-sequencing
481	GATTCCACTTCTGAGACG	re-sequencing
482	GCAGGAATACATCAGGAC	re-sequencing
483	GAGGGTTACCTGACTTAA	re-sequencing
484	CAGTTCACTTACCTGAAAT	re-sequencing
485	CTGAATATCCACGCCCAAAT	re-sequencing
486	CTGAGCGGGTCATGGGC	re-sequencing
487	CAGGGCGTTCGCAGAGCG	re-sequencing
488	GTTATCCGCCGTCCAATC	re-sequencing
489	CGATGGCTGTTATGATAT	re-sequencing
490	CATTGCTCCGTGTATTCAC	re-sequencing
491	CGGTTGTCACGGAGCCAT	re-sequencing
492	CGTAGTCGATGCGTTCTG	re-sequencing
493	GATGTAGCCGATGAACAC	re-sequencing
495	CAAACGCAGGAGTGAAACA	re-sequencing
496	GAGTGTGATGAATACCTG	re-sequencing
497	CGGCAGATGAAGGTGATG	re-sequencing
498	GTTGTAGGCGTTCAGGAAG	re-sequencing

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 $MgSO_4$.

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 $MgCl_2$), 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 -80ºC 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} 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.

Strain	Genotype
LE392	*supE*44 *supF*58 *hsdR*514 *galK*2 *galT*22 *metB*1 *trpR*55 *lacY*1 [9]
RK898	MG1655; wild-type
10G Elite	F- mcrA D(mrr-hsdRMS-mcrBC) φ80dlacZΔM15 ΔlacX74 endA1 recA1araD139 Δ(ara, leu)7697 galU galK rpsL nupG λ- tonA [43]
RK1212	HK106 lysogen of RK898

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.

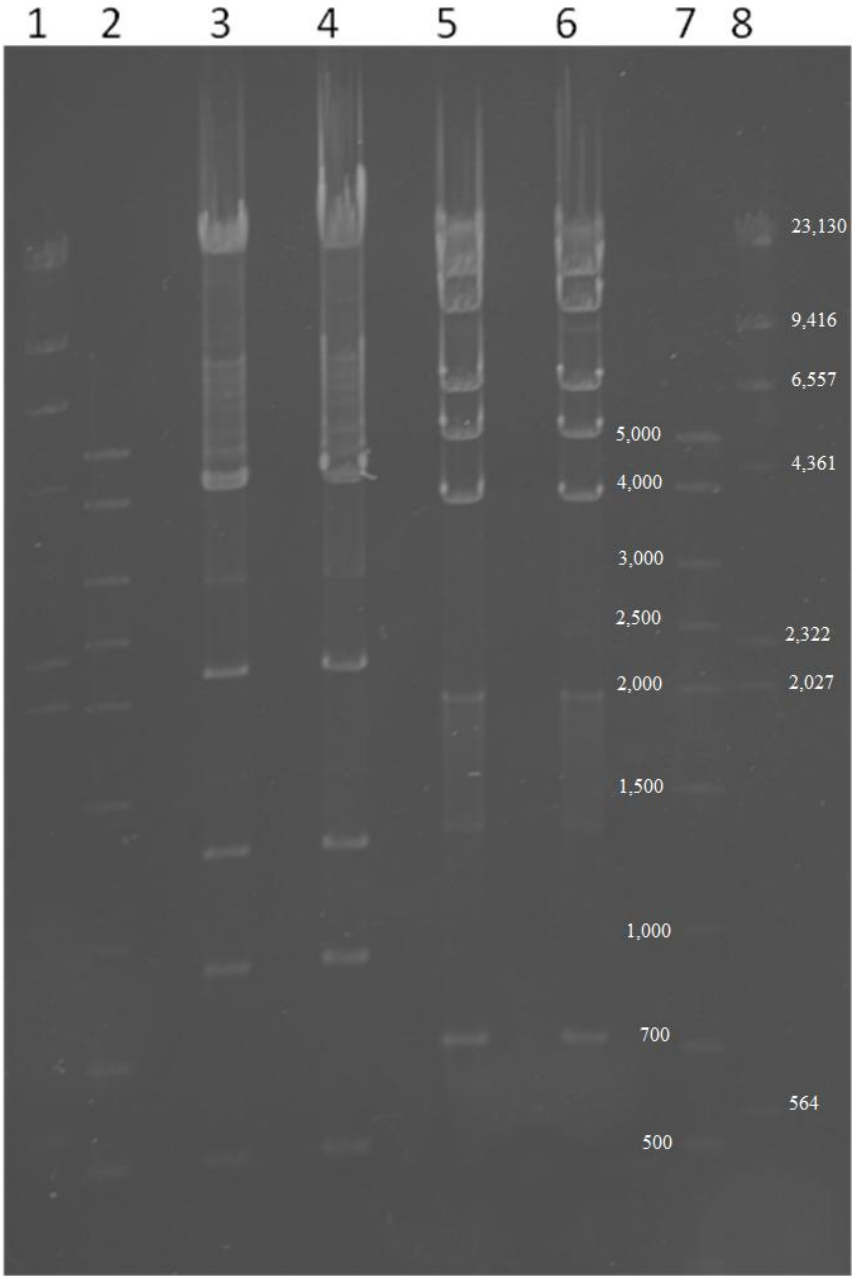


Figure 4. Restriction digests of HK239 genomic DNA. Lanes 1 and 8 contain a Lambda HindIII ladder. Lanes 2 and 7 contain a mid-range ladder (Fisher). Lanes 3 and 4 contain HK239 genomic DNA digested with MfeI, heat inactivated and untreated respectively. Lanes 5 and 6 contain HK239 genomic DNA digested 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.

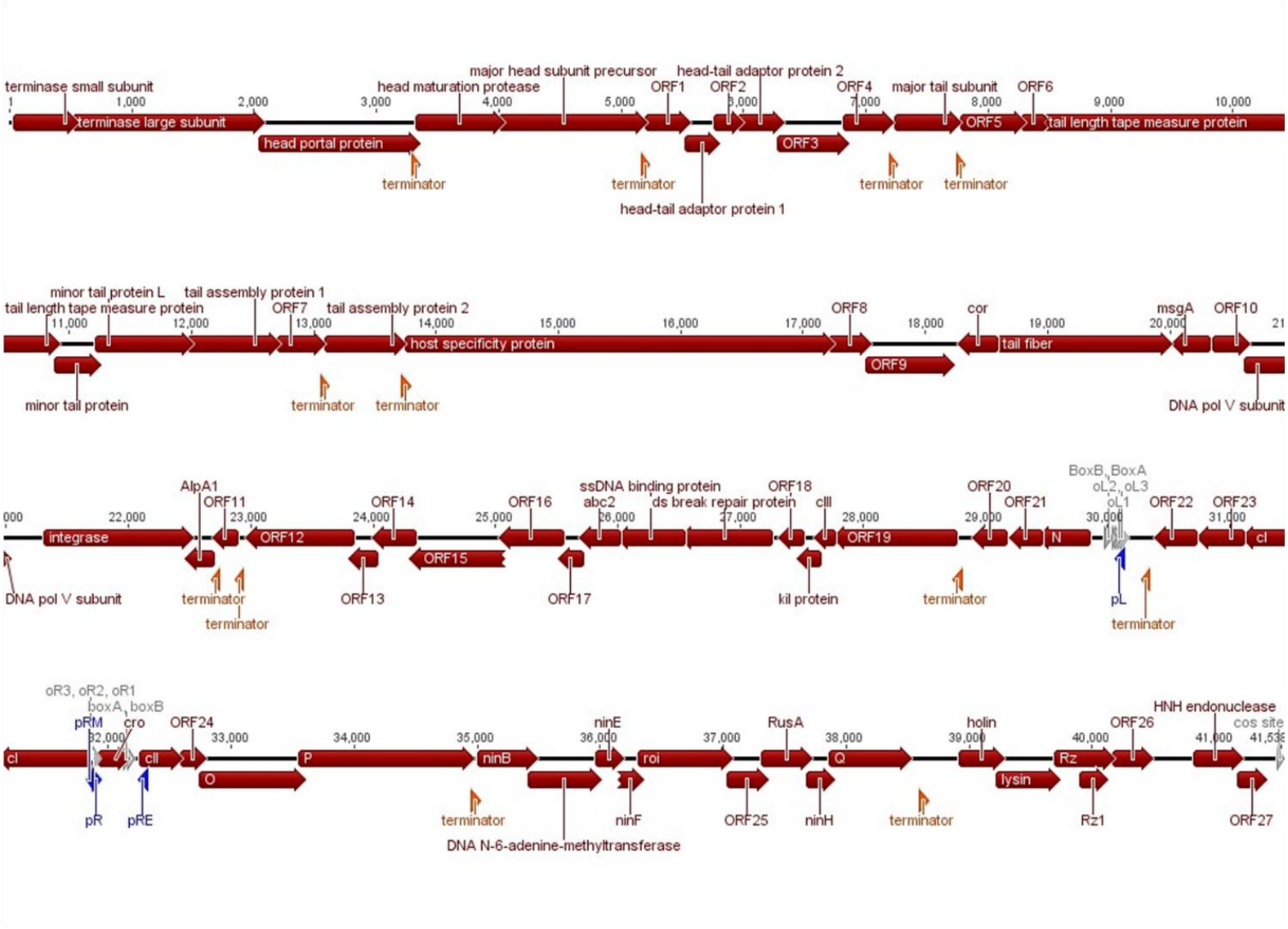

terminase small subunit
head maturation protease
major head subunit precursor
head-tail adaptor protein 2
ORF1
ORF2
ORF4
major tail subunit
ORF6
terminase large subunit
head portal protein
ORF3
ORF5
tail length tape measure protein
terminator
terminator
terminator
terminator
head-tail adaptor protein 1
minor tail protein L
tail length tape measure protein
tail assembly protein 1
ORF7
tail assembly protein 2
ORF8
cor
msgA
ORF10
host specificity protein
tail fiber
ORF9
minor tail protein
terminator
terminator
DNA pol V subunit
AlpA1
ORF11
ORF14
ORF16
ssDNA binding protein
abc2
ds break repair protein
ORF18
cIII
ORF20
ORF21
BoxB, BoxA
oL2, oL3
oL1
ORF22
ORF23
integrase
ORF12
ORF15
ORF19
N
cI
DNA pol V subunit
terminator
terminator
ORF13
ORF17
kil protein
terminator
pL
terminator
oR3, oR2, oR1
boxA, boxB
pRM
cro
ORF24
ninE
RusA
holin
ORF26
HNH endonuclease
cos site
cI
cII
P
ninB
roi
Q
Rz
O
lysin
pR
pRE
terminator
ninF
ORF25
ninH
terminator
Rz1
ORF27
DNA N-6-adenine-methyltransferase

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.

Gene	Strand	Left end	Right End	Length	Nucleotide e value	Protein e value	Homology
Terminase small subunit	(+)	50	535	485	0	3.00×10^{-88}	HK022, HK97
Terminase large subunit	(+)	542	2056	1514	0	0	HK022, HK97
Head portal protein	(+)	2056	3330	1274	0	0	HK022, HK97
Head maturation protease	(+)	3348	4025	677	0	4.00×10^{-129}	HK022, HK97
Major head subunit precursor	(+)	4028	5185	1157	0	0	HK022, HK97
ORF1	(+)	5219	5545	326	3.00×10^{-168}	6.00×10^{-57}	HK022, HK97
Head-tail adaptor protein 1	(+)	5545	5781	236	2.00×10^{-118}	7.00×10^{-38}	HK022
ORF2	(+)	5778	5975	197	4.00×10^{-95}	4.00×10^{-28}	HK022
Head-tail adaptor protein 2	(+)	5977	6309	332	3.00×10^{-173}	4.00×10^{-58}	HK022
ORF3	(+)	6302	6841	539	0	1.00×10^{-99}	HK022
ORF4	(+)	6838	7203	365	0	9.00×10^{-65}	HK022
Major tail subunit	(+)	7258	7758	500	0	4.00×10^{-92}	HK022
ORF5	(+)	7797	8282	485	0	6.00×10^{-88}	HK022
ORF6	(+)	8288	8455	168	8.00×10^{-82}	2.00×10^{-25}	HK022
Tail length tape measure protein	(+)	8478	10898	2420	0	0	HK022
minor tail protein	(+)	10898	11236	338	1.00×10^{-176}	2.00×10^{-60}	HK022
minor tail protein L	(+)	11233	11988	755	0	5.00×10^{-147}	HK022
tail assembly protein 1	(+)	11990	12700	710	0	2.00×10^{-135}	mostly HK022
ORF7	(+)	12730	13071	341	N/A	7.00×10^{-12}	*Y. pestis* phage lipoprotein

Table 3. Continued

Gene	Strand	Left end	Right End	Length	Nucleotide e value	Protein e value	Homology
tail assembly protein 2	(+)	13115	13723	608	1.00×10^{-164}	3.00×10^{-78}	HK022 partial
host specificity protein	(+)	13776	17231	3455	0	0	HK97 partial
ORF8	(+)	17233	17535	302	6.00×10^{-110}	1.00×10^{-47}	HK022
ORF9	(+)	17535	18209	674	2.00×10^{-122}	1.00×10^{-96}	HK022 partial
cor	(-)	18288	18572	284	3.00×10^{-133}	9.00×10^{-45}	phi80
tail fiber	(+)	18615	19988	1373	1.00×10^{-162}	4.00×10^{-77}	HK97, beginning only, partial to tail fiber
msgA	(-)	20035	20307	272	N/A	3.00×10^{-32}	virulence protein
ORF10	(+)	20372	20626	254	N/A	6.00×10^{-27}	*Enterobacter cancerogenus* hypothetical
DNA pol V subunit	(+)	20626	21015	389	N/A	7.00×10^{-50}	DNA polymerase subunit V
Integrase	(+)	21329	22507	1178	0	0	stx2 converting phage 1717
AlpA	(-)	22488	22679	192	4.00×10^{-95}	2.00×10^{-95}	*E. coli*
ORF11	(-)	22710	22877	167	2.00×10^{-78}	7.00×10^{-25}	Lambda
ORF12	(-)	22978	23817	839	N/A	N/A	None
ORF13	(-)	23817	24017	201	N/A	N/A	None
ORF14	(-)	24010	24324	315	N/A	N/A	None
ORF15	(-)	24321	25055	734	0	4.00×10^{-110}	partial HK620,
ORF16	(-)	25052	25534	482	2.00×10^{-151}	1.00×10^{-61}	HK620
ORF17	(-)	25531	25695	164	7.00×10^{-77}	8.00×10^{-23}	HK97
abc2	(-)	25706	25999	293	3.00×10^{-138}	2.00×10^{-49}	HK97, HK022
ssDNA binding protein	(-)	26013	26528	515	0	8.00×10^{-79}	CP-1639 partial

Table 3. Continued

Gene	Strand	Left end	Right End	Length	Nucleotide e value	Protein e value	Homology
ds break repair protein	(-)	26529	27236	707	0	4.00×10^{-134}	Sf6
ORF18	(-)	27336	27494	158	7.00×10^{-77}	1.00×10^{-22}	Sf6
kil protein	(-)	27491	27643	152	7.00×10^{-62}	6.00×10^{-20}	Sf6, HK97 kil protein
cIII	(-)	27628	27759	131	6.00×10^{-62}	2.00×10^{-16}	Sf6
ORf19	(-)	27784	28752	968	0	5.00×10^{-167}	Sf6
ORF20	(-)	28920	29162	242	N/A	N/A	None
ORF21	(-)	29227	29451	224	1.00×10^{-111}	1.00×10^{-34}	E. coli,
N	(-)	29455	29838	383	0	8.00×10^{-69}	Stx2 converting phage 1
ORF22	(-)	30409	30714	305	N/A	2.00×10^{-32}	none, kilA like protein
ORF23	(-)	30754	31095	341	N/A	8.00×10^{-30}	none, kilA like protein
cI	(-)	31127	31840	713	0	4.00×10^{-135}	HK97
cro	(+)	31941	32141	200	2.00×10^{-88}	1.00×10^{-30}	Lambda
cII	(+)	32279	32575	296	3.00×10^{-143}	7.00×10^{-49}	HK022, HK97
ORF24	(+)	32608	32769	161	1.00×10^{-78}	3.00×10^{-22}	Sf6
O	(+)	32756	33577	821	0	3.00×10^{-160}	HK97
P	(+)	33574	34950	1376	0	0	HK97
ninB	(+)	35024	35464	440	0	7.00×10^{-79}	HK97, ninB
DNA N-6-adenine methyl-transfer-ase	(+)	35431	35988	557	0	4.00×10^{-103}	HK97, most/E. coli
ninE	(+)	35985	36167	183	4.00×10^{-85}	1.00×10^{-26}	HK620
ninF	(+)	36164	36334	171	4.00×10^{-80}	1.00×10^{-23}	HK620

Table 3. Continued

Gene	Strand	Left end	Right End	Length	Nucleotide e value	Protein e value	Homology
roi	(+)	36327	37049	722	0	9.00×10^{-136}	HK620
ORF25	(+)	37049	37339	290	1.00×10^{-126}	5.00×10^{-48}	HK620
RusA	(+)	37336	37698	362	2.00×10^{-176}	5.00×10^{-62}	HK620
ninH	(+)	37695	37883	188	2.00×10^{-88}	3.00×10^{-28}	HK620
Q	(+)	37880	38503	623	0	1.00×10^{-118}	HK620
holin	(+)	38937	39620	683	1.00×10^{-166}	2.00×10^{-53}	HK620
lysin	(+)	39244	39720	476	0	1.00×10^{-88}	HK022
Rz	(+)	39717	40154	437	0	5.00×10^{-75}	HK97
Rz1	(+)	39916	40101	186	9.00×10^{-87}	3.00×10^{-26}	HK022
ORF26	(+)	40191	40466	275	N/A	N/A	None
HNH endonuclease	(+)	40846	41199	353	0	2.00×10^{-53}	HK022, mostly
ORF27	(+)	41199	41402	203	4.00×10^{-100}	N/A	HK022, HK97

Table 4. Predicted Rho-independent transcription terminators in the HK239 genome.

DNA Sequence	Strand	Left end	Right end	Length
GCCCCGTAATAACGGGGCTTAATTTTT	(+)	3309	3335	27
GGGCGGGGAAACCCGCCCTTTT	(+)	5192	5213	21
CCGGCCTTGAGCCGGTTTTTTT	(+)	7220	7241	22
AGGGCGGCAACGCCCTTATTAATCAGGATT	(+)	7763	7792	29
CCCGCTTCGGCGGGTTTTTTT	(+)	13083	13103	20
GCCACCTTCGGGTGGCTTTTTTAT	(+)	13743	13766	23
CCGCCTGATGGCGGTTTCTTTTT	(-)	22695	22717	22
ACTCGCTACGGCGAGTTTTGTTTT	(-)	22891	22914	23
TTGCCCTCCAGTGTGAGGGCGATTTTTTT	(-)	28765	28793	28
CCCGGCCACAGAGCCGGGTTTTCTTT	(-)	30302	30327	25
AGGCCTGCTGGTAATCGCAGGCCTTTTTATTT	(+)	34976	35007	31
GCCCTGAGTTAATATCTCGGGGCTTTTTGCGTTTT	(+)	38611	38645	34

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

Element name	Strand	Left end	Right End	Length	Homology
boxB left	(-)	30004	30018	14	HK106
boxA left	(-)	30062	30070	8	HK106
O_{L1}	(-)	30089	30105	16	HK106
P_L	(-)	30087	30115	28	HK106
O_{L2}	(-)	30113	30129	16	HK106
O_{L3}	(-)	30133	30149	16	HK106
P_{RM}	(-)	31847	31875	28	HK106
O_{R3}	(+)	31851	31867	16	HK106
O_{R2}	(+)	31874	31890	16	HK106
O_{R1}	(+)	31898	31914	16	HK106
P_R	(+)	31888	31916	28	HK106
boxA right	(+)	32149	32157	8	HK106
boxB right	(+)	32180	32194	14	HK106
P_{RE}	(+)	32270	32297	27	HK97

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).

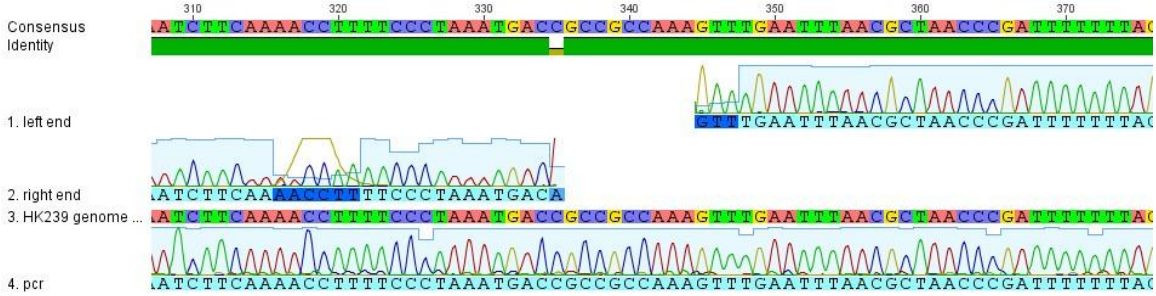


Figure 6. Verification of the ends of HK239 and identification of the cos site. Sequences from top to bottom: consensus sequence, sequence from left end, sequence from right end, sequence of PCR products generated from ligated genomic DNA, and HK239 genomic sequence. The highlighted sequence on the PCR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

Bioinformatics Analysis

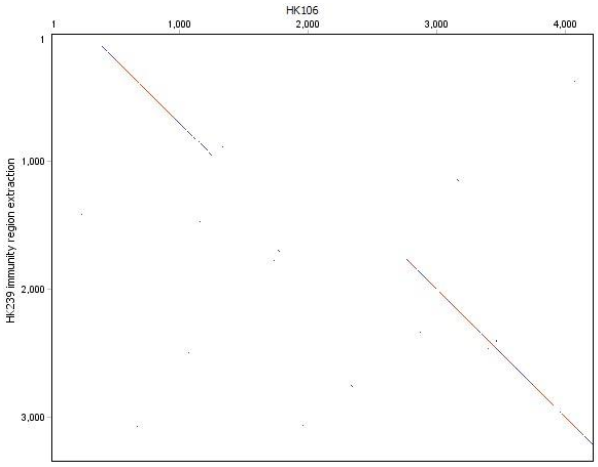
Once the genome of HK239 had been completed, it was compared to other phage genomes. It is known that lambdoid phages are genetic mosaics of one another [34], and this also holds true for HK239. Among the information revealed by HK239's mosaicism, two pieces were particularly important: HK239's potential 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, cI, 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 cI, 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.



B.

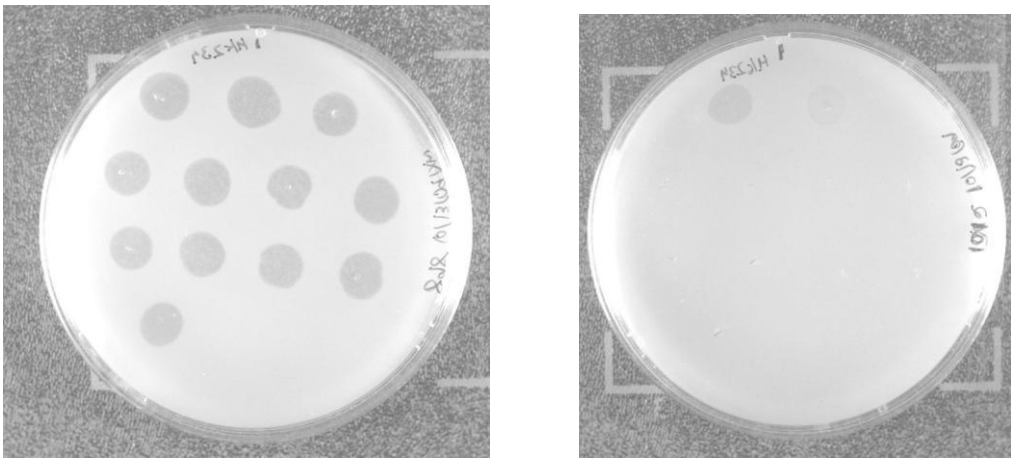


Figure 7. A. Dot plot of HK239 and HK106 immunity regions (HK106 accession number: EF120461) [41]. Regions of homology are indicated by the diagonal line. B. Immunity test: Plating of HK239 on RK898 (left) and an HK106 lysogen (right). A 5 μL aliquot of each serial dilution ranging from 10^{-1} to 10^{-12} 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.

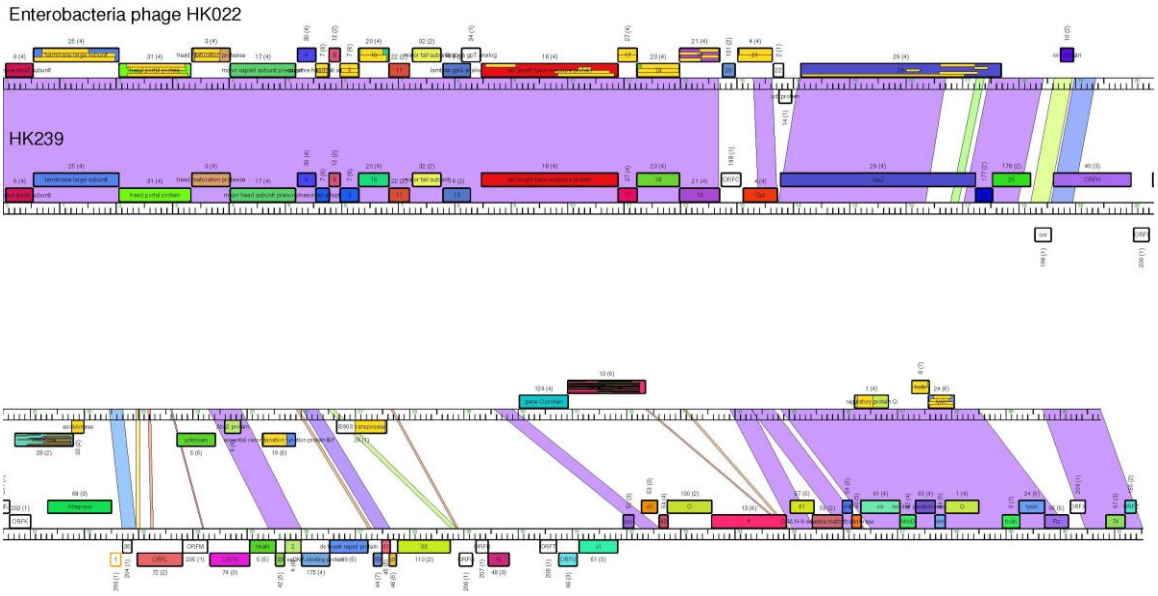


Figure 8. Phamerator output of HK022 (accession number AF069308) and HK239 genomes [42]. The purple shading and other colors indicate homology. The absence of shading indicates no homology.

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).

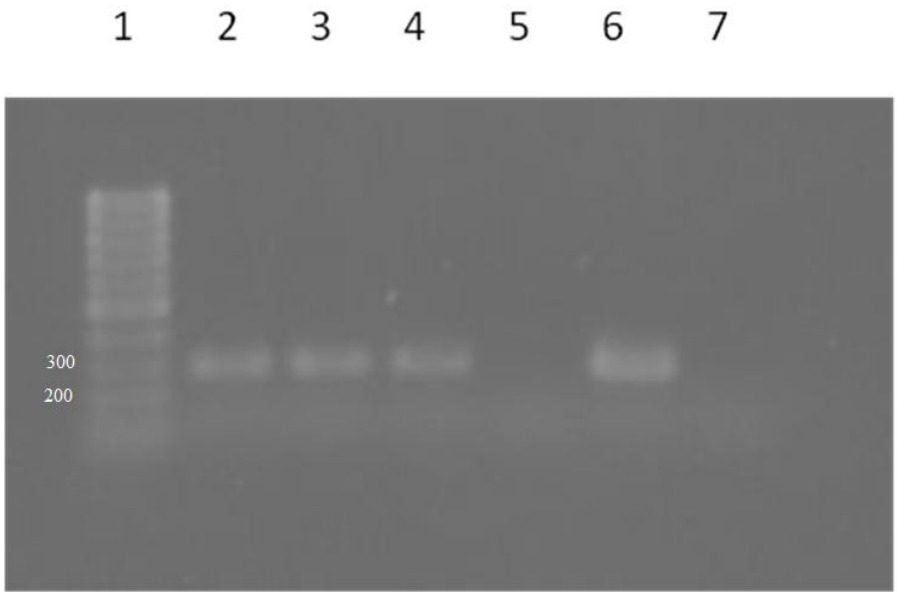
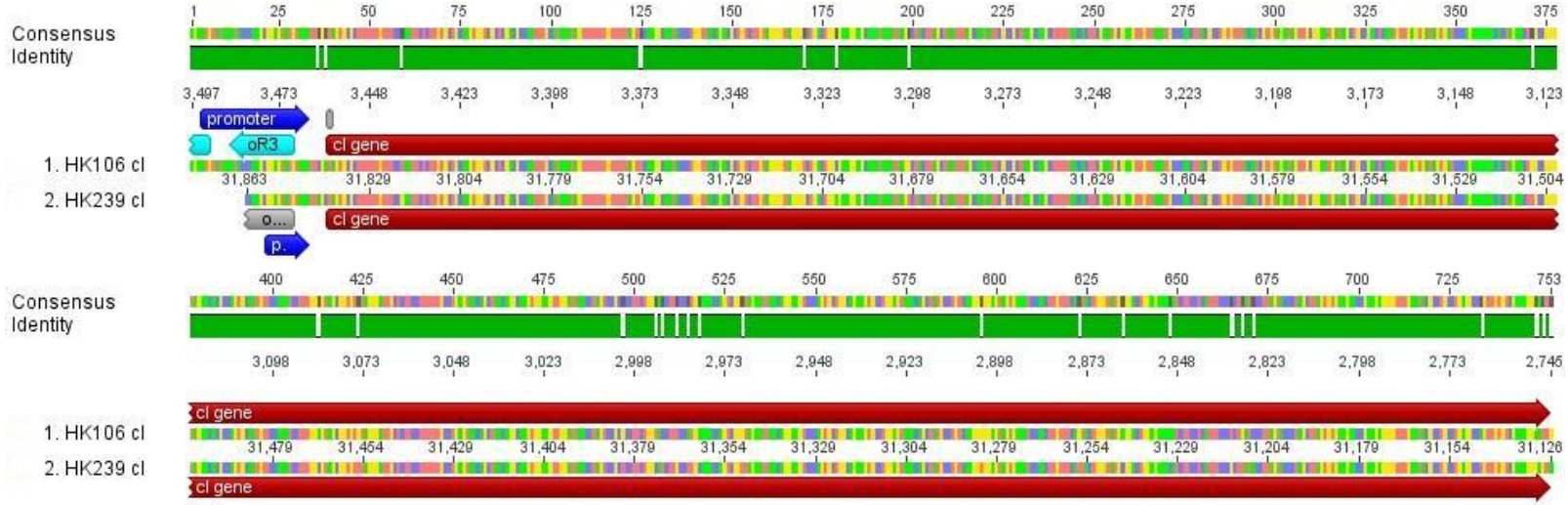


Figure 9. Confirmation of HK106 lysogen. Purified suspected lysogens were screened for the presence of the HK106 prophage by PCR with primers 424 and 469. Lane 1 contains a 100 bp ladder. Lanes 2-4, show an expected band of 287 bp. Lane 5 did not contain any product, indicating the colony did not contain 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.

A


Consensus
Identity
promoter
oR3
cl gene
1. HK106 cl
2. HK239 cl
o...
p.
cl gene
Consensus
Identity
cl gene
1. HK106 cl
2. HK239 cl
cl gene

B

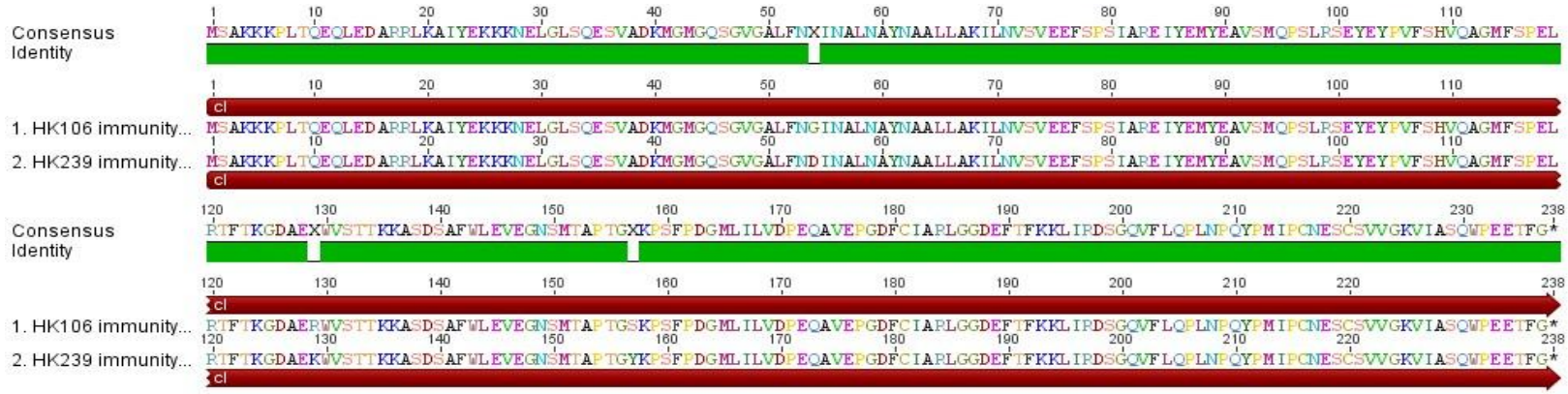

Consensus
Identity
cl
1. HK106 immunity...
2. HK239 immunity...
cl
Consensus
Identity
cl
1. HK106 immunity...
2. HK239 immunity...
cl

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).

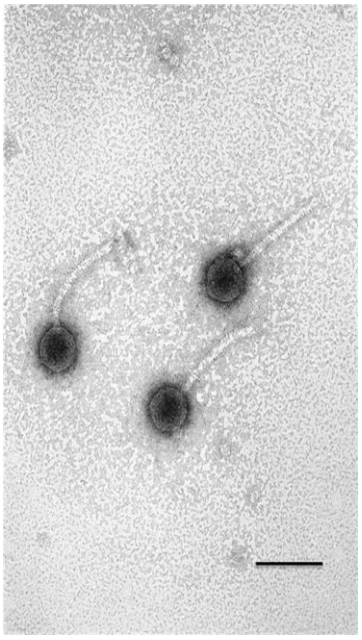
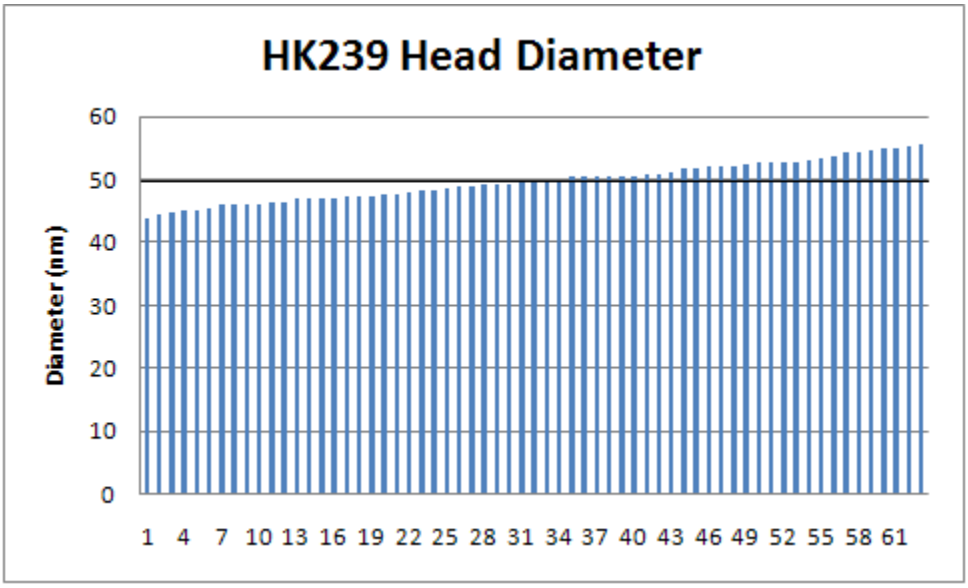


Figure 11. Electron micrographs of Bacteriophage HK239. The phage were deposited on formvar coated grids and stained with 1% uranyl acetate. The scale bar is approximately 100 nm.

A



B

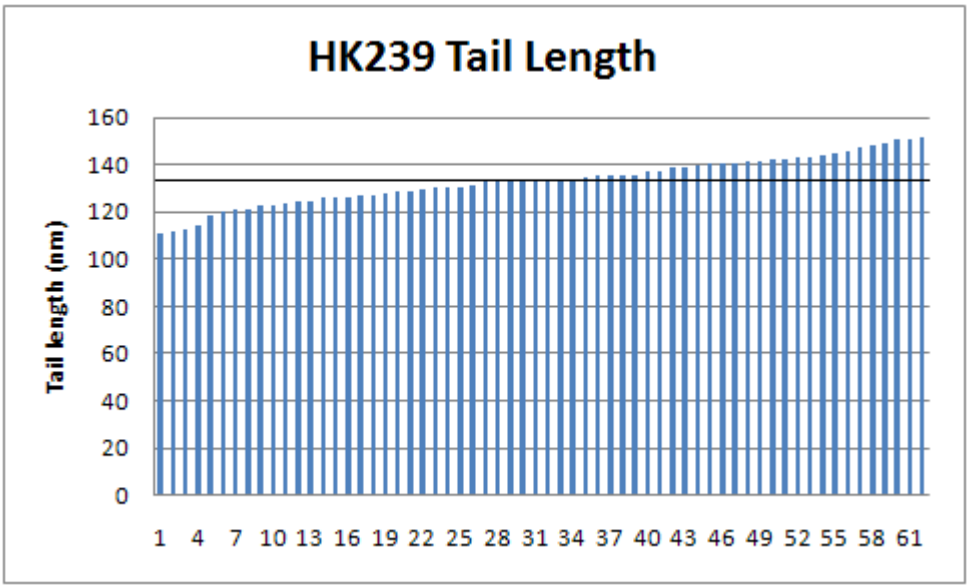


Figure 12. Head and tail measurements of 61 HK239 bacteriophage. The average head length, as indicated by the black line, is 49.7 nm. The average tail length, as indicated by the black line, is 133.2 nm.

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, P_L and P_R, are identical. The promoters that drive repressor synthesis, P_{RM} and P_{RE}, 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 know 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 homoimmune 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 homoimmune, 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

1. **Shigemitsu Hara, Kazuki Terauchi, and Isao Koike** (1991). Abundance of Viruses in Marine Waters: Assessment by Epifluorescence and Transmission Electron Microscopy. *Applied And Environmental Microbiology, Vol. 57, No. 9, pp. 2731-2734.*
2. **Benjamin Lewin** (2006). Essential Genes. *Pearson Education, Inc.*
3. **Max E. Gottesman and Robert A. Weisberg** (2004). Little Lambda, Who Made Thee? *Microbiology and Molecular Biology Reviews, Vol. 68, No. 4, pp. 796-813.*
4. **Robert A. Weisberg, Max E. Gottesman, Roger W. Hendrix, and John W. Little** (1999). Family Values in the Age of Genomics: Comparative Analyses of Temperate Bacteriophage HK022. *Annual Reviews in Genetics, Vol. 33, pp. 565-602.*
5. **A. Skalka** (1966). Regional and Temporal Control of Genetic Transcription in Phage lambda. *Proceedings of the National Academy of Sciences, USA, Vol. 55, No. 5, pp. 1190-1195.*
6. **Louis Reichardt and A. D. Kaiser** (1971). Control of lambda Repressor Synthesis. *Proceedings of the National Academy of Sciences, USA, Vol. 68, No. 9, pp. 2185-2189.*
7. **Tom Maniatis and Mark Ptashne** (1973). Multiple Repressor Binding at the Operators in Bacteriophage lambda. *Proceedings of the National Academy of Sciences, USA, Vol. 70, No. 5, pp. 1531-1535.*
8. **H. Eisen, P. Brachet, L. Pereira da Silva, and F. Jacob** (1970). Regulation of Repressor Expression in Lambda. *Proceedings of the National Academy of Sciences, Vol. 66, No. 3 pp. 855-862.*
9. **Joseph Sambrook and David W. Russell** (2001). Molecular Cloning: A Laboratory Manual 3rd Ed. *Cold Spring Harbor Laboratory Press.*
10. **J. F. Conway, R. L. Duda, N. Cheng, R. W. Hendrix, and A. C. Steven** (1995). Proteolytic and Conformational Control of Virus Capsid Maturation: The Bacteriophage HK97 System. *Journal of Molecular Biology, Vol. 253, pp. 86-99.*
11. **R. Y. Young** (1992). Bacteriophage Lysis: Mechanism and Regulation. *Microbiological Reviews Vol. 56, No. 3, pp. 430-481.*
12. **Matthew K. Waldor and John T. Mekalanos** (1996). Lysogenic Conversion by a Filamentous Phage Encoding Cholera Toxin. *Science, Vol. 272, No. 5270, pp. 1910-1914.*
13. **Itai Sharon, Areilla Alperovitch, Forest Rohwer, Matthew Haynes, Fabian Glaser, Nof Atamna-Ismaeel, Ron Y. Pinter, Frederic Partensky, Eugene V. Koonin, Yuri I. Wolf, Nathan Nelson, and Oded Beja** (2009). Photosystem I Gene Cassettes Are Present in Marine Virus Genomes. *Nature, Vol. 461, pp. 258-262.*

14. **Bettina Hofer, Monika Ruge, and Brigitte Dreiseikelmann** (1995). The Superinfection Exclusion Gene (sieA) of Bacteriophage P22: Identification and Overexpression of the Gene Product. *Journal of Bacteriology, Vol. 177, No. 11, pp. 3080-3086.*
15. **Augusto Uc-Mass, Eva Jacinto Loeza, Mireya de la Garza, Gabriel Guarneros, Javier Hernandez-Sanchez, Luis Kameyama** (2004). An Orthologue of the *cor* Gene Is Involved In the Exclusion of Temperate Lambdoid Phages. Evidence that Cor Inactivates FhuA Receptor Functions. *Virology, Vo. 329, pp. 425-433.*
16. **Alexander A. Vostrov, Olga A. Vostrukhina, Alexander N. Svarchevsky, and Valentin N. Ribchin** (1996). Proteins Responsible for Lysogenic Conversion Caused by Coliphages N15 and φ80 Are Highly Homologous. *Journal of Bacteriology, Vol. 178, No. 5, pp. 1484-1486.*
17. **Janine Robert, Sieghild Bohmer Sloan, Robert A. Weisberg, Max E. Gottesman, Renato Robledo, and Douglas Harbrecht** (1987). The Remarkable Specificity of a New Transcription Termination Factor Suggests that the Mechanisms of Termination and Antitermination Are Similar. *Cell, Vol. 51, pp. 483-492.*
18. **Cornelius Faber, Manuela Scharpf, Thomas Becker, Heinrich Sticht, and Paul Rosch** (2001). The Structure of the Coliphage HK022 Nun Protein-λ-phage *boxB* RNA Complex. *The Journal of Biological Chemistry, Vol. 276, No. 34, pp. 32064-32070.*
19. **Jeffrey Baron and Robert A. Weisberg** (1992). Mutations of the Phage λ *nut*L Region That Prevent the Action of Nun, a Site-Specific Transcription Termination Factor. *Journal of Bacteriology, Vol. 174, No. 6, pp. 1983-1989.*
20. **Hyeong C. Kim, Jian-guang Zhou, Helen R. Wilson, Grigoriy Mogilnitskiy, Donald L. Court, and Max E. Gottesman** (2003). Phage HK022 Nun Protein Represses Translation of Phage λ N (Transcription Termination/Translation Repression). *Proceedings of the National Academy of Sciences USA, Vol. 100, No. 9, pp. 5308-5312.*
21. **Alan Greener and C. W. Hill** (1980). Identification of a Novel Genetic Element in *Escherichia coli* K-12. *Journal of Bacteriology, Vol. 144, No. 1, pp. 312-321.*
22. **Preeta Mehta, Sherwood Casjens, and Sankaran Krishnaswamy** (2004). Analysis of the Lambdoid prophage element e14 in the *E. coli* K-12 genome. *BMC Microbiology, Vol. 4, No. 4.*
23. **Cheng Kao and Larry Snyder** (1988). The *lit* Gene Product Which Blocks Bacteriophage T4 Late Gene Expression Is a Membrane Protein Encoded by a Cryptic DNA Element, e14. *Journal of Bacteriology, Vol. 170, No. 5, pp. 2056-2062.*

24. **Cheng Kao, Effie Gumbs, and Larry Snyder** (1987). Cloning and Characterization of the *Escherichia coli lit* Gene, Which Blocks Bacteriophage T4 Late Gene Expression. *Journal Of Bacteriology, Vol. 169, No. 3, pp. 1232-1238.*
25. **T. Georgiou, Y.-T. N. Yu, S. Ekunwe, M. J. Buttner, A.-M. Zuurmond, B. Kraal, C. Kleanthous, and L. Snyder** (1998). Specific Peptide-Acitvated Proteolytic Cleavage of *Escherichia coli* Elongation Factor Tu. *Proceedings of the National Academy of Sciences, USA, Vol. 95, pp. 2891-2895.*
26. **Yuen-Tsu Nicco Yu and Larry Snyder** (1994). Translation Elongation Factor Tu Cleaved By a Phage-Exclusion System. *Proceedings of the National Academy of Science USA, Vol. 91, pp. 802-806.*
27. **Ryan Bingham, Stephen I. N. Ekunwe, Sherry Falk, Larry Snyder, and Colin Kleanthous** (2000). The Major Head Protein of Bacteriophage T4 Binds Specifically to Elongation Factor Tu. *Journal of Biological Chemistry, Vol. 275, No. 30, pp. 23219-23226.*
28. **Larry Snyder** (1995). Phage-exclusion Enzymes: A Bonanza of Biochemical and Cell Biology Reagents? *Molecular Microbiology, Vol. 15, No. 3, pp. 415-420.*
29. **Karen Matz, Margaret Schmandt, and Gary N. Gussin** (1982). The Rex gene of bacteriophage λ Is Really Two Genes. *Genetics, Vol 102, pp. 319-327.*
30. **Lazarus Astrachan and Joan F. Miller** (1972). Regulation of λ rex Expression After Infection of *Escherichia* coli K by Lambda Bacteriophage. *Journal of Virology, Vol. 9, No. 3, pp. 510-518.*
31. **David H. Parma, Marlene Snyder, Sergey Sobolevski, Myra Nawroz, Edward Brody, and Larry Gold** (1992). The Rex system of bacteriophage lambda: tolerance and altruistic cell Death. *Genes and Development, Vol 6., pp. 497-510.*
32. **Graham F. Hatfull** (2010). Bacteriophage Research: Gateway to Learning Science. *Microbe, Vol. 5, No. 6, pp. 243-250.*
33. **Graham F. Hatfull** (2008). Bacteriophage Genomics. *Current Opinion in Microbiology, Vol. 11, No. 5, pp. 447-453.*
34. **Andrew M. Kropinski, Mark Borodovsky, Tim J. Carver, Ana M. Cerdeno-Tarraga, Aaron Darling, Alexandre Lomsadze, Padmanabhan Mahadevan, Paul Sotthard, Donald Seto, Gary Van Domselaar, and David S. Wishart** (2009). In silico Identification of Genes in Bacteriophage DNA. *Methods in Molecular Biology, pp. 57-89.*
35. **Elvera K. S. Dhillon and T. S. Dhillon** (1973). HK239: A P2 Related Temperate Phage Which Excludes rII Mutants of T4. *Virology, Vol. 55, pp. 136-142*.
36. **T. S. Dhillon and Elvera K. S. Dhillon** (1973). Mutants of Phage HK239 Defective in Excluding Phages λ, T4rII, P1, and P2. *Molecular & General Genetics, Vol. 127, pp. 249-254.*

37. **Ali Wright and Rodney King** (2006) Identifying Exclusion Genes of Bacteriophage HK239. Honors Thesis. Western Kentucky University.
38. **Chris Upton, Duncan Hogg, David Perrin, Matthew Boone, and Nomi L. Harris** (2000). Viral Genome Organizer: A System for Analyzing Complete Viral Genomes. *Virus Research, Vol. 70, pp. 55-64.*
39. **Alexander V. Lukashin and Mark Borodovsky** (1998). GeneMark.hmm: New Solutions for Gene Finding. *Nucleic Acids Research, Vol. 26, No. 4, pp. 1107-1115.*
40. **Carleton L. Kingsford, Kunmi Ayanbule, and Steven L. Salzberg** (2007). Rapid, Accurate, Computational Discovery of Rho-Independent Transcription Terminatiors Illuminates Their Relationship to DNA Uptake. *Genome Biology, Vol. 8, Issue 2, pp. R22*.
41. **A. J. Drummond, B. Ashton, S. Buxton, M. Cheung, J. Heled, M. Kearse, R. Moir, S. Stones-Havas, T. Thierer, and A Wilson** (2009). Geneious v4.8.
42. **Steve Cresawn.** James Madison University. Personal communication.
43. https://www.vwrcanlab.com/literature/products/pdf/12570_lucigen_e._cloni.pdf
44. **John S. Gunn, Celia M. Alpuche-Aranda, Wendy P. Loomis, William J. Belden, and Samuel I. Miller** (1994). Characterization of the *Salmonella typhimurium pagC/pagD* Chromosomal Region. *Journal of Bacteriology, Vol. 177, No. 17, pp. 5040-5047.*
45. **Harald Brüssow and Roger W. Hendrix** (2002). Phage Genomics: Small Is Beautiful. *Cell, Vol. 108, No. 1, pp. 13-16.*
46. **Robert J. Juhala, Michael E. Ford, Robert L. Duda, Anthony Youlton, Graham F. Hatfull, and Roger W. Hendrix** (2000). Genomic Sequences of Bacteriophages HK97 and HK022: Pervasive Genetic Mosaicism in the Lambdoid Bacteriophages. *Journal of Molecular Biology, Vol. 299, pp. 27-51.*
47. **International Committee on Taxonomy of Viruses** (2006). http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/index.htm

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)
AUTHORS Wright,A.A., King,R.A. and Christie,G.E.
TITLE Direct Submission
JOURNAL 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
IRSKALDSWTPADLLAAVELANNQLYITVLRKDLRKEERIRGEERDEGLIKDLRK
QIVELQRTILAQRRDLQIHSHATNGESRDQKKRNQNDRDARNTKNEHQDQDDNL
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"

/translation="MTRGERVIAFIERFCIVPEGKLIGQPMRLDPFQKDFILAVYDNPAGT
DMAILSIARKNGKTGLIAGILLAHLVGPEAVQNTQIVSGALSREQAAIVFNLAVK
MVNLNPKLQEIVHITPSGKKLIGLPCNVEYKALSAEGKTTHGLSPILAILDETGQV
RGPQDDFIDAITTAQGAHENPLLIVISTQAANDADLLSIWIDDAVKSKDPHIVCHV
YEAPKDADISKRESWLAANPALGTFRSEKDMARQAEKAGRMPSFENTFRNLNL
NQRVSTVSPFISRSVWELCGEMPINTPRKWYAGLDLSARNDLTALVIAGEADDG
VWDVFPFFWTPQKTLEERTKTDRAPYDVWVREGLLRTTPGASVDYSFVVADIAE
IIGDFDLTSMAFDRWRIDQFRKDADAIGLSLPLVEFGQGFKDMGPAVDTLESLML
NGRVRHGMHPVLTMCAVNAVVVKDAAGNRKLDKSKATGRIDGMVAMTMSVG
AANGEVTEQGGDFDDFIFRPLSM"

gene 2056..3330
/gene="head portal protein"

CDS 2056..3330
/gene="head portal protein"
/codon_start=1
/transl_table=11
/product="head portal protein"

/translation="MEEPKYTIDLRTNNGWWARLKSWFVGGRLVTPNQGSQTGPVSAH
GYLGDSSINDERILQISTVWRCVSLISTLTACLPLDVFETDQNDNRKKVDLSNPLA
RLLRYSPNQYMTAQEFREAMTMQLCFYGNAYALVDRNSAGDVISLLPLQSANM
DVKLVGKKVVYRYQRDSEYADFSQKEIFHLKGFGFTGLVGLSPIAFACKSAGVA
VAMEDQQRDFFANGAKSPQILSTGEKVLTEQQRSQVEENFKEIAGGPVKKRLWI
LEAGFSTSAIGVTPQDAEMMASRKFQVSELARFFGVPPHLVGDVEKSTSWGSGIE
QQNLGFLQYTLQPYISRWENSIQRWLIPAKDVGRIHAEHNLDGLLRGDSASRAAF
MKAMGESGLRTINEMRRTDNLPPLPGGDVAMRQSQYVPITDLGTNKEPRNNGA"

terminator 3309..3335
/standard_name="terminator"

gene 3348..4025
/gene="head maturation protease"

CDS 3348..4025
/gene="head maturation protease"
/codon_start=1
/transl_table=11
/product="head maturation protease"

/translation="MPEIVKTLSFDETEIKFTGDGKQGIFEGYASVFNNTDSDGDIILPGAF
KNALANQTRKVAMFFNHRTWELPVGKWDSLAEDEKGLYVRGQLTPGHSGAAD
LKAAMQHGTVEGMSVGFSVSKDDYTIIPTGRIFKNIQALREISVCTFPANEQAGIA

AMKSVDGIETIRDVENWLRDSVGLTKSQAVGLIARFKSAIRSESEGDGNEAQINA
LLQSIKSFPSNLGK"
gene 4028..5185
/gene="major head subunit precursor"
CDS 4028..5185
/gene="major head subunit precursor"
/codon_start=1
/transl_table=11
/product="major head subunit precursor"

/translation="MSELALIQKAIEESQQKMTQLFDAQKAEIESTGQVSKQLQSDLMKV
QEELTKSGTRLFDLEQKLASGAENPGEKKSFSERAAEELIKSWDGKQGTFGAKTF
NKSLGSDADSAGSLIQPMQIPGIIMPGLRRLTIRDLLAQGRTSSNALEYVREEVFT
NNADVVAEKALKPESDITFSKQTANVKTIAHWVQASRQVMDDAPMLQSYINNR
LMYGLALKEEGQLLNGDGTGDNLEGLNKVATAYDTSLNATGDTRADIIAHAIYQ
VTESEFSASGIVLNPRDWHNIALLKDNEGRYIFGGPQAFTSNIMWGLPVVPTKAQ
AAGTFTVGGFDMASQVWDRMDATVEVSREDRDNFVKNMLTILCEERLALAHY
RPTAIIKGTFSSGS"
terminator 5192..5213
/standard_name="terminator"
gene 5219..5545
/gene="ORF1"
CDS 5219..5545
/gene="ORF1"
/note="Similar to Bacteriophage HK022 gp6"
/codon_start=1
/transl_table=11
/product="unknown"

/translation="MAIDVLDVISLSLFKQQIEFEEDDRDELITLYAQAAFDYCMRWCDEP
AWKVAADIPAAVKGAVLLVFADMFEHRTAQSEVQLYENAAAERMMFIHRNWR
GKAESEEGS"
gene 5545..5781
/gene="head-tail adaptor protein 1"
CDS 5545..5781
/gene="head-tail adaptor protein 1"
/codon_start=1
/transl_table=11
/product="head-tail adaptor protein 1"

/translation="MEPGRFRHRVKILTFTTSRDPSGQPVESWTGGNPVPAEVKRISGREQ
LSGGAETAQATIRVWMRATERVDKKVGKNSL"
gene 5778..5975
/gene="ORF2"
CDS 5778..5975
/gene="ORF2"
/note="Similar to Bacteriophage HK022 gp8"
/codon_start=1
/transl_table=11
/product="unknown"

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

/translation="MQAGKLRHRITLQEPVKVQNPTTGAVINTWRDVATVRAEVSPLSA
REFIAAQASQGEITTRIVIRYRAGVTRKHRILFRGAVYNIHGVLPDPKSGREYLTL
PCSEGVNDG"
gene 6302..6841
/gene="ORF3"
CDS 6302..6841
/gene="ORF3"
/note="Similar to Bacteriophage HK022 gp10"
/codon_start=1
/transl_table=11
/product="unknown"

/translation="MADSVEVSLTGLESLLGKMEAVSEVTRNKAGRFALRKAANIIRDRA
RSNASRVDDPLTKEAIHKNIVASFSSKQFRRTGDLAFRVGVMGGARQYANTKAN
VRKGRAGKTYKTSGDKGNPGGDTWYWRFLEFGTEHTSARPILRPAMNGVDND
VINVFSTEMGKAIDRAIRLAMKKGTTA"
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="MIAPIFSVCASSPEVTDLLGSNPVRIYPFGIQDDNVVYPYVVWQNIT
GSPENYIAQRPDADFFTLQVDAYADTVDEVIAVATALRDAIEPHAHITRWGGQE
RDPETKRYRYSFDVDWIVTR"
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="MSVLTQGTQLFVLVKGKVSEVECITAFSPGSNPADQIEDTCLSERFD
RSYKRGLRTPGTASLTLNADPKNTSHIMLYNLSISDDEEDQDLTFAIGWSDGTAS
PTAAENGASGAVDGLVLPDSRTWFVFKGYVSDFPFDFSANTVVSTSASIQRSGSA
VWVPKVVTP"
terminator 7764..7785
/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="MKLTLDALKESGAFTGRPVEKEITWTQGDKKITATVYVRPMGYHT
ATSDVLAFGGKVDGVAGRIAASICDEHGKPIFTPADITGEADPERGALDGGLTVA
LLLAIQEVNDLGKTSSSAPKTNSGASSSSTESEAEQSPKRGKSSHSKSPNSGQSTG
NDTGV"
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="MLRVEWGAGLVSSMIANVNRDPKRPPFNPTDFTLHFTKVKAADGPI
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
AAAMTGFATAVGAAAIGAGVAGYNLLKTTSRQITESDRWAKSLNMSTQSLLAW
QYAAEKAGVSGDQMADIFKDVGDKIGDAVLNKSGEAVGALDALGLSAKKLAG
ESPDKQLLAISDALEKVKSNAEKTTILESLGNDLSKMLPLLDNGSEKLRQYMDAA
KKFGVAPDDADIEKLVRVNALFEDMETQVNGVKIELAAGLASVDLSGLQKSIGD
MGDVFKDPAVIQGLTDLVGGVVDLATWLVRVGAEAGKLIDQYKGGSSVGLNAS
IPEIERRIKNLNADLDDKGILASFNRIGMDVSGKEAERAELQRRLAFLRNSQSTLP
KIKLPEPVKTNYTLGAGETNGKPQKNTSGQKLDSAFKSAERSYMRQIELIDTTGK
KTAVVTEQQKLQFDIADGKLQGLNETQKKRLASLAQEVDRLNAVKKANEENAK
VAAFVANLQEQNENARADLGVDIQGAGLGDKQRERLRERLSIERSYLDQQRDLQ
KQYQSGDISQTVYDRETQALKDAQAERLGIQEDYYSQIDALQSDWVTGARDGL
ADWVDDSTNYATLAADAMKSALSGISSNIVDMLNGNKASWKDWGVSVLKIIEQ
VMVNMMIANAASSIGSLFGGAASSSASSGTAIQSYGASLQFNAKGGVYSSADLS
QYSNSVVSSPTLFAFAKGAGLMGEAGPEAIMPLTRAADGSLGVRAMGISGLTPG
GSSAPQVSIQIDGNGNTQTQASGGYEQFGREVGSFVDRRYRELIGRDLSPGGAV
WNLAKGGR"
gene 10898..11236
/gene="minor tail protein"
CDS 10898..11236
/gene="minor tail protein"
/codon_start=1
/transl_table=11

/product="minor tail protein"

/translation="MAIETFSWCPRPNAEQEVTFRRRTAQFGDGYQQVSGDGINPRSQK
WTLQFTGTETYIGAIKDFLDRHAGVTAFQWRPPLEPLGLYRCDTYTPTPLGAGLF
NLSATFEQAYKP"

gene 11233..11988
/gene="minor tail protein L"

CDS 11233..11988
/gene="minor tail protein L"
/codon_start=1
/transl_table=11
/product="minor tail protein L"

/translation="MSLNADFQKLEPGDVVRLFEVDGTAFGTGDVLRFHSYSLAHSEAEII
AAGGDENKLPAKSIWWQGEEYKAWPCQIEGIEASTSGSSAQPKLSVANLDSSITA
LCLAYDDMLQAKVTIHDTLGKYLDARNFTGGNPTADPTQEKLKVFYIDAKSSEN
NEVVEFTLSSPMDLQGLMIPTRQLHSLCTWCIRNKYRTGDGCDYAGTRYFDKNN
NQVSDPSLDECNGTLTACKLRFGENNELSFGGFPGTSLIRS"

gene 11990..12700
/gene="tail assembly protein 1"

CDS 11990..12700
/gene="tail assembly protein 1"
/codon_start=1
/transl_table=11
/product="tail assembly protein 1"

/translation="MRQKTIDAIMAHAAAEYPRECCGVVAQKSRVEKYFPCSNLATEPTE
HFHLSPEDYAAAEDWGTVIAIVHSHPDATTQPSELDKAQCDATLLPWHIVSWPE
GDLRTIQPRGELPLLERPFVLGHFDCWGLVMSYFRQTHGIELHDYRVDYPWWEN
AYPDNFYQDCWYECGFREFDGPPQEGDLVIMQVQADKWNHAGILLEGNMLLH
HLYGHLSQRVPYGGYWQERTMKILRYKSLC"

gene 12730..13071
/gene="ORF7"

CDS 12730..13071
/gene="ORF7"
/codon_start=1
/transl_table=11
/product="unknown"

/translation="MKKTLLTLSLIIMAGCSSMQDLRKEPASNTFQSKKQIDAVAECILSG

WQEESQKYGSVFIQPYDGGKTVFTQSQLEMVDLISDGGITKIEFRHQGGLFAYRI
NSRIKVIERCI"
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="MKEVMTTIQLGGVLGKTFGRTHQRLIARTGEAAIALSKTLPGFESF
MISSKRRGLTFAVFKGKRNIAADEMGFPSEGDVVRIMPVIIGSKRAGLLQTILGAV
LITAAVLTGPGGIGAAFAAGGLTGFAAATGASLVLGGVIQLLSPQPSGIASKQSAD
NRASYAFGGVTNTAAQGYPVPLLYGKRRIGGAIISAGIYVEDQQ"
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="MATDKVLKGRKGGSSSSRTPTEQPDDLQSVAKAKILVALGEGEFAG
QLTGKDIYLDGTALENADGSQNFSGVTWEFRAGTQAQKYIQGIPGTENEISVGTE
VSSATAWTRTFTNTQLSAVRLRLKWPSLFKQEDDGDLVGYSVNYAIDLQTDGGT
WQTVLNTSVTGKTTSGYERSHRIDLPQAGSTWTIRLRKITSDANSAKIGDTMMLQ
SFTEVIDAKLRYPNTALLYVEFDSSQFNGSIPQISCEPRGRVIRVPDTYDPETRTYS
GTWTGAFKWAWTDNPAWIFYDLVVSDRFGLGHRLTAANIDKWTLYQVAQYCD
QMVPDGKGGNGTEPRYTCNVYIQDRNDAYTVLRDFAAIFRGMTYWGGDQIVAL
ADMPRDVDYSYTRANVVGGRFTYSSSTTKSRYTTALVSWSDPGNAYADAMEPV
FEQALVARYGFNQLEMTAIGCTRQSEANRKGRWGILTNNKDRVVSFDVGLDGNI
PQPGYIIAVADELLSGKVMGGRISAVNGRVIKLDRVADAAPGDRLILNLPSGASQ
SRTIQAVNGESVTVTTAYSETPQAEAVWVVESDELYAQQYRVVSVSDNNDGTFS
ITGAWHDPDKYACIDTGAIIDQRPVSVIPPGNQSPPANIVISSFSVVQQNISVETMR
VSWDQAQNAIAYEAQWRRNDGNWVNVPRTSTTSFDVPGIYAGRYLVRVRAINA
AEISSGWGYSEEKTLTGKVGNPPKPVGFIASENVVFGIELNWGFPANTDDTLKTEI

QYSLTGSEDDAILLSDVPYPQRKYQQMGLKAGQIFWYRAQLVDRTGNESGYTD
WVRGQASIDVSDITDVILEDIKESDTFKELIESAVDSNEKIAGMADDIRQNADDLE
QQALAIKENADGLAQAEVKIDEISVSMDGMTGGVKNSSIAVIQNSLAQVTSRRSQ
TATNAGNSASIDRIDTTIADTSQAVARALVTLDASAGGNVSNATDLTETLADFTQ
ASATKINSLTVTVNGQTAAINQTAQAVADVNGNLSAMYNIKVGVSSNGQYYAA
GMGIGVENTPSGMQSQVIFLADRFAVTTMVGGTVTLPFVIQNGQAIIRDTVIGDG
TISNAKIGNYIQSNNYVAGSVGWKLDKSGTFENYGSTAGEGAMKQTNQTISVKD
DNNVLRVQFGRLTGVF"

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="MAYGIQTWDASGKPNNYGIKPVSVVGRIQLAAGQNSGSWSFTVPS
GMKVGFALSLDEGGNSVGRSIVASGNTITVTAASSVGLGNYPASKCEVVVFMEK
A"

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="MAEFGAMILMDNGNPFVTPQSTPFCLYGKYTFNSSANGSSQQVAQ
NIALNADYPVMVFIKTTNTAQPTPVMSYRNGGNVYVAGVNPYNQSFTLTSYVFA
IFPQILPKWGLAIWDASGKLVLTNESRVLSDLQTVGTPGANGGINIDQTLSGSWA
VAPAQLGQTIIVNNSTKPPTIYTINAYSSCRFDGANTRINAGGTSTGAGSPGGGTN
TGISLTAINTAAYD"

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

/product="cor"

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

/translation="MLYNTGTIAINGNTATGKGTNWTAPASQVRAGQTIIVMSNPVQLFQ
ISSVNSATSMTVTPAASPALSGQKYAILVSDIISVDGLAQAMSQLIKEYDENIGAW
ETFATTSANQTITVTINGTTVTIPGIGKLAQKGSNGAVTVADGGTGATTAEGSRT
NLGLGNSATRNVGTAAGTVAAGDDSRFEKIAKLGSAATKDTGEGEGNVLITGSF
GVGSKILPVISDIWDKSQGSRFCNVTPATLGGPGIYGSGIRLSDRNIGSGATYAAQ
QSFFALIFSGKVIQFMGMTDGVDTGWMKIYHTGNTTRASDGTLKAASPIVRLFG
NGKCQLNDESEGCTVTRLATGKYLVEGCEGLNSDAAWGGIDGGFDIPTDRNKQP
LIWLDYEVNADGSVLIQTYHRTHPSAPAYARNERDGINDGEPIDIPSDQFVSVRV
EMPANSIWNQKQKAIEEAAKSASEEVQ"
gene complement(20035..20307)
/gene="msgA"
CDS complement(20035..20307)
/gene="msgA"
/codon_start=1
/transl_table=11
/product="msgA"

/translation="MVMFVELVYDKRNVEGLKGAREIILAELKKRVHQIFPDAEVKVKP
MQANGLNSDASKSDREKLNRMLEEMFEESDMWLVSEFPTVLQNGV"
gene 20372..20626
/gene="ORF10"
CDS 20372..20626
/gene="ORF10"
/codon_start=1
/transl_table=11
/product="unknown"

/translation="MRAILMPRKSDIHAAFLASIEQNQKGYLCLNTNKFINKLREKSWHFS
QADANTWIERYQPDFADKTTNGSQNRYWILRNMGRVF"
gene 20626..21015
/gene="DNA pol V subunit"
CDS 20626..21015
/gene="DNA pol V subunit"
/codon_start=1
/transl_table=11
/product="DNA pol V subunit"

/translation="MGFPSPATDYVEQRISLDERIITRPAATYFMRAGATHYREGILNGAL
LVVDASMSPCDGSLLVCTDSGEFRIKRYRTHPRPHLENLENGKRESLLDKDEVSD
TSRPVFGVITYIINDARSGEFDDYPLK"
gene 21329..22507
/gene="integrase"
CDS 21329..22507
/gene="integrase"
/codon_start=1
/transl_table=11
/product="integrase"

/translation="MAISDTKLRTIYGKPYSGPQEVADADGLSVRISPKGVIQFQYRYRW
HGKPNRLGLGRYPSLSLKDARQITADLRNLYFSGTDPRTYFEEKVENSMTVAQC
LDYWFDNYVSTTLREKTQALYRSTVMKRMHDAFPNRPASSITVKQWVDLLTEE
ERDNPRRARQVLSQLRSAISWCMRRQLIDSCAIMSIQPRDFGSRAEVGDRVLSYH
ELAKIWLAIERSRASTSNKLLHQMLMLWGARLSELRLAKKTEFDLLENVWTVPK
EHSKMGNVIRRPIFEQIKPFLEKAMTTYNDVLFPGEDINKPISIAAANRFVNRIRGG
MDLGYWRTHDFRRTLVTRLSEMNVEPHVTERMLGHELGGIMSVYNKHDWIEA
QRKAYELHADKLFWHIRSISD"
gene complement(22488..22679)
/gene="AlpA1"
CDS complement(22488..22679)
/gene="AlpA1"
/codon_start=1
/transl_table=11
/product="AlpA1"

/translation="MTDTSLIPEKEVMNKLGVSSRQTIWNYTKRHGFPKPVRTHPKSYLR
EAVEGWILNGGVNQKCS"

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="MHFRVTGEWNGEPFNRVIEAENINDCYDHWMIWAQIAHADVTNIRI
EELKEHQAA"
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="MELSRDKIKEIINDDWLLMDDCEGNKNCDVVKEMARIALASLEAEP
VAWLLSGGGAKNNVSFDSGNAYADPLREVTPLYTAPPAPVSVPAAMEIDDDFDS
AFEHGKAVGWNAYRASMLQAEPVSNSDELPLDYLQGHKDGLEWAAQLAEANH
PQTGDWLYDDQIDLARAIRKGPDMPTVQGGNSPVIPDGWISCSEQMPVIGELNW
RTSFPLLVTCEIGVIPAYYGFVSVNGDRHYGFMESLKYGDDNGNHPQTNEYGLIS
NVTHWMPLPEPPQEVNRG"
gene complement(23817..24017)
/gene="ORF13"
CDS complement(23817..24017)
/gene="ORF13"
/codon_start=1
/transl_table=11
/product="unknown"

/translation="MANSLLETCNDWQILRAEILARNPNMAMTLQKLDIMIEHSVRAAIEI
SHRVDWDFSEAERKAKAVN"
gene complement(24010..24324)

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

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

/translation="MSQIDYQALRAKAEKATCGEWSLEYGDGRFDGDDALIHREAAGYI
PICRIEGAHPESGFDEDFQMEQQANAEFIAASNPATVLALLDERERNQQYIKRRD
QENEEIALTVGKLRVELEGKDSKIANLTAERDALREGEMGDARHSNTRAAADIY
FQLVEECEIPAGGSLVEYVDDMREKLEAAEKRIAELELREVVLPQCYSMLHRVD
FDEPYHTEMVYRQHQVLEALHNAGINVTEACKGEAS"
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"

/translation="MKQMSLIEMDGFLKGKCIPRDLKVNETNAEYLVRKFGELESKLETA
LRECRSAGITIDNLEAKCAALDAEKEKFAVECAATKIAIAYLKSGRQDFSLNTTA
TDAFLAEVRAQGVEMYADNLDNGADDAERGGFDYAVRFLRSEASGVRLFADQ
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="MRGLAYNPGILPAEMIIRQRVKPMPSREELLKRNSFGSVNDNKYLN
AMWRSGKK"
gene complement(25706..25999)
/gene="abc2"
CDS complement(25706..25999)
/gene="abc2"
/codon_start=1
/transl_table=11
/product="abc2"

/translation="MPAPLYGADDARRCSGNSVSEVLDKFRKNYDLIMSLPQETKDEKEF
RHCIWLAEKEERERIYQTSIRPFRKATYTHFPEIDPRLRNYRSRYGAISND"
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="MASRGVNKVIIIGHLGHDPEIRYSPSGTAFANITVATSEQWRDKQTG
EQKEQTEWHRVVMSGKLAEIASEYLRKGSEVYLEGKLRTRKWQDQSGQDRFTT
EVIVGVGGTMQMLGGKQGGNEQSSPQRNNGQQQRQQSQQQGNHSEPPMDFDD
DIPFAPVTLPFPRHAIHAI"
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="MDLNKFDEPFCPEDIEWRIQQSGKTRDGKVWAMVLAYVTNRAIM
KRLDDVCGKAGWRNEYRDIPNNGGVECGISIKIDSEWVTKWDAAENTQVEAVK
GGRSGAMKRAAVQWGIGRYLYKLEEGFAQTSLDKKHGWHRAKLKDGTGFYW

LPPSLPGWAIPASDNKPSPENTNQKSPSVDCEQILKDFSDYASTETDKKKLIERYQ
RDWQLMAGNEEAQAKCVQVMNIRVNELKQAA"
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="MSFTDNWSDEEFIRQMKELIGNEGDIHVTCNHSEGEQVTETHVHAE
SSLVSP"
gene complement(27491..27643)
/gene="kil protein"
CDS complement(27491..27643)
/gene="kil protein"
/codon_start=1
/transl_table=11
/product="kil protein"

/translation="MRNEIAINHQMLRAAQNKAVIARFIGDSKMWFEANKAMKSAINIP
WYRRK"
gene complement(27628..27759)
/gene="cIII"
CDS complement(27628..27759)
/gene="cIII"
/codon_start=1
/transl_table=11
/product="cIII"

/translation="MIYAIAGGARMGAFQLNESLLERITRKLRDGWKRVEVLLCAMK"
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="MSEVTDLVVIEKANAMIVFQSADQIEEILQKVEREVMSFVPDITTAK
GRKEIASLAYKVAQTKTYLDGLGKDLVAELKEIPKLIDANRKTVRDRLDELKAK
ARQPLTDYEEEQARIKAEEEAKAAAEALAKQIESDHEIAILMDREFDRQREEARL
KAEQEKREHEERLKREAEEKARAEAEAKAKAEIEAAARREAEAKAAAERAERE
RIEAEQRAQREAKEAAERAEREKQAAIEAERRKAQEEAERIRRDAEAKEQARIAE
EKRIKEEEERRAKDKAHRKEVNNKILADLIKVGASEDVAKNIITAIVKGEVFATKI
TY"

terminator complement(28765..28793)
/standard_name="terminator"

gene complement(28920..29162)
/gene="ORF20"

CDS complement(28920..29162)
/gene="ORF20"
/codon_start=1
/transl_table=11
/product="unknown"

/translation="MRVKTMGASPLSGRIFQGTLNTEKGMWVGKKEDVTEQAVKAVAE
HLMIKDQKYAYETKDGKWLIISHQLVDKLPEDFIAD"

gene complement(29227..29451)
/gene="ORF21"

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

/translation="MNQTYIPSCLRNLPKQKAKPRKQAIKDAKAEVIDQAIQLLREELRSG
KLEGMMMPYQRGYLSAISKLEVLKSEL"

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="MTRRTQFKGNSRSRRRERLKAKALANGVLAREEAISSEVLHRPTLS

RAQIQAKGTHETPERIEDAKPIKFMAQDVIWQQKEYRRNLERAAIVYANEFGHK
QPETGVCLPNVAIYAAGYRKSKQLTAR"
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="MTNRGRVPHTHFSMLNELTFNLVAPLEQAGYTLPEKMVPDISEGRV
FSQWLRDNRGVEPKTFPTYNHEYPDGRTFPVRLYPNEYCRFQTILQRSVAASVR
S"
gene complement(30754..31095)
/gene="ORF23"
CDS complement(30754..31095)
/gene="ORF23"
/codon_start=1
/transl_table=11
/product="unknown"

/translation="MTQFQLALIAREVDGEVIHLRTKDGYINATAMCKSAGKLLADYTRL
KTTQDFFDELSRDMGIPISELIQSFKGGRAENQGTWVHPDIAINLAQWLSPKFAV
QVSRWVREWMSG"
gene complement(31127..31840)
/gene="cI"
CDS complement(31127..31840)
/gene="cI"
/note="Similar to Bacteriophage HK106 cI"
/codon_start=1
/transl_table=11
/product="cI"

/translation="MSAKKKPLTQEQLEDARRLKAIYEKKKNELGLSQESVADKMGMG
QSGVGALFNDINALNAYNAALLAKILNVSVEEFSPSIAREIYEMYEAVSMQPSLR
SEYEYPVFSHVQAGMFSPELRTFTKGDAEKWVSTTKKASDSAFWLEVEGNSMT
APTGYKPSFPDGMLILVDPEQAVEPGDFCIARLGGDEFTFKKLIRDSGQVFLQPLN
PQYPMIPCNESCSVVGKVIASQWPEETFG"
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=11

/product="cro"

/translation="MEQRITLKDYAIRFGQTKTAKDLGVYQSAINKAIHAGRKIFLTINAD
GSVYAEEIKPFPSNKKTTA"
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"
CDS 32279..32575
/gene="cII"
/note="Similar to Bacteriophage HK106 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="MTKRRKKYQEKEEIRHPDSPEGLVVAAANNRAFAERLVGVYRLAK
AGVKHGRR"
gene 32756..33577
/gene="O"
CDS 32756..33577
/gene="O"
/codon_start=1

/transl_table=11
/product="O"

/translation="MGVVKLADYRPQLEVVEHRVADTEDGFMRVANEITDSLLMADLT
VRQMKVMLAIMRKTYGFNKPMDRLTNTQIAAMTGIHHTHVCAAKRQLIERKFLI
ADGVKIGVNKVVSQWISQDSLTLAKTANKTLAKSANGYKPSQLNTKDNIQKTIN
TNTPLPPNGGGDGQVKPERRKAERIDYESFLNAYNTEVGDRLPHAVAVNEKRKR
RLKKIIPQLKTPNVDGFRAYVRAFVHQAKPFYFGDNDTGWTADFDYLLREDSLT
GVREGKFADRGIA"
gene 33574..34950
/gene="P"
CDS 33574..34950
/gene="P"
/codon_start=1
/transl_table=11
/product="P"

/translation="MKQDIEASVIGGLLIGGLTPTASDVLATLEPEAFSIPLYRKAFEVIRK
QARNRNLIDALMVAEACGEEHFTSILMTSKNCPSAANLKGYAGMVADNYHRRL
VLEIMDEMREPIQSGTIDASSQAMDELVKRLSAIRKPRDEVKPVRLGEIITDYTDT
LDRRLRNGEESDTLKTGIEELDAITGGMNAEDLVIIAARPGMGKTELALKIAEGV
ASRVIPGSDVRRGVLIFSMEMSALQIAERSIANAGRMSVSVLRNPAAMDDEGWA
RVANGMSQLADLDVWVVDASRLSVEEIRSIAERHKQENPNLSLIMVDYLGLIEKP
KADRNDLAIAHISGSLKAMAKDLKTPVISLSQLSRDVEKRPNKRPTNADLRDSGS
IEQDADSIIMLYREAVYDENSSAAPFAEIIVTKNRFGSLGTVYQRFCNGHFVACD
QDEARQICTSSNAPAARGRRYAQGADV"
terminator 34976..35007
/standard_name="terminator"
gene 35024..35464
/gene="ninB"
CDS 35024..35464
/gene="ninB"
/codon_start=1
/transl_table=11
/product="ninB"

/translation="MKKLTFEIRSPAHQQNAIHAVQQILPDPTKPIVVTIQERNRSLDQNR
KLWACLGDVSRQVEWHGRWLDAESWKCVFTAALKQQDVVPNLAGNGFVVIG
QSTSRMRVNEFAELLELIQAFGTERGVKWSDEARLALEWKARWGDRAA"
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
AASDKNALCAHWLTEADDALNSEWVSHGAIWNNPPYSNIRPWVEKAAEQCIQQ
RQTVVMLVPEDMSVGWFSKALESVDEVRIITDGRINFIEPSTGLEKKGNSKGSML
LIWRPFISPRRMFTTVSKSALMAIGQGARRAA"
gene 35985..36167
/gene="ninE"
CDS 35985..36167
/gene="ninE"
/codon_start=1
/transl_table=11
/product="ninE"

/translation="MRRQRRSITDIICENCKYLPTKRSRNKRKPIPKESDVKTFNYTAHLW
DIRWLRERARKTR"
gene 36164..36334
/gene="ninF"
CDS 36164..36334
/gene="ninF"
/codon_start=1
/transl_table=11
/product="ninF"

/translation="MIDPNRSYEQQSVERALTCANCGQKLHVLEVHVCEHCCAELMSDP
NSSMYEEEDDE"
gene 36327..37049
/gene="roi"
CDS 36327..37049
/gene="roi"
/codon_start=1
/transl_table=11
/product="roi"

/translation="MNELINGNAIKMTSIEIAELVGKRHDNVKRTIETLAKNGVIRLPQIEV

SERINNLGFNVQYEHYVFEGEQGKRDSIVVVAQLSPEFTARLVDRWRDLEEAAV
NIPKTLPEALRLAADLAEQKMQLENQLAIAAPKVEFADRVGEASGILIGNFAKVV
GIGPNKLFAWMRDHKILIASGSRRNVPMQEYMDRGYFTVKETAVNTNHGIQISF
TTKITGRGQQWLTRKLLDNGMLKVTGEAA"
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="MANLRKEARGRECQVRIYGVCNGNPETTVLAHYRMAGICGTGMK
PDDLIGAWACSACHDEIDRRTHNLDNKDAKLYHLEGVIRTQAILLKEGKIKP"
gene 37336..37698
/gene="RusA"
CDS 37336..37698
/gene="RusA"
/codon_start=1
/transl_table=11
/product="RusA"

/translation="MNEYQFVLPYPPSVNTYWRRRGSQYYISAKGQKYRKDVQQIIRQL
KLDIFTKSRLRIKVIADVPDSRRRDLDNILKGLLDSLIHAGFAEDDEQFDDIRVIRG
VKVPGGRLGIKITELENA"
gene 37695..37883
/gene="ninH"
CDS 37695..37883
/gene="ninH"
/codon_start=1
/transl_table=11
/product="ninH"

/translation="MNATIQTIPELLIQTRGNQTEVARMLSCARGTVLKYNRDSKGERHV
IVNGVLMVKQGKRGRP"
gene 37880..38503
/gene="Q"
CDS 37880..38503
/gene="Q"

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

/translation="MRLESVAKFHSPKSPMMSDSPRATASDSLSGTDVMAAMGMAQSQ
AGFGMAAFCGKHELSQNDKQKAINYLMQFAHKVSGKYRGVAKLEGNTKAKVL
QVLATFAYADYCRSAATPGARCRDCHGTGRAVDIAKTERWGRVVEKECGRCK
GVGYSRMPASATYRAVTMLIPNLTQPTWSRTVKPLYDALVVQCHKEESIADNIL
NAVTR"
terminator 38611..38645
/standard_name="terminator"
gene 38937..39260
/gene="holin"
CDS 38937..39260
/gene="holin"
/codon_start=1
/transl_table=11
/product="holin"

/translation="MKMPEKNDLLAAILAAKEQGIGAILAFAMAYLRGRYNGGAFTKT
VIDATMCAIIAWFIRDLLDFAGLSSNLAYITSVFIGYIGTDSIGSLIKRFAAKKAGV
EDGGNQ"
gene 39244..39720
/gene="lysin"
CDS 39244..39720
/gene="lysin"
/codon_start=1
/transl_table=11
/product="lysin"

/translation="MVEINNQRKAFLDMLAWSEGTDNGRQKTRNHGYDVIVGGELFTD
YSDHPRKLVTLNPKLKSTAAGRYQLLSRWWDAYRKQLGLKDFSPKSQDAVALQ
QIKERGALPMIDRGDIRQAIDRCSNIWASLPGAGYGQFEHKADNLIAKFKEAGGT
VREIEV"
gene 39717..40154
/gene="Rz"
CDS 39717..40154
/gene="Rz"
/codon_start=1
/transl_table=11

/product="Rz"

/translation="MSRVTAIISALVICIIVCLSWAVNHYRDNAITYKAQRDKATYIIADM
QKRQRDVAELDARYTKELADANATIESLRADVSAGRKRLQVAATCAKSTTGAS
SMGDGESPRLTADAELNYYRLRSGIDRITAQVNYLQEYIRTQCLK"

gene 39916..40101
/gene="Rz1"

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

/translation="MLTRLSKVSVLMFLLGVSACKSPPPVQSQRPEPAAWAMEKAQDLQ
QMLNSIITVSEVESTG"

gene 40191..40466
/gene="ORF26"

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

/translation="MTKLYHRISTFLSGCWAFITSISFAIFSFGSTACSLSRGLWRAISALAP
KFLPEKAVWRIVERMCSESVREKINVFGRHPRNTGALCSPLL"

gene 40846..41199
/gene="HNH endonuclease"

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

/translation="METGMNKELRVYGSRWDKARLRFLQQHPLCVMCEQQGRITPATV
VDHIVPHKLKDALKSGNPLAISKAQLLFWSKENWQPLCKAHHDSTKQRMEKSG
AVIGCDANGYPLDPASHWST"

gene 41199..41402
/gene="ORF27"

CDS 41199..41402
/gene="ORF27"

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

/translation="MKDLIIEYRDGKFVQLAIDGVEMKRVTSIQFSHTVGEDVPTLTVSGH
VWSEYGKGDHKLEQVDKHSA"
misc_feature 41529..41538
/standard_name="cos site"
BASE COUNT 10667 a 10205 c 10617 g 10049 t
ORIGIN
1 gtttgaattt aacgctaacc cgatttttt agttttaagg tgttgacata tggcagataa
61 acgaatccgt tccgacagtt cggcggcagc ggttcaggcc atgaaaaatg cagctgtgga
121 caccatcgat ccgccgtccc atgcaggttt ggagaaaaaa gccgaaccat tctggcatga
181 caatatcaga tcgaaagctc tggacagctg gacgcctgcc gaccttctag ccgccgtaga
241 acttgctaat aatcagctct atatcaccgt tttacgcaag gatttacgca aagaagagcg
301 catacgcggg gaggagcgag acgaaggcct tataaaagac ctccgcaagc aaattgttga
361 gctacaacga actatcctgg ctcagcgccg cgacctccag atccactccc acgcaaccaa
421 cggcgaaagt cgcgaccaga agaaacgtaa tcagaacgat cgtgatgcac ggaataccaa
481 aaacgagcat caggaccagg acgataacct gatcgccttt cccaagcacg gataaaagac
541 tatgacgcga ggtgagcgtg taatagcgtt cattgagcgc ttttgcatcg tgccagaagg
601 caagcttatc ggccaaccta tgcggttgga ccccctttcag aaagatttca tcctggcggt
661 ttacgacaat ccagccggaa cggatatggc gatcctctcc atcgcccgaa aaaatggtaa
721 gactggcctg attgccggaa ttctgctggc tcacctggtg gggcctgaag cggtgcagaa
781 cacgcagatt gtcagcggtg cactcagccg ggaacaggcg gccatcgttt ttaacctcgc
841 ggtgaagatg gttaacctga accccaagct gcaggagatt gtgcacatta cgcccagcgg
901 caaaaagctg atcggcctgc cgtgtaacgt cgaatacaag gctttatccg cagaaggtaa
961 gacgacgcac ggcctttccc ccattctggc cattctcgat gaaacagggc aggttagggg
1021 cccgcaggat gattttatcg atgcaataac tacggcgcag ggggctcatg aaaacccgct
1081 gctaatcgtt atcagtacgc aggcagcaaa cgatgctgac ctgcttagca tctggattga
1141 tgatgcggtc aaatcgaaag atccgcacat cgtgtgccac gtttatgaag cgccaaaaga
1201 cgctgatatc agtaaacgcg agtcctggct ggctgcgaac ccggcactgg gaacattcag
1261 gtcagaaaaa gacatggcgc gccaggctga gaaagctggc cgaatgccaa gcttcgaaaa
1321 caccttccga aacctcaacc tcaatcagcg cgtgtctacc gtatcgccgt ttatctcccg
1381 cagcgtgtgg gagctttgcg gagagatgcc gattaacacc ccgaggaagt ggtacgcggg
1441 gctggatctg tcagccagga acgacttaac ggcgctggtt atcgctggtg aagcagatga
1501 tggtgtctgg gatgttttcc ccttcttctg gacaccgcaa aagactcttg aagagcgaac
1561 caaaacggac cgagcaccat atgacgtttg ggtgagggag gggctgctgc gcaccacgcc
1621 aggagcttcg gtggattact cattcgtcgt tgcggatatc gctgaaatta tcggtgattt
1681 cgaccttacc tcgatggctt ttgaccgctg gcgcattgac cagttcagga aggatgccga
1741 tgccattggg ctgagcctcc cgctggtcga gttcggccag ggctttaagg atatggggcc

1801 agctgtagac acgctggagt ctctgatgct taacgggcgc gtgaggcatg gcatgcaccc
1861 cgtattaacg atgtgtgctg tgaatgcggt ggtggtgaaa gatgctgctg gcaaccgcaa
1921 gctcgataaa tccaaagcaa cgggccgtat tgatggcatg gtcgcaatga caatgtccgt
1981 tggtgctgct aatggggaag ttaccgaaca gggtggtgac ttcgacgatt tcatttccg
2041 accgctgagc atgtgatgga agaacctaaa tacacgattg acctgcgaac caataacggc
2101 tggtgggcaa ggctgaagtc ctggtttgtc ggcgggcgtt tagtcacccc aaatcagggc
2161 tcacagacgg ggcctgtttc ggcccacgga tacctgggcg attcatccat taacgatgaa
2221 cggatactgc aaatttcgac tgtgtggcgc tgcgtgagcc tgatttcaac gctcacggca
2281 tgcttaccgc ttgatgtctt cgaaacagac cagaatgaca accgtaaaaa agtggatttg
2341 agcaatccac tggcgcgact gctgcgctac tcaccgaatc agtacatgac cgcccaagaa
2401 ttcagggagg ccatgacgat gcagctctgt ttctacggta acgcatatgc acttgtggac
2461 cgcaacagcg caggtgacgt gatcagccta ctcccgcttc agtctgccaa tatggatgtg
2521 aaactcgtcg gaaaaaaagt ggtttatcgc tatcaacgcg acagcgaata cgccgacttt
2581 tcgcagaaag agatttttca ccttaaaggc ttcggattca ccgggcttgt aggcctgtca
2641 cccattgctt ttgcctgtaa atcggcaggt gtggcagttg cgatggagga ccagcagcga
2701 gatttctttg ccaacggcgc caagtctccg caaatcctct caactggcga aaaagtgcta
2761 actgaacagc agcgctccca ggtcgaagag aacttcaaag agatcgccgg cggcccggta
2821 aaaaaacgcc tctggattct ggaagcgggt ttttctacat cggcaattgg cgtaactccg
2881 caggatgccg aaatgatggc gtcccgaaaa tttcaggtaa gtgaactggc gcgattcttt
2941 ggcgtaccgc ctcaccttgt cggcgacgtc gagaaatcaa cgagctgggg atcgggcatc
3001 gagcagcaga atctcggctt cctgcagtac accctgcagc cctatatctc caggtgggaa
3061 aacagcattc agcggtggct tattcctgct aaggatgttg gccgcattca tgctgagcac
3121 aacctcgacg gcctgctgag gggagattcg gcatcccgcg ctgcctttat gaaggcaatg
3181 ggagagtcag ggctacgcac catcaacgag atgcgacgaa cggacaacct cccgccattg
3241 ccgggtggcg atgtggcaat gcgccagtcg caatacgtgc cgatcaccga tttaggaacc
3301 aacaaagagc cccgtaataa cggggcttaa tttttatggg ggccgtaatg cctgaaatcg
3361 taaaaacgct gtccttcgac gagacagaaa tcaaattcac cggtgacggt aaacagggga
3421 tttttgaagg ctacgcctct gtttttaata acaccgattc cgatggcgac atcattctgc
3481 ccggggcgtt taaaaacgcg ctggcgaacc agacccgaaa agtggcgatg ttttttaacc
3541 acaggacgtg ggaactgccg gttggcaaat gggacagcct ggctgaggac gaaaaaggcc
3601 tgtatgtgcg cggtcaactt accccaggac acagcggcgc cgctgacctg aaagcggcaa
3661 tgcagcacgg tacggttgaa ggtatgtcgg ttggcttttc cgtttcgaaa gacgattaca
3721 ccatcattcc aacaggccgc atttttaaga atatccaggc tctgcgcgaa atcagcgtct
3781 gcactttccc cgccaacgaa caggctggca tcgcagccat gaaaagtgtc gatggtattg
3841 aaacgattcg tgatgtggag aactggctga gggattcagt cggcctcacc aaatcacagg
3901 cagttgggct aattgcccgg tttaagtcag cgattcggag cgagtccgag ggcgacggaa
3961 acgaagcaca aataaacgct ctgcttcaga gcatcaaatc tttcccttct aacttaggta
4021 aataattatg tctgaactcg ctctcattca aaaagctatc gaagaatccc agcagaaaat
4081 gacccagctt ttcgatgcgc agaaagcaga aatcgaaagc acaggccagg tttccaaaca
4141 gttgcagtcc gacctgatga aagtacagga agagctgacc aaatccggca ctcgcctctt

4201 cgatctggaa cagaaactgg catccggcgc tgagaatcct ggtgagaaga aatccttctc
4261 tgaacgggct gctgaagagc ttattaagtc atgggacggt aaacagggca cctttggcgc
4321 taaaacgttt aacaagtcac taggcagtga cgctgattcg gctggctcac tgatccagcc
4381 tatgcagatc ccaggcatca tcatgccagg cctgcgccgt cttaccattc gtgatctgct
4441 ggctcagggc cgcacttcca gtaacgctct ggaatatgtg cgtgaagagg tgtttaccaa
4501 taacgccgac gtggtggcag agaaagcact gaagccagaa tcggatatca ccttcagcaa
4561 acaaaccgcg aacgtgaaga ccatcgcgca ctgggtgcag gcatcacgtc aggtgatgga
4621 tgatgcgcca atgcttcagt cctacattaa caaccgcctc atgtacggcc tggcactgaa
4681 ggaagagggc cagctgctga acggcgacgg taccggggat aacctggaag ggctgaacaa
4741 agtggcaacc gcctatgaca cctcgctgaa tgccaccggc gacacccgcg ctgacattat
4801 cgctcacgct atttatcagg tgaccgagtc tgagttcagc gcttccggta tcgtcctgaa
4861 cccgcgcgac tggcacaaca tcgcgttgct gaaagacaat gaaggccgct atatcttcgg
4921 tggtcctcag gcattcacca gtaacatcat gtggggcttg ccagtggttc cgactaaggc
4981 gcaggccgcc ggcaccttta ccgtaggcgg tttcgatatg gcctcacagg tctgggatcg
5041 catggatgcc accgtggaag ttagccgtga agaccgcgat aacttcgtga aaaacatgct
5101 gaccatcctg tgcgaagagc gtctggcgct ggcgcattat cgcccgacag caatcatcaa
5161 gggcaccttc tcttctggct catgatggag ggggcgggga aacccgccct tttaacgtat
5221 ggcgatagat gttctggatg taatttccct cagtctgttt aagcagcaga ttgagtttga
5281 ggaagacgac agggacgagc tgatcacgct gtacgcccag gccgcttttg actactgcat
5341 gcgctggtgc gatgaaccag catggaaggt tgcggctgat attcctgccg ccgttaaggg
5401 cgccgttctg cttgtctttg ctgacatgtt tgaacaccgg acggcacaaa gcgaagtgca
5461 gctttatgag aatgcagccg ccgaacgcat gatgttcatt catcgcaact ggcgcggtaa
5521 agccgaatca gaggagggct cctgatggaa ccaggacgat tcaggcaccg ggtaaaaatt
5581 ctcaccttca cgacttcgcg cgatccatct ggtcagccgg ttgaatcgtg gacaggtggc
5641 aacccggtcc cggctgaggt aaagaggatc agcggcagag agcagctttc aggcggcgcg
5701 gaaacggcgc aggcaacgat tcgcgtctgg atgcgcgcaa cggagcgagt ggataaaaaa
5761 gtgggcaaaa atagcttatg aacgagcaac tggtaaacga actgatagca gtgatacggg
5821 aacaaatcac agcgcagagg gagcagacgg aagcgataag ccgtcttgct gaatcaaata
5881 tggcgctttg cgatgtcatt atccagtcgc tggctggtga actagatgaa actgcagagc
5941 aacaaaccta tctgagcggt aaacccaggg ggtaacatgc aggccgggaa attgcgccac
6001 aggattactc tgcaggagcc agtaaaagtg cagaacccga caacgggagc cgtaattaat
6061 acctggcgag atgttgcgac tgtccgggcc gaagtttcac cattgtcagc gcgtgaattt
6121 attgctgcgc aggcttctca gggtgaaata accacccgta ttgttattcg ctaccgtgcc
6181 ggagttacca gaaaacatcg catcttgttt cgtggtgctg tgtataacat ccatggtgtt
6241 ctgcccgatc caaaaagcgg acgtgaatac ctgacgcttc cctgttctga gggtgttaat
6301 gatggctgat agtgttgagg taagtcttac cgggcttgaa tcacttcttg ggaaaatgga
6361 ggccgtatca gaagtcaccc gtaataaagc cggtcggttt gcattgcgta aggcggcaaa
6421 cattataagg gatcgggcca gaagcaacgc gtcacgagtt gatgatcctc tgactaaaga
6481 agcaatccac aaaaatatcg tcgccagctt cagcagcaag caattccgca ggacaggtga
6541 tctggcattc cgtgtcgggg taatgggcgg tgccaggcag tatgcaaata caaaggcaaa

6601 cgtcagaaaa ggcagagcgg gtaaaacata caaaacctca ggagataaag gcaatccagg
6661 aggagatacc tggtactggc gttttcttga gtttggtact gagcacacat cagcgcggcc
6721 tatcctccga cctgcaatga acggcgtaga taatgacgtg attaatgttt tttctacgga
6781 aatgggaaag gctattgatc gcgctatcag gctggccatg aagaaaggaa ccactgcatg
6841 attgccccta ttttttctgt ttgcgcgtcg agtcctgaag tcactgactt actcggaagc
6901 aatccggtaa gaatttatcc tttcggaatt caggacgaca acgtggtcta tccctatgtg
6961 gtctggcaga acatcaccgg ttctccagag aactacattg cccagcgccc tgacgcagac
7021 tttttcacgc tgcaggtgga tgcatatgcc gacaccgtgg atgaagtgat tgccgtggct
7081 actgcgctgc gggatgccat tgagccgcac gcgcatatca cgcgctgggg cggacaggaa
7141 agagacccag aaacaaagcg ctatcgctac tcattcgatg ttgactggat agtcactcga
7201 taaacgtatt attcaaccac cggccttgag ccggtttttt tatgaccgga gataacaatg
7261 tctgtattga cgcaaggtac gcagctcttt gtgctcgtaa aaggcaaggt gagcgaagtt
7321 gaatgtatca ctgcattttc acccggcagc aatccggctg accagattga agacacctgt
7381 ctttctgaac gctttgatcg cagctataag cgtggtcttc gaacgcctgg cacagcatca
7441 ctgacgctta acgctgatcc taaaaatacc agccacatca tgctctacaa cctgtccatt
7501 tcggacgacg aagaggatca ggacctgacc ttcgcgattg gatggtcaga cggaaccgca
7561 tcgccaactg cggctgaaaa tggtgcatcc ggtgcagtcg atggcctggt gctacctgat
7621 agccgcacat ggttcgtatt caaaggctat gtgtccgact tcccatttga tttttcagca
7681 aacacggttg tttctacttc tgcatctatc cagcgctccg gctctgctgt atgggtgcca
7741 aaagtcgtga ccccataaaa tcagggcggc aacgccctta ttaatcagga ttaataatga
7801 aattaacact tgatgcactc aaggagtccg gcgcgtttac cggtcgcccg gttgagaaag
7861 aaatcacctg gacgcaaggc gataaaaaaa tcaccgcgac cgtgtatgtt cgcccgatgg
7921 gttatcacac agcaacgtcc gatgtgctgg cttttggtgg taaggtcgat ggtgtggcag
7981 ggcgtatcgc cgcatccatc tgcgatgaac atggtaaacc catcttcacc ccggctgaca
8041 ttactggtga ggcagatcct gaacgtggtg cactggatgg cggcctgacg gtagcactgc
8101 ttctggccat tcaggaagtt aacgatctgg gaaagacttc gagctcagcg ccgaagacga
8161 attctggtgc gagctcgtcc tcaacggaat cggaggcaga acaatcgccg aagcgcggga
8221 agtcctctca ttcaaagagt cccaactctg ggcaaagtac cgggaacgat acgggagtct
8281 gaaccctatg ttgcgggttg agtggggtgc cgggctggtg agcagcatga tagccaacgt
8341 taacagagac cccaagcgcc caccattcaa cccgaccgat ttcacacttc actttaccaa
8401 agtcaaagct gctgatggac caatctcgtt agaggaagcc agagccagct ggacataatg
8461 ccgccaacgg agagtttatg gcttccaaat cactgggcac gctgacgatc gacctgattg
8521 ccaaagtggg cggttttgtc tcaggcctct cgcaggctga gcgagcttca caaaaatggc
8581 gtaagcaggt aaaggaggat gccgctgccg ctgcagctgc catgaccgga ttcgcaacag
8641 cagtcggggc tgcagccatc ggagccggag tggcggggta taacctgctc aaaaccactt
8701 cacgtcagat taccgaatca gaccgctggg caaagtcact taacatgtcc acgcagtccc
8761 tgttagcctg gcaatatgct gcagaaaaag caggtgtttc cggcgatcag atggccgata
8821 tctttaaaga tgtcggagac aaaatcggtg atgccgttct taataaatct ggtgaggcgg
8881 tcggtgcgct ggacgctctg ggtttgtcgg ctaagaaact ggccggagaa tctcccgaca
8941 agcaactcct ggctatcagc gacgcactgg agaaagttaa gtccaacgcc gagaagacta

9001 ccatcctgga gagtctgggt aatgacctgt caaagatgct gccgctcctg gacaatggca
9061 gtgaaaagct gcgtcagtat atggatgccg cgaagaagtt tggtgtggcg cccgatgacg
9121 cagatattga aaagctggtc agagtcaacg ccctgtttga ggatatggag acgcaggtca
9181 acggcgtcaa aattgaactt gctgcagggc tcgccagcgt agatttgagt gggctccaga
9241 agtctattgg agatatgggg gatgtattta aagacccggc tgttattcag ggtctgactg
9301 atctggttgg tggggtagtg gaccttgcca cctggctggt aagggtagga gccgaggccg
9361 gaaagctgat agaccagtac aaaggtggga gttcggtcgg cctgaatgcc tccattccag
9421 aaattgaacg gcgcatcaag aacctgaatg cagatctcga tgataaaggc atactggcga
9481 gtttcaacag aataggtatg gacgtatccg gtaaggaggc cgaaagggct gagctccaga
9541 ggcgactggc ctttttgaga aactcccaat ctactctacc gaaaataaaa ctgcccgagc
9601 cagtcaaaac aaactatacc cttggcgccg gggaaacaaa cggtaagccg caaaaaaata
9661 cctctggtca gaaactggat tcagcgttta aaagcgctga gcgcagttac atgcgtcaga
9721 ttgagctgat cgataccacc ggcaaaaaaa ctgctgttgt gaccgagcag caaaaactgc
9781 agttcgatat agctgacggc aagttgcagg ggcttaacga aacccagaag aaacgacttg
9841 cgtctctggc tcaggaagtt gatcgcctta atgccgtcaa aaaagctaat gaagaaaacg
9901 cgaaggtagc ggcgttcgtg gcaaatctgc aggagcagaa cgagaatgca cgtgcagatt
9961 tgggcgtgga tattcagggg gccggactcg gtgacaagca gcgtgaaagg ctgagggaaa
10021 ggctgagtat agagcgcagt tatctcgatc agcagcgcga tctgcaaaag cagtatcagt
10081 caggagatat cagccagaca gtttatgacc gcgaaaccca ggcattaaag gatgcacaag
10141 ctgagaggct gggcatccag gaggattatt acagtcaaat tgatgcgctg cagtctgact
10201 gggtgactgg cgccagagac ggtctcgctg actgggtaga tgattcaact aactatgcga
10261 cgctggcggc ggacgctatg aaaagcgcgc tttccggtat cagcagcaat atcgtcgaca
10321 tgctcaacgg caataaagcg agctggaaag actggggtgt cagtgttctg aaaatcatcg
10381 aacaggtgat ggttaacatg atgatcgcca atgcagccag ctcgataggc tcattgttcg
10441 gtggtgccgc atcgtcttcc gccagcagcg gtactgcgat tcagtcctac ggggcgagcc
10501 tgcaattcaa cgccaaaggc ggcgtttact cttctgccga tctcagtcag tacagtaact
10561 ctgtcgtcag ctctccaaca ctgtttgcct ttgccaaagg ggccggattg atgggggagg
10621 ctgggccgga ggctattatg ccactgaccc gcgccgctga cggatcgctg ggcgtgcgtg
10681 ctatgggaat ctcaggctta acaccgggtg gtagtagtgc tcctcaggtc agtattcaga
10741 ttgatggcaa tgggaatacc cagactcagg ctagcggtgg atatgagcaa ttcgggcggg
10801 aggtaggtag ttttgttgat cggcgatacc gcgagttgat aggccgtgac ctttcacccg
10861 gtggcgcggt ctggaatctg gcaaaaggag ggcggtgatg gctatagaaa cattcagctg
10921 gtgtccgcgc ccgaacgcgg agcaggaagt gacattccga cggcgcaccg cgcagttcgg
10981 tgacggatat cagcaggttt ccggcgacgg gattaatccc cgctcgcaga aatggactct
11041 ccagtttacc ggtaccgaaa cgtatatcgg ggcgattaaa gactttctcg atcgccatgc
11101 gggcgtgacg gcgtttcagt ggcgcccgcc gcttgagcct ctcgggcttt accgctgcga
11161 cacctacaca ccaacgccac ttggtgccgg gttatttaat ctttctgcaa cttttgaaca
11221 ggcctacaaa ccatgagtct taacgcagat tttcagaagc tcgaacctgg cgatgtggtc
11281 aggcttttcg aagtggatgg cacggcattt ggtacaggtg atgtgctgcg atttcacagc
11341 tacagtcttg cgcactctga agccgaaatt atcgctgctg gcggtgatga aaataaactg

11401 ccagcaaaat ctatctggtg gcagggtgag gagtataaag cgtggccatg ccagattgag
11461 gggatcgaag cttcaacctc gggaagtagt gcccagccaa aactttcggt tgctaacctc
11521 gacagttcga tcaccgcgct ttgcctggct tacgacgata tgctgcaggc gaaggtgacg
11581 atacatgaca cccttggtaa gtatctcgat gcgagaaact tcaccggagg caatccgaca
11641 gccgatccga cccaggagaa gctgaaggtt ttctacatcg atgcaaagag cagtgaaaac
11701 aatgaagttg ttgagttcac actatccagc cctatggatc tgcaggggct gatgatacct
11761 acccgccagc ttcattctct gtgcacatgg tgcatccgga acaagtaccg aaccggcgac
11821 ggctgcgact atgcaggtac gcgctatttc gacaaaaaca acaaccaggt aagcgatccg
11881 tcactggatg aatgcaacgg aacgctgacg gcctgcaaac ttcggttcgg tgaaaacaac
11941 gaactctcgt ttggtgggtt cccgggtacg tcgctgatca ggagttgata tgcgtcagaa
12001 aaccatcgat gcgattatgg cgcatgctgc agctgaatat cctcgtgagt gttgcggcgt
12061 ggtggcgcaa aagagccggg tagagaaata ttttccatgt agtaacctcg ctaccgagcc
12121 gacagaacat tttcacctgt cgccagaaga ctatgcagct gctgaggact ggggtacggt
12181 gatcgccatc gttcacagtc accctgacgc cactacgcag ccgagcgaac tggataaagc
12241 gcaatgcgac gcaacgcttt taccctggca tattgtgagc tggccggagg gggatttacg
12301 caccatccag ccgcgcggag aactgccact gctggagcgt ccgtttgtgc ttggacactt
12361 cgactgctgg gggctggtaa tgagctattt ccggcaaacg catggtatcg agctccacga
12421 ttaccgggtg gattatccct ggtgggaaaa cgcctatccg gacaattttt atcaggattg
12481 ctggtatgaa tgcggattca gggagtttga cggcccgcct caggaagggg acctcgtcat
12541 catgcaggtg caggccgata agtggaatca tgccgggatt ttactggagg gtaacatgct
12601 gctgcaccac ctgtacggac atctgagcca gcgcgtgccg tatggtggct actggcagga
12661 aaggacgatg aagattctac gttacaaatc tctgtgctaa ccttttgcaa aaccaaaggg
12721 gatagggata tgaaaaaaac attattgaca ctttctttga taatcatggc tggttgttcg
12781 agtatgcagg atctccggaa ggaaccagcg tcaaatactt ttcaatcgaa gaaacaaatt
12841 gacgcagtag ctgaatgcat actctctggc tggcaagaag aaagccaaaa atatggaagc
12901 gtttttattc agccttatga cggtggtaaa actgtattta cacaatctca acttgagatg
12961 gttgatttaa tatcggacgg cggaattacc aagattgaat tccgtcatca aggtgggcta
13021 tttgcttatc gaataaacag ccggattaaa gtaatagaac gctgtatcta accaagactt
13081 aacccgcttc ggcgggtttt tttatggtga gaatatgaaa gaagtaatga caacaattca
13141 gctcggcgga gtgttaggaa agacattcgg tagaacccat caacgactga tagcgcgaac
13201 tggtgaagct gctattgctt taagtaaaac attgcccggt ttcgaaagtt tcatgatcag
13261 cagtaaacgt cgcggattaa ctttcgcggt gttcaaaggg aaaagaaata ttgccgcaga
13321 tgagatgggt tttccctctg aaggcgatgt agtaaggatt atgcctgtaa ttatcggtag
13381 taaacgcgct ggtcttttgc agacaatcct gggggccgtg cttattactg ctgctgtatt
13441 gactggccct ggcggtattg gcgctgcttt cgctgctggt ggattgacgg ggtttgctgc
13501 tgccactggc gcctcgttgg tcctcggtgg ggttattcag cttctttcac cgcagccatc
13561 aggcatagcc agtaaacaaa gcgcagataa ccgtgcatcg tatgcgtttg gcggggtaac
13621 caacactgca gcgcaaggct acccggtgcc tctgctttat ggtaagcgtc gaatcggcgg
13681 agcgattatt tctgccggaa tttatgtcga agatcagcag tagataacta acctttttc
13741 tggccacctt cgggtggctt ttttatgggc gcaatatggc tacagataaa gtgttaaaag

13801 gccgcaaggg cggcagctca agttcccgaa ctcctaccga acagcctgat gatctgcaat
13861 ctgtagcgaa ggcaaaaatc ctcgttgcgc ttggggaagg tgagtttgca gggcagctca
13921 ccggcaaaga tatctacctg gacggaacag cgctggagaa tgctgacggc tcccaaaact
13981 tcagcggggt aacgtgggag tttcgcgcgg gaactcaggc gcaaaaatat attcagggta
14041 ttcccggtac cgaaaacgag atcagcgtag gaactgaggt atcaagtgcc acagcctgga
14101 cgcgcacgtt taccaatacg cagctttcag cagttcgcct gcgtcttaaa tggccctcgc
14161 ttttcaaaca ggaagacgac ggcgatctgg tgggttactc ggtcaattat gcgattgacc
14221 tacagacgga cggcggcaca tggcagacgg tactcaatac cagcgtgacc ggtaaaacga
14281 cctcaggtta tgagcgcagc caccgtatcg atttaccgca ggctggcagc acctggacaa
14341 tacgtctgcg taagattacc tctgacgcca acagcgcgaa gatcggcgac acgatgatgc
14401 tgcagagctt caccgaggtg attgacgcca aactgcgcta cccgaacacc gcgctgctct
14461 acgtcgaatt cgactcaagc cagttcaacg gctctattcc tcaaatttca tgcgaaccgc
14521 gcggccgcgt tatccgcgtt ccagatacct acgaccctga aacccgcact tatagcggta
14581 catggaccgg tgcgtttaag tgggcatgga cggataaccc tgcgtggatt ttttacgacc
14641 tggttgtttc tgaccggttc ggccttgggc accgtttgac cgctgcgaat attgataaat
14701 ggacgcttta tcaggttgct cagtattgtg atcagatggt accagacggc aaaaggggca
14761 atggtacaga accacgttat acctgcaacg tgtacattca ggaccggaac gacgcctaca
14821 cagtcctgcg tgattttgcc gctatcttcc gtggcatgac ctactggggc ggggatcaga
14881 ttgtggccct ggctgacatg ccgcgcgatg ttgattacag ctatacgcgc gctaacgttg
14941 ttggcggtcg cttcacctat tcgagcagca ccacgaaaag ccgctacacc acagcgctgg
15001 tttcatggtc agacccgggt aacgcttatg ccgacgcgat ggagccggta tttgagcagg
15061 cgctggtggc gcggtacggc ttcaatcagc tggaaatgac agccatcggc tgcaccaggc
15121 agtcagaggc gaaccgaaag gggcgctggg gtattctcac caataacaag gatcgcgttg
15181 tttcgtttga tgtcgggctg gacggaaaca ttccgcagcc gggctacatc atcgccgtgg
15241 cagacgagct gctttccgga aaggttatgg gcggccgcat cagcgccgtt aacggtcgcg
15301 ttatcaaact tgaccgcgta gctgatgcag caccaggtga tcgccttatt ctcaaccttc
15361 cctccggagc gtcgcagagc aggaccattc aggccgtgaa cggggaatca gtcacagtca
15421 ccacggcata cagtgagacg ccacaggccg aagctgtttg ggtggttgaa tctgacgagc
15481 tttacgcgca gcagtatcgt gttgtcagcg tttccgataa caatgatggc actttctcga
15541 ttaccggcgc atggcacgac ccggataaat atgcctgtat cgataccgga gccatcattg
15601 accaacggcc ggtgagcgtg atcccgccgg gcaaccagtc gccgcctgcg aatatcgtga
15661 tcagctcgtt ttccgtggtt cagcaaaata tcagcgtcga aacaatgcgc gtgagctggg
15721 accaggcgca gaacgctatc gcctatgaag cgcaatggcg ccgcaacgac gggaactggg
15781 ttaacgtgcc gcgcacctcc accacgtcat tcgacgtccc ggggatttat gccgggcgct
15841 acctggtgcg ggtgcgcgca atcaatgccg cagaaatttc atccggatgg ggctattcag
15901 aagagaaaac gctgacgggt aaagtgggca atccaccgaa gccggttggc tttatcgcct
15961 ctgaaaacgt ggtgttcggt atcgagctga actggggatt cccggcgaat accgacgaca
16021 cgctgaagac ggaaattcag tacagcctga ccgggagcga agatgatgcc attcttctga
16081 gcgatgttcc ctatccgcag cgcaagtatc agcagatggg cctgaaggcg gggcaaattt
16141 tctggtaccg ggcgcagctg gtggacagga caggcaatga gtcgggttat accgactggg

16201 tacgtggaca ggccagtatc gatgtgtcgg atatcaccga tgttatcctg gaggacatta
16261 aagaatcgga cacgttcaag gaactgatcg aaagcgcagt agacagcaac gaaaaaattg
16321 ctggtatggc tgacgatatc agacagaacg ctgacgatct ggagcaacag gcgctggcca
16381 tcaaggaaaa cgccgatggg ctcgcccagg ccgaggtgaa gattgacgaa atctctgtct
16441 cgatggatgg catgacggga ggcgttaaaa actcctctat cgcggttatt cagaacagcc
16501 tcgcgcaggt caccagccgt cgatcccaga cagccaccaa cgccgggaac agcgccagca
16561 tcgaccgtat cgataccacc attgcagata ccagccaggc ggttgcccgt gcgctggtta
16621 cgcttgatgc ttctgccggt ggtaatgtct caaacgcgac cgatctcacc gaaacccttg
16681 ctgatttcac gcaggcctcg gccacgaaaa tcaactccct gacggttacg gtaaacggcc
16741 agacagcggc tattaaccag accgcgcagg cggtggctga tgtgaacggt aacctcagcg
16801 cgatgtacaa catcaaggtt ggtgtctcca gcaacgggca gtattacgcc gcggggatgg
16861 gtatcggcgt ggagaatacg ccatccggca tgcagtcaca ggtaatcttc ctggcagacc
16921 gtttcgccgt gactactatg gtcggcggga ctgtaacact gccgttcgtt atccagaatg
16981 gccaggccat tatcagggat acagtcattg gcgacgggac cataagcaat gcaaaaattg
17041 gcaattacat ccagtcgaac aattatgttg ccggctctgt tggctggaaa ctggataagt
17101 ccgggacgtt tgagaactac ggttcgacag ctggggaggg agccatgaag cagaccaacc
17161 agacaatcag cgtcaaggat gacaacaacg ttctgagagt gcagtttggc cgattaacgg
17221 gggtgttctg atatggctta tggaatacag acctgggatg cttcaggaaa acccaacaac
17281 tatggcatca aacccgtttc cgtcgttgga cgtatacaac ttgccgccgg acaaaactcc
17341 ggcagctggt ctttcactgt accctcagga atgaaggtcg gttttgcgct ctcacttgat
17401 gaaggaggta acagcgtagg gaggagcatt gtcgcgtcag ggaacacaat aaccgtaacc
17461 gcagcctctt ctgtgggcct gggtaattac ccggcctcta aatgtgaggt ggtcgttttc
17521 atggagaaag cataatggcc gaatttggcg cgatgatatt aatggataac gggaatccct
17581 ttgtaacgcc acaatcaacg cctttctgtc tttacgggaa gtataccttc aattcatccg
17641 cgaatggcag ttctcagcag gttgctcaaa atatcgcttt aaacgctgat tacccagtga
17701 tggtttttat caagaccacc aatacagcac agcccactcc ggttatgtct tacaggaacg
17761 gcggtaatgt gtatgttgct ggtgttaatc cctacaacca aagtttcact ttaacgtcgt
17821 acgtttttgc catattcccg cagatattac cgaaatgggg tttggcaata tgggatgcga
17881 gcggaaagct tgtgttaact aatgagtccc gtgtgctatc agacctgcag acggttggca
17941 cgcctggtgc aaacggcggg ataaatattg accagacgct gagcgggtca tgggccgttg
18001 cacctgctca gttgggtcag accatcattg tgaataattc aaccaagcct ccgactatct
18061 acacgataaa tgcttattct tcatgcaggt ttgacggggc caatacgagg ataaacgcag
18121 gggggacctc cactggggca ggttcacctg gaggggggaac gaatactggc atttcattaa
18181 ccgccataaa tacagcggcc tatgattgat tgatcgtttt tagcgatcaa taacataata
18241 ttgatctatc caatcaatta tacccaccag aattgtattg gtatcgtcta agatactgaa
18301 ttcctctgga tactatcaaa atgagaaaac tgattatctg catggcaggc gctgtcatgc
18361 ttacaggatg cgctggcgta attgagaaac aggaaccagt ttgcagcggc actgcaatcg
18421 ttggcggtca ggaaactacg gttcagattt acggtgtgcg taaacaaaac aaccagacgc
18481 agtaccgggc tggatatcct ttcagctggc gctgggtaag tgcgaataca tttaccgaaa
18541 caacctgcaa ataacccact acgcttaaac ataaacctcg ctccggcggg gttttttat

18601 tgcctggaga aaatatgctt tataacaccg gcaccatcgc cattaatgga aataccgcca
18661 ccgggaaggg tacaaactgg acggcaccgg ccagccaggt tcgcgctggc cagacaatta
18721 tcgtgatgtc taacccggtg caactgttcc agatttcatc cgtgaacagc gccacgtcaa
18781 tgacggttac gcctgccgct tccccggcgc tgagcggcca gaaatacgcc attctggtgt
18841 cagacattat ctccgtggac ggactggccc aggcaatgtc gcagctcatc aaagagtatg
18901 acgagaacat tggcgcgtgg gagacgttcg ccactacctc agcaaaccag accatcaccg
18961 ttaccatcaa cgggaccacc gtaaccatcc ctggcatcgg taaactggca cagaaaggga
19021 gcaacggtgc ggttactgtc gcagacggcg gaaccggggc aacaacggca gaagggagtc
19081 gcacaaacct cggtttagga aacagcgcca caaggaacgt cggaacggct gccggaactg
19141 tggcggccgg ggatgattcg cgattcgaaa agattgccaa attaggaagt gcagcaacaa
19201 aagatactgg tgagggtgaa ggaaatgtgc tgataacggg gtcatttggt gtgggttcta
19261 aaattttacc agtcatcagc gatatttggg ataaaagcca aggctcgcgc ttctgcaacg
19321 ttacccctgc gacattggga ggtcctggta tttacggctc aggcatccgg ctgtctgatc
19381 gaaacatcgg aagtggagca acctatgccg cacagcaatc ctttttgct ctgattttta
19441 gcggcaaagt cattcaattt atggggatga ctgatggcgt cgacaccggg tggatgaaga
19501 tttaccacac agggaacacg accagggcat ctgatggcac gctgaaggcc gcctccccga
19561 ttgtacggtt gtttgggaat ggcaaatgcc agcttaacga cgaatcagag gggtgcactg
19621 taacccgctt ggctacggga aaatatctgg ttgaagggtg tgaggggctt aactctgatg
19681 ccgcctgggg cggtattgat ggcggtttcg acatcccaac cgaccgcaac aaacaaccgc
19741 tcatctggct ggactacgaa gtaaacgctg atggttctgt gctgattcaa acctaccacc
19801 gcacacatcc ttcggcacct gcttatgcca gaaatgaacg tgacgggatc aatgacggag
19861 agcctattga tatcccgtca gatcagttcg tcagcgttcg agtagaaatg cctgctaaca
19921 gcatctggaa tcagaaacag aaagcaatag aagaagctgc gaaaagcgcg tccgaagagg
19981 ttcaataaaa taattgtgac ggcagtaaag cgactctttt actgccgtca aatatcacac
20041 tccattctgg agaacggtcg ggaactcaga aaccagccac atatcggact cttcaaacat
20101 ttcctccagc atgcggttca gtttttcccg atcacttttg ctggcatcgc tatttaaacc
20161 gttcgcctgc attggcttca ccttcacttc ggcatcgggg aaaatctggt gcacccgctt
20221 cttcaactcg gccagtatga tctctctggc ccctttgagc ccctcaacat tacgcttgtc
20281 ataaactagt tcaacgaaca taaccatccc tctcgcttac ttgatctctg caaataaaaa
20341 tactactgta ttcatataca gtcaatgtgt catgagggct atacttatgc ctcgtaaatc
20401 agatattcat gcagcttttt tagcctctat agaacagaat caaaagggtt acctttgtct
20461 caatacaaat aaattcatca ataagttgcg cgagaagagc tggcatttca gccaggcgga
20521 tgcaaataca tggatcgagc gataccagcc ggacttcgcc gataagacga caaacggcag
20581 ccagaaccgg tactggatcc tgcgtaacat ggggagggtt ttctaatggg ctttccttcg
20641 ccggctacgg attacgttga acagcgtata tcgcttgacg agcgcatcat caccaggcca
20701 gcggctacgt actttatgcg ggccggtgca acgcattacc gagaaggtat cctcaatggt
20761 gctttgctgg ttgtcgacgc gtcaatgtct ccatgtgatg gttcattgct ggtttgcaca
20821 gatagcggtg agtttaggat taagcggtat cgcacacacc cgcggccaca cctggaaaac
20881 cttgagaacg gtaaacggga gagtttgcta gataaggatg aggtatccga cacttcgcgt
20941 ccggtatttg gggtgatcac gtatatcatc aacgacgcgc ggtctggtga gtttgatgac

21001 tacccgctga agtgaaaata gtgttgtgta ccaaattgcg taccaaacta aaatcacaaa
21061 tcatgaaacc cttgttcatg gcggttctca ggggtattgc gcgtaatcgt gaaagaggaa
21121 ggtaggttgt ttatctgtgt ctgtcattgt ctagcattgt ccactgtagt taattaactg
21181 cctgttataa ctcgaaaaac aaccattttt tcacctgagt atgtctacga acgtcaaata
21241 tagttgttta ctgtgcgctg ttgtccatgt gatattgtgt accagatgtg taccaaacca
21301 ttatttctga gcgtaccaaa tattatttat ggctataagc gacacaaaac tgcgtactat
21361 ttatggtaaa ccatattcgg gcccacaaga agtggctgat gccgatggcc tcagcgtacg
21421 aatttcaccg aaggggggtca ttcagttcca gtaccgttat cgctggcatg gcaagcctaa
21481 tcgacttggg cttggtcgat acccatccct gtctttgaag gatgccagac agatcactgc
21541 tgacttgcga aatctctatt tctcaggaac ggatccacgc acctattttg aagagaaggt
21601 ggagaactcc atgacggtcg cccagtgtct cgactactgg ttcgacaact acgtctctac
21661 aactctcaga gaaaagaccc aggcacttta ccgatcaacg gttatgaagc gcatgcatga
21721 cgcctttcct aatcgtccgg catcttctat cacggttaag caatgggttg acctgcttac
21781 cgaagaagaa agagataatc cacgccgagc aaggcaggtg ctaagtcaac taagatcagc
21841 aataagttgg tgcatgcggc gtcagttgat agatagttgc gcaattatga gcatccaacc
21901 aagggacttc ggctcccgcg ctgaggtagg ggatcgggta ctgtcgtatc acgaactggc
21961 taagatttgg cttgctattg aaagaagccg tgcgtctacg tcaaataagc tacttcatca
22021 gatgcttatg ctgtgggggg cgaggctctc agagcttagg ctggcaaaaa agacagaatt
22081 tgacctgctg gaaaacgtat ggaccgtacc gaaagagcat agcaagatgg gtaatgttat
22141 ccgccgtcca atcttcgaac aaattaagcc tttcctcgaa aaggccatga caacgtacaa
22201 tgatgttctt ttccctggag aagacataaa caaaccgatc agcatcgctg cagccaaccg
22261 attcgtaaat agaataaggg gagggatgga tctaggttac tggcgaacac acgatttcag
22321 aagaacgctt gttacacgtc tgtccgagat gaatgtcgag cctcatgtta ctgagcgaat
22381 gctcggtcat gaacttggcg ggataatgtc cgtgtacaat aaacacgact ggatagaggc
22441 tcagcgcaaa gcgtatgagc ttcacgctga taaattgttc tggcacatca ggagcatttc
22501 tgattaacgc caccgttaag aatccaccct tcaacagctt cacgaaggta tgatttgggg
22561 tgggttctga ctggcttcgg aaatccgtgc cgtttggtat agttccagat tgtctgacgt
22621 gatgaaacac cgagcttgtt catcacttct ttctcaggaa tcaggctggt atcggtcatc
22681 ttaattctcc aggcaaaaag aaaccgccat caggcggctt ggtgttcttt cagttcttca
22741 attcgaatat tggttacgtc tgcatgtgct atctgcgccc atatcatcca gtggtcatag
22801 cagtcgttga tgttctccgc ttcgataact ctgttgaatg gctctccatt ccattcacct
22861 gtgactcgga agtgcattta tcatctccat aaaacaaaac tcgccgtagc gagttcagat
22921 aaaagaaatc cccgagagtg cgaggattgt tattcacctt taacggcaag ttgcaggtta
22981 gccacggttt acctcctgcg gcggttctgg tagtggcatc cagtgtgtga cattgctaat
23041 cagaccatat tcattagttt gaggatggtt gccgttatcg tctccgtatt taagactctc
23101 cataaagcca taatgcctat caccattaac gctcacaaag ccgtaataag caggtataac
23161 gccgatctca cacgtaacca gcaaaggaaa actagttctc caatttagct cgccgattac
23221 aggcatctgc tcactacagc ttatccaacc atctggaatc accggagagt tgccaccctg
23281 aacagtaggc atatccggac ctttgcgaat cgccctggca agatcgattt ggtcgtcgta
23341 caaccagtca cctgtttgcg gatgattggc ttctgccaat tgtgcagccc actccaggcc

23401 gtctttgtgt ccttgcagat agtccagcgg taactcatca ctattactta caggttcggc
23461 ctgaagcatg gaggcgcgat aggcgttcca gccgacagct tttccgtgtt caaacgcgct
23521 gtcaaagtca tcatcaattt ccatcgcagc gggcacagat accggcgctg gcggggcggt
23581 gtaaagcggc gttacttctc gcagcgggtc ggcataagca ttgccactat cgaagctgac
23641 gttgtttttt gccccgcctc ctgacagcaa ccacgccacc ggctccgctt cgagcgatgc
23701 caacgctata cgcgccatct ctttgacaac gtcacagttt ttatttcctt cgcaatcatc
23761 catcaaaagc caatcgtcat tgattatttc tttaatctta tcccgtgaaa gttccattaa
23821 ttcaccgcct tagctttacg ctctgcttca ctaaaatccc agtcaacacg atgcgatatt
23881 tcaattgccg ccctaacgga atgctcaatc atgatgtcga gcttttgaag tgtcatggcc
23941 atatttggat ttcgagccaa tatttctgcg cgaagaattt gccagtcatt gcatgtttca
24001 agtaatgaat tagccattat ctttaccccg cgctccatat gcagataagc actcttcaaa
24061 tccagcctga ttatccgttt gaccaaaact gaagccatgc tgaagaccat gacgaaatgc
24121 gctatcttgc aatttatctg cgctatcgag cttcgcttcc agttcagcga ttcgcttctc
24181 tgcggcttcc agcgatgcca gtgcaatacg cagggcagcc aggatattcc cctggtaatc
24241 atcaagcccg aaaggaagct catcacgaac gctttcataa tcgttgatag cctgctgaag
24301 ccattctctt gtgatagtgc tcatgatgcc tctcctttac atgcttcggt gacgtttatt
24361 ccagcgttgt gcagtgcctc aagaacctga tgctgcctgt aaaccatttc agtgtggtaa
24421 ggctcgtcaa aatcgacgcg atgcaacatg ctatagcatt gcggaagcac aacctcccgc
24481 aactccagct cagcaatgcg cttctctgcg gcttccagct tctcgcgcat atcgtcaacg
24541 tactcgacca gagatccgcc agcaggaatt tcgcactcct cgaccagttg gaagtagata
24601 tcagctgcgg cccgtgtgtt gctatgccta gcgtcgccca tctcaccttc acgaagagca
24661 tcgcgttcgg cggtaagatt ggctattttg ctgtctttgc cttccagctc aacgcgcagc
24721 ttccctaccg ttagcgcaat ttcctcgttc tcctggtcac ggcgttttat gtattgctgg
24781 tttctttccc gttcatccag cagtgccagc acagtagccg gattggatgc ggcgatgaat
24841 tcagcattgg cctgctgttc catttggaaa tcttcatcga aaccgctttc tggatgcgcg
24901 ccttcaattc tgcaaatggg aatatatcca gcagcctcgc gatgaattag cgcatcatca
24961 ccatcaaatc ggccatctcc atattcgagc gaccactcgc cacacgttgc tttttctgcc
25021 ttagcacgca gtgcctgata atcaatctgg ctcactggct gcctcctttg cgaagctggt
25081 cggcgaacaa acgtacacca gacgcttcac tgcgtagaaa cctaacggca taatcaaaac
25141 cacctcgttc tgcgtcgtct gctccgttgt cgaggttatc tgcgtacatc tctacccccct
25201 gcgcccgcac ttcagccagg aaggcgtcgg ttgctgtggt gttaagtgaa aagtcctgtc
25261 ggcctgattt cagatacgca atggcgatct ttgttgctgc gcattcgaca gcgaatttct
25321 ctttttcagc atccagcgcc gcgcacttgg cctcaaggtt atcaatcgtg attccagcag
25381 aacgacactc ccgcaacgcc gtttctagtt ttgattcaag ttcaccgaac ttacgcacca
25441 gatattcagc gtttgtttcg ttaaccttta aatctcgggg gatgcattta cctttcagaa
25501 aaccatccat ctcaattagt gacatttgtt tcatttcttc ccactccgcc acattgcatt
25561 cagatatttg ttgtcattaa cagagccgaa actatttctc ttaagcaatt cctctctcga
25621 tggcattggc tttacgcgtt ggcgaataat catttctgcc ggaagaatgc cgggattgta
25681 tgcaagtcct ctcatggtaa atttctcagt cattactgat agcgccataa cgtgagcggt
25741 aattacgcag gcgcgggtca atttcaggga agtgggtata tgtggctttg cggaatggtc

25801 ggattgatgt ctggtaaatt cgctcgcgct cttctttctc tgcaagccat atacaatgac
25861 gaaattcctt ttcatctttc gtttcctgcg gtagcgacat tatcaggtcg tagtttttc
25921 tgaatttatc cagcacctcc gatacggaat tgccggaaca gcggcgcgcg tcatccgcac
25981 catacagagg cgctggcatg attttctcct gattaaattg cgtgaatagc gtgacgaggg
26041 aaggggagag ttactggtgc aaagggtata tcgtcgtcaa aatccatagg tggttcgctg
26101 tgattcccct gctgctgaga ttgctgtctt tgttgctgac cgttatttcg ctgaggtgaa
26161 gactgttcat tgcctccttg cttgccacca agcatttgca tggttccacc aacgcctacg
26221 ataacttcgg tagtgaaccg atcctgtccg ctttgatcct gccattttct tgtccgcaat
26281 ttgccttcaa gataaacctc agagcctttt cgcagatatt cgctggcaat ttctgccagt
26341 ttcccgctca ttaccacacg gtgccactcc gtctgctcct tttgctctcc agtttgctta
26401 tcacgccatt gttctgacgt agcaactgta atgtttgcaa atgccgttcc tgatggtgaa
26461 tatctgattt ctggatcatg cccaaggtga ccaataatga tcaccttatt tacgcctctg
26521 cttgccattt atgccgcctg ttttagttcg ttaactctga tgttcattac ctgaacgcat
26581 ttagcctgcg cctcctcgtt gccagccatt aattgccagt cacgttgata acgctcgatg
26641 agtttttct tgtcagtttc tgttgacgca taatcgctga agtctttcag gatttgttcg
26701 cagtcaaccg atggagattt ctggttggta ttttctggtg atggtttgtt atctgatgct
26761 ggtattgccc atcccggcag cgatggaggg agccagtaaa atcctgttcc atccttgagt
26821 tttgccctgt gccacccatg ctttttatcg agagatgttt gtgcgaaacc ttcctcaagt
26881 ttatacagat accgaccgat tccccactga acggcagcac gcttcattgc acctgaacga
26941 ccacctttga cggcttctac ctgcgtgttt tcagcagcat cccatttggt tacccattcg
27001 gaatcaatct ttattgatat gccgcattca acgccgccgt tgttgggaat atcgcggtat
27061 tcattgcgcc atcctgcttt gccgcaaaca tcgtccaggc gtttcatgat tgcccggttc
27121 gttacataag ccagcaccat agcccacacc ttgccatcgc gtgttttacc gctttgctgt
27181 attcgccatt cgatatcttc agggcagaat ggctcatcga atttgttcaa atccataatt
27241 cacctcagaa tggacacagc ccaaggaaat aacgctgatt taatacttcg actcgggaca
27301 aattaaggca tacccgcatt ccttcgcggt cgccattatg gcgataccag agagctttct
27361 gcgtgtacat gcgtctctgt aacttgctct ccttcactgt ggttgcaagt gacatgaata
27421 tctccttcgt taccgattaa ttctttcatc tgacgaatga attcttcgtc tgaccagtta
27481 tctgtaaaac tcatttcctg cgataccacg gaatgttgat cgctgatttc atcgctttat
27541 ttgcttcaaa ccacattttg gaatcaccaa taaatctggc tattactgct ttgttttgtg
27601 cagcacgaag catctgatga tttatggcta tttcattgcg cataataaga cctcaactct
27661 tttccatccg tcacgtaatt tacgggtgat tcgttcaagt aaagattcat ttagttggaa
27721 ggcacccatg cgagcgcctc ccgcgattgc gtaaatcatg ggtggttcct tatgttggtt
27781 ttattagtag gttatttttg ttgcgaatac ttcgcctttt acgatggctg ttatgatatt
27841 tttagcaaca tcttctgatg caccaacctt gataaggtca gcaagtattt tgttatttac
27901 ttctttccgg tgagctttat cctttgctct acgctcttct tcttccttga ttctttttc
27961 ttctgctatt ctggcttgct cttttgcttc agcatcgcga cggattcgtt cagcctcctc
28021 ctgtgctttt ctgcgttctg cttcaattgc tgcctgcttt tctctttcag ctcgttctgc
28081 tgcctctttt gcttcgcgtt gtgctcgctg ctcggcttca atgcgttcac gctctgcacg
28141 ttccgctgcg gccttagctt ctgcttctcg ccttgctgct gcttcaattt cggcttttgc

28201 ctttgcttcg gcttctgctc tggctttctc ttcagcttct ctttttaagc gttcttcatg
28261 ctctcgcttt tcctgctccg ctttgagtct tgcctcttct ctttggcggt caaattcgcg
28321 atccatcaaa atagctattt catggtcaga ctcaatttgc tttgcgagag cttcagctgc
28381 tgccttagct tcttcttcgg ctttaatccg cgcctgttct tcctcataat cagtaagagg
28441 ctggcgtgcc ttggctttca gctcatcaag gcgatcacgc actgtcttgc ggttggcatc
28501 aattagcttt ggaatttcct tcagttcagc aacaaggtct ttgccaagac catcgagata
28561 tgttttcgtc tgcgcaactt tatacgccag agaagcgatc tcctttctgc cctttgccgt
28621 tgtgatatca ggcacaaagg acataacttc acgttcaacc ttttggagaa tttcttcaat
28681 ctggtcagca gactgaaata caatcattgc atttgctttt tcaataacaa ctaaatctgt
28741 tacttcactc atatatcctc catcaaaaaa atcgccctca cactggaggg caaagaagat
28801 ttccaataat cagaacaagt cggctcctgt ttagttacga gcgacattgc tccgtgtatt
28861 cactcgttgg aatgaataca cagtgcagtg tttattctgt tatttatgcc aaaaaaattt
28921 taatcagcaa taaaatcttc tggtaattta tcaaccagtt gatggcttat tatcagccat
28981 ttgccatcct tcgtttcgta tgcgtatttc tggtctttta tcatcaggtg ttcagctact
29041 gccttaactg cctgttcggt gacatcttct ttctttccta cccacattcc tttttcagtg
29101 tttaatgttc cttgaaaaat acgaccgctt aatgggcttg cgcccatagt ttttactctc
29161 atgtatcagc tctcaaataa gtggtttgct gccaaaacaa tgaaccatcc ggaaattcca
29221 gatagttcat aattcactct tcaatacttc cagcttacta atcgccgata gatatccgcg
29281 ctgatagggc atcatcattc cttcgagctt gccacttctt aactcctccc tgagcaattg
29341 tattgcttga tcaataacct ctgccttagc gtcctttatg gcttgcttgc ggggctttgc
29401 tttctgcttt ggcagatttc tcaagcatga tggaatgtat gtctgattca tcacttacct
29461 cgccgtcagt tgttttgatt tccggtagcc tgccgcgtaa atggctacgt ttggaagaca
29521 tacaccagtt tctggttgct tatgtccaaa ctcattcgcg tacacaatgg ccgctcgctc
29581 cagattgcgt ctgtattctt tctgttgcca gatcacgtcc tgtgccatga acttaattgg
29641 cttagcgtct tctatgcgct caggcgtttc gtgagtacct ttagcctgaa tctgcgctct
29701 gcttagagta gggcggtgta atacttctga acttattgct tcttcgcggg ccagcacacc
29761 gttagctaat gcctttgcct ttaaacgctc acgacgacga gaacgtgaat tgcctttgaa
29821 ctgagttctg cgtgtcatat agacctcctg atgaactttg gtggtgtggt aggtgggaga
29881 cccatttcga cctgtttcgg cctacttcaa ttcggcaata gtcccgcagg cctcgccgct
29941 ttacgtgcga catattcccg tccatgaacc cttcaccaca ccccaaagtt cactttggtt
30001 attgcgcttt gtcagcgccg tagattcata tttgaatcgt tgtatattca ctgccatggt
30061 gagtagtgcg tcctgctaat gtgtttagta tcaccgccag tggtatttat gtcaacaccg
30121 ctagagataa ttcatcaccg cagatggtta tctgtatgtt ttttatatga atttattttt
30181 tgcaggggtg tattgtttgg taggtgagag atctgaactg ctatgtttag tgagttgtat
30241 ctatttattt tcaaatagat acaattggtt atgtgttctt gggtgagggg gatcgtgagg
30301 caaagaaaac ccggctctgt ggccgggaat aacatttagg aatcaaggtc aggtagcatg
30361 attttctcaa tcaacgtcaa tgccctttgg tctcgttcag caaaatattt aggagcgtac
30421 tgaggcagcc acacttcgtt gaagtattgt ttgaaatctg caatattcgt ttgggtatag
30481 acgtaccggg aatgtccggc catctgggta ctcatggtta tatgttggga atgtcttcgg
30541 ctcaaccccc cggttgtcac ggagccattg cgagaaaacc ctaccctctg aaatatcagg

30601 gaccattttt tctggcagcg tatatccggc ctgttcaagt ggcgcaacca agttaaacgt
30661 cagttcatta agcatagaaa agtgcgtatg aggaactctg cctcggtttg tcatatacct
30721 cttaaggtgg atagggagtt cggcaggcgc tctttaacct gacatccact cacgcaccca
30781 tctcgatact tgcactgcaa attttggaga tagccactga gctaaattaa ttgcgatgtc
30841 tggatgaacc caagtccctt gattctctgc tcttccgcct ttaaatgatt gaattaactc
30901 cgatatggga atccccatat cgcgtgataa ttcatcaaaa aaatcttgtg ttgtttttag
30961 tcgtgtatag tcagcaagta gcttccccgc agacttgcac atcgcggtgg cattgatgta
31021 tccgtctttg gtgcgaagat ggatgacttc tccatcaact tctctagcga ttaatgcaag
31081 ttgaaattgt gtcataaatc atcctattac gtgaaaaata aaacaatcac ccaaacgtct
31141 cttctggcca ctggctggcg ataactttcc ccacaacgga acaactctca ttgcatggga
31201 tcattggata ttgcgggttt agtggttgta gaaacacctg accgctatcc ctgattagtt
31261 tcttgaaggt aaactcatca cctccaagtc tggctatgca gaaatcacct ggctcaacag
31321 cctgctcagg gtcaaccaga attaacattc cgtcaggaaa gcttggcttg tagcctgttg
31381 gtgcggtcat ggaattacct tcaacctcaa gccagaatgc agaatcactg gctttttgg
31441 ttgtgcttac ccatttctcc gcatcgcctt tggtaaaggt tctaagctca ggcgagaaca
31501 tcccggcttg aacatgagaa aaaacagggt actcatactc acttctaagt gacggctgca
31561 tactaaccgc ttcatacatc tcgtagattt ctctggcgat tgaagggcta aattcttcaa
31621 cgctaacgtt gagaattttt gcaagcaatg cggcgttata agcatttaat gcattgatgt
31681 cattaaataa agcaccaaca cctgactgcc ccatccccat cttgtctgcg acagattcct
31741 gggataagcc aagctcattt ttctttttt cataaatagc tttaaggcga cgtgcgtcct
31801 caagctgctc ttgtgttaac ggtttctttt ttgcgctcat gcattaaatc tatcaccgca
31861 agggataaac ttctaacacc gtgcgtgttg actattttac ctctagcggt gataatggtt
31921 gcctgtacta aggaggttgt atggaacaac gcataaccct gaaagattat gcaatacgct
31981 ttgggcaaac caagacggct aaagatctcg gcgtatatca aagcgcgatt aacaaggcca
32041 ttcatgcagg ccgaaagatt tttttaacta taaacgctga tggaagcgtt tatgcggaag
32101 aaataaagcc cttcccaagt aacaaaaaaa caactgcata agtaacaccg ctcttttcac
32161 aatggacatt cgtcctacgt cgctgacaaa gcgagcccca agatatctga ccaactaagg
32221 ccatatgcgt ttccacgcat acctttcaac taactattca ctattggaaa attaacaaat
32281 gacacaagca agttatagca agccaacaca gcgagaaatt gatcgcgctg aaactgattt
32341 actcatcaac ctgtcaacgc ttacccagcg cggtctggca aagatgattg gctgtcatga
32401 atcgaagata agcagaacgg actggagatt tattgcttcg gtcttgtgtg ctttcggaat
32461 ggcatcagac atcagtccga ttagcagggc ttttaagtat gcattggatg gaatcacaaa
32521 gaaaaaatcc ccggctgcca ccgaggattc tgagcaaatt gatatgcaat tctgagggaa
32581 ttactggatc aatccacagg agtaattatg acaaaacgtc gtaagaaata ccaggaaaaa
32641 gaagagattc gacaccctga ttcacctgag ggattagtgg tagccgcagc aaataacagg
32701 gcgttcgcag agcgccttgt tggtgtttac agactagcca aagcaggagt gaaacatggg
32761 cgtcgttaag ttagctgatt acaggcctca actggaggtc gtggagcatc gcgtggcaga
32821 taccgaagat ggtttcatgc gcgttgctaa cgagattacc gacagtctgc tgatggctga
32881 tttaaccgtc cggcagatga aggtgatgct cgctatcatg cgcaagacat acggattcaa
32941 taagccgatg gatcgactca caaacacgca gatagcagcc atgacaggta ttcatcacac

33001 tcatgtttgc gctgccaagc gccagcttat tgagcgtaaa ttcctcattg ctgatggcgt
33061 gaaaatcgga gtgaacaagg tggtttctca gtggattagc caggacagct taacattagc
33121 taaaacagct aataaaacat tagccaagtc ggctaatggg tataagccaa gtcagctaaa
33181 cacaaaagac aatatacaaa agacaataaa tacaaatacc cccttacccc ctaacggggg
33241 aggcgatggg caggttaaac ctgaacgtcg caaggcagaa cgcatcgact acgaatcctt
33301 cctgaacgcc tacaacaccg aagtcggtga cagacttcca cacgctgttg cggtcaacga
33361 gaaacgcaaa cgccgtctga agaaaatcat tccgcaactg aaaacgccaa acgtggacgg
33421 tttcagagcg tatgtcaggg cgtttgtgca tcaggccaag ccgttttact tcggagacaa
33481 cgacacgggc tggacggctg attttgatta cctgctgagg gaagactcgt taacgggagt
33541 tcgggaaggg aagtttgcag acaggggat tgcatgaaac aggatatcga agcgagcgtt
33601 atcggtggcc tgctgattgg tggattaact ccaaccgcca gcgacgttct ggcaacgctt
33661 gagccggaag cgttttcaat tccgctctac cggaaagcct tcgaggttat ccgcaagcag
33721 gcgagaaaca gaaacctaat cgacgcgctg atggttgccg aggcgtgcgg agaggagcat
33781 ttcacctcaa tcctgatgac cagcaaaaac tgcccgagtg ccgcaaacct gaagggatat
33841 gccggaatgg tcgcggataa ctatcaccgc cgtctggtgc tggaaatcat ggatgaaatg
33901 cgtgaaccaa ttcagagcgg aaccatcgac gcatcgagtc aggcgatgga tgaacttgta
33961 aagcgtcttt cagccatcag aaagccccgt gacgaggtaa aacctgtacg gttaggggaa
34021 atcatcactg actacactga cacgcttgac aggcgtctga ggaacggaga agagtccgat
34081 accctgaaga ccggaatcga agaacttgat gccatcaccg gagggatgaa cgcggaagac
34141 ctggtgataa tcgctgctcg tcctggtatg gggaaaaccg aactggcgct gaagattgcc
34201 gaaggcgttg caagccgcgt tattcctggt tctgacgtcc ggcgcggggt attgattttc
34261 tcaatggaaa tgagcgcatt gcagattgca gagcgaagca ttgccaacgc cgggaggatg
34321 tcggttagtg tgctgcgaaa tcctgcagcg atggatgacg aaggctgggc gcgcgttgct
34381 aacggcatga gtcagcttgc agatttggat gtatgggtag tcgatgcctc gcggttatcg
34441 gtcgaagaaa tacgctcaat cgcagaacgg cacaaacagg aaaatccaaa cctctcactc
34501 atcatggtgg attatcttgg cctgattgag aagccgaaag cagatcgcaa tgacctcgca
34561 attgctcaca tctccggaag ccttaaggcg atggcgaaag acctgaaaac gcctgtgatc
34621 tccctaagtc agctttcgcg cgatgttgag aagcgaccaa acaaacgccc gacaaacgca
34681 gatttgcgtg attcaggaag cattgagcag gacgcagact caatcatcat gctctatcgg
34741 gaagcggtat atgacgagaa cagtagcgcc gcgccatttg ctgaaatcat cgtaacgaaa
34801 aaccgttttg gctcgcttgg tacggtttac cagcggttct gcaacggaca ctttgttgcc
34861 tgtgaccagg atgaagccag acagatttgc acatcatcaa atgcacctgc tgcgcgtggc
34921 agacgatatg cacaaggggc tgacgtatga ccatctacat cactgagcta ataacaggcc
34981 tgctggtaat cgcaggcctt tttatttggg ggagagggaa gtcatgaaaa aactaacctt
35041 tgaaattcga tctccagcac atcagcaaaa cgctattcac gcggtacagc aaattcttcc
35101 agacccaacc aaaccaatcg tagtaaccat tcaggaacgc aaccgcagct tagaccagaa
35161 tcgaaagctt tgggcttgcc ttggtgacgt ctctcgtcag gttgaatggc atggtcgctg
35221 gctggatgca gaaagctgga agtgtgtgtt taccgcagca ttaaagcagc aggatgttgt
35281 tcctaacctt gccgggaatg gctttgtggt aataggccag tcaaccagca ggatgcgtgt
35341 aaacgaattt gcggagctat tagagcttat acaggcattc ggtacagaac gtggcgttaa

35401 gtggtcagac gaagcgcgac tggctctcga atggaaagcg cgatggggag atcgggctgc
35461 atgactatca aatcaaatac gccagcacac gacaaggact gctggcaaac gccgctttgg
35521 ctttttgatg cactggatat tgagtttgga ttctgtctgg attcggcagc gagcgacaaa
35581 aatgctctgt gcgctcactg gctaactgag gccgacgacg cgctcaattc tgagtgggta
35641 agccacggtg caatctggaa taacccaccg tacagcaata tcaggccgtg ggtggaaaaa
35701 gccgctgagc agtgcataca acagcgacag acggtagtga tgcttgtgcc agaggatatg
35761 tcagtcggat ggttcagcaa ggctctggag agtgtcgacg aagttcgtat tatcactgat
35821 ggacggatta attttatcga accatcgaca gggctagaga agaagggaaa cagcaaaggt
35881 tccatgctgc tgatttggcg accgttcatc agtcctcgac ggatgtttac taccgtatcc
35941 aaatcggcat tgatggcgat cggacagggc gccaggaggg cggcatgaga cgacagcgac
36001 gaagtatcac cgacataatc tgcgaaaact gcaaatacct tccaacgaaa cgctccagaa
36061 ataaacgcaa gccaatcccg aaagaatctg acgtaaaaac cttcaactac acggctcacc
36121 tgtgggatat ccggtggctt agagaacgtg cgaggaaaac aaggtgattg acccaaatcg
36181 aagttacgaa caacaaagcg tcgagcgggc tttaacgtgc gctaactgcg gtcagaagct
36241 gcatgtgctg gaagttcacg tgtgtgagca ctgctgcgca gaactgatga gcgatccgaa
36301 tagctcaatg tacgaggaag aagacgatga atgagttaat aaatggcaat gccatcaaaa
36361 tgacaagcat tgaaatcgct gagttggtgg gtaagcgtca tgacaatgtg aaacgtacca
36421 tcgaaacgct ggctaaaaat ggtgttatcc ggcttcctca aattgaggtt tccgaaagaa
36481 tcaataactt agggttcaat gttcagtacg agcattacgt cttcgaaggc gaacaaggta
36541 agcgagatag tattgttgtt gttgcccagt tgtcgccaga gttcaccgct cgccttgttg
36601 accgctggcg agatcttgaa gaagctgcgg ttaatatccc caaaacgcta ccggaagcgt
36661 tgcgccttgc tgctgatctt gctgagcaga aaatgcaact ggaaaaccag cttgcaattg
36721 ccgcacctaa agttgagttt gccgatcgcg ttggcgaggc cagcggaatt ttgattggaa
36781 actttgcaaa ggttgtcggt attggtccaa acaaactgtt tgcgtggatg cgcgatcaca
36841 aaatccttat tgcttcaggt tcccggcgca atgtgccaat gcaggaatat atggatcgcg
36901 gctatttcac agtgaaagaa acagcggtca acacaaatca cggaatacag atatcgttca
36961 ccacaaaaat caccgggcgt ggtcaacagt ggctgaccag aaagctgctc gataatggaa
37021 tgctgaaagt aacaggggag gctgcttaat ggctaatcta cgcaaagaag cacgcggcag
37081 agaatgccag gtacgtattt acggcgtatg caatggtaat cctgaaacta cagttctggc
37141 acattaccgg atggctggaa tttgcggaac gggaatgaag cctgacgacc tgatcggcgc
37201 atgggcttgt agcgcgtgtc acgatgaaat cgaccgacgc acccataatc tcgacaacaa
37261 agacgccaaa ctttaccacc tcgaaggcgt gatcaggacg caggcgatac tgctgaagga
37321 ggggaagatt aagccatgaa cgaatatcag tttgtgcttc catacccgcc gtcggtgaac
37381 acctactggc gaagacgggg aagccaatac tacatcagcg ctaaaggcca gaaataccga
37441 aaagatgtac agcaaataat ccgccaactc aagttagaca ttttcaccaa atcacgactc
37501 cgcatcaaag tcatcgcaga cgttccagac tcccgccgcc gcgacctcga caacatcctg
37561 aaaggtttac tcgactccct tatccacgcc ggatttgcgg aagacgacga gcaattcgat
37621 gacattcgtg taattcgtgg tgtgaaagta ccaggcggaa ggcttggaat aaaaatcacc
37681 gaactggaga acgcatgaac gccacaattc aaacgatacc agagcttctt atccagacac
37741 gaggcaatca gaccgaagtg gcgaggatgc tttcctgcgc aagaggaaca gtgctcaagt

37801 acaaccgaga cagcaaaggc gagcgtcacg taatagttaa cggcgtcctg atggtcaaac
37861 agggcaagag gggaagacca tgagactcga aagcgtagct aaatttcatt cgccaaaaag
37921 cccgatgatg agcgactcac cacgggctac ggcttctgac tctctttccg gtactgatgt
37981 gatggctgct atggggatgg cgcaatcaca agccggattc ggaatggctg cattctgtgg
38041 caagcacgaa ctcagccaga acgacaaaca aaaggctatc aactatctga tgcaatttgc
38101 acacaaggta tcggggaaat accgtggtgt ggcaaagctc gaaggaaata ctaaggcaaa
38161 ggtactgcaa gtgctcgcaa cattcgctta tgcggattat tgccgtagtg ccgcgacgcc
38221 gggcgcaaga tgcagagatt gtcacggtac tggccgggcc gttgatattg ccaaaacaga
38281 acggtggggg agagttgtcg agaaagagtg cggaagatgc aaaggcgtcg gctattcaag
38341 gatgccagca agcgcaacat atcgcgctgt gacgatgcta atcccaaacc ttactcaacc
38401 cacctggtca cgcactgtta agccgctgta tgacgctctg gtggttcaat gccacaagga
38461 agagtcaatc gcagacaaca ttttgaacgc ggtcacacgt tagcagcatg attgccacgg
38521 atggcaacat attaacggca taatattgac tttttgaata aaattgggta aatttgactc
38581 aacgatggat aaatgcactc gttaaataaa gccctgagtt aatatctcgg ggctttttgc
38641 gttttaagca cggcctttct gaaagcacat caaaccaaat accagacaga caaaaataat
38701 caccttatcc gctgtggcta cggtgcggtg tgctttgcat aaaagaaaac cagcgcaatg
38761 gctggcttcg tgaaagcggg tggcatgagg ttgccctaac aacctcctgc cgttttgccc
38821 gtgcatatcg gtcacgaaca aatctgatta ctaaacacag tagcctggat ttgttctatc
38881 agtaatcgac cttattccta attaaataga gcaaatcccc ttattggggg taagacatga
38941 agatgccaga aaaaaatgac ctgttagccg ccattctcgc ggcaaaggaa caaggcatcg
39001 gggcaatcct tgcgtttgca atggcgtacc ttcgcggcag atataatggc ggtgcgttta
39061 caaaaacagt aatcgacgca acgatgtgcg ccattatcgc ctggttcatt cgtgaccttc
39121 tcgacttcgc cggactaagt agcaatctcg cttatataac gagcgtgttc atcggctaca
39181 tcggtactga ctcgattggt tcgcttatca aacgcttcgc tgctaaaaaa gccggagtag
39241 aagatggtgg aaatcaataa tcaacgtaag gcgttcctcg atatgctggc gtggtcagag
39301 ggaacagata acggacgtca gaaaaccaga aatcatggtt atgacgtcat tgtaggagga
39361 gagctattca ctgattactc cgatcaccct cgcaaacttg tcacgctaaa cccaaaactc
39421 aaatcaacag cagccggacg ctaccagctt ctttcccgtt ggtgggatgc ctaccgcaag
39481 cagcttgggc tgaaagactt ctctccgaaa agccaggacg ctgttgcact acagcagatt
39541 aaagagcgtg gcgctttacc gatgattgat cgcggtgata ttcgtcaggc tatcgaccgt
39601 tgcagcaata tctgggcttc actgccgggc gctggttatg gtcagttcga gcataaggct
39661 gacaacctga ttgcaaaatt caaagaagca ggcggaacgg tcagagagat tgaggtatga
39721 gcagagtaac cgcgattatc tccgctctgg ttatctgcat catcgtttgc ctgtcatggg
39781 ctgttaatca ttaccgtgat aacgccatta cctacaaagc ccagcgcgat aaagccacgt
39841 acatcatcgc tgacatgcag aagcgtcaac gtgatgtagc agaactcgac gccagataca
39901 caaaggagct tgctgatgct aacgcgacta tcgaaagtct ccgtgctgat gtttctgctg
39961 ggcgtaagcg cctgcaagtc gccgccacct gtgcaaagtc aacgaccgga gccagcagca
40021 tgggcgatgg agaaagccca agacttacag cagatgctga actcaattat taccgtctca
40081 gaagtggaat cgacaggata accgcgcagg ttaactacct gcaggaatac atcaggacgc
40141 aatgcctgaa atgatcgggc gatgaaaacc aaaaaaaaca ggagcaatac atgactaagc

40201 tttatcaccg catctcaact tttctctctg gttgctgggc gtttatcacg tctatttcgt
40261 tcgccatctt tagtttcggc agcacagcgt gctcgcttag tcggggtctg tggcgtgcta
40321 tttcagcact agcgccgaaa tttttacctg aaaaggctgt ttggcgaatt gtagagcgaa
40381 tgtgtagtga gagcgttcgc gagaagatta acgtatttgg acgtcatcct cgaaatacag
40441 gcgcattgtg cagtccgtta ctgtagtcat tacaaagccc atctacgggt gggcttgata
40501 atgaaaccgg aatttattct tggcaaccag ttacggcagt accacgaaac aacccaagcc
40561 agaaagtgtg gaaataacac tggcagccac tgaaagataa acctcctgcc ttatggcaaa
40621 aaagattctt tgtggtggcg gactgatgga aagacatcgg ttattgcaga gaccattcaa
40681 tgagtggtct cgacaatggc ttatacccta cacgggataa cttaactgat atccctttta
40741 acggataaac ggagccaaca atggcagaga ttattcccat gactgaagaa cagaaattcc
40801 agttagagat ttacaaactg gtcatgaacc agaacgccgc cgtgcatgga gactggcatg
40861 aacaaagagc tccgcgtata tggcagccga tgggataagg cccgtctgcg ttttctccag
40921 cagcacccac tgtgtgtgat gtgcgagcag caggggcgca taacaccagc aacggtggtc
40981 gaccatatcg tgccccacaa actgaaagat gcgcttaagt caggtaaccc tctggccata
41041 tcgaaagcac agctcctgtt ctggagtaaa gagaactggc agccactgtg caaagcgcat
41101 catgactcaa cgaaacagag aatggagaag agcggcgcgg taataggttg tgatgccaac
41161 ggctacccgc tcgatcctgc gtctcactgg agcacgtaat gaaagacctc atcattgaat
41221 accgagacgg taagtttgtt cagctggcga ttgatggcgt ggagatgaag cgcgtaacgt
41281 ctatccagtt ctcccacacc gtaggcgagg acgtaccgac attgaccgtc tcaggacatg
41341 tgtggtccga gtatgggaaa ggcgatcaca aactcgaaca ggtagacaaa cattcggcat
41401 agcgcggcgg cggcaagtcg attatctatc atgtgaaatc atttcaaatg caacgatatc
41461 aaatgagaat gaatcgcatt catggcaggg ggggggatca aatcttcaaa acctttttccc
41521 taaatgaccg ccgccaaa