Project Number: MQP-BC-DSA-0639

RECOMBINANT CLONING AND EXPRESSION OF VIRAL PROTEINS FROM WEST NILE AND DENGUE VIRUSES

A Major Qualifying Project Report

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

in

Biochemistry

By

Sheila Ginés

April 28, 2005

APPROVED:

Irene Bosch, Ph.D.
Center for Infectious Diseases
UMass Medical Center
Major Advisor

Dave Adams, Ph.D.
WPI Project Advisor

ABSTRACT

Recent studies with Dengue (DEN) and West Nile (WN) viruses have focused on structural analyses, and host cellular and immunological responses to viral infection. In this project, various DEN and WN proteins were expressed in *E. coli* using recombinant expression plasmids. Expression was verified by RT-PCR and immunoblots. The expressed proteins will be used in the future to test T-cell reactivities against viral proteins.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I owe the privilege of working at the Center of Infectious Diseases and Vaccine Research (CIDVR) in the University of Massachusetts Medical School (Worcester, MA) to Dr. Alan Rothman and Dr. Irene Bosch. I am appreciative of Kris Giaya and Rajas Warke for taking their time to instruct and direct me in the lab. In addition, all the researchers at CIDVR were generous and helpful with their time, knowledge, workspace, and materials. I would also like to thank the Wadsworth Center of the New York State Department of Health for providing the Replicon of the West Nile Virus of the New York strain. Finally, I am very grateful to Dr. Dave Adams at WPI for assisting with the project initiation, and guidance with the writing of this report.

BACKGROUND

Flaviviruses

The Flaviviridea family has a multitude of viruses that cause disease in humans. The International Committee of Taxonomy of Viruses (ICTV) has subdivided the Flaviviridea family into three genera: Flavivirus, Pestivirus, and Hepacivirus. Flaviviruses include 69 pathogens such as Dengue (DEN), West Nile (WN), and Yellow Fever viruses. Some of these pathogens are dangerous arthropod-borne diseases, which cause millions of human illnesses a year. Carriers are typically mosquitoes, sand flies, ceratopogonids, and ticks. Several of these dangerous viruses are being studied for defense against bioterrorist attacks (Gaidos, 2002). The Pestivirus genus consists of three serotypes of bovine viral diarrhea, but there are no known human pathogens. The Hepacivirus genus includes viruses such as Hepatitis C.

The focus of this project is two pathogens in the flavivirus family, DEN and WN. Flavivirus genomes consist of monopartite linear, single stranded, positive sense RNA. Because of the positive sense RNA, the nucleic acid is directly capable of initiating infection in an appropriate host cell. The virons are known to be spherical with a diameter of 40-65 nm (see Figure 1).

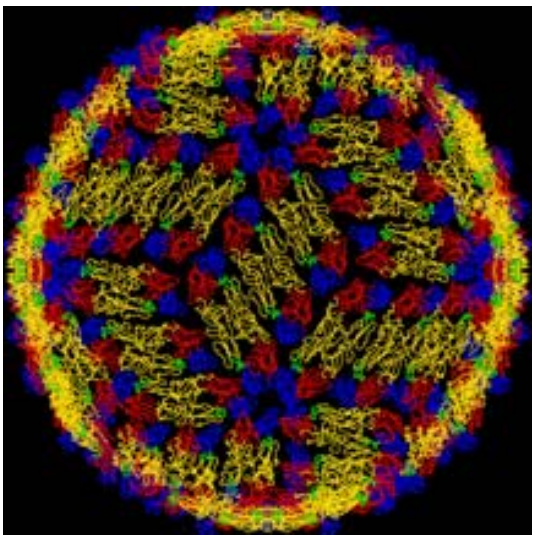

Figure 1: First Diagram of the Structure of a Flavivirus.
Purdue University and California Institute of Technology were the first to first to solve the three-dimensional structure of a Flavivirus, in this case DEN. According to Richard Kuhn, a Purdue University researcher, the structure "reveals an architectural structure that is different from any other virus that has been seen"(Gaidos, 2002). This illustration shows the E protein organizing itself to form a protective shell around the virus (Purdue University computer illustration).

All members of the flavivirus family are transmitted by arthropods (mosquitoes and ticks). The virus is spread to humans by the bite of an infected female mosquito, but is not spread through contaminated body fluids. Flaviviruses are capable of reproducing in their vector, which is very important for viral transmission because it allows the virus to pass from one host to the next (Flavivirus, 1999).

Dengue Virus (DEN)

Dengue fever is a vector-borne infectious disease from the tropical and subtropical urban areas of the world. The vector for DEN is mainly the *Aedes aegypti* mosquito, but sometimes *Aedes albopictus*. The *Aedes aegypti* mosquito is a daytime biting mosquito that usually only feeds on humans and it's the most common DEN vector

(CDC Dengue Fever Home Page, 2005). However, *Aedes albopictus* was responsible for the 2001 outbreak in Hawaii (Dengue Q&A, 2005).

There are four different serotypes of DEN (DEN-1, DEN-2, DEN-3, and DEN-4), with genetic variation within each serotype. Some of the serotypes appear to be more virulent than others (Dengue Fever Slides, 2003). Infection and recovery from one serotype only provides immunity for that specific serotype (Fact Sheet, 2003). However there is sometimes short-term cross-immunity against the other serotype, but this may only last for a couple of months (Dengue Fever Slides, 2003). An individual can experience four different DEN infections in a lifetime (Fact Sheet, 2003). The severe form of infection usually occurs after a secondary infection (Tassaneetriethep et al, 2003). Although this process is not fully understood, the generally accepted reason for this is an antibody-dependent enhancement process in which the antibodies induced in the primary infection facilitate the uptake of the secondary virus into white blood cells, and crossreacts with the serotype of the secondary infection contributing to a more severe form of infection (Se-Thoe et al, 2000).

DEN transmission occurs as a cycle (Figure 2). The cycle begins when a female mosquito feeds on a person who has already been infected with the virus, however, the mosquitoes are only infected by feeding on a viremic person with viruses present in the bloodstream. When a person is first infected with the virus, they go through an average incubation period of four to seven days but it can range from three to fourteen days. Viremia begins at the end of incubation, slightly before the onset of symptoms and lasts up to five days. The symptom onset lasts about three to ten days, which is several days after viremia has ended. After the female mosquito has fed on a person during the

viremic period, that mosquito becomes infected with the virus. If that mosquito feeds on another person, it can transmit the virus to that person. Viral transmission can only occur between mosquito and human, not from human to human.

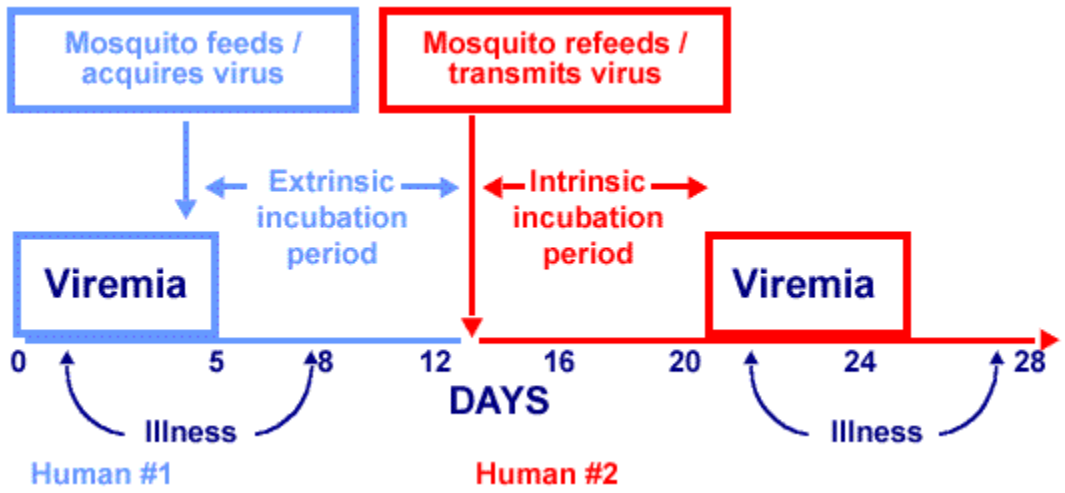

Figure 2: DEN Transmission Cycle. A mosquito becomes infected with virus by feeding on viremic human #1 with virus in the bloodstream. The infected mosquito then feeds on human #2 infecting him with the virus. When first infected with the virus, there is an intrinsic incubation period of four to seven days. The viremic period occurs toward the end of the incubation period. The onset of symptoms begins a day or two after viremia has begun, and lasts for three to ten days (CDC image).

Once the virus is inoculated into a person by a mosquito, the virus targets specific organs, such as the lymph nodes and the liver. Then the virus infects white blood cells and lymphatic tissues, then is released into the bloodstream and spread throughout the body.

The principal symptoms of DEN are high fever, severe headaches, backaches, joint pains, nausea and vomiting, eye pain, and rash. The more severe symptoms include shock and hemorrhage. Most DEN infections are mild, but some result in dengue hemorrhagic fever (DHF) or dengue shock syndrome (DSS). DHF is characterized by a fever that last 2-7 days, followed by hemorrhagic manifestations, skin hemorrhage,

bleeding of gums or nose, and possible internal bleeding. When the small blood vessels (capillaries) become excessively permeable, fluid components escape from blood vessels, and this may cause failure to the circulatory system and shock, closely followed by death if the circulatory problem is not corrected (Dengue Q&A, 2005).

There are millions of incidences of DEN worldwide. It is estimated that there are 50 to 100 million cases of DEN each year, and hundreds of thousands of DHF cases. The average fatality rate for DHF is 5% (Fact Sheet, 2003).

There is no specific treatment or medication for DEN. People who believe they have DEN are recommended to contact a physician, drink plenty of fluids, get plenty of rest, take pain relievers with acetaminophen, and avoid those that contain aspirin (Dengue Q&A, 2005). Recently, a vaccine candidate for DEN was developed in Thailand. Research is also being conducted to develop a vaccine for secondary infection. These vaccines are still in testing mode and will not become available for the public for another 5 to 10 years (CDC Dengue Fever Home Page, 2005).

Much remains unknown about DEN virus. Numerous clinical studies are in progress to further understand this virus. Research is also focusing on the development of a tetravalent vaccine that will target all four serotypes of DEN. There is also a need to develop rapid, sensitive, and specific diagnostic tests, and more effective prevention programs.

West Nile Virus

WN is a mosquito-borne virus from the temperate regions of the world, most recently in Europe and North America. WN transmission and infection (Figure 3) occurs when a female mosquito feeds on a nonhuman vertebrate (like a bird). Then the infected mosquito feeds on a person and transmits the virus through its saliva. WN is primarily transmitted by members of the *Culex* mosquito species.

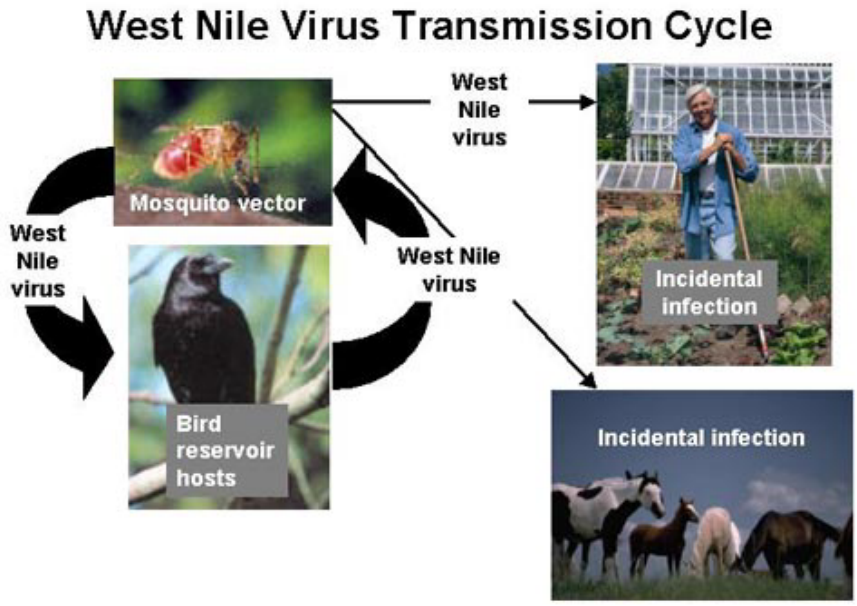

Figure 3: Transmission Cycle of the WN Virus. A mosquito becomes infected by feeding on a viremic reservoir host (usually a bird), and then the mosquito feeds on humans, domestic animals, or horses transmitting the virus to them. The end hosts are known as incidental host because they do not develop infectious levels of viremia. Mosquitos are not capable of acquiring virus from incidental hosts, only from reservoir hosts (CDC image).

WN is mainly spread from mosquito to vertebrate, rarely from vertebrate to vertebrate. Mosquito's are the primary vector, but the WN virus does not grow well there, so vertebrates are the host. Hosts include humans, domestic animals (dogs, birds, and horses), and wild birds. Humans and most other mammals are known to develop

infectious levels of viremia (virus in the bloodstream), but aren't reservoir hosts. They are usually known as incidental-hosts or "dead ends." If a mosquito feeds on a WN infected mammal, the mosquito doesn't become infected with the virus, it merely passes it on to the next host bitten. Because bird viremia is thousands of times higher than for humans, the mosquito usually acquires the virus from birds (West Nile Virus: Entomology, 2004).

A very small number of cases of WN are known to be spread through blood transfusion, organ transplants, and mother to child. Mother to child instances occur during pregnancy and breastfeeding (West Nile Fact Sheet, 2004).

WN affects the central nervous system, and symptoms usually vary from case to case. WN is mainly known to cause encephalitis (inflammation of the brain) and meningitis (inflammation of the lining of the brain and the spinal cord). About 80% of the people infected with the virus don't show any symptoms. The other 20% of people infected display symptoms including fever, headache, body aches, nausea, and vomiting. Sometimes symptoms include swollen lymph glands or skin rash on the chest, back, and stomach. In the more severe cases of WN, symptoms include high fever, headaches, neck stiffness, stupor, disorientation, coma, tremors, convulsions, muscle weakness, vision loss, numbness, and paralysis. In the milder cases, symptoms last a few days but can last longer. In the more severe cases, symptoms last several weeks, and neurological effects can be permanent (West Nile Fact Sheet, 2004).

WN virus has become a predominant health risk in North America, primarily in the US. According to the CDC, in 2002 there were 4,156 reported cases of WN human disease in the US. 3000 of these cases were central nervous system (CNS) disease cases,

and nearly 10% of them were fatal. In 2003, there were 9862 human cases of WN disease, and 264 cases were fatal (West Nile: Q&A, 2004). WN cases decreased in 2004, 2470 cases were reported to the CDC and 88 of them were fatal (West Nile Virus: Statistics, Surveillance, and Control, 2005). Less than 1% of the people infected with WN develop encephalitis, which is the more severe symptom of WN. Of those who develop encephalitis, 3% to 15% are fatality cases (West Nile Virus, 2002).

There is no specific treatment or therapy for WNV. People who develop symptoms are recommended to contact a physician, drink plenty of fluids and get plenty of rest. In the more severe cases hospitalization is highly recommended. Patients will need intravenous fluids and nutrition, airway management, ventilatory support, and prevention of secondary infections (West Nile Virus, 2002).

Much remains unknown about WNV. Numerous clinical studies are in progress to further understand this virus. Research is focusing on the development of a vaccine; rapid, sensitive, and specific diagnostic tests; and more effective prevention programs.

Viral Proteins

DEN and WN positive-stranded RNAs encode one polyprotein. The polyprotein produces ten smaller proteins, three structural proteins and seven non-structural proteins. Structural proteins include: C, prM, and E. Non-structural proteins include: NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 (Figures 4 and 5).

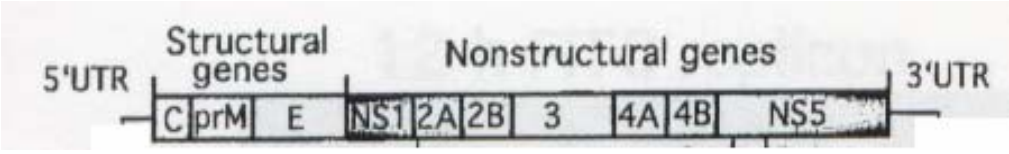

Figure 4: General Structure of the WN and DEN Genome. Image illustrates the orientation of the genes in both DEN and WN. Structural proteins come first [C (capsid), prM (membrane/premembrane), and E (envelope)], followed by the nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5).

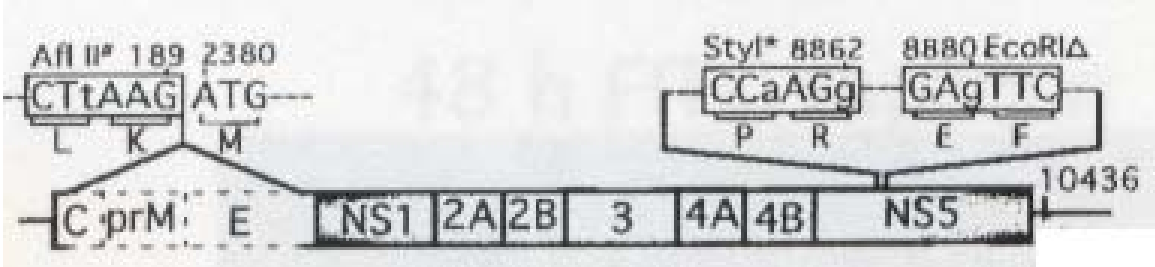

Figure 5: WN Replicon Structure. WN replicon was used to target the nonstructural proteins. Replicon is a WN structure with deletion of structural proteins (deletion indicated by dotted lines).

The C protein is known as the capsid protein (or core protein), it's a highly basic component of the nucleocapsid (Figure 6). It is believed that the C protein functions in neutralizing the negatively charged viral RNA. There are two types of membrane (M) protein: prM and M. M resides in the extracellular mature virions, and prM resides in the intracellular mature virions. M and prM are usually grouped together as the prM (membrane) protein (Chang, 1997). prM is cleaved when the virus exits the cell, cleavage forms "pr" and "M" segments. The hydrophilic "pr" segment gets secreted into the extracellular medium; and "M" becomes a structural part of the lipid envelope (Campbell et al, 2002). Cleavage of prM results in the formation of mature virion with the ability to infect other cells (the non primary target cells). E is the envelope protein, the major part of the virion surface. E protein is associated with several biological functions, receptor binding and membrane fusion. It is also a key target of the immune response to flaviviruses (Chang, 1997). The E protein organizes itself to form a

protective shell around the virus; the shell serves as a cage for the genetic material inside the virus (Gaidos, 2002). The viral envelope and the membrane proteins are responsible for many of the important properties of the virus; host range, tissue tropism, replication, assembly, and stimulation of B and T cell immune responses (Campbell et al, 2002).

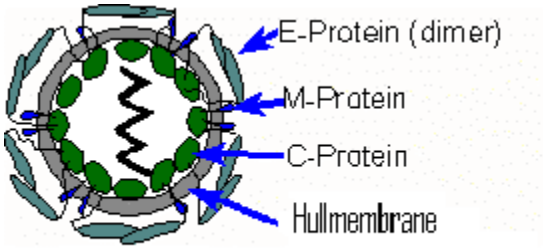

Figure 6: Spherical Structure of Flaviviruses. This image illustrates the spherical structure of flaviviruses and the general orientation of the structural proteins. The surface is usually smooth; its membrane is enclosed by a protein shell created by the envelope protein. (Universtat Wien image)

The functions of the non-structural proteins are not as well classified as the functions of the structural proteins. NS1 is expressed on the surface of infected cells, thus it is believed to be a target of the immune response. It's also believed to be a soluble complement-fixing antigen (Chang, 1997). NS2A works in tandem with NS3 as a protease. NS3 is a flavivirin protease. NS2B is a flavivirin protease of regulatory subunits. NS4A is a catalytic subunit. NS5 is an RNA dependent RNA polymerase (Castle et al, 1986). NS5 activities include RNA capping and viral RNA replication (Chang, 1997). In DEN-2, NS2A, NS4A, and NS4B have recently been associated with interferon antagonists (Munoz-Jordan et al, 2003).

Gateway Technology

Gateway Technology (Gateway) by Invitrogen Corp. was the cloning technology used in this project. It was used because of its highly efficient and rapid route to

functional analysis, protein expression, and cloning/subcloning of DNA segments. Gateway maximizes compatibility and flexibility, and it also minimizes the planning process. There is no need of restriction enzymes, gel purification, or ligation when using Gateway. Gateway allows site specific recombination, transfer of one or more genes in one or more expression vectors in one experiment, yet any vector can be made compatible with the technology.

One of Gateway's greatest strengths is the capability of going directly from cloning to protein expression. It provides a rapid optimized recombinant protein expression. Gateway allows DNA segments to be transferred between different cloning vectors while maintaining its orientation and reading frames. This is possible because of the site specific recombination system of phage 1, which effectively replaces the use of restriction endonuclease and ligase. A main goal of Gateway is to move gene(s) from one vector to another. There are various host systems in Gateway. The best host system needs to be chosen for each specific experiment. Once the gene is cloned in a system, the sequence can be moved freely without concern of restriction enzyme site, reading frame, or insert orientation.

The Gateway reactions are of high efficiency, fast, and simple. They represent the in vitro version of integration and excision reactions. The two main Gateway reactions are the BP recombination reaction and the LR recombination reaction.

BP Recombination Reaction

An entry clone is created by using the BP recombination reaction. The reaction transfers the gene of interest to a donor vector to create an entry clone. The BP

recombination reaction is a lysogenic pathway, catalyzed by a bacteriophage λ integrase (Int) and *E. coli* Integration Host Factor (IHF) protein, that represent the BP Clonase enzyme mix. The BP reaction is:

attB x attP → attL x attR.

The BP reaction is used to facilitate the recombination of attB DNA (i.e. PCR product or linearized expression clone) with an attP DNA (i.e. donor vector, pDONOR 201 or pDONOR 221) to create an attL (substrate containing an entry clone). The BP reaction is catalyzed by the BP Clonase enzyme mix (Figure 7).

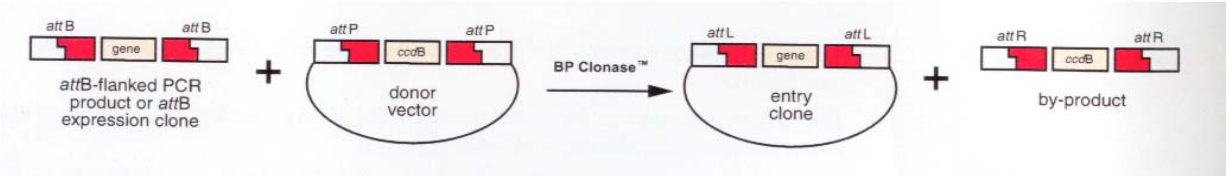

Figure 7: The Gateway BP Recombination Reaction. BP reaction facilitates the recombination of attB substrate with an attP substrate to create and attL clone. Reaction is catalyzed by BP clonase (Gateway image).

LR Recombination Reaction

An expression clone is created by using the LR recombination reaction. The reaction transfers the gene of interest to a destination vector to create an expression clone. The LR reaction is a lytic pathway, catalyzed by a bacteriophage λ integrase (Int), Excisionase (Xis), and *E. coli* Integration Host Factor (IHF) protein, each of which are included in the LR Clonase enzyme mix. The LR reaction is:

attL x attR → attB x attP

The LR reaction is used to facilitate the recombination of attL substrate (entry clone) with a attR substrate (destination vector, pDEST17, pDEST26) to create an attB

(substrate containing an expression clone) clone. The LR reaction is catalyzed by the LR Clonase enzyme mix (Figure 8).

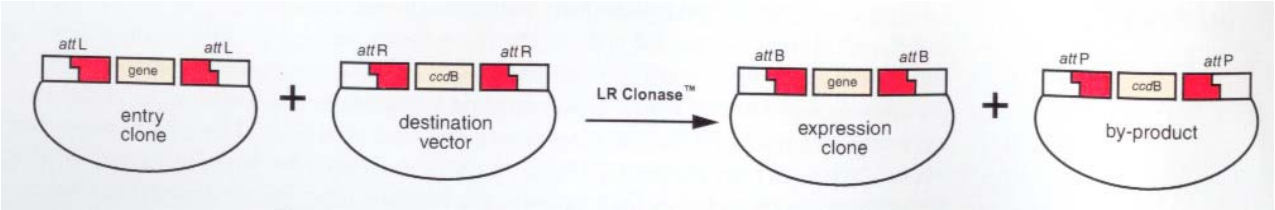

Figure 8: The LR Recombination Reaction. The LR reaction facilitates the recombination of an attL substrate with an attR substrate to create and attB clone. The reaction is catalyzed by LR clonase (Gateway image).

Site	Length	Found in
AttB	25 bp	Expression vectors Expression clone
AttP	200 bp	Donor vector
AttL	100 bp	Entry vector Entry clone
AttR	125 bp	Destination vector

Table 1 Characteristics of att Sites. attB1 sites only react with attP1 sites, attB2 sites only react with attP2 sites, attL1 sites only react with attR1 sites, and attL2 sites only react with attR2 sites.

UMASS Center of Infectious Disease and Vaccine Research

The research for this project was conducted at the University of Massachusetts Medical School at the CIDVR department under the supervision of Irene Bosch. CIDVR's mission "is to conduct research in and training for the study of the molecular basis of human diseases caused by infectious agents and the development of safe and effective vaccines against these agents (Center for Infectious Disease and Vaccine Research)." The ultimate goal of the research at CIDVR is to understand the molecular mechanisms of disease pathogenesis, to define treatment of viral diseases, and develop

effective vaccines to prevent infection. CIDVR's clinical base research focuses on defining human T lymphocyte responses to viral infection and their role in protective immunity and disease pathogenesis. CIDVR's research focuses on a variety of pathogens, flaviviruses (DEN, WN, yellow fever, and Japanese encephalitis virus), poxviruses (vaccinia virus), hantaviruses, influenza virus, hepatitis C virus, and HIV.

DEN and WN studies focus on defining immunological responses responsible for the increase severity of DEN disease during secondary infection. CIDVR collaborates with several international clinical studies, such as Thailand, Hawaii, Venezuela and Puerto Rico. In Thailand, clinical studies are currently analyzing viral replication of DEN and the immune response of children with acute DEN infection. They are also studying viral transmission and the disease in school children (Research Projects, CIDVR).

Dr. Bosch's lab focuses on understanding the cellular response to viral infection from RNA viruses (DEN, WN, and yellow fever). The main focus of the lab is monocyte infections and the exploration of inflammatory responses (Bosch Research, 2005).

Recently the lab researched the increased production of interleukin-8 in primary human monocytes and in human epithelial and endothelial cell lines after dengue virus challenge (Bosch et al, 2002). Another recent study was that dengue virus induces novel changes in gene expression of human umbilical vein endothelial cells (Warke et al, 2003).

Bosch's lab also participates in the transfer of technology to third world countries for the advancement of science. There is a constant active participation in exchanges and activities outside the US (Bosch Research, 2005).

PROJECT OBJECTIVE

The pathogenesis of Dengue (DEN) and West Nile (WN) virus are not well understood at this time, and therapeutic interventions do not exist. The UMASS CIDVR has been focusing on understanding the cellular and immunological responses of these viruses to allow further understanding of the pathogenesis of these viruses and hopefully lead to the development of a treatment. In order to understand a virus as a whole, one needs to learn about its components, thus the focus of this project is the viral proteins. Dr. Bosch's lab is interested in cloning and expressing all the viral proteins for the DEN and WN viruses that will be used to test T-cell reactivities from infected patients, or to create antibodies to facilitate further analysis of protein function. The goal of this MQP was to finish cloning and expressing several remaining viral proteins of DEN-2 and WN virus in bacterial and mammalian cells.

METHODS

RT-PCR

The reverse transcriptase-polymerase chain reaction (RT-PCR) technique was used to obtain the cDNA of several viral proteins. The viral RNA was provided by the CIDVR. Forward and reverse primers were also provided by CIDVR.

RT reactions were carried out in 20 μL reactions with 2 μL of 10x Buffer, 2 μL of dNTP (10 μM), 2 μL of specific forward primer, 2 μL of specific reverse primer, 1 μL of RT enzyme, 0.5 μL of RNase inhibitor, 7.5 μL of RNase free water, and 3 μL of RNA. The RT condition was 60 minutes at 37°C, 5 minutes at 93°C, and 4°C pause. PCR reactions were carried out in 50 μL reactions with 23 μL of H_2O, 15 μL of rtTH buffer, 1 μL dNTP (10 μM), 2 μL of specific forward primer, 2 μL of specific reverse primer, 1.8 μL of MgOAc (25 μM), 1 μL of rtTH polymerase, and 4 μL of cDNA from RT reaction. The PCR condition was 60 seconds at 93°C; then 35 cycles of 15 seconds at 94°C and 6 minutes at 60°C; then 10 minutes at 72°C, and 4°C pause.

RT-PCR Product Confirmation

A 1.5% agarose gel in TBE buffer was run with 20 μL of samples. 20 μL samples were prepared by combining 16 μL of RT-PCR product and 4 μL of loading dye. 2 μL of Molecular Weight Marker (100 bp ladder) was also loaded on a separate lane as a reference point.

Gel Extraction and DNA Purification

Positive DNA bands from RT-PCR product confirmations were excised from the gel for DNA purification. The QIAGEN QIAquick Gel Extraction Kit, a microcentrifugation protocol, was used for DNA extraction and purification. (See Appendix 3 for protocol.) After gel extraction and DNA purification, samples were quantified by UV absorbance.

BP Reaction

DNA mass values determined by UV absorbance were converted to fmol using the following formula:

$$ng = (fmol)(N)(\frac{660fg}{fmol})(\frac{1ng}{10^6fg})$$

N represents the length of the DNA in base pair (bp).

In a 1.5 mL microcentrifuge tube the following components were mixed at room temperature: 1-10 µL (40-100 fmol) attB-PCR product, 2 µL pDONOR vector (150 ng/µL), 4 µL 5X BP Clonase Reaction Buffer, and enough TE Buffer (pH 8.0) to make a solution volume of 16 µL. The solution was mixed well, 4 µL of BP Clonase enzyme (stored at -80°C) was added to each sample, and samples were mixed again by vortexing. Samples were later incubated at 25°C for 1 hour. After incubation, 2 µL of Proteinase K solution was added to each sample and incubated for 10 minutes at 37°C. Then samples were transformed in competent *E. coli* cells.

Cell Transformation

An appropriate antibiotic based vector (entry vector or destination vector) was needed for transforming competent *E. coli* to select for entry clones. Entry vectors were used in the BP reaction, which required kanamycin (KAN) antibiotic. Samples from the BP reaction were transformed in 50 μL of competent *E. coli* cells, DH5α or Top 10 cells.

1 μL of BP recombination reaction was added to E. coli. Samples were incubated for 10 minutes on ice. After ice incubation, samples were heat shocked for 30 seconds at 42°C. Tubes were immediately transferred to ice after heat shock. Then 450 μL of SOC medium was added to the samples. Samples were then incubated for an hour at 37°C while shaking horizontally at 200 rpm. After incubation, 100 μL of transformation was spread on pre-warmed KAN plates and incubated overnight at 37°C. Colonies were picked the following day and cultured overnight in LB-KAN medium.

Plasmid Purification

Plasmid DNA was prepared from the overnight cultures of the cell transformation using a QIAGEN QIAprep Miniprep Kit, microcentrifuge protocol. See Appendix 4 for protocol.

Plasmid Confirmation

The purified plasmids were analyzed by polymerase chain reaction (PCR). PCR reactions were carried out in 50 μL with 39.5 μL of H_2O, 5 μL of PCR buffer ($MgCl_2$), 1 μL dNTP (10 mM), 1 μL of specific forward primer, 1 μL of specific reverse

primer, 1.5 μL of Taq Polymerase (5 μ/μL), and 2 μL of plasmid from Miniprep. The PCR condition was 60 seconds at 94°C; followed by a 28 cycles of 60 seconds at 58°C, 45 seconds at 72°C, and 30 seconds at 94°C; then 60 seconds at 58°C, 10 minutes 72°C, and a 4°C pause.

After PCR, plasmids were visualized by running a 1.5% agarose gel in TBE buffer. 20 μL samples were prepared by combining 16 μL of PCR product and 4 μL of loading dye. 2 μL of Molecular Weight Marker was also loaded on a separate lane as a reference point.

Plasmids were also confirmed by sequencing. Samples were sent to the UMASS sequencing center for sequencing. Sequence confirmations were reviewed by Irene Bosch and Kris Giaya.

LR Reaction

In a 1.5 mL microcentrifuge tube the following components were mixed at room temperature, 1-11 μL (100-220 ng/reaction) entry clone, 1-11 μL pDEST vector (300 ng/reaction), 4 μL 5X LR Clonase Reaction Buffer, and enough TE Buffer (pH 8.0) to make solution volume 16 μL. The solution was mixed well, 4 μL of LR Clonase enzyme (stored at -80°C) was added to each sample, and samples were mixed well by vortexing. Reaction was incubated at 25°C for 1 hour. After incubation, 2 μL of Proteinase K solution was added to each sample and incubated for 10 minutes at 37°C. Then samples were transformed into competent *E. coli* cells.

Cell Transformation

An appropriate antibiotic based vector (entry vector or destination vector) was needed for transforming competent *E. coli* to select for entry clones. Destination vectors were used in an LR reaction, which required ampicillin (AMP) antibiotic. Samples from the LR reaction were transformed in 50 μL of competent *E. coli* cells, (DH5α or Top 10).

1 μL of LR recombination reaction was added to competent *E. coli*. Samples were incubated for 10 minutes on ice. After ice incubation, samples were heat shocked for 30 seconds at 42°C. Tubes were immediately transferred to ice after heat shock. Then 450 μL of SOC medium was added to the samples. Samples were then incubated for an hour at 37°C while shaking horizontally at 200 rpm. After incubation, 100 μL of transformation was spread on pre-warmed AMP plates and incubated overnight at 37°C. Colonies were picked the following day and cultured overnight in LB-AMP medium.

Plasmid Purification

Plasmid DNA was isolated using a QIAGEN QIAprep Miniprep Kit. See Appendix 4 for protocol.

Plasmid Confirmation

The purified plasmids were analyzed by polymerase chain reaction (PCR). PCR reactions were carried out in 50 μL with 39.5 μL of H_2O, 5 μL of PCR buffer ($MgCl_2$), 1 μL dNTP (10 mM), 1 μL of specific forward primer, 1 μL of specific reverse primer, 1.5 μL of Taq Polymerase (5 μ/μL), and 2 μL of plasmid from Miniprep. The PCR condition was 60 seconds at 94°C; followed by a 28 cycles of 60 seconds at 58°C,

45 seconds at 72°C, and 30 seconds at 94°C; then 60 seconds at 58°C, 10 minutes 72°C, and a 4°C pause.

After PCR, plasmids were visualized by running a 1.5% agarose gel in TBE buffer. 20 μL samples were prepared by combining 16 μL of PCR product and 4 μL of loading dye. 2 μL of Molecular Weight Marker was also loaded on a separate lane as a reference point.

Plasmids were also confirmed by sequencing. Samples were sent to the UMASS Sequencing Center for sequencing. Sequence confirmations were reviewed by Irene Bosch and Kris Giaya.

Expression of Recombinant Protein

Cell Transformation

Samples from an LR miniprep (1 μL) were transformed in 50 μL of competent BL21-A1 cell. Samples were incubated on ice for 10 minutes. After ice incubation, samples were heat shocked for 30 seconds at 42°C. Tubes were immediately transferred to ice after heat shock. Then 250 μL of SOC medium was added to the samples. Samples were then incubated for 30 minutes at 37°C while shaking horizontally at 200 rpm. After incubation, 100 μL of transformation was added to 1 mL of LB-AMP medium and incubated overnight at 37°C.

After the cultures were grown overnight, 50 μL of culture was plated on pre-warmed AMP plates and incubated overnight at 37°C. Colonies were picked the

following day and cultured overnight in LB-AMP medium at 37°C while shaking horizontally at 200 rpm.

Induction

10 μL of colony culture and 3 mL of LB-AMP medium was mixed and incubated at 37°C while shaking horizontally at 200 rpm for about an hour. Samples were incubated until their OD_{600} was greater than 0.4 nm. The 3 mL of the culture was aliquotted into separate 1mL samples for induction. For the IPTG induction, 1 μL of IPTG was added to the 1 mL culture. For the arabinose induction, 50 μL of arabinose was added to the 1 mL culture. All samples were mixed well, and then they were incubated for about 4-6 hours at 37°C while shaking horizontally at 200 rpm. The incubation was run until the OD_{600} was greater than 0.7 nm, preferably around 1.5 nm.

After induction, samples were centrifuged at high speed for 5 minutes. The supernatant was discarded and the pellet was resuspended in 100 μL of PBS protease inhibitor buffer mix (Roche). Cells were then lysed by sonication on ice for 30 seconds. 20 μL of sample and 20 μL of sample buffer were mixed and boiled at 82°C for 5 minutes.

Protein Electrophoresis

Cell lysate proteins were separated by SDS-PAGE gel electrophoresis. The resolving gel consisted of 11.25 mL of 1M Tris pH 8.8, 10 mL of 30% acrylamide, 8.75 mL of dH_2O, 150 μL of fresh 10% APS, and 30 μL of TEMED. The solution was mixed well and poured onto the gel apparatus. Before the stacking gel was added onto the

resolving gel, dH_2O was added to the top of the resolving gel to level the resolving gel, it was allowed to dry for 10 minutes, and then the water was removed. The stacking gel consisted of 1.25 mL of 1M Tris pH 6.8, 1.25 mL of 30 % acrylamide, 7.41 mL of dH_2O, 75 µL of fresh 10% APS, and 18 µL of TEMED. The solution was mixed well and poured on top of the resolving gel. A comb was placed on the stacking gel in the gel apparatus. The gel was allowed to polymerize for about an hour, then the comb was removed once gel solidified.

Proteins were run through the gel in Tris/Glycine/SDS Buffer (10x) at 200 volts, 3 amps, and 200 watts for 2-4 hours. Gel was run until the bromophenol blue dye was approximately 2 cm from the base of the gel. The gel was removed from apparatus and stained in Safe Stain solutions for 1 hour at room temperature while shaking. Destaining was performed in dH_2O overnight.

RESULTS

The purpose of this project was to use genetic engineering to express a variety of DEN-2 and WN virus genes for future analysis of the produced proteins against patient T-cell responses. Table 2 below lists the sizes of the known DEN and WN genes used to verify RT-PCR DNA amplicon sizes on agarose gels during cloning experiments.

Gene Name	Gene Lengths (bp)	
	DEN-2	WN
C	341	369
prM	497	278
E	1484	1503
NS1	1055	1056
NS2A	653	693
NS2B	389	393
NS3	1844	1857
NS4A	858	447
NS4B	336	765
NS5	2699	2715
Total Genome	10156	10060

Table 2: DEN-2 and WN Genes. Table consists of all the DEN and WN genes and their genome sizes in base pairs. This table was used to verify appropriate RT-PCR amplicon band sizes on agarose gels.

Cloning DEN-2 Genes

DEN-2 entry clone plasmids (C, prM, NS1, NS2A, NS2B, NS4A, NS4B, and NS5) were provided to Bosch's lab by Rothman's lab at the UMASS Center for Infectious Diseases and Vaccine Research. Entry clone plasmids were used to subclone the DEN-2 genes into expression plasmid pDEST 17 by using a LR reaction. Reactions were transformed into Top 10 *E. coli* cells, then plasmid DNA was prepared from amp^r

colonies, and screened by PCR (see Figure 9). Table 3 shows the PCR results of the expression plasmid screening. Positive pDEST 17 expression plasmids were obtained for 6 of the 8 DEN-2 genes tested (clones were not obtained for NS1 and NS4B).

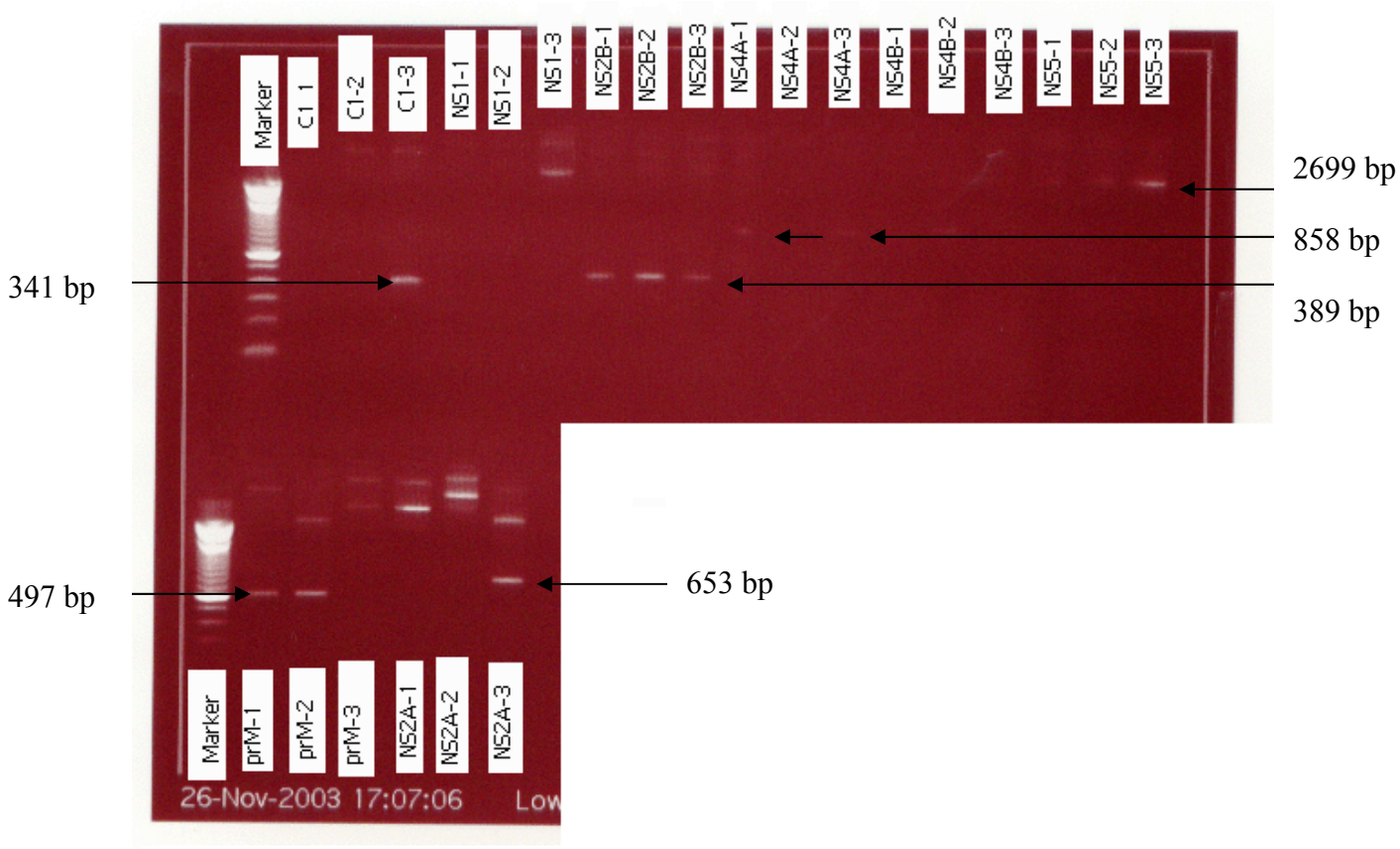

Figure 9: DEN-2 Gene PCR Screening of LR Reaction pDEST 17 Plasmids. LR reactions were transformed into Top 10 E. coli cells. Plasmid DNA was extracted from amp^r colonies, then screened for DEN-2 genes by PCR. Arrows denote the sizes in bp of positive clones.

Plasmid	Vector	Cell line	OD (ng/μL)	Gel Lane	Clone result (+/-)
C1-1	pDEST 17	Top 10	30.6	2	-
C1-2	pDEST 17	Top 10	65.2	3	-
C1-3	pDEST 17	Top 10	37.3	4	+
NS1-1	pDEST 17	Top 10	102.4	5	-
NS1-2	pDEST 17	Top 10	102.5	6	-
NS1-3	pDEST 17	Top 10	190.4	7	-
NS2B-1	pDEST 17	Top 10	78.6	8	+
NS2B-2	pDEST 17	Top 10	108.1	9	+
NS2B-3	pDEST 17	Top 10	79.4	10	+
NS4A-1	pDEST 17	Top 10	101.4	11	+

NS4A-2	pDEST 17	Top 10	112	12	-
NS4A-3	pDEST 17	Top 10	158.9	13	+
NS4B-1	pDEST 17	Top 10	102.8	14	-
NS4B-2	pDEST 17	Top 10	96.2	15	-
NS4B-3	pDEST 17	Top 10	93	16	-
NS5-1	pDEST 17	Top 10	100	17	+
NS5-2	pDEST 17	Top 10	134.6	18	+
NS5-3	pDEST 17	Top 10	128.5	19	+
Prm-1	pDEST 17	Top 10	32.9	20	+
Prm-2	pDEST 17	Top 10	64.6	21	+
Prm-3	pDEST 17	Top 10	41.2	22	-
NS2A-1	pDEST 17	Top 10	225.7	23	-
NS2A-2	pDEST 17	Top 10	205.8	24	-
NS2A-3	pDEST 17	Top 10	178.1	25	+

Table 3: PCR Screening of the LR Reaction Cloning into Expression Plasmid pDEST 17. LR reactions were transformed into the TOP10 cell strain of *E.coli*, then plasmid DNA was isolated from amp^r colonies, and screened by PCR for specific DEN-2 genes. Shown are the sample concentrations and whether cloning was successful.

The positive DEN-2 expression plasmid samples were sent to the UMASS Nucleic Acid Facility (NAF) for sequencing.

Expression of Recombinant DEN-2 Proteins

Recombinant pDEST 17 plasmids samples were expressed in *E. coli* BL21-A1 cells using IPTG or arabinose induction. Purified plasmids from an LR recombinant reaction were transformed into BL21-A1 cells. Amp^r colonies were grown overnight, picked, then cultured overnight again. Overnight cultures were induced by IPTG or arabinose, using the experimental format listed in Table 4.

Sample	Induction type	OD_{600} (nm)
E18	Non-induced	1.578

E18	IPTG	0.701
E18	arabinose	1.549
E20	Non-induced	1.572
E20	IPTG	0.714
E20	Arabinose	1.534
NS3-15	Non-induced	1.575
NS3-15	IPTG	0.734
NS3-15	arabinose	1.541

Table 4: Induction of DEN-2 Protein Production from Recombinant pDEST 17 Plasmids. Clones E18, E20, and NS3-15 were induced until OD_{600} nm was greater than 0.7.

Incubation for induction was for 5 hours, until the $OD_{600\ nm}$ were greater than 0.7, preferably around 1.5. After the 5 hour incubation period, the non-induced and arabinose samples were above 1.5, so incubation was stopped. However, IPTG induction samples were around 0.7, these samples should have been incubated longer. Cell lysates were prepared, then analyzed by SDS-PAGE. The E18 and E20 clones were positive (data not shown) producing protein bands of the expected 50 kDa E protein sizes. The NS3-15 clone produced protein bands that were faint but of the expected 60 kDa NS3 size (data not shown). The NS3-15 samples should have been incubated longer to allow a more enhanced induction.

Cloning WNV Genes

RT-PCR was performed on WNV RNA to obtain genes NS2A, NS2B, and NS4B (Figure 10). Amplicons were purified from the gel, then used in a BP reaction in an attempt to create "entry clone" plasmids. Plasmid DNA was prepared from amp^r colonies, then screened by PCR for WN genes. Unfortunately the screening results were negative (data not shown) probably due to a low amplicon DNA yield.

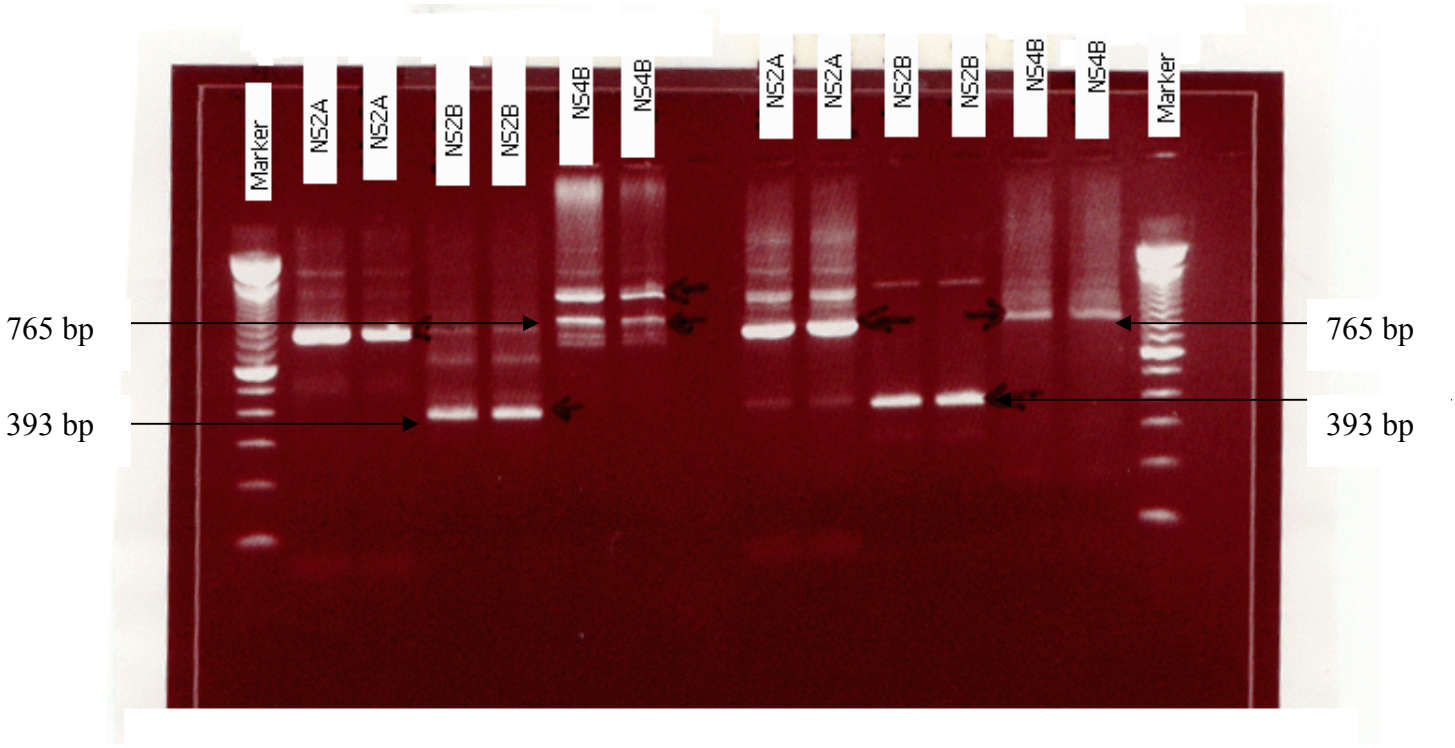

Figure 10: WN RT-PCR. RT-PCR was performed to obtain amplicons for specific WN genes. NS2A, NS2B, and NS4B were the targeted proteins. Two of the 3 produced amplicons of the expected sizes (NS2B and NS4B). The NS2A band was too high, it should have been 693 bp.

RT-PCR was redone on WN RNA to obtain the same genes as in Figure 10 plus NS5 and NS4A (data not shown), but two attempts failed to obtain the desired amplicons, probably because the viral RNA was contaminated. So a WN DNA replicon was obtained, it lacked the structural proteins, and was used instead of the RNA. WN

replicon was used to target the nonstructural proteins NS2A, NS2B, NS4A, NS4B, and NS5 (Figure 11). Correct sized amplicons were obtained for 3 of the 5 genes tested (NS2B, NS4A, NS4B), whose DNA was purified. Yields of DNA purification were tested by agarose electrophoresis (Figure 12) to confirm purification; visible bands were obtained for NS2B and NS4A, but not for NS4B. It was discovered that there was an incorrect primer sequence for NS2A, NS4B, and NS5. Primers were re-designed and ordered for later experiments.

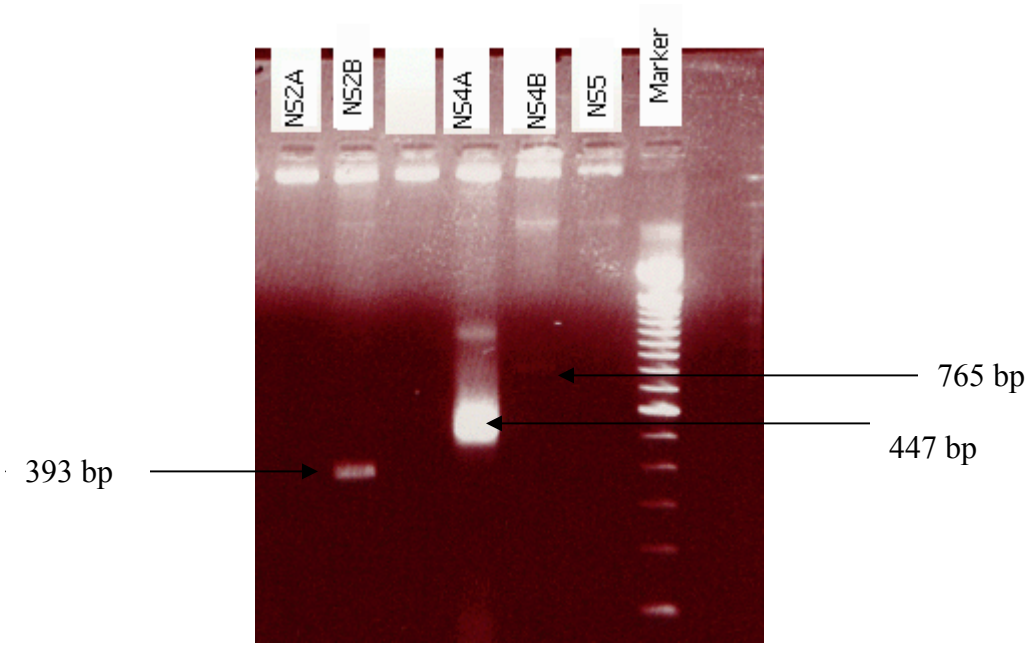

Figure 11: PCR of WNV Non-Structural Genes from a WN DNA Replicon. PCR was performed on a WN DNA replicon in an attempt to obtain cDNAs for NS2A, NS2B, NS4A, NS4B, and NS5. Positives were obtained for NS2B, NS4A, and NS4B. Arrows denote the expected gene sizes in bp. There were no visible bands for NS2A and NS5.

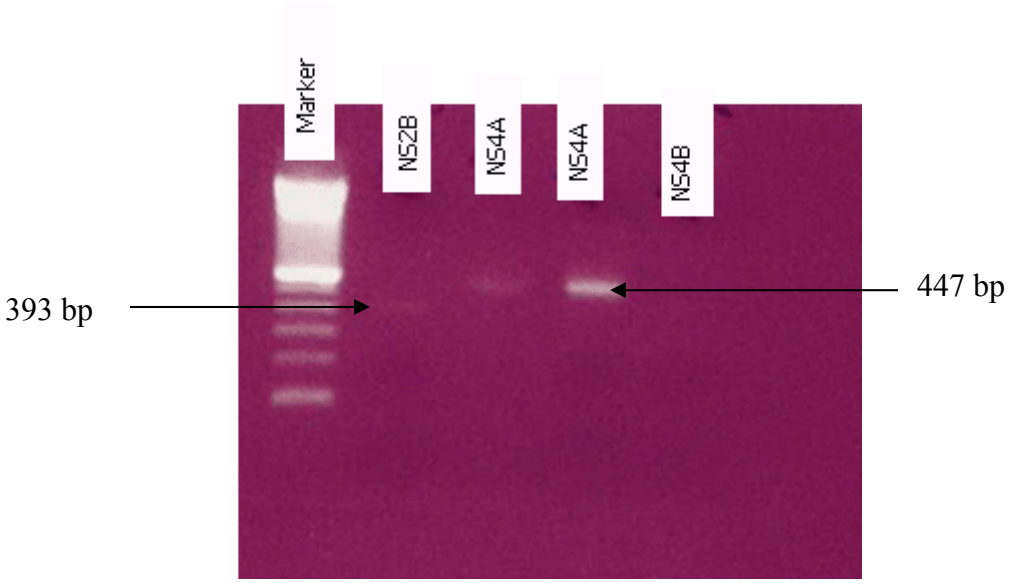

Figure 12: DNA Purification Confirmation Gel. The yield of DNA from the PCR experiment in the previous figure was assayed by agarose electrophoresis. NS2B and NS4A produced visible bands, but NS4B was negative.

BP and LR recombinant reactions were performed on the NS2B and NS4A samples (data not shown) obtaining successful positives for each, but the SDS-PAGE results were negative.

PCR of the WN replicon was redone for genes NS2A, NS2B, NS4A, NS4B, and NS5. The PCR results were all positive (data not shown), the bands were extracted, DNA was purified, and the BP and LR recombinant reactions were performed. All samples were transformed in TOP10 cells and plasmids were screened by PCR. Figure 13 shows the successful cloning of NS5 into expression plasmid pDEST 17, while Figure 14 shows the successful cloning of NS4A into both pDONOR 221, and expression plasmid pDEST 17.

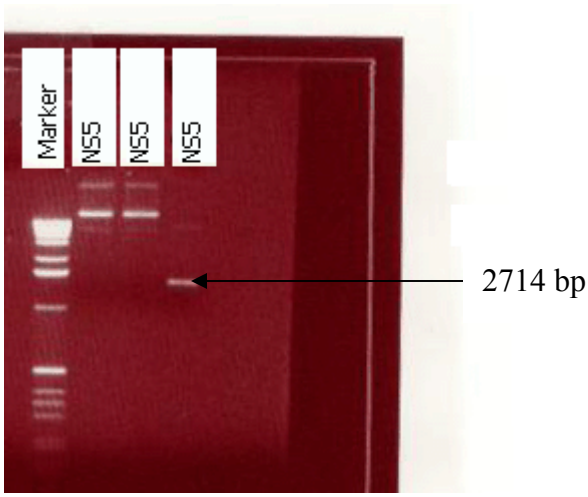

Figure 13: Cloning of WN NS5 into Expression Plasmid pDEST 17 by Using the LR Reaction. The NS5 LR reaction was transformed into TOP10 cells, selected with Ampicillin, and plasmid DNA purified. Plasmids were screened by PCR. Lane 4 was the only positive sample.

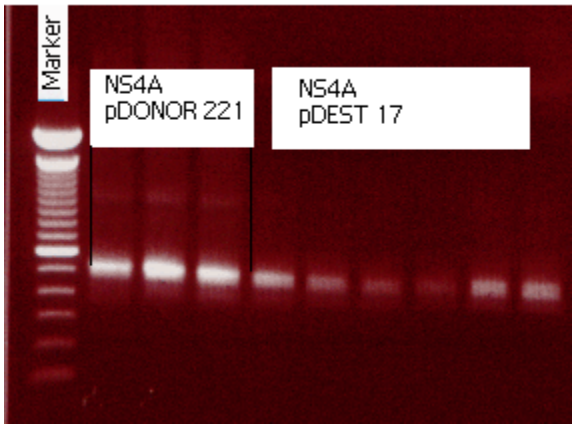

Figure 14: Plasmid Screening for WN NS4A. Lanes 2-4 were NS4A entry clones in plasmid pDONOR 221, from a BP recombinant reaction (TOP10 cells, selected with Kanamycin, and plasmids isolated by minipreps). All 3 lanes were positive. Lanes 5-10 represent the subcloning of NS4A into expression plasmid pDEST 17 using the LR recombinant reaction (TOP10 cells, selected with Ampicillin, and plasmid DNA purified by minipreps). All lanes contained positives of the correct size, 447 bp.

Unfortunately no expression test of the successfully cloned NS5 and NS4A pDEST 17 clones could be completed since Bosch's lab was not allowed to work with WN due to NIH contract issues.

DISCUSSION

The data shows that 6 DEN-2 genes, C, prM, NS2A, NS2B, NS4A, and NS5, were amplified and cloned successfully into an expression vector. Figure 9 and Table 3 clearly demonstrate the PCR screening of the cloned DEN-2 genes into expression plasmids pDEST 17. Two DEN-2 genes, NS3 and E, were successfully expressed in *E. coli* cells. Data also shows that WN genes, NS4A (Figure 14) and NS5 (Figure 13), were amplified and cloned successfully.

The project objective was to finish the cloning and expression of several remaining viral proteins of DEN-2 and WN virus in bacterial and mammalian cells. Unfortunately, the objective wasn't fully accomplished. 6 out of the 10 DEN-2 viral proteins (C, prM, NS2A, NS2B, NS4A, and NS5) were successfully cloned. Two other viral genes (E and NS3) that were previously cloned were successfully expressed in bacterial cells. 2 (NS4A and NS5) out of the 5 targeted WN viral proteins (NS2A, NS2B, NS4A, NS4B, and NS5) were successfully cloned. Unfortunately, expression in bacterial cells wasn't performed due to NIH contract issues. Due to lack of time and NIH contract issues, expression in mammalian cells was not feasible for any of the cloned viral proteins.

Bosch's lab continues to study DEN and WN to further understand the viruses. They will need to complete the cloning and expression of all remaining DEN-2 and WN viral proteins in bacterial and mammalian cells. They will also need to clone and express DEN-1, DEN-3, and DEN-4 viral proteins. All cloned and expressed genes will be used to test T-cell reactivity from infected patients, or to create antibodies to facilitate further

analysis of protein function. The next step of the research is to investigate the role of the viral proteins in inflammation by measuring the levels of cytokines secreted by cells transfected with the expression plasmids.

A more complete understanding of the viral proteins will bring a vast knowledge to the pathogenesis of DEN and WN. Hopefully all this newfound knowledge can be used to develop a vaccine for DEN and WN; and to develop a more rapid, sensitive, and specific diagnostic test.

BIBLIOGRAPHY

Bosch, Irene. "Irene Bosch, PhD: Research Interests." University of Massachusetts, Center for Infectious Disease and Vaccine Research. http://www.umassmed.edu/cidvr/faculty/bosch.cfm

Bosch, I., Xhaja, K., Estevez, L., Raines, G., Melichar, H., Warke, R., Fournier, M.V., Ennis, F.A., Rothman, A.L. "Increased production of interleukin-8 primary human monocytes and in human epithelial and endothelial cell lines after dengue virus challenge." *Journal of Virology*. (June 2002): 76-11

Campbell, G.L., Marfin, A.A., Lanciotti, R.S., Gubler, D.J. "West Nile virus." The Lancet. (September 2002): 2.

Castle, E., Leinder, U., Nowak, T., Wengler, G., and Wengler, G. "Primary structure of the West Nile flavivirus genome region coding for all nonstructural proteins." *Virology.* (February 1986) 149-1.

"CDC Dengue Fever Home Page." Center of Disease Control. February 2, 2005. http://www.cdc.gov/ncidod/dvbid/dengue/

"Center for Infectious Disease and Vaccine Research" University of Massachusetts, Center for Infectious Disease and Vaccine Research. http://www.umassmed.edu/cidvr/

Chang, G. Dengue and Dengue Hemorrhagic Fever: Molecular biology of dengue virus. Pg. 175-189. New York: D.J Gubler & G. Kuno,, 1997.

"Dengue and Dengue Hemorrhagic Fever: Question and Answers." Center of Disease Control. January 13, 2005. http://www.cdc.gov/ncidod/dvbid/dengue/dengue-qa.htm

"Dengue Fever Slide Sets." Center of Disease Control. November 10, 2003. http://www.cdc.gov/ncidod/dvbid/dengue/slideset/index.htm

"Fact Sheet: Dengue and Dengue Hemorrhagic Fever." Center of Disease Control. November 10, 2003. http://www.cdc.gov/ncidod/dvbid/dengue/facts.htm

"Flavivirus." Stanford University. http://www.stanford.edu/group/virus/1999/asb-flavi/overview.htm

Gaidos, Susan. "Scientists solve structure of dengue virus." Purdue News. March 7, 2002. http://news.uns.purdue.edu/UNS/html4ever/020307.Kuhn.dengue.html

Munoz-Jordan, J.L., Sanchez-Burgos, G.G., Rolle-Laurent, M., and Garcia-Sastre, A. "Inhibition of interferon signaling by dengue virus." *PNAS.* (November 25, 2003): 100-24

"Research Projects: Human Immune Responses in Emerging Viral Diseases." University of Massachusetts, Center for Infectious Disease and Vaccine Research. http://www.umassmed.edu/cidvr/research.cfm

Se-Thoe, S.Y., Ling, A.E., and Ng, M.M.L. "Alteration of virus entry mode: a neutralization mechanism of dengue-2 virus." *Journal of Medical Virology.* (2000): vol. 62.

Tassaneetrithep, B., Burgess, T.H., Granelli-Piperno, A., Trumpheller, C., Finke, L., Sun, W., Eller, M.A., Pattanapanyasay, K., Sarasombath, S., Birx, D.L., Steinman, R.M., Schlesinger, S., and Marovich, M.A. "DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells." *Journal of Experimental Medicine.* (2003): 197-7.

Warke, R.V., Xhaja, K., Martin, K.J., Fournier, M.F., Shaw, S.K., Brizuela, N., de Bosch, N., Lapointe, D., Ennis, F.A., Rothman, A.L., Bosch, I. "Dengue virus induces novel changes in gene expression of human umbilical vein endothelial cells." *Journal of Virology.* (November, 2003) 77-21

"West Nile Virus." West Nile Fever. 2000-2002. http://www.westnilefever.com/west_nile_virus_symptoms.htm

"West Nile Virus: Entomology." Center of Disease Control. March 22, 2004. http://www.cdc.gov/ncidod/dvbid/westnile/insects.htm

"West Nile Virus: Questions and Answers." Center of Disease Control. August 20, 2004. http://www.cdc.gov/ncidod/dvbid/westnile/qa/cases.htm

"West Nile Virus: Statistics, Surveillance, and Control." Center of Disease Control. January 11, 2005. http://www.cdc.gov/ncidod/dvbid/westnile/surv&controlCaseCount04_detailed.htm

"West Nile: What You Need To Know. CDC Fact Sheet." Center of Disease Control. August 10, 2004. http://www.cdc.gov/ncidod/dvbid/westnile/wnv_factsheet.htm

APPENDIX 1- West Nile Genome Sequence

West Nile virus NY-99 flamingo (AF 196835)

```
1 agtagttcgc ctgtgtgagc tgacaaactt agtagtgttt gtgaggatta acaacaatta
61 acacagtgcg agctgtttct tagcacgaag atctcgatgt ctaagaaacc aggagggccc
121 ggcaagagcc gggctgtcaa tatgctaaaa cgcggaatgc cccgcgtgtt gtccttgatt
181 ggactgaaga gggctatgtt gagcctgatc gacggcaagg ggccaatacg atttgtgttg
241 gctctcttgg cgttcttcag gttcacagca attgctccga cccgagcagt gctggatcga
301 tggagaggtg tgaacaaaca aacagcgatg aaacaccttc tgagttttaa gaaggaacta
361 gggaccttga ccagtgctat caatcggcgg agctcaaaac aaaagaaaag aggaggaaag
421 accggaattg cagtcatgat tggcctgatc gccagcgtag gagcagttac cctctctaac
481 ttccaaggga aggtgatgat gacggtaaat gctactgacg tcacagatgt catcacgatt
541 ccaacagctg ctggaaagaa cctatgcatt gtcagagcaa tggatgtggg atacatgtgc
601 gatgatacta tcacttatga atgcccagtg ctgtcggctg gtaatgatcc agaagacatc
661 gactgttggt gcacaaagtc agcagtctac gtcaggtatg gaagatgcac caagacacgc
721 cactcaagac gcagtcggag gtcactgaca gtgcagacac acggagaaag cactctagcg
781 aacaagaagg gggcttggat ggacagcacc aaggccacaa ggtatttggt aaaaacagaa
841 tcatggatct tgaggaaccc tggatatgcc ctggtggcag ccgtcattgg ttggatgctt
901 gggagcaaca ccatgcagag agttgtgttt gtcgtgctat tgcttttggt ggccccagct
961 tacagcttca actgccttgg aatgagcaac agagacttct tggaaggagt gtctggagca
1021 acatgggtgg atttggttct cgaaggcgac agctgcgtga ctatcatgtc taaggacaag
1081 cctaccatcg atgtgaagat gatgaatatg gaggcggcca acctggcaga ggtccgcagt
1141 tattgctatt tggctaccgt cagcgatctc tccaccaaag ctgcgtgccc gaccatggga
1201 gaagctcaca atgacaaacg tgctgaccca gcttttgtgt gcagacaagg agtggtggac
1261 aggggctggg gcaacggctg cggactattt ggcaaaggaa gcattgacac atgcgccaaa
1321 tttgcctgct ctaccaaggc aataggaaga accatcttga aagagaatat caagtacgaa
1381 gtggccattt ttgtccatgg accaactact gtggagtcgc acggaaacta ctccacacag
1441 gttggagcca ctcaggcagg gagattcagc atcactcctg cggcgccttc atacacacta
1501 aagcttggag aatatggaga ggtgacagtg gactgtgaac cacggtcagg gattgacacc
1561 aatgcatact acgtgatgac tgttggaaca aagacgttct tggtccatcg tgagtggttc
1621 atggacctca acctcccttg gagcagtgct ggaagtactg tgtggaggaa cagagagacg
1681 ttaatggagt ttgaggaacc acacgccacg aagcagtctg tgatagcatt gggctcacaa
1741 gagggagctc tgcatcaagc tttggctgga gccattcctg tggaattttc aagcaacact
1801 gtcaagttga cgtcgggtca tttgaagtgt agagtgaaga tggaaaaatt gcagttgaag
1861 ggaacaacct atggcgtctg ttcaaaggct ttcaagtttc ttgggactcc cgcagacaca
1921 ggtcacggca ctgtggtgtt ggaattgcag tacactggca cggatggacc ttgcaaagtt
1981 cctatctcgt cagtggcttc attgaacgac ctaacgccag tgggcagatt ggtcactgtc
2041 aacccttttg tttcagtggc cacggccaac gctaaggtcc tgattgaatt ggaaccaccc
2101 tttggagact catacatagt ggtgggcaga ggagaacaac agatcaatca ccattggcac
2161 aagtctggaa gcagcattgg caaagccttt acaaccaccc tcaaaggagc gcagagacta
2221 gccgctctag gagacacagc ttgggacttt ggatcagttg gaggggtgtt cacctcagtt
2281 gggaaggctg tccatcaagt gttcggagga gcattccgct cactgttcgg aggcatgtcc
2341 tggataacgc aaggattgct gggggctctc ctgttgtgga tgggcatcaa tgctcgtgat
2401 aggtccatag ctctcacgtt tctcgcagtt ggaggagttc tgctcttcct ctccgtgaac
2461 gtgcacgctg acactgggtg tgccatagac atcagccggc aagagctgag atgtggaagt
2521 ggagtgttca tacacaatga tgtggaggct tggatggacc ggtacaagta ttaccctgaa
2581 acgccacaag gcctagccaa gatcattcag aaagctcata aggaaggagt gtgcggtcta
2641 cgatcagttt ccagactgga gcatcaaatg tgggaagcag tgaaggacga gctgaacact
2701 cttttgaagg agaatggtgt ggaccttagt gtcgtggttg agaaacagga gggaatgtac
2761 aagtcagcac ctaaacgcct caccgccacc acggaaaaat tggaaattgg ctggaaggcc
2821 tggggaaaga gtattttatt tgcaccagaa ctcgccaaca acacctttgt ggttgatggt
2881 ccggagacca aggaatgtcc gactcagaat cgcgcttgga atagcttaga agtggaggat
```

2941 tttggatttg gtctcaccag cactcggatg ttcctgaagg tcagagagag caacacaact
3001 gaatgtgact cgaagatcat tggaacggct gtcaagaaca acttggcgat ccacagtgac
3061 ctgtcctatt ggattgaaag caggctcaat gatacgtgga agcttgaaag ggcagttctg
3121 ggtgaagtca aatcatgtac gtggcctgag acgcatacct tgtggggcga tggaatcctt
3181 gagagtgact tgataatacc agtcacactg gcgggaccac gaagcaatca caatcggaga
3241 cctgggtaca agacacaaaa ccagggccca tgggacgaag gccgggtaga gattgacttc
3301 gattactgcc caggaactac ggtcaccctg agtgagagct gcggacaccg tggacctgcc
3361 actcgcacca ccacagagag cggaaagttg ataacagatt ggtgctgcag gagctgcacc
3421 ttaccaccac tgcgctacca aactgacagc ggctgttggt atggtatgga gatcagacca
3481 cagagacatg atgaaaagac cctcgtgcag tcacaagtga atgcttataa tgctgatatg
3541 attgaccctt ttcagttggg ccttctggtc gtgttcttgg ccacccagga ggtccttcgc
3601 aagaggtgga cagccaagat cagcatgcca gctatactga ttgctctgct agtcctggtg
3661 tttgggggca ttacttacac tgatgtgtta cgctatgtca tcttggtggg ggcagctttc
3721 gcagaatcta attcgggagg agacgtggta cacttggcgc tcatggcgac cttcaagata
3781 caaccagtgt ttatggtggc atcgtttctc aaagcgagat ggaccaacca ggagaacatt
3841 ttgttgatgt tggcggctgt tttctttcaa atggcttatc acgatgcccg ccaaattctg
3901 ctctgggaga tccctgatgt gttgaattca ctggcggtag cttggatgat actgagagcc
3961 ataacattca caacgacatc aaacgtggtt gttccgctgc tagccctgct aacacccggg
4021 ctgagatgct tgaatctgga tgtgtacagg atactgctgt tgatggtcgg aataggcagc
4081 ttgatcaggg agaagaggag tgcagctgca aaaaagaaag gagcaagtct gctatgcttg
4141 gctctagcct caacaggact tttcaacccc atgatccttg ctgctggact gattgcatgt
4201 gatcccaacc gtaaacgcgg atggcccgca actgaagtga tgacagctgt cggcctaatg
4261 tttgccatcg tcggagggct ggcagagctt gacattgact ccatggccat tccaatgact
4321 atcgcggggc tcatgtttgc tgctttcgtg atttctggga aatcaacaga tatgtggatt
4381 gagagaacgg cggacatttc ctgggaaagt gatgcagaaa ttacaggctc gagcgaaaga
4441 gttgatgtgc ggcttgatga tgatggaaac ttccagctca tgaatgatcc aggagcacct
4501 tggaagatat ggatgctcag aatggtctgt ctcgcgatta gtgcgtacac cccctgggca
4561 atcttgccct cagtagttgg attttggata actctccaat acacaaagag aggaggcgtg
4621 ttgtgggaca ctccctcacc aaaggagtac aaaaaggggg acacgaccac cggcgtctac
4681 aggatcatga ctcgtgggct gctcggcagt tatcaagcag gagcgggcgt gatggttgaa
4741 ggtgttttcc acaccctttg gcatacaaca aaaggagccg ctttgatgag cggagagggc
4801 cgcctggacc catactgggg cagtgtcaag gaggatcgac tttgttacgg aggaccctgg
4861 aaattgcagc acaagtggaa cgggcaggat gaggtgcaga tgattgtggt ggaacctggc
4921 aagaacgtta agaacgtcca gacgaaacca ggggtgttca aaacacctga aggagaaatc
4981 ggggccgtga ctttggactt ccccactgga acatcaggct caccaatagt ggacaaaaac
5041 ggtgatgtga ttgggcttta tggcaatgga gtcataatgc ccaacggctc atacataagc
5101 gcgatagtgc agggtgaaag gatggatgag ccaatcccag ccggattcga acctgagatg
5161 ctgaggaaaa aacagatcac tgtactggat ctccatcccg gcgccggtaa aacaaggagg
5221 attctgccac agatcatcaa agaggccata aacagaagac tgagaacagc cgtgctagca
5281 ccaaccaggg ttgtggctgc tgagatggct gaagcactga gaggactgcc catccggtac
5341 cagacatccg cagtgcccag agaacataat ggaaatgaga ttgttgatgt catgtgtcat
5401 gctaccctca cccacaggct gatgtctcct cacagggtgc cgaactacaa cctgttcgtg
5461 atggatgagg ctcatttcac cgacccagct agcattgcag caagaggtta catttccaca
5521 aaggtcgagc taggggaggc ggcggcaata ttcatgacag ccaccccacc aggcacttca
5581 gatccattcc cagagtccaa ttcaccaatt tccgacttac agactgagat cccggatcga
5641 gcttggaact ctggatacga atggatcaca gaatacaccg ggaagacggt ttggtttgtg
5701 cctagtgtca agatggggaa tgagattgcc ctttgcctac aacgtgctgg aaagaaagta
5761 gtccaattga acagaaagtc gtacgagacg gagtacccaa aatgtaagaa cgatgattgg
5821 gactttgtta tcacaacaga catatctgaa atgggggcta acttcaaggc gagcagggtg
5881 attgacagcc ggaagagtgt gaaaccaacc atcataacag aaggagaagg gagagtgatc
5941 ctgggagaac catctgcagt gacagcagct agtgccgccc agagacgtgg acgtatcggt
6001 agaaatccgt cgcaagttgg tgatgagtac tgttatgggg ggcacacgaa tgaagacgac
6061 tcgaacttcg cccattggac tgaggcacga atcatgctgg acaacatcaa catgccaaac
6121 ggactgatcg ctcaattcta ccaaccagag cgtgagaagg tatataccat ggatggggaa
6181 taccggctca gaggagaaga gagaaaaaac tttctggaac tgttgaggac tgcagatctg
6241 ccagtttggc tggcttacaa ggttgcagcg gctggagtgt cataccacga ccggaggtgg
6301 tgctttgatg gtcctaggac aaacacaatt ttagaagaca acaacgaagt ggaagtcatc

6361 acgaagcttg gtgaaaggaa gattctgagg ccgcgctgga ttgacgccag ggtgtactcg
6421 gatcaccagg cactaaaggc gttcaaggac ttcgcctcgg gaaaacgttc tcagataggg
6481 ctcattgagg ttctgggaaa gatgcctgag cacttcatgg ggaagacatg ggaagcactt
6541 gacaccatgt acgttgtggc cactgcagag aaaggaggaa gagctcacag aatggccctg
6601 gaggaactgc cagatgctct tcagacaatt gccttgattg ccttattgag tgtgatgacc
6661 atgggagtat tcttcctcct catgcagcgg aagggcattg gaaagatagg tttgggaggc
6721 gctgtcttgg gagtcgcgac ctttttctgt tggatggctg aagttccagg aacgaagatc
6781 gccggaatgt tgctgctctc ccttctcttg atgattgtgc taattcctga gccagagaag
6841 caacgttcgc agacagacaa ccagctagcc gtgttcctga tttgtgtcat gacccttgtg
6901 agcgcagtgg cagccaacga gatgggttgg ctagataaga ccaagagtga cataagcagt
6961 ttgtttgggc aaagaattga ggtcaaggag aatttcagca tgggagagtt tcttttggac
7021 ttgaggccgg caacagcctg gtcactgtac gctgtgacaa cagcggtcct cactccactg
7081 ctaaagcatt tgatcacgtc agattacatc aacacctcat tgacctcaat aaacgttcag
7141 gcaagtgcac tattcacact cgcgcgaggc ttccccttcg tcgatgttgg agtgtcggct
7201 ctcctgctag cagccggatg ctggggacaa gtcaccctca ccgttacggt aacagcggca
7261 acactccttt tttgccacta tgcctacatg gttcccggtt ggcaagctga ggcaatgcgc
7321 tcagcccagc ggcggacagc ggccggaatc atgaagaacg ctgtagtgga tggcatcgtg
7381 gccacggacg tcccagaatt agagcgcacc acacccatca tgcagaagaa agttggacag
7441 atcatgctga tcttggtgtc tctagctgca gtagtagtga acccgtctgt gaagacagta
7501 cgagaagccg gaattttgat cacggccgca gcggtgacgc tttgggagaa tggagcaagc
7561 tctgtttgga acgcaacaac tgccatcgga ctctgccaca tcatgcgtgg gggttggttg
7621 tcatgtctat ccataacatg gacactcata aagaacatgg aaaaaccagg actaaaaaga
7681 ggtggggcaa aaggacgcac cttgggagag gtttggaaag aaagactcaa ccagatgaca
7741 aaagaagagt tcactaggta ccgcaaagag gccatcatcg aagtcgatcg ctcagcggca
7801 aaacacgcca ggaaagaagg caatgtcact ggagggcatc cagtctctag gggcacagca
7861 aaactgagat ggctggtcga acggaggttt ctcgaaccgg tcggaaaagt gattgacctt
7921 ggatgtggaa gaggcggttg gtgttactat atggcaaccc aaaaaagagt ccaagaagtc
7981 agagggtaca caaagggcgg tcccggacat gaagagcccc aactagtgca aagttatgga
8041 tggaacattg tcaccatgaa gagtggagtg gatgtgttct acagaccttc tgagtgttgt
8101 gacaccctcc tttgtgacat cggagagtcc tcgtcaagtg ctgaggttga agagcatagg
8161 acgattcggg tccttgaaat ggttgaggac tggctgcacc gagggccaag ggaattttgc
8221 gtgaaggtgc tctgccccta catgccgaaa gtcatagaga agatggagct gctccaacgc
8281 cggtatgggg ggggactggt cagaaaccca ctctcacgga attccacgca cgagatgtat
8341 tgggtgagtc gagcttcagg caatgtggta cattcagtga atatgaccag ccaggtgctc
8401 ctaggaagaa tggaaaaaag gacctggaag ggaccccaat acgaggaaga tgtaaacttg
8461 ggaagtggaa ccagggcggt gggaaaaccc ctgctcaact cagacaccag taaaatcaag
8521 aacaggattg aacgactcag gcgtgagtac agttcgacgt ggcaccacga tgagaaccac
8581 ccatatagaa cctggaacta tcacggcagt tatgatgtga agcccacagg ctccgccagt
8641 tcgctggtca atggagtggt caggctcctc tcaaaaccat gggacaccat cacgaatgtt
8701 accaccatgg ccatgactga cactactccc ttcgggcagc agcgagtgtt caaagagaag
8761 gtggacacga aagctcctga accgccagaa ggagtgaagt acgtgctcaa tgagaccacc
8821 aactggttgt gggcgttttt ggccagagaa aaacgtccca gaatgtgctc tcgagaggaa
8881 ttcataagaa aggtcaacag caatgcagct ttgggtgcca tgtttgaaga gcagaatcaa
8941 tggaggagcg ccagagaagc agttgaagat ccaaaatttt gggagatggt ggatgaggag
9001 cgcgaggcac atctgcgggg ggaatgtcac acttgcattt acaacatgat gggaaagaga
9061 gagaaaaaac ccggagagtt cggaaaggcc aagggaagca gagccatttg gttcatgtgg
9121 ctcggagctc gctttctgga gttcgaggct ctgggttttc tcaatgaaga ccactggctt
9181 ggaagaaaga actcaggagg aggtgtcgag ggcttgggcc tccaaaaact gggttacatc
9241 ctgcgtgaag ttggcacccg gcctgggggc aagatctatg ctgatgacac agctggctgg
9301 gacacccgca tcacgagagc tgacttggaa aatgaagcta aggtgcttga gctgcttgat
9361 ggggaacatc ggcgtcttgc cagggccatc attgagctca cctatcgtca caaagttgtg
9421 aaagtgatgc gcccggctgc tgatggaaga accgtcatgg atgttatctc cagagaagat
9481 cagaggggga gtggacaagt tgtcacctac gccctaaaca ctttcaccaa cctggccgtc
9541 cagctggtga ggatgatgga aggggaagga gtgattggcc cagatgatgt ggagaaactc
9601 acaaaaggga aaggacccaa agtcaggacc tggctgtttg agaatgggga agaaagactc
9661 agccgcatgg ctgtcagtgg agatgactgt gtggtaaagc ccctggacga tcgctttgcc
9721 acctcgctcc acttcctcaa tgctatgtca aaggttcgca aagacatcca agagtggaaa

9781 ccgtcaactg gatggtatga ttggcagcag gttccatttt gctcaaacca tttcactgaa
9841 ttgatcatga aagatggaag aacactggtg gttccatgcc gaggacagga tgaattggta
9901 ggcagagctc gcatatctcc aggggccgga tggaacgtcc gcgacactgc ttgtctggct
9961 aagtcttatg cccagatgtg gctgcttctg tacttccaca gaagagacct gcggctcatg
10021 gccaacgcca tttgctccgc tgtccctgtg aattgggtcc ctaccggaag aaccacgtgg
10081 tccatccatg caggaggaga gtggatgaca acagaggaca tgttggaggt ctggaaccgt
10141 gtttggatag aggagaatga atggatggaa gacaaaaccc cagtggagaa atggagtgac
10201 gtcccatatt caggaaaacg agaggacatc tggtgtggca gcctgattgg cacaagagcc
10261 cgagccacgt gggcagaaaa catccaggtg gctatcaacc aagtcagagc aatcatcgga
10321 gatgagaagt atgtggatta catgagttca ctaaagagat atgaagacac aactttggtt
10381 gaggacacag tactgtagat atttaatcaa ttgtaaatag acaatataag tatgcataaa
10441 agtgtagttt tatagtagta tttagtggtg ttagtgtaaa tagttaagaa aattttgagg
10501 agaaagtcag gccgggaagt tcccgccacc ggaagttgag tagacggtgc tgcctgcgac
10561 tcaaccccag gaggactggg tgaacaaagc cgcgaagtga tccatgtaag ccctcagaac
10621 cgtctcggaa ggaggacccc acatgttgta acttcaaagc ccaatgtcag accacgctac
10681 ggcgtgctac tctgcggaga gtgcagtctg cgatagtgcc ccaggaggac tgggttaaca
10741 aaggcaaacc aacgccccac gcggccctag ccccggtaat ggtgttaacc agggcgaaag
10801 gactagaggt tagaggagac cccgcggttt aaagtgcacg gcccagcctg actgaagctg
10861 taggtcaggg gaaggactag aggttagtgg agaccccgtg ccacaaaaca ccacaacaaa
10921 acagcatatt gacacctggg atagactagg agatcttctg ctctgcacaa ccagccacac
10981 ggcacagtgc gccgacaatg gtggctggtg gtgcgagaac acaggatct

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=11597239

APPENDIX 2- Dengue Virus Genome Sequence

Dengue virus type 2 New Guinea-C (M29095)

```
1 nnnnnnnagt ctacgtggac cgacaaagac agattctttg agggagctaa gctcaacgta
61 gttctaacag ttttttaatt agagagcaga tctctgatga ataaccaacg aaaaaaggcg
121 agaaatacgc ctttcaatat gctgaaacgc gagagaaacc gcgtgtcgac tgtacaacag
181 ctgacaaaga gattctcact tggaatgctg cagggacgag gaccattaaa actgttcatg
241 gccctggtgg cgttccttcg tttcctaaca atcccaccaa cagcagggat actgaagaga
301 tggggaacaa ttaaaaaatc aaaagccatt aatgttttga gagggttcag gaaagagatt
361 ggaaggatgc tgaacatctt gaacaggaga cgcagaactg caggcatgat cattatgctg
421 attccaacag tgatggcgtt ccatttaacc acacgtaacg gagaaccaca catgatcgtc
481 agtagacaag agaaagggaa aagtcttctg tttaaaacag aggatggtgt gaacatgtgt
541 accctcatgg ccatggacct tggtgaattg tgtgaagata caatcacgta caagtgtcct
601 tttctcaagc agaatgaacc agaagacata gattgttggt gcaactctac gtccacatgg
661 gtaacttatg ggacgtgtac caccacagga gaacacagaa gagaaaaaag atcagtggca
721 ctcgttccac atgtgggaat gggactggag acacgaactg aaacatggat gtcatcagaa
781 ggggcctgga aacatgccca gagaattgaa acttggatct tgagacatcc aggctttacc
841 ataatggcag caatcctggc atacaccata ggaacgacac atttccaaag agccctgatt
901 ttcatcttac tgacagctgt cgctccttca atgacaatgc gttgcatagg aatatcaaat
961 agagactttg tagaaggggt ttcaggagga agctgggttg acatagtctt agaacatgga
1021 agctgtgtga cgacgatggc aaaaaacaaa ccaacattgg attttgaact gatagaaaca
1081 gaagccaaac aacctgccac tctaaggaag tactgtatag aggcaaagct gaccaacaca
1141 acaacagatt ctcgctgccc aacacaagga gaacccagcc taaatgaaga gcaggacaaa
1201 aggttcgtct gcaaacactc catggtggac agaggatggg gaaatggatg tggactattt
1261 ggaaaaggag gcattgtgac ctgtgctatg ttcacatgca aaaagaacat gaaaggaaaa
1321 gtcgtgcaac cagaaaactt ggaatacacc attgtgataa cacctcactc aggggaagag
1381 catgcagtcg gaaatgacac aggaaaacat ggcaaggaaa tcaaaataac accacagagt
1441 tccatcacag aagcagagtt gacaggctat ggcactgtca cgatggagtg ctctccgaga
1501 acgggcctcg acttcaatga gatggtgttg ctgcaaatgg aaaataaagc ttggctggtg
1561 cacaggcaat ggttcctaga cctgccgttg ccatggctgc ccggagcgga cacacaagga
1621 tcaaattgga tacagaaaga gacattggtc actttcaaaa atccccatgc gaagaaacag
1681 gatgttgttg ttttgggatc ccaagaaggg gccatgcaca cagcactcac aggggccaca
1741 gaaatccaga tgtcatcagg aaacttactg ttcacaggac atctcaagtg caggctgagg
1801 atggacaaac tacagctcaa aggaatgtca tactctatgt gcacaggaaa gtttaaagtt
1861 gtgaaggaaa tagcagaaac acaacatgga acaatagtta tcagagtaca atatgaaggg
1921 gacggttctc catgtaagat ccctttgag ataatggatt tggaaaaaag acatgtttta
1981 ggtcgcctga ttacagtcaa cccaatcgta acagaaaaag atagcccagt caacatagaa
2041 gcagaacctc cattcggaga cagctacatc atcataggag tagagccggg acaattgaag
2101 ctcaactggt ttaagaaagg aagttctatc ggccaaatga ttgagacaac aatgagggga
2161 gcgaagagaa tggccatttt aggtgacaca gcttgggatt ttggatccct gggaggagtg
2221 tttacatcta taggaaaggc tctccaccaa gttttcggag caatctatgg ggctgccttc
2281 agtggggtct catggattat gaaaatcctc ataggagtca ttatcacatg gataggaatg
2341 aattcacgca gcacctcact gtctgtgtca ctagtattgg tgggagtcgt gacgctgtat
2401 ttgggagtta tggtgcaggc cgatagtggt tgcgttgtga gctggaaaaa caaagaactg
2461 aagtgtggca gtgggatttt catcacagac aacgtgcaca catggacaga acaatacaag
2521 ttccaaccag aatccccttc aaaactagct tcagctatcc agaaagctca tgaagagggc
2581 atttgtggaa tccgctcagt aacaagactg gaaaatctga tgtggaaaca aataacacca
2641 gaattgaatc acattctatc agaaaatgag gtgaagttga ctattatgac aggagacatc
2701 aaaggaatca tgcaggcagg aaaacgatct ctgcagcccc agcccactga gctgaagtat
2761 tcatggaaaa catggggcaa agcgaaaatg ctctctacag agtctcataa ccagaccttt
2821 ctcattgatg gccccgaaac agcagaatgc cccaacacaa acagagcttg gaattcgctg
2881 gaagttgaag actatggctt tggagtattc accaccaata tatggctaaa gttgagagaa
```

2941 aagcaggatg tattctgcga ctcaaaactc atgtcagcgg ccataaaaga caacagagcc
3001 gtccatgccg atatgggtta ttggatagaa agtgcactca atgacacatg gaagatagag
3061 aaagcctctt tcatcgaagt taaaagctgc cactggccaa agtcacacac cctctggagt
3121 aatggagtgt tagaaagtga gatgataatt ccaaagaatt tcgctggacc agtgtcacaa
3181 cacaactaca gaccaggcta ccatacacaa acagcaggac catggcatct aggtaagctt
3241 gagatggact ttgatttctg cgaaggaacc acagtggtgg tgactgagga ctgtggaaat
3301 agaggaccct ctttaagaac aactactgcc tctggaaaac tcataacaga atggtgctgc
3361 cgatcttgca cattaccacc gctaagatac agaggtgagg acggatgctg gtacgggatg
3421 gaaatcagac cattgaaaga gaaagaagag aatttggtca actccttggt cacagccgga
3481 catgggcaga ttgacaactt ttcactagga gtcttgggaa tggcattgtt cctggaagaa
3541 atgctcagga cccgagtagg aacgaaacat gcaatactac tagttgcagt ttcttttgtg
3601 acattgatca cagggaacat gtcctttaga gacctgggaa gagtgatggt tatggtgggc
3661 gctactatga cggatgacat aggtatgggc gtgacttatc ttgccctact agcagccttc
3721 aaagtcagac caacttttgc agctggacta ctcttgagaa agttgacctc caaggaattg
3781 atgatgacta ccataggaat cgtactcctc tcccagagca ccataccaga gaccattctt
3841 gaactgactg atgcgttagc cttgggcatg atggttctta aaatggtgag aaaaatggaa
3901 aagtatcaat tggcagtgac tatcatggct atcttgtgcg tcccaaatgc agtgatatta
3961 caaaacgcat ggaaagtgag ttgcacaata ttggcagtgg tgtccgtttc cccactgttc
4021 ttaacatcct cacagcagaa agcggattgg ataccattag cattgacgat caagggtctc
4081 aatccaacag ctatttttct aacaaccctt tcaagaacca acaagaaaag gagctggcca
4141 ctaaatgagg ctatcatggc agtcgggatg gtgagcattt tggccagttc actcctaaag
4201 aatgacattc ccatgacagg accattagtg gctggagggc tcctcactgt gtgctacgtg
4261 ctcactggac gatcggccga tttggaactg gagagagccg ccgatgtcaa atgggaagat
4321 caggcagaga tatcaggaag cagtccaatc ctgtcaataa caatatcaga agatggtagc
4381 atgtcgataa aaaacgaaga ggaagaacaa acactgacca tactcatcag aacaggattg
4441 ctggtgatct caggactttt tcctgtatca ataccaatca cggcagcagc atggtacctg
4501 tgggaagtga agaaacaacg ggctggagta ttgtgggatg tcccttcacc cccacccgtg
4561 ggaaaggctg aactggaaga tggagcctat agaatcaagc aaaaagggat tcttggatat
4621 tcccagatcg gagccggagt ttacaaagaa ggaacattcc atacaatgtg gcatgtcaca
4681 cgcggcgctg ttctaatgca taaaggaaag aggattgaac catcatgggc ggacgttaag
4741 aaagacctaa tatcatatgg aggaggctgg aagctagaag gagaatggaa ggaaggagaa
4801 gaagtccagg tcttggcatt ggagcctgga aaaaatccaa gagccgtcca aacaaaacct
4861 ggtcttttca aaaccaacgc cggaaccata ggtgccgtat ctctggactt ttctcctgga
4921 acctcaggat cgccaatcat cgacaaaaaa ggaaaagttg tgggtcttta tggtaatggt
4981 gttgttacaa ggagtggagc atatgtgagt gctatagccc agactgaaaa aagtattgaa
5041 gacaatccag agatcgaaga tgacattttt cgaaagagaa aattgaccat catggacctc
5101 cacccaggag cgggaaagac gaagagatac cttccggcca tagtcagaga ggctataaaa
5161 cggggcctga ggacattaat cctggccccc actagagtcg tggcagctga aatggaggaa
5221 gccctaagag gacttccaat aagataccaa accccagcca tcagagctga gcacaccggg
5281 cgggagattg tggacctaat gtgtcatgcc acattcacta tgaggctgct atcaccagtt
5341 agagtgccaa attacaacct gatcatcatg gacgaagccc atttcacaga cccagcaagt
5401 atagcggcta gaggatacat ctcaactcga gtagagatgg gtgaggcagc tgggattttc
5461 atgacagcca ctcctccggg aagcagagac ccattccctc agagcaatgc accaatcatg
5521 gatgaagaaa gagaaatccc tgaacgttcg tggagttctg gacatgagtg ggtcacggat
5581 tttaaaggga agactgtttg gttcgttcca agtataaaag caggaaatga tatagcagct
5641 tgcctgagaa aaaatggaaa gaaagtgata caactcagta ggaagacctt tgattctgag
5701 tatgtcaaga ctagaaccaa tgattgggac ttcgtggtca caactgacat ttcagaaatg
5761 ggtgccaact tcaaggctga gagggttata gaccccagac gctgcatgaa accagttata
5821 ctaacagatg gtgaagagcg ggtgatcctg gcaggaccta tgccagtgac ccactctagt
5881 gcagcacaaa gaagagggag aataggaaga aatccaaaaa atgaaaatga ccagtacata
5941 tacatggggg aacctctgga aaatgatgaa gactgtgcac actggaaaga agctaaaatg
6001 ctcctagata acatcaacac acctgaagga atcattccta gcatgttcga accagagcgt
6061 gaaaaggtgg atgccattga tggtgaatac cgcttgagag gagaagcaag gaaaaccttt
6121 gtggacctaa tgagaagagg agacctacca gtctggttgg cctacagagt ggcagccgaa
6181 ggcatcaact acgcagacag aaggtggtgt tttgatggaa ttaagaacaa ccaaatcttg
6241 gaagaaaatg tggaggtgga aatctggaca aaagaagggg aaaggaagaa attaaaaccc
6301 agatggttgg atgccaagat ctactctgac ccactggcgc taaaggaatt caaggagttt

6361 gcagctggaa gaaagtccct gaccctgaac ctaatcacag aaatgggtag gcttccaact
6421 ttcatgactc agaaggcaag agacgcactg gacaacttag cagtgctgca cacggctgaa
6481 gcaggaggaa gggcgtacaa tcatgctctc agtgaactgc cggagaccct ggagacattg
6541 cttttactga cacttctggc tacagtcaca ggaggaatct ttttattctt gatgagcgga
6601 aggggtatag ggaagatgac cctgggaatg tgctgcataa tcacggctag tattctccta
6661 tggtacgcac aaatacagcc acactggata gcagcttcaa taatactgga gttttttctc
6721 atagttttgc ttattccaga accagaaaag cagagaacac cccaagataa ccaattgacc
6781 tacgttgtca tagccatcct cacagtggtg gccgcaacca tggcaaacga gatgggtttc
6841 ctggaaaaaa cgaagaaaga tctcggattg ggaagcatta caacccagca acccgagagc
6901 aacatcctgg acatagatct acgtcccgca tcagcatgga cgctgtatgc tgtggccaca
6961 acatttgtca caccaatgtt gagacacagc attgaaaatt cctcagtgaa cgtgtcccta
7021 acagctattg ccaaccaagc cacagtgtta atgggtcttg ggaaaggatg gccattgtca
7081 aagatggaca tcggagttcc ccttctcgcc attggatgct actcacaagt caaccccata
7141 actctcacag cagctctttt cttactggta gcacattatg ccatcatagg gccaggactc
7201 caagcaaaag caaccagaga agctcagaaa agagcagcag cgggcatcat gaaaaaccca
7261 actgtcgatg gaataacagt gattgaccta gatccaatac cctatgatcc aaagtttgaa
7321 aagcagttgg gacaagtaat gctcctagtc ctctgcgtga ctcaagtgtt gatgatgagg
7381 actacatggg ctctgtgtga ggctttaacc ttagcgaccg ggcctatctc cacattgtgg
7441 gaaggaaatc cagggaggtt ttggaacact accattgcag tgtcaatggc taacattttt
7501 agagggagtt acttggccgg agctggactt ctcttttcca tcatgaagaa cacaaccaac
7561 acgagaaggg gaactggcaa cataggagag acgcttggag agaaatggaa aagccgattg
7621 aacgcattgg ggaaaagtga attccagatc tacaagaaaa gtggaatcca ggaagtggat
7681 agaaccttag caaaagaagg cattaaaaga ggagaaacgg accatcacgc tgtgtcgcga
7741 ggctcagcaa aactgagatg gttcgtcgag agaaatatgg tcacaccaga agggaaagta
7801 gtggacctcg gttgcggcag aggaggctgg tcatactatt gtgggggact aaagaatgta
7861 agagaagtca aaggcctaac aaaaggagga ccaggacatg aagaacccat ccccatgtca
7921 acatatgggt ggaatctagt acgtcttcaa agtggagttg acgttttctt cactccgcca
7981 gaaaagtgtg acacattgtt gtgtgacata ggggagtcgt caccaaatcc cacggtagaa
8041 gcaggacgaa cactcagagt ccttaactta gtggaaaatt ggttgaacaa caacacccaa
8101 ttttgcataa aggttctcaa cccatacatg ccctcagtca tagaaaaaat ggaagcacta
8161 caaaggaaat atggaggagc cttagtgagg aatccacttt cacgaaactc cacacatgag
8221 atgtactggt tatccaatgc ctccgggaac atagtgtcat cagtgaacat gatttcaagg
8281 atgttgatca acagattcac aatgagacac aagaaagcca cttacgagcc agatgtagac
8341 ctcggaagcg gaacccgcaa catcggaatt gaaagtgaga taccaaacct agacataatc
8401 gggaaaagaa tagaaaaaat aaaacaagag catgaaacat catggcacta tgaccaagac
8461 cacccataca aaacgtgggc ttaccatggc agctatgaaa caaaacaaac tggatcagca
8521 tcatccatgg gtaacggagt ggtcagactg ctgacaaaac cttgggacgt cgtccccatg
8581 gtaacacaga tggcaatgac agacacgact ccatttggac aacagcgcgt ttttaaagaa
8641 aaagtggaca cgagaaccca agaaccgaaa gaaggcacaa agaaactaat gaaaatcacg
8701 gcagagtggc tttggaaaga actagggaag aaaaagacac ctaggatgtg cactagagaa
8761 gaattcacaa gaaaggtgag aagcaatgca gccttggggg ccatattcac tgatgagaac
8821 aagtggaagt cggcacgtga ggctgttgaa gatagtaggt tttgggagct ggttgacaag
8881 gaaaggaatc tccatcttga aggaaagtgt gaaacatgtg tgtataacat gatgggaaaa
8941 agagagaaga agctagggga gttcggcaag gcaaaaggca gcagagccat atggtacatg
9001 tggcttggag cacgcttctt agagtttgaa gccctaggat tcttgaatga agatcactgg
9061 ttctccagag agaactcctt gagtggagtg gaaggagaag ggctgcacaa gctaggttac
9121 attttaagag acgtgagcaa gaaagaggga ggagcaatgt atgccgatga caccgcagga
9181 tgggacacaa gaatcacact agaagaccta aaaaatgaag aaatggtaac aaaccacatg
9241 gaaggagaac acaagaaact agccgaggcc attttcaaat taacgtacca aaacaaggtg
9301 gtgcgtgtgc aaagaccaac accaagaggc acagtaatgg atatcatatc gagaagagac
9361 caaagaggta gtggacaagt tggtacctat ggactcaata ctttcaccaa tatggaagcc
9421 caactaatca gacagatgga gggagaagga gtcttcaaaa gcattcagca cctgacagtc
9481 acagaagaaa tcgccgtgca aaactggtta gcaagagtag ggcgcgaaag gttatcaaga
9541 atggccatca gtggagatga ttgtgttgtg aaacctttag atgacaggtt cgcaagcgct
9601 ttaacagctc taaatgacat gggaaaggtt aggaaagaca tacaacaatg ggaaccttca
9661 agaggatgga acgattggac acaagtgccc ttctgttcac accatttcca tgagttaatc
9721 atgaaagacg gccgcgtact tgtagttcca tgcagaaacc aagatgaact gattggtaga

9781 gcccgaattt cccaaggagc tgggtggtct ttgcgagaga cggcctgttt ggggaagtcc
9841 tacgcccaaa tgtggagctt gatgtacttc cacagacgtg acctcaggct ggcggctaat
9901 gctatttgct cggcagtccc atcacattgg gttccaacaa gtagaacaac ctggtccata
9961 cacgccaaac atgaatggat gacaacggaa gacatgctga cagtctggaa cagggtgtgg
10021 attcaagaaa acccatggat ggaagacaaa actccagtgg aatcatggga ggaaatccca
10081 tacttgggga aaagagaaga ccaatggtgc ggctcattga ttgggctaac aagcagggcc
10141 acctgggcaa agaacatcca aacagcaata aatcaagtta gatcccttat aggcaatgag
10201 gaatacacag attacatgcc atccatgaaa agattcagaa aagaagagga agaggcagga
10261 gtcctgtggt agaaggcaaa actaacatga aacaaggcta gaagtcaggt cggattaagc
10321 catagtacgg aaaaaactat gctacctgtg agccccgtcc aaggacgtta aaagaagtca
10381 ggccattaca aatgccatag cttgagtaaa ctgtgcagcc tgtagctcca cctgagaagg
10441 tgtaaaaaat ctgggaggcc acaaaccatg gaagctgtac gcatggcgta gtggactagc
10501 ggttagagga gacccctccc ttacaaatcg cagcaacaat gggggcccaa ggtgagatga
10561 agctgtagtc tcactggaag gactagaggt tagaggagac ccccccaaaa caaaaaacag
10621 catattgacg ctgggaaaga ccagagatcc tgctgtctcc tcagcatcat tccaggcaca
10681 gaacgccaga aaatggaatg gtgctgttga atcaacaggt tct

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=323447

APPENDIX 3- QIAGEN QIAquick Gel Extraction Protocol

Protocol

QIAquick PCR Purification Kit Protocol

using a microcentrifuge

This protocol is designed to purify single- or double-stranded DNA fragments from PCR and other enzymatic reactions (see page 8). For cleanup of other enzymatic reactions, follow the protocol as described for PCR samples or use the new MinElute Reaction Cleanup Kit. Fragments ranging from 100 bp to 10 kb are purified from primers, nucleotides, polymerases, and salts using QIAquick spin columns in a microcentrifuge.

Notes:
- Add ethanol (96–100%) to Buffer PE before use (see bottle label for volume).
- All centrifuge steps are at ≥10,000 x *g* (~13,000 rpm) in a conventional tabletop microcentrifuge.

1. **Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix. It is not necessary to remove mineral oil or kerosene.**

   For example, add 500 µl of Buffer PB to 100 µl PCR sample (not including oil).
2. **Place a QIAquick spin column in a provided 2 ml collection tube.**
3. **To bind DNA, apply the sample to the QIAquick column and centrifuge for 30–60 s.**
4. **Discard flow-through. Place the QIAquick column back into the same tube.**

   Collection tubes are re-used to reduce plastic waste.
5. **To wash, add 0.75 ml Buffer PE to the QIAquick column and centrifuge for 30–60 s.**
6. **Discard flow-through and place the QIAquick column back in the same tube. Centrifuge the column for an additional 1 min at maximum speed.**

   IMPORTANT: Residual ethanol from Buffer PE will not be completely removed unless the flow-through is discarded before this additional centrifugation.
7. **Place QIAquick column in a clean 1.5 ml microcentrifuge tube.**
8. **To elute DNA, add 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) or H_2O to the center of the QIAquick membrane and centrifuge the column for 1 min. Alternatively, for increased DNA concentration, add 30 µl elution buffer to the center of the QIAquick membrane, let the column stand for 1 min, and then centrifuge.**

   IMPORTANT: Ensure that the elution buffer is dispensed directly onto the QIAquick membrane for complete elution of bound DNA. The average eluate volume is 48 µl from 50 µl elution buffer volume, and 28 µl from 30 µl elution buffer.

   Elution efficiency is dependent on pH. The maximum elution efficiency is achieved between pH 7.0 and 8.5. When using water, make sure that the pH value is within this range, and store DNA at –20°C as DNA may degrade in the absence of a buffering agent. The purified DNA can also be eluted in TE (10 mM Tris·Cl, 1 mM EDTA, pH 8.0), but the EDTA may inhibit subsequent enzymatic reactions.

APPENDIX 4- QIAGEN QIAprep Miniprep Protocol

QIAprep Spin

QIAprep Spin Miniprep Kit Protocol

using a microcentrifuge

This protocol is designed for purification of up to 20 µg of high-copy plasmid DNA from 1–5 ml overnight cultures of *E. coli* in LB (Luria-Bertani) medium. For purification of low-copy plasmids and cosmids, large plasmids (>10 kb), and DNA prepared using other methods, refer to the recommendations on page 35.

Please read Important Notes for QIAprep Procedures on pages 18–19 before starting.

Note: All protocol steps should be carried out at room temperature.

Procedure

1. **Resuspend pelleted bacterial cells in 250 µl Buffer P1 and transfer to a microcentrifuge tube.**

   Ensure that RNase A has been added to Buffer P1. No cell clumps should be visible after resuspension of the pellet.

2. **Add 250 µl Buffer P2 and gently invert the tube 4–6 times to mix.**

   Mix gently by inverting the tube. Do not vortex, as this will result in shearing of genomic DNA. If necessary, continue inverting the tube until the solution becomes viscous and slightly clear. Do not allow the lysis reaction to proceed for more than 5 min.

3. **Add 350 µl Buffer N3 and invert the tube immediately but gently 4–6 times.**

   To avoid localized precipitation, mix the solution gently but thoroughly, immediately after addition of Buffer N3. The solution should become cloudy.

4. **Centrifuge for 10 min at 13,000 rpm (~17,900 x *g*) in a table-top microcentrifuge.**

   A compact white pellet will form.

5. **Apply the supernatants from step 4 to the QIAprep Spin Column by decanting or pipetting.**

6. **Centrifuge for 30–60 s. Discard the flow-through.**

7. **(Optional): Wash the QIAprep Spin Column by adding 0.5 ml Buffer PB and centrifuging for 30–60 s. Discard the flow-through.**

   This step is necessary to remove trace nuclease activity when using *endA*+ strains such as the JM series, HB101 and its derivatives, or any wild-type strain, which have high levels of nuclease activity or high carbohydrate content. Host strains such as XL-1 Blue and DH5α™ do not require this additional wash step.

8. **Wash QIAprep Spin Column by adding 0.75 ml Buffer PE and centrifuging for 30–60 s.**

9. **Discard the flow-through, and centrifuge for an additional 1 min to remove residual wash buffer.**

   IMPORTANT: Residual wash buffer will not be completely removed unless the flow-through is discarded before this additional centrifugation. Residual ethanol from Buffer PE may inhibit subsequent enzymatic reactions.

10. **Place the QIAprep column in a clean 1.5 ml microcentrifuge tube. To elute DNA, add 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) or water to the center of each QIAprep Spin Column, let stand for 1 min, and centrifuge for 1 min.** (Add H2O instead of Buffer EB)