AN ABSTRACT OF THE THESIS OF

Younshim Park for the Master of Science

in Biological Sciences presented on May 15, 2014

Title: Detection of Antibiotic Resistance in the Environment Using Functional Metagenomics

Abstract approved:

Metagenomics has the potential to discover novel phenotypes that are not recognizable by DNA sequence analysis. In this study, functional metagenomic analysis of Niangua River soil detected spectinomycin (NR-YP1) and nalidixic acid (NR-YP3) resistance expressed in *E. coli* EPI300. Subclones of each were generated and screened on spectinomycin or nalidixic acid containing media for NR-YP1 and NR-YP3, respectively. No subclones were identified conferring resistance to either of the antibiotics. Random DNA sequencing and BLAST analysis of both NR-YP1 and NR-YP3 subclones matched an environmental clone, termed zdt-9n2, with almost 100% sequence identity. However, proteomic analysis of zdt-9n2 yielded no proteins with the potential to confer antibiotic resistance. Thus, the two clones, NR-YP1 and NR-YP3, appear to be nearly 100% identical in their cloned DNA but yet express two different antibiotic resistance phenotypes. Genetic loci responsible for the resistance could not be identified in this study.

Keywords: metagenomics, antibiotic resistance, spectinomycin, nalidixic acid

DETECTION OF ANTIBIOTIC RESISTANCE IN THE ENVIRONMENT USING FUNCTIONAL METAGENOMICS

A Thesis

Presented to

The Department of Biological Sciences

EMPORIA STATE UNIVERSITY

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Younshim Park

May 2014

_____

Approved by Major Advisor

_____

Approved by Committee Member

_____

Approved by Committee Member

_____

Approved for Department of Biological Sciences

_____

Approved for Dean of Graduate Studies and Research

ACKNOWLEDGMENTS

I would like to give my thanks to my advisor, Dr. Scott Crupper, for his assistance and guidance in both the laboratory and the writing of this thesis. I would also like to thank my committee members, Dr. Yixin Eric Yang and Dr. John Richard Schrock, for their suggestions and assistance in the writing of this thesis. Additionally, I would like to thank my lab mates, Cailin Wilson, Changjun Lin, Lan Rong, and Peng Shi for their assistance in the lab as well as their kindness and understanding throughout my graduate studies at ESU. Lastly, I would like to thank my family in Korea and my friends in Emporia for their encouragement, support, and love. Without them I would not have been able to accomplish what I have.

PREFACE

This thesis was prepared following the publication style of the American Society for Microbiology.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

Introduction

Bacteria of all genera are showing unprecedented levels of antibiotic resistance in clinical settings. Even though the use of antibiotics has decreased the consequences of infectious diseases in clinical settings, some human pathogens have developed resistance to multiple antibiotics (13). Moreover, commensal bacteria are also showing resistance to many of the existing antibiotics (17). To deal with a current and impending health crisis, many scientists have focused on the origin, selection, and dissemination of antibiotic resistance genes as a means of confronting this problem. Identifying potential sources of antibiotic resistance genes will aid the effort to track their movement from natural environments to clinical settings in addition to providing information about known and unknown mechanisms of resistance.

It is well-known that microorganisms in the environment, such as fungi and bacteria, naturally carry antibiotic resistance elements as a means of self protection against other antibiotic producers (7, 31). These organisms could serve as the original reservoir of antibiotic resistance elements for dissemination into human pathogens (7, 31, 40, 48). Some studies suggest the overuse of antibiotics in medicine and livestock stimulates the acquisition of resistance genes either via vertical evolution in terms of mutation of chromosomal genes, or via horizontal gene transfer (HGT) (28). Acquiring resistance genes carried by mobile genetic elements, including plasmids, bacteriophages, and transposons, may also play a role (24, 26, 42). Regardless, environmental reservoirs of antibiotic resistance genes (34) are extremely complex and warrant study (1, 12, 16, 27, 33, 36).

The environment is an important genetic pool of undiscovered bacteria and potentially undocumented antibiotic resistance genes (9, 37). However, approximately 98% of environmental microorganisms are undiscovered and unculturable in the laboratory (4, 6, 21, 47, 52). Thus, traditional microbial techniques, such as culturing, do not adequately reflect the total diversity in the environment. In addition, artificial conditions provided in the laboratory may change the primary role of many genes (35). Even though the polymerase chain reaction (PCR) has identified unculturable antibiotic resistance genes from environmental samples, data obtained is limited within known gene families (5, 43, 44, 51).

To circumvent these problems, a culture-independent technique called metagenomics has been developed in the last two decades (18, 19, 20). Metagenomics is a generic term used to describe the entire collective genomic DNA, called the metagenome, from natural environments (20). Unlike traditional microbiology and molecular techniques dependent on culturing, metagenomics is independent from culturing so it allows scientists to directly study unknown environmental microorganisms for the presence of novel antibiotic resistance genes.

In metagenomics, DNA is isolated from an environmental sample and cloned into a host organism to generate a metagenomic library that contains both cultured and uncultured bacterial DNA. In functional metagenomics, genes contained in the surrogate host are expressed to discover the function of the molecule it encodes. Since there is no DNA sequence bias in identifying an activity, functional metagenomics increases the chances of discovering novel activities of cloned genes (46). This technique has

successfully identified and characterized novel antibiotic resistance genes in several studies (2, 11, 32).

In this study, the goal was to discover novel antibiotic resistance genes using functional metagenomics. To accomplish this goal, a metagenomic library was constructed and screened against 11 different antibiotics to identify potential antibiotic resistance genes. Two clones were obtained that conferred resistance independently to either spectinomycin or nalidixic acid. Although these two antibiotics have a very different mechanism of action, the cloned DNA obtained appears to be the same in each organism.

Materials and Methods

Soil Sample Collection

Soil used in this study was collected from random locations along the Niangua River in Missouri by Dr. Scott Crupper (Emporia State University) in 2011. Soil samples were transported at room temperature to the laboratory and stored at 4˚C until needed.

Bacterial Strains, Vectors, and Culture Conditions

Escherichia coli EPI300 (Epicentre; Madison, WI) was used as the host strain for transformation of the metagenomic library. The fosmid, pCC2FOS, (Epicentre) was used for construction of the metagenomic library. Metagenomic subclones were prepared in pBluescript and maintained in *E. coli* DH10b. Both *E. coli* strains were routinely grown either in Luria-Bertani (LB) (Difco Laboratories; Sparks, MD) broth media at 37˚C with shaking at 250 rpm or on LB agar (LBA) plates containing 15 μg ml^{-1} chloramphenicol for fosmid clones and 100 μg ml^{-1} ampicillin for pBluescript subclones.

Soil DNA Isolation

Total DNA was extracted from soil samples using a ZR Soil Microbe DNA MicroPrep™ (Zymo Research; Irvine, CA) by following the manufacturer's instructions.

Metagenomic Library Construction

To construct the metagenomic library, DNA extracted from the soil sample (~40 kb) was inserted into pCC2FOS and transformed into *E. coli* EPI300 using the CopyControl Fosmid Library Production Kit (Epicentre) by following the manufacturer's

instructions. Transformed clones were grown overnight on LBA plates containing 15 µg ml^{-1} chloramphenicol to select for fosmid containing clones. Colonies were scraped from agar plates and stored in LB containing 20% glycerol at -80˚C.

Fosmid DNA Isolation

Fosmid clones were grown overnight at 37˚C in 250 ml of LB containing 15 µg ml^{-1} chloramphenicol and 500 µl autoinducer (Epicentre) with shaking at 250 rpm. Fosmid DNA was extracted using the Endofree Plasmid Maxi Kit (Qiagen; Germantown, MD) by following the manufacturer's instructions and stored at 4˚C.

Minimal Inhibitory Concentration

Antibiotics used for minimal inhibitory concentration (MIC) testing are listed in Table 1. Nineteen different antibiotics were used to determine the MIC for *E. coli* EPI300. *E. coli* EPI300 was grown in 3 ml of LB overnight and 30 µl of the overnight culture inoculated into fresh 3 ml of LB containing different concentrations of each antibiotic. Cultures were incubated for 48 hours at 37˚C with shaking and compared to uninoculated LB. In a separate experiment, 100 µl of *E. coli* EPI300 grown overnight in LB was plated on LBA containing each antibiotic. Plates were incubated at 37˚C for 48 hours to determine the MIC on agar plates. Eight antibiotics were eliminated from the original 19 due to the natural resistance of *E. coli* EPI300.

Screening the Metagenomic Library

The metagenomic library was screened on LBA plates containing the 11 antibiotics listed in Table 1. Briefly, each metagenomic library pool was thawed on ice and 0.5 ml inoculated into 50 ml LB containing 15 µg ml^{-1}chloramphenicol. After inoculation at 37˚C with shaking at 250 rpm for 4 hrs, 100 µl of the bacterial culture was plated on each antibiotic containing media. Three plates of each antibiotic were incubated at 28˚C and 3 plates incubated at 37 ˚C for 72 hours to select for resistance phenotypes.

Testing of Antibiotic Resistant Clones

Clones NR-YP1 (spectinomycin resistance) and NR-YP3 (nalidixic acid resistance) were tested in triplicate experiments (triplicate data points) for growth in various concentrations of spectinomycin or nalidixic acid, respectively. *E. coli* EPI300 without a fosmid was also included as a negative control. Overnight cultures were grown for 16 to 18 hrs at 37˚C with shaking at 250 rpm and the absorbance was measured at λ_{600}. Cultures containing 3 ml LB and a variable concentration of either spectinomycin or nalidixic acid were inoculated with each overnight bacterial culture to an initial $\lambda_{600} = 0.05$. Subsequently, cultures were incubated for 16 to 18 hrs at 37˚C with shaking at 250 rpm followed by λ_{600} measurement.

Restriction Enzyme Digestion, Ligation, and Transformation

Two plates containing subclones digested by the *Rsa*I restriction enzyme (Promega; Madison, WI) were generated for screening and sequencing. Briefly, 2 µl *Rsa*I digested NR-YP1 and NR-YP3 were mixed with 1 µl of 50 ng/µl pJET1.2 (Fermentas;

Pittsburgh, PA), 1 µl of 10X Ligase Buffer (Fisher Scientific; St. Louis, MO), 1 µl of T4 DNA Ligase (Fisher Scientific), and 5 µl of H_2O. Ligation mixtures were incubated at room temperature for 3 hours prior to incubation with 100 µl competent *E. coli* TG1 cells and put on ice for 15 min. Subsequently, incubation at 42 ˚C for 90 seconds followed by addition of 1 ml sterile LB broth media without ampicillin was performed. After incubation at 37˚C for 45 min, 100 µl was spread on a LBA/100 µg ml^{-1}ampicillin plates and incubated overnight at 37˚C.

Subcloning and Screening for Spectinomycin and Nalidixic Acid Resistance

NR-YP1 and NR-YP3 clones were randomly sheared by hydroshearing at the Clemson University Genomic Institute (Clemson, SC). Resulting 3-5 Kb pieces of DNA were end repaired, ligated into *Eco*RV (Promega) digested pBluescript, and transformed into *E. coli* DH10b using standard methologies (41). Transformation reactions were subsequently screened on LBA plates containing either 100 µg ml^{-1} spectinomycin, 30 µg ml^{-1} nalidixic acid or 100 µg ml^{-1}ampicillin.

Polymerase Chain Reaction

DNA was amplified using the GoTaq® Green Master Mix (Promega) according to the manufacturer's instructions with gene specific primers using a Bio-Rad T100 Thermal Cycler (Bio-Rad; Hercules, CA). Primers and amplification conditions are listed in Appendix 1.

Agarose Gel Electrophoresis

To confirm the size of purified and amplified DNA, agarose gel electrophoresis was performed according to standard conditions (41). 30 ml of 1X TAE buffer diluted from a 50X TAE stock, 0.3 g of agarose (Agarose UNLIMITED, Gainesville, FL), and 2 µl of 10 mg/ml ethidium bromide (EtBr) were mixed and heated in a microwave until the agarose completely dissolved. The mixture was poured and solidified in a gel tray containing a gel comb and 1X TAE used to fill the gel tray after the comb was removed. The gel was electrophoresed using a EC 105 power supply (Thermo Fisher Scientific; St. Louis, MO) at ~100 volts for 1 hour. Pictures were taken using a FluorChem E photodocumentation system (ProteinSimple; Santa Clara, CA).

Purification of DNA from Agarose Gels

Sized confirmed DNA was excised from electrophoresed agarose gels using a razor blade and purified using a Zymoclean™ Gel DNA Recovery Kit (Zymo Research) according to the manufacturer's instructions.

Measurement of DNA Concentration

The concentration of isolated and amplified DNA was measured using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific™) according to the manufacturer's instructions and stored at 4˚C.

DNA Sequencing

Plasmid DNA was isolated using a Zyppy^{TM}Plasmid Miniprep Kit (Zymo Research) by following the manufacturer's instructions. DNA sequencing was performed at the University of Arkansas for Medical Sciences, in Little Rock, Arkansas.

BLAST Analysis

The Basic Local Alignment Search Tool (BLAST) (3) analysis was performed to analyze metagenomic DNA sequence information for homology to sequence information contained in GenBank.

Table 1. Minimal Inhibitory Concentration of Antibiotics Used In This Study

Function	**Antibiotic**	**MIC for *E. coli* EPI300**	**Mechanism**
Cell wall synthesis inhibitor	Ampicillin*	50 μg ml^{-1}	Binds to specific penicillin-binding proteins (PBPs)
	Penicillin G*	110 μg ml^{-1}	Inhibits the formation of peptidoglycan cross-links
	Imipenem	> 14 μg ml^{-1}	Binds to penicillin-binding proteins (PBPs)
	Phosphomycin	> 300 μg ml^{-1}	Inhibits the enzyme UDP-*N*-acetylglucosamine-3-enolpyruvyltransferase
Protein synthesis inhibitor	Erythromycin	> 200 μg	Binds to the 50 S subunit of bacterial ribosomes or near the "P" or donor site
	Gentamicin*	30 μg ml^{-1}	Irreversibly binds to specific 30S subunit
	Kanamycin*	40 μg ml^{-1}	Irreversibly binds to specific 30S subunit
	Neomycin sulfate*	120 μg ml^{-1}	Binds to a specific receptor protein on the 30 S subunit of bacterial ribosomes
	Streptomycin	> 400 μg	Irreversibly binds to specific 30S subunit
	Spectinomycin*	80 μg ml^{-1}	Binds to the 30S ribosomal subunit
	Tetracycline*	10 μg ml^{-1}	Reversibly binds to the 30S ribosomal subunit
Nucleic acid synthesis inhibitor	Ciprofloxacin*	0.5 μg ml^{-1}	Inhibits the enzymes topoisomerase II (DNA gyrase) and topoisomerase IV
	Nalidixic acid*	15 μg ml^{-1}	Binds strongly, but reversibly, to DNA gyrase and topoisomerase IV
	Nitrofurantoin*	12 μg ml^{-1}	Damages ribosomal proteins or other macromolecules, especially DNA
	Rifampicin	> 130 μg ml^{-1}	Inhibits DNA-dependent RNA polymerase
Cell membrane function inhibitor	Bacitracin	> 120 μg ml^{-1}	Interferes with the dephosphorylation of the 55-carbon, biphosphate lipid transport molecule C55-isoprenyl pyrophosphate (undecaprenyl pyrophosphate) and binds to divalent transition metal ions (Mn(II), Co(II), Ni(II), Cu(II), and Zn(II))
	Polymyxin B*	5 μg ml^{-1}	Interacts with the lipopolysaccharide of the cytoplasmic outer membrane of Gram-negative bacteria
Anti-metabolites	Sulfonamide	> 400 μg ml^{-1}	Inhibits bacterial enzyme dihydropteroate synthetase
	Trimethoprim	> 250 μg ml^{-1}	Binds to dihydrofolate reductase and inhibits the reduction of dihydrofolic acid (DHF) to tetrahydrofolic acid (THF)

* indicates antibiotics used for metagenomic library screening on LBA plates

Results

Construction of Niangua River Soil Metagenomic Library

A soil metagenomic library was constructed in *E. coli* EPI300 containing approximately 40 Kb pieces of DNA. The library contained approximately 1970 colonies comprising an estimated 78.8 Megabases of metagenomic DNA extracted from a Niangua River soil sample. The library was divided into pools (Appendix 2) and stored at -80 ˚C until needed.

Minimal Inhibitory Concentration of *E. coli* EPI300 to Antibiotics

The minimal inhibitory concentration (MIC) of *E. coli* EPI300 was determined against the antibiotics listed in Table 1. Among 19 antibiotics examined, 8 were ruled out due to the natural resistance of *E. coli* EPI300. The remaining 11 antibiotics (Table 1) were used to screen the entire metagenomic library to select clones that express antibiotic resistance. Two clones were identified that showed resistance to spectinomycin and nalidixic acid. These clones were named NR-YP1 and NR-YP3, respectively.

Antibiotic Resistance of NR-YP1 and NR-YP3

Clones conferring resistance to spectinomycin (NR-YP1) and nalidixic acid (NR-YP3) were further examined to determine their antibiotic resistance levels. As shown in Figure 1, clone NR-YP1 demonstrated resistance to spectinomycin while NR-YP3 and *E. coli* EPI300 were susceptible. Clone NR-YP3 was resistant to nalidixic acid while NR-YP1 and *E. coli* EPI300 were susceptible (Figure 2). These results confirmed that NR-

YP1 and NR-YP3 contain DNA inserts conferring resistance to spectinomycin and nalidixic acid, respectively.

Restriction Enzyme Digestion and Subcloning of NR-YP1 and NR-YP3

NR-YP1 and NR-YP3 fosmid DNA were randomly digested with *Rsa*I for varying lengths of time, subcloned into pJET1.2, and transformed into competent *E. coli* TG1. Two plates containing subclones were generated for screening and sequencing. Transformation mixtures were also screened on LBA plates containing either 100 µg ml^{-1} spectinomycin or 30 µg ml^{-1} nalidixic acid. No subclones showed resistance to either of the antibiotics (data not shown).

Random shearing and subcloning of NR-YP1 and NR-YP3

NR-YP1 and NR-YP3 fosmid DNA were randomly sheared into 3-5 Kb fragments, ligated into pBluescript, and transformed into *E. coli* DH10b. None of the subclones (NR-YP1 and NR-YP3) showed resistance to spectinomycin or nalidixic acid (data not shown). Random NR-YP1 and NR-YP3 subclones were subjected to DNA sequencing to generate a partial genetic profile. BLAST analysis revealed genetic similarity nearing 100% among both NR-YP1 and NR-YP3. Furthermore, the DNA sequence obtained matched an environmental clone zdt-9n2 (GenBank: AC150248.3) with almost 100% sequence identity (Appendix 3 and 4).

BLASTp analysis of Enterobacteria phage DE3 and *E. coli* K-12 MG1655

Additional BLAST analysis confirmed that NR-YP1 and NR-YP3 subclones and environmental clone zdt-9n2 also match Enterobacteria phage DE3 (GenBank: EU078592.1) and *E. coli* str. K-12 substr. MG1655 (GenBank: U00096.3) (Figure 3). DNA sequence encoding Enterobacteria phage DE3 and *E. coli* str. K-12 substr. MG1655 was analyzed by BLASTp. Proteins encoded by these sequences are listed in Appendix 5.

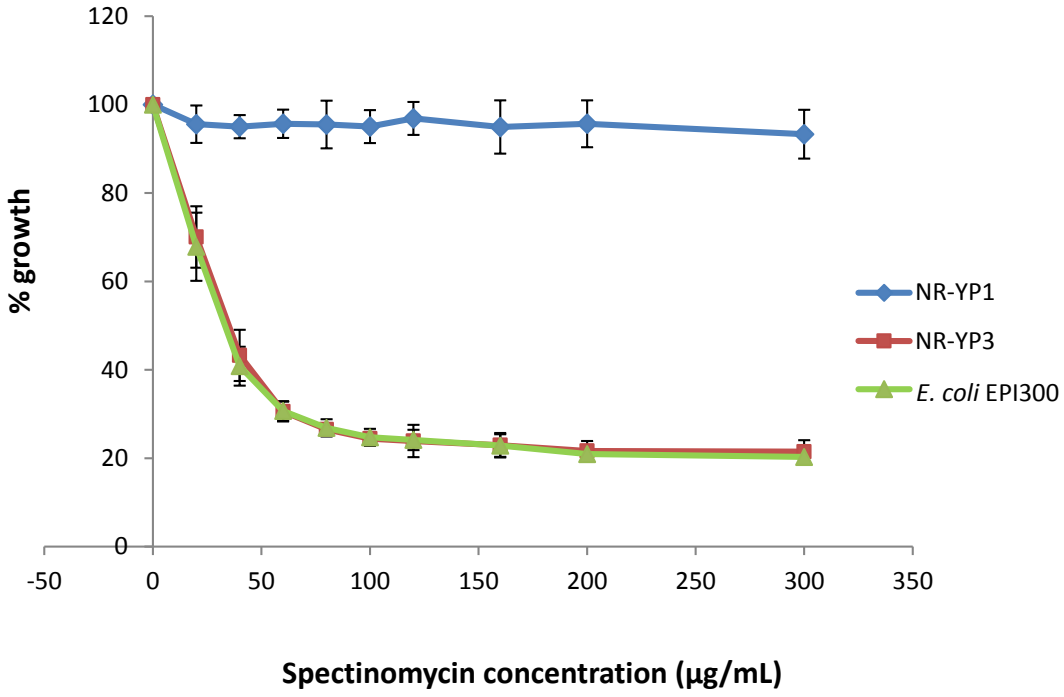
120
100
80
60
40
20
0
% growth
-50
0
50
100
150
200
250
300
350
Spectinomycin concentration (µg/mL)
NR-YP1
NR-YP3
E. coli EPI300

Figure 1. Spectinomycin resistance of NR-YP1. Growth percentage was calculated by comparing the growth of *E. coli* EPI300 (calculated at 100%) in media without spectinomycin.

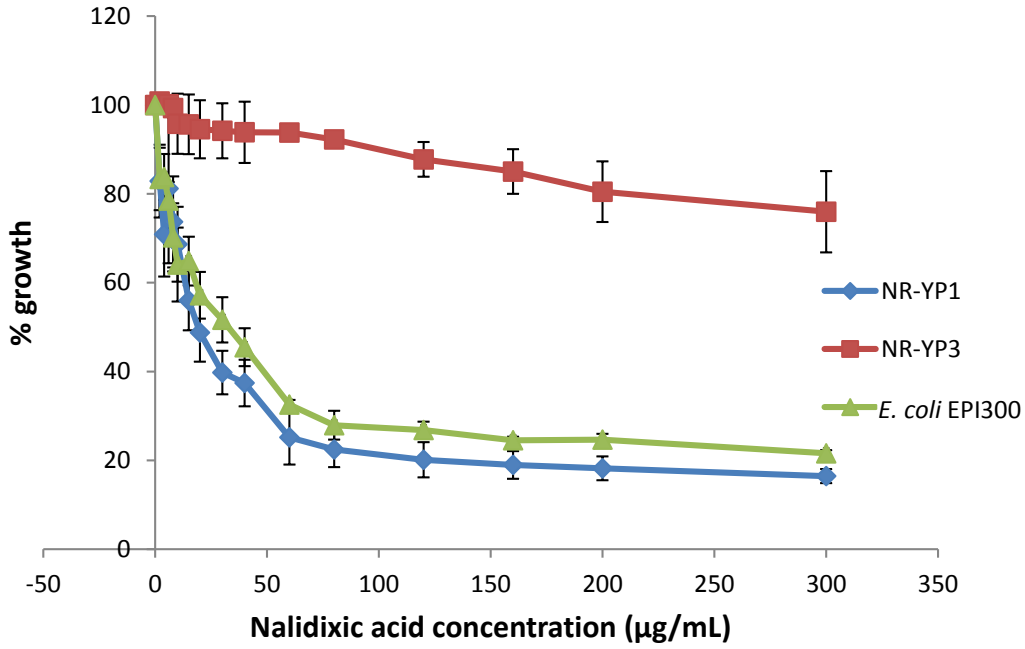
120
100
80
60
40
20
0
% growth
-50
0
50
100
150
200
250
300
350
Nalidixic acid concentration (µg/mL)
NR-YP1
NR-YP3
E. coli EPI300

Figure 2. Nalidixic acid resistance of NR-YP3. Growth percentage was calculated by comparing the growth of *E. coli* EPI300 (calculated at 100%) in media without nalidixic acid.

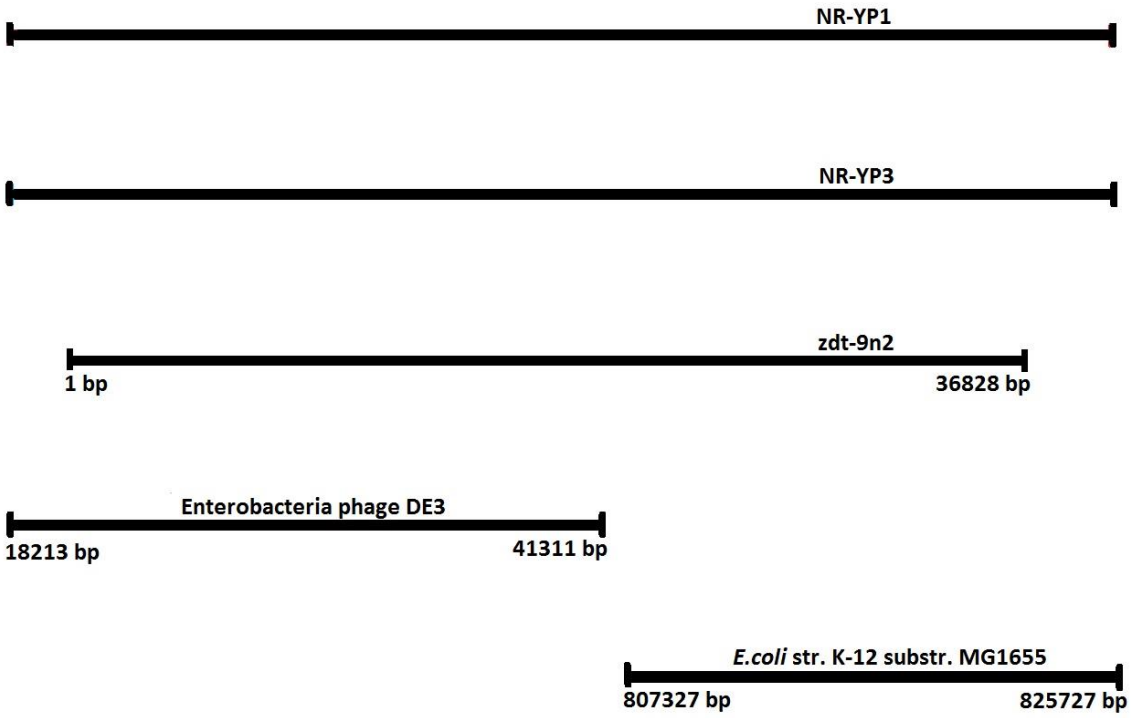
NR-YP1
NR-YP3
zdt-9n2
1 bp
36828 bp
Enterobacteria phage DE3
18213 bp
41311 bp
E.coli str. K-12 substr. MG1655
807327 bp
825727 bp

Figure 3. Illustration of DNA sequence relationship among NR-YP1, NR-YP3, zdt-9n2, Enterobacteria phage DE3, and *E. coli* str. K-12 substr. MG1655. This illustration was generated from DNA sequence data obtained from random subclones of NR-YP1 and NR-YP3. Also, DNA primers based on Enterobacteria phage DE3 and *E. coli* str. K-12 substr. MG1655 sequences were used to define the terminal DNA sequence of both NR-YP1 and NR-YP3.

Discussion

As infectious diseases have become the second-leading cause of millions of deaths throughout all human society (14), antibiotics have become important therapeutic agents to treat a range of diseases. However, the rapidly emerging problem of antibiotic resistance is greatly challenging and threatening public health globally (13). To confront this problem, one important strategy is to understand of how bacteria develop antibiotic resistance. As a step toward this goal, this study explored environmental microorganisms that serve as natural reservoirs of antibiotic resistance genes.

The realization that the majority of environmental microorganisms cannot be cultured in the laboratory forced scientists to develop a culture-independent technique called metagenomics. Metagenomics has successfully demonstrated that the uncultured microbial world can be directly investigated (23, 38). Briefly, metagenomic analysis consists of DNA isolation from environmental samples, cloning the DNA into a vector, transformation into a host strain, and screening transformants for various functions (18).

Functional metagenomic analysis aims to detect expression of specific traits such as antibiotic resistance or various other enzymatic activities leading to the discovery of novel phenotypes not recognizable by sequence analysis (10, 25, 29). However, identification of clones expressing a desired function is the primary challenge needed to be overcome due to the lack of suitable heterologous expression systems useful for all genes (8, 15). For example, Henne et al. screened 730,000 clones to identify lipolytic activity and detected only 1 clone with the desired phenotype (22).

In this study, functional metagenomic analysis of Niangua River soil detected spectinomycin (NR-YP1) and nalidixic acid (NR-YP3) resistance expressed in *E. coli*

EPI300. The approach used to detect these resistant clones consisted of cloning ~40 Kb pieces of environmental DNA into a fosmid followed by screening on LBA plates containing either 100 μg ml^{-1} spectinomycin or 30 μg ml^{-1} nalidixic acid. Fosmids are well suited for this type of study since large pieces of DNA (~40 Kb) can be packaged into a viral vector to infect host cells. The infection process is much more efficient than transformation. Once the fosmid is inside the host cell, it replicates like a plasmid. Numerous studies have employed fosmids for functional metagenomics, including those aiming to discover novel antibiotic resistance genes (11, 25, 50). By selecting clones based on antibiotic resistance (i.e. functional phenotype), sequence bias is eliminated. Both NR-YP1 and NR-YP3 were tested in multiple experiments (Figure 1 and 2) to conclusively demonstrate increased antibiotic resistance levels.

NR-YP1 demonstrated increased resistance to spectinomycin. This antibiotic binds to a specific ribosomal protein and inhibits protein synthesis. A common bacterial mechanism to resist spectinomycin is to inactivate it using a modification enzyme called aminoglycoside adenylyltransferase (30, 45). Another strategy is to eliminate spectinomycin binding sites on the target protein by mutational alteration (49). The second clone, NR-YP3, was resistant to nalidixic acid. This antibiotic targets bacterial DNA gyrase and topoisomerase IV to inhibit nucleic acid synthesis. Mutation of genes encoding these target proteins is common in nalidixic acid resistance (39).

Since spectinomycin and nalidixic acid each has a distinct mechanism of action, it was not surprising that each clone was resistant to only one of the antibiotics. Furthermore, it was reasonable to assume at this point that unique genetic loci existed for each of the clones. To test this assumption, subclones of NR-YP1 and NR-YP3 were

generated based on restriction enzyme digestion and hydroshearing. DNA fragments generated were cloned into a plasmid, transformed into *E. coli*, and screened for growth on spectinomycin or nalidixic acid containing media for NR-YP1 and NR-YP3, respectively. This approach is commonly done to define the genetic loci in large fosmid clones responsible for an activity (37, 50).

No clones were identified conferring resistance to either of the antibiotics. A possible interpretation includes the gene conferring the resistance was digested. Although probable with the restriction enzyme used (*Rsa*I is a 4 bp cutter), it is less likely with hydroshearing, which digests randomly. Since both restriction enzyme and hydroshearing resulted in 3-5 Kb fragments, this may be too small if the resistance gene is present on a larger or multiple genes. However, fragment size limitations of plasmid cloning make subcloning larger fragments difficult.

Since no subclones could be obtained by screening on antibiotic containing media, random clones were subjected to DNA sequencing followed by BLAST analysis. Random sequences of both NR-YP1 and NR-YP3 subclones matched an environmental clone, named zdt-9n2, nearly 100%. This was unexpected because each clone expressed a different antibiotic resistance phenotype. Since fosmid clones both matched zdt-9n2, it was hypothesized that the zdt-9n2 clone may contain an antibiotic resistance gene. However, detailed proteomic analysis yielded no proteins similar to known antibiotic resistance (Appendix 5).

Since proteomic analysis was inconclusive, DNA sequencing efforts were directed toward defining the DNA sequence of the ends. It was hypothesized that the ends could be different between the two fosmid clones and these differences could

account for the differences in antibiotic resistance profiles. However, PCR combined with DNA sequence analysis showed the ends are also nearly 100% similar between two clones (Figure 3). Based on the sequence analysis, it was concluded that the DNA sequences of the two fosmid clones, NR-YP1 and NR-YP3, are nearly identical and no known antibiotic resistance genes were present as determined by DNA sequence analysis.

In conclusion, this study has demonstrated the power of functional metagenomics for the discovery of novel antibiotic resistance genes. The two clones appear to be nearly 100% identical in their DNA sequence but yet express two different antibiotic resistance phenotypes. Due to limitations of time and money, no data were obtained to conclusively determine the genetic loci responsible for the resistance.

References

1. **Allen, H.K., K.A. Cloud-Hansen, J.M. Wolinski, C. Guan, S. Greene, S. Lu, M. Boeyink, N.A. Broderick, K.F. Raffa, and J. Handelsman.** 2009. Resident microbiota of the gypsy moth midgut harbors antibiotic resistance determinats. DNA Cell Biol. **28**:109-117.

2. **Allen, H.K., L.A. Moe, J. Rodbumrer, A. Gaarder, and J. Handelsman.** 2009. Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. ISME J. **3**:243-251.

3. **Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403-410.

4. **Amann, R.I., W. Ludwig, and K.H. Schleifer.** 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol. Rev. **59**:143-169.

5. **Aminov, R.I., N. Garrigues-Jeanjean, and R.I. Mackie.** 2001. Molecular ecology of tetracycline resistance: development and validation of primers for detection of tetracycline resistance genes encoding ribosomal protection proteins. Appl. Environ. Microbiol. **67**:22-32.

6. **Béjà, O., M.T. Suzuki, J.F. Heidelberg, W.C. Nelson, C.M. Preston, T. Hamada, J.A. Eisen, C.M. Fraser, and E.F. DeLong.** 2002. Unsuspected diversity among marine aerobic anoxygenic phototrophs. Nature **415**:630-633.

7. **Courvalin, P.** 2008. Predictable and unpredictable evolution of antibiotic resistance. J. Intern. Med. **264**:4-16.

8. **Craig, J.W., F. Chang, J.H. Kim, S.C. Obiajulu, and S.F. Brady.** 2010. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. Appl. Environ. Microbiol. **76**:1633-1641.

9. **D'Costa, V.M., K.M. McGrann, D.W. Hughes, and G.D. Wright.** 2006. Sampling the antibiotic resistome. Science **311**:374-377.

10. **Diaz-Torres, M. L., R. McNab, D.A. Spratt, A. Villedieu, N. Hunt, M. Wilson, and P. Mullany.** 2003. Novel tetracycline resistance determinant from the oral metagenome. Antimicrob. Agents Chemother. **47**:1430–1432.

11. **Donato, J.J., L.A. Moe, B.J. Converse, K.D. Smart, F.C. Berklein, P.S. McManus, and J. Hendelsman.** 2010. Metagenomic analysis of apple orchard soil reveals antibiotic resistance genes encoding predicted bifunctional proteins. Appl. Environ. Microbiol. **76**:4396-4401.

12. **Fajardo, A., J.F. Linares, and J.L Martinez.** 2009. Towards an ecological approach to antibiotics and antibiotic resistance genes. Clin. Microbiol. Infect. **15**:14-16.

13. **Fauci, A.S.** 2001. Infectious diseases: considerations for the 21st century. Clin. Infect. Dis. **32**:675-685.

14. **Fauci, A.S., N.A. Touchette, and G.K. Folkers.** 2005. Emerging infectious diseases: A 10-year perspective from the National Institute of Allergy and Infectious Diseases. Emerging Infect. Dis. **11**:519-525.

15. **Gabor, E.M., W.B. Alkema, and D.B. Janssen.** 2004. Quantifying the accessibility of the metagenome by random expression cloning techniques. Environ. Microbiol. **6**:879–886.

16. **Groh, J.L., Q. Luo, J.D. Ballard, and L.R. Krumholz.** 2007. Genes that enhance the ecological fitness of Shewanella oneidensis MR-1 in sediments reveal the value of antibiotic resistance. Appl. Environ. Microbiol. **73**:492-498.

17. **Hancock, R.E.W.** 2007. The end of an era? Nat. Rev. Drug Discovery **6**:28.

18. **Handelsman, J.** 2004. Metagenomics: application of genomics to uncultured microorganisms. Mol. Biol. Rev. **68:**669-685.

19. Handelsman, J. 2008. Metagenomics is not enough. DNA Cell Biol. **27**:219-221.

20. Handelsman, J., M.R. Rondon, S. Brady, J. Clardy, and R.M. Goodman. 1998. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. Chem. Biol. **5**:R245-249.

21. Head, I.M., J.R. Saunders, and R.W. Pickup. 1998. Microbial evolution, diversity, and ecology: A decade of ribosomal RNA analysis of uncultivated microorganisms. Microb. Ecol. **35**:1-21.

22. Henne, A., R.A. Schmitz, M. Bomeke, G. Gottschalk, and R. Daniel. 2000. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. Appl. Environ. Microbiol. **66**:3113–3116.

23. Henne, A., R. Daniel, R.A. Schmitz, and G. Gottschalk. 1999. Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Appl. Environ. Microbiol. **65**:3901–3907.

24. Jeters, R.T., G. Wang, K. Moon, N.B. Shoemaker, and A.A. Salyers. 2009. Tetracycline-associated transcriptional regulation of transfer genes of the *Bacteroids* conjugative transposon CTnDOT. J. Bacteriol. **191**:6374-6382.

25. **Lang, K.S., J.M. Anderson, S. Schwarz, L. Williamson, J. Handelsman, and R.S. Singer.** 2010. Novel florfenicol and chloramphenicol resistance gene discovered in Alaskan soil by using functional metagenomics. Appl. Environ. Microbiol. **76**:5321-5326.

26. **Maiques, E., C. Ubeda, M.A. Tormo, M.D. Ferrer, I. Lasa, R.P. Novick, and J.R. Penades.** 2007. Role of staphylococcal phage and SaPIintegrase in intra-and interspecies SaPI transfer. J. Bacteriol. **189**:5608-5616.

27. **Martinez, J.L.** 2009. The role of natural environments in the evolution of resistance traits in pathogenic bacteria. Proc. R. Soc. Lond. [Biol] **276**:2521-2530.

28. **Martinez, J.L., and F. Baquero.** 2000. Mutation frequencies and antibiotic resistance. Antimicrob. Agents Chemother. **44**:1771-1777.

29. **Martiny, A.C., J.B.H. Martiny, C. Weihe, A. Field, and J.C. Ellis.** 2011. Functional metagenomics reveals previously unrecognized diversity of antibiotic resistance genes in gulls. Front. Microbiol. **2**:1-11.

30. **Mazel, D., B. Dychinco, V.A. Webb, and J. Davies.** 2000. Antibiotic ressitance in the ECOR collection: integrons and indentification of a novel *aad* gene. Antimicrob. Agents Chemother. **44**:1568-1574.

31. Mazodier, P., and J. Davies. 1991. Gene transfer between distantly related bacteria. Annu. Rev. Genet. **25**:147-171.

32. Mori, T., S. Mizuta, H. Suenaga, and K. Miyazaki. 2008. Metagenomic screening for bleomycin resistance genes. Appl. Environ. Microbiol. **74**:6803-6805.

33. Nies, D.H. 2003. Efflux-mediated heavy metal resistance in prokaryotes. FEMS Microbiol. Rev. **27**:313-339.

34. Nwosu, V.C. 2001. Antibiotic resistance with particular reference to soil microorganisms. Res. Microbiol. **152**:421-430.

35. Piepersberg, W., J. Distler, P. Heinzel, and J. Perez-Gonzalez. 1988. Antibiotic resistance by modification: Many resistance genes could be derived from cellular control genes in antinomycetes. Actinomycetologica **2**:83-98.

36. Poole, K. 2005. Efflux-mediated antimicrobial resistance. J. Antimicrob. Chemother. **56**:20-51.

37. Riesenfeld, C.S., R.M. Goodman, and J. Handelsman. 2004. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. Environ. Microbiol. **6**:981-989.

38. **Rondon, M.R., P.R. August, A.D. Bettermann, S.F. Brady, T.H. Grossman, M.R. Liles, K.A. Loiacono, B.A. Lynch, I.A. MacNeil, C. Minor, C.L. Tiong, M. Gilman, M.S. Osburne, J. Clardy, J. Handelsman, and R.M. Goodman.** 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. **66**:2541–2547.

39. **Sáenz, Y., M. Zarazaga, L. Briñas, F. Ruiz-Larrea, and C. Torres.** 2003. Mutations in *gyrA* and *parC* genes in nalidixic acid-resistant *Escherichia coli* strains from food products, humans and animals. J. Antimicrob. Chemother. **51**:1001-1005.

40. **Salyers, A.A., A. Gupta, and Y. Wang.** 2004. Human intestinal bacteria as reservoirs for antibiotic resistance genes. Trends Microbiol. **12**:412-416.

41. **Sambrook, J., E.F. Fritsch, and T. Maniatis.** 1989. Molecular cloning, vol. 2. Cold Spring Harbor Laboratory Press New York.

42. **Schjorring, S., and K.A. Krogfelt.** 2011. Assessment of bacterial antibiotic resistance transfer in the gut. Int. J. Microbiol. doi:10.1155/2011/312956.

43. **Smalla, K., E. Krögerrecklenfort, H. Heuer, W. Dejonghe, E. Top, M. Osborn J. Niewint, C. Tebbe, M. Barr, M. Bailey, A. Greated, C. Thomas, S. Turner,**

P. Young, D. Nikolakopoulou, A. Karagouni, A. Wolters, J. D. Elsas, K. Dronen, R. Sandaa, S. Borin, J. Brabhu, E. Grohmann, and P. Sobecky. 2000. PCR-based detection of mobile genetic elements in total community DNA. Microbiology **146**:1256-1257.

44. **Stokes, H.W., A.J. Holmes, B.S. Nield, M.P. Holley, K.M. Nevalainen, B.C. Mabbutt, and M.R. Gillings.** 2001. Gene cassette PCR: Sequence-independent recovery of entire genes from environmental DNA. Appl. Environ. Microbiol. **67**:5240-5246.

45. **Sunde, M., and M. Norström.** 2005. The genetic background for streptomycin resistance in *Escherichia coli* influences the distribution of MICs. J. Antimicrob. Chemother. **56**:87-90.

46. **Torres-Cortés, G., V. Millán, H.C. Ramírez-Saad, R. Nisa-Martínez, N. Toro, and F. Martínez-Abarca.** 2011. Characterization of novel antibiotic resistance genes identified by functional metagenomics on soil samples. Environ. Microbiol. **13**:1101-1114.

47. **Torsvik, V., F.L. Daae, R.A. Sandaa, and L. Ovreas.** 1998. Novel techniques for analyzing microbial diversity in natural and perturbed environments. J. Biotech. **64**:53-62.

48. Tsubakishita, S., K. Kuwahara-Arai, S. Takashi, and K. Hiramatsu. 2010. Origin and molecular evolution of the determinant of methicillin resistance in Staphylococci. Antimicrob. Agents Chemother. **54**:4352-4359.

49. Unemo, M., D. Golparian, V. Skogen, A.O. Olsen, H. Moi, G. Syversen, and S.O. Hjelmevoll. 2013. *Neisseria gonorrhoeae* strain with high-level resistance to spectinomycin due to a novel resistance mechanism (mutated ribosomal protein S5) verified in Norway. Antimicrob. Agents Chemother. **57**:1057.

50. Uyaguari, M.I., E.B. Fichot, G.I. Scott, and R.S. Norman. 2011. Characterization and quantitation of a novel β-lactamase gene found in a wastewater treatment facility and the surrounding coastal ecosystem. Appl. Environ. Microbiol. **77**:8226-8233.

51. Waters, B., and J. Davies. 1997. Amino acid variation in the GyrA subunit of bacteria potentially associated with natural resistance to fluoroquinolone antibiotics. Antimicrob. Agents Chemother. **41**:2766-2769.

52. Whitman, W.B., D.C. Coleman, and W.J. Wiede. 1998. Prokaryotes: The unseen majority. Proc. Natl. Acad. Sci. U.S.A. **95**:6578-6583.

Appendix 1

Appendix 2

Niangua River soil metagenomic library

Pool #	# of clones	Mb
NR-1	44	1.76
NR-2	48	1.92
NR-3	45	1.8
NR-4	51	2.04
NR-5	54	2.16
NR 1-5 mix	242	9.68
NR-6	51	2.04
NR-7	50	2
NR-8	50	2
NR-9	50	2
NR-10	50	2
NR-11	50	2
NR-12	51	2.04
NR 6-12 mix	352	14.08
NR-13	102	4.08
NR-14	100	4
NR-15	101	4.04
NR-16	100	4
NR-17	101	4.04
NR 13-17 mix	504	20.16
NR-18	109	4.36
NR-19	103	4.12
NR-20	100	4
NR-21	100	4
NR-22	100	4
NR 18-22 mix	512	20.48
NR-23	115	4.6
NR-24	115	4.6
NR-25	130	5.2
NR 23-25 mix	360	14.4
total	**1970**	**78.8**

Appendix 3

NR-YP1 and NR-YP3 subclones matching an environmental clone zdt-9n2. DNA sequence comparison of NR-YP1 and NR-YP3 subclones to the environmental clone zdt-9n2 (GenBank: AC150248.3). The solid line represents the DNA sequence of zdt-9n2, while the shorter lines represent subclones of either NR-YP1 (aka YP1) or NR-YP3 (aka YP3).

Appendix 4

DNA sequence of subclones NR-YP1 (aka YP1) and NR-YP3 (aka YP3) matching an environmental clone zdt-9n2, Enterobacter phage DE3, and *E. coli* str. K-12 substr. MG1655

YP1 1 R

CCCCCTCGAGGTCGACGGTATCGATAAGCTTGATGTTGCTGTTCCATACTGAC
TCCAGCCAGAACTGTTCATCCTTAAACCACTTGTGTGGGCATGAGCACCCGC
GGCCCCTGTTGAACCGCTCAGACTGTGAGCATGAGCCCCCGTGTTATTCGTCG
ATTTGGTGCCGTAATCGAAACTGCCTGTTGTTT

YP1 4 F

GCAGGAATTCGATCTTTACCCGTGGCAATGCCCGCGCAGACGATCTGGTACG
CAATAACGGCTATGCCGCCAACGCCATCCAGCTGCATCAGGATCATATCGTC
GGGTCTTTTTTCCGGCTCAGTCATCGCCCAAGCTGGCGCTATCTGGGCATCGG
GGAGGAAGAAGCCCGTGCCTTTTCCCGCGAGGTTGAAGCGGCATGGAAAGA
GTTTGCCGAGGATGACTGCTGCTGCATTGACGTTGAGCGAAAACGCACGTTT
ACCATGATGATTCGGGAAGGTGTGGCCATGCACGCCTTTAACGGTGAACTGT
TCGTTCAGGCCACCTGGGATACCAGTTCGTCGCGGCTTTTCCGGACACAGTTC
CGGATGGTCAGCCCGAAGCGCATCAGCAACCCGAACAATACCGGCGACAGC
CGGAACTGCCGTGCCGGTGTGCAGATTAATGACAGCGGTGCGGCGCTGGGAT
ATTACGTCAGCGAGGACGGGTATCCTGGCTGGATGCCGCAGAAATGGACATG
GATACCCCGTGAGTTACCCGGCGGGCGCGCCTCGTTCATTCACGTTTTTGAAC
CCGTGGAGGACGGGCAGACTCGCGGTGCAAATGTGTTTTACAGCGTGATGGA
GCAGATGAAGATGCTCGACACGCTGCAGAACACGCAGCTGCAGAGCGCCATT
GTGAAGGCGATGTATGCCGCCACCATTGAGAGTGAGCTGGATACGCAGTCAG
CGATGGATTTTATTCTGGGCGCGAACAGTCAGGAGCAGCGGGAAAGGCTGAC
CGGCTGGATTGGTGAAATTGCCGCGTATTACGCCGCAGCGCCGGTCCGGCTG
GGAGGCGCAAAAGTACCGCACCTGATGCCGGGTGACTCACTGAACCT

YP1 4 R

CCTCGAGGTCGACGGTATCGATAAGCTTGATAGACCGCGTGCCTAAGTGTTC
CCCAGCACCATCGTGTTGTCCGGCAGGAAGTTCTTTTTGACGCCGTTTTCCAC
GTACTGTCCGGAATACACGACGATGGCCACATCGCCATACATCCCCTTATAG
GACACCGCTTTGCCCAGGTCTTCCACCGCTGTCTCCAGCTCGGAATTAGAGCC
ACGACGGGTATCCAGCTTCTCCTTGACGGCTTTGAAGGAACGGAACAGCGCC
CAGCCTTTCGGATCGAACACGATGATATTCACCACACCGCTGGCGTTCAGCG

CGTAGGCTTCGATATCGTCGGTCGGGTCATACGTGGACTTGTCACGCTTGCTC
CACTCCGTGCCGCCGGACTGCGTGATGTTATTCTCCTCACTGCGGCCCATATC
CACCTCAACCGGATCGAAGGCTTCACCGGTCATGGTGTATTTGCCCTTAAGCA
CGGCAGAAACTGCCTGCATCTCTTCGACCTGAGCAATGGCCAGCTCTTCGTCA
CGCATGTTCTGCATGATGATGCGACGGCGGCGGTAAGCCGGGTCCGCCAGAT
TCTGCGGATCTTCATCCGGCAGGCGACGCAGGGTCATCTGCGGATTCACTTCA
TGCTTCGGCTTGACA

YP1 5 F

CAGGAATTCGATAGGACTCCTTCGGCGGGATGATCCCCATGTTCAGGGGGCT
TGCCGGTGCGATCACCCTGCCGATGGTGGGGGCCACCTCGCTGGCGGTGGCG
ACCGGTGCGCTGGCGTATGCCTGGTATCAGGGCAACTCAACCCTGTCCGATTT
CAACAAAACGCTGGTCCTTTCCGGCAATCAGGCGGGACTGACGGCAGATCGT
ATGCTGGTCCTGTCCAGAGCCGGGCAGGCGGCAGGGCTGACGTTTAACCAGA
CCAGCGAGTCACTCAGCGCACTGGTTAAGGCGGGGGTAAGCGGTGAGGCTCA
GATTGCGTCCATCAGCCAGAGTGTGGCGCGTTTCTCCTCTGCATCCGGCGTGG
AGGTGGACAAGGTCGCTGAAGCCTTCGGGAAGCTGACCACAGACCCGACGTC
GGGGCTGACGGCGATGGCTCGCCAGTTCCATAACGTGTCGGCGGAGCAGATT
GCGTATGTTGCTCAGTTGCAGCGTTCCGGCGATGAAGCCGGGGCATTGCAGG
CGGCGAACGAGGCCGCAACGAAAGGGTTTGATGACCAGACCCGCCGCCTGA
AAGAGAACATGGGCACGCTGGAGACCTGGGCAGACAGGACTGCGCGGGCAT
TCAAATCCATGTGGGATGCGGTGCTGGATATTGGTCGTCCTGATACCGCGCA
GGAGATGCTGATTAAGGCAGAGGCTGCGTATAAGAAAGCAGACGACATCTG
GAATCTGCGCAAGGATGATTATTTTGTTAACGATGAAGCGCGGGCGCGTTAC
TGGGATGATCGTGAAAAGGCCCGTCTTGCGCTTGAAGCCGCCCGAAAGAAGG
CTGAGCAGCAGACTCAACAGGACAAAAATGCGCAGCAGCAGAGCGATAC

YP1 5 R

CGACGGTATCGATAAGCTTGATCCCCGGCGACTCTGGGAACAATATGAATTA
CAGCGCCATCAGGCAGAGTCTCATGTAACTGCGCCGTTAACCCGGACGTGCT
GACGTCCCGCCCGGCAATCCGTACCTGATACCAGCCGTCGCTCAGTTTCTGAC
GAAACGCCGGGAGCTGTGTGGCCAGTGCCCGGATGGCTTCAGCCCCCGTTTT
CACACGAAGGTCGATGCGGCGACCAAATCGTTGTAAATCCCCGTAAAGGCAG
ATGCGCGCCATGCCCGGTGACGCCAGAGGGAGTGTGTGCGTCGCTGCCATTT
GTCGGTGTACCTCTCTCGTTTGCTCAGTTGTTCAGGAATATGGTGCAGCAGCT
CGCCGTCGCCGCAGTAAATTGCGGCGTGATTCGGCACTGATGAACCAAAACA
GCACAGCAGCACATCGCCCGGCTGTGCCGCTGACAACGGCACCTGATACAGC
CCCGTCGCCTCCAGATTATCCAGATAGAGATTCTGGCCGTTACGCCACCAGTC
ATCCTCACGATGAAAGTCCGGCATCTCAATCCCCGCCAGATGATAAGCATCC
CGGAACAGTGTGTAACAGTCCGTCACACCGTGCTCAAAGCGCCGCCCGGTGA
GATGCGGCACACAGCGGAACTTATGAATCGTCCCCCGGCAGACCAGCCACCA

CGGCAAATCACTCTGCACCTGCAGCCGCCGGTCGGCCTCACTCAGCCAGGGC
AGACCACCGGGGTGGCTGTGGACCAGCGCCACAATCTCACCCTGCA

YP1 6 R

GACGGTATCGATAAGCTTGATCAGGTGCCTCTCACTTTTGGCGTGGAGCCACG
GATGGACTTTGGCCAGACCGAGCCTTCAATACGCTTGTCACCCAGGAACGTC
GGAGAGCCTTCCTGTTCAATATCATCATCAAAAGCAGCAAGTTCATCATAAC
CCGCCACATCCACCGACTTTTCACGGTAGTTTTTTGCCGCTTTACCGCCCAGG
CACCAGAAGCCACGCCCATTAGTGAAACGCTTCATGGTGAGCGTGTTATCCC
GGTGCTTTTTGCCATACCACGGGGCCAGCGCCAGCAGCGACGGAATATCACG
AATAGTCGGCTCAACGTGGGTTTTCATAAAGTTCTCGGCATCACCATCCGTCG
GCAACCAGATAAGGGTGTTGCGCTGCTTATGCTCTATAAAGTAGGCATAAAC
ACCCAGCAGCATTTTGGAATAACCGACACGGGCAGACTTCACCACATTCACC
TCACGGATGTAGTCGCTGCCCATCGCATTCATGATGGCCCGCTGAAAGGGCA
GTGTTTCCCAGCGCCCTTCCTGGTATGCGGATTCTTTCGGGAGATAGTAATTA
GCATCCGCCCATTCAACGGCGGTCTGTGGCTCCGGCCTGAACAGTGAGCGAA
GCCCGGCGCGGACAAAATGCCGCAGCCTGTTAACCTGACTGTTCGATATATT
CACTCAGCAACC

YP1 7 F

TTGCAGGCCAGCTTGGGATCAGCAGCCTGACGGATGCGGTGTCCGGCGACAG
CCTGACTGCCCAGGAGGCACTCGCGACGCTGGCATTATCCGGTGATGATGAC
GGACCACGACAGGCCCGCAGTTATCAGGTCATGAACGGCATCGCCGTGCTGC
CGGTGTCCGGCACGCTGGTCAGCCGGACGCGGGCGCTGCAGCCGTACTCGGG
GATGACCGGTTACAACGGCATTATCGCCCGTCTGCAACAGGCTGCCAGCGAT
CCGATGGTGGACGGCATTCTGCTCGATATGGACACGCCCGGCGGGATGGTGG
CGGGGGCATTTGACTGCGCTGACATCATCGCCCGTGTGCGTGACATAAAACC
GGTATGGGCGCTTGCCAACGACATGAACTGCAGTGCAGGTCAGTTGCTTGCC
AGTGCCGCCTCCCGGCGTCTGGTCACGCAGACCGCCCGGACAGGCTCCATCG
GCGTCATGATGGCTCACAGTAATTACGGTGCTGCGCTGGAGAAACAGGGTGT
GGAAATCACGCTGATTTACAGCGGCAGCCATAAGGTGGATGGCAACCCCTAC
AGCCATCTTCCGGATGACGTCCGGGAGACACTGCAGTCCCGGATGGACGCAA
CCCGCCAGATGTTTGCGCAGAAGGTGTCGGCATATACCGGCCTGTCCGTGCA
GGTTGTGCTGGATACCGAGGCTGCAGTGTACAGCGGTCAGGAGGCCATTGAT
GCCGGACTGGCTGATGAACTTGTTAACAGCACCGATGCGATCACCCGTCATG
CGTGATGCACTGGATGCACG

YP1 7 R

ACGGTATCGATAAGCTTGATGGCACCTTTCACCGGCATTGTAGGATTTGGTAC
AGGCATAGCGTCCTCACATTTCATAGGTAATGACATAAGTCAGATCGGCTGA

ACTCCACAAGCCCGCATCATCGTCGCGCCGGTAGTCATAGCCGCTGGCCACC
ATACTGGTGATCAAATCTGACAGTGCCGGGATATCGCTCATCACCGGATAAA
TCCGGGACTCCATCCACGCATCCAGCTCTGAATCCGGCACCTGAGCAGGCAG
GAAAACTTCGATATGCAGCTCCGCCTGCCAGGTATCGCTGTCCAGCTCTTCGC
CCGTGTATTCAGCGCCGGTGAGATAAACGGCAACTGCCGGAAAATCCGCCTC
ATCAAAAACAGCGGGGCGACCATCAAAAAACGTCGCCCCGGTGTCATGCTTC
TCCAGTGCATCCAGTACGGCTGCACGGAGTTCAGTATGTTTCATCGCTTTATT
ACCATCCTCAGTTGATGCTGCAGCGCATAGCCCAGCTCTTTCGGAAGACGTTC
ACGCCGTATCCGCTCAATATTTTGTTTAAACGCCGTGGTCAGCGGCACCGC

YP1 9 F

AAAACGCGGTGACGACTATCAGGAAATTTTTGCCCAGCAGGTCCGTGAAACG
ATGGAGCGCCGTGCAGCCGGTCTTAAACCGCCCGCCTGGGCGGCTGCAGCAT
TTGAATCCGGGCTGCGACAATCAACAGAGGAGGAGAAGAGTGACAGCAGAG
CTGCGTAATCTCCCGCATATTGCCAGCATGGCCTTTAATGAGCCGCTGATGCT
TGAACCCGCCTATGCGCGGGTTTTCTTTTGTGCGCTTGCAGGCCAGCTTGGGA
TCAGCAGCCTGACGGATGCGGTGTCCGGCGACAGCCTTGACTGCCCAGGAGG
CACTTCGCGGACGCTGGCATTAATCCGGTTGATTGATTGACACGGACCACAC
GGACCAGGCCCCNGAAGTTAATCCAGGGTCANATGGACACGCCATATCGCGC
GTGTGCTNTGCCGTTGTCCCCGCNACGCGCTGTGTNTAGAGCCNGAGCACNC
GCGGGCGCGCTGCGCAGCCCGTGACNCTCGGGGGTANGACACGGTGTACAC
ACNCGCGCATATTATACGGCGCCGTCTCTGCGCACACAGCTGTGCACACGCG
ATCTCGATATGTGTGACACGCGACTATCTGCGCTCCGATAATGGGACACACG
CGCCCGGCGCGGATAGTGTGTGCGCGGGGCACTATTGACACGCGCGCTGTAC
ACANNNTACGCGCCGTGTGTCGCTGTACACTATAAAACACGGTATAGGGGCG
GCGTTTGCGCAACACAGCANTATAAACTGCGAGATGTCGCAGGGCACGTTTG
TCTCTTGTGCACATGGCNCGCCTCTCCCGCCGTNCTGTGTTACAGCAGAGCAC
GCGCCGGACACGAGCTCTCATCNGGCGTGNATANAGGTGCTCTACAC

YP1 9 R

GACGGTATCGATAAGCTTGATTCCAGTCTGCATCTTCATCATCGAGATAGCTG
TCGTCATAGGACTCAGCGGTCAGTTCGCCGGGCGTCAGGTCTTTAACTTTTGC
CAGACGCGACCAGTCAACGTCTGAAAGCGGATTCGCGTAAGGGTCACCGCTC
CCCTTATAAACCCACAGGGTGGTCCCGGCACCTTTCACCGGCATTGTAGGATT
TGGTACAGGCATAGCGTCCTCACATTTCATAGGTAATGACATAAGTCAGATC
GGCTGAACTCCACAAGCCCGCATCATCGTCGCGCCGGTAGTCATAGCCGCTG
GCCACCATACTGGTGATCAAATCTGACAGTGCCGGGATATCGCTCATCACCG
GATAAATCCGGGACTCCATCCACGCATCCAGCTCTGAATCCGGCACCTGAGC
AGGCAGGAAAACTTCGATATGCAGCTCCGCCTGCCAGGTATCGCTGTCCAGC
TCTTCGCCCGTGTATTCAGCGCCGGTGAGATAAACGGCAACTGCCGGAAAAT
CCGCCTCATCAAAAACAGCGGGGCGACCATCAAAAAACGTCGCCCCGGTGTC
A

YP1 1-10 F

AACTAGTGGATCCCCCGGGCTGCAGGAATTCGATGTCAGGAAGAACTGAACA
AGGCACTGAAAGACGGGAAAATCCTGCAGGCGGATTACAACACGCTGATGG
CGGCGGCGAAAAAGGATTATGAAGCGACGCTGAAAAAGCCGAAACAGTCCA
GCGTGAAGGTGTCTGCGGGCGATCGTCAGGAAGACAGTGCTCATGCTGCCCT
GCTGACGCTTCAGGCAGAACTCCGGACGCTGGAGAAGCATGCCGGAGCAAA
TGAGAAAATCAGCCAGCAGCGCCGGGATTTGTGGAAGGCGGAGAGTCAGTTC
GCGGTACTGGAGGAGGCGGCGCAACGTCGCCAGCTGTCTGCACAGGAGAAA
TCCCTGCTGGCGCATAAAGATGAGACGCTGGAGTACAAACGCCAGCTGGCTG
CACTTGGCGACAAGGTTACGTATCAGGAGCGCCTGAACGCGCTGGCGCAGCA
GGCGGATAAATTCGCACAGCAGCAACGGGCAAAACGGGCCGCCATTGATGC
GAAAAGCCGGGGGCTGACTGACCGGCAGGCAGAACGGGAAGCCACGGAACA
GCGCCTGAAGGAACAGTATGGCGATAATCCGCTGGCGCTGAATAACGTCATG
TCAGAGCAGAAAAAGACCTGGGCGGCTGAAGACCAGCTTCGCGGGAACTGG
ATGGCAGGCCTGAAGTCCGGCTGGAGTGAGTGGGAAGAGAGCGCCACGGAC
AGTATGTCGCAGTAAAAGTGCAGCCACGCAGACCTTTGATGGTATTGCACAG
AATATGGCGGCGATGCTGACCGGCAGTGAGCAGAACTGGCGCAGCTTCACCC
GTTCCGTGCTGTCCATGATGACAGAAATTCTGCTTAAGCAGCAATGGTGGGG
ATTGTCGGAGTATCGCAGCCGCCATTGGCGGGC

YP1 1-11 F

GCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATTAATGGAAAGG
AAGCCGCCAAAGTTGCCGACGTTATTGCGGAACTTACAACCGCTCAGGCATT
TGCTGCATTTATCCTTCGTGATATCGGACGTTGGCTGGTCATATTCATCCGCG
ACAGCCGGACCGCTATAACCGCACTCGTCACCGCGATAGGTCCAGGTGCAGG
TGTTGGCCAGCATGATACGTCCCGGAAAAACAGCGCCATCCGTTTCCGTCGG
CGTGGACAGTACAAAGGAGGCACTCACCGCGCTCAGTTCGCTGCACTGCTCA
ATGCGCCAGCGGCTGATCACCTCCTGCTCCGGATCGGCGTAACTGTTTCCGTT
GACGAAGTTCACCGCATCCAGAAAACGGGCGTAAACCTTACGCCGGACCACC
GTTCCGCCGACCAGACTCTGCATATCATCAAGCTTATCGATACCGTCGACCTC
GAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGC
GCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTC
ACAATTCCACACAACATACGAGCCGGGAGCATAAAGTGTAAAGCCTGGGGTG
CCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTC
CAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGG
GGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCG
CTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
TAATACGGTTATCCACAGAATCAGGGATAACGC

YP1 1-13 F

CGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATGGGCAGATAG
GTCCGGCTGCTCTGAAGGCGGTGTATGACATGGCCCGCAAGGGTGCCCGTGA
TGAAATTCAGACACAGATGCGTGATGGTGGCCTGTTCTCCGGAGGTGGACGA
TGAAGACCTTCCGCTGGAAAGTGAAACCCGGTATGGATGTGGCTTCGGTCCC
TTCTGTAAGAAAGGTGCGCTTTGGTGATGGCTATTCTCAGCGAGCGCCTGCCG
GGCTGAATGCCAACCTGAAAACGTACAGCGTGACGCTTTCTGTCCCCCGTGA
GGAGGCCACGGTACTGGAGTCGTTTCTGGAAGAGCACGGGGGCTGGAAATCC
TTTCTGTGGACGCCGCCTTATGAGTGGCGGCAGATAAAGGTGACCTGCGCAA
AATGGTCGTCGCGGGTCAGTATGCTGCGTGTTGAGTTCAGCGCAGAGTTTGA
ACAGGTGGTGAACTGATGCAGGATATCCGGCAGGAAACACTGAATGAATGC
ACCCGTGCGGAGCAGTCGGCCAGCGTGGTGCTCTGGGAAATCGACCTGACAG
AGGTCGGTGGAGAACGTTATTTTTTCTGTAATGAGCAGAACGAAAAAGGTGA
GCCGGTCACCTGGCAGGGGCGACAGTATCAGCCGTATCCCATTCAGGGGAGC
GGTTTTGAACTGAATGGCAAAGGCACCAGTACGCGCCCCACGCTGACGGTTT
CTAACCTGTACGGTATGGTCACCGGGATGGCGGAAGATATGCAGAGTCTGGT
CGGCGGAACGGTGGTCCGGCGTAAGGTTTACGCCCGTTTTCTGGATGCGGTG
AACTTCGTCAACGGAAACAGTTACGCCGATCCGGAGCAGGAGGTGATCAGCC
GCTGGCGCATTGAGCAGTGCAGCGAACTGAGCGCGGTGAGTGCCTCCTTTGT
ACTGTCCACGCCGA

YP1 1-19 F

GCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATCGCCGCAGGCG
TCGTAAAAAGGGGCAGCGTTCATCCCTGAAAGGTGGCGGCAGCGTGCTTGTG
GTGGGTAACCGTCGTATTCCCGGCGCGTTTATTCAGCAACTGAAAAATGGCC
GGTGGCATGTCATGCAGCGTGTGGCTGGGAAAAACCGTTACCCCATTGATGT
GGTGAAAATCCCGATGGCGGTGCCGCTGACCACGGCGTTTAAACAAAATATT
GAGCGGATACGGCGTGAACGTCTTCCGAAAGAGCTGGGCTATGCGCTGCAGC
ATCAACTGAGGATGGTAATAAAGCGATGAAACATACTGAACTCCGTGCAGCC
GTACTGGATGCACTGGAGAAGCATGACACCGGGGCGACGTTTTTTGATGGTC
GCCCCGCTGTTTTTGATGAGGCGGATTTTCCGGCAGTTGCCGTTTATCTCACC
GGCGCTGAATACACGGGCGAAGAGCTGGACAGCGATACCTGGCAGGCGGAG
CTGCATATCGAAGTTTTCCTGCCTGCTCAGGTGCCGGATTCAGAGCTGGATGC
GTGGATGGAGTCCCGGATTTATCCGGTGATGAGCGATATCCCGGCACTGTCA
GATTTGATCACCAGTATGGTGGCCAGCGGCTATGACTACCGGCGCGACGATG
ATGCGGGCTTGTGGAGTTCAGCCGATCTGACTTATGTCATTACCTATGAAATG
TGAGGACGCTATGCCTGTACCAAATCCTACAATGCCGGTGAAAGGTGCCGGG
ACCACCCTGTGGGTTTATAAGGGGAGCGGTGACCCTTACGCGAATCCGCTTTC
AGACGTTGACTGGTCGCGTCTGGCAAAAGTTAAAGACCTGACGCCCGGCGAA
CTGACCGCTGAGTCCTATGACGACAGCTATCTCGATGAT

YP1 1-20 F

CGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATTTCAGG
GGAGCGGTTTTGAACTGAATGGCAAAGGCACCAGTACGCGCCCCACGCTGAC
GGTTTCTAACCTGTACGGTATGGTCACCGGGATGGCGGAAGATATGCAGAGT
CTGGTCGGCGGAACGGTGGTCCGGCGTAAGGTTTACGCCCGTTTTCTGGATGC
GGTGAACTTCGTCAACGGAAACAGTTACGCCGATCCGGAGCAGGAGGTGATC
AGCCGCTGGCGCATTGAGCAGTGCAGCGAACTGAGCGCGGTGAGTGCCTCCT
TTGTACTGTCCACGCCGACGGAAACGGATGGCGCTGTTTTTCCGGGACGTATC
ATGCTGGCCAACACCTGCACCTGGACCTATCGCGGTGACGAGTGCGGTTATA
GCGGTCCGGCTGTCGCGGATGAATATGACCAGCCAACGTCCGATATCACGAA
GGATAAATGCAGCAAATGCCTGAGCGGTTGTAAGTTCCGCAATAACGTCGGC
AACTTTGGCGGCTTCCTTTCCATTAACAAACTTTCGCAGTAAATCCCATGACA
CAGACAGAATCAGCGATTCTGGCGCACGCCCGGCGATGTGCGCCAGCGGAGT
CGTGCGGCTTCGTGGTAAGCACGCCGGAGGGGGAAAGATATTTCCCCTGCGT
GAATATCTCCGGTGAGCCGGAGGCGTATTTCCGTATGTCGCCGGAAGACTGG
CTGCAGGCAGAAATGCAGGGTGAGATTGTGGCGCTGGTCCACAGCCACCCCG
GTGGTCTGCCCTGGCTGAGTGAGGCCGACCGGCGGCTGCAGGTGCAGAGTGA
TTTGCCGTGGTGGCTGGTCTGCCGGGGGACGATTCATAAGTTCCGCTGTGTGC
CGCATCTCACCGGGCGGCGCTTTGAGCACGGTGTGACGGACTGTTACACACT
GTTCCGGGATGCTTATCATCTGGCGGGGATTGAGATGCCCGGACTTTCATCGT
GAGGATGACTGGTGGCGTA

YP1 1-22 F

GCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATTGATG
CTGCAGCGCATAGCCCAGCTCTTTCGGAAGACGTTCACGCCGTATCCGCTCA
ATATTTTGTTTAAACGCCGTGGTCAGCGGCACCGCCATCGGGATTTTCACCAC
ATCAATGGGGTAACGGTTTTTCCCAGCCACACGCTGCATGACATGCCACCGG
CCATTTTTCAGTTGCTGAATAAACGCGCCGGGAATACGACGGTTACCCACCA
CAAGCACGCTGCCGCCACCTTTCAGGGATGAACGCTGCCCCTTTTTACGACGC
CTGCGGCGCGAAAGGACAACCCGCGCATTACCCAGCTTGATTACGGGCAAAT
CCCCCCGGTTAACTTTGATTCTGGCCTGCGGATTTTTGACCGTGGCCCTTTTCA
GCCTGGCCCTTTCCTTTACCAGTTTCCGGCGTACCTTTGTCTCACGGGCAACC
TGTGACGCCGACTGCGATATCGCGGATGAAGCAACGCGGTTAATGGCCATTG
CGGCGGCACCAGGCACCGCCGTTTTGCTGATACGGCTGAGGTTTTCAACGGC
CTGCTCAAGACCTTTTATGGCCATACATCCCCCTTTCAGCGGCGACGGTTAAC
GGCAGGCGGTACGCCCCGTCCAAGCCAGAGATGACAACTTCCGCCATCATCC
GGCGAAACCCGATCTACCCAGAAATTTTCCTCACCGATGGTCAGCGTGTCTCC
ACGCCGCAGCTGCCGCACCTCATCAGTCCGGACAAACAGGGACGGGCTGGA
GCCTTCAACGCGCACGCCCTGTCCGGCATAGCTGATATTTTCAGGGTCATCAA
AAACACCACGTATCACCGCACCTGACTGCTCACCGGATGTAATGGTGGCTGA
CGTTCCCATGTACCCGCGTATCGTTTCATCGGCGCGGGCAATGGCAGCATCG
AACAGGTTATCGAAATCAGCCACAGCGCCTCCCGTTATTGCATTCTGGCCAG
GCCGCGCTCTGT

YP1 1-25 F

CGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATGAACCA
ACACACAGGGTATTGCTTATTTATACTAACTTGAGCGAAACGGGAAGGTAAA
AAGACAAAAAGTTGTTTTTAATACCTTTAAGTGATACCAGATGGCATTGCGCC
ATCTGGCAGAGTGATTAACTAAACATCGCAGTAATCGAGGCGCTTGCCAGAG
AGTGGAAATGAACGTTAAACCCGACCATCGCGCCGCTGGCACCTTCATCGAC
ATCAATACGTTCTATATCCAGCGCGTGAACGGTAAAAATGTAGCGATGAGTT
TCGCCTTTCGGCGGTGCTGCGCCATCGTACCCGGTTTTACCAAAGTCGGTACG
CGTCTGCAAAACGCCGTCTGGCATTGCTACCAGACCAGAGCCAAACCCTTGC
GGTAATACGCGGGTATCAGCGGGTAAGTTAACAACTACCCAGTGCCACCAGC
CGGAGCCGGTTGGCGCATCCGGGTCGTAGCAGGTGACAACAAAACTTTTCGT
TCCCGCAGGAACATCATCCCACGCCAGATGCGGTGAAATATTATCGCCATCG
TAACCCATGCCGTTAAAGACATGACGATGCGGCAATTTATCGCCATCGCGCA
GATCGTTACTGATGAGTTTCATGAACCCTCCTTTCTTGTTTGCAGAAAGTGTA
GCCAGAAACCCTCACGCGGACTTCTCGTTATTGGCAAAAAAATGTTTCATCCT
GTACCGCGCGGTTAACCGCTGCGGTCAGACGCTGCAACTGTTGCGGGAGAAT
AATATAGGGCGGCATCAGGTAAATCAGTTTGCCAAAAGGCCGGATCCAGACA
CCCTGTTCGACAAAGAATTTTTGCAGCGCCGCCATATTCACCGGATGAGTGGT
TTCGACCACGCCAATGGCCCCCAGTACGCGCACATCGGCAACCATTTCGGCA
TCACGGGCGGGGGC

YP1 1-32 F

CCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAA
TTCGATGGCTTTTTTGTGGGGTGAATATGGCAGTAAAGATTTCAGGAGTCCTG
AAAGACGGCACAGGAAAACCGGTACAGAACTGCACCATTCAGCTGAAAGCC
AGACGTAACAGCACCACGGTGGTGGTGAACACGGTGGGCTCAGAGAATCCG
GATGAAGCCGGGCGTTACAGCATGGATGTGGAGTACGGTCAGTACAGTGTCA
TCCTGCAGGTTGACGGTTTTCCACCATCGCACGCCGGGACCATCACCGTGTAT
GAAGATTCACAACCGGGGACGCTGAATGATTTTCTCTGTGCCATGACGGAGG
ATGATGCCCGGCCGGAGGTGCTGCGTCGTCTTGAACTGATGGTGGAAGAGGT
GGCGCGTAACGCGTCCGTGGTGGCACAGAGTACGGCAGACGCGAAGAAATC
AGCCGGCGATGCCAGTGCATCAGCTGCTCAGGTCGCGGCCCTTGTGACTGAT
GCAACTGACTCAGCACGCGCCGCCAGCACGTCCGCCGGACAGGCTGCATCGT
CAGCTCAGGAAGCGTCCTCCGGCGCAGAAGCGGCATCAGCAAAGGCCACTG
AAGCGGAAAAAAGTGCCGCAGCCGCAGAGTCCTCAAAAAACGCGGCGGCCA
CCAGTGCCGGTGCGGCGAAAACGTCAGAACGAATGCTGCAGCGTCACAACA
ATCAGCCGCCACGTCTGCCTCCACCGCGGCCACGAAAGCGTCAGAGGCCGCC
ACTTCAGCACGAGATGCGGTGGCCTCAAAGAGGCAGCAAAATCATCAGAAA
CGAACGCATCATCAAGTGCCGGTCGTGCAGCTTCCTCGGCAACGGCGGCAGA
AAATTCTGCCAGGGCGGCAAAAACGTCCGAGACGAATGCCAGGTCATCTGAA
ACAG

YP1 2-5 F

CTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATCGTACAGCGTGA
CGCTTTCTGTCCCCCGTGAGGAGGCCACGGTACTGGAGTCGTTTCTGGAAGA
GCACGGGGGCTGGAAATCCTTTCTGTGGACGCCGCCTTATGAGTGGCGGCAG
ATAAAGGTGACCTGCGCAAAATGGTCGTCGCGGGTCAGTATGCTGCGTGTTG
AGTTCAGCGCAGAGTTTGAACAGGTGGTGAACTGATGCAGGATATCCGGCAG
GAAACACTGAATGAATGCACCCGTGCGGAGCAGTCGGCCAGCGTGGTGCTCT
GGGAAATCGACCTGACAGAGGTCGGTGGAGAACGTTATTTTTTCTGTAATGA
GCAGAACGAAAAAGGTGAGCCGGTCACCTGGCAGGGGCGACAGTATCAGCC
GTATCCCATTCAGGGGAGCGGTTTTGAACTGAATGGCAAAGGCACCAGTACG
CGCCCCACGCTGACGGTTTCTAACCTGTACGGTATGGTCACCGGGATGGCGG
AAGATATGCAGAGTCTGGTCGGCGGAACGGTGGTCCGGCGTAAGGTTTACGC
CCGTTTTCTGGATGCGGTGAACTTCGTCAACGGAAACAGTTACGCCGATCCG
GAGCAGGAGGTGATCAGCCGCTGGCGCATTGAGCAGTGCAGCGAACTGAGC
GCGGTGAGTGCCTCCTTT

YP1 2-17 F

GGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATTGGTTTTC
GTCATCCGGTGAAGAGATTGAGCCACCTGACAGTGTGACCTTTCACATCTGG
ACAGCGTACAGCCCGTTCACCACCTGGGTGCAGATTGTCAAAGACTGGATGA
AAACGAAAGGGGATACGGGAAAACGTAAAACCTTCGTAAACACCACGCTCG
GTGAGACGTGGGAGGCGAAAATTGGCGAACGTCCGGATGCTGAAGTGATGG
CAGAGCGGAAAGAGCATTATTCAGCGCCCGTTCCTGACCGTGTGGCTTACCT
GACCGCCGGTATCGACTCCCAGCTGGACCGCTACGAAATGCGCGTATGGGGA
TGGGGGCCGGGTGAGGAAAGCTGGCTGATTGACCGGCAGATTATTATGGGCC
GCCACGACGATGAACAGACGCTGCTGCGTGTGGATGAGGCCATCAATAAAAC
CTATACCCGCCGGAATGGTGCAGAAATGTCGATATCCCGTATCTGCTGGGAT
ACTGGCGGGATTGACCCGACCATTGTGTATGAACGCTCGAAAAAACATGGGC
TGTTCCGGGTGATCCCCATTAAAGGGGCATCCGTCTACGGAAAGCCGGTGGC
CAGCATGCCACGTAAGCGAAACAAAAACGGGGTTTACCTTACCGAAATCGGT
ACGGATACCGCGAAAG

YP1 2-18F

GCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATCCGGATGA
CCCCTCCAGCGTGTTTTATCTCTGCGAGCATAATGCCTGCGTCATCCGCCAGC
AGGAGCTGGACTTTACTGATGCCCGTTATATCTGCGAAAAGACCGGGATCTG
GACCCGTGATGGCATTCTCTGGTTTTCGTCATCCGGTGAAGAGATTGAGCCAC
CTGACAGTGTGACC

YP1 2-22 F

GCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATTTGAC
TGCGCTGACATCATCGCCCGTGTGCGTGACATAAAACCGGTATGGGCGCTTG
CCAACGACATGAACTGCAGTGCAGGTCAGTTGCTTGCCAGTGCCGCCTCCCG
GCGTCTGGTCACGCAGACCGCCCGGACAGGCTCCATCGGCGTCATGATGGCT
CACAGTAATTACGGTGCTGCGCTGGAGAAACAGGGTGTGGAAATCACGCTGA
TTTACAGCGGCAGCCATAAGGTGGATGGCAACCCCTACAGCCATCTTCCGGA
TGACGTCCGGGAGACACTGCAGTCCCGGATGGACGCAACCCGCCAGATGTTT
GCGCAGAAGGTGTCGGCATATACCGGCCTGTCCGTGCAGGTTGTGCTGGATA
CCGAGGCTGCAGTGTACAGCGGTCAGGAGGCCATTGATGCCGGACTGGCTGA
TGAACTTGTTAACAGCACCGATGCGATCACCGTCATGCGTGATGCACTGGAT
GCACGTAAATCCCGTCTCTC

YP1 2-23 F

CTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATAGCGCGCTTCAG
TTTTCAGGAAGCCCGCAGTGCCTGGGGGAACTGCGACTGGATAGGCTCCGGT
CGTATGGCCATCGATGGTCTGAAAGAAGTTCAGGAAGCGGTGATGCTGATAG
AAGCCGGACTGAGTACCTACGAGAAAGAGTGCGCAAAACGCGGTGACGACT
ATCAGGAAATTTTTGCCCAGCAGGTCCGTGAAACGATGGAGCGCCGTGCAGC
CGGTCTTAAACCGCCCGCCTGGGCGGCTGCAGCATTTGAATCCGGGCTGCGA
CAATCAACAGAGGAGGAGAAGAGTGACAGCAGAGCTGCGTAATCTCCCGCA
TATTGCCAGCATGGCCTTTAATGAGCCGCTGATGCTTGAACCCGCCTATGCGC
GGGTTTTCTTTTGTGCGCTTGCAGGCCAGCTTGGGATCAGCAGCCTGACGGAT
GCGGTGTCCGGCGACAGCCTGACTGCCCAGGAGGCACTCGCGACGCTGGCAT
TATCCGGTGATGATGACGGACCACGACAGGCCCGCAGTTATCAGGTCATGAA
CGGCATCGCCGTGCTGCCGGTGTCCGGCACGCTGGTCAGCCGGACGCGGGCG
CTGCAGCCGTACTCGGGGATGACCGGTTACAACGGCATTATCGCCCGTCTGC
AACAGGCTGCCAGCGAT

YP1 2-24F

GCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATTCGGTTAT
AAATTCTGATTAGCCAGGTAACACAGTGTTATGACAGCCCGCCGGAACCGGT
GGGCTTTTTTGTGGGGTGAATATGGCAGTAAAGATTTCAGGAGTCCTGAAAG
ACGGCACAGGAAAACCGGTACAGAACTGCACCATTCAGCTGAAAGCCAGAC
GTAACAGCACCACGGTGGTGGTGAACACGGTGGGCTCAGAGAATCCGGATG
AAGCCGGGCGTTACAGCATGGATGTGGAGTACGGTCAGTACAGTGTCATCCT
GCAGGTTGACGGTTTTCCACCATCGCACGCCGGGACCATCACCGTGTATGAA
GATTCACAACCGGGGACGCTGAATGATTTTCTCTGTGCCATGACGGAGGATG
ATGCCCGGCCGGAGGTGCTGCGTCGTCTTGAACTGATGGTGGAAGAGGTGGC
GCGTAACGCGTCCGTGGTGGCACAGAGTACGGCAGACGCGAAGAAATCAGC
CGGCGATGCCAGTGCATCAGCTGCTCAGGTCGCGGCCCTTGTGACTGATGCA

ACTGACTCAGCACGCGCCGCCAGCACGTCCGCCGGACAGGCTGCATCGTCAG
CTCAGGAAGCGTCCTCCGGCGCAGAAGCGGCATCAGCAAAGGCCACTGAAG
CGGAAAAAAGTGCCGCAGCCGCAGAGTCCTCAAAAAACGCGGCGGCCACCA
GTGCCGGTGCGGCGAAAACGTCAGAAACGAATGCTGCAGCGT

YP1 2-26F

CTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATTGCGCCCTGGAC
AATCGGCTCAATGATCACCGCCGCGATTTCATGACGATGCGCCGCCATCAGG
CGGGCAAAGCCCACCATATCGCGCTCATCCCATTCGCCATCCATGCGGCTTTG
CGGGGCGGGAGCAAACAGGTTTTCTGGCAGGTAGCCTTTCCACAGACTGTGC
ATTGAGTTATCCGGATCGCACACCGACATCGCGCCAAAGGTATCGCCATGAT
AACCATTGCGGAAGGTCAGAAAACGCTGGCGCGCTTCGCCTTTGGCTTGCCA
GTACTGCAACGCCATTTTCATCGCCACTTCCACCGCTACGGAACCGGAGTCCG
CGAGAAAAACGCACTCCAGCGGTTGCGGCGTCATCGCCACCAGTTTGCGGCA
CAGCTCAATGGCTGGCGCATGGGTGATACCGCCAAACATCACATGCGACATG
GCATCAATTTGCGACTTCATCGCCGCATTAAGCTGCGGGTGATTGTAGCCGTG
GATCGATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCA
GCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCA
TAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACAC

YP1 2-27F

AACTAGTGGATCCCCCGGGCTGCAGGAATTCGATCCAGATTGTCCTGGGGGC
TGCCGCCATTGCCGGATCATTCTTTACCGCCGGAGCCACCCTTGCAGCATGGG
GGGCAGCCATTGGGGCCGGTGGTATGACCGGCATCCTGTTTTCTCTCGGTGCC
AGTATGGTGCTCGGTGGTGTGGCGCAGATGCTGGCACCGAAAGCCAGAACTC
CCCGTATACAGACAACGGATAACGGTAAGCAGAACACCTATTTCTCCTCACT
GGATAACATGGTTGCCCAGGGCAATGTTCTGCCTGTTCTGTACGGGGAAATG
CGCGTGGGGTCACGCGTGGTTTCTCAGGAGATCAGCACGGCAGACGAAGGGG
ACGGTGGTCAGGTTGTGGTGATTGGTCGCTGATGCAAAATGTTTTATGTGAAA
CCGCCTGCGGGCGGTTTTGTCATTTATGGAGCGTGAGGAATGGGTAAAGGAA
GCAGTAAGGGGCATACCCCGCGCGAAGCGAAGGACAACCTGAAGTCCACGC
AGTTGCTGAGTGTGATCGATGCCATCAGCGAAGGGCCGATTGAAGGTCCGGT
GGATGGCTTAAAAAGCGTGCTGCTGAACAGTACGCCGGTGCTGGACACTGAG
GGGAATACCAACATATCCGGTGTCACGGTGGTGTTCCGGGCTGGTGAGCAGG
AGCAGACTCCGCCG

YP1 2-29F

GCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATCATATTGTTCCC
AGAGTCGCCGGGGCCAAGTCAGGTGGCGTATTCCAGATTGTCCTGGGGGCTG
CCGCCATTGCCGGATCATTCTTTACCGCCGGAGCCACCCTTGCAGCATGGGG

GGCAGCCATTGGGGCCGGTGGTATGACCGGCATCCTGTTTTCTCTCGGTGCCA
GTATGGTGCTCGGTGGTGTGGCGCAGATGCTGGCACCGAAAGCCAGAACTCC
CCGTATACAGACAACGGATAACGGTAAGCAGAACACCTATTTCTCCTCACTG
GATAACATGGTTGCCCAGGGCAATGTTCTGCCTGTTCTGTACGGGGAAATGC
GCGTGGGGTCACGCGTGGTTTCTCAGGAGATCAGCACGGCAGACGAAGGGG
ACGGTGGTCAGGTTGTGGTGATTGGTCGCTGATGCAAAATGTTTTATGTGAAA
CCGCCTGCGGGCGGTTTTGTCATTTATGGAGCGTGAGGAATGGGTAAAGGAA
GCAGTAAGGGGCATACCCCGCGCGAAGCGAAGGACAACCTGAAGTCCACGC
AGTTGCTGAGTGTGATCGATGCCATCAGCGAAGGGCCGATTGAAGGTCCGGT
GGATGGCTTAAAAAGCGTGCTGCTGAACAGTACGCCGGTGCTGGACACTGAG
GGGAATACCAACATATCCGGTGTCACGGTGGTGTTCCGGG

YP1 2-30F

CGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATTGATGCGCTTC
GGGCTGACCATCCGGAACTGTGTCCGGAAAAGCCGCGACGAACTGGTATCCC
AGGTGGCCTGAACGAACAGTTCACCGTTAAAGGCGTGCATGGCCACACCTTC
CCGAATCATCATGGTAAACGTGCGTTTTCGCTCAACGTCAATGCAGCAGCAG
TCATCCTCGGCAAACTCTTTCCATGCCGCTTCAACCTCGCGGGAAAAGGCAC
GGGCTTCTTCCTCCCCGATGCCCAGATAGCGCCAGCTTGGGCGATGACTGAG
CCGGAAAAAAGACCCGACGATATGATCCTGATGCAGCTGGATGGCGTTGGCG
GCATAGCCGTTATTGCGTACCAGATCGTCTGCGCGGGCATTGCCACGGGTAA
AGTTGGGCAACAGGGCTGCATCCACACTTTCACTCGGTGGGTTCCAC

YP1 3-1F

TAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATCTTTCTGTCCCCCGTG
AGGAGGCCACGGTACTGGAGTCGTTTCTGGAAGAGCACGGGGGCTGGAAATC
CTTTCTGTGGACGCCGCCTTATGAGTGGCGGCAGATAAAGGTGACCTGCGCA
AAATGGTCGTCGCGGGTCAGTATGCTGCGTGTTGAGTTCAGCGCAGAGTTTG
AACAGGTGGTGAACTGATGCAGGATATCCGGCAGGAAACACTGAATGAATG
CACCCGTGCGGAGCAGTCGGCCAGCGTGGTGCTCTGGGAAATCGACCTGACA
GAGGTCGGTGGAGAACGTTATTTTTTCTGTAATGAGCAGAACGAAAAAGGTG
AGCCGGTCACCTGGCAGGGGCGACAGTATCAGCCGTATCCCATTCAGGGGAG
CGGTTTTGAACTGAATGGCAAAGGCACCAGTACGCGCCCCACGCTGACGGTT
TCTAACCTGTACGGTATGGTCACCGGGATGGCGGAAGATATGCAGAGTCTGG
TCGGCGGAACGGTGGTCCG

YP1 3-3F

CTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATAAACAGATTGAG
CAGGAAGTGCTTACCACCTGGCCCACGGAGGCAATTTCTCATGCTGAAAACG
TGGTGTACCGGCTGTCTGGTATGTATGAGTTTGTGGTGAATAATGCCCCTGAA

CAGACAGAGGACGCCGGGCCCGCAGAGCCTGTTTCTGCGGGAAAGTGTTCGA
CGGTGAGCTGAGTTTTGCCCTGAAACTGGCGCGTGAGATGGGGCGACCCGAC
TGGCGTGCCATGCTTGCCGGGATGTCATCCACGGAGTATGCCGACTGGCACC
GCTTTTACAGTACCCATTATTTTCATGATGTTCTGCTGGATATGCACTTTTCCG
GGCTGACGTACACCGTGCTCAGCCTGTTTTTCAGCGATCCGGATATGCATCCG
CTGGATTTCAGTCTGCTGAACCGGCGCGAGGCTGACGAAGAGCCTGAAGATG
ATGTGCTGATGCAGAAAGCGGCAGGGCTTGCCGGAGGTGTCCGCTTTGGCCC
GGACGGGAATGAAGTTATCCCCGCTTCCCCGGATGTGGCG

YP1 3-5F

TAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATCGCTCGCCGCCTTTA
CAATGTCCCCGACGATTTTTTCCGCCCTCAGCGTACCGTTTATCGTACAGTTTT
CAGCTATCGTCACATTACTGAGCGTCCCGGAGTTCGCATTCACACTGCCACTG
ATATCCGCATTTTTAGCGGTCAGCTTTCCGTCCGGTGTCAGGGAAAAGGCCGG
AGGATTGCCGCCGCTGGTAATGGTGGGGGCCGTCAGGCGCTTCAGGAACACG
TCGTTCATGAATATCTGGTTGCCCTGCGCCACAAACATCGGCGTTTCATTCCC
GTTTGCCGGGTCAATAAATGCGATACGATTGGCGGCAACCAGAAACTGGCTC
AGTTTGCCTTCCTCCGTGTCCTCCATGCTGAGGCCAATACCCGCGACATAATG
TTTGCCGTCTTTGGTCTGCTCAATTTTGACAGCCCACATGGCATTCCACTTATC
ACTGGCATCCTTCCACTCTTTCGAAAACTCCTCCAGTCTGCTGGCGTTATCCT
CCGTCAGCTCGACTTTTTCCAGCAGCTCCTTGCCGAGATGGGATTCGGTTATC
TTGCCTTTGAAAAAATCCAGGTAACCTTCCGCATC

YP1 3-8F

CTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATTGCAGCCTCGGTAT
CCAGCACAACCTGCACGGACAGGCCGGTATATGCCGACACCTTCTGCGCAAA
CATCTGGCGGGTTGCGTCCATCCGGGACTGCAGTGTCTCCCGGACGTCATCCG
GAAGATGGCTGTAGGGGTTGCCATCCACCTTATGGCTGCCGCTGTAAATCAG
CGTGATTTCCACACCCTGTTTCTCCAGCGCAGCACCGTAATTACTGTGAGCCA
TCATGACGCCGATGGAGCCTGTCCGGGCGGTCTGCGTGACCAGACGCCGGGA
GGCGGCACTGGCAAGCAACTGACCTGCACTGCAGTTCATGTCGTTGGCAAGC
GCCCATACCGGTTTTATGTCACGCACACGGGCGATGATGTCAGCGCAGTCAA
ATGCCCCCGCCACCATCCCGCCGGGCGTGTCCATATCGAGCAGAATGCCGTC
CACCATCGGATCGCTGGCAGCCTGTTGCAGACGGGCGATAATGCCGTTGTAA
CCGGTCATCCCCGAGTACG

NR-YP1 with M8

AATAAGCAGGGCCAGCGCAGTAGCGAGTAGCATTTTTTTCATGGTGTTATTCC
CGATGCTTTTTGAAGTTCGCAGAATCGTATGTGTAGAAAATTAAACAAACCCT
AAACAATGAGTTGAAATTTCATATTGTTAATATTTATTAATGTATGCCAGGTG

CGATGAATCGTCATTGTATTCCCGGATTAACTATGTCCACAGCCCTGACGGGG
AACTTCTCTGCGGGAGTGTCCGGGAATAATTAAAAACGATGCACACAGGGTT
TAGCGCGTACATGTATTGTATTATGCCAACACCCCGGTGCTGACACGGAAGA
AACCGGACGTTATGATTTAGCGTGGAAAGATTTGTGTAGTGTTCTGAATGCTC
TCAGTAAATAGTAATGAATTATCAAAGGTATAGTAATATCTTTTATGTTCGTG
GATATTTGTAATCCATCGGAAAACTCCTGCTTTAGCAAGATTTTCCCTGTATT
GCTGAAATGTGATTTCTCTTGATTTCAACCTATCATAGGACGTTTCTATAAGA
TGCGTATTTCTTGAGAATTTAACATTTACAACCTTTTTAAGTCCTTTTATTAAC
ACGGTGTTATCGTTTTCTAACACAATGTGAATATTATCTGTGGCTAGATAGTA
AATATAATGTGAGACATTGTGACGTTTTAGTTCAGAATAAAACAATTCACAG
TTTAAATCT

NR-YP1 with M17

GCGCCGATTGTCGAAGATATTCTCCAGTTTGAGCTGGAAAACCTGCCTGGAC
AGAGCGCGGCACGACGCTGGTGGCGACGTCATCACAAAGCGGAAGAGAACC
GCCAGCCGGGAGAAGCGCAGGTATTGCTGGTCTGGCGCGATAACGAAGAAC
ATCGCGATGATATTGAACGCCATTATTTGAAAATGCTCACTCAGGCGCGGCG
GGAAGTGATTATCGCCAACGCCTACTTCTTCCCCGGCTATCGATTTTTACACG
CCTTGCGTAAAGCGGCACGGCGCGGGGTGCGGATCAAACTGATCATTCAGGG
CGAACCGGATATGCCGATTGTCAGAGTCGGTGCGCGCTTGCTGTATAACTATC
TGGTTAAAGGCGGCGTTCAGGTTTTTGAGTACCGCCGCCGCCCGCTCCACGG
CAAAGTGGCATTGATGGACGATCACTGGGCGACAGTAGGGTCCAGTAATCTC
GATCCGCTCAGTTTGTCACTGAATCTCGAAGCAAATGTCATCATCCACGATCG
TCATTTTAACCAGACGCTGCGCGATAATCTGAACGGCATTATTGCCGCAGATT
GTCAGCAGGTGGATGAAACCATGCTGCCCAAACGCACCTGGTGGAACCTGAC
CAAAAGCGTGCTGGCGTTCCACTTTTTACGCCACTTCCCGGCGCTGGTTGGCT
GGCTTCCGG

NR-YP1 with M18

CGTCAGCGCCGATATTGTTTGCCTGCAGGAAGTGATGGGCGCGCACGAAGTT
CATCCGCTGCATGTGGAAAACTGGCCCGATACCTCGCACTACGAGTTTCTCGC
CGACACTATGTGGAGCGATTTTGCCTACGGTCGCAATGCCGTATACCCGGAA
GGGCATCACGGCAACGCCGTACTGTCGCGTTATCCCATTGAACATTATGAGA
ATCGCGATGTTTCGGTCGATGGTGCGGAAAAGCGCGGCGTGCTCTACTGCCG
CATTGTGCCGCCGATGACCGGAAAAGCGATTCATGTGATGTGCGTACATCTG
GGCCTGCGTGAGGCGCACCGTCAGGCGCAGCTTGCGATGCTCGCCGAATGGG
TGAATGAGCTACCGGACGGCGAACCGGTATTGGTGGCGGGTGATTTCAATGA
CTGGCGGCAAAAAGCTAATCATCCGTTAAAAGTGCAGGCCGGACTGGATGAG
ATTTTTACCCGCGCCCACGGACGCCCGGCGCGCACGTTTCCGGTGCAATTTCC
TCTACTACGACTGGACAGGATCTACGTCAAAAATGCCAGCGCCAGCGCGCCA
ACCGCGTTGCCGCTGCGGACATGGCGACACCTTTCTGATCATGCCCCTTTAAG
TGCGGAGATTCATTTATGAAATGTAGCTGGCGCGAAGGCAATAAGATCCAGT

TGCTGGAAAACGGCGAGCAATATTATCCCGCGGTGTTTAAGGCGATTGGCGA
GGCACAAGAACGCATCATTCTTGAAACGTTTATCTGGTTTGAGGAT

NR-YP1 with M19

CTGCTGGAAGGAAAACCGAAACCTAACAACAGCGAGCTGGTGGTACGCAAC
TGGTATAACCCGAATCTCGACTACAAATGGTTTGTGGTGCCGTCACTGATCGC
CATGATCACCACTATCGGCGTAATGATCGTCACTTCACTTTCCGTCGCCCGCG
AACGTGAACAAGGTACGCTCGATCAGCTACTGGTTTCGCCGCTCACCACCTG
GCAGATCTTCATCGGCAAAGCCGTACCGGCGTTAATTGTCGCCACCTTCCAG
GCCACCATTGTGCTGGCGATTGGTATCTGGGCGTATCAAATCCCCTTCGCCGG
ATCGCTGGCGCTGTTCTACTTTACGATGGTGATTTATGGTTTATCGCTGGTGG
GATTCGGTCTGTTGATTTCATCACTCTGTTCAACACAACAGCAGGCGTTTATC
GGCGTGTTTGTCTTTATGATGCCCGCCATTCTCCTTTCCGGTTACGTTTCTCCG
GTGGAAAACATGCCGGTATGGCTGCAAAACCTGACGTGGATTAACCCTATTC
GCCACTTTACGGACATTACCAAGCAGATTTATTTGAAGGATGCGAGTCTGGA
TATTGTGTGGAATAGTTTGTGGCCGCTACTGGTGATAACGGCCACGACAGGG
TCAGCGGCGTACGCGATGTTTAGACGTAAGGTGATGTAACTTCTTATCTTTCG
CCAGCAAAGACA

YP3 1-3 F

AGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATGTGCAGCCAGTGAATC
CCCGCATTTTATGCGTTTTCATGTTGCCTGCCCGCATTGCGGGGAGGAGCAGT
ATCTTAAATTTGGCGACAAAGAGACGCCGTTTGGCCTCAAATGGACGCCGGA
TGACCCCTCCAGCGTGTTTTATCTCTGCGAGCATAATGCCTGCGTCATCCGCC
AGCAGGAGCTGGACTTTACTGATGCCCGTTATATCTGCGAAAAGACCGGGAT
CTGGACCCGTGATGGCATTCTCTGGTTTTCGTCATCCGGTGAAGAGATTGAGC
CACCTGACAGTGTGACCTTTCACATCTGGACAGCGTACAGCCCGTTCACCACC
TGGGTGCAGATTGTCAAAGACTGGATGAAAACGAAAGGGGATACGGGAAAA
CGTAAAACCTTCGTAAACACCACGCTCGGTGAGACGTGGGAGGCGAAAATTG
GCGAACGTCCGGATGCTGAAGTGATGGCAGAGCGGAAAGAGCATTATTCAGC
GCCCGTTCCTGACCGTGTGGCTTACCTGACCGCCGGTATCGACTCCCAGCTGG
ACCGCTACGAAATGCGCGTATGGGGATGGGGGCCGGGTGAGGAAAGCTGGC
TGATTGACCGGCAGATTATTATGGGCCGCCACGACGATGAACAGACGCTGCT
GCGTGTGGATGAGGCCATCAATAAAACCTATACCCGCCGGAATGGTGCAGAA
ATGTCGATATCCCGTATCTGCTGGGATACTGGCGGGATTGACCCGACCATTGT
GTATGA

YP3 1-3 R

GGTCGACGGTATCGATAAGCTTGATGGTGATCTGCGCGTTCACGTCCGGCTGC
GCCGCGCTGGCGTTCTCGCCCTCCGTCGCTGGCACCACGTCAGTAACGTCAGC

CTGCGAAGCAGTGGCTGAAACAGTTGTTGATTGAGTCTCTTTGGTCATTCGCC
CTCCTGAGAGACGGGATTTACGTGCATCCAGTGCATCACGCATGACGGTGAT
CGCATCGGTGCTGTTAACAAGTTCATCAGCCAGTCCGGCATCAATGGCCTCCT
GACCGCTGTACACTGCAGCCTCGGTATCCAGCACAACCTGCACGGACAGGCC
GGTATATGCCGACACCTTCTGCGCAAACATCTGGCGGGTTGCGTCCATCCGG
GACTGCAGTGTCTCCCGGACGTCATCCGGAAGATGGCTGTAGGGGTTGCCAT
CCACCTTATGGCTGCCGCTGTAAATCAGCGTGATTTCCACACCCTGTTTCTCC
AGCGCAGCACCGTAATTACTGTGAGCCATCATGACGCCGATGGAGCCTGTCC
GGGCGGTCTGCGTGACCAGACGCCGGGAGGCGGCACTGGCAAGCAACTGAC
CTGCACTGCAGTTCATGTCGTTGGCAAGCGCCCATACCGGTTTTATGTCACGC
ACACGGGCGATGATGTCAGCGCAGTCAAATGCCCCCGCCACCATCCCGCCGG
GCGTGTCCATATCGAGCAGAATGCCGTC

YP3 1-4 F

TCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATACTGCGGGCTTCCT
GAAAACTGAAGCGCGCTTTTGAAGGTAACGTCACCACGCGGCGAACGATGGC
CTCTTCCAGCCAGCACAGAAACATCTGGCTCGCCTGACGGGATGCGACGAAT
TTTCGCCGCCCCATAAAGTACGCCCACGACTCGTTCGCACTGGCCCGTGCCGT
GGAGTAGCTCATCTGGGCGTAATTCCGGGAAAGCTGCTCATACGAGACACCC
AGCCCGGCAGCGATATACCGCAGCAGTGACTGCTCAAACACGGAGTAGCCGT
TATCCGTATCCTGAGCCGTCTGCAGGTTCAGTGAGTCACCCGGCATCAGGTGC
GGTACTTTTGCGCCTCCCAGCCGGACCGGCGCTGCGGCGTAATACGCGGCAA
TTTCACCAATCCAGCCGGTCAGCCTTTCCCATCAAGCTTATCGATACCGTCGA
CCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATT
GCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCC
GCTCACAATTCCACACAACATACGAGCCGGGAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC
TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGA

YP3 7 R

CGACGGTATCGATAAGCTTGATAGCGCGGCTGCTTTGTTCATGGCTTTGATGA
TATCCCGTTTCAGGAAATCAACATGTCGGTTTTCCAGTTCCGGAAAACGCCGC
TGCACCGACAGGGGGAGCCCGTCGAGAATACTGGCAATTTCACCTGCGATCC
GCGACAGCACGAAAGTACAGAATGCGGTTTCCACCACTTCAGCGGAGTCTCT
GGCATTCTTCAGTTCCTGTGCGTCGGCCTGCGCACGCGTAAGTCGATGGCGTT
CGTACTCAATAGTTCCTGGCTGGAGATCTGCCTCGCTGGCCTGCCGCAGTTCT
TCAACCTCCCGGCGCAGCTTTTCGTTCTCAATTTCAGCATCCCTTTCGGCATA
CCATTTTATGACGGCGGCAGAGTCATAAAGCACCTCATTACCCTTGCCACCGC
CTCGCAGAACGGGCATTCCCTGTTCCTGCCAGTTCTGAATGGTACGGATACTC
GCACCGAAAATGTCAGCCAGCTGCTTTTTGTTGACTTCCATTGTTCATTCCAC
GGACAAAAACAGAGAAAGGAAACGACAGAGGCCAAAAAGCTCGCTTTCAGC

ACCTGTCGTTTCCTTTCTTTTCAGAGGGTATTTTAAATAAAAACATTAAGTTAT
GACGAAGAAGAACGGAAACGC

YP3 1-7 F

AACTAGTGGATCCCCCGGGCTGCAGGAATTCGATTGTTTGAAGCGCAGCAGG
TGCATCGCCAGCATTTCGATCCTCGTCAGGTGCAGGTCAGCACGTTGCTGTCG
ATTAAGACCGGAGCTTGTCCGGAAGATTGCAAATACTGCCCGCAAAGCTCGC
GCTACAAAACCGGGCTGGAAGCCGAGCGGTTGATGGAAGTTGAACAGGTGCT
GGAGTCGGCGCGCAAAGCGAAAGCGGCAGGATCGACGCGCTTCTGTATGGG
CGCGGCGTGGAAGAATCCCCACGAACGCGATATGCCGTACCTGGAACAAATG
GTGCAGGGGGTAAAAGCGATGGGGCTGGAGGCGTGTATGACGCTGGGCACG
TTGAGTGAATCTCAGGCGCAGCGCCTCGCGAACGCCGGGCTGGATTACTACA
ACCACAACCTGGACACCTCGCCGGAGTTTTACGGCAATATCATCACCACACG
CACTTATCAGGAACGCCTCGATACGCTGGAAAAAGTGCGCGATGCCGGGATC
AAAGTCTGTTCTGGCGGCATTGTGGGCTTAGGCGAAACGGTAAAAGATCGCG
CCGGATTATTGCTGCAACTGGCAAACCTGCCGACGCCGCCGGAAAGCGTGCC
AATCAACATGCTGGTGAAGGTGAAAGGCACGCCGCTTGCCGATAACGATGAT
GTCGATGCCTTTGATTTTATTCGCACCATTGCGGTCGCGCGGATCATGATGCC
AACCTCTTACGTGCGCCTTTCTGCCGGACGCGAGCAGATGA

YP3 1-8 F

ACTAGTGGATCCCCCGGGCTGCAGGAATTCGATATATCCCGTATCTGCTGGG
ATACTGGCGGGATTGACCCGACCATTGTGTATGAACGCTCGAAAAAACATGG
GCTGTTCCGGGTGATCCCCATTAAAGGGGCATCCGTCTACGGAAAGCCGGTG
GCCAGCATGCCACGTAAGCGAAACAAAAACGGGGTTTACCTTACCGAAATCG
GTACGGATACCGCGAAAGAGCAGATTTATAACCGCTTCACACTGACGCCGGA
AGGGGATGAACCGCTTCCCGGTGCCGTTCACTTCCCGAATAACCCGGATATTT
TTGATCTGACCGAAGCGCAGCAGCTGACTGCTGAAGAGCAGGTCGAAAAATG
GGTGGATGGCAGGAAAAAAATACTGTGGGACAGCAAAAAGCGACGCAATGA
GGCACTCGACTGCTTCGTTTATGCGCTGGCGGCGCTGCGCATCAGTATTTCCC
GCTGGCAGCTGGATCTCAGTGCGCTGCTGGCGAGCCTGCAGGAAGAGGATGG
TGCAGCAACCAACAAGAAAACACTGGCAGATTACGCCCGTGCCTTATCCGGA
GAGGATGAATGACGCGACAGGAAGAACTTGCCGCTGCCCGTGCGGCACTGC
ATGACCTGATGACAGGTAAACGGGTGGCAACAGTACAGAAAGACGGACGAA
GGGTGGAGTTTACGGCCACTTCCGTGTCTGACCTGAAAAAATATATTGCAGA
GCTGGAAGTGCAGACCGGCATGACACAGCGACGCAGGGGACCTGCAGGATT
TTATGTATGAAAACGCCCACCATTCCCACCCTTCTGGGGCCGGACGGCATG

YP3 1-9 F

ATCGGATTGAGTGCGAAAGCGCCTGCAATGACCCCGCTGATGCTGGACACCT
CCAGCCGTAAGCTGGTTGCGTGGGATGGCACCACCGACGGTGCTGCCGTTGG
CATTCTTGCGGTTGCTGCTGACCAGACCAGCACCACGCTGACGTTCTACAAGT
CCGGCACGTTCCGTTATGAGGATGTGCTCTGGCCGGAGGCTGCCAGCGACGA
GACGAAAAAACGGACCGCGTTTGCCGGAACGGCAATCAGCATCGTTTAACTT
TACCCTTCATCACTAAAGGCCGCCTGTGCGGCTTTTTTTACGGGATTTTTTTAT
GTCGATGTACACAACCGCCCAACTGCTGGCGGCAAATGAGCAGAAATTTAAG
TTTGATCCGCTGTTTCTGCGTCTCTTTTTCCGTGAGAGCTATCCCTTCACCACG
GAGAAAGTCTATCTCTCACAAATTCCGGGACTGGTAAACATGGCGCTGTACG
TTTCGCCGATTGTTTCCGGTGAGGTTATCCGTTCCCGTGGCGGCTCCACCTCT
GAATTTACGCCGGGATATGTCAAGCCGAAGCATGAAGTGAATCCGCAGATGA
CCCTGCGTCGCCTGCCGGATGAAGATCCGCAGAATCTGGCGGACCCGGCTTA
CCGCCGCCGTCGCATCATCATGCAGAACATGCGTGACGAAGAGCTGGCCATT
GCTCAGGTCGAAGAGATGCAGGCAGTTTCTGCCGTGCTTAAGGGCAAATACA
CCATGACCGGTGAAGCCTTCGATCCGGTTGAGGTGGATATGGGCCGCAGTGA
GGAGAATAACATCACGCAGTCCGGCGGCACGGAGTGGAGCAAGCGTGACAA
GTCCACGTATGACCCGACCGACGATATCGAAGCCTACGCGCTGAACGCCAGC
GGTGT

YP3 10 F

GCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATTGGCCTGGCGC
ACCAGACGTTACTTGGCGTGACTGGCTCAGGGAAAACCTTCACCATTGCCAA
TGTCATTGCTGACCTTCAGCGCCCAACCATGGTACTTGCGCCCAACAAAACG
CTGGCGGCCCAGCTGTATGGCGAAATGAAAGAGTTCTTCCCGGAAAACGCGG
TGGAATATTTCGTTTCCTACTACGACTACTATCAGCCGGAAGCCTATGTACCG
AGTTCCGACACTTTCATTGAGAAAGATGCCTCGGTTAACGAACATATTGAGC
AGATGCGTTTGTCCGCCACCAAAGCGATGCTGGAGCGGCGTGATGTGGTTGT
GGTGGCGTCTGTTTCCGCGATTTATGGTCTGGGCGATCCTGATTTATATCTCA
AGATGATGCTCCATCTCACGGTCGGTATGATTATCGATCAGCGCGCGATTCTG
CGCCGACTGGCGGAGCTGCAATACGCTCGTAATGATCAAGCATTCCAGCGTG
GTACTTTCCGCGTTCGTGGCGAGGTGATAGATATCTTCCCGGCAGAATCGGAT
GACATTGCACTTC

YP3 1-10 F

CGATAACAGAAAAGCCCACTGGACAGTCCGGCACTGACCGGAACGCCAACA
GCACCAACCGCGCTCAGGGGAACAAACAATACCCAGATTGCGAACACCGCTT
TTGTACTGGCCGCGATTGCAGATGTTATCGACGCGTCACCTGACGCACTGAAT
ACGCTGAATGAACTGGCCGCAGCGCTCGGGAATGATCCAGATTTTGCTACCA
CCATGACTAACGCGCTTGCGGGTAAACAACCGAAGAATGCGACACTGACGGC
GCTGGCAGGGCTTTCCACGGCGAAAAATAAATTACCGTATTTTGCGGAAAAT

GATGCCGCCAGCCTGACTGAACTGACTCAGGTTGGCAGGGATATTCTGGCAA
AAAATTCCGTTGCAGATGTTCTTGAATACCTTGGGGCCGGTGAGAATTCGGCC
TTTCCGGCAGGTGCGCCGATCCCGTGGCCATCAGATATCGTTCCGTCTGGCTA
CGTCCTGATGCAGGGGCAGGCGTTTGACAAATCAGCCTACCCAAAACTTGCT
GTCGCGTATCCATCGGGTGTGCTTCCTGATATGCGAGGCTGGACAATCAAGG
GGAAACCCGCCAGCGGTCGTGCTGTATTGTCTCAGGAACAGGATGGAATTAA
GTCGCACACCCACAGTGCCAGTGCATCCGGTACGGATTTGGGGACGAAAACC
ACATCGTCGTTTGATTACGGGACGAAAACAACAGGCAGTTTCGATTACGGCA
CCAAATCGACGAATAACACGGGGGCTCATGCTCACAGTCTGAGCGGTTCAAC
AGGGGCCGCGGGTGCTCATGCCCACACAAGTGGTTTAAGGATGAACAGTTCT
GGCTGGAGTCAGTATGGAACAGCAACCATATCAAGCTTATCGATACCGTCGA
CCTC

YP3 1-10 R

TCGACGGTATCGATAAGCTTGATGTTACATAACGCGCTGGAAAAGGCGTTAC
AGGCTTTAGGTTAACGGAAATCCGAAGGAAAATTCCGGCTTCCTATTGAAGA
CAAAGTGCGCGTTGTTTATGCCGGATGCGGCGTGAACGCCTTATCCGGCCTAC
AAACCGCGCAAATTCAATATATTGCGGAGAAAATGTAGGCCTGATAAGCGTA
GCGCATCAGGCTGTTTTCCGTTTGTCATCAGTCTTCTTCGCTATCCTGTTACGA
TGCCGCGATAAACAGCTCACGCAGCTGATGCAACTGGTCACGAATTTGCGCC
GCTTCTTCGAACTCCAGATTCTGCGCGTGTTGCATCATCAACCCTTCCAGCTC
ATGGATTTTCTGCTGCAACGCTTTAGGCGACATATCCATCGGCACATTATCCG
GCTCAACAATCGGGCGCGATTTTCCTCTGCCCTTCGCTTTGGTTTTGGCAATG
TTCTGCCCCAGCGCCAGGATATCGACCACTTTCTTGTTCAAGCCTTGCGGCGT
AATTCCGTGTTCCTCGTTGTACTTCTGCTGTTTCTCACGGCGACGTTC

YP3 1-13 F

ACTAGTGGATCCCCCGGGCTGCAGGAATTCGATGCAAAACGGCGGTGCCTGG
TGCCGCCGCAATGGCCATTAACCGCGTTGCTTCATCCGCGATATCGCAGTCGG
CGTCACAGGTTGCCCGTGAGACAAAGGTACGCCGGAAACTGGTAAAGGAAA
GGGCCAGGCTGAAAAGGGCCACGGTCAAAAATCCGCAGGCCAGAATCAAAG
TTAACCGGGGGGATTTGCCCGTAATCAAGCTGGGTAATGCGCGGGTTGTCCTT
TCGCGCCGCAGGCGTCGTAAAAAGGGGCAGCGTTCATCCCTGAAAGGTGGCG
GCAGCGTGCTTGTGGTGGGTAACCGTCGTATTCCCGGCGCGTTTATTCAGCAA
CTGAAAAATGGCCGGTGGCATGTCATGCAGCGTGTGGCTGGGAAAAACCGTT
ACCCCATTGATGTGGTGAAAATCCCGATGGCGGTGCCGCTGACCACGGCGTT
TAAACAAAATATTGAGCGGATACGGCGTGAACGTCTTCCGAAAGAGCTGGGC
TATGCGCTGCAGCATCAACTGAGGATGGTAATAAAGCGATGAAACATACTGA
ACTCCGTGCAGCCGTACTGGATGCACTGGAGAAGCATGACACCGGGGCGACG
TTTTTTGATGGTCGCCCCGCTGTTTTTGATGAGGCGGATTTTCCGGCAGTTGCC
GTTTATCTCACCGGCGCTGAATACACGGGCG

YP3 1-14 F

CTAGTGGATCCCCCGGGCTGCAGGAATTCGATAACTCCCCGTATACAGACAA
CGGATAACGGTAAGCAGAACACCTATTTCTCCTCACTGGATAACATGGTTGC
CCAGGGCAATGTTCTGCCTGTTCTGTACGGGGAAATGCGCGTGGGGTCACGC
GTGGTTTCTCAGGAGATCAGCACGGCAGACGAAGGGGACGGTGGTCAGGTTG
TGGTGATTGGTCGCTGATGCAAAATGTTTTATGTGAAACCGCCTGCGGGCGGT
TTTGTCATTTATGGAGCGTGAGGAATGGGTAAAGGAAGCAGTAAGGGGCATA
CCCCGCGCGAAGCGAAGGACAACCTGAAGTCCACGCAGTTGCTGAGTGTGAT
CGATGCCATCAGCGAAGGGCCGATTGAAGGTCCGGTGGATGGCTTAAAAAGC
GTGCTGCTGAACAGTACGCCGGTGCTGGACACTGAGGGGAATACCAACATAT
CCGGTGTCACGGTGGTGTTCCGGGCTGGTGAGCAGGAGCAGACTCCGCCGGA
GGGATTTGAATCCTCCGGCTCCGAGACGGTGCTGGGTACGGAAGTGAAATAT
GACACGCCGATCACCCGCACCATTACGTC

YP3 1-16 F

GGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATT
AAATCGCGCTTCGTGGTGTAACCGTACAGGCTCATTGCGCCGAACATCCCGG
CAGTAACGACGAAAGTACTGGCGATAGAAGCAGCGGTATAGACAATGAATA
TACTGGAAAGCGTAAGACCCGTCAGCGCCGAATAAAGCATAAAGAGCATCGT
CGTTACACCTGCGCTCAGCTTTTGAATCATCGCTGATAACACAATAACCAATG
CTAATTGCGCGATGATCAGACCGATTAAAAAGACACGGTTAGTGAACAACAG
CTCCATCACGGCCGCGGAATTAGCCGCATACCAGGCAACAAATGCGGTCAGC
AACAAGCCAACGGTCATCCAGCCATAGACTTGAGCCATATAAGTTTGCAAGC
CAGCCCGGGGTTGTACGATTGAATCAGAACGTGGGAATCTGTCCATGACGAT
CTCCTGAAGATATAAGGAATATCTTAAGGATACTGCAAAATGATGAGGCTGT
GCATCGACGCAGCGTAAAC

YP3 1-16 R

CGACGGTATCGATAAGCTTGATAGCCGCAGGTAAACTCAACTCACGCCCCAG
ATTGCCGATATAAACCATCGGCGCTGGCGTGCGGCGTAATGCCTGGGCGATT
TCCTTCAGCAGCAGAATTGGCATCAGGCTGGTATAAAAACTGCCAGGCCCAA
TAATGATGAGATCCGCTTCATTGATAGCGTGAACCGCCTCACGCGTTGCGGGT
ACATTAGGCGTTAACAATAACTCTTGAATCGGCGTAGTTAACTGGTCGATATT
GACCTCGCCGTAAACTTCATGCCCCTGATCGTCAATCGCCATCAGATCAACA
GGATGCTCTGACATTGGAATCAAATGCGTATCCACTTTCAGCAGATTACGAAT
TAAATTGATGGCTTCCAGAGGCCGCACGCTAAGGTGATCCAGCGCCTTTAAC
ATCAAGTTTCCGAGATTATGACCGGAAAGTTCGCCATTGCCACCAAAACGGT
ATTCAAACATCGCGGAGGCGACGCTCGGTTCCGTTATCAGCTGGTTGAGGCA
GTTGCGCATATCGCCCCAGGCAATGCCGCCTTCTGAACGGCGAATACGCCCC
GTCGAGCCACCATTATCGGTGGTGGTGACGATACCCGTTAAACGAGAACCCA

AAGACGAAAGTGATGAGAGAACGCGTCCCAGTCCATGCCCTCCGCCGAGAG
CAACGACACGAT

YP3 1-18 F

GAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATGGCAGAGGCTGCGTATA
AGAAAGCAGACGACATCTGGAATCTGCGCAAGGATGATTATTTTGTTAACGA
TGAAGCGCGGGCGCGTTACTGGGATGATCGTGAAAAGGCCCGTCTTGCGCTT
GAAGCCGCCCGAAAGAAGGCTGAGCAGCAGACTCAACAGGACAAAAATGCG
CAGCAGCAGAGCGATACCGAAGCGTCACGGCTGAAATATACCGAAGAGGCG
CAGAAGGCTTACGAACGGCTGCAGACGCCGCTGGAGAAATATACCGCCCGTC
AGGAAGAACTGAACAAGGCACTGAAAGACGGGAAAATCCTGCAGGCGGATT
ACAACACGCTGATGGCGGCGGCGAAAAAGGATTATGAAGCGACGCTGAAAA
AGCCGAAACAGTCCAGCGTGAAGGTGTCTGCGGGCGATCGTCAGGAAGACA
GTGCTCATGCTGCCCTGCTGACGCTTCAGGCAGAACTCCGGACGCTGGAGAA
GCATGCCGGAGCAAATGAGAAAATCAGCCAGCAGCGCCGGGATTTGTGGAA
GGCGGAGAGTCAGTTCGCGGTACTGGAGGAGGCGGCGCAACGTCGCCAGCT
GTCTGCACAGGAGAAATCCCTGCTGGCGCATAAAGATGAGACGCTGGAGTAC
AAACGCCAGCTGGCTGCAC

YP3 1-20 F

AGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATCCTGAAAACCGAATCA
TTTGAACATAACGGTGTGACCGTCACGCTTTCTGAACTGTCAGCCCTGCAGCG
CATTGAGCATCTCGCCCTGATGAAACGGCAGGCAGAACAGGCGGAGTCAGA
CAGCAACCGGAAGTTTACTGTGGAAGACGCCATCAGAACCGGCGCGTTTCTG
GTGGCGATGTCCCTGTGGCATAACCATCCGCAGAAGACGCAGATGCCGTCCA
TGAATGAAGCCGTTAAACAGATTGAGCAGGAAGTGCTTACCACCTGGCCCAC
GGAGGCAATTTCTCATGCTGAAAACGTGGTGTACCGGCTGTCTGGTATGTATG
AGTTTGTGGTGAATAATGCCCCTGAACAGACAGAGGACGCCGGGCCCGCAGA
GCCTGTTTCTGCGGGAAAGTGTTCGACGGTGAGCTGAGTTTTGCCCTGAAACT
GGCGCGTGAGATGGGGCGACCCGACTGGCGTGCCATGCTTGCCGGGATGTCA
TCCACGGAGTATGCCGACTGGCACCGCTTTTACAGTACCCATTATTTTCATGA
TGTTCTGCTGGATATGCACTTTTCCGGGCTGACGTACACC

YP3 1-25 F

GCTCCACCGCGGAGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAG
GAATTCGATTGAGAGCCTGCGTGGACGTTATGTGAGCGTGATGGCCGGACCG
GTTTTACAAATCAGTAAG

YP3 1-28 F

GCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGATCTCGCGGAGTG
CCGTGGATAAATTACCGCACCACTGCACTGCGAGATTGCTCCATGCAAGATC
GAACGTCGCAGTCGCTAACGGCAGGGATTCGATATCTCCCGCCAGATAATGG
TCTGCGGCATCCTTCTGGCGTGCCTGAACAAGCATTGGCGGCGAGAGATCTA
AGGCCGTCACCTGCGCGTGACGTTCCCGCCAGTGGCGGCTCATCCAGCCAGG
TCCACAACCCGCGTCCAGTACGTGGGTGTATTTACGCTGTGGAAGCATTGCCA
GTAAGGCGTCAGCACTCTGGCGCTGTAGATCTGCATGTTGCTCATAGTGTGCG
GCTGCCCGACCAAATGCCGCTGCAATGGCTTGTTTATTAACCGTTGCCATGCA
GCACCTCCAGCAGACGGTCGATATCCTGCATTTCATGCGCAGCGGTTAGCGTT
AAGCGCAGTCGCGCAGTACCAGCGGGTACGGTTGGCGGGCGAATCGCCGTG
ACCCAGCAGCCTTGCTGACGCAGTTTTTCTGCCAGTTGTAACGCACGGCTGTT
ATCACCGACAAATCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGG
TACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCA
TGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGC

NR-YP3 with M8

AGGGCCAGCGCAGTAGCGAGTAGCATTTTTTTCATGGTGTTATTCCCGATGCT
TTTTGAAGTTCGCAGAATCGTATGTGTAGAAAATTAAACAAACCCTAAACAA
TGAGTTGAAATTTCATATTGTTAATATTTATTAATGTATGCCAGGTGCGATGA
ATCGTCATTGTATTCCCGGATTAACTATGTCCACAGCCCTGACGGGGAACTTC
TCTGCGGGAGTGTCCGGGAATAATTAAAAACGATGCACACAGGGTTTAGCGC
GTACATGTATTGTATTATGCCAACACCCCGGTGCTGACACGGAAGAAACCGG
ACGTTATGATTTAGCGTGGAAAGATTTGTGTAGTGTTCTGAATGCTCTCAGTA
AATAGTAATGAATTATCAAAGGTATAGTAATATCTTTTATGTTCGTGGATATT
TGTAATCCATCGGAAAACTCCTGCTTTAGCAAGATTTTCCCTGTATTGCTGAA
ATGTGATTTCTCTTGATTTCAACCTATCATAGGACGTTTCTATAAGATGCGTA
TTTCTTGAGAATTTAACATTTACAACCTTTTTAAGTCCTTTTATTAACACGGTG
TTATCGTTTTCTAACACAATGTGAATATTATCTGTGG

NR-YP3 with M9

GGAGCGACAAAATGAATAAAGAACAATCTGCTGATGATCCCTCCGTGGATCT
GATTCGTGTAAAAAATATGCTTAATAGCACCATTTCTATGAGTTACCCTGATG
TTGTAATTGCATGTATAGAACATAAGGTGTCTCTGGAAGCATTCAGGGCAATT
GAGGCAGCGTTGGTGAAGCACGATAATAATATGAAGGATTATTCCCTGGTGG
TTGACTGATCACCATAACTGCTAATCATTCAAACTACTTAACCTGTGACAGAG
CCAACACGCAGTCTGTCACTGTCAGGAAAGTGGTAAAACTGCAACTCAATTA
CTGCAATGCCCTCGTAATTAAGTGAATTTACAATATCGTCCTGTTCGGAGGGA
AGAACGCGGGATGTTCATTCTTCATCACTTTTAATTGATGTATATGCTCTCTTT
TCTGACGTTAGCCTCCGACGGCAGGCTTCAATGACCCAGGCTGAGAAATTCC

CGGACCCTTTTTGCTCAAGAGCGATGTTAATTTGTTCAATCATTTGGTTAGGA
AAGCGGATGTTGC

NR-YP3 with M9 (2)

AGCGACAAAATGAATAAAGAACAATCTGCTGATGATCCCTCCGTGGATCTGA
TTCGTGTAAAAAATATGCTTAATAGCACCATTTCTATGAGTTACCCTGATGTT
GTAATTGCATGTATAGAACATAAGGTGTCTCTGGAAGCATTCAGGGCAATTG
AGGCAGCGTTGGTGAAGCACGATAATAATATGAAGGATTATTCCCTGGTGGT
TGACTGATCACCATAACTGCTAATCATTCAAACTACTTAACCTGTGACAGAGC
CAACACGCAGTCTGTCACTGTCAGGAAAGTGGTAAAACTGCAACTCAATTAC
TGCAATGCCCTCGTAATTAAGTGAATTTACAATATCGTCCTGTTCGGAGGGAA
GAACGCGGGATGTTCATTCTTCATCACTTTTAATTGATGTATATGCTCTCTTTT
CTGACGTTAGCCTCCGACGGCAGGCTTCAATGACCCAGGCTGAGAAATTCCC
GGACCCTTTTTGCTCAAGAGCGATGTTAATTTGTTCAATCATTTGGTTAGGAA
AGCGGATGTTGCGGGTTGTTGTTCTGCGGGTTCTGTTCTTAGTTGACATGAGG
TTGCCCCGTATTCAGTGTCGCTGATTTGTATTGTCTGAAGTTGTTTTTACGTTA
AGTTGATGCAGATCAATTAATACGATACCTGCGTCATAATTGATTATTTGACG
TGGTTTGATGGCGTAGATGCACGTTGTGACATGTAGATGATAATTATTATCAT
TTTGTGGGTCCTTTCCGGCGATCCGACAGGTTACGGGGCGGCGACCTCGCGG
GTTTTCGCTATTTA

NR-YP3 with M17

TTGTCGAAGATATTCTCCAGTTTGAGCTGGAAAACCTGCCTGGACAGAGCGC
GGCACGACGCTGGTGGCGACGTCATCACAAAGCGGAAGAGAACCGCCAGCC
GGGAGAAGCGCAGGTATTGCTGGTCTGGCGCGATAACGAAGAACATCGCGAT
GATATTGAACGCCATTATTTGAAAATGCTCACTCAGGCGCGGCGGGAAGTGA
TTATCGCCAACGCCTACTTCTTCCCCGGCTATCGATTTTTACACGCCTTGCGTA
AAGCGGCACGGCGCGGGGTGCGGATCAAACTGATCATTCAGGGCGAACCGG
ATATGCCGATTGTCAGAGTCGGTGCGCGCTTGCTGTATAACTATCTGGTTAAA
GGCGGCGTTCAGGTTTTTGAGTACCGCCGCCGCCCGCTCCACGGCAAAGTGG
CATTGATGGACGATCACTGGGCGACAGTAGGGTCCAGTAATCTCGATCCGCT
CAGTTTGTCACTGAATCTCGAAGCAAATGTCATCATCCACGATCGTCATTTTA
ACCAGACGCTGCGCGATAATCTGAACGGCATTATTGCCGCAGATTGTCAGCA
GGTGGATGAAACCATGCTGCCCAAACGCACCTGGTGGAACCTGACCAAAAGC
GTGCTGGCGTTCCACTTTTTACGCCACTTCCCGGCGCTGGTTGGCTGGCTTCC
GGCACACACGCCACGTCTGGCGCAGGTTGATCCGCCCGCACAACCGACAATG
GAAACGCAGGATCGGGTAGAAACTGAAAACACGGGGGTAAAACCCTGATGA
GTAAATCACACCCGCGCTGGCGCTTAGCAAAGAAGATCCTCACCTGGCTGTT
TTTTATCGCGGTGAT

NR-YP3 with M18

GATATTGTTTGCCTGCAGGAAGTGATGGGCGCGCACGAAGTTCATCCGCTGC
ATGTGGAAAACTGGCCCGATACCTCGCACTACGAGTTTCTCGCCGACACTAT
GTGGAGCGATTTTGCCTACGGTCGCAATGCCGTATACCCGGAAGGGCATCAC
GGCAACGCCGTACTGTCGCGTTATCCCATTGAACATTATGAGAATCGCGATGT
TTCGGTCGATGGTGCGGAAAAGCGCGGCGTGCTCTACTGCCGCATTGTGCCG
CCGATGACCGGAAAAGCGATTCATGTGATGTGCGTACATCTGGGCCTGCGTG
AGGCGCACCGTCAGGCGCAGCTTGCGATGCTCGCCGAATGGGTGAATGAGCT
ACCGGACGGCGAACCGGTATTGGTGGCGGGTGATTTCAATGACTGGCGGCAA
AAAGCTAATCATCCGTTAAAAGTGCAGGCCGGACTGGATGAGATTTTTACCC
GCGCCCACGGACGCCCGGCGCGCACGTTTCCGGTGCAATTTCCTCTACTACG
ACTGGACAGGATCTACGTCAAAAATGCCAGCGCCAGCGCGCCAACCGCGTTG
CCGCTGCGGACATGGCGACACCTTTCTGATCATGCCCCTTTAAGTGCGGAGAT
TCATTTATGAAATGTAGCTGGCGCGAAGGCAATAAGATCCAGTTGCTGGAAA
ACGGCGAGCAATATTATCCCGCGGTGTTTAAGGCGATTGGCGAGGCACAAGA
ACGCATCATTCTTGAAACGTTTATCTGGTTTGAGGATGACGTCGGCAAACAAC
TGCATGCGGCACTACTGGCAGCAGCGCAACGCGGGGTTAAAGCGGAAGTCTT
GCTGGATGGCTACGGTTCGCCGGATCTCAGCGATGAGTTTGTCAATGAACTG
ACGGCAGCTGGCGTAGTGTTCCGCTACTACGATCCCCGCCCTCGC

NR-YP3 with M19

CTGGAAGGAAAACCGAAACCTAACAACAGCGAGCTGGTGGTACGCAACTGG
TATAACCCGAATCTCGACTACAAATGGTTTGTGGTGCCGTCACTGATCGCCAT
GATCACCACTATCGGCGTAATGATCGTCACTTCACTTTCCGTCGCCCGCGAAC
GTGAACAAGGTACGCTCGATCAGCTACTGGTTTCGCCGCTCACCACCTGGCA
GATCTTCATCGGCAAAGCCGTACCGGCGTTAATTGTCGCCACCTTCCAGGCCA
CCATTGTGCTGGCGATTGGTATCTGGGCGTATCAAATCCCCTTCGCCGGATCG
CTGGCGCTGTTCTACTTTACGATGGTGATTTATGGTTTATCGCTGGTGGGATT
CGGTCTGTTGATTTCATCACTCTGTTCAACACAACAGCAGGCGTTTATCGGCG
TGTTTGTCTTTATGATGCCCGCCATTCTCCTTTCCGGTTACGTTTCTCCGGTGG
AAAACATGCCGGTATGGCTGCAAAACCTGACGTGGATTAACCCTATTCGCCA
CTTTACGGACATTACCAAGCAGATTTATTTGAAGGATGCGAGTCTGGATATTG
TGTGGAATAGTTTGTGGCCGCTACTGGTGATAACGGCCACGACAGGGTCAGC
GGCGTACGCGATGTTTAGACGTAAGGTGATGTAACTTCTTATCTTTCGCCAGC
AAAGACACTACCGCCGGGCCGGCAAGGATTGCCAGCCCTGCAATTGCCAGCA
AAAAGTTGTTTTCCAGAACGCGGGGACAACAGCCACAGCACAATCAGCGCTG
CGCCAAAATACCAGGTCGTGGTCAGCTCTTCCAGCACGTCTGCCACCTCACG
CCAGCG

Appendix 5

BLASTp analysis of Enterobacteria phage DE3 (GenBank: EU078592.1) and *E. coli* str. K-12 substr. MG1655 (GenBank: U00096.3).

Gene products and functions encoded in Enterobacter phage DE3

Gene Name	Coding Region	Gene Product	Note
	18037..18330	Bor protein precursor	similar to lambdap77
	18620..19153	putative envelope protein	similar to lambdap78
	19316..19522	hypothetical protein	similar to lambdap79
	20081..20092		cos site
nu1	20270..20815	DNA packaging protein	similar to lambdap01
A	20790..22715	DNA packaging protein	similar to lambdap02
W	22712..22918	head-tail joining protein	similar to lambdap03
B	22915..24516	capsid component	similar to lambdap04
C	24497..25816	capsid component	similar to lambdap05
D	25826..26158	head-DNA stabilization protein	similar to lambdap07
E	26214..27239	capsid component	similar to lambdap08
Fi	27281..27679	DNA packaging protein	similar to lambdap09
Fii	27691..28044	head-tail joining protein	similar to lambdap10
Z	28056..28634	tail component	similar to lambdap11
U	28631..29026	tail component	similar to lambdap12
V	29034..29774	tail component	similar to lambdap13
G	29790..30212	tail component	similar to lambdap14
T	30194..30628	tail component	similar to lambdap15
H	30621..33182	tail component	similar to lambdap16
M	33179..33508	tail component	similar to lambdap17
L	33508..34206	tail component	similar to lambdap18
K	34356..34955	tail component	similar to lambdap19
I	34853..35524	tail component	similar to lambdap20
J	35585..38983	tail:host specificity protein	similar to lambdap21
lom	39045..39665	outer host membrane	similar to lambdap26
	39730..40935	tail fiber protein	orf-401; similar to lambdap27
	41109..41372	tail fiber	orf-314; similar to lambdap28; C-terminal extended due to deletion

Gene products and functions encoded in *E. coli* K-12 MG1655

Gene Name	Coding Region	Gene Product	Note
ybhB	807433..807909	kinase inhibitor homolog, UPF0098 family	
bioA	807968..809257	7,8-diaminopelargonic acid synthase, PLP-dependent	enzyme; Biosynthesis of cofactors, carriers:Biotin
bioB	809344..810384	biotin synthase	enzyme; Biosynthesis of cofactors, carriers: Biotin
bioF	810381..811535	8-amino-7-oxononanoate synthase	enzyme; Biosynthesis of cofactors, carriers: Biotin
bioC	811522..812277	malonyl-ACP O-methyltransferase, SAM-dependent	enzyme; Biosynthesis of cofactors, carriers: Biotin
bioD	812270..812947	dethiobiotin synthetase	enzyme; Biosynthesis of cofactors, carriers: Biotin
uvrB	813526..815547	excinulease of nucleotide excision repair, DNA damage recognition component	enzyme; Degradation of DNA
ybhK	815739..816647	putative NAD(P)-binding transferase	putative structure; Not classified
moaA	817044..818033	molybdopterin biosynthesis protein A	enzyme; Biosynthesis of cofactors, carriers: Molybdopterin
moaB	818055..818567	inactive molybdopterin adenylyltransferase	enzyme; Biosynthesis of cofactors, carriers: Molybdopterin
moaC	818570..819055	molybdopterin biosynthesis, protein C	enzyme; Biosynthesis of cofactors, carriers: Molybdopterin
moaD	819048..819293	molybdopterin synthase, small subunit	enzyme; Biosynthesis of cofactors, carriers: Molybdopterin
moaE	819295..819747	molybdopterin synthase, large subunit	enzyme; Biosynthesis of cofactors, carriers: Molybdopterin
ybhL	819884..820588	UPF0005 family inner membrane protein	
ybhM	820793..821506	UPF0005 family inner membrane protein	
ybhN	821542..822498	UPF0104 family inner membrane protein	
clsB	822498..823739	cardiolipin synthase 2	Enzyme; Macromolecule synthesis: Phospholipids
ybhP	823736..824497	endo/exonuclease/phosphatase family protein	
ybhQ	824630..825040	inner membrane protein	
ybhR	825002..826108	inner membrane putative ABC superfamily transporter permease	

Appendix 1

Primers used for PCR in this study

Name of Primer	Sequence (5′ − 3′)	Base Sequence	Reaction Condition
YP3 1-3F PCR-R (M1)	5′GGCGGATGACGCAGGCATTATGCT-3′	YP3 1-3	30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/2 min
YP3 1-16F PCR-R (M2)	5′CGGGTCTTACGCTTTCCAGTATAT-3′	YP3 1-16	30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/2 min
zdt-9n2 184-161 PCR (M3)	5′CCCGGCGCAGCTTTTCGTTCTCAA-3′	zdt-9n2	30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/2 min
zdt-9n2 36408-36431 PCR (M4)	5′GCGTCAGGTTGAAGGCGTAGCAGA-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/2 min
phage DE3 14857-14880 F (M5)	5′AACGGGAAGGAAAGATGAGCACGA-3′	Enterobacteria phage DE3	30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/2 min
phage DE3 16107-16130 F (M6)	5′GGCAACATATTAACGGCATGATAT-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/2 min
phage DE3 17115-17138 F (M7)	5′GGCGTGGTCGGAGGGAACTGATAA-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/2 min
phage DE3 18217-18240 (M8)	5′GGTGATGGTTTCCTTTGGTGCTAC-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/2 min
phage DE3 19237-19260 F (M9)	5′CGTGCTCAAATCTTCATACAGAAA-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/2 min

phage DE3 20276-20253 R (M10)	5′CTTCCATTGTTCATTCCACGGACA-3′	Enterobacteria phage DE3	30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
phage DE3 20460-20437 R (M11)	5′CGGCGCAGCTTTTCGTTCTCAATT-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
pCC2FOS T7 modified F (M12)	5′TAATACGACTCACTATAGGGCGAA-3′	pCC2FOS	30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
pCC2FOS R (476-453) (M13)	5′CGCCAAGCTATTTAGGTGAGACTA-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
pCC1Fos R (580-560) (M14)	5′GCGGGCAGTGAGCGCCAACGCA-3′	pCC1FOS	30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
MG1655 820161-820183 (M15)	5′CGGCGCTGACGGGTCTTACGCTT-3′	*E. coli* str. K-12 substr. MG1655	30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
MG1655 822530-822551 F (M16)	5′CCCGATCCTGCGTTTCCATTGTCG-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
MG1655 823360-823337 (M17)	5′GCGGGCTGAATTACTCCGCCGAGC-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
MG1655 824440-824417 (M18)	5′GGCTTTACCGCGTTTAACCGACGC-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
MG1655 825750-825727 (M19)	5′TCCTCGACGGGCGTAACTCCAACA-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
E.coli MG1655 M18F (M20)	5′GCGTCGGTTAAACGCGGTAAAGCC-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min

E.coli MG1655 M19F (M21)	5′TGTTGGAGTTACGCCCGTC GAGGA-3′	*E. coli* str. K-12 substr. MG1655	30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
pCC2FOS-F 261-284 (M22)	5′CGCCAGGGTTTTCCCAGTC ACGAC-3′	pCC2FOS	30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
pCC2FOS-R 520-497 (M23)	5′AGCGGATAACAATTTCACA CAGGA-3′		30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
#88 Modified Revers (M24)	5′GGAAACAGCTATGACCATG ATTACG-3′	pBluescript	30 cycles; 94˚C/1 min, 52˚C/1 min, 72˚C/3 min
#89 MODIFIED M13-20 (M25)	5′TTGTAAAACGACGGCCAGT GAATTG-3′		30 cycles; 94˚C, 52˚C/1 min, 72˚C/3 min
#4 pSK-M13minus20 (M26)	5′GTAAAACGACGGCCAGTGA ATTG-3′		30 cycles; 94˚C, 52˚C/1 min, 72˚C/2 min
#3 pSK-Reverse (M27)	5′GGAAACAGCTATGACCA TGATTAC-3′		30 cycles; 94˚C, 52˚C/1 min, 72˚C/2 min

PERMISSION TO COPY STATEMENT

I, Younshim Park, hereby submit this thesis/report to Emporia State University as partial fulfillment of the requirements for an advanced degree. I agree that the Library of the University may make it available to use in accordance with its regulations governing materials of this type. I further agree that quoting, photocopying, digitizing or other reproduction of this document is allowed for private study, scholarship (including teaching) and research purposes of a nonprofit nature. No copying which involves potential financial gain will be allowed without written permission of the author. I also agree to permit the Graduate School at Emporia State University to digitize and place this thesis in the ESU institutional repository.

_____

Signature of Author

_____

Date

_____

Title of Thesis

_____

Signature of Graduate Office Staff

_____

Date Received