Discovery and characterization of antibacterial compounds expressed by soil microorganisms using culture-dependent and -independent approaches

by

Shamima Nasrin

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy
Auburn, Alabama
August 01, 2015

Keywords:
MRSA, Metagenomics, Antibiotics, Culture-dependent,
Culture-independent, Actinobacteria

Approved by

Mark R. Liles, Chair, Associate Professor of Biological Sciences
Peter Panizzi, Assistant Professor of Department of Drug Discovery and Development
Paul A. Cobine, Associate Professor of Biological Sciences
Elizabeth H. Schwartz, Assistant Professor of Biological Sciences

Abstract

The emergence of multidrug-resistant pathogens has increased the demand for discovery of novel antibiotics. Both culture-dependent and culture-independent approaches were used to discover antibiotics against methicillin-resistant clinical isolates of *Staphylococcus aureus* (MRSA). A collection of 548 bacterial and fungal isolates were isolated from soil using low-strength (1/200th) nutrient agar supplemented with soil extract incubated for more than three months. Two isolates, designated as A115 and F4, were found to inhibit the growth of pathogenic MRSA strains. The isolate A115, member of the genus *Streptomyces*, produces pink pigments after extended incubation. The isolate F4, identified as *Nonomuraea*, produces a high molecular weight (>100kDa), heat-stable reddish pigment with anti-MRSA activity. Whole genome sequencing using a combination of shotgun and mate-pair next-generation sequencing resulted in the complete assembled genome for each isolate, with the size of the A115 and F4 genomes at 8.6 Mbp and 10.3 Mbp, respectively. The %G+C contents of strains A115 and F4 were determined to be 71% and 70.4%, respectively. Phylogenetic analysis using six housekeeping genes revealed that strain A115 was most closely related to *Streptomyces afghaniensis* and *Streptomyces olindensis*; however, the low level of average nucleotide identity (ANI) values in comparing the A115 genome were 89.76% and 89.14% for *S. afghaniensis* and *S. olindensis*, respectively. These genomic results, combined with differentiation of strain A115 from other *Streptomyces* species by morphological and physiological characteristics, led to the conclusion that strain A115 is a novel species of the genus *Streptomyces*, for which the name

Streptomyces alburnustigris sp. nov. is proposed. *In silico* analysis using anti-SMASH predicts that A115 and F4 genomes encode many genetic clusters for secondary metabolite biosynthesis, including the synthesis of terpene, aminoglycoside, thiopeptide, bacteriocin, oligosaccharide, phenazine, butyrolactone, siderophore, melanine and potentially other bioactive compounds produced by non-ribosomal peptide synthetase and polyketide synthetase pathways. Both the genomes of *S. alburnustigris* A115 and *Nonomuraea* spp. strain F4 are predicted to encode Type I, II, and III PKS pathways. A large collection of plant growth-promoting rhizobacteria (PGPR) (n=147) isolates were screened for anti-MRSA activity, among which five *Bacillus* strains were identified with anti-MRSA activity. One of these five, *B. amyloliquefaciens* strain AP183, was found to produce a novel macrodiolide compound described herein as bacillusin A with potent anti-MRSA activity of a minimum inhibitory concentration of 0.6 µg/mL. Based on itsnovel biochemistry and strong *in vivo* anti-MRSA activity, strain AP183 was selected for evaluation as a skin probiotic to prevent MRSA infection using a mouse wound model. *In vivo* studies showed that co-administration of secondary metabolites and AP183 spores resulted in a significant reduction in the number of *S. aureus* that colonized mouse skin compared to a negative control. Analysis of 16S rRNA genes PCR amplified from skin samples revealed a significant reduction in the relative abundance of *S. aureus* after AP183 application while the relative abundance of other bacterial taxa increased in the skin microbiome as a result of probiotic administration. Using a culture-independent approach to identify antibacterial compounds, a large-insert soil metagenomic library was constructed that comprised of 19,200 *E. coli* clones with an average insert size of 110 kb. The library clones were screened for anti-MRSA activity using a 96-well microtiter plate. *In situ* lysis of the *E. coli* host enabled detection of both intra- and extracellular compounds, yielding a total of 28 clones that consistently inhibited MRSA growth.

Transformation of naïve *E. coli* with BAC DNA isolated from anti-MRSA clones confirmed the presence of their anti-MRSA activity. Seven of the clones were capable of modifying chloramphenicol added to the *E. coli* culture medium, thereby resulting in modification of an existing antimicrobial scaffold. LC-MS analysis of the organic extract of the clones revealed three new chloramphenicol derivatives. Chemical synthesis of these derivatives showed antimicrobial activities against diverse group of pathogens including MRSA, *Mycobacterium intracellulare* and *M. tuberculosis*. Together with all these results demonstrate that both culture-dependent and –independent approaches can be used to identify previously undescribed bioactive compounds with antimicrobial activity that can be used to control multidrug-resistant pathogens.

Acknowledgments

At this time, I would like to give all honor and thanks to God for the many blessings that he has bestowed upon me, in addition to the endurance and strength given to complete my graduate school journey. I would also like to whole-heartedly thank my loving parents and husband for all of the effort, time and money that they invested in me. Without them, my journey definitely would have been cut short a long time ago. The many times I was indecisive, they were always there to either offer words of wisdom or countless words of encouragement.

I express my sincere and heartfelt gratitude to my major advisor Dr. Mark R. Liles. His timeless and well-tested patience, encouragement and never-ending support have helped me to reach new-found heights as well as expand both my research and general knowledge. Because of Dr. Liles' tremendous efforts, I have learned the true meaning of the word "independent thinking" that will definitely help me to build up my career as a researcher. I would like to gratefully thank my advisory committee member Dr. Peter Panizzi of Auburn University Harrison School of Pharmacy, who consistently guided me in conducting my research using mouse model and in vivo imaging. Very special thanks go to my advisory committee member Dr. Paul A Cobine for his valuable advice and help in completion of this journey. I would also send appreciation to Dr. Elizabeth H. Schwartz and Dr. Joseph W. Kloepper for their much needed assistance in addition to their participation as members of my dissertation committee.

I would also like to thank members of Dr. Cobine's lab, Dr. Panizzi's lab and Dr. Kloepper's lab, who helped many ways to conduct my experiments and make my time in the Auburn more fun and interesting.

Special thanks go to my colleagues in Dr. Liles lab, Nancy Capps, Ran Chao, Malachi Williams, Dr. Abel Carrias, Dr. Kavida Kakirde, Dr. Molli Newman, Dr. Dawei Sun, Charles Thurlow, Jinglie Zhou, Alinne Pereira, Cody Rasmussen-Ivey, Alysa Moore and Erin Schmale for their constant support and encouragement.

I am very thankful to my brother-in-law Mohammad Kamrul Hassan for his inspiration and help. I am also very thankful to my Bangladeshi friends here at Auburn who made my life enjoyable in abroad.

I would like to express my sincere love and utmost gratitude to my husband Mohammad Jahangir Hossain for all the support he has given me to continue my graduate school journey. Without his support and encouragement I might not be able to cross this long path. I would like to express my unconditional love to my daughter Sarina J. Naureen. Her smiley face always cheers me up for conducting my experiments over and over.

Table of Contents

List of Tables

List of Figures

Chapter 2

Chapter 4

Chapter 1

Literature review

1.1 **Overview of multi-drug resistant pathogens**

Bacterial pathogens that are resistant to multiple drugs represent a growing public health threat as there are fewer, or even sometimes no, effective antimicrobial agents available for treating infections caused by these bacteria. Infectious diseases caused by multi-drug resistant (MDR) bacteria are a significant global public health concern. In the United States, the mortality rate due to infectious diseases was significantly low only 59 deaths per 100,000 (Armstrong et al. 1999) in 1996 but is a significant healthcare burden with 4.5 million hospital days costing $96 billion in 2008. Smolinski et al. (Smolinski et al. 2003) and Morens et al. (Morens et al. 2010) reported the increased risk of emerging infections in the USA and highlighted the nation's crumbling public health infrastructure, and called for substantial improvements in the USA's capacity to address these mounting challenges.

Six species of bacteria are responsible for two thirds of all health care-associated infections (HAI) in the USA (Bandeira et al. 2014) including *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella* species, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*,

and *Enterobacter* sp. (also known as the ESKAPE pathogens) (Bandeira et al. 2014). Strains from these pathogens were found to be drug resistance for all known drug classes and poses a significant challenge for anti-infective therapy (Pendleton et al. 2013; Boucher et al. 2009). In addition, most of the ESKAPE bacteria have the ability to form biofilms which can worsen the situation by making these bacteria to persist in hospital units and to increase antibiotic resistance (Bales et al. 2013).

Currently, the most notorious bacterial pathogen is the Gram-positive organism *Staphylococcus aureus*. *S. aureus* is a facultative anaerobe and present as a nasal commensal in 30% of the population (Fair et al. 2014). Though traditionally opportunistic, in recent years, many *S. aureus* strains have evolved as aggressively pathogenic,causing the major nosocomial infections (Davies et al. 2010). National Nosocomial Infection Surveillance (NNIS) System data indicate a steady increase in the incidence of nosocomial infections caused by methicillin-resistant *S. aureus* (MRSA) among intensive care units (ICUs) patients over time and now accounts for over 60% of *S. aureus* isolates in US hospital ICUs (NNIS report 2004). *S. aureus* was also identified as one of the most frequent nosocomial pathogens in European ICUs (Vincent et al. 1995). MRSA is associated with significant morbidity and mortality and is the causative agent for both the hospital care and community associated infections (Klein, Smith et al. 2007). In the US, hospital acquired infections are the fourth cause of death, taking 100,000 lives and adding $30 billion to hospital costs per year (McCaughey 2006; Hassan, Tuckman et al. 2010). *S. aureus* infections alone cost $14.5 billion in 2003 (Noskin, Rubin et al. 2007) and took approximately 19,000 lives in 2005 (Klevens R and et al. 2007). The estimated number of deaths due to MRSA infections exceeds that due to HIV/AIDS (Klevens R and et al. 2007).

1.2 **General features of *S. aureus***

S. aureus is a member of the phylum *Firmicutes,* belongs to the family Staphylococcaceae and appears as cocci in clusters under microscopy. *S. aureus* can be distinguished from other staphylococcal species on the basis of the golden pigmentation of colonies and biochemical characteristics (i.e. positive for coagulase, mannitol-fermentation, and deoxyribonuclease) (Wilkinson, 1997). *S. aureus* genome consists of a circular chromosome with approximately 2.8 Mbp nucleotides (Davis et al. 2013). Its genome contains prophages, transposons and plasmids which play an important role in virulence and antibiotic resistance mechanisms.

1.3 **Disease caused by *S. aureus***

S. aureus is the most commonly isolated bacterial pathogen of humans, causing infections which can be divided into three categories: a) superficial infections such as mild skin and soft tissue infections (Vincent et al. 2008), wound lesions, b) toxinoses such as toxic shock syndrome and food poisoning, and c) life-threatening systemic infections, including pneumonia (Rubinstein et al. 2008), endocarditis (Panizzi et al. 2011), bacteremia, sepsis, metastatic infections in distal part of the body, and osteomyelitis (Lowy, 1998). *S. aureus* can also cause diseases in several other animals including reptiles, birds and non-human mammals (van Leeuwe et al. 2005), particularly, dogs, cows, horses, goats, sheep and camel (Sung et al. 2008). *S. aureus* is the most common cause of mastitis in cattle and comparative genomic of *S. aureus* genomes revealed that the animal-associated *S. aureus* strains clustered into ten lineages which are unique to animals (Sung et al. 2008).

1.4 **Virulence factors produced by *S. aureus***

Pathogenic *S. aureus* produces a plethora of virulence factors for attachment of bacterial cells to the host cell, colonization in the host, evading host immune defense, persistence and penetration in host tissue. Virulence factors produced by *S. aureus* include surface-associated adhesins, extracellular enzymes and exo-toxins (Lowy, 1998). It is difficult to sort out the role of individual virulence factor in the pathogenic process of staphylococcal infection because of the functional redundancy of *S. aureus* virulence factors and multifactorial nature of *S. aureus* infections (Novick, 2000).

1.4.1 **Cell surface-associated virulence factors**

1.4.1.1. **Adhesins**

Proteinaceous Staphylococcal Adhesins

Initial infection begins when cells of *S. aureus* adhere to the components of extracellular matrix or host tissue or to abiotic surface such as medical devises and the adherence is mediated by proteinaceous surface adhesins known as microbial surface components recognizing adhesive matrix molecules family (MSCRAMMs) and non-proteinaceous adhesins (Lowy, 1998). Among all the members of MSCRAMMs family, collagen-binding protein (Can), fibrinogen-binding clumping factor A and B (ClfA and ClfB), and fibronectin binding protein A and B (FnBPA and FnBPB) are the most important adhesins of *S. aureus* (Jonsson et al. 1991; Patti et al. 1992; Menzies, 2003). MSCRAMMs family proteins are one of the major classes of *S. aureus* adhesins that covalently linked to the cell wall peptidoglycan via the threonine residue in the signal motif at their C-terminus (Foster et al. 1998; Speziale et al. 2009; Marraffini et al. 2006). These adhesin proteins specifically mediate attachment of bacteria to the plasma or extracellular matrix (ECM) components including fibronectin (Fn), fibrinogen (Fg), vitronectin (Vn),

thrombospondin, bone-sialoprotein, elastin, collagen, and von Willebrand factor which facilities direct adherence to host tissue (Flock et al. 1987; Cheung et al. 2002).

In addition to bacterial cell wall anchored adhesins, several non-covalently linked surface associated proteins also play important role during attachment of *S. aureus* in host tissue, such as autolysin (Hirschhausen et al. 2010), secretable expanded repertoire adhesive molecule SERAMs (Chavakis et al. 2005) and membrane-spanning proteins (Clarke et al. 2006). These non-covalently bound proteins are associated with the surface by ionic, hydrophobic or so far unknown interactions. In case of autolysin, AtlE which was first identified as *S. epidermidis* surface associated component, later found homology to *S. aureus* autolysin AtlE has an important role in attachment to polystyrene, biofilm formation and adherence to Vn (Heilmann et al. 1997; Biswas et al. 2006). Hirschhausen et al. demonstrated that Atl/AtlE can also binds to Fg, Fn, Vn and human endothelial cells and functioning staphylococcal internalization by endothelial cells (Hirschhausen et al. 2010). The Atl- or AtlE-mediated internalization mechanism is the sole mechanism involved in the internalization of coagulase-negative staphylococci. However, in coagulase-positive, *S. aureus* it might be a backup internalization mechanism (Hirschhausen et al. 2010).

Another example of non-covalently linked surface-associated proteins are SERAMs that include fibrinogen binding protein A (FbpA), coagulase (Coa), von Willebrand factor binding protein, extracellular fibrinogen binding protein (Efb), extracellular matrix binding protein (Emp) and extracellular adhesive protein (Eap) both have broad binding spectrum to host ligands and play an important role in endovascular infection. Among all the SERAMs, Eap and Emp are the two main molecules that bind to the components of the extracelluar matrix and mediate attachment and colonization of *S. aureus* cells on host tissue. The role of these proteins was

confirmed in deletion mutants with loss of *eap* and *emp* genes showed reduced ability to colonize and invade host tissue (Chavakis et al. 2005).

Other secretable proteins such as *S. aureus* coagulase (Coa), fibrinogen binding protein A (FbpA) and von Willebrand factor binding protein have homologous domain to Coa and play significant role in bacterial pathogenesis of endovascular infections. Panizzi et al. demonstrated the role of staphylocoagulase as a virulence factor in endocarditis and present evidence for its regulation by bacterial quorum sensing mechanism (Panizzi et al. 2011). The Efb protein interacts with the α chain of fibrinogen thus inhibiting aggregation of platelets which impaired fibrin formation, resulting disruption of wound healing in experimental wound model (Palma et al. 2001; Shannon et al. 2004). Literature showed that Efb not only bind to fibrinogen but also involved in counter-acting host defense by binding to complement component C3b, thus inhibiting complement mediated phagocytosis of invading bacterial cells (Lee et al. 2004; Lee et al. 2004; Lambris et al. 2008).

Additional non-covalently anchored cell surface proteins are extracellular matrix-binding protein homologue (Ebh) for *S. aureus* and extracellular matrix-binding protein homologue (Embp) for *S. epidermidis* (Clarke et al. 2006; Williams et al. 2002). The proteins Ebh and Embp bind with Fn; however, binding sites of Ebh and Embp are not similar and they were encoded from the largest genetic regions of *S. aureus* and *S. epidermidis* genomes, respectively. Further non-covalently anchored surface protein is elastin-binding protein of *S. aureus* (EbpS), is an integral membrane that binds with a major component of the extracellular matrix, elastin (Downer et al. 2002).

Non-proteinaceous Staphylococcal Adhesins

Polysaccharide intercellular adhesin (PIA) was first identified in *S. epidermidis*, and later it was also found in *S. aureus* (Mack et al. 1996; Cramton et al. 1999) with a clear role in biofilm formation. PIA is encoded by the *icaADBC* operon and is present in most of the *S. aureus* strains (Rohde et al. 2001). The role of PIA as a virulence factor has been confirmed in *S. epidermidis* mediated foreign-body infection model (Gotz, 2002; Heilmann et al. 2010); however, its role in *S. aureus* virulence was not clearly demonstrated, and conflicting results were later reported (Kristina et al. 2004).

Wall teichoic acids (WTA) and lipoteichoic acids (LPA) are non-proteinaceous highly charged cell wall polymers that have role in *S. aureus* colonization, infection, and immune evasion (Brown et al, 2013; Xia et al, 2010). The WTA is covalently linked to the peptidoglycan, whereas the lipoteichoic acid is attached to the cytoplasmic membrane via glycolipid. Biosynthesis of lipoteichoic acid is catalyzed by the enzyme glycolipid synthase, YpfP, and mutation in *ypfP* markedly decreased the production of lipoteichoic acid resulting in reduced ability to form biofilm on a polystyrene surface (Fedtke at al. 2007).

1.4.1.2 **Capsular polysaccharides**

Karakawa et al (Karakawa et al. 1988) found that most of the *S. aureus* strains were encapsulated, and a quarter of recovered human isolates belonged to the serotype CP5 and CP8 (Arbeit et al. 1984). Encapsulated bacteria are highly resistant to phagocytosis as compared to their counterpart, resulting in bacterial persistence in the bloodstream of the infected host (Karakawa et al. 1988; O'Riordan et al. 2004). Further roles of capsular adherence to endothelial surfaces resulting in involvement in colonization and persistence on mucosal surfaces were confirmed by several *in vitro* and *in vivo* studies, respectively (O'Riordan et al. 2004; Nanra et al. 2013).

1.4.2 **Secreted virulence factors**

The secreted virulence factors of *S. aureus* are usually produced at the stationary phase of growth and are involved in detoxifying various innate immune mechanisms (Schlievert et al. 2010). They also make nutrients available to the bacteria, thus promoting growth inside the host. A number of secreted factors are produced by *S. aureus* including various exoenzymes, exotoxins, superantigens, staphylococcal enterotoxins (SE) like proteins, toxic shock syndrome toxin-1(TSST-1) and exfoliative toxins A and B (Schlievert et al. 2010).

1.4.2.1 **Exoenzymes**

Various exoenzymes are produced by *S. aureus* such as proteases, lipase, nucleases, hyaluronidase and aureolysin (Costa et al. 2013). Exoenzymes of *S. aureus* are mainly involved in destruction of connective tissue, making nutrients available for the bacteria, facilitating spread of infection by detaching bacterial cells from initial colonization site, inhibiting the activity of clotting and kinin systems, promoting bacterial growth in host fatty acids and directly interacting with host immune cells via inactivation of PMN's mediated host defense (McAleese et al. 2001; Shaw et al. 2004; Kantyka et al. 2011).

1.4.2.2 **Superantigens**

Staphylococcal superantigens are a family of potent immunostimulatory enterotoxin that play major role in invasive infections including toxic shock syndrome (TSS), food poisoning, atopic dermatitis (AD), Kawasaki disease (KD), and chronic rhinosinusitis (CRS) (White et al. 1989; Xu et al. 2012). The superantigens of *S. aureus* include staphylococcal enterotoxins (SEs; A, B, C, D, E, G, H, I, and R), SE-like proteins (SEls; J, K, L, M, N, N, O, P, Q, S, U, V, and X), TSST-1 and exfoliative toxins A and B (Schlievert et al. 2010; Xu et al. 2012). Most of the clinical *S. aureus* strains possess at least one superantigen encoding gene in their genome which

functions to stimulate T cells (Barsumian et al. 1978; Schlievert et al. 1981) and macrophages (Marrack et al. 1990) to produce massive amounts of cytokines such as interleukin-1β (IL-1β), IL-2, tumor necrosis factor-α (TNF-α), TNF-β and interferon-γ (IFN-γ) (Schlievert et al. 2010). Superantigens stimulate T cells by bypassing the usual pathway of antigen-mediated immune response, usually antigens are processed by antigen presenting cells and processed antigen fragments then expressed on the surface of the major histocompatibility complex type II (MHCII). The resulting antigen-MHCII complex then interacts with receptor of T cell thus activating specific T cell mediated immune response. However, superantigens are not presented by antigen-presenting cells. Instead, they bind directly to the MHCII complex where they can interact with T cell receptors. Since superantigens are not presented by antigen- presenting cells, they activate T cells non-specifically. This non-specific activation of large number of T cells results in massive production of cytokins. Several studies have shown that these cytokines can mediate clinical symptoms such as fever, rash, hypotension, tissue injury, and shock (Stevens et al. 1989; Hackett et al. 1992; Leung et al. 1995; Bronze et al. 1996; Johnson et al. 1996).

1.4.2.3 **Exotoxins**

S. aureus produces a large repertoire of exotoxins that possess highly inflammatory cytolytic activity. Cytolytic toxins cause cell death by forming pores in the membrane of the target cell (Bien et al. 2011). Single *S. aureus* strain can secret a number of cytolytic toxins including hemolysins (α, β, γ), leukocidin and Panton-Valentine leukocidin (PVL) (DuMont et al. 2013). Hemolysins, including α and β have the ability to damage the membrane of host immune cells by osmotic lysis through the formation of pores in the membrane of platelets and monocytes. Leukocidins such as LukED and LukAB are cytotoxic for innate immune cells, including lymphocytes, macrophages, and dendritic cells and play an important role in *S. aureus*

infections (Alonzo et al, 2013). LukAB leukocidin was identified in 2011 (Dumont et al. 2011) and found that it specifically targets phagocytic cells such as polymorphonuclear (PMN) cells or neutrophils. In 2013, DuMont et al. identify cellular receptors for LukAB-mediated cytotoxicity and demonstrated their role in species specificity (DuMont et al. 2013). PVL, hetero-chain containing heptamer pore-forming cytolytic toxin is highly cytotoxic to human PMNs (Schlievert et al. 2010). However, PVL showed minimum cytotoxicity to mouse PMNs in a pulmonary disease model. Furthermore, both PVL^+ and PVL^- *S. aureus* strains were lethal for mice when administered intra-bronchially (Schlievert et al. 2010).

Biofilm formation is another virulence strategy which allows *S. aureus* to persist on host tissue or to the surface of a medical device and resist host defenses or antibiotics (Foster, 2005). Generation of small colony variants is another virulence mechanism which helped *S. aureus* survival in a metabolically inactive state under harsh conditions. Small colony variants have an association with chronic, recurrent and persistent infections such as chronic osteomyelitis and persistent skin and soft tissue infection (von Eiff et al. 2006).

1.5 **Epidemiology of *S. aureus* disease**

Humans are a natural reservoir for *S. aureus*, which is a commensal bacterium known to asymptomatically colonize the human skin, nares, and gastrointestinal tract (Lowy 1998). Individuals colonized with *S. aureus* are at increased risk to develop infections. Risk factors associated with *S. aureus* colonization include intravenous drug use, use of intravascular devices, surgery, immunocompromised or immunosuppressed patient, and patient with type 1 diabetes (Lowy, 1998; Naber, 2009). Transmission of infection occurs mainly by direct contact to a colonized carrier (Chambers, 2001). The rates of infections caused by *S. aureus* have been increased and the treatment options for these diseases are becoming limited because of the

increased resistance of *S. aureus* strains to antibiotics (Corey, 2009). Thus the ability of these bacteria to spread in both community and hospital settings has increased substantially. Earlier investigations on methicillin-resistant strains of *S. aureus* (MRSA) revealed that these strains are largely confined to hospitals and long-term care facilities; however, their prevalence in the community is now well recognized (Saravolatz et al. 1982; David et al. 2010).

1.6 **Pathogenesis of HA-MRSA**

In 1959, methicillin was first introduced and thought to be an antibiotic of choice for almost all of the penicillin resistant *S. aureus* strains. However, in 1961, soon after the introduction of methicillin, the first methicillin-resistant strains were identified in England (Jevons, 1961). Since then, MRSA has become a major nosocomial pathogen that causing severe morbidity and mortality worldwide (Haddadin et al. 2002). Methicillin resistant gene *mecA* is a part of mobile genetic element called the staphylococcal cassette chromosome (SCC) which encodes a penicillin-binding protein 2A (PBP2A). PBP2A binds to the β-lactam antibiotics with lower affinity than the regular PBP of *S. aureus* (Chambers 1997). Hospital-acquired MRSA (HA-MRSA) strains possess large SCCmec types I to III which encode one or multiple antibiotic resistance genes and allow the bacterium to survive under the pressure of antibiotics (Liu, 2009). It was found that HA-MRSA strains rarely causes disease in healthy individuals; however, they can cause invasive infections including pneumonia, bacteremia in people who are exposed to the health care setting; older people and have one or more predisposing conditions (Chambers, 2001). Therefore, it has been suggested that HA-MRSA strains might be less robust than other strains of *S. aureus* in terms of its pathogenesis and virulent determinants. HA-MRSA strains are more susceptible to neutrophils mediated killing and are less virulent when administered to mice systemically (Voyich et al. 2005). Furthermore, HA-MRSA strains expressed low levels of

phenol-soluble modulin (PSM) peptides which raise the possibility that PSMs might have role in reduced virulence of HA-MRSA. In addition, many of the HA-MRSA isolates exhibit a *agr*$^{-}$ or a mixed *agr*$^{+}$ and *agr*$^{-}$ genotype (Shopsin et al. 2008), and the presence of this genotype could explain the relative nonpathogenic nature of HA-MRSA toward immunocompetent hosts and beneficial for HA-MRSA survival in the healthcare setting where bacterial competition is limited by the positive pressure of antibiotics. Vuong et al. (Vuong et al. 2000) also found that *agr*$^{-}$ genotype facilitates biofilm formation and proliferation of HA-MRSA isolates on plastic tubing. Virulence of HA-MRSA isolate is largely depending upon to its ability to form biofilm on indwelling medical devices (Ferreira et al, 2013).

1.7 **Pathogenesis of CA-MRSA**

Until 1990s, MRSA infections were mainly confined to immunocompromised individuals or individuals with healthcare exposure and it rarely caused infections among community members without exposure to the health care setting. In the late 1990's four healthy children died in the USA from MRSA sepsis and pneumonia (CDC report 1999). That was the first report of MRSA infections which was occurring among healthy people in the community without health care exposure. Since then a number of reports had been published which had described the emergence of new community-acquired MRSA (CA-MRSA) strains (Groom et al. 2001; Baggett et al. 2003; Liu et al. 2009) that were markedly different from HA-MRSA. CA-MRSA strains are genetically different from HA-MRSA strains, shared a small sized type IV SCC*mec* cassette, and encoded the genes for the Panton-Valentine Leukocidin (PVL) (Vandenesch et al. 2003).

CA-MRSA strains largely caused skin and soft tissue infections; however, several reports have found their connection with more severe infections such as necrotizing pneumonia, necrotizing fasciitis, and myositis (Bradley, 2003; Miller et al. 2005; Pannaraj et al 2006). One

CA-MRSA clone, named USA300 spreads rapidly and is inked to more severe bone, skin and soft tissue infection (Miller et al. 2008). Furthermore, the rate of skin colonization is significantly higher in CA-MRSA infected individuals than the HA-MRSA and MSSA infected individuals. Together, these observations suggest that clone USA300 is more virulent because of its rapid transmission capacity, ability to colonize efficiently and unique mode of pathogenicity. Though the clone USA300 has been proven to be the most virulent isolate but it is not known whether other clones of CA-MRSA are equally virulent. The major virulence determinants of the CA-MRSA epidemic strain include PVL, *arc* gene in arginine deiminase system, *opp-3* gene in ABC-transporter and PSM peptides (Boyle-Vavra et al. 2007; Diep et al. 2006; Wang et al. 2007). However, the role of each virulent factor in the pathogenicity of the specific strain is not known.

1.8 Host defense mechanisms to MRSA infections

S. aureus can survive both outside and inside of host cells. Host cells that provide support for intracellular growth of *S. aureus* are epithelial cells, endothelial cells, and even macrophages (Kubica et al. 2008). Neutrophils play critical roles by providing first line of defense against invading *S. aureus*. The roles of neutrophils in providing a rapid, non-specific, and potent response to infections are manifested by increased incidence of recurrent invasive *MRSA* infections of neutrophil dysfunction immune system (Dinauer et al. 2000). PMN are part of the innate immune system, rapidly migrate to sites of infection where they bind and engulf invading *S. aureus (*Rigby et al. 2012*)* for phagocytosis which triggers potent oxidative and non-oxidative antimicrobial killing mechanisms, including generation of reactive oxygen species (ROS) and antimicrobial proteins that serve to limit pathogen survival and dissemination (Babior et al. 1973; Bainton et al. 1968). PMN activation is intimately linked with the production of

superoxide and other secondarily derived ROS, such as hypochlorous acid, hydroxyl radical, chloramines, and singlet oxygen that have proven microbicidal activity (Rigby et al. 2012). Furthermore, neutrophil mediated phagocytosis triggers synthesis of a number of immunomodulatory factors (Kobayashi et al. 2003; Borjesson et al. 2005; Scapini et al. 2000) which recruit additional neutrophils, modulate subsequent neutrophil responses, and coordinate early responses of other cells types such as monocytes, macrophages, dendritic cells and lymphocytes thereby providing an important link between innate and acquired immune responses (Rigby et al. 2012).

The progression of disease followed by infection triggers acquired immune responses which are responsible for clearance of invading pathogens and providing long term immunologic memories. B-cell and T-cell mediated responses, which are part of adaptive immune system, are involved in antibody production and cell-mediated defense against specific antigen of *S. aureus*, respectively (Girardi, 2007). In B-cell mediated immune responses against *S. aureus*/MRSA infections, many antibodies are generated against toxins, cell-wall proteins, capsular polysaccharides and other virulence factors (Holtfreter et al. 2010). These newly generated antibodies then opsonize the bacteria to facilitate complement-mediated ingestion of *S. aureus* (Holtfreter et al. 2010). In addition to opsonization and phagocytic responses of antibodies, other host defense system includes inhibition of cytolytic activity of α-toxin and PVL toxins and inhibition of *S. aureus* nutrient uptake (Kennedy et al. 2010; Bubeck et al. 2008; Brown et al. 2009). Furthermore, T helper cell subsets (CD4+ T cells) play an important role in the pathogenesis of *S. aureus* skin infections by producing a number of cytokins including IFN-γ, IL-4, IL-13, IL-17, IL-21 and IL-22 (O'Shea et al. 2010). These cytokins promote neutrophil recruitment, cell-mediated and antibody mediated immune responses (O'Shea et al. 2010;

Krishna et al. 2012). It is important to understand the mechanisms of protective immune responses against *S. aureus* infections for the future development of immunomodulatory therapies and vaccination strategies to prevent infections caused by *S. aureus*.

1.9 **Current treatment options for MRSA infections**

The increasing antimicrobial resistance and a variety of diseases, including invasive and noninvasive infections, caused by MRSA are limiting factors for treating MRSA infection. The treatment of MRSA infections is also limited by the ability of MRSA strains to produce biofilm in tissues and medical devices (Mazaitis et al. 2011). A number of antibiotics are currently being used for the treatment of MRSA infections. For the treatment of minor skin infections including small furuncles and abscesses, surgical incision and drainage of pus combine with oral antibiotic therapy is the first choice of treating MRSA infection. Oral antibiotics for the treatment of CA-MRSA infections include linezolid, rifampin with/without fusidic acid, trimethoprim-sulfamethoxazole, doxycycline, minocycline, clindamycin (Moellering, 2008). Linezolid is the only oral agent that is currently used for the treatment of CA-MRSA infections in outpatients (Moellering, 2008). In case of serious skin and skin-structure infections and pneumonia due to CA-MRSA, linezolid might be particularly useful because of its ability to impair toxin production (Moellering, et al. 2008; Ramirez et al. 2012). Though the application of linezolid for the treatment of MRSA provided apparent advantages (Micek 2007), certain safety concerns including serotonin toxicity and thrombocytopenia limit the applicability of this antibiotic for treating MRSA infections (Kishimoto 1995; Lawrence, Adra et al. 2006). Additional parenteral antimicrobials such as vancomycin, teicoplanin, daptomycin, linezolid, and tigecycline can be used for treating severe infections due to CA-MRSA. Most recently the U.S. Food and Drug Administration (FDA) approved dalbavancin and tedizolid phosphate, a lipoglycopeptide and

protein synthesis inhibitor antibiotics respectively, for treating adult patients with acute bacterial skin and skin structure infections (ABSSSI) (Walker, 2014). For the treatment of nosocomial pneumonia, vancomycin, linezolid, and teicoplanin antibiotics are currently being used (Kalil et al. 2013; Rivera et al. 2011). Daptomycin is not recommended for use in pneumonia due to MRSA because of its inactivation by pulmonary surfactant (Liu et al. 2011; Ramirez et al. 2012). However, daptomycin is recommended for patient with bacteremia and endocarditis (Liu et al. 2011). Vancomycin was introduced into clinical practice in 1958, and was first isolated from the soil Actinobacterium *Amycolatopsis orientalis*. It is a member of the glycopeptide antibiotic family that inhibits the late stages of peptidoglycan assembly by forming complexes with the *D*-alanyl-*D*-alanine (*D*-Ala-*D*-Ala) in C termini of the peptidoglycan precursors on the external side of the cell membrane (Neu et al. 2002). Cell wall assembly is hampered by the formation of these complexes which prevents the cross-linking reaction catalyzed by transglycosylases, *D,D*-transpeptidases, and *D,D*-carboxypeptidases (Beltrametti et al. 2010). Though vancomycin is the first choice of treatment for most of the MRSA infections, its application is limited by the emergence of strains with reduced antimicrobial susceptibility (Davies et al. 2010) and the occurrence of vancomycin treatment failure and mortality in patient with methicillin-sensitive *S. aureus* (MSSA) bacteremia (Graves et al. 2008; Marco et al. 2008). The MIC creep, the incremental vancomycin MIC, is a frequently observed phenomenon in MRSA infected patients with vancomycin treatment (White et al. 2007). Until now eleven vancomycin-resistant MRSA strains have been identified, nine in the USA(7 from Michigan, 1 from Pennsylvania, and 1 from New York) (Périchon et al. 2009). Wide spread use of vancomycin for the treatment of MRSA infection has led to the emergence of two types of glycopeptide-resistant *S. aureus* strains that include vancomycin-intermediate-resistant *S. aureus* (VISA) and vancomycin-resistant *S. aureus*

(VRSA). In addition to antibiotic resistance to vancomycin, side effects caused by vancomycin treatment also impede treatment of MRSA infections. Patients treated with vancomycin may experience nephrotoxicity (Lodise, Patel et al. 2009), “red man” syndrome (Sivagnanam and Deleu 2003) and anaphylaxis (Polk 1991; Neughebauer, Negron et al. 2002). Although MRSA causing significant health care-associated and community-associated infections, the current therapeutic alternatives for MRSA infections are limited to few antibiotics.

1.10 **Resistance mechanisms to antimicrobial agents**

S. aureus is the most versatile bacterial pathogen that has a unique ability to evolve quickly in response to new antibiotic (Lowy, 2003; Reygaert, 2013). It has developed resistance towards a number of antibiotics including penicillin, methicillin, cephalosporins, amikacin, clindamycin, fluoroquinolones, gentamycin, tobramycin, trimethoprim/sulfamethoxazole, vancomycin, linezolid and daptomycin. *S. aureus* confers antibiotic resistance by four general mechanisms including i) enzymatic inactivation of the antibiotic ii) alteration of the target side which led to decrease affinity for the antibiotic iii) trapping of the antibiotic and iv) efflux pumps. Bacteria develop these resistance mechanisms by intrinsic (presence of SCC *mec* gene or *vanA* operon in their chromosome), acquired (through horizontal gene transfer via plasmid, bacteriophage) or through spontaneous mutations and positive selection (Reygaert, 2013).

The mortality rate of patients with *S. aureus* bacteremia before the introduction of antibiotics was more than 80%. However, the mortality rate decreased dramatically after the introduction of penicillin in the early 1940s. In 1944, Kirby (Kirby 1944) was first identified seven *S. aureus* strains that were resistant to penicillin. First penicillin-resistant staphylococci were recognized in hospitals and subsequently in the community (Rammelkamp et al. 1942). Today greater than 95% of all *S. aureus* isolates are resistant to penicillin and the resistance is

primarily mediated by the enzyme *β*-lactamase which is encoded by the gene *blaZ*. The enzyme β-lactamase hydrolyzes the *β*-lactam ring presents in this antibiotic for enzymatic inactivation (Bondi et al. 1945). Resistance to *β*-lactam antibiotics can also be confirmed by the production of an extra penicillin-binding protein, PBP 2a, a transpeptidase encoded by the gene *mecA (Zhang et al. 2001).* The *mecA* gene, which is part of a mobile genetic element found in all MRSA strains, confer resistance not only to methicillin, but also to all other *β*-lactam antibiotics (Lowy, 2003).

With the increasing antimicrobial resistance in *S. aureus*, vancomycin has become the drug of choice for treating infections caused by *S. aureus*. However, indiscriminate use of vancomycin created two types of resistant strains dubbed as vancomycin intermediate *S. aureus* (VISA) and vancomycin resistant *S. aureus* (VRSA). Both strains showed complete resistance to vancomycin but the development of resistance mechanisms were different (Lowy, 2003). The VRSA strains acquire resistance by conjugal transfer of the *vanA* operon from an *Enterococcus faecalis (*Périchon et al. 2004*),* whereas chromosomal resistance was found in VISA strains (Saito et al. 2014). The glycopeptide-intermediate-resistant is associated with a thickened and poorly cross-linked cell wall, resulting in an accumulation of *D*-alanyl-*D*-alanine (*D*-Ala-*D*-Ala) targets in the periphery that sequester glycopeptides(Périchon et al. 2009). Sieradzki et al. (Sieradzki et al. 1999) showed that large amounts of vancomycin become sequestered in the abnormal peptidoglycan that are unable enter into the cell and confer its antimicrobial activity.

Daptomycin is a lipopeptides antibiotic first introduced into market in 2003 for skin and invasive infections (Bayer et al. 2013). Since 2005 an alarming number of reports have been published indicating the *in vivo* development of daptomycin resistance (Vikram et al. 2005; Julian et al. 2007; Murthy et al. 2008; Chambers et al. 2009; Bayer et al. 2013). Daptomycin

resistant strains of *S. aureus* showed altered structure of both cell wall and cell membrane which resulted in cell membrane depolarization, reduced surface binding and permeability of daptomycin. Furthermore, cell wall modifications by increased expression of the *dlt* operon which is involved in D-alanylation of teichoic acids and by thickening of cell wall are found to play important role in daptomycin resistance phenomenon (Ho et al. 2008; Straus et al. 2006; Scott et al. 2007; Bayer et al. 2013).

Linezolid is a synthetic protein synthesis inhibitor, used for the treatment of infections caused by MRSA and VRSA (Tsiodras et al. 2001). Bacteria acquire resistance by spontaneous point mutations at the drug target site or by acquisition of a natural resistance gene, *cfr (*Morales et al. 2010*).* The *cfr* gene is present in plasmids and may be horizontally transferred to S. aureus strains since it was first identified in a bovine *S. sciuri* isolate (Schwarz et al. 2000; Morales et al. 2010).

1.11 **Discovery of new antibiotics for MRSA infections**

Since MRSA infections are continuously challenging the medical and scientific community due to limited treatment options and the emergence of new antibiotic-resistant strains, it is essential to find new antibiotics that will be able to cure MRSA infections. According to the Infectious Diseases Society of America (IDSA), at least ten new antibiotics are needed by 2020 to combat multi drug resistant (MDR) bacterial pathogens (IDSA 2010). To discover new antibiotics with novel mechanisms of action we need to expand our knowledge in conventional culturing approach, novel culture methods, heterologous DNA-based methods, metagenomics, combinatorial biosynthesis and fragment-based drug design.

1.11.1 Discovery of novel antibiotics using a culture-based approach

Several new methods for developing antibiotic have been introduced recently, including genome mining, novel culturing methods, and metagenomics. Decoding the genomes of antibiotic producing microbes has revealed a surprisingly large number of biosynthetic pathways (Bentley et al. 2002; Wilkinson et al. 2007; Nikolouli et al. 2012). *Streptomyces coelicolor* is known to produce only three antimicrobial compounds, even though its genome encodes 20 secondary metabolites (Bentley et al. 2002). Nearly every antibiotic-producing microbial genome examined bioinformatically contains multiple pathways for secondary metabolites (Nikolouli et al. 2012). Unfortunately these cryptic pathways are mostly silent and efforts to turn them on have succeeded, but not as a large scale platform (Lewis, 2013). Culture-based approaches have recently gotten new momentum due to technology advancements such as whole genome sequencing, high-throughput screening and cultivating as-yet unculturable microorganisms using new cultivation approaches (Baltz 2008). culture-based strategies are still the most successful method for discovering clinically useful antibiotics (Fleming 1929; Debono, Abbott et al. 1988; Nam et al. 2013) since alternative strategies such as metagenomics, combinatorial biosynthesis and fragment-based drug design have yet to yield large numbers of novel chemical entities with antimicrobial activity. It was estimated that natural products from only bacteria account for half of all commercially available pharmaceuticals (Stewart, 2012). Both culturable and unculturable bacteria from diverse environmental sources have potential to produce antimicrobial compounds. New cultivation strategies using diffusion chambers to mimic the natural enviromnent have been devised to coax uncultured bacteria to grow (Ling et al. 2015). A new class of antibiotic, teixobactin, was recently discovered by screening 10,000 soil bacteria using this approach (Ling et al. 2015). Though the majority of bioactive natural products have been found in soil-borne

bacteria especially from actinomycetes, other environment such as marine ecosystem is a promising source of novel antibiotic producers.

Bacteria from the phylum *Actinobacteria* are the most prominent source for clinically important antibiotics (Baltz 2007). Among them microorganisms belonging to the order Actinomycetales are fascinatingly diverse for their ability to produce biologically active secondary metabolites. It has been estimated that the order Actinomycetales has yielded ~3000 antibiotics since after its first reporting of streptomycin in 1942 (Waksman et al. 1941; Watve et al. 2001). The genus *Streptomyces* itself produced about 90% of these antibiotics including cephalosporins, chloramphenicol, neomycin, erythromycin, tetracycline, novobiocin, vancomycin, kanamycin, fosfomycin and daptomycin (Watve et al. 2001; de Lima Procópio et al. 2012). It has also been reported that *Streptomyces* spp. have the ability to synthesize important antifungal (amphotericin B) (Caffrey, Aparicio et al. 2008), anticancer (mitomycin C) (Olano, Mendez et al. 2009), antiparasitic (ivermectin) (Nett et al. 2009) and immunosuppressive (rapamycin) agents (Graziani 2009).

Single species of *Streptomyces* can produce several antibiotics and antibiotic profile is specific to each species. *Streptomyces coelicolor* is the model species of this genus produces at least five different antibiotics including two pigmented antibiotics actinorhodin and undecylprodigiosins, as well as a polyketide cryptic polyketide, a calcium-dependent ionophore antibiotic, and an unusual cyclopentanone antibiotic methylenomycin (Liu et al. 2013). David A. Hopwood was the pioneer of the *S. coelicolor* genetic studies (Chater, 1999; Ruddy et. al. 1979; Hopwood, 2006) provided the first evidence that the genes for biosynthesis of any particular antibiotic are clustered on the chromosome or in plasmids (Rudd et al. 1979; Ruddy et al. 1980). *S. coelicolor* has a complex life cycle consisting of the substrate or vegetative mycelium to the

formation of sporulating aerial hyphae. The germination of spores occurs in the presence of soluble nutrients, and when nutrients begin to run out, an aerial mycelium forms. Autolysis of the vegetative and substrate mycelium generates energy for the aerial growth. Antibiotic production usually takes place at this stage when nutrient concentration is limited. Antibiotic production might provide protection for the nutrients being released or against invaders (Chater et. al. 1979).

Complete genome sequence of *S. coelicolor* strain A3(2) (Bentley et al. 2002) revealed that the organism has a linear chromosome of 8.6Mbp encoding 7,825 predicted genes with more than 20 known or predicted secondary metabolites biosynthesis gene clusters (Bentley et al. 2002). These clusters are generally large and usually contain several operons. Genome mining also revealed that large portion of the genome is dedicated to encode regulatory genes that control physiology, developmental state, population density, and the level of antibiotic production. Detailed molecular analysis of sequenced genes of *S. coelicolor* demonstrated the complex developmental inter play of antibiotic production with morphological differentiation providing information for increased level of antibiotic production and also suggested ways to activate silent genes for its production. The physiological signals and regulatory mechanisms play important roles in the activation of many cryptic secondary biosynthetic gene clusters thus understanding these mechanisms will unleash the full biosynthetic potential of this organism. The proposed methods for activating cryptic gene clusters in *Streptomyces* include manipulation of fermentation conditions, genome mining, genetic manipulation of the regulatory gene clusters, regulation of signaling molecules, ribosome engineering and heterologous expression of gene clusters (Liu et al. 2013).

Antibiotic production is dependent on a number of nutritional factors including efficient sources of carbon, nitrogen, phosphate, several metals such as zinc, iron, and manganese (Coisne et al. 1999; Owen et al. 2007). In addition to nutritional effects, the pH, temperature and dissolved oxygen level are also important for antibiotic production (Desai et al. 2002).

In addition to *Streptomyces*, other *Actinobacteria* including *Nonomuraea* produce secondary metabolites with antimicrobial activity (Cornaglia et a. 2009). For instance, *Nonomuraea* sp. ATCC 39727 produces the teicoplanin-like glycopeptide antibiotic A40926 which is a precursor for semi-synthetic derivative dalbavancin, an antibiotic that is currently approved by FDA for treating MRSA infections. The genus *Nonomuraea* was described by Zhang et al. (Zhang et al. 1998) based on spore formation and 16S rRNA gene sequences. The genus comprises 27 species and *Nonomuraea pusilla* is the type species (Zhao, Li et al. 2011). The members of this genus can be isolated from diverse natural habitats including soil, plants, caves, marine and river sediments (Nakaew, Sungthong et al. 2012). Like most *Actinobacteria*, *Nonomuraea* spp. also produce various shades of blue, violet, red, rose, yellow, green, brown and black pigments on natural and synthetic media, and the pigments maybe dissolved into the medium or it may be retained in the mycelium. The pigments produced from *Actinobacteria* have been used on cotton shades (Perumal, Stalin et al. 2009) and in medicine, pharmacology and cosmetic preparations (Perumal, Stalin et al. 2009). Moreover, several reports indicated that antimicrobial activities and/or antibiotic production of *Actinobacteria* are associated with pigment production (Miyaura and Tatsumi 1961). Watve et al (Watve, Tickoo et al. 2001) estimated that members of this phylum potentially produce around 100,000 antimicrobial metabolites, and this estimate demonstrates that only a small percentage of the extant antibiotics that have already been

discovered. To discover the vast number of remaining antibiotics, we need to get access to a greater diversity of bacteria.

Members of other bacterial phyla including *Firmicutes, Bacteroidetes* and *Proteobacteria* are well known to be a prolific source of bioactive natural products (Hamdache et al. 2011; Stein 2005; Figueiredo et al. 2011). Within these diverse bacteria, the genus *Bacillus* represents particular interest in terms of its ability to produce wide range of natural products with potential antimicrobial activity. The species of genus *Bacillus* is common in soil and plays an important role in plant growth as they are considered to be a member of plant growth-promoting rhizobacteria (PGPR). The PGPR bacteria can promote plant growth directly by helping plants acquire nutrition from soil, or indirectly by controlling phytopathogens to prevent plant diseases (Kloepper et al. 1980). Application of PGPR results in significant enhancement of plant growth and increased yields of agronomically important crops. PGPR produce different types of metabolites that include antibiotics (Fuller, Mellows et al. 1971), cell wall- degrading enzymes (Ramos-Solano, Lucas García et al. 2010), sidorophores (Kloepper, Leong et al. 1980) and HCN (Askeland and Morrison 1983). A wide variety of antibiotics that include polyketides, heterocyclic nitrogenous compounds, phenylpyrrole, cyclic lipopeptides, noncyclic lipopeptides and aminopolyols are also produced by PGPR strains (Fernando, Nakkeeran et al. 2006).

The majority of antibiotics produced by *Bacillus* spp. are low molecular weight polypeptides that are synthesized by ribosomal or non-ribosomal mechanisms. Surfactins, iturins, fengycins macrolactins, difficidins and lantibiotics are major antibiotics that are produced by *B. subtilis* (Hamdache et al. 2011; Stein 2005; Sumi et al. 2014). In addition to antibiotic production, *Bacillus* spp. can be used as probiotics in animals and plants for growth promotion and disease control (Ran et al. 2012; Ahmed et al 2014; González-Ortiz et al. 2013).

The beneficial effects of probiotics include antagonism to pathogens, enhancement of immune response and restoration of the body's normal microflora (Sun et al. 2010; Casula et al. 2002). Many *Bacillus* strains from species *B. clausii, B. cereus, B. pumilus* and *B. amyloliquefaciens* are currently being used as probiotic strains for human nutrition, as animal feed supplements and in aquaculture (Krober et al. 2014; Lee et al. 2012; Verschuere et al. 2000; Larsen et al. 2014). Probiotics contain live microorganisms that, when administered in adequate amounts (e.g., in case of *Bacillus* a single dose contains up to 10^9 spores/g or 10^9 spores/ml), confer a health benefit on the host (Duc et al, 2004; Sanders, 2008). Probiotics are generally taken as a prophylactic agents, but their application as therapeutic agents has also been described in literature (Mazza, 1994). However probiotics are used, the following three basic mechanisms are involved their beneficial effects: 1) secretion of antimicrobial compounds which inhibit the growth of pathogens, 2) competitive exclusion of pathogens (e.g., competition for adhesion sites), and 3) immunomodulation (e.g., stimulation of lymphocytes and induction of cytokines). The ability of probiotic *Bacillus* strains to produce antimicrobial agents is well documented and more than 80 different types of antimicrobial compound have been identified from different *Bacillus* species (Mazza, 1994). These antimicrobial agents can completely inhibit the growth of pathogenic bacteria or contribute to the competitive exclusion of pathogens. The immune stimulation mechanism of probiotic involves in induction of proinflammatory cytokines that increase phagocytosis of pathogens (by macrophages or dendritic cells) and also stimulation of B cells and cytotoxic T cells. Duc et al. (Duc et al. 2004) demonstrated that *Bacillus* spore is immunogenic, producing 10-fold higher spore-specific IgG responses when mice were orally administered with Biosubtyl^{NT} , a strain of *B. pumilus*. *In vivo* cytokine expression profile revealed the production of TNF-α and IFN-γ in the secondary lymphoid organs and gut

associated lymphoid tissue of mice after inoculation of probiotic (Duc et al. 2004). However, in their *in vitro* assay *TNF-α and IL-1α* were absent instead they found proinflammatory cytokine IL-6. Together, these inflammatory responses enhance the innate immune system and activate macrophage for phagocytosis after application of probiotic. Similar results were observed when probiotic *Lactobacillus* species were administered orally (Schiffrin et al. 1995).

Although probiotic use of *Bacillus* for growth promotion and prevention of gastrointestinal infection has been studied extensively, the probiotic effect of *Bacillus* strains against MRSA has not been previously explored. Therefore, exploring *Bacillus* spp. for identifying active compounds with anti-MRSA activity and their potential use as a probiotic to prevent skin colonization by MRSA will be of special interest.

1.11.2 Culture-independent approach to discover antibiotics

Prokaryotes are the most abundant biological entities in soils (except perhaps viruses) and comprise most of the soil biomass (Hassink, Bouwman et al. 1993). The metabolic and functional versatility of soil microorganisms makes this environment a good source for the discovery of novel natural products, including antibiotics (Marcia S. Osburne, Trudy H. Grossman et al. 2000). A vast number of antibiotics in use today to treat patients with infectious diseases are derived from soil bacteria or fungi (Newman and Cragg 2007). However, only a small fraction of the environmental microbes has been successfully explored for natural product discovery (Thomason et al. 2007) indicating that a wealth of microbial diversity still exists in nature and the remaining untapped natural product diversity offers tremendous potential for discovering novel therapeutics. A recent strategy for antibiotic discovery is metagenomic analysis of uncultured microbes.

Metagenomic libraries are constructed by extracting environmental DNA (eDNA) directly, allowing analysis of the collective genomes of all the resident organisms (Yadav et al. 2003; Kapur et al. 2008). Any environmental source such as soil (Sangwan, Lata et al. 2012), sediments (Havelsrud, Haverkamp et al. 2011), activated sludge (Liaw, Cheng et al. 2010), and hot thermal vent sediments (Wemheuer, Taube et al. 2013) can be used for the isolation of genomic DNA for metagenomic studies. However, isolation of high molecular weight (HMW) DNA from environment is always a significant issue due to the inherent conflict between the need to recover DNA from diverse microorganisms while preserving DNA integrity. A significant progress has been made for isolation of high quality of DNA from diverse environmental samples (Liles et al. 2008; Voget et al. 2006; Stein et al. 1996; Abulencia et al. 2006; Uchiyama et al. 2005; Rhee et al. 2005). The isolated eDNA fragments then directly cloned into a vector and the resulting cloned libraries can be investigated for the microbial diversity analysis and drug discovery.

To construct a metagenomic library, eDNA fragments can be cloned into plasmid, cosmid, fosmid or BAC (Bacterial Artificial Chromosome) vector. These cloning vectors have limitation for example plasmid, cosmid and fosmid vectors can carry only small fragments of DNA: plasmids, <20 kb; cosmids, 37 - 52 kb; and fosmids, <42 kb. However, BAC vector can carry longer DNA fragments of up to 300 kb (Wang et al. 2014). This is particularly advantageous since many of the bioactive compounds (e.g. antibiotics, multimodular polyketide, or nonribosomal peptide) are encoded by a gene cluster whose length typically exceeds beyond the limits of DNA sizes carried by a plasmid, cosmid, or fosmid vector (Piel, 2011). Other advantages of BAC clone libraries include screening a smaller number of BAC clones to identify desired cluster and stability of metagenomic DNA fragments in the BAC vector. However, BAC

libraries are technically more difficult to construct than other types of clone libraries, the choice of the vector depends on the DNA quality, targeted genes, and library screening strategy (Daniel, 2005).

A metagenomic library can be screened by functional screening and sequence based screening. Recent advancement of next-generation sequencing techniques allows large-scale analysis of microbial communities with novel applications. Sequence-based screening is a widely used approach to find genes or gene clusters involved in particular functions within a metagenomic library. Sequence-based screening first identify clones that contain know pathways for secondary metabolites and then express these pathways in a suitable host for activity. This method allows to identify novel analogs of known metabolites. Sequence-based screening has led to the successful identification of pathways and genes that encode novel enzymes, such as type II polyketide synthases (PKS) biosynthetic pathway was first identified in cosmid clones (Feng et al. 2011), discovery of new pentangular polyphenols including calixanthomycin A, arenimycins C and D with antiproliferative and antibacterial activity respectively (Kang et al. 2014), and several enzymes including nitrite reductases (Bartossek et al. 2010), glycerol dehydratases (Knietsch et al. 2003), chitinases (Hjort et al. 2010). The first bacterial proteorhodopsin, a light-driven proton pump, was also discovered by sequence analysis of specific BAC clones containing 16S rRNA gene sequence and other essential genes for function (Beja et al.2000). Though sequence-based screening showed potential to discover novel compounds and pathways, large-scale metagenomic projects are limited with respect to data handling, data integration and data analysis.

Though there are a number of outstanding challenges in functional metagenomics such as 1) the insert size of the library, 2) efficient screening methods for the massive libraries generated,

and 3) barriers to heterologous gene expression in the host species (Uchiyama et al. 2009) which can limit natural product discovery, functional metagenomics is still the only strategy that bears the potential to identify entirely novel classes of genes encoding known or novel functions.

The functional diversity of metagenomic clones can be determined phenotypically and in most cases, phenotypical detection employs by applying an indicator substrate of the enzymes of interest into the growth medium, where it confers the specific metabolic capabilities of individual clones. The active clones can be detected by visual inspection of an indicator agar plate, by flow cytometry, a spectrophotometer, or fluorescent microtitre plate readerdepending on the assay type (Taupp et al. 2011).

Enzymatic activities expressed by metagenomic clones can be identified by function-based screening. The active clones can be detected via many different methods, including colorimetric or fluorescent assays, as well as indicator media (Taupp et al. 2011). For example, lipase-producing BAC clones were identified using 1.0% tributyrin on LB agar plates. Tributyrin hydrolysis forms a halo around clones which can produce lipase enzyme. Other enzymes identified from BAC clones include cellulase, xylanase, esterase, alcohol dehydrogenase, amidase, amylase, protease, chitinase, dehydratase, and β-lactamase (Lorenz et al. 2005).

Identification of secondary metabolites with antimicrobial activities using functional metagenomic approach is well documented such as isolation of turbomycin A and B (Gillespie et al. 2002), the identification of antibacterial activities expressed in cosmid libraries in different proteobacterial hosts (Craig et al. 2010), identification of gene clusters involved in synthesis of antifungal activities (Chung et al. 2008) and identification of anticancer agent (Pettit 2004; Banik and Brady 2010). MacNeil et al (MacNeil et al. 2001) isolated a small molecule, called indirubin,

from a soil metagenomic library and later it was found that indirubin and its derivatives inhibit tumor growth by antitumor angiogenesis (Zhang et al. 2011; Shin et al. 2012).

The selection of a host for cloning, maintenance and screening of a metagenomic library influences the success of the metagenomic project since functional expression of genes from an environmental source might be limited due to the application of a single heterologous host. *E. coli* is the most widely used host for expressing metagenomic libraries, due to the ease of cloning DNA but not necessarily improved expression of metagenomic DNA. Other bacterial heterologous hosts may include *Streptomyces*, *Agrobacterium tumefaciens*, *Burkholderia graminis*, *Caulobacter vibrioides*, *Pseudomonas putida*, or *Ralstonia metallidurans*, all of which have been used for the discovery of natural products from soils using a functional metagenomic approach (Wang, Graziani et al. 2000; Craig et al. 2010).

Reference

(1999). "Four pediatric deaths from community-acquired methicillin-resistant Staphylococcus aureus - Minnesota and North Dakota, 1997-1999." MMWR Morb Mortal Wkly Rep **48**(32): 707-710.

(2004). "National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2004, issued October 2004." Am J Infect Control **32**(8): 470-485.

Abulencia, C. B., D. L. Wyborski, et al. (2006). "Environmental whole-genome amplification to access microbial populations in contaminated sediments." Appl Environ Microbiol **72**(5): 3291-3301.

Alonzo, F., 3rd, L. Kozhaya, et al. (2013). "CCR5 is a receptor for Staphylococcus aureus leukotoxin ED." Nature **493**(7430): 51-55.

Arbeit, R. D., W. W. Karakawa, et al. (1984). "Predominance of two newly described capsular polysaccharide types among clinical isolates of Staphylococcus aureus." Diagn Microbiol Infect Dis **2**(2): 85-91.

Armstrong, G. L., L. A. Conn, et al. (1999). "Trends in infectious disease mortality in the United States during the 20th century." Jama **281**(1): 61-66.

Babior, B. M., R. S. Kipnes, et al. (1973). "Biological defense mechanisms. The production by leukocytes of superoxide, a potential bactericidal agent." J Clin Invest **52**(3): 741-744.

Baggett, H. C., T. W. Hennessy, et al. (2003). "An outbreak of community-onset methicillin-

resistant Staphylococcus aureus skin infections in southwestern Alaska." Infect Control Hosp Epidemiol **24**(6): 397-402.
Bainton, D. F. and M. G. Farquhar (1968). "Differences in enzyme content of azurophil and specific granules of polymorphonuclear leukocytes. I. Histochemical staining of bone marrow smears." J Cell Biol **39**(2): 286-298.
Bales, P. M., E. M. Renke, et al. (2013). "Purification and Characterization of Biofilm-Associated EPS Exopolysaccharides from ESKAPE Organisms and Other Pathogens." PLoS One **8**(6).
Bandeira, M., P. Almeida Carvalho, et al. (2014). "Exploring Dangerous Connections between Klebsiella pneumoniae Biofilms and Healthcare-Associated Infections." Pathogens **3**(3): 720-731.
Barsumian, E. L., P. M. Schlievert, et al. (1978). "Nonspecific and specific immunological mitogenicity by group A streptococcal pyrogenic exotoxins." Infect Immun **22**(3): 681-688.
Bartossek, R., G. W. Nicol, et al. (2010). "Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context." Environ Microbiol **12**(4): 1075-1088.
Bayer, A. S., T. Schneider, et al. (2013). "Mechanisms of daptomycin resistance in Staphylococcus aureus: role of the cell membrane and cell wall." Ann N Y Acad Sci: 5.
Beja, O., L. Aravind, et al. (2000). "Bacterial rhodopsin: evidence for a new type of phototrophy in the sea." Science **289**(5486): 1902-1906.
Bentley, S. D., K. F. Chater, et al. (2002). "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)." Nature **417**(6885): 141-147.
Bentley, S. D., K. F. Chater, et al. (2002). "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)." Nature **417**(6885): 141-147.
Bien, J., O. Sokolova, et al. (2011). "Characterization of Virulence Factors of Staphylococcus aureus: Novel Function of Known Virulence Factors That Are Implicated in Activation of Airway Epithelial Proinflammatory Response." J Pathog **601905**(10): 14.
Biswas, R., L. Voggu, et al. (2006). "Activity of the major staphylococcal autolysin Atl." FEMS Microbiol Lett **259**(2): 260-268.
Bondi, A., Jr. and C. C. Dietz (1945). "Penicillin resistant staphylococci." Proc Soc Exp Biol Med **60**: 55-58.
Borjesson, D. L., S. D. Kobayashi, et al. (2005). "Insights into pathogen immune evasion mechanisms: Anaplasma phagocytophilum fails to induce an apoptosis differentiation program in human neutrophils." J Immunol **174**(10): 6364-6372.
Boucher, H. W., G. H. Talbot, et al. (2009). "Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America." Clin Infect Dis **48**(1): 1-12.
Boyle-Vavra, S. and R. S. Daum (2007). "Community-acquired methicillin-resistant Staphylococcus aureus: the role of Panton-Valentine leukocidin." Lab Invest **87**(1): 3-9.
Bradley, S. F. (2005). "Staphylococcus aureus pneumonia: emergence of MRSA in the community." Semin Respir Crit Care Med **26**(6): 643-649.
Bronze, M. S. and J. B. Dale (1996). "The reemergence of serious group A streptococcal infections and acute rheumatic fever." Am J Med Sci **311**(1): 41-54.
Brown, E. L., O. Dumitrescu, et al. (2009). "The Panton-Valentine leukocidin vaccine protects mice against lung and skin infections caused by Staphylococcus aureus USA300." Clin Microbiol Infect **15**(2): 156-164.
Brown, S., J. P. Santa Maria, et al. (2013). "Wall Teichoic Acids of Gram-Positive Bacteria."

Annual review of microbiology **67**: 10.1146/annurev-micro-092412-155620.
Bubeck Wardenburg, J. and O. Schneewind (2008). "Vaccine protection against Staphylococcus aureus pneumonia." J Exp Med **205**(2): 287-294.
Chambers, H. F. (1997). "Methicillin resistance in staphylococci: molecular and biochemical basis and clinical implications." Clin Microbiol Rev **10**(4): 781-791.
Chambers, H. F. (2001). "The changing epidemiology of Staphylococcus aureus?" Emerg Infect Dis **7**(2): 178-182.
Chambers, H. F. (2001). "The changing epidemiology of Staphylococcus aureus?" Emerg Infect Dis **7**(2): 178-182.
Chambers, H. F., L. Basuino, et al. (2009). "Relationship between susceptibility to daptomycin in vitro and activity in vivo in a rabbit model of aortic valve endocarditis." Antimicrob Agents Chemother **53**(4): 1463-1467.
Chater, K. (1999). "David Hopwood and the emergence of Streptomyces genetics." Int Microbiol **2**(2): 61-68.
Chater K.F, M. M. J. (1979). Streptomyces. In Developmental Biology of Prokaryotes Parish J.H, UK:Blackwell.
Chavakis, T., K. Wiechmann, et al. (2005). "Staphylococcus aureus interactions with the endothelium: the role of bacterial "secretable expanded repertoire adhesive molecules" (SERAM) in disturbing host defense systems." Thromb Haemost **94**(2): 278-285.
Cheung, A. L., S. J. Projan, et al. (2002). "The Genomic Aspect of Virulence, Sepsis, and Resistance to Killing Mechanisms in Staphylococcus aureus." Curr Infect Dis Rep **4**(5): 400-410.
Chung, E. J., H. K. Lim, et al. (2008). "Forest soil metagenome gene cluster involved in antifungal activity expression in Escherichia coli." Appl Environ Microbiol **74**(3): 723-730.
Clarke, S. R. and S. J. Foster (2006). "Surface adhesins of Staphylococcus aureus." Adv Microb Physiol **51**: 187-224.
Clarke, S. R. and S. J. Foster (2006). "Surface adhesins of Staphylococcus aureus." Adv Microb Physiol **51**: 187-224.
Coisne, S., M. Bechet, et al. (1999). "Actinorhodin production by Streptomyces coelicolor A3(2) in iron-restricted media." Lett Appl Microbiol **28**(3): 199-202.
Corey, G. R. (2009). "Staphylococcus aureus bloodstream infections: definitions and treatment." Clin Infect Dis **15**(48): 598186.
Costa, A. R., D. W. F. B. , et al. (2013). "Staphylococcus aureus virulence factors and disease " : 702-710.
Craig, J. W., F.-Y. Chang, et al. (2010). "Expanding Small-Molecule Functional Metagenomics through Parallel Screening of Broad-Host-Range Cosmid Environmental DNA Libraries in Diverse Proteobacteria." Applied and Environmental Microbiology **76**(5): 1633-1641.
Cramton, S. E., C. Gerke, et al. (1999). "The intercellular adhesion (ica) locus is present in Staphylococcus aureus and is required for biofilm formation." Infect Immun **67**(10): 5427-5433.
Daniel, R. (2005). "The metagenomics of soil." Nat Rev Microbiol **3**(6): 470-478.
David, M. Z. and R. S. Daum (2010). "Community-Associated Methicillin-Resistant Staphylococcus aureus: Epidemiology and Clinical Consequences of an Emerging Epidemic." Clinical Microbiology Reviews **23**(3): 616-687.
Davies, J. and D. Davies (2010). "Origins and Evolution of Antibiotic Resistance." Microbiology

and Molecular Biology Reviews : MMBR **74**(3): 417-433.
Davis, R., M. J. Hossain, et al. (2013). "Complete Genome Sequence of Staphylococcus aureus Tager 104, a Sequence Type 49 Ancestor." Genome Announcements **1**(5): e00706-00713.
de Lima Procópio, R. E., I. R. da Silva, et al. (2012). "Antibiotics produced by Streptomyces." The Brazilian Journal of Infectious Diseases **16**(5): 466-471.
Desai, R. P., T. Leaf, et al. (2002). "Enhanced production of heterologous macrolide aglycones by fed-batch cultivation of Streptomyces coelicolor." J Ind Microbiol Biotechnol **28**(5): 297-301.
Diep, B. A., S. R. Gill, et al. (2006). "Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant Staphylococcus aureus." Lancet **367**(9512): 731-739.
Dinauer, M. C., J. A. Lekstrom-Himes, et al. (2000). "Inherited Neutrophil Disorders: Molecular Basis and New Therapies." Hematology Am Soc Hematol Educ Program: 303-318.
Downer, R., F. Roche, et al. (2002). "The elastin-binding protein of Staphylococcus aureus (EbpS) is expressed at the cell surface as an integral membrane protein and not as a cell wall-associated protein." J Biol Chem **277**(1): 243-250.
Duc, L. H., H. A. Hong, et al. (2004). "Characterization of Bacillus Probiotics Available for Human Use." Applied and Environmental Microbiology **70**(4): 2161-2171.
Dumont, A. L., T. K. Nygaard, et al. (2011). "Characterization of a new cytotoxin that contributes to Staphylococcus aureus pathogenesis." Mol Microbiol **79**(3): 814-825.
DuMont, A. L., P. Yoong, et al. (2013). "Staphylococcus aureus LukAB cytotoxin kills human neutrophils by targeting the CD11b subunit of the integrin Mac-1." Proc Natl Acad Sci U S A **110**(26): 10794-10799.
Fair, R. J. and Y. Tor (2014). "Antibiotics and Bacterial Resistance in the 21st Century." Perspectives in Medicinal Chemistry **6**: 25-64.
Fedtke, I., D. Mader, et al. (2007). "A Staphylococcus aureus ypfP mutant with strongly reduced lipoteichoic acid (LTA) content: LTA governs bacterial surface properties and autolysin activity." Molecular Microbiology **65**(4): 1078-1091.
Feng, Z., D. Kallifidas, et al. (2011). "Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites." Proceedings of the National Academy of Sciences of the United States of America **108**(31): 12629-12634.
Ferreira, F. A., R. R. Souza, et al. (2013). "Impact of agr dysfunction on virulence profiles and infections associated with a novel methicillin-resistant Staphylococcus aureus (MRSA) variant of the lineage ST1-SCCmec IV." BMC Microbiol **13**(93): 1471-2180.
Flock, J. I., G. Fröman, et al. (1987). "Cloning and expression of the gene for a fibronectin-binding protein from Staphylococcus aureus." The EMBO Journal **6**(8): 2351-2357.
Foster, T. J. (2005). "Immune evasion by staphylococci." Nat Rev Microbiol **3**(12): 948-958.
Foster, T. J. and M. Höök (1998). "Surface protein adhesins of Staphylococcus aureus." Trends in microbiology **6**(12): 484-488.
Gillespie, D. E., S. F. Brady, et al. (2002). "Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA." Appl Environ Microbiol **68**(9): 4301-4306.
Girardi, M. (2007). "Cutaneous perspectives on adaptive immunity." Clin Rev Allergy Immunol **33**(1-2): 4-14.
Gotz, F. (2002). "Staphylococcus and biofilms." Mol Microbiol **43**(6): 1367-1378.

Groom, A. V., D. H. Wolsey, et al. (2001). "Community-acquired methicillin-resistant Staphylococcus aureus in a rural American Indian community." Jama **286**(10): 1201-1205.
Hackett, S. P. and D. L. Stevens (1992). "Streptococcal toxic shock syndrome: synthesis of tumor necrosis factor and interleukin-1 by monocytes stimulated with pyrogenic exotoxin A and streptolysin O." J Infect Dis **165**(5): 879-885.
Haddadin, A., S. Fappiano, et al. (2002). "Methicillin resistant Staphylococcus aureus (MRSA) in the intensive care unit." Postgraduate Medical Journal **78**(921): 385-392.
Heilmann, C., M. Hussain, et al. (1997). "Evidence for autolysin-mediated primary attachment of Staphylococcus epidermidis to a polystyrene surface." Mol Microbiol **24**(5): 1013-1024.
Heilmann, C. a. G., F (2009). Cell–Cell Communication and Biofilm Formation in Gram-Positive Bacteria, in Bacterial Signaling, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.
Hirschhausen, N., T. Schlesier, et al. (2010). "A novel staphylococcal internalization mechanism involves the major autolysin Atl and heat shock cognate protein Hsc70 as host cell receptor." Cell Microbiol **12**(12): 1746-1764.
Hjort, K., M. Bergstrom, et al. (2010). "Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen-suppressive soil." FEMS Microbiol Ecol **71**(2): 197-207.
Ho, S. W., D. Jung, et al. (2008). "Effect of divalent cations on the structure of the antibiotic daptomycin." Eur Biophys J **37**(4): 421-433.
Holtfreter, S., J. Kolata, et al. (2010). "Towards the immune proteome of Staphylococcus aureus - The anti-S. aureus antibody response." Int J Med Microbiol **300**(2-3): 176-192.
Holtfreter, S., J. Kolata, et al. (2010). "Towards the immune proteome of Staphylococcus aureus - The anti-S. aureus antibody response." Int J Med Microbiol **300**(2-3): 176-192.
Hopwood, D. A. (2006). "Soil to genomics: the Streptomyces chromosome." Annu Rev Genet **40**: 1-23.
Jevons, M. P. (1961.). ""“Celbenin”-resistant staphylococci." BMJ **1**: 124-125.
Johnson, H. M., B. A. Torres, et al. (1996). "Superantigens: structure and relevance to human disease." Proc Soc Exp Biol Med **212**(2): 99-109.
Jonsson, K., C. Signas, et al. (1991). "Two different genes encode fibronectin binding proteins in Staphylococcus aureus. The complete nucleotide sequence and characterization of the second gene." Eur J Biochem **202**(3): 1041-1048.
Julian, K., K. Kosowska-Shick, et al. (2007). "Characterization of a daptomycin-nonsusceptible vancomycin-intermediate Staphylococcus aureus strain in a patient with endocarditis." Antimicrob Agents Chemother **51**(9): 3445-3448.
Kalil, A. C., M. Klompas, et al. (2013). "Treatment of hospital-acquired pneumonia with linezolid or vancomycin: a systematic review and meta-analysis." BMJ Open **3**(10): e003912.
Kang, H. S. and S. F. Brady (2014). "Mining soil metagenomes to better understand the evolution of natural product structural diversity: pentangular polyphenols as a case study." J Am Chem Soc **136**(52): 18111-18119.
Kantyka, T., L. N. Shaw, et al. (2011). "Papain-like proteases of Staphylococcus aureus." Adv Exp Med Biol **712**: 1-14.
Kapur, S., A. Worthington, et al. (2008). "Mechanism based protein crosslinking of domains from the 6-deoxyerythronolide B synthase." Bioorg Med Chem Lett **18**(10): 3034-3038.

Karakawa, W. W., A. Sutton, et al. (1988). "Capsular antibodies induce type-specific phagocytosis of capsulated Staphylococcus aureus by human polymorphonuclear leukocytes." Infection and Immunity **56**(5): 1090-1095.
Kennedy, A. D., J. Bubeck Wardenburg, et al. (2010). "Targeting of alpha-hemolysin by active or passive immunization decreases severity of USA300 skin infection in a mouse model." J Infect Dis **202**(7): 1050-1058.
Ki, V. and C. Rotstein (2008). "Bacterial skin and soft tissue infections in adults: A review of their epidemiology, pathogenesis, diagnosis, treatment and site of care." The Canadian Journal of Infectious Diseases & Medical Microbiology **19**(2): 173-184.
Kirby, W. M. (1944). "Extraction of a Highly Potent Penicillin Inactivator from Penicillin Resistant Staphylococci." Science **99**(2579): 452-453.
Knietsch, A., S. Bowien, et al. (2003). "Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures." Appl Environ Microbiol **69**(6): 3048-3060.
Kobayashi, S. D., K. R. Braughton, et al. (2003). "Bacterial pathogens modulate an apoptosis differentiation program in human neutrophils." Proc Natl Acad Sci U S A **100**(19): 10948-10953.
Krishna, S. and L. S. Miller (2012). "Innate and adaptive immune responses against Staphylococcus aureus skin infections." Semin Immunopathol **34**(2): 261-280.
Kristian, S. A., T. Golda, et al. (2004). "The ability of biofilm formation does not influence virulence of Staphylococcus aureus and host response in a mouse tissue cage infection model." Microb Pathog **36**(5): 237-245.
Kubica, M., K. Guzik, et al. (2008). "A potential new pathway for Staphylococcus aureus dissemination: the silent survival of S. aureus phagocytosed by human monocyte-derived macrophages." PLoS One **3**(1): 0001409.
Lambris, J. D., D. Ricklin, et al. (2008). "Complement evasion by human pathogens." Nature reviews. Microbiology **6**(2): 132.
Larsen, N., L. Thorsen, et al. (2014). "Characterization of Bacillus spp. strains for use as probiotic additives in pig feed." Appl Microbiol Biotechnol **98**(3): 1105-1118.
Lee, L. Y., M. Hook, et al. (2004). "Inhibition of complement activation by a secreted Staphylococcus aureus protein." J Infect Dis **190**(3): 571-579.
Lee, L. Y., X. Liang, et al. (2004). "Identification and characterization of the C3 binding domain of the Staphylococcus aureus extracellular fibrinogen-binding protein (Efb)." J Biol Chem **279**(49): 50710-50716.
Leung, D. Y., J. B. Travers, et al. (1995). "The role of superantigens in skin disease." J Invest Dermatol **105**(1 Suppl): 37S-42S.
Lewis, K. (2013). "Platforms for antibiotic discovery." Nat Rev Drug Discov **12**(5): 371-387.
Liles, M. R., L. L. Williamson, et al. (2008). "Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms." Appl Environ Microbiol **74**(10): 3302-3305.
Ling, L. L., T. Schneider, et al. (2015). "A new antibiotic kills pathogens without detectable resistance." Nature **517**(7535): 455-459.
Liu, C., A. Bayer, et al. (2011). "Clinical Practice Guidelines by the Infectious Diseases Society of America for the Treatment of Methicillin-Resistant Staphylococcus Aureus Infections in Adults and Children." Clinical Infectious Diseases.

Liu, G., K. F. Chater, et al. (2013). "Molecular regulation of antibiotic biosynthesis in streptomyces." Microbiol Mol Biol Rev **77**(1): 112-143.
Liu, G. Y. (2009). "Molecular Pathogenesis of Staphylococcus aureus Infection." Pediatric research **65**(5 Pt 2): 71R-77R.
Lorenz, P. and J. Eck (2005). "Metagenomics and industrial applications." Nat Rev Microbiol **3**(6): 510-516.
Lowy, F. D. (1998). "Staphylococcus aureus infections." N Engl J Med **339**(8): 520-532.
Lowy, F. D. (2003). "Antimicrobial resistance: the example of Staphylococcus aureus." Journal of Clinical Investigation **111**(9): 1265-1273.
Mack, D., W. Fischer, et al. (1996). "The intercellular adhesin involved in biofilm accumulation of Staphylococcus epidermidis is a linear beta-1,6-linked glucosaminoglycan: purification and structural analysis." J Bacteriol **178**(1): 175-183.
Manders, S. M. (1998). "Toxin-mediated streptococcal and staphylococcal disease." J Am Acad Dermatol **39**(3): 383-398.
Marrack, P. and J. Kappler (1990). "The staphylococcal enterotoxins and their relatives." Science **248**(4956): 705-711.
Marraffini, L. A., A. C. DeDent, et al. (2006). "Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria." Microbiology and Molecular Biology Reviews **70**(1): 192-221.
Mazza, P. (1994). "The use of Bacillus subtilis as an antidiarrhoeal microorganism." Boll Chim Farm **133**(1): 3-18.
McAleese, F. M., E. J. Walsh, et al. (2001). "Loss of clumping factor B fibrinogen binding activity by Staphylococcus aureus involves cessation of transcription, shedding and cleavage by metalloprotease." J Biol Chem **276**(32): 29969-29978.
Menzies, B. E. (2003). "The role of fibronectin binding proteins in the pathogenesis of Staphylococcus aureus infections." Curr Opin Infect Dis **16**(3): 225-229.
Miller, L. G. and B. A. Diep (2008). "Clinical practice: colonization, fomites, and virulence: rethinking the pathogenesis of community-associated methicillin-resistant Staphylococcus aureus infection." Clin Infect Dis **46**(5): 752-760.
Miller, L. G., F. Perdreau-Remington, et al. (2005). "Necrotizing fasciitis caused by community-associated methicillin-resistant Staphylococcus aureus in Los Angeles." N Engl J Med **352**(14): 1445-1453.
Moellering, R. C. (2008). "Current Treatment Options for Community-Acquired Methicillin-Resistant Staphylococcus aureus Infection." Clinical Infectious Diseases **46**(7): 1032-1037.
Morales, G., J. J. Picazo, et al. (2010). "Resistance to linezolid is mediated by the cfr gene in the first report of an outbreak of linezolid-resistant Staphylococcus aureus." Clin Infect Dis **50**(6): 821-825.
Morens, D. M. and A. S. Fauci (2012). "Emerging infectious diseases in 2012: 20 years after the institute of medicine report." MBio **3**(6): 00494-00412.
Morens, D. M., G. K. Folkers, et al. (2010). "The challenge of emerging and re-emerging infectious diseases." Nature **463**(7277): 122-122.
Murthy, M. H., M. E. Olson, et al. (2008). "Daptomycin non-susceptible meticillin-resistant Staphylococcus aureus USA 300 isolate." J Med Microbiol **57**(Pt 8): 1036-1038.
Naber, C. K. (2009). "Staphylococcus aureus bacteremia: epidemiology, pathophysiology, and management strategies." Clin Infect Dis **15**(48): 598189.

Nanra, J. S., S. M. Buitrago, et al. (2013). "Capsular polysaccharides are an important immune evasion mechanism for Staphylococcus aureus." Human Vaccines & Immunotherapeutics **9**(3): 480-487.
Nikolouli, K. and D. Mossialos (2012). "Bioactive compounds synthesized by non-ribosomal peptide synthetases and type-I polyketide synthases discovered through genome-mining and metagenomics." Biotechnol Lett **34**(8): 1393-1403.
Novick, R. P. (2000). Pathogenicity factors and their regulation. In: Gram-Positive Pathogens Washington, D.C, American Society for Microbiology,.
O'Riordan, K. and J. C. Lee (2004). "Staphylococcus aureus Capsular Polysaccharides." Clinical Microbiology Reviews **17**(1): 218-234.
O'Shea, J. J. and W. E. Paul (2010). "Mechanisms underlying lineage commitment and plasticity of helper CD4+ T cells." Science **327**(5969): 1098-1102.
Owen, G. A., B. Pascoe, et al. (2007). "Zinc-responsive regulation of alternative ribosomal protein genes in Streptomyces coelicolor involves zur and sigmaR." J Bacteriol **189**(11): 4078-4086.
Palma, M., O. Shannon, et al. (2001). "Extracellular fibrinogen-binding protein, Efb, from Staphylococcus aureus blocks platelet aggregation due to its binding to the alpha-chain." J Biol Chem **276**(34): 31691-31697.
Panizzi, P., M. Nahrendorf, et al. (2011). "In vivo detection of Staphylococcus aureus endocarditis by targeting pathogen-specific prothrombin activation." Nat Med **17**(9): 1142-1146.
Pannaraj, P. S., K. G. Hulten, et al. (2006). "Infective pyomyositis and myositis in children in the era of community-acquired, methicillin-resistant Staphylococcus aureus infection." Clin Infect Dis **43**(8): 953-960.
Patti, J. M., H. Jonsson, et al. (1992). "Molecular characterization and expression of a gene encoding a Staphylococcus aureus collagen adhesin." J Biol Chem **267**(7): 4766-4772.
Pendleton, J. N., S. P. Gorman, et al. (2013). "Clinical relevance of the ESKAPE pathogens." Expert Rev Anti Infect Ther **11**(3): 297-308.
Périchon, B. and P. Courvalin (2004). "Heterologous Expression of the Enterococcal vanA Operon in Methicillin-Resistant Staphylococcus aureus." Antimicrobial Agents and Chemotherapy **48**(11): 4281-4285.
Piel, J. (2011). "Approaches to capturing and designing biologically active small molecules produced by uncultured microbes." Annu Rev Microbiol **65**: 431-453.
Ramirez, P., L. Fernandez-Barat, et al. (2012). "New therapy options for MRSA with respiratory infection/pneumonia." Curr Opin Infect Dis **25**(2): 159-165.
Rammelkamp, C. H. and T. Maxon (1942). "Resistance of Staphylococcus aureus to the Action of Penicillin." Exp Biol Med **51**(3): 386-389
Reygaert, W. C. (2013). Antimicrobial resistance mechanisms of Staphylococcus aureus.
Rhee, J. K., D. G. Ahn, et al. (2005). "New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library." Appl Environ Microbiol **71**(2): 817-825.
Rigby, K. M. and F. R. DeLeo (2012). "Neutrophils in innate host defense against Staphylococcus aureus infections." Seminars in Immunopathology **34**(2): 237-259.
Rivera, A. M. and H. W. Boucher (2011). "Current Concepts in Antimicrobial Therapy Against Select Gram-Positive Organisms: Methicillin-Resistant Staphylococcus aureus,

Penicillin-Resistant Pneumococci, and Vancomycin-Resistant Enterococci." Mayo Clinic Proceedings **86**(12): 1230-1243.
Rohde, H., J. K. Knobloch, et al. Correlation of Staphylococcus aureus icaADBC genotype and biofilm expression phenotype, J Clin Microbiol. 2001 Dec;39(12):4595-6.
Rubinstein, E., M. H. Kollef, et al. (2008). "Pneumonia caused by methicillin-resistant Staphylococcus aureus." Clin Infect Dis **1**(46): 533594.
Rudd, B. A. and D. A. Hopwood (1979). "Genetics of actinorhodin biosynthesis by Streptomyces coelicolor A3(2)." J Gen Microbiol **114**(1): 35-43.
Rudd, B. A. and D. A. Hopwood (1980). "A pigmented mycelial antibiotic in Streptomyces coelicolor: control by a chromosomal gene cluster." J Gen Microbiol **119**(2): 333-340.
Saito, M., Y. Katayama, et al. (2014). "“Slow VISA,” a Novel Phenotype of Vancomycin Resistance, Found In Vitro in Heterogeneous Vancomycin-Intermediate Staphylococcus aureus Strain Mu3." Antimicrobial Agents and Chemotherapy **58**(9): 5024-5035.
Sanders, M. E. (2008). "Probiotics: definition, sources, selection, and uses." Clin Infect Dis **1**(46): 523341.
Saravolatz, L. D., N. Markowitz, et al. (1982). "Methicillin-resistant Staphylococcus aureus. Epidemiologic observations during a community-acquired outbreak." Ann Intern Med **96**(1): 11-16.
Scapini, P., J. A. Lapinet-Vera, et al. (2000). "The neutrophil as a cellular source of chemokines." Immunol Rev **177**: 195-203.
Schiffrin, E. J., F. Rochat, et al. (1995). "Immunomodulation of human blood cells following the ingestion of lactic acid bacteria." J Dairy Sci **78**(3): 491-497.
Schlievert, P. M., K. N. Shands, et al. (1981). "Identification and characterization of an exotoxin from Staphylococcus aureus associated with toxic-shock syndrome." J Infect Dis **143**(4): 509-516.
Schlievert, P. M., K. L. Strandberg, et al. (2010). "Secreted virulence factor comparison between methicillin-resistant and methicillin-sensitive Staphylococcus aureus, and its relevance to atopic dermatitis." Journal of Allergy and Clinical Immunology **125**(1): 39-49.
Schlievert, P. M., K. L. Strandberg, et al. (2010). "Secreted Virulence Factor Comparison Between Methicillin-Resistant and Methicillin-Sensitive Staphylococcus aureus, and its Relevance to Atopic Dermatitis." The Journal of allergy and clinical immunology **125**(1): 39.
Schwarz, S., C. Werckenthin, et al. (2000). "Identification of a plasmid-borne chloramphenicol-florfenicol resistance gene in Staphylococcus sciuri." Antimicrob Agents Chemother **44**(9): 2530-2533.
Scott, W. R., S. B. Baek, et al. (2007). "NMR structural studies of the antibiotic lipopeptide daptomycin in DHPC micelles." Biochim Biophys Acta **12**(26): 15.
Shannon, O. and J. I. Flock (2004). "Extracellular fibrinogen binding protein, Efb, from Staphylococcus aureus binds to platelets and inhibits platelet aggregation." Thromb Haemost **91**(4): 779-789.
Shaw, L., E. Golonka, et al. (2004). "The role and regulation of the extracellular proteases of Staphylococcus aureus." Microbiology **150**(Pt 1): 217-228.
Shopsin, B., A. Drlica-Wagner, et al. (2008). "Prevalence of agr dysfunction among colonizing Staphylococcus aureus strains." J Infect Dis **198**(8): 1171-1174.
Sieradzki, K., R. B. Roberts, et al. (1999). "The development of vancomycin resistance in a patient with methicillin-resistant Staphylococcus aureus infection." N Engl J Med **340**(7):

517-523.
Smolinski, M. S., Hamburg, and J. M.A. and Lederberg (2003). Microbial threats to health: emergence, detection, and response,. Institu
te of Medicine of the National Academies, Washington, DC, The National Academy Press.
Speziale, P., G. Pietrocola, et al. (2009). "Structural and functional role of Staphylococcus aureus surface components recognizing adhesive matrix molecules of the host." Future microbiology **4**(10): 1337-1352.
Stein, J. L., T. L. Marsh, et al. (1996). "Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon." J Bacteriol **178**(3): 591-599.
Stevens, D. L., M. H. Tanner, et al. (1989). "Severe group A streptococcal infections associated with a toxic shock-like syndrome and scarlet fever toxin A." N Engl J Med **321**(1): 1-7.
Stewart, E. J. (2012). "Growing unculturable bacteria." J Bacteriol **194**(16): 4151-4160.
Straus, S. K. and R. E. Hancock (2006). "Mode of action of the new antibiotic for Gram-positive pathogens daptomycin: comparison with cationic antimicrobial peptides and lipopeptides." Biochim Biophys Acta **9**(23): 3.
Sung, J. M., D. H. Lloyd, et al. (2008). "Staphylococcus aureus host specificity: comparative genomics of human versus animal isolates by multi-strain microarray." Microbiology **154**(Pt 7): 1949-1959.
Taupp, M., K. Mewis, et al. (2011). "The art and design of functional metagenomic screens." Curr Opin Biotechnol **22**(3): 465-472.
Thomason, L. C., N. Costantino, et al. (2007). "Multicopy plasmid modification with phage lambda Red recombineering." Plasmid **58**(2): 148-158.
Tracey, W. (2014). "FDA approves a new antibiotic to treat MRSA."
Tsiodras, S., H. S. Gold, et al. Linezolid resistance in a clinical isolate of Staphylococcus aureus, Lancet. 2001 Jul 21;358(9277):207-8.
Uchiyama, T., T. Abe, et al. (2005). "Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes." Nat Biotechnol **23**(1): 88-93.
Uchiyama, T. and K. Miyazaki (2009). "Functional metagenomics for enzyme discovery: challenges to efficient screening." Curr Opin Biotechnol **20**(6): 616-622.
van Leeuwen, W. B., D. C. Melles, et al. (2005). "Host- and Tissue-Specific Pathogenic Traits of Staphylococcus aureus." Journal of Bacteriology **187**(13): 4584-4591.
Vandenesch, F., T. Naimi, et al. (2003). "Community-acquired methicillin-resistant Staphylococcus aureus carrying Panton-Valentine leukocidin genes: worldwide emergence." Emerg Infect Dis **9**(8): 978-984.
Verschuere, L., G. Rombaut, et al. (2000). "Probiotic Bacteria as Biological Control Agents in Aquaculture." Microbiology and Molecular Biology Reviews **64**(4): 655-671.
Vikram, H. R., N. L. Havill, et al. (2005). "Clinical progression of methicillin-resistant Staphylococcus aureus vertebral osteomyelitis associated with reduced susceptibility to daptomycin." J Clin Microbiol **43**(10): 5384-5387.
Vincent, J. L., D. J. Bihari, et al. (1995). "The prevalence of nosocomial infection in intensive care units in Europe. Results of the European Prevalence of Infection in Intensive Care (EPIC) Study. EPIC International Advisory Committee." Jama **274**(8): 639-644.
Voget, S., H. L. Steele, et al. (2006). "Characterization of a metagenome-derived halotolerant cellulase." J Biotechnol **126**(1): 26-36.

von Eiff, C., G. Peters, et al. (2006). "The small colony variant (SCV) concept -- the role of staphylococcal SCVs in persistent infections." Injury **37**(2): S26-33.
Voyich, J. M., K. R. Braughton, et al. (2005). "Insights into mechanisms used by Staphylococcus aureus to avoid destruction by human neutrophils." J Immunol **175**(6): 3907-3919.
Vuong, C., H. L. Saenz, et al. (2000). "Impact of the agr quorum-sensing system on adherence to polystyrene in Staphylococcus aureus." J Infect Dis **182**(6): 1688-1693.
Wang, L., Nasrin, S., Liles, M. R. and Yu, Z. (2012). "Use of Bacterial Artificial Chromosomes in Metagenomics Studies, Overview " Encyclopedia of Metagenomics: 1-12.
Wang, R., K. R. Braughton, et al. (2007). "Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA." Nat Med **13**(12): 1510-1514.
White, J., A. Herman, et al. (1989). "The V beta-specific superantigen staphylococcal enterotoxin B: stimulation of mature T cells and clonal deletion in neonatal mice." Cell **56**(1): 27-35.
Wilkinson, B. and J. Micklefield (2007). "Mining and engineering natural-product biosynthetic pathways." Nat Chem Biol **3**(7): 379-386.
Wilkinson BJ (1997). The staphylococci in human disease. New York: Churchill Livingstone.
Williams, R. J., B. Henderson, et al. (2002). "Identification of a fibronectin-binding protein from Staphylococcus epidermidis." Infect Immun **70**(12): 6805-6810.
Xia, G., T. Kohler, et al. (2010). "The wall teichoic acid and lipoteichoic acid polymers of Staphylococcus aureus." Int J Med Microbiol **300**(2-3): 148-154.
Xu, S. X. and J. K. McCormick (2012). "Staphylococcal superantigens in colonization and disease." Frontiers in Cellular and Infection Microbiology **2**: 52.
Yadav, G., R. S. Gokhale, et al. (2003). "Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases." J Mol Biol **328**(2): 335-363.
Zhang, H. Z., C. J. Hackbarth, et al. (2001). "A proteolytic transmembrane signaling pathway and resistance to beta-lactams in staphylococci." Science **291**(5510): 1962-1965.

Chapter 2

Isolation of novel soil *Actinobacteria* that express antibacterial activity against methicillin resistant *Staphylococcus aureus*

1. Abstract

The emergence of multidrug-resistant pathogens has increased the demand for discovery of novel antibiotics. Soil microbial communities are a great resource for natural products but a majority of them have not been explored due to the recalcitrance of many bacterial taxa to laboratory cultivation. We isolated a collection of 548 bacterial and fungal isolates from soil using low-strength (1/200th) nutrient agar supplemented with soil extract incubated for more than three months at room temperature. Bacterial diversity analysis using 16S rRNA gene sequences of newly cultured isolates revealed that they represent diverse bacterial genera affiliated with the phyla *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*. Two isolates, designated as A115 and F4, were found to inhibit the growth of pathogenic methicillin-resistant *Staphylococcus aureus* (MRSA). The isolate A115, member of the genus *Streptomyces*, produces pink pigments after incubation for more than ten days. The isolate F4, identified as a *Nonomuraea* spp., produces a high molecular weight (>100kDa), heat stable reddish pigment associated with anti-MRSA activity. Genome sequencing using a combination of shotgun and mate-pair next-generation sequencing resulted in the complete assembled genome for each isolate, with the size of the A115 and F4 genomes at 8.6 Mb and 10.3 Mb, respectively. The %G+C contents of strains A115 and F4 were determined to be 71% and 70.4%, respectively. Phylogenetic analysis using multilocus sequence analysis with six housekeeping genes revealed that strain A115 was most closely related to *Streptomyces afghaniensis* and *Streptomyces olindensis*; however, the low level of average nucleotide identity (ANI) values in comparing the

A115 genome were 89.76% and 89.14% for *S. afghaniensis* and *S. olindensis*, respectively. These genomic results, combined with differentiation of strain A115 from other *Streptomyces* species by morphological and physiological characteristics, led to the conclusion that strain A115 represents a novel species of the genus *Streptomyces*, for which the name *Streptomyces alburnustigris* sp. nov. is proposed. Phylogenetic analysis based on 16S rRNA gene sequence revealed that the closest phylogenetic relative of F4 strain was *Nonomuraea antimicrobica* YIM 61105. *In silico* analysis using anti-SMASH predicts that the A115 and F4 genomes encode many gene clusters for secondary metabolite biosynthesis, including the synthesis of terpene, aminoglycoside, thiopeptide, bacteriocin, oligosaccharide, phenazine, butyrolactone, siderophore, melanine and potentially other bioactive compounds produced by non-ribosomal peptide synthetase and polyketide synthetase pathways. Both *S. alburnustigris* A115 and *Nonomuraea* spp. strain F4 genomes are predicted to encode Type I, II, and III PKS pathways. The biochemical structure of the active anti-MRSA compounds are currently being characterized using liquid chromatography–mass spectrometry (LC/MS). This study identified novel bacterial isolates with anti-MRSA activity and demonstrates the utility of novel cultivation techniques in obtaining previously uncultured and phylogenetically diverse soil microorganisms, some of which express potent bioactive secondary metabolites.

2. Introduction

The phylum *Actinobacteria* is comprised of diverse Gram-positive taxa that have a high %G+C content and exhibit varied morphologies, physiologies, and metabolic properties such as the production of extracellular enzymes and the synthesis of a wide variety of secondary metabolites. Members of *Actinobacteria* have very diverse lifestyles, ranging from pathogens (e.g., *Corynebacterium*, *Mycobacterium*, *Nocardia*, *Tropheryma*, and *Propionibacterium*) to

antibiotic-producing soil inhabitants (*Streptomyces*, *Nonomureae*) and plant commensals *(Leifsonia*), or gastrointestinal commensals (*Bifidobacterium*). Furthermore, they adopt complex life cycles that consist of vegetative growth followed by the formation of aerial hyphae and finally exospore formation. Spore formation allows them to disperse and persist in diverse environmental niches. Most members of this phylum are ubiquitous in terrestrial and aquatic ecosystems and play an important role in the production of natural products with pharmaceutical applications and recycling of organic matter by decomposition.

Actinobacteria constitute a significant portion of soil microflora, estimating that a gram of fresh soil contains about 10^9 CFU/g of bacteria and of which 10^7 are *Actinobacteria* taxa (Steffan et al. 1988; Weinbauer et al. 1998). Therefore, diverse soil samples can be an overwhelmingly rich reservoir for discovering of bioactive compounds. The microorganisms belonging to the order Actinomycetales are fascinatingly diverse for their ability to produce biologically active secondary metabolites. It has been estimated that the order Actinomycetales has yielded ~3000 antibiotics since after its first reporting of streptomycin in 1942 (Waksman and Woodruff 1941; Watve, Tickoo et al. 2001). The genus *Streptomyce*s itself produced about 90% of these antibiotics including vancomycin, erythromycin and tetracycline (Watve, Tickoo et al. 2001). It has also been reported that *Streptomyces* spp. have the ability to synthesize important antifungal (amphotericin B) (Caffrey, Aparicio et al. 2008), anticancer (mitomycin C) (Olano, Mendez et al. 2009), antiparasitic (ivermectin) (Nett, Ikeda et al. 2009) and immunosuppressive (rapamycin) agents (Graziani 2009).

In addition to *Streptomyces*, some rare *Actinobacteria* including *Nonomuraea* (Tiwari et al. 2012) can be potential producers of secondary metabolites with antimicrobial activity (Cornaglia and Rossolini 2009). For instance, *Nonomuraea* sp. ATCC 39727 produces the

teicoplanin-like glycopeptide antibiotic A40926 which is a precursor for semi-synthetic derivative dalbavancin, an antibiotic that is currently in clinical use (Boucher et al. 2014). Like most *Actinobacteria*, *Nonomuraea* spp. also produce various shades of blue, violet, red, rose, yellow, green, brown and black pigments on natural and synthetic media and the pigments maybe dissolved in to the medium or it may be retained in the mycelium. The pigments produced from *Actinobacteria* have been used on cotton shades (Perumal, Stalin et al. 2009) and in medicine, pharmacology and cosmetic preparations (Perumal, Stalin et al. 2009). Moreover, several reports indicated that antimicrobial activities and/or antibiotic production of *Actinobacteria* are associated with pigment production (Miyaura and Tatsumi 1961). It has been estimated that (Watve, Tickoo et al. 2001) members of this phylum can potentially produce around 100,000 antimicrobial metabolites and this estimate demonstrated that only a small percentage (~3%) of the extant antibiotics that have already been discovered (Watve et al. 2001). To discover the remaining antibiotics, we need to get access to a greater diversity of bacteria. Those diverse bacterial sources can be tapped for antibiotics by expanded conventional culturing approach, novel culture methods, heterologous DNA-based methods and metagenomics.

The discovery of new antibiotics using a culture-based approach is still the most historically successful approach (Fleming 1929; Debono, Abbott et al. 1988; Jang, Nam et al. 2013), since alternative strategies such as metagenomics, combinatorial biosynthesis and fragment-based drug design have yet to yield large numbers of novel chemical entities with antimicrobial activity. However, although there has been a wealth of antibiotics discovered from cultured soil microbes (Thiele-Bruhn 2003), the past several decades have experienced diminishing success rates for antibiotic discovery using a culture-based approach. This is due to the very high rate of antibiotic re-discovery (>99%) when screening cultured bacteria grown

under “normal” laboratory conditions, i.e. high nutrient levels with short incubation times. Culture-based approaches have recently gotten new momentum due to technology advancements such as employing novel culturing techniques, i.e. low-nutrient media with extended incubation time (Hamaki et al. 2005) or use of a multichannel device iChip for simultaneous isolation and culturing of as-yet unculturable bacteria (Ling et al. 2015), applying next-generation sequencing and high-throughput screening (Baltz 2008).

The recent advancement of genome sequencing technologies has revealed that actinomycetes have a much greater potential for secondary metabolite production than first assumed. Complete genome sequences of available *Actinobacteria* indicated that they contain many more secondary metabolite biosynthetic gene clusters than the number of actually identified metabolites would suggest. Genome mining revealed that a single isolate has the genetic potential to synthesize more than one secondary metabolite; however, the probability of discovering a novel compound can be far greater if unique isolates are screened simultaneously. Thus, in this research project a combination of selective isolation and screening procedures for the collection of novel and/or rare Actinobacteria from unexplored soil samples were used for discovery of novel compounds with antibacterial activity against methicillin resistant *Staphylococcus aureus*.

3. Materials and methods

Collection of soil samples

To isolate soil microorganisms with antibacterial activity against clinical MRSA strain, soil samples were collected from three different sources including the Cullars Rotation, the Auburn University Arboretum and a long-leaf pine forest sample that represented three characteristic soils in the State of Alabama. The Cullars Rotation has a sandy agricultural soil

and for this study, the soil sample was collected from a soil plot that had not been amended with fertilizers for the past 100 years. Soil from Arboretum was the Black Belt soil which is very rich and alkaline in nature, and contained a considerable amount of montmorillonite clay. The forest soil obtained from Long Leaf Pine forest at Auburn contained high clay abundance characteristic of many Appalachian soils. All samples were collected from the top 10-cm layer of soil, homogenized, and sieved to eliminate plant roots and other debris. Soil samples were diluted in sterile water and extracts were plated immediately into 1/200 strength Nutrient Agar (NA) plates after sampling.

Cultivation of soil microorganisms

Soil microorganisms were isolated based on the method described previously (George et al. 2011). Briefly, two types of solid media were used for the cultivation of soil microorganisms: 1) dilute nutrient broth (0.065g/L) which is 200 times diluted than the manufacturer's instructions ($1/200^{th}$ NA) supplemented with 15 g/L agar; and 2) $1/200^{th}$ diluted nutrient agar (NA) supplemented soil extract (SE) ($1/200^{th}$). The soil extract was prepared from three soil samples according to the protocol described by (George et al. 2011). To cultivate soil microorganisms, 1.0 g of soil from Cullars rotation, arboretum or long-leaf pine forest was added into 99.0 mL of sterile milliQ water and stirred for 2 hours at 200 rpm. After settling for 1 hour, the supernatant was serially diluted up to 10^{-7} dilutions and aliquots of 100 µL of each dilution were spread onto $1/200^{th}$ NA and $1/200^{th}$ NA+SE plates with 4 replicates per dilution. Plates were incubated for up to 3 months at room temperature (approximately 22°C). After incubation, each colony was subcultured onto $1/200^{th}$ and $1/100^{th}$ diluted NA plates in duplicates. Each isolate was then subjected for phylogenetic analysis using 16S rRNA gene sequences and screening for antibacterial activity against clinical MRSA strain 30.

Phylogenetic analysis of soil isolates

Each of the pure culture isolate grown on 1/200th NA was subjected to phylogenetic analysis using 16S rRNA gene-specific sequences. Colony PCR was performed on each individual colony with universal bacterial primers 27F and 1492R which generated approximately 1.5 kb products. PCR reactions were performed in 50 µL reaction volumes which contained 25.0 µL of 2× EconoTaq plus Green DNA Polymerase (Lucigen Co. WI), 0.2 µM of each primer and sterile nuclease free water for adjusting to 50 µL. Amplification of the 16S rRNA gene was carried out under the following conditions: denaturation at 94°C for 2 min followed by 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 2.0 min and final extension at 72°C for 10 min. Amplified PCR products of bacterial isolates were analyzed by electrophoresis with 0.7% agarose gel run at 200V for an hour. After electrophoresis gel was stained by ethidium bromide and then visualized and photographed under UV transilluminator, Gel Doc XR system (Biorad USA). The PCR products were purified by E.Z.N.A cycle pure kit (Omega bio-tak USA) according to the manufacturer's instructions and purified PCR products were then sequenced in both directions using primers 27F and 1492R (Lucigen Corp, Middleton, WI USA). The sequences were trimmed for quality using ChromasPro (Technelysium, Australia). Trimmed sequences were then assembled and compared against sequences available in the National Center for Biotechnology Information (NCBI data base), Genbank, using the BLASTn algorithm.

Isolation of soil microorganisms with antibacterial activity

The antibiosis assay was carried out by double-layer soft agar method with minor modifications (Jack et al. 1996). Briefly, each of the soil microorganisms were grown on 1/200th NA plates for 3 months at room temperature. After incubation, soft agar (0.7% w/v agar) was

prepared with NA was melted, cooled and seeded with a freshly prepared inoculum of log-phase MRSA strain 30 to achieve the absorbance at 600 nm (OD_{600}) of 0.5. The bacterial cell suspension in soft agar was immediately poured over the $1/200^{th}$ NA plates and incubated for 24 hours at 37°C. After incubation, the presence of zones of inhibition in the growth of MRSA strain 30 were recorded (in mm) as evidence of growth inhibition caused by soil isolates. Soil bacterial isolates that inhibited the growth of MRSA strain 30 were subcultured on to $1/200^{th}$, $1/100^{th}$, $1/50^{th}$, and $1/10^{th}$ NA and nutrient broth (NB) for further analysis.

To verify the anti-MRSA activity, all soil isolates that showed antibacterial activity in primary screen were re-tested via drop assay. In this assay soil isolates that showed antibacterial activities against MRSA strain 30 from the preliminary selection were grown in $1/10^{th}$ NB for one month at room temperature. After incubation, supernatants were collected and frozen for further analysis. A broth culture of actively growing MRSA strain 30 was adjusted to OD_{600} of 0.5 and evenly swabbed onto TSA plates. Then 10 µL of supernatants derived from soil cultures were added onto the MRSA strain 30 culture. Zones of inhibition were measured after 24 hours of incubation at 37°C. The soil isolates which showed repeated antibacterial activities in drop assay were selected as positive isolates and maintained the cultures in $1/10^{th}$ NB for further experiments. An aliquot of supernatants obtained from each soil isolates were also shipped to the National Center for Natural Products Research (NCNPR), the University of Mississippi for screening against a larger collection of bacterial and fungal pathogens.

Antibacterial activity of *Streptomyces* sp. strain A115 and *Nonomuraea* sp. strain F4

The effect of incubation time and concentrations of A115 and F4 supernatants on the growth of MRSA strain 30 was determined by the microtiter plate assay as described previously (Rufián-Henares et al. 2008). Both A115 and F4 isolates were grown in 10 mL of $1/10^{th}$ NB at room temperature. Each week 2 mL of A115 and F4 supernatants were evaluated for anti-MRSA activity by using double dilution method. For antibacterial assay, the overnight culture of MRSA strain 30 was diluted 1:100 into a fresh broth and incubated further (~2 to 3 hours) to achieve absorbance at 600 nm (OD_{600}) of 0.5 which are equivalent to 1.4×10^8 CFU/mL. Then 50µL of bacterial cell suspension was inoculated into a sterile 96-well microtiter plate containing A115 or F4 supernatant and plates were incubated for 24 hrs at 37°C. Turbidity of MRSA strain 30 was measured as absorbance at 600 nm by Gen5 spectrophotometer (BioTek Instruments, VT, USA).

Extraction of anti-MRSA compounds

Two of the bacterial isolates *Streptomyces* sp. A115 and *Nonomuraea* sp. F4 with anti-MRSA activity were grown in $1/10^{th}$ NB at room temperature for one month. The active supernatant of A115 isolate was collected by centrifugation followed by filtration. Filtered supernatants were stored at -80°C and an aliquot of the supernatant was sent to the National Center for Natural Product Research (NCNPR, Oxford, MS) according to the standard protocol established by the NCNPR for determining the biochemical structure of the active compound(s) and further screening against a large collection of bacterial and fungal pathogens. The active supernatant of F4 isolate was extracted with five different amberlite resins including XAD-4, XAD-7, XAD-16, XAD-1180N and an anionic resin DE52 (Sigma-Aldrich, USA). For separating active compound(s), a 25 mL syringe was packed with approximately 8 g of XAD resin. Then 50 mL of F4 supernatant was passed through the XAD column and then column was

washed 3 times with water. Finally, adsorbed compounds were eluted with 90% ethanol and 10% acetic acid. Ten fractions with approximately 10 mL of each were manually collected followed by concentrated with Vacufuge (Eppendorf, USA). Each concentrated (500×) fraction was tested for Anti-MRSA activity before they were analyzed by HPLC method. The active fractions were separated by HPLC and each fraction was tested separately or pooled together for anti-MRSA activity.

Morphological characteristics of anti-MRSA compound producing soil bacteria

Morphological characteristics of *Streptomyces* sp. A115 and *Nonomuraea* sp. F4 were examined using light and scanning electron microscopy of colonies grown on $1/10^{th}$ NA and $1/10^{th}$ NB after incubated for 21 days at room temperature. For high-resolution scanning electron microscopy, a colony with agar or 10 µL of bacterial suspension was added onto a double sided sticker coated aluminum stubs and dried completely by air drying. The dried samples were coated with gold alloy by a 108 Auto/SE Sputter Coater (Cressington EM Vacuum Coating Systems, USA), and examined with a Zeiss EVO 50VP (Germany) scanning electron microscope.

The colony morphology and anti-MRSA activity of strains A115 and F4 were observed on several standard media that included i) International Streptomyces Project 4 (ISP 4), ii) ISP 2 supplemented with 5% NaCI, iii) $1/10^{th}$ NA supplemented with 0.2% N-acetylglucosamine, iv) $1/10^{th}$ NA with or without 1.5% NaCI, v) SYZ media with or without artificial sea water, vi) R2YE (Shepherd et al. 2010) and vii) modified YEME media (Shepherd et al. 2010) after 21 days of incubation at room temperature.

Biochemical and physiological characteristics of actinobacterial strains A115 and F4

Biochemical and physiological characteristics of the actinobacterial isolates A115 and F4 were determined based on the method described previously (Shirling et al. 1966). Biochemical tests including Gram's reaction, MR-VP, H_2S production, nitrate reduction, oxidase, catalase, urease, blood hemolysis, TSI, citrate utilization, starch, casein and gelatin hydrolysis were performed as protocol described by Meena et al. (Meena et al. 2013). Ability of the isolates to utilize various carbon sources, i.e. sucrose, lactose, glucose, ribose, xylose, mannitol, maltose, arabinose, raffinose and salicin were performed in ISP-2 agar medium with phenol red as an indicator (Biehle et al. 1996). Physiological characteristics such as tolerance to salt (5-30% NaCl), pH (5–11) and survival at 25-45°C with and without shaking of isolates A115 and F4 were also examined. Strains A115 and F4 were grown in $1/10^{th}$ NB and incubated on a rotatory shaker at 85 rpm at room temperature for 21 days. Cells were harvested by centrifugation at 8000rpm for 10 mins. Aliquots of cell pellets were shipped to MIDI labs for whole cell fatty acid profiling.

Genetic characterization of actinobacterial isolates

Isolation of genomic DNA

Actinobacterial isolates A115 and F4 were grown in 500 mL of $1/10^{th}$ nutrient broth for 25 days at room temperature before cells were harvested by centrifugation. Genomic DNA was extracted according to the protocol described previously (Nikodinovic et al. 2003) with modifications. Briefly, the cell pellets were collected by centrifugation and pellets were crushed five times with liquid nitrogen. After crushing, the pellets were resuspended in 10 mL of TE buffer which contained 150mg of lysozyme and 50mg of achromopeptidase (Wako chemicals USA). Cell pellets were mixed thoroughly and incubated at 37°C for 4 hours. Then 1.0 mL of

10% SDS and 100 mg of proteinase K (Sigma-Aldrich USA) were added into the suspension and incubated at 55°C for an hour. Tubes were shaking in every 15 min. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added and the resulting suspension was mixed thoroughly before spinning at 10,000 rpm at 4°C for 10 min. The aqueous phase was collected into a fresh tube and 0.6 volume of isopropanol was added to precipitate the DNA. Shake the tube back and forth until a stringy white DNA precipitate becomes clearly visible. DNA was collected by centrifugation at 12,000 rpm at room temperature for 10 min. The DNA pellet was washed with 70% ethanol and dried the pellet using vacufuge concentrator (Eppendorf USA). Re-suspended the pellet with 10 mL of TE buffer containing 100 μL of RNase A(10 mg/mL) and incubated the solution at room temperature for 30 minutes. Re-extracted DNAs with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1), after mixing thoroughly the DNAs were collected by centrifugation at 10,000 rpm at 4°C for 10 min. The aqueous phase was transferred into a fresh tube and $1/10^{th}$ volume of 3M Na-acetate (pH 5.2) and 2.5 volume of 100% ethanol were added. Resulting solution was incubated at -20°C for 30 min prior to collecting the DNA pellet by spinning at 12,000 rpm for 10 min at 4°C. The pellet was washed with 70% ethanol and dried the pellet using Vacufuge concentrator. After drying, the pellet was dissolved in 500 μL of nuclease free water and concentration of DNA was measured by using Qubit 2.0 fluoromoter (Life technologies, USA).

Whole genome sequencing, assembly and annotation

To complete the genomes of A115 and F4, long span 10-20 kb NGS mate pair libraries and a conventional 600 bp paired end fragment libraries were constructed at Lucigen Corp. (Middleton, WI). Genomic DNAs of strains A115 and F4 were used to make 8 kb mate-pair

libraries. The library was constructed for Illumina sequencing using a NxSeq® Long Mate Pair Library Kit according to manufacturers protocol (Lucigen, Middleton, WI). The 20 kb Long Mate Pair Libraries were constructed using the NxSeq® Long Mate Pair Library Kit with the 10-20 kb insert Supplementary Protocol (Lucigen). In brief, the 20kb libraries were constructed by shearing genomic DNA to 25 kb with a Covaris G-Tube (15 µg DNA was centrifuged at 3200 rpm in a G-Tube for 8 min in both orientations) followed by end repair, A-tailing and ligation to adaptors. Adapted DNA was then size-selected on a 0.3% SeaKem Gold agarose gel (Lonza, Basel, Switzerland) and eluted with an Elutrap Electrophoresis Chamber (GE Healthcare Life Sciences, Pittsburgh, PA). Size-selected DNA was then ligated to a coupler, exonuclease treated, digested with restriction endonucleases and purified prior to circularization with a junction adaptor and amplification with KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Boston, MA). A115 and F4 fragment libraries were constructed from genomic DNA sheared to 50-1000 bp (peak ~500 bp) by g-Tube (Covaris), using the NxSeq Fragment Library kit (Lucigen, Middleton WI) according to manufacturer's protocol. All libraries were size-selected with Agencourt AMPure XP beads (Beckman Coulter, Inc., Brea CA) and sequenced on an Illumina MiSeq using MiSeq Reagent kit version 3. Genomic sequence assemblies were carried out with SPAdes 3.5 (bioinf.spbau.ru/spades) using a range of Kmer values (22 to 77). Fragment library reads were assembled independently, in combination with 8kb mate-pair library reads, or in combination with both 8kb and 20kb mate-pair library reads. For each genome, the assembly with the largest scaffold(s) and fewest unassigned contigs was imported into CLC Genomics Workbench 7.5 for gap filling and finishing the genome. The remaining gaps in the genomes were closed by Sanger sequencing of PCR amplicons and unscaffolded gaps were closed by sequencing of single-primer PCR amplicons according to the method described previously

(Karlyshev et al. 2000). For annotation of the A115 and F4 genomes, the final closed-circle version of the genome sequences were submitted to the NCBI's Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (Angiuoli et al. 2008), followed by submission to GenBank. Gene prediction and annotation were performed using GeneMark (Lukashin and Borodovsky 1998) and RAST annotation server (Aziz, Bartels et al. 2008), respectively. The identity of individual ORFs from secondary metabolite biosynthesis gene clusters was confirmed by BLASTx against the GenBank database.

Prediction of secondary metabolite biosynthesis gene clusters in *Actinobacteria* strains A115 and F4

Secondary metabolite biosynthesis gene clusters for strains A115 and F4 were predicted using the secondary metabolite identification tool antiSMASH3.0 (Blin, Medema et al. 2013). Gene prediction and annotation were carried out by GeneMark (Lukashin and Borodovsky 1998) and BLASTx (NCBI), respectively.

Phylogenetic analysis

The 16S rRNA gene was predicted within the complete genome sequences of isolates F4 and A115 using RNammer (Lagesen et al. 2007), software for ribosomal RNA prediction. For multi-locus sequence analysis (MLSA), sequences specific to 16S rRNA, DNA gyrase subunit B (*gyrB*), RNA polymerase beta factor (*rpoB*), bacterial DNA recombination protein (*recA*), ATP synthase beta subunit (*atpD*) and tryptophan synthase beta subunit (*trpB*) were retrieved from 59 *Streptomyces* strains and *Mycobacterium tuberculosis* strain H37Rv using BLASTn tool available in the NCBI web server. Locus specific sequences were aligned, trimmed and concatenated using CLC Genomics Workbench (version 8.0.1). ClustalW algorithm in MEGA

6.0 was used to align 16S rRNA sequences for phylogenetic analysis. The evolutionary history of the *Nonomureae* and *Streptomyces* isolates that included anti-MRSA isolates F4 and A115, respectively, were inferred using Neighbor-Joining method in MEGA 6.0 (Tamura et al. 2011). The confidence of the evolutionary relationships was assessed using the bootstrap method with 1000 replicates. Similar approaches were taken to reconstruct the phylogenetic tree for concatenated six house-keeping genes of 59 strains of *Streptomyces* species.

Average nucleotide identity (ANI)

The average nucleotide identities (ANI) of anti-MRSA isolate A115 against 40 *Streptomyces* strains were determined using JSpecies (version 1.2.1) (Richter et al. 2009)

Mutagenesis of A115 genome to identify secondary metabolite gene cluster(s) for anti-MRSA activity

To identify the secondary metabolites gene clusters responsible for anti-MRSA activity, the genome of A115 was randomly mutagenized using mariner-based transposon Himar1 according to the methods described by Bilyk et al. (Bilyk et al. 2013). Briefly, the *himar1* transposon was conjugally transferred to strain A115 by mating with *E. coli* strain SM10λpir bearing the plasmid pHTM and pHAM. Randomly transposon mutagenized A115 mutants were screened for the loss of anti-MRSA activity.

4. Results

Isolation and identification of soil bacteria

A total of 548 bacterial and fungal isolates were recovered after incubation of three soil suspensions on $1/200^{th}$ NA and $1/200^{th}$ NA +SE plates. It was observed that the addition of soil

extract to the medium was slightly beneficial to all strains, as CFU counts increased 1.5 to 3.2-fold. Each of the pure culture isolates grown on 1/200thNA was subjected to a phylogenetic analysis using 16S rRNA gene. These bacterial isolates (fungal isolates were not ribotyped) represent diverse bacterial phyla that include *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Firmicutes* and *Proteobacteria*. Bacteria isolated from black belt and forest soils were predominantly affiliated with the class α-*Proteobacteria*, representing 55% and 68% of total isolated bacterial phyla for the Black Belt and Forest soils, respectively (Figure 1 A and C). Whereas bacteria isolated from the Cullars agricultural soil were mostly affiliated with the class γ-Proteobacteria, 32% of isolates were belonged to this phylum (Figure 1 B). It was also noted that the forest soil harbored less diverse bacterial phyla than the soils of black belt and agriculture.

Screening of soil isolates with anti-MRSA activity

Each of the isolates collected from 1/200th NA was screened for anti-MRSA activity using a soft agar overlay. After primary screening, a total of 22 isolates were found to inhibit MRSA growth. These bacterial isolates were tested twice for antibacterial and four of them showed repeated antibacterial activities against multiple MRSA strains (Figure 2). Out of the four soil isolates with anti-MRSA, two designated as C3 and C4 were isolated from agricultural soil, other two isolates as A115 and F4 were recovered from black belt and forest soils, respectively. Phylogenetic analysis based on 16S rRNA gene revealed that isolates C3 and C4 were affiliated with the genus *Burkholderia*, and A115 and F4 were identified as *Streptomyces* sp. and *Nonomuraea* sp., respectively. Unfortunately, two of these bacterial isolates, C3 and C4 were not maintained in culture, but the isolates A115 and F4 have been consistently grown on 1/10th

strength NA and exhibit potent anti-MRSA activity against multiple MRSA strains (Figure 2 and 3).

Antibacterial activities of *Streptomyces* sp. strain A115 and *Nonomuraea* sp. strain F4

Using a double-dilution method, culture supernatants of A115 and F4 were screened for *in vitro* antibacterial activity against MRSA strain 30. The expression of secondary metabolites active against MRSA during a time course of bacterial growth was determined and it was found that both A115 and F4 express detectable anti-MRSA compounds as early as 10 days post-inoculation into $1/10^{th}$ NB at room temperature with shaking at 85 rpm. Highest antibacterial activity of A115 was observed after 30 days of incubation (Figure 4) whereas F4 showed highest antibacterial activity after 50 days of incubation (Figure 5). Antibacterial activity of A115 and F4 strains decreased drastically at 40 and 60 days of incubations respectively (Figure 4 and 5).

Extraction of anti-MRSA compounds

It was found that the strain A115 produced a pink, cell-associated pigment. However, the role of pigment in antibacterial activity was not determined. The extraction methods including different resin treatments and organic solvent partitioning were investigated at the NCNPR for efficient recovery of anti-MRSA activity from supernatants and cell lysates. Ethyl acetate and methanol extractions of crud extracts of A115 supernatant showed IC_{50} of 4.42 and 4.17 µg/mL, respectively, against a MRSA strain. Preliminary LC-MS analysis of A115 supernatant failed to identify a UV-active compound with anti-MRSA activity. An investigation is ongoing to scale-up cultures in media that promote high anti-MRSA activity in order to obtain sufficient compound for biochemical structure elucidation.

It was observed that strain F4 produced a reddish pigment and its anti-MRSA activity is associated with a size fraction greater than 100kDa. The anti-MRSA activity is maintained even

after 10 min of boiling (Figure 3B). However, antimicrobial activity was not found from the F4 supernatant after filtration with 0.22 or even 0.45µm filter (Figure 3C). This finding demonstrated that the anti-MRSA compound(s) produced from F4 strain might be a high molecular weight compound(s) which might be (associated with) the reddish pigment. The extraction method including XAD-7 resin treatment revealed efficient extraction of active compound(s) of F4 supernatant (Figure 6) but subsequent HPLC analysis did not identify any active fraction.

Morphological and biochemical characteristics of *Actinobacteria* strains A115 and F4

On $1/10^{th}$ NA plates, strain A115 took 7-10 days to form visible colonies on a plate and the colony size gradually increased with increasing incubation times (up to a month). Highly filamentous colonies with massive aerial hyphae were observed after a month of incubation. Increasing the incubation time also increased the pink pigment production and the center of the colony had apparent aerial hyphae. In the case of strain F4, colonies typically appeared after 3-4 weeks of incubation on $1/10^{th}$ NA at room temperature. The colonies were initially white in color and completely grown into the agar, colony was collected by cutting the entire agar surrounding the colony. Increasing incubation time up to 3 months increased the production of brown or reddish color of pigment. Scanning electron microscopic observations demonstrated very different morphological features of two actinobaterial isolates (Figure 7 and 8). Biochemical and physiological characteristics of A115 and F4 were presented in Table 1. The fatty acid analysis of both strains revealed no significant match with any previously identified *Actinobacteria* species. The major fatty acid components of A115 were C15:0 iso, anteiso and C16:0 iso. For F4

strain, the major fatty acids were C15:0 iso, C16:1 iso and 10 methyl C17:0. Based on unique morphological, physiological and biochemical properties, these two isolates are both predicted to be novel *Actinobacteria* species. In order to improve the production of secondary metabolites by A115 and F4 strains in different laboratory conditions, several growth media and culture incubation conditions were tested (Table 2). It was established that the SYZ and $1/10^{th}$ NB media were the best for antibiotic production by A115 and F4, respectively.

Phylogenetic analysis

To determine the evolutionary relationships of the anti-MRSA isolate A115, a *16S rRNA gene sequence* based phylogenetic analysis was conducted using a total of 60 *Streptomyces* strains that included isolate A115 and 59 other *Streptomyces* strains sequences available in the GenBank of NCBI. The phylogenetic tree reconstructed from these near-complete (1,389 bp) 16S rRNA gene sequences (Figure 9) showed that isolate A115 clusters with members of the genus *Streptomyces*. The 16S rRNA gene sequence of isolate A115 is most closely related to that of *S. afghaniensis* strain 772 and *S. olindensis* DAUFPE 5622. A clear branching of isolate A115 from this species was strongly supported on the basis of 100% bootstrap values. BLASTn searching of the GenBank database using 1,389 bp 16S rRNA gene sequence revealed that isolate A115 is most closely related to strains belonging to the genus *Streptomyces* (84-95% similarity). The greatest % identity of the isolate A115 16S rRNA gene was with *S. afghaniensis* 772 and *S. olindensis* DAUFPE 5622, both which were observed to have 95% identity with their respective 16S rRNA gene sequences (Table 3). The low % identity of the isolate A115 16S rRNA gene with any other *Streptomyces* sp., below the 97% cutoff that is commonly used, suggested that this isolate is a novel species of *Streptomyces*. Because bacterial species cannot be defined solely on the basis of a 16S rRNA gene, a more robust phylogenetic analysis was

required to establish whether strain A115 was truly a representative of a novel *Streptomyces* species.

To provide a more refined phylogenetic placement of the anti-MRSA isolate A115, a multilocus sequence-based phylogeny was conducted that included 8,419 bp of concatenated nucleotide sequences from six housekeeping genes such as 16S rRNA, DNA gyrase subunit B (*gyrB*), RNA polymerase beta factor (*rpoB*), bacterial DNA recombination protein (*recA*), ATP synthase beta subunit (*atpD*) and tryptophan synthase beta subunit (*trpB*). MLSA clearly distinguished isolate A115 from the others and represented a distinct lineage. The multi-locus sequence analysis clearly segregated strains from individual species according to the established species delineation of *Streptomyces* genus (Figure 10), supported by 100% bootstrap value. Sequences of isolate A115 clustered separately from those of all previously described *Streptomyces species*. This multilocus phylogeny demonstrated that the anti-MRSA isolate A115 formed one independent lineage clearly separated from other 58 *Streptomyces* isolates used in this study (Figure 10). MLSA phylogenetic tree revealed that isolate A115 is closely related but independently branched from *S. afghaniensis* strain 772 and *S. olindensis* DAUFPE 5622. This finding is concordant with the 16S rRNA based phylogeny that demonstrated that isolate A115 is a separate clade from the other *Streptomyces* species described previously. Taken together, these findings support the hypothesis that isolate A115 is a novel *Streptomyces* species.

In addition, the evolutionary relationships of the anti-MRSA isolate F4 was determined by using *16S rRNA gene sequence.* Phylogenetic tree was constructed using a total of 97 *Nonomuraea* strains that included isolate F4 and 95 other *Nonomuraea* strains sequences available in the GenBank of NCBI. The phylogenetic tree reconstructed from these near-

complete (1,304 bp) 16S rRNA gene sequences (Figure 11) showed that isolate F4 clusters with members of the genus *Nonomuraea*. The 16S rRNA gene sequence of isolate F4 is most closely related to that of *N. antimicrobica*.

Average nucleotide identity (ANI)

Since average nucleotide identity (ANI) is a most reliable measurement of genomic relatedness between strains to determine their species demarcation (Kim et al. 2014), the genome sequence of anti-MRSA isolate A115 was compared against 39 strains of *Streptomyces*. All the strains that were compared against A115 genome showed ANI values less than 90% (Table 4). The genome of *S. afghaniensis* 772 showed highest similarity with A115 strain with ANI of 89.76% which is lower than ANI of 95% to be considered as member of the same species. This finding is consistent with the 16S rRNA gene based phylogenetic analysis which showed *S. afghaniensis* 772 as a closest neighbor to isolate A115. The A115 ANI of lower than 95% determined against a large number of *Streptomyces* species including its closest neighbor suggests that A115 isolate is a novel species of *Streptomyces*.

Prediction of secondary metabolite biosynthesis gene clusters in *Actinobacteria* strains A115 and F4

Secondary metabolite biosynthesis gene clusters for strains A115 and F4 were predicted using the secondary metabolite identification tool antiSMASH3.0 (Blin, Medema et al. 2013). Primary analysis of the A115 and F4 contig sequences revealed that A115 and F4 encode 69 and 66 secondary metabolite biosynthetic gene clusters respectively. However, using whole genome sequences of A115 and F4 suggested that these genomes encode 34 and 24 biosynthetic gene clusters respectively (Figure 12 and 13). This observation suggested the power of using whole

genome sequencing for identifying complete biosynthetic pathways for antibiotic production, some of which will not be identified otherwise.

5. Discussion

The emergence of multidrug-resistant pathogens has increased the demand for discovery of novel antibiotics (Bush 2011; Shlaes et al. 2004; Falkinham et al. 2009). New antibiotics are in high demand for the treatment of *Staphylococcus aureus* infections particularly due to the emergence of methicillin-resistant *S. aureus* (MRSA) in communities (**Hageman et al. 2006**; Lewis 2012). MRSA infections were initially restricted to hospitals, but are now widely present in the community (Lewis 2012).

The best possible source for new antibiotics with potentially novel mechanisms of action is within natural environments, particularly soils (Ling et al. 2015), which have the greatest diversity of microbial life. Soil is densely populated with microorganisms that produce small bioactive molecules, including antibiotics, anticancer compounds, immunosuppressive agents, insecticides, and others (**Handelsman et al. 1998;** Omura et al. 1992; Paradkar et al. 2003; Pettit et al. 2004). Over two-thirds of clinically-used antibiotics are natural products or semi-synthetic derivatives (Fair et al. 2014). Most of these have been derived from cultured microorganisms, which represent <1% of the total soil community (Torsvik, Goksøyr et al. 1990; Amann et al. 1995). Because soil is estimated to harbor $>10^5$ species per gram (Hartmann et al. 2006), there exists extensive undiscovered functional diversity (Lewis et al. 2010). However, although there has been a wealth of antibiotics discovered from cultured soil microbes (Thiele-Bruhn 2003), the past several decades have experienced diminishing success rates for antibiotic discovery using a culture-based approach. This is due to the very high rate of antibiotic re-discovery (>99%) when

screening cultured bacteria grown under “normal” laboratory conditions, i.e. high nutrient levels with short incubation times. Novel culturing techniques can help to access a greater diversity of antibiotics produced by previously uncultured microorganisms.

In this study, a low-nutrient medium supplemented with a soil extract combined with very long-term incubation approaches were used to avoid the re-isolation of previously identified burden of background for screening of common bacteria. Results of present study suggesting that the addition of soil extract to medium and longer incubation time are helpful for the isolation of bacteria from soils. Phylogenetic analysis revealed the presence of diverse bacterial phyla in soil and the overall bacterial diversity showed significant variability among samples. In this study, only the pure culture isolates were ribotyped which might not represent the true bacterial diversity that were present in the soil samples since majority of soil bacteria are recalcitrant to cultivate in laboratory conditions. Out of 548 soil isolates two of them have potential to inhibit the growth of clinical MRSA strains. These two bacteria, isolates A115 and F4, are affiliated with the phylum *Actinobacteria* and a phylogenetic analysis based on *16S rRNA gene sequence* analysis using *Streptomyces* strains sequences available in the GenBank showed that isolate A115 clusters with members of the genus *Streptomyces*. The greatest % identity of the isolate A115 16S rRNA gene was with *S. afghaniensis* 772 and *S. olindensis* DAUFPE 5622, both of which were observed to have 95% identity with their respective 16S rRNA gene sequences. Since the % sequence similarity of 16S rRNA gene sequence is less than 97% which has been widely used as a threshold for bacterial species delineation, the isolate A115 considered as a novel species. More rigorous phylogenetic analysis with six housekeeping genes using MLSA revealed that the isolate A115 is closely related but independently branched from *S. afghaniensis* strain 772 and *S. olindensis* DAUFPE 5622. Furthermore, the comparative study based on ANI

values of whole genomes indicating lower than 95% identity that can be considered A115 as member of the same species. Therefore, the isolate A115 can be considered as a novel *Streptomyces* species. The identification of *Streptomyces* isolates has been extremely important historically for antibacterial screening, as this genus is proven to be the prolific producers of novel antibiotics (Watve et al. 2001) with approximately 75% of commercially useful antibiotics being derived from them (Guo et al. 2015).

The genus *Nonomuraea* belongs to rare *Actinobacteria* that promise a raise in the prospect of discovering novel compounds with potential antimicrobial activities (Tiwari et al. 2012). Here in this study, strain F4 was isolated from forest soil which belonged to the genus *Nonomuraea,* had strong antibacterial activity against clinical MRSA strain. Whole genome sequencing of F4 strain revealed high G+C% of 70.4 and relatively larger genome of 10.3 Mbp which is similar to three other available complete genomes of *Nonomuraea* species such as *N. coxensis* (G+C% 72; 9.0Mb), *N. candida* (G+C% 72.1; 11.01Mb) and *N. kuesteri* (G+C% 70.5; 13.36Mb). Phylogenetic analysis based on 16S rRNA gene sequence revealed *N. antimicrobica* YIM 61105 as the closest neighbor of *Nonomuraea* sp. strain F4. In the phylogenetic tree, 16S rRNA sequence of F4 strain produced a monophyletic branch with *N. antimicrobica* with bootstrap value of 93%, however, these two isolates showed significant differences in their morphological, physiological and biochemical characteristics. Because of few available whole genomes sequences of *Nonomuraea* spp. in public database, we were not able to calculate ANI values of F4 strain with other *Nonomuraea* spp. to determine the genetic relatedness between strains. ANI is one of the most robust measurements of genomic relatedness between strains and has potential to use as an alternative of DNA-DNA hybridization (DDH) technique for species delineation (Kim et al. 2014). Therefore, we cannot confirm whether strain F4 should be

belonged to the species *N. antimicrobica* only based on 16S rRNA gene specific phylogenetic tree. Although 16S rRNA gene is the most widely used phylogenetic marker in microbial ecology, however, the presence of multiple copies of this gene in a single bacterium can influence the phylogenetic resolution and operational taxonomic unit estimation at the species level or below (Case et al. 2007). Since 16S rRNA gene is evolutionarily conserved, using this sequence for delineating species designation using phylogenetic analysis is difficult. Therefore, alternative molecular marker such as MLSA studies using several housekeeping genes should be used to classify bacterial isolates to the species level.

Several attempts have been taken for biochemically characterize antibacterial compound(s) produced by this strain was unsuccessful; the reason behind this may be due to the large size of the active compound and failure to use appropriate extraction procedures for compound separation. Thus, future research to develop an efficient and sensitive analytic system for active compound analysis from this strain will be of special interest.

Taken together, our findings indicated that novel culturing can be applied to identify novel species of *Actinobacteria* from soil and some of which can be used as therapeutics to prevent infections caused by MRSA.

Table 1. Morphological, biochemical and physiological characteristics of A115 and F4 strains

Properties	***Streptomyces* sp. A115**	***Nonomuaea* sp. F4**
Morphological characteristics		
Color of aerial mycelium	Whiteish pink	Non
Color of substrate mycelium	Red	Brown
Soluble pigment	Red	Brown
Biochemical characteristics		
Gram staining	+	+
KOH	-	-
Indole production	-	-
Methyl Red	Nd	Nd
Voges Proskauer	-	-
Citrate utilization	+	-
H_2S production	+	-
Nitrate reduction	-	
Urease	+	-

Catalase	+	+
Oxidase	-	-
Motility	-	-
Starch hydrolysis	+	-
Triple sugar iron	alk/alk H_2S +	-
Survival at 50°C	Slight	-
Carbon source utilization		
PR arabinose	-	-
PR Lactose	-	-
PR Salicin	-	-
PR Ribose	-	+
PR Raffinose	-	-
PR Maltose	-	-
PR Mannitol	-	-
PR Xylose	-	-
Glucose	+	+
Starch	+	-
Dextrose	+	-
O/F Trehalose	Oxidation +	-
Phenylalanine deaminase	-	-
pH		
5	+	+
6	+	+

7	+	+
8	-	-
9	-	-
10	-	-
11	-	-
NaCl tolerence (%)		
5	Slight	-
10	-	-
15	-	-
20	-	-
25	-	-
30	-	-

+ indicates positive, - indicates negative and Nd means not determined.

Table 2. Effect of growth media on antibiotic production

Media	Antibacterial activity of A115 supernatant	Antibacterial activity of F4 supernatant
1/10th NB	+	++
1/10th NB + Salt	---	---
1/10th NB + NAGA	---	---
ISP 2	---	---
ISP 4	---	---
SYZ	+++	---

Antimicrobial activity of A115 and F4 strains in different media type. Note that (+) indicates a zone of inhibition up to 5 mm, (++) indicates a zone of inhibition from 5 mm to 10 mm, (+++) indicates a zone of inhibition greater than 15mm, and (---) indicates no observable zone of inhibition.

Table 3. ANI value and % similarity of 16S rRNA Sequence of A115

Strains of *Streptomyces*	ANI of A115	% Similarity of 16S rRNA Sequence of A115
Streptomyces afghaniensis 772	89.76	95
Streptomyces viridochromogenes DSM40736	89.32	Nd
Streptomyces olindensis DAUFPE 5622	89.14	95
Streptomyces sviceus ATCC29083	82.96	92
Streptomyces davawensis JCM4913	82.80	93
Streptomyces canus 299MFChir4	82.72	Nd
Streptomyces lividans TK24	82.51	92
Streptomyces coelicolor A3(2)	82.45	92
Streptomyces gancidicus BKS13-15	82.12	Nd
Streptomyces collinus Tu365	81.81	92
Streptomyces avermitilis MA-4680	81.79	91
Streptomyces prunicolor NBRC13075	81.72	91
Streptomyces turgidiscabies Car8	81.36	Nd
Streptomyces achromogenes subsp *achromogenes* NRRL B 2120	81.30	91

Streptomyces bottropensis ATCC25435	81.23	92
Streptomyces scabiei 87.22	81.22	92
Streptomyces acidiscabies 84 104	79.68	92
Streptomyces aurantiacus JA4570	79.19	Nd
Streptomyces albus DSM41398	77.88	89
Streptomyces pristinaespiralis ATCC25486	77.85	88
Streptomyces thermolilacinus SPC6	77.84	Nd
Streptomyces lavendulae	77.64	89
Streptomyces flavidovirens DSM40150	77.61	88
Streptomyces exfoliates	77.59	88
Streptomyces roseosporus NRRL 11379	77.57	Nd
Streptomyces fulvissimus DSM40593	77.49	89
Streptomyces globisporus subsp	77.45	89
Streptomyces griseus XylebKG-1	77.41	89
Streptomyces clavuligerus ATCC27064	77.35	87
Streptomyces scopuliridis RB72	77.27	Nd
Streptomyces himastatinicus ATCC53653	77.07	Nd
Streptomyces bingchenggensis BCW-1	76.93	Nd
Streptomyces niveus NCIMB 11891	76.87	88
Streptomyces pratensis ATCC33331	76.86	88
Streptomyces hygroscopicus subsp	76.78	87
Streptomyces rapamycinicus NRRL5491	76.75	86
Streptomyces rimosus	76.73	88

Streptomyces yeochonensis CN732	75.19	84
Candidatus Streptomyces massiliensis AP10	74.71	85
Streptomyces xinghaiensis S187	76.76	88
Streptomyces mutabilis	82.67	93
Streptomyces mobaraensis NBRC13819	76.19	88
Streptomyces griseoaurantiacus M045	80.29	90
Streptomyces peucetius	76.77	88
Streptomyces ghanaensis ATCC14672	83.73	94
Streptomyces xanthophaeus	77.06	Nd
Streptomyces nodosus ATCC14899	80.49	88
Streptomyces griseoflavus Tu4000	83.39	92
Streptomyces glaucescens GLA.O	82.46	92
Streptomyces galbus	81.29	92
Streptomyces_griseorubens	82.14	Nd
Streptomyces vietnamensis GIM4.0001	77.73	89
Streptomyces sulphureus DSM 40104	73.65	87
Streptomyces coelicoflavus ZG0656	82.59	92
Streptomyces purpureus KA281	78.24	88
Streptomyces tsukubaensis NRRL18488	77.17	87
Streptomyces ipomoeae 91-03	81.54	92
Streptomyces vitaminophilus DSM 41686	73.67	Nd
Streptomyces cattleya DSM 46488	75.57	85
Streptomyces somaliensis DSM 40738	77.66	88

Streptomyces auratus AGR0001	76.66	89
Streptomyces venezuelae ATCC10712	77.78	Nd
Streptomyces leeuwenhoekii C34	82.99	93
Streptomyces scabrisporus DSM41855	71.74	Nd
Streptomyces aureus	80.89	90
Streptomyces erythrochromogenes	77.21	Nd
Streptomyces chartreusis NRRL12338	89.65	Nd
Streptomyces zinciresistens K42	82.21	92

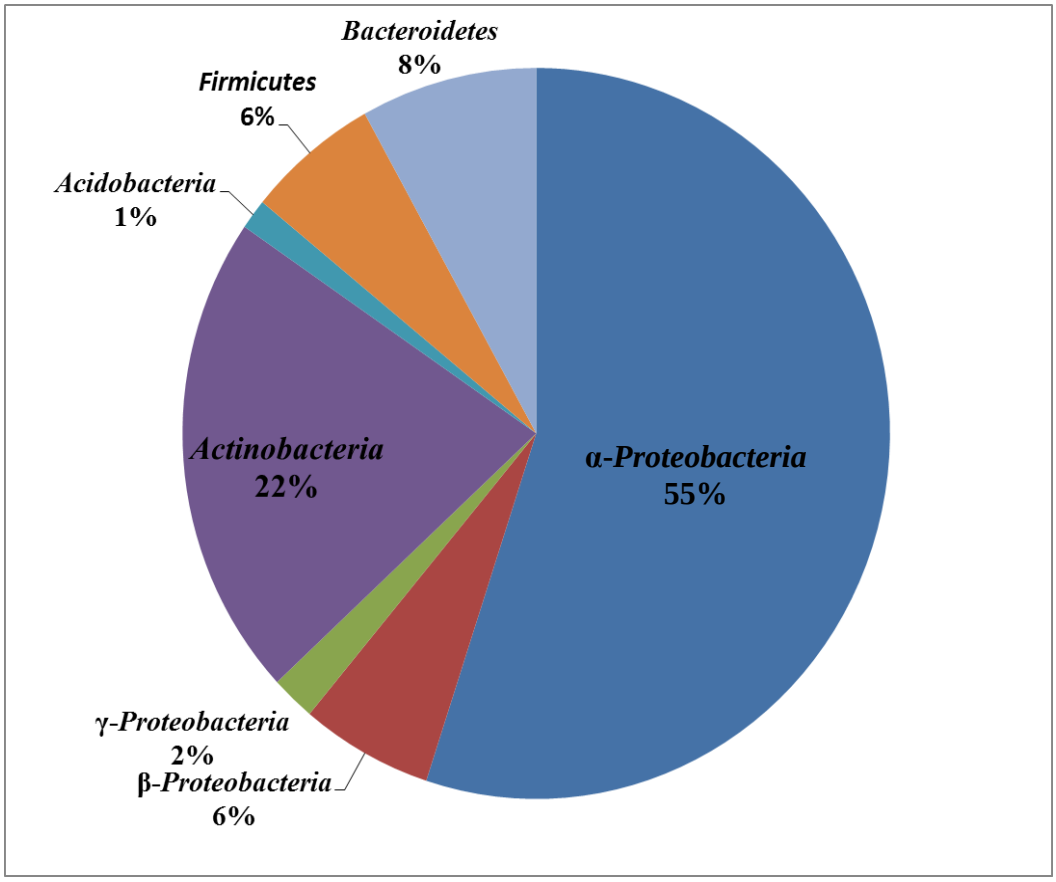


Figure 1A. Bacterial Phyla representation of cultured isolates from the sample of Black belt soil (Auburn University Arboretum, from soil removed during the construction of the Auburn University-Montgomery campus).

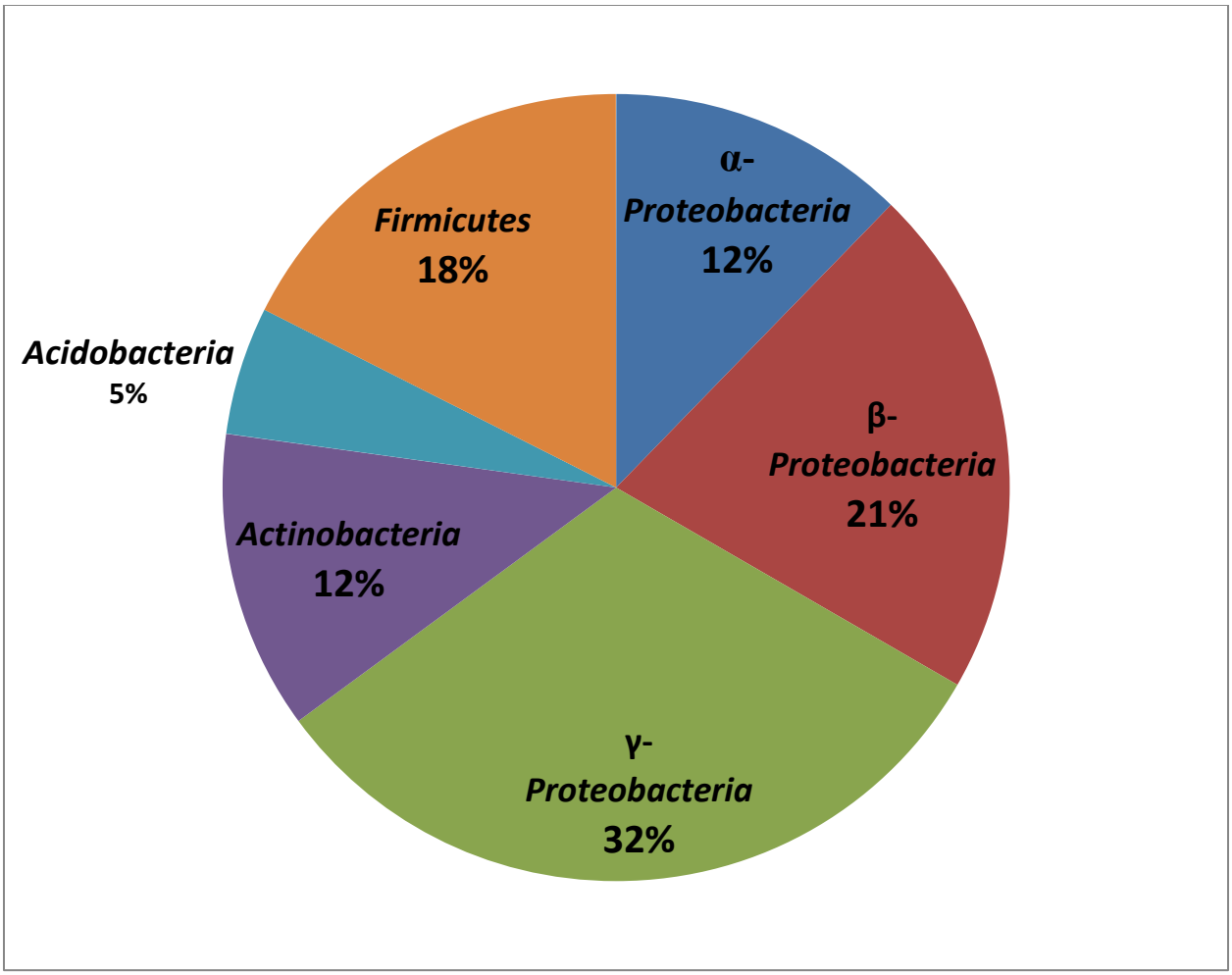


Figure 1B. Bacterial Phyla representation of cultured isolates from the sample from the Cullars Rotation agricultural soil (Auburn, AL).

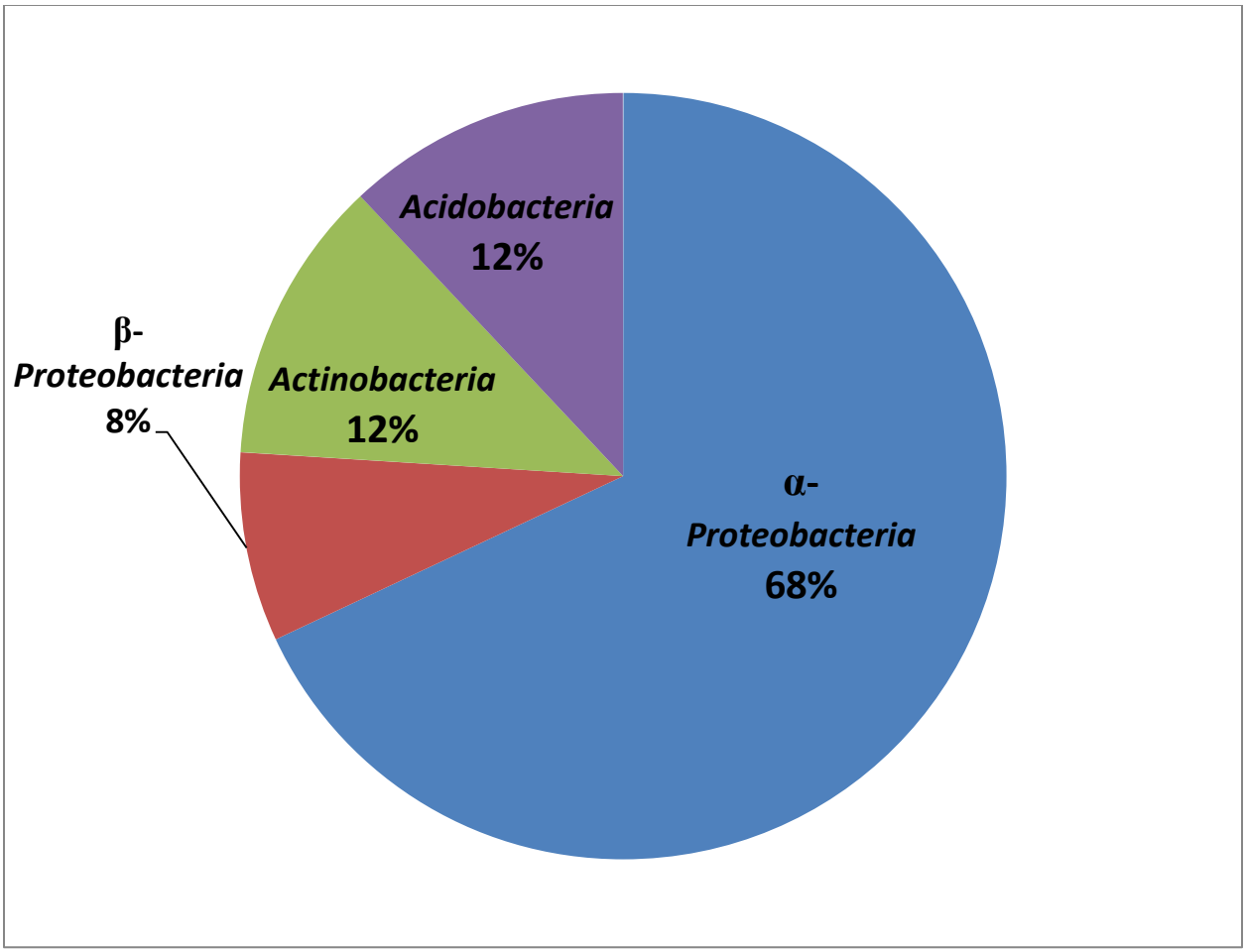


Figure 1C. Bacterial Phyla representation of cultured isolates from a sample from a forest soil adjacent to a long-leaf pine tree (Auburn, AL).

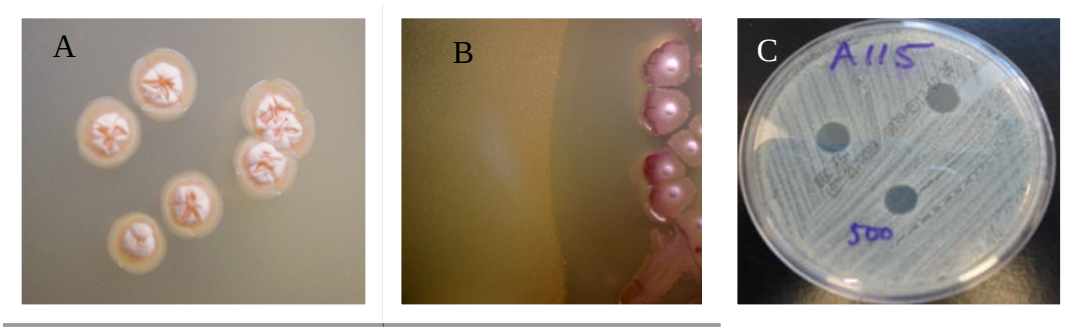


Figure 2. Culture A115 (Phylum *Actinobacteria*, genus *Streptomyces*) produces a pink pigment that is cell-associated (Panel A) and excretes a potent anti-MRSA activity (Panel B) demonstrated here by a soft agar overlay with a clinical MRSA strain 30. The anti-MRSA activity is present in cell-free bacterial supernatants when grown in 1/10th strength Nutrient Broth (Panel C).

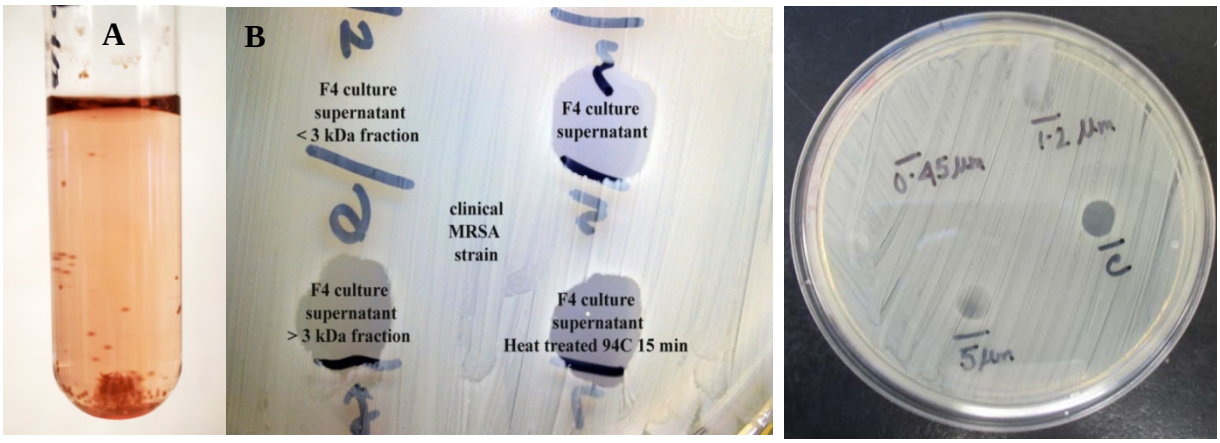


Figure 3. Culture F4 (Phylum *Actinobacteria*, genus *Nonomuraea*) produces a high molecular weight, reddish pigment (panel A) and this same fraction has a heat stable (Panel B), anti-MRSA activity (Panel C) as shown on a lawn of a clinical MRSA strain 30.

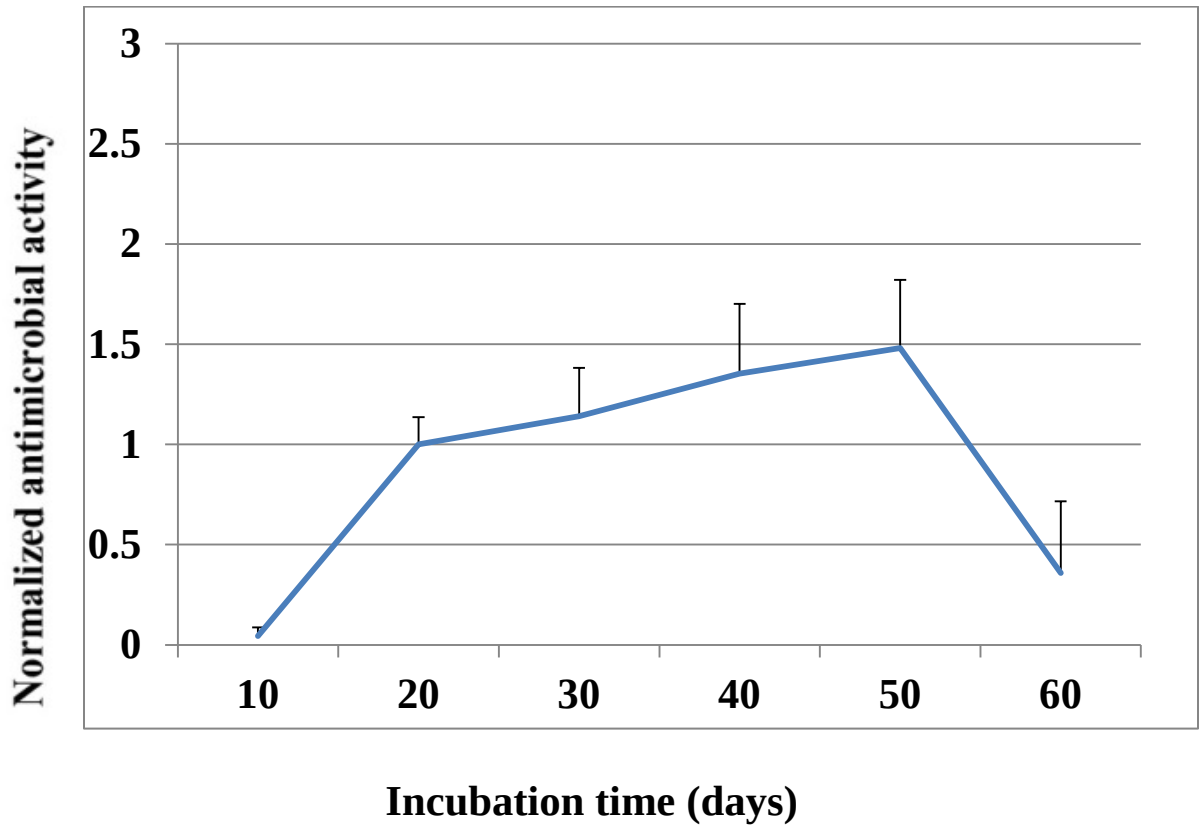


Figure 4. Normalized antibacterial activity of A115 supernatant against MRSA strain 30. A115 strain was grown for 2 months in 1/10th NB at room temperature and supernatant was tested from different time points of incubation.

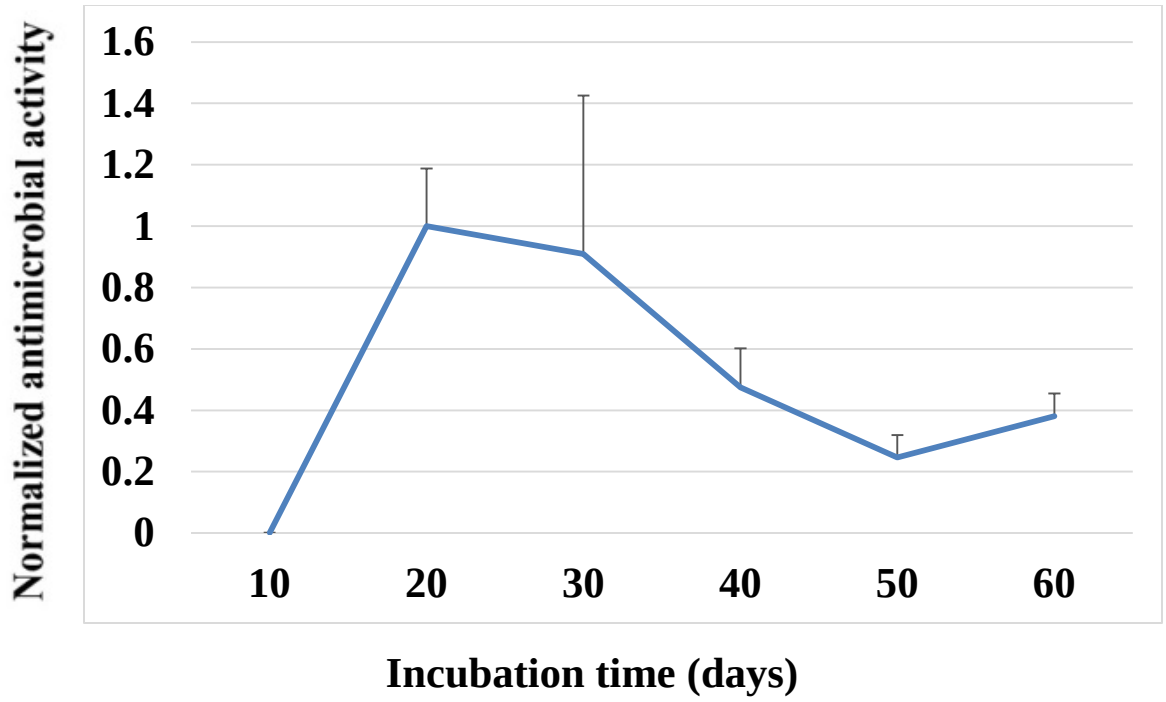


Figure 5. Normalized antibacterial activity of F4 supernatant against MRSA strain 30. F4 strain was grown for 2 months in 1/10th NB at room temperature and supernatant was tested from different time points of incubation.

Figure 6. Extraction of anti-MRSA compound(s) of F4 supernatant by XAD-7 resin treatment. Fractions 2 to 6 labeled as EL-2 to EL-6 contained active compound that was demonstrated by drop assay.

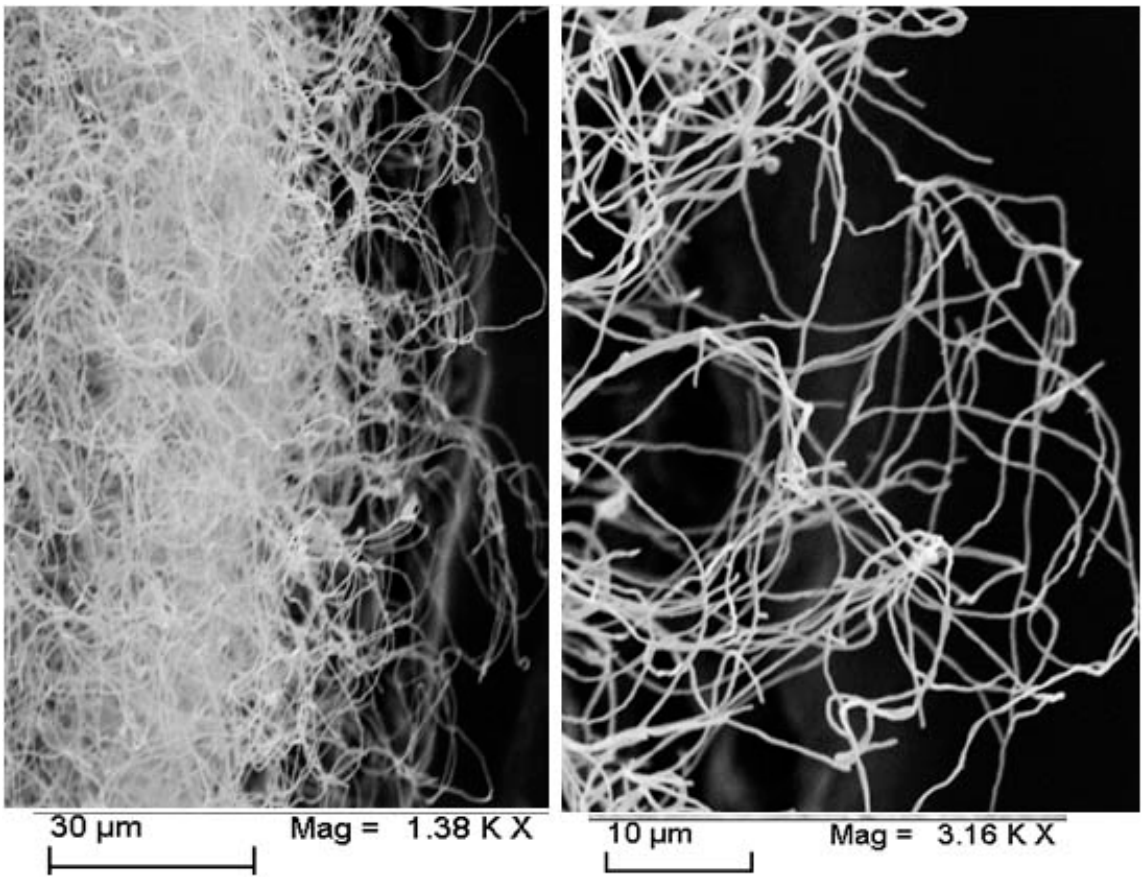


Figure 7. Scanning electron micrograph of *Streptomyces alburnustigris* strain A115 showing mycelial structures after 3 weeks of incubation at room temperature in 1/10th NB.

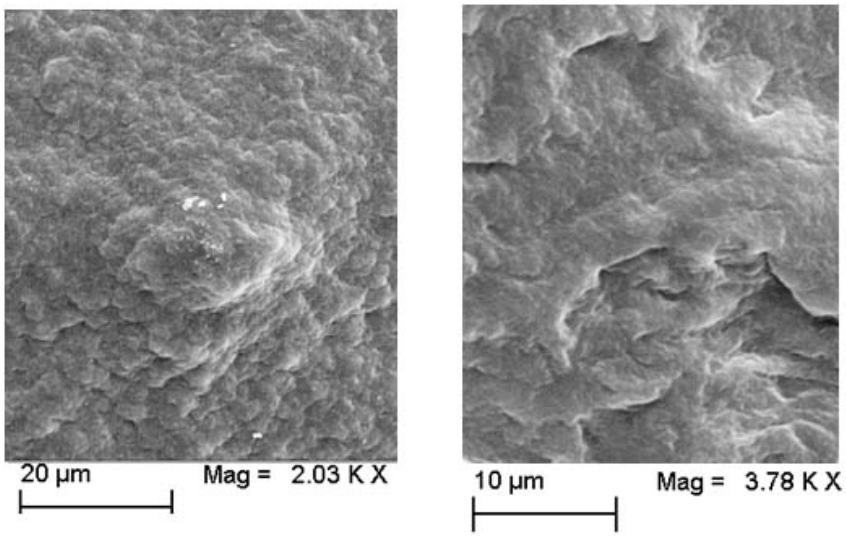


Figure 8. Scanning electron micrograph of *Nonomuraea* sp. strain F4 showing compact structures of cells after 3 weeks of incubation at room temperature on 1/10th NA.

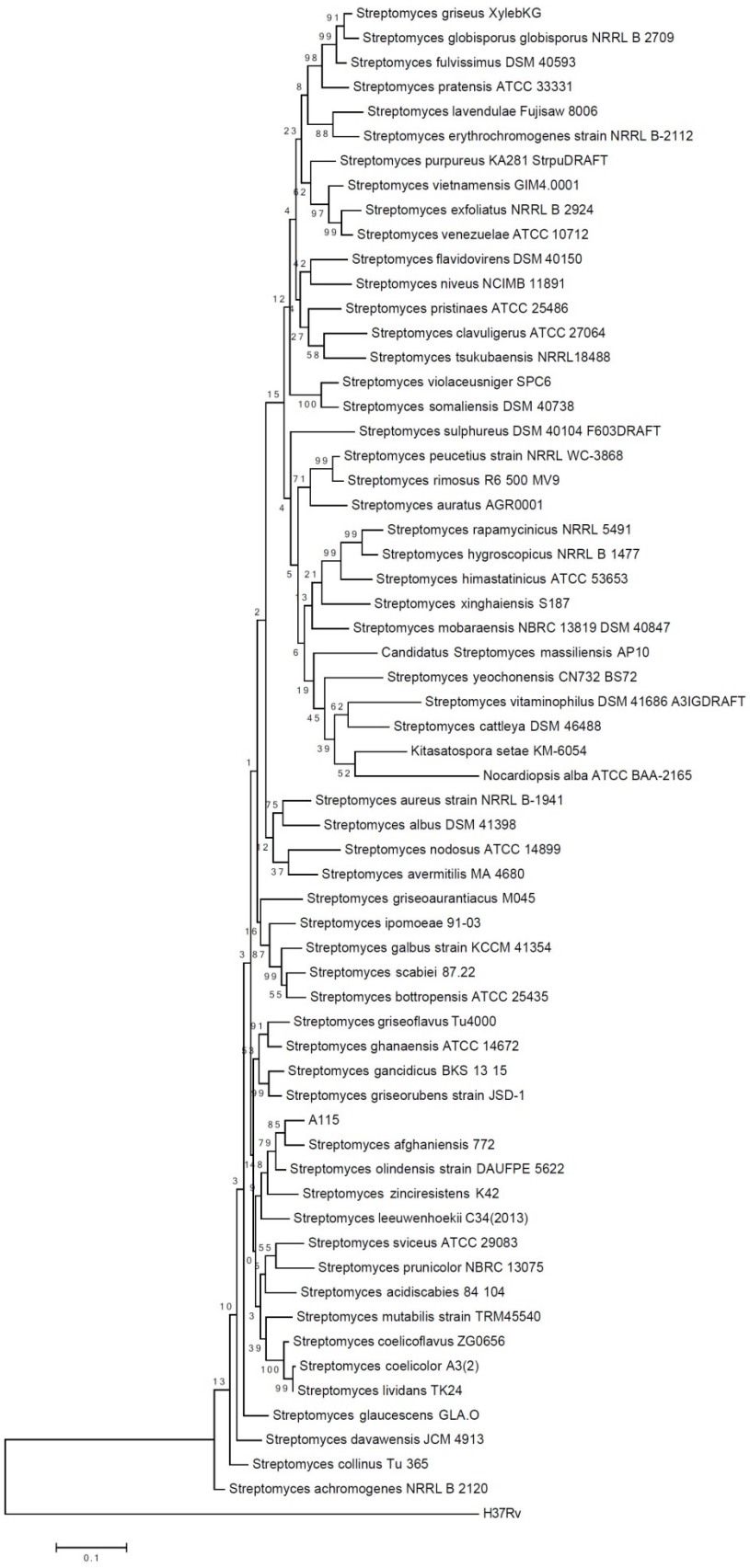

Streptomyces griseus XylebKG
Streptomyces globisporus globisporus NRRL B 2709
Streptomyces fulvissimus DSM 40593
Streptomyces pratensis ATCC 33331
Streptomyces lavendulae Fujisaw 8006
Streptomyces erythrochromogenes strain NRRL B-2112
Streptomyces purpureus KA281 StrpuDRAFT
Streptomyces vietnamensis GIM4.0001
Streptomyces exfoliatus NRRL B 2924
Streptomyces venezuelae ATCC 10712
Streptomyces flavidovirens DSM 40150
Streptomyces niveus NCIMB 11891
Streptomyces pristinaes ATCC 25486
Streptomyces clavuligerus ATCC 27064
Streptomyces tsukubaensis NRRL18488
Streptomyces violaceusniger SPC6
Streptomyces somaliensis DSM 40738
Streptomyces sulphureus DSM 40104 F603DRAFT
Streptomyces peucetius strain NRRL WC-3868
Streptomyces rimosus R6 500 MV9
Streptomyces auratus AGR0001
Streptomyces rapamycinicus NRRL 5491
Streptomyces hygroscopicus NRRL B 1477
Streptomyces himastatinicus ATCC 53653
Streptomyces xinghaiensis S187
Streptomyces mobaraensis NBRC 13819 DSM 40847
Candidatus Streptomyces massiliensis AP10
Streptomyces yeochonensis CN732 BS72
Streptomyces vitaminophilus DSM 41686 A3IGDRAFT
Streptomyces cattleya DSM 46488
Kitasatospora setae KM-6054
Nocardiopsis alba ATCC BAA-2165
Streptomyces aureus strain NRRL B-1941
Streptomyces albus DSM 41398
Streptomyces nodosus ATCC 14899
Streptomyces avermitilis MA 4680
Streptomyces griseoaurantiacus M045
Streptomyces ipomoeae 91-03
Streptomyces galbus strain KCCM 41354
Streptomyces scabiei 87.22
Streptomyces bottropensis ATCC 25435
Streptomyces griseoflavus Tu4000
Streptomyces ghanaensis ATCC 14672
Streptomyces gancidicus BKS 13 15
Streptomyces griseorubens strain JSD-1
A115
Streptomyces afghaniensis 772
Streptomyces olindensis strain DAUFPE 5622
Streptomyces zinciresistens K42
Streptomyces leeuwenhoekii C34(2013)
Streptomyces sviceus ATCC 29083
Streptomyces prunicolor NBRC 13075
Streptomyces acidiscabies 84 104
Streptomyces mutabilis strain TRM45540
Streptomyces coelicoflavus ZG0656
Streptomyces coelicolor A3(2)
Streptomyces lividans TK24
Streptomyces glaucescens GLA.O
Streptomyces davawensis JCM 4913
Streptomyces collinus Tu 365
Streptomyces achromogenes NRRL B 2120
H37Rv
0.1

Figure 9. Phylogenetic tree reconstructed based on 16S rRNA gene sequences showing the evolutionary relationship among anti-MRSA isolate A115 and members of the genus *Streptomyces*. The tree was inferred using the maximum likelihood method. The 16S rRNA sequences of *Mycobacterium tuberculosis* strain H37Rv was used as an outgroup. Numbers at each branch nodes indicate bootstrap percentages based on the maximum likelihood method (1000 replicates) that was calculated using MEGA 6.0 software (Tamura et al. 2011). Bar indicates 5 substitutions per 100 positions.

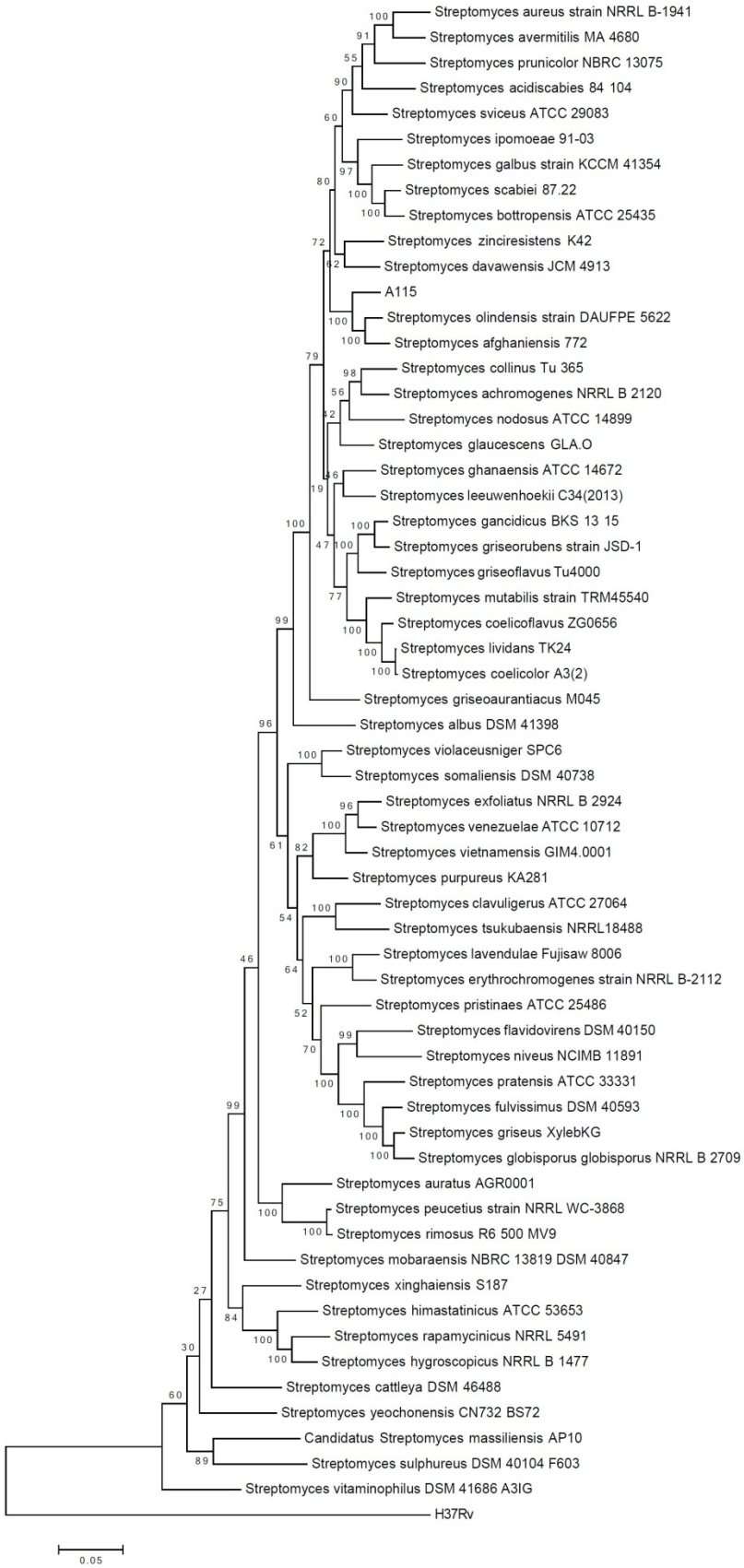

Streptomyces aureus strain NRRL B-1941
Streptomyces avermitilis MA 4680
Streptomyces prunicolor NBRC 13075
Streptomyces acidiscabies 84 104
Streptomyces sviceus ATCC 29083
Streptomyces ipomoeae 91-03
Streptomyces galbus strain KCCM 41354
Streptomyces scabiei 87.22
Streptomyces bottropensis ATCC 25435
Streptomyces zinciresistens K42
Streptomyces davawensis JCM 4913
A115
Streptomyces olindensis strain DAUFPE 5622
Streptomyces afghaniensis 772
Streptomyces collinus Tu 365
Streptomyces achromogenes NRRL B 2120
Streptomyces nodosus ATCC 14899
Streptomyces glaucescens GLA.O
Streptomyces ghanaensis ATCC 14672
Streptomyces leeuwenhoekii C34(2013)
Streptomyces gancidicus BKS 13 15
Streptomyces griseorubens strain JSD-1
Streptomyces griseoflavus Tu4000
Streptomyces mutabilis strain TRM45540
Streptomyces coelicoflavus ZG0656
Streptomyces lividans TK24
Streptomyces coelicolor A3(2)
Streptomyces griseoaurantiacus M045
Streptomyces albus DSM 41398
Streptomyces violaceusniger SPC6
Streptomyces somaliensis DSM 40738
Streptomyces exfoliatus NRRL B 2924
Streptomyces venezuelae ATCC 10712
Streptomyces vietnamensis GIM4.0001
Streptomyces purpureus KA281
Streptomyces clavuligerus ATCC 27064
Streptomyces tsukubaensis NRRL18488
Streptomyces lavendulae Fujisaw 8006
Streptomyces erythrochromogenes strain NRRL B-2112
Streptomyces pristinaes ATCC 25486
Streptomyces flavidovirens DSM 40150
Streptomyces niveus NCIMB 11891
Streptomyces pratensis ATCC 33331
Streptomyces fulvissimus DSM 40593
Streptomyces griseus XylebKG
Streptomyces globisporus globisporus NRRL B 2709
Streptomyces auratus AGR0001
Streptomyces peucetius strain NRRL WC-3868
Streptomyces rimosus R6 500 MV9
Streptomyces mobaraensis NBRC 13819 DSM 40847
Streptomyces xinghaiensis S187
Streptomyces himastatinicus ATCC 53653
Streptomyces rapamycinicus NRRL 5491
Streptomyces hygroscopicus NRRL B 1477
Streptomyces cattleya DSM 46488
Streptomyces yeochonensis CN732 BS72
Candidatus Streptomyces massiliensis AP10
Streptomyces sulphureus DSM 40104 F603
Streptomyces vitaminophilus DSM 41686 A3IG
H37Rv
0.05

Figure 10. Phylogenetic tree of concatenated sequences of genes 16S rRNA, *atpD*, *gyrB*, *rpoB*, *recA*, and *trpB* of species of the genus *Streptomyces*. The tree was inferred using the maximum likelihood method. *Mycobacterium tuberculosis* strain H37Rv was used as an outgroup to root the tree. Bar indicates proportion of nucleotide substitutions. Values at the nodes denote bootstrap support (in percentage) obtained based on 1000 replicates.

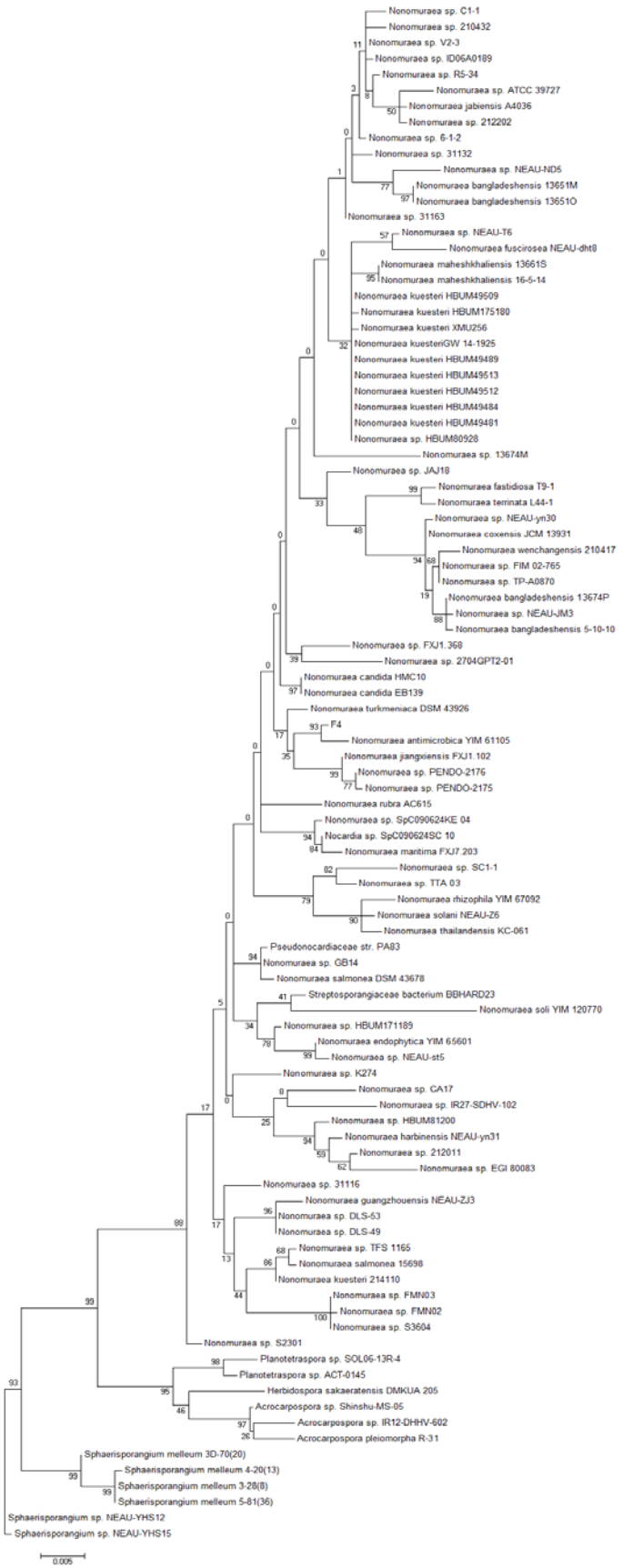
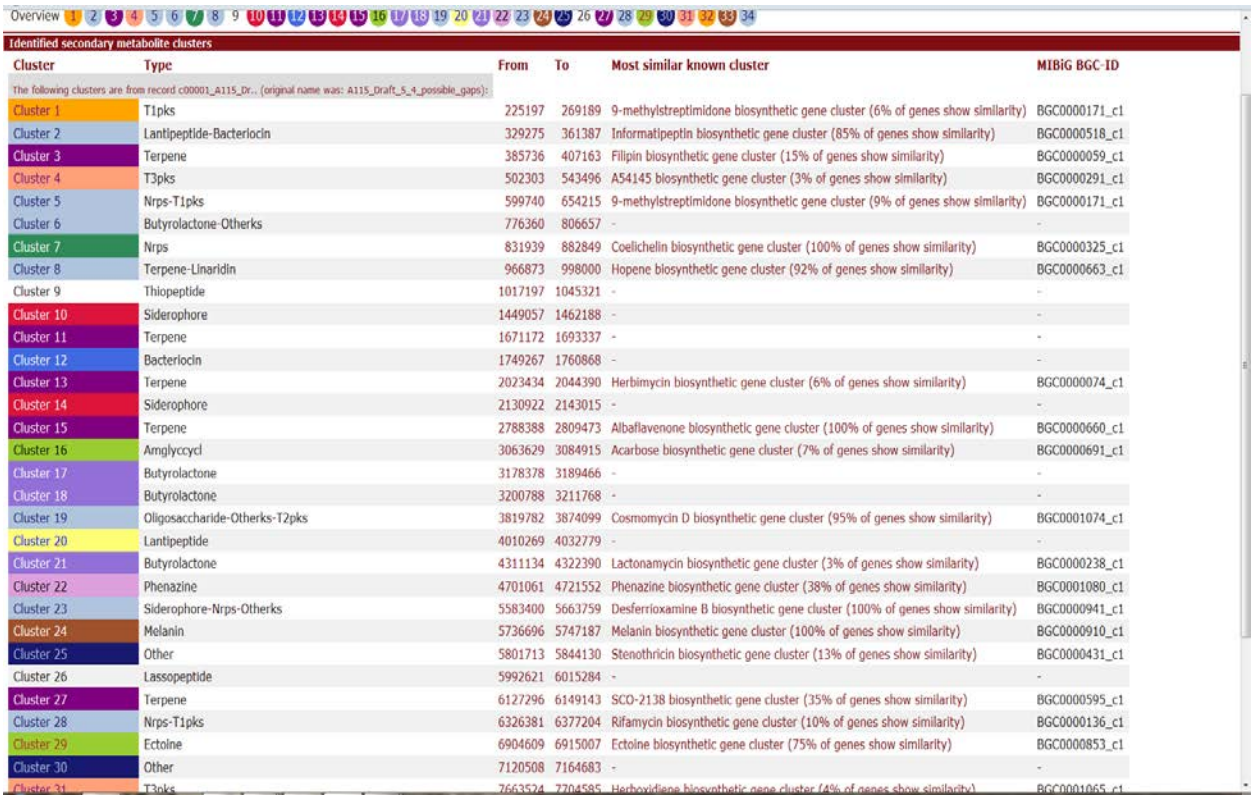

Nonomuraea sp. C1-1
Nonomuraea sp. 210432
Nonomuraea sp. V2-3
Nonomuraea sp. ID06A0189
Nonomuraea sp. R5-34
Nonomuraea sp. ATCC 39727
Nonomuraea jabiensis A4036
Nonomuraea sp. 212202
Nonomuraea sp. 6-1-2
Nonomuraea sp. 31132
Nonomuraea sp. NEAU-ND5
Nonomuraea bangladeshensis 13651M
Nonomuraea bangladeshensis 13651O
Nonomuraea sp. 31163
Nonomuraea sp. NEAU-T6
Nonomuraea fuscirosea NEAU-dht8
Nonomuraea maheshkhaliensis 13661S
Nonomuraea maheshkhaliensis 16-5-14
Nonomuraea kuesteri HBUM49509
Nonomuraea kuesteri HBUM175180
Nonomuraea kuesteri XMU256
Nonomuraea kuesteriGW 14-1925
Nonomuraea kuesteri HBUM49489
Nonomuraea kuesteri HBUM49513
Nonomuraea kuesteri HBUM49512
Nonomuraea kuesteri HBUM49484
Nonomuraea kuesteri HBUM49481
Nonomuraea sp. HBUM80928
Nonomuraea sp. 13674M
Nonomuraea sp. JAJ18
Nonomuraea fastidiosa T9-1
Nonomuraea terrinata L44-1
Nonomuraea sp. NEAU-yn30
Nonomuraea coxensis JCM 13931
Nonomuraea wenchangensis 210417
Nonomuraea sp. FIM 02-765
Nonomuraea sp. TP-A0870
Nonomuraea bangladeshensis 13674P
Nonomuraea sp. NEAU-JM3
Nonomuraea bangladeshensis 5-10-10
Nonomuraea sp. FXJ1.368
Nonomuraea sp. 2704GPT2-01
Nonomuraea candida HMC10
Nonomuraea candida EB139
Nonomuraea turkmeniaca DSM 43926
F4
Nonomuraea antimicrobica YIM 61105
Nonomuraea jiangxiensis FXJ1.102
Nonomuraea sp. PENDO-2176
Nonomuraea sp. PENDO-2175
Nonomuraea rubra AC615
Nonomuraea sp. SpC090624KE 04
Nocardia sp. SpC090624SC 10
Nonomuraea maritima FXJ7.203
Nonomuraea sp. SC1-1
Nonomuraea sp. TTA 03
Nonomuraea rhizophila YIM 67092
Nonomuraea solani NEAU-Z6
Nonomuraea thailandensis KC-061
Pseudonocardiaceae str. PA83
Nonomuraea sp. GB14
Nonomuraea salmonea DSM 43678
Streptosporangiaceae bacterium BBHARD23
Nonomuraea soli YIM 120770
Nonomuraea sp. HBUM171189
Nonomuraea endophytica YIM 65601
Nonomuraea sp. NEAU-st5
Nonomuraea sp. K274
Nonomuraea sp. CA17
Nonomuraea sp. IR27-SDHV-102
Nonomuraea sp. HBUM81200
Nonomuraea harbinensis NEAU-yn31
Nonomuraea sp. 212011
Nonomuraea sp. EGI 80083
Nonomuraea sp. 31116
Nonomuraea guangzhouensis NEAU-ZJ3
Nonomuraea sp. DLS-53
Nonomuraea sp. DLS-49
Nonomuraea sp. TFS 1165
Nonomuraea salmonea 15698
Nonomuraea kuesteri 214110
Nonomuraea sp. FMN03
Nonomuraea sp. FMN02
Nonomuraea sp. S3604
Nonomuraea sp. S2301
Planotetraspora sp. SOL06-13R-4
Planotetraspora sp. ACT-0145
Herbidospora sakaeratensis DMKUA 205
Acrocarpospora sp. Shinshu-MS-05
Acrocarpospora sp. IR12-DHHV-602
Acrocarpospora pleiomorpha R-31
Sphaerisporangium melleum 3D-70(20)
Sphaerisporangium melleum 4-20(13)
Sphaerisporangium melleum 3-28(8)
Sphaerisporangium melleum 5-81(36)
Sphaerisporangium sp. NEAU-YHS12
Sphaerisporangium sp. NEAU-YHS15
0.005

Figure 11. Phylogenetic tree reconstructed based on 16S rRNA gene sequences showing the evolutionary relationship among anti-MRSA isolate F4 and members of the genus *Nonomuraea*. The tree was inferred using the maximum likelihood method. Numbers at each branch nodes indicate bootstrap percentages based on the maximum likelihood method (1000 replicates) that was calculated using MEGA 6.0 software (Tamura et al. 2011). Bar indicates 5 substitutions per 100 positions.

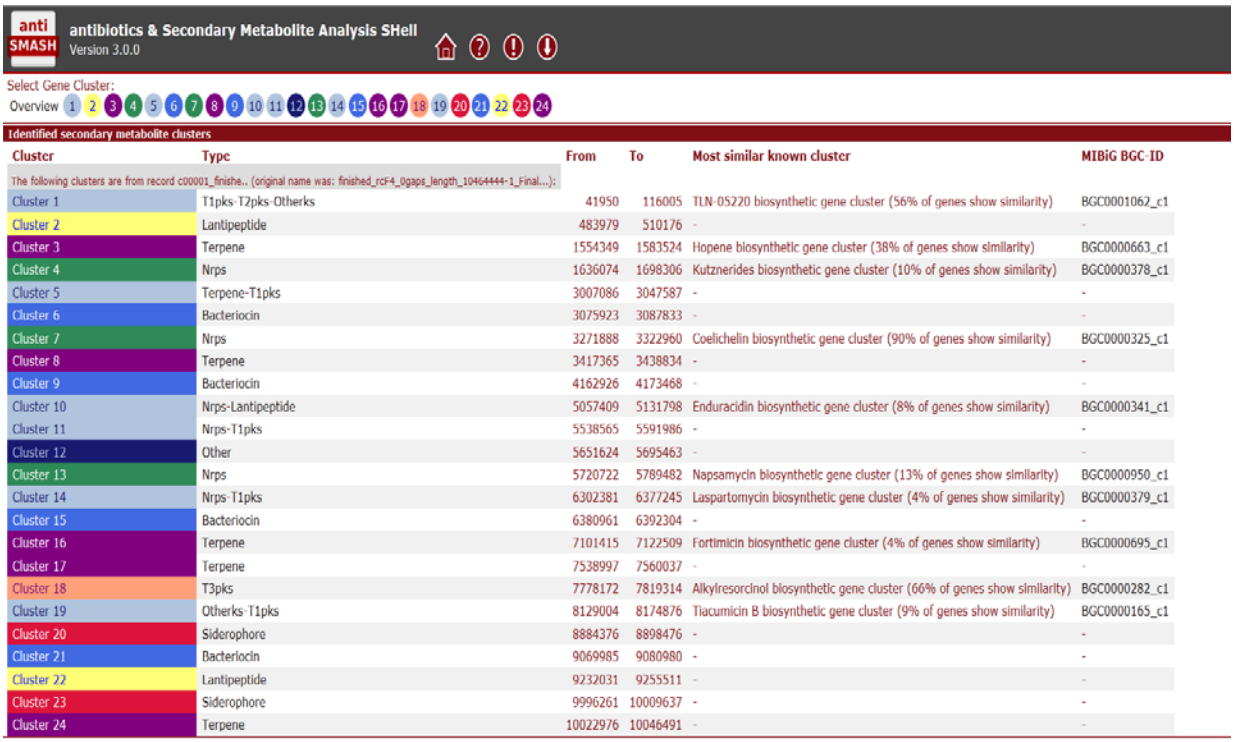


Overview 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34

Identified secondary metabolite clusters

Cluster	Type	From	To	Most similar known cluster	MIBiG BGC-ID
The following clusters are from record c00001_A115_Dr.. (original name was: A115_Draft_5_4_possible_gaps):					
Cluster 1	T1pks	225197	269189	9-methylstreptimidone biosynthetic gene cluster (6% of genes show similarity)	BGC0000171_c1
Cluster 2	Lantipeptide-Bacteriocin	329275	361387	Informatipeptin biosynthetic gene cluster (85% of genes show similarity)	BGC0000518_c1
Cluster 3	Terpene	385736	407163	Filipin biosynthetic gene cluster (15% of genes show similarity)	BGC0000059_c1
Cluster 4	T3pks	502303	543496	A54145 biosynthetic gene cluster (3% of genes show similarity)	BGC0000291_c1
Cluster 5	Nrps-T1pks	599740	654215	9-methylstreptimidone biosynthetic gene cluster (9% of genes show similarity)	BGC0000171_c1
Cluster 6	Butyrolactone-Otherks	776360	806657	-	-
Cluster 7	Nrps	831939	882849	Coelichelin biosynthetic gene cluster (100% of genes show similarity)	BGC0000325_c1
Cluster 8	Terpene-Llnaridin	966873	998000	Hopene biosynthetic gene cluster (92% of genes show similarity)	BGC0000663_c1
Cluster 9	Thiopeptide	1017197	1045321	-	-
Cluster 10	Siderophore	1449057	1462188	-	-
Cluster 11	Terpene	1671172	1693337	-	-
Cluster 12	Bacteriocin	1749267	1760868	-	-
Cluster 13	Terpene	2023434	2044390	Herbimycin biosynthetic gene cluster (6% of genes show similarity)	BGC0000074_c1
Cluster 14	Siderophore	2130922	2143015	-	-
Cluster 15	Terpene	2788388	2809473	Albaflavenone biosynthetic gene cluster (100% of genes show similarity)	BGC0000660_c1
Cluster 16	Amglyccycl	3063629	3084915	Acarbose biosynthetic gene cluster (7% of genes show similarity)	BGC0000691_c1
Cluster 17	Butyrolactone	3178378	3189466	-	-
Cluster 18	Butyrolactone	3200788	3211768	-	-
Cluster 19	Oligosaccharide-Otherks-T2pks	3819782	3874099	Cosmomycin D biosynthetic gene cluster (95% of genes show similarity)	BGC0001074_c1
Cluster 20	Lantipeptide	4010269	4032779	-	-
Cluster 21	Butyrolactone	4311134	4322390	Lactonamycin biosynthetic gene cluster (3% of genes show similarity)	BGC0000238_c1
Cluster 22	Phenazine	4701061	4721552	Phenazine biosynthetic gene cluster (38% of genes show similarity)	BGC0001080_c1
Cluster 23	Siderophore-Nrps-Otherks	5583400	5663759	Desferrioxamine B biosynthetic gene cluster (100% of genes show similarity)	BGC0000941_c1
Cluster 24	Melanin	5736696	5747187	Melanin biosynthetic gene cluster (100% of genes show similarity)	BGC0000910_c1
Cluster 25	Other	5801713	5844130	Stenothricin biosynthetic gene cluster (13% of genes show similarity)	BGC0000431_c1
Cluster 26	Lassopeptide	5992621	6015284	-	-
Cluster 27	Terpene	6127296	6149143	SCO-2138 biosynthetic gene cluster (35% of genes show similarity)	BGC0000595_c1
Cluster 28	Nrps-T1pks	6326381	6377204	Rifamycin biosynthetic gene cluster (10% of genes show similarity)	BGC0000136_c1
Cluster 29	Ectoine	6904609	6915007	Ectoine biosynthetic gene cluster (75% of genes show similarity)	BGC0000853_c1
Cluster 30	Other	7120508	7164683	-	-
Cluster 31	T3pks	7663524	7704585	Herboxidiene biosynthetic gene cluster (4% of genes show similarity)	BGC0001065_c1

Figure 12. Secondary metabolite biosynthesis gene clusters were predicted for strain A115 using the secondary metabolite identification tool antiSMASH3.0.



Cluster	Type	From	To	Most similar known cluster	MIBiG BGC ID
The following clusters are from record c00001_finishe.. (original name was: finished_rcF4_0gaps_length_10464444-1_Final...):					
Cluster 1	T1pks-T2pks-Otherks	41950	116005	TLN-05220 biosynthetic gene cluster (56% of genes show similarity)	BGC0001062_c1
Cluster 2	Lantipeptide	483979	510176	-	-
Cluster 3	Terpene	1554349	1583524	Hopene biosynthetic gene cluster (38% of genes show similarity)	BGC0000663_c1
Cluster 4	Nrps	1636074	1698306	Kutznerides biosynthetic gene cluster (10% of genes show similarity)	BGC0000378_c1
Cluster 5	Terpene-T1pks	3007086	3047587	-	-
Cluster 6	Bacteriocin	3075923	3087833	-	-
Cluster 7	Nrps	3271888	3322960	Coelichelin biosynthetic gene cluster (90% of genes show similarity)	BGC0000325_c1
Cluster 8	Terpene	3417365	3438834	-	-
Cluster 9	Bacteriocin	4162926	4173468	-	-
Cluster 10	Nrps-Lantipeptide	5057409	5131798	Enduracidin biosynthetic gene cluster (8% of genes show similarity)	BGC0000341_c1
Cluster 11	Nrps-T1pks	5538565	5591986	-	-
Cluster 12	Other	5651624	5695463	-	-
Cluster 13	Nrps	5720722	5789482	Napsamycin biosynthetic gene cluster (13% of genes show similarity)	BGC0000950_c1
Cluster 14	Nrps-T1pks	6302381	6377245	Laspartomycin biosynthetic gene cluster (4% of genes show similarity)	BGC0000379_c1
Cluster 15	Bacteriocin	6380961	6392304	-	-
Cluster 16	Terpene	7101415	7122509	Fortimicin biosynthetic gene cluster (4% of genes show similarity)	BGC0000695_c1
Cluster 17	Terpene	7538997	7560037	-	-
Cluster 18	T3pks	7778172	7819314	Alkylresorcinol biosynthetic gene cluster (66% of genes show similarity)	BGC0000282_c1
Cluster 19	Otherks-T1pks	8129004	8174876	Tiacumicin B biosynthetic gene cluster (9% of genes show similarity)	BGC0000165_c1
Cluster 20	Siderophore	8884376	8898476	-	-
Cluster 21	Bacteriocin	9069985	9080980	-	-
Cluster 22	Lantipeptide	9232031	9255511	-	-
Cluster 23	Siderophore	9996261	10009637	-	-
Cluster 24	Terpene	10022976	10046491	-	-

Figure 13. Predicted secondary metabolites biosynthetic gene clusters present in F4 genome using the secondary metabolite identification tool antiSMASH3.0.

6. Reference

Angiuoli, S. V., A. Gussman, et al. (2008). "Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation." Omics **12**(2): 137-141.

Biehle, J. R., S. J. Cavalieri, et al. (1996). "Novel method for rapid identification of Nocardia species by detection of preformed enzymes." J Clin Microbiol **34**(1): 103-107.

Bilyk, B., S. Weber, et al. (2013). "In vivo random mutagenesis of streptomycetes using mariner-based transposon Himar1." Appl Microbiol Biotechnol **97**(1): 351-359.

Bush, K. (2011). "Introduction to Antimicrobial Therapeutics Reviews: antibiotics that target the ribosome." Ann N Y Acad Sci **10**(06367).

Case, R. J., Y. Boucher, et al. (2007). "Use of 16S rRNA and rpoB Genes as Molecular Markers for Microbial Ecology Studies." Applied and Environmental Microbiology **73**(1): 278-288.

Fair, R. J. and Y. Tor (2014). "Antibiotics and bacterial resistance in the 21st century." Perspect Medicin Chem **6**: 25-64.

Falkinham, J. O., 3rd, T. E. Wall, et al. (2009). "Proliferation of antibiotic-producing bacteria and concomitant antibiotic production as the basis for the antibiotic activity of Jordan's red soils." Appl Environ Microbiol **75**(9): 2735-2741.

George, I. F., M. Hartmann, et al. (2011). "Recovery of as-yet-uncultured soil acidobacteria on dilute solid media." Appl Environ Microbiol **77**(22): 8184-8188.

Guo, X., N. Liu, et al. (2015). "Red soils harbor diverse culturable actinomycetes that are promising sources of novel secondary metabolites." Appl Environ Microbiol **81**(9): 3086-3103.

Hageman, J. C., T. M. Uyeki, et al. (2006). "Severe community-acquired pneumonia due to Staphylococcus aureus, 2003-04 influenza season." Emerg Infect Dis **12**(6): 894-899.

Handelsman, J., M. R. Rondon, et al. (1998). "Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products." Chem Biol **5**(10): R245-249.

Hartmann, M. and F. Widmer (2006). "Community structure analyses are more sensitive to differences in soil bacterial communities than anonymous diversity indices." Appl Environ Microbiol **72**(12): 7804-7812.

Jankowitsch, F., J. Schwarz, et al. (2012). "Genome sequence of the bacterium Streptomyces davawensis JCM 4913 and heterologous production of the unique antibiotic roseoflavin." J Bacteriol **194**(24): 6818-6827.

Karlyshev, A. V., M. J. Pallen, et al. (2000). "Single-primer PCR procedure for rapid identification of transposon insertion sites." Biotechniques **28**(6).

Kim, M., H. S. Oh, et al. (2014). "Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes." Int J Syst Evol Microbiol **64**(Pt 2): 346-351.

Lagesen, K., P. Hallin, et al. (2007). "RNAmmer: consistent and rapid annotation of ribosomal RNA genes." Nucleic Acids Res **35**(9): 3100-3108.

Lewis, K. (2012). "Antibiotics: Recover the lost art of drug discovery." Nature **485**(7399): 439-440.

Lewis, K., S. Epstein, et al. (2010). "Uncultured microorganisms as a source of secondary metabolites." J Antibiot **63**(8): 468-476.

Ling, L. L., T. Schneider, et al. (2015). "A new antibiotic kills pathogens without detectable

resistance." Nature **517**(7535): 455-459.

Meena, B., L. A. Rajan, et al. (2013). "Novel marine actinobacteria from emerald Andaman & Nicobar Islands: a prospective source for industrial and pharmaceutical byproducts." BMC Microbiol **13**(145): 1471-2180.

Nikodinovic, J., K. D. Barrow, et al. (2003). "High yield preparation of genomic DNA from Streptomyces." Biotechniques **35**(5): 932-934.

Omura, S. (1992). "Thom Award Lecture. Trends in the search for bioactive microbial metabolites." J Ind Microbiol **10**(3-4): 135-156.

Paradkar, A., A. Trefzer, et al. (2003). "Streptomyces genetics: a genomic perspective." Crit Rev Biotechnol **23**(1): 1-27.

Pettit, R. K. (2004). "Soil DNA libraries for anticancer drug discovery." Cancer Chemother Pharmacol **54**(1): 1-6.

Richter, M. and R. Rossello-Mora (2009). "Shifting the genomic gold standard for the prokaryotic species definition." Proc Natl Acad Sci U S A **106**(45): 19126-19131.

Rufián-Henares, J. A. and F. J. Morales (2008). "Microtiter plate-based assay for screening antimicrobial activity of melanoidins against E. coli and S. aureus." Food Chemistry **111**(4): 1069-1074.

Shirling, E. B. and D. Gottlieb (1966). "Methods for characterization of Streptomyces species." International Journal of Systematic Bacteriology **16**(3): 313-340.

Tamura, K., D. Peterson, et al. (2011). "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods." Mol Biol Evol **28**(10): 2731-2739.

Tiwari, K. and R. K. Gupta (2012). "Rare actinomycetes: a potential storehouse for novel

antibiotics." Crit Rev Biotechnol **32**(2): 108-132.

Torsvik, V. and L. Ovreas (2002). "Microbial diversity and function in soil: from genes to ecosystems." Curr Opin Microbiol **5**(3): 240-245.

Watve, M. G., R. Tickoo, et al. (2001). "How many antibiotics are produced by the genus Streptomyces?" Arch Microbiol **176**(5): 386-390.

Chapter 3

Isolation and characterization of anti-MRSA active compounds from *Bacillus* spp. strains and their potential use as a skin probiotic.

1. Abstract

Skin and soft tissues are the most common sites of *Staphylococcus aureus* infection and inhibition of *S. aureus* skin colonization can potentially prevent life-threatening bacteremia. The emergence of methicillin-resistant *S. aureus* (MRSA) strains has increased the demand for new strategies to combat and prevent infection. We screened a library of bacterial isolates to identify novel chemical compounds for MRSA control. We identified five *Bacillus* strains that expressed metabolites with anti-MRSA activity and used LC-MS to characterize antibacterial compounds expressed by these strains. The *B. amyloliquefaciens* strain AP183 was found to produce a novel macrodiolide compound described herein as bacillusin A with potent anti-MRSA activity of a minimum inhibitory concentration of 0.6 µg/mL. Because Bacillusin A has a short half-life after extraction, and we hypothesized that it may not persist within living tissue and would therefore be suitable for *in vivo* application. AP183 was tested *in vivo* as a skin probiotic to prevent MRSA infection using a mouse model. Mice were simultaneously challenged with bioluminescent *S. aureus* strain Xen29 with and without AP183 spores in two separate wounds. In additional experiments, we tested the effects of AP183 spores with and without accompanying secondary metabolites. After challenge, skin wound healing was monitored for one week and *S. aureus* growth was assessed by bioluminescent imaging. After one week, mice were sacrificed and wounds were homogenized and plated to determine culturable bacterial counts and confirmed by

conducting a culture-independent skin microbiome analysis. Our *in vivo* studies showed that co-administration of secondary metabolites and AP183 spores resulted in a significant reduction in the number of *S. aureus* colonization compared to a negative control. Molecular analysis has also shown a significant reduction in *S. aureus* relative abundance when AP183 was applied while the relative abundance of other bacterial taxa increased in the skin microbiome as a result of probiotic administration. In future work, we will determine the *in vivo* efficacy and safety for the application of strain AP183 and its active metabolites.

2. Introduction

The indiscriminate use of antibiotics has led to an increase bacterial resistance especially for Gram-positive pathogens, *Staphylococcus aureus*, *Enterococcus* and coagulase-negative *Staphylococcus* that pose serious problem in treating infections caused by these pathogens (Tarai et al. 2013). *S. aureus* is among the most common causative agents involved in skin infections in the United States (Edelsberg et al. 2009). An untreated minor skin infection caused by *S. aureus* can spread quickly and progress into more serious condition including distal abscesses of the kidneys and spleen, sepsis and endocarditis (DeLeo et al. 2009). Since skin and soft tissues are the most common sites of *S. aureus* infection, inhibition of *S. aureus* skin colonization can potentially prevent life-threatening bacteremia. The emergence of (MRSA) strains and the life-threatening diseases accompanying MRSA strains has subsequently increased the demand for new antibiotics (DeLeo et al. 2009; Wright 2015).

Members of the genus *Bacillus* are well known to be a prolific source of bioactive natural products (Hamdache et al. 2011; Stein 2005). The *Bacillus* genus is comprised of Gram-positive aerobic or facultative anaerobic, spore forming, rod shaped bacteria that are common in soil. *Bacillus* spp. that have been screened for antimicrobial activities over the past few decades

(Mannanov et al. 2001; Chen et al. 2009). Strains within the *B. subtilis* group, which includes the species *B. amyloliquefaciens,* have been used as biocontrol agents against plant and animal pathogens in agriculture and aquaculture (Ongena et al. 2008; Cook et al. 1995; Ran et al. 2012; Mohammad et al. under review). Plant growth-promoting rhizobacteria (PGPR) the bacteria can promote plant growth directly by helping plants acquire nutrition from soil, or indirectly by controlling phytopathogens to prevent plant diseases (Kloepper et al. 1980). PGPR result in significant enhancement of plant growth and increases yields of agronomically important crops. In addition to plant growth promotion and disease control, several publications suggest that *B. amyloliquefaciens* can also improve well-being of animals through use asa potential probiotic and/or as a curative agent (Islam et al. 2011; Ahmed et al. 2014; González-Ortiz et al. 2013). The majority of antibiotics produced by *Bacillus* spp. are low molecular weight polypeptides that are synthesized by ribosomal or non-ribosomal mechanisms. Many *Bacillus* spp. are known to produce polyketides with antibiotic activities such as macrolactins, difficidins, and oxidifficidins, as well as lipopeptides such as surfactins, iturins, and fengycins (Hamdache et al. 2011; Stein 2005; Sumi et al. 2014). *B. subtilis* strains are the prominent producers of surfactin antibiotics, which exhibit antibacterial and antiviral activity (Wang et al. 2008; Ongena et al. 2007). Iturins produced by various strains of *B. subtilis* are amphiphilic compounds with a peptide ring of seven amino acid residues, including an invariable D-Tyr2 residue (Maget-Dana et al. 1994). The members of iturins family exhibit potent antifungal activity. Iturins A showed strong antimicrobial activity against fungal pathogens *Phythium ultimum*, *Rhizoctonia solani*, *Fusarium oxysporum*, *Sclerotinia sclerotiorum* and *Macrophomina phaseolina* (Li et al. 2014). In addition to iturins, *Bacillus* species produce several other antibiotics that include lantibiotics (Stein et al.

2002), kanosamine (Milner et al. 1996), zwittermycin A (Silo-Suh et al. 1998), bacillomycin (Volpon et al. 1999), plipastatins (Volpon et al. 2000), and bacillusin A (Ravu et al. 2015).

Bacillusin A is a recently discovered macrocyclic polyene antibiotic from *B. amyloliquefaciens* strain AP183 which showed potent antibacterial activities against MRSA and vancomycin-resistant *Enterococcus faecium* with minimum inhibitory concentrations in a range of 0.6 to 1.2 µg/mL (Ravu et al. 2015). Bacillusin A has a short half-life after extraction, which might prove pharmaceutically beneficial for infected wounds where rapid elimination of antibiotic residues can be an advantage.

Following determination of bacillusin A as a potent antibacterial agent capable of inhibiting MRSA growth, genome sequencing of its producer *B. amyloliquefaciens* strain AP183 was performed to determine the gene(s) responsible for the synthesis of bacillusin A. From an analysis of the AP183 genome, the trans-AT polyketide synthases (PKS) pathways were predicted to be responsible for synthesis of bacillusin A (Nasrin et al. 2015).

Strains of *B. amyloliquefaciens* have been used previously as probiotics in animals and plants and are not associated with disease (Ahmed et al 2014; González-Ortiz et al. 2013). The beneficial effects of probiotics include antagonism to pathogens, enhancement of immune response and restoration of body's normal flora (Sun et al. 2010; Casula et al. 2002). There are many strains of *B. amyloliquefaciens* that have already been developed for use with crops, livestock, or products for human consumption (Krober et al. 2014; Lee et al. 2012); however,the use of*B. amyloliquefaciens* strains as a skin probiotic to inhibit cutaneous wound colonization of MRSA has not been reported previously. Since *Bacillus* spores are highly stable and large-scale production of spores is well described (Monteiro et al. 2014), it may be possible to incorporate *Bacillus* spores into different formulations for topical application (e.g., Band-aid, lotions). The

practical considerations for scale-up and eventual application(s) are highly favorable for use in preventing or treating skin infections caused by MRSA. Thus, we developed we developed a novel and clinically applicable method for combating dermal MRSA infections by utilizing spores and metabolites of AP183.

3. Materials and methods

Microorganisms and growth conditions

In this study, a collection of 177 *Bacillus* spp. strains were screened *in vitro* for antimicrobial activities against number of bacterial and fungal pathogens (Table 1). Out of 177 *Bacillus* spp. strains, 160 strains were PGPR *Bacillus* spp. strains which were provided by the laboratory of Dr. Joseph Kloepper (Department of Entomology and Plant Pathology, Auburn University). An additional 17 *Bacillus* spp. strains were isolated from catfish gut (Ran et al. 2012). The tester bacteria methicillin-resistant *Staphylococcus aureus* strain number EAMC30 was a clinical MRSA strain obtained from East Alabama Medical Center, Opelika, AL, provided by Dr. James Barbaree (Department of Biological Sciences, Auburn University). Bioluminescent *S. aureus* strain Xen29 was purchased from PerkinELmer (USA). In addition, methicillin-susceptible *Staphylococcus aureus* ATCC 29213, methicillin-resistant *Staphylococcus aureus* ATCC 29213, *Enterococcus faecalis* ATCC 51299, *Enterococcus faecalis* ATCC 29212, *Enterococcus faecium* ATCC 700221, and the yeast *Candida albicans* ATCC 90028 were purchased from the American Type Culture Collection (Manassas, VA). Both *Bacillus* spp. and *S. aureus* strains were routinely grown and maintained in Tryptic Soy Broth (TSB) and /or in TSB media supplemented with 15 g/L agar (TSA) at 30°C and 37°C, respectively. All bacterial isolates were cryopreserved at -80°C.

Screening of *Bacillus* spp. strains with antimicrobial activities

Antibiosis assays were carried out with double-layer soft agar method with minor modifications (Jack et al. 1996). Briefly, each of the *Bacillus* isolates were grown in TSB media on a rotatory shaker at 200 rpm for 24 hours at 30°C. Prior to the inoculation of bacilli, a sterile cork borer of 10 mm diameter was used to bore wells in those water agar plates and wells were filled with ~1.0 mL of TSA. 10 µL of each *Bacillus* culture was then spotted onto duplicate plates of water agar and incubated for 48 hours at 30°C. After 48 hours of incubation, soft agar (0.7% w/v agar) prepared with TSB was melted, cooled, and seeded with a freshly prepared inoculum of log-phase MRSA strain 30 to achieve the absorbance at 600 nm (OD_{600}) of 0.5. The bacterial cell suspension in soft agar was immediately poured over the water agar plates and incubated for 24 hours at 37°C. After incubation, the zones of clearing in the growth of MRSA strain 30 were recorded (in mm) as evidence of growth inhibition by corresponding *Bacillus* spp. strains.

To verify antimicrobial activity, the entire collection of *Bacillus* strains was further screened by drop assay. In this assay *Bacillus* strains were grown in TSB on a rotatory shaker at 200 rpm for 48 hours at 30°C. Cells were harvested by centrifugation at 8000 rpm for 10 min. Supernatants were filtered and frozen for further analysis. A broth culture of actively growing MRSA strain 30 was adjusted to OD_{600} of 0.5 and evenly swabbed onto TSA plates. Next 10 µL of cell-free supernatant derived from *Bacillus* culture was added onto the MRSA strain 30 culture. Zones of inhibition were measured after 24 hours of incubation at 37°C. An aliquot of supernatants from different *Bacillus* strains were also shipped to the National Center for Natural Products Research (NCNPR) at the University of Mississippi for screening against a larger collection of bacterial and fungal pathogens.

Phylogenetic analysis of *Bacillus* spp. strains producing anti-MRSA compound

Five *Bacillus* isolates that include AP143, AP183, AP191, AP218 and AB01 with consistently strong antimicrobial activities in three independent *in vitro* antibiosis assays were selected for phylogenetic analysis using 16S rRNA gene-specific sequences. For colony PCR, universal bacterial primers 27F and 1492R, which generated approximately 1.5kb products, were used. PCR reactions were performed in 50µL reaction volumes which contained 25.0 µL of 2x EconoTaq plus Green DNA Polymerase (Lucigen Co. WI), 0.2µM of each primer and sterile distilled water for adjusting to 50 µL. Amplification of 16S rRNA gene was carried out under the following conditions: denaturation at 94°C for 2 min, followed by 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 2.0 min, and final extension at 72°C for 10 min. Amplified PCR products of *Bacillus* isolates were analyzed by electrophoresis with 0.7% agarose gel run at 200V for one hour. After electrophoresis, the gel was stained by ethidium bromide and then visualized and photographed under UV transilluminator using Gel Doc XR system (Biorad USA). The PCR products were purified by E.Z.N.A cycle pure kit (Omega bio-tak USA) according to the manufacturer's instructions and purified PCR products were then sequenced in both directions using primers 27F and 1492R (Lucigen Corp, Middleton, WI USA). The sequences were trimmed for quality using ChromasPro (Technelysium, Australia). Trimmed sequences were assembled and compared against sequences available in the National Center for Biotechnology Information (NCBI data base), GenBank using the BLASTn algorithm.

Preliminary LC-MS analysis

Supernatants from five anti-MRSA compound-producing bacilli were analyzed by LC-MS method at the National Center for Natural Product Research (NCNPR, Oxford, MS) according to the standard protocol established by the NCNPR (Ravu et al. 2015). Initial LC-MS

data revealed the presence of a potential novel compound in the supernatant derived *Bacillus* strain AP183.

Purification of active secondary metabolites from strain AP183

To purify anti-MRSA compounds from strain AP183, preliminary fractionation, isolation of fractions, and *in vitro* antimicrobial activity of the purified compound were determined as described by Ravu et al. 2015. Large-scale cultures of AP183 (≥1L) were grown in TSB for 48 hours at 30°C. Supernatant was filtered and shipped to the NCNPR for biochemical analyses, including LC/MS (liquid chromatography/mass spectrometry) and NMR (nuclear magnetic resonance) for biochemical structural elucidation.

Isolation of genomic DNA from strain AP183

To determine the whole genome sequence of *B. amyloliquefaciens* subsp. *plantarum* strain AP183, genomic DNA was extracted according to the methods described previously (Wilson 2001). The DNA concentration was measured by Qubit 2.0 fluorometer (Life technologies, USA) following the manufacturer's instructions.

Phylogenetic analysis of strain AP183

The *gyrB* gene from strain AP183 was PCR amplified and sequenced using *Bacillus* spp. specific universal primer sets UP-1 and UP-2r according to the methods described previously. The *gyrB* sequence-based phylogenetic tree was inferred with MEGA5.05 (Tamura, Peterson et al. 2011) using the Maximum Likelihood (ML) method (Felsenstein 1981) with 1000 iterations for bootstrap support.

Whole genome sequencing, assembly, and annotation

Next-generation sequencing of *Bacillus* strain AP183 was performed using the Illumina MiSeq sequencing platform. An indexed Illumina library was prepared using Nextera DNA

Sample Prep Kit (Epicentre, Madison, WI) and sequences were generated using an Illumina MiSeq with a 2 × 250 paired end sequencing kit. Sequence reads were trimmed for quality and assembled *de novo* using the CLC Genomics Workbench (CLCBio, Cambridge, MA). Gene prediction and annotation were performed using GeneMark (Lukashin and Borodovsky 1998) and RAST annotation server (Aziz, Bartels et al. 2008), respectively. The identity of individual ORFs from secondary metabolite biosynthesis gene clusters was confirmed by BLASTx against the GenBank database. The whole genome shotgun of AP183 was deposited at DDBJ/EMBL/GenBank under the accession no. JXAM00000000.

Prediction of secondary metabolite biosynthesis gene clusters in *Bacillus* strainAP183

Secondary metabolite biosynthesis gene clusters for strain AP183 were predicted using the secondary metabolite identification tool antiSMASH2.0 (Blin, Medema et al. 2013). Gene prediction and annotation were carried out by GeneMark (Lukashin and Borodovsky 1998) and BLASTx (NCBI), respectively.

Antibiotic resistance profile

The susceptibility of AP183 to broad range of antibiotics was determined by Kirby-Bauer disc diffusion method, outlined by National Committee for Clinical Laboratory Standards (CLSI 2012). A log-phase culture of AP183, diluted to a concentration of approximately 1×10^8 CFU/mL, was seeded onto a Mueller-Hinton agar plate. Antibiotic-impregnated discs (BD Biosciences) were placed onto the seeded plates with three replicates. The zone of inhibition was measured and recorded after 18 hours of incubation at 30°C.

Preparation of *Bacillus* spore formulation for mouse challenge

Five Bacillus species including AP143, AP183, AP191, AP218 and AB01were evaluated as a topical probiotic to inhibit the colonization of S. aureus on mouse skin wounds. Bacillus

spores were prepared according to the method described by Ran et al. (Ran et al. 2012). The concentration of the spore suspension was determined by serially diluting the spore suspension in sterile water and plating them in TSA plates for overnight incubation at 30°C. Spores were preserved at 4°C and the final concentration of the spore suspension for challenge studies were adjusted to 1.0×10^8 CFU/mL by diluting with 30% glycerol.

Preparation of *Bacillus* spores and metabolites formulation for mouse challenge

For this *in vivo* cutaneous infection model, a combination of AP183 spores and metabolites were used. For preparing the spores and metabolites formulation, a colony of strain AP183 was inoculated into a 20 mL culture tube containing 5.0 mL of TSB and incubated on a rotatory shaker at 200 rpm at 30°C for five days. Before mouse challenge studies, an aliquot of cell free supernatant was tested for anti-MRSA activity *in vitro* and the presence of spores were confirmed by spore staining. AP183 culture was dispensed into 1.0 mL aliquots containing 30% glycerol.

Preparation of *S. aureus* strain Xen29 inoculum for *in vivo* cutaneous wound challenge

S. aureus Xen29 used for *in vivo* cutaneous wound challenge studies was derived from the parental strain of *S. aureus* 12600 and possesses a stable copy of the *luxABCDE* operon at a single integration site on the chromosome. For mouse challenge studies, Xen 29 was grown overnight in Brain Heart Infusion (BHI) broth containing 50 µg/mL kanamycin with shaking at 225 rpm at 37°C. The overnight culture was then washed twice and diluted 1:100 into 1× sterile PBS with 10 % glycerol to achieve absorbance at 600 nm (OD_{600}) of 0.4 which are equivalent to 1.4×10^8 CFU/mL. For creating each wound, 10 µL of injection volume was used which contained ~1.0×10^6 CFU/mL of Xen29 cells.

Mice challenge studies

Experimental protocols were reviewed, approved and performed under regulatory supervision of Auburn University's Institutional Biosafety Committee (IBC) and Institutional Animal Care and Use Committee (IACUC). For *in vivo* challenge studies, female C57BL/65 mice of 6 to 8 weeks old were purchased from Jackson Laboratories (Bar Harbor, ME) and housed 3 to 5 animals per cage. These mice were fed standard alfalfa free rodent diet and distilled water.

In the first challenge, spores of five *Bacillus* spp. (strains AP143, AP191, AP183, AP218 and AB01) were selected for evaluation of their ability to inhibit colonization of *S. aureus* Xen29 in mouse cutaneous wound model. To induce skin infections, the mice were first anesthetized with isoflurane (2% (vol/vol)/2 liters O_2) (Panizzi et al. 2011) and the hair on the back of mice were shaved (electric clipper) and depilated (Veet; Reckitt Benckiser, Germany). Each mouse in the treatment group was simultaneously challenged with bioluminescent *S. aureus* strain Xen29 in two independent cutaneous wounds on each mouse back. The suspension of spores and/or Xen29 was injected subcutaneously in to 3 mice per treatment group. The concentrations of Xen29 and *Bacillus* spp. spores per wound were in the range of 1.0×10^7 and 1.0×10^8 CFU, respectively. Four to six hours of post challenge, mice were images for bioluminescence using IVIS Lumina XRms (Parkin Elmer) imaging system for monitoring the progress of Xen29 establishment in infection sites. The progression of skin abscess lesions was monitored daily for one week after challenge by Bioluminescent Imaging (BLI).

Three different formulations were evaluated in this *in vitro* model: a) spores, b) secondary metabolites and, c) combination of spores and metabolites of AP183. Bioluminescent *S. aureus* strain Xen29 was injected subcutaneously in mouse wound model (n=10 mice per treatment). After challenge, skin wound healing was monitored for one week and mice in each

treatment group were assessed for *S. aureus* growth by BLI flux (rate of photon per second, p/s) analysis.

Determination of staphylococcal survival in mouse wound

Mice were euthanized by induction of anesthesia with isofluran (2% (vol/vol)/2 liters O_2) followed by cervical dislocation. Then wounds including few millimeters of surrounding skin were excised with half of the excised wounds homogenized in PBS prior to serial dilution and spread plating. Plates were incubated for 18 hours at 37°C followed by bioluminescent imaging and counting of staphylococcal colonies present in each treatment group. The other half of the wound was prepared for histological analysis by Gram stain and H&E.

Microbial diversity analysis of the mouse skin

To assess the wound site-associated microbiota, genomic DNA was extracted from mouse wound tissue homogenates of each treatment group using genomic DNA isolation kit (Ultraclean microbial DNA isolation kit, MO BIO). Genomic DNA was extracted according to the manufacturer instructions and the extracted genomic DNAs were used as templates for PCR amplification of 16S rRNA genes with bar-coded “universal Bacteria” primer sets. Pooled amplicons were sequenced using paired-end reads on an Illumina MiSeq and several thousand 16S rRNA sequences were generated per sample. The trimmed sequence reads were analyzed using the QIIME pipeline and operational taxonomic units (OTU) were generated at 97% cutoff using BLASTn and compared to the curated database at the ribosomal database project (DeSantis et al. 2006).

Metabolomic analysis of mice wounds

Filtered wound tissue homogenates were ethyl acetate extracted prior to analysis by a nano LC-MS system (Acquity, Waters, Milford, MA, USA) according to the standard protocol

established by NCNPR. The reversed phase analysis was performed using an Acquity UPLC BEH C_{18} column (2.1 × 150 mm, 1.7 μm) with the injection volume of 5.0 μL. The mobile phase was consisted of solvent A and solvent B that contain water with 0.05 % formic acid and CH_3CN with 0.05% formic acid respectively. The flow rate for BEH-C_{18} column was 0.2 mL/min in gradient mode and gradient elution from 1% to 100% CH_3CN in H_2O. Gradient solvent system starts from solvent B 1% to 100% in 25 mins, then wash the column with solvent B 100% until 30 mins. The separated compounds were detected by a diode array detector (Agilent technologies, CA) of the UV wavelength of 220, 254, 325 and 380 nm. MS/MS studies were conducted by positive and negative electrospray ionization (ESI) conditions with scan mode of *m/z* of 100-150.

Statistical analysis

Statistical significance was determined by Student's *t* test or one way ANOVA. *P* values of <0.05 were considered significant.

4. Results

Characterization and antimicrobial activity of *Bacillus* spp strains

A total of five *Bacillus* spp. strains AP143, AP183, AP191, AP218 and AB01 showed strong *in vitro* antimicrobial activity against a number of bacterial and fungal pathogens (Table 1). The anti-MRSA activity of *Bacillus* spp. strains was confirmed by soft agar overlay and drop assay (Figure 1). Strains AP143, AP183, AP191 and AP218 were initially isolated from the plant rhizosphere as PGPR strains that can inhibit phytopathogens and promote plant growth. Each of the *Bacillus* spp. strains that exhibited inhibitory activity to MRSA strain 30 was capable of endospore formation. A phylogenetic analysis based on 16S rRNA and *gyrB* gene sequences of each of the *Bacillus* isolate indicated that four of the *Bacillus* strains were within the *B. subtilis*

group (inclusive of *B. amyloliquefaciens*). The *gyrB*-based phylogenetic approach demonstrated that strain AP183 is affiliated with *B. amyloliquefaciens* subsp. *plantarum* with strong bootstrap support (Figure 2). The phylogenetic affiliation of strain AP191 was determined as *B. methylotrophicus.*

In vitro antibacterial activities of _B. amyloliquefaciens_ subsp. _plantarum_ strain AP183

Using a double-dilution method, culture supernatant of AP183 was screened for *in vitro* antibacterial activity against *S. aureus* strain Xen29 (Figure 3) with IC_{50} of <1/32th dilution. The expression of secondary metabolites active against MRSA during a time course of bacterial growth was determined and it was found that AP183 strain starts to produce anti-MRSA compounds as early as 8 hours post-inoculation into TSB at 30°C and shaking with aeration at 200 rpm.

The extraction methods including different resin treatments and organic solvent partitioning were also investigated at the NCNPR for efficient recovery of anti-MRSA activity from supernatants and cell lysates by growing bulk cultures (≥1L) in appropriate growth conditions. The solvent partitioning method was the most effective extraction method. For example, the ethyl acetate extract of the supernatant showed IC_{50}s of 18.0 and 7.7 µg/mL against *S. aureus* ATCC 29213 and MRSA ATCC 33591, respectively. The methanol extract of the AP183 cell pellet also showed potent activity with IC_{50}s of 7.7 and 13.5 µg/mL, respectively, against *S. aureus* ATCC 29213 and MRSA ATCC 33591 (Figure 4). Preliminary fractionation of the ethyl acetate extract by reversed-phase silica gel chromatography generated 16 fractions, with the 9^{th} fraction being the most active with IC_{50} values of <1.1 µg/mL against the two tested strains (Figure 5). This activity-enriched fraction was predicted to contain a new compound (Ravu et al. 2015).

LC-MS analysis of the supernatant and cell lysate of *B. amyloliquefaciens* strain AP183

A correlation between the assays using the volume-based liquid culture supernatant and the weight-based solid extracts was established at the NCNPR. LC-MS analysis indicated that the two extracts displayed similar chemical profiles and the known antibiotics surfactins, iturins, and fengycins that are previously discovered from *Bacillus* spp. were also identified in this strain (Figure 6). Since the three classes of antibiotics are weakly active against *S. aureus* and MRSA, which has been confirmed by testing the commercially available surfactins and iturins in our assays, the potent activity of the AP 183 extracts suggested the presence of new antibiotic compounds.

Identification and structure determination of a novel antibacterial compound

Scale-up fermentation of a 30 L culture at the NCNPR led to the isolation of a new microcyclic polyene antibiotic, designated bacillusin A. The structure of bacillusin A was assigned by interpretation of NMR and MS spectroscopic data as a novel macrodiolide composed of dimeric 4-hydroxy-2-methoxy-6-alkenyl-benzoic acid lactones with conjugated pentaene-hexahydroxy polyketide chains (Figure 7). The presence of conjugated aromatic-pentaene system as indicated by LC-MS suggests that bacillusin A is chemically unstable in organic solvents (Figure 8). Additional compounds with the same molecular weight were detected when bacillusin A was allowed to stand at room temperature or even at 4^{o}C for a few days (Figure 8). Reduced antibacterial activity was also observed when bacillusin A was stored in DMSO for more than one week.

***In vitro* antibacterial activities of bacillusin A**

Bacillusin A showed strong *in vitro* antibacterial activities against multiple drug-sensitive and -resistant *S. aureus* and *Enterococcus* spp., and its potency was compared against vancomycin, ciprofloxacin, and methicillin (Table 2).

Bacillusin A exhibited a minimum inhibitory concentration (MIC) of 0.6 µg/mL against *E. faecium* ATCC 700221, which is was resistant to three of the aforementioned antibiotics tested at100 µg/mL. Another significant feature is that bacillusin A is a bactericidal compound exhibiting antibacterial activity at extremely low concentrations. Bacillusin A shows an IC_{50} of <0.02 µg/mL against the clinical isolate MRSA strain 30 compared to 0.2, 29.4, and 10.1µg/mL for vancomycin, ciprofloxacin, and methicillin, respectively (Table 2).

Whole genome sequencing of strain AP183

To determine the genomic basis of antibacterial mechanisms, the genome of AP183 was sequenced by Illumina Miseq sequencing technology. The sequence reads were trimmed for quality and assembled using the CLC Genomics Workbench (CLC bio, Cambridge, MA), obtaining 1,331,792 sequence reads, with an average coverage of 36×. *De novo* assembly of strain AP183 genome sequences resulted in 40 contigs larger than 500 bp, with an N50 of 190,739 bp, and the largest contig was 541,177 bp.

The estimated genome size was ~3.99 Mbp, with an average G+C% of 46.4%. The genome contains a total of 4,005 predicted open reading frames (ORF), of which 74% had a significant BLAST hit (E value of 0.001). The RAST server predicted 41 tRNA genes in this genome. It was found that the genome of AP183 contains two genes predicted to encode resistance to the antibiotics fosfomycin and fluoroquinolone, but no genes were predicted to encode virulence factors within this genome.

Gene clusters encoding secondary metabolite biosynthesis in strain AP183

Analysis of the AP183 contig sequences using the antiSMASH2.0 secondary metabolite prediction program suggested that AP183 encodes 18 predicted secondary metabolite biosynthesis gene clusters containing a total of 566 genes (Figure 9). AP183 is predicted to encode five trans-acyltransferase (AT) polyketide synthases (PKS), three nonribosomal peptide synthetases (NRPS), two hybrid PKS-NRPS, one hybrid trans-AT PKS, one type I PKS, one type II PKS, one type III PKS, and two terpene and two bacteriocin biosynthesis gene clusters. The trans-AT PKSs have emerged recently as an important group of biosynthetic enzymes involved in the production of many structurally complex, bioactive compounds. In 1993, bacillaene 1 was identified as the first member of the products of trans-AT PKS from the genome of *Bacillus subtilis* 168 (Piet 2009; Matilla et al. 2012). Since then, a number of clinically used antibiotics, mupirocin, virginiamycin M and anti-cancer agent bryostatins A were identified to be the products of trans-AT PKS enzyme (Davison et al. 2014). Although we have not yet confirmed the biosynthetic route of bacillisin A, its structure suggests a polyketide origin. We hypothesized that a trans-AT PKS biosynthetic gene cluster could potentially be involved in the synthesis of bacillusin A.

The AP183 genome is also predicted to contain a cluster with ORFs with homology to genes in the bacilysin biosynthetic cluster. In addition, an NRPS biosynthetic gene cluster was predicted in the AP183 genome with no known homology to that of other *Bacillus* species but with homologs to the genome of *Cyanothece* spp. strain PCC 7424.

Antibiotic resistance of AP183

Antibiotic resistance analysis revealed that the strain AP183 is susceptible to most of the tested antibiotics to varying degrees except colistin. It is highly susceptible to ampicillin, chloramphenicol, cephalothin and erythromycin (> 25mm diameter inhibition zone). Antibiotics

kanamycin, rifampin and sulfadiazine also inhibited their growth effectively (15-20mm zones of inhibition) whereas penicillin, vancomycin, novobiocin, neomycin, spectinomycin, gentamicin, ciprofloxacin and nalidixic acid showed moderate inhibition (10-15mm zones of inhibition).

***In vivo* cutaneous wound challenge**

Preliminary *in vivo* studies demonstrated that direct administration of AP183 spores to wounds significantly inhibit *S. aureus* Xen29 growth. A 46.91% reduction in the number of viable *S. aureus* cells was observed in response to AP183 spores application when the wound tissue was homogenized and plated for CFU numbers (Figure 10). Similarly, spores of AB01, AP143, AP191 and AP218 also showed anti-MRSA activities with this *in vivo* model.

The novel chemical structure, *in vitro* and *in vivo* antibacterial activities of AP183 was the basis for further use of this strain in mouse model to understand the mechanisms of its antibacterial action. We have found that applications of AP183 spores alone are capable to reduce *S. aureus* growth. However, the effectiveness was enhanced by including both the spores and the metabolites expressed by AP183 (Figure 11 A and B). Our *in vivo* results with more than 10 replicates demonstrated a significant inhibition of *S. aureus*-derived bioluminescence (Figure 12 A and B) as a result of AP183 spore and metabolite application ($P < 0.05$). Furthermore, we observed a significant 69.5% reduction in the number of viable *S. aureus* cells in response to AP183 and metabolite application when the wound tissue was homogenized and plated for CFUs (Figure 13).

We observed little or no bioluminescence associated with injection sites that contained both viable AP183 spores and secondary metabolites (Figure 12 A). These results were consistently observed with >10 mice and no incidence of inflammation, rash, or skin discoloration

at the site of injection were observed when animals were injected with AP183 spores and/or its associated metabolites.

Microbial diversity analysis of the mouse skin

In addition to identifying a beneficial bacterium capable of inhibiting MRSA growth, it was desired to find a probiotic strain that does not indiscriminately inhibit a broad spectrum of skin-associated microbiota. Therefore a culture-independent approach was adapted to determine the percent relative abundance of bacterial genera present in each treatment group. Our results indicated that the microbial community associated with an active *S. aureus* infection was dominated by *S. aureus*, and that the administration of AP183 and its metabolites resulted in a significant reduction in *S. aureus* percent relative abundance (Figure 14). Furthermore, inoculation of AP183 and its metabolites increased the relative abundance of other skin-associated bacterial taxa. These other residents of the mouse skin microbiome were not detected in the absence of AP183 inoculation, suggesting the growth promotion of other members of skin microbiome,

Metabolomic analysis of wound

The original LC-ESI-MS of both positive and negative detection chromatograms indicated the presence of surfactins from the mouse tissues after administration of AP183 and its metabolites. The compound was detected with retention time of ~26 min and molecular weight of 1035 (Figure 15). However, we were unable to detect other metabolites such as iturins, fengycin and bacillusin A from mice wounds.

5. Discussion

The results of this study led to the identification of *B.amyloliquefaciens* subsp. *plantarum* strain AP183 as a promising probiotic candidate for preventing skin infection caused by MRSA. Although *S. aureus* can cause life-threatening systemic infection, skin and soft tissues are the most common sites of *S. aureus* infection. It comprises more than 75% of MRSA-related diseases (Cohen et al. 2007). Inhibition of the colonization of MRSA on skin could prevent the bacteria entering the bloodstream and ultimately reduce the occurrence of sepsis.

Through screening of 277 *Bacillus* strains for *in vitro S. aureus* inhibition, and conducting preliminary studies using a mouse wound model, it was observed that *B. amyloliquefaciens* strain AP183 was the most effective inhibitor of pathogenic *S. aureus* growth. The antibacterial activity of AP183 is partly due to the production of secondary metabolites. In this study AP183 was found to produce a novel antibiotic, bacillusin A that exhibits strong *in vitro* activities against MRSA and vancomycin resistant *E. faecium* (Ravu et al. 2015). Bacillusin A is a newly discovered macrocyclic polyene antibiotic composed of dimeric 4-hydroxy-2-methyl-6-alkenylbenzoic acid lactons with conjugated pentaene-hexahydroxy polyketide chains.

A review of the literature suggests that bacillusin A is structurally analogous to marinomycins A-D that was isolated from a marine Actinomycete, *Marinispora* (Kwon et al. 2006) and SIA7248 isolated from the marine isolate *Stretomyces* sp. A7248 (Zou et al. 2013). Compared to the marinomycins and SIA7248, bacillusin A possesses a larger macrocyclic ring with four additional C2 extender units, making it more difficult to determine the absolute or even relative configuration for the stereogenic carbons. Several attempts to crystallize bacillusin A in different organic solvents for X-ray crystallography failed, partly due to the limited amount of purified compound obtained from the isolation process. The limited amount (0.247mg/L) of compound prevented chemical derivatization that may provide further stereochemical

information. In addition, bacillusin A was found to be chemically unstable in organic solvents as indicated by LC-MS analysis. Additional compounds with the same molecular weight were detected by LC-MS when bacillusin A was allowed to stand at room temperature or even at 4^{o}C for few days. This is similar to marinomycin A that can be photoisomerized to marinomycins B and C with reduced antibacterial activities through the conversion of the *trans* double bond attached to the aromatic ring to a *cis* double bond when exposed to light (Kwon et al. 2006). This isomerization and possible decomposition during the isolation process resulted in a low isolation yield of the compound (0.247 mg/L). It was also observed that the antibacterial activity of bacillusin A decreased when it was stored in DMSO at 4°C for more than one week. The biochemical instability of this compound might be medically beneficial in the use of these antibiotics to protect against bacterial infections, e.g., infected wounds where rapid elimination of antibiotic residues can be an advantage.

The primary objective of this research project was to identify beneficial bacterial strain that can be applied as a probiotic to inhibit skin infections caused by MRSA and restored beneficial skin microbiota. A probiotic can be used either internally or externally to restore the balance of beneficial microorganisms to pathogen. The key considerations for use of a beneficial bacterium must be safety and efficacy. For these reasons we have focused on identifying endospore-forming members of the genus *Bacillus* that have efficacy in preventing *S. aureus* infections, without any known potential for pathogenicity. Specifically, this study focused on identifying gram-positive *Bacillus* strains that lack endotoxin or other known virulence factors and were highly effective at inhibiting *S. aureus* growth under *in vitro* conditions. This led to the discovery of a specific strain, AP183, which is the leading candidate for a skin probiotic for MRSA inhibition. The annotation of the genome sequence of the AP183 strain revealed the

absence of predicted virulence factor encoding genes in this genome which supports the safety of this strain for topical application. Further studies of repeated applications of strain AP183 in mouse and minipig models of MRSA inhibition should be conducted for safety assessment, along with studies to assess any cytotoxicity of AP183 secondary metabolites (e.g., bacillusin A).

Strains of *B. amyloliquefaciens*, as members of the *B. subtilis* group, have been used previously as probiotics in animals and plants and are not associated with disease (Ahmed et al. 2014, Gonzalez-Ortiz et al. 2013). There are many strains of *B. amyloliquefaciens* already developed for use with crops, livestock or products for human consumption (Krober et al. 2014, Lee et al. 2012), and *Bacillus* spores are currently been using topically to prevent infections caused by yeast, fungus, bacteria or Herpes simplex virus (Farmer and Mikhail 1998). Farmer and Mikhail (Farmer and Mikhail 1998) described the potential use of *Bacillus* spores for the prevention and control of bacterial infections, wherein spores of *B. coagulans* inhibited growth of *S. aureus* in skin. The key differences with the previous study is the use of a different species of *Bacillus* (*amyloliquefaciens* strain AP183) and the observation that the use of *Bacillus* spores alone is not sufficient to inhibit the establishment of MRSA, whereby this strain also requires the presence of AP183 metabolites. The proposed mechanisms of action include both direct inhibition of *S. aureus* growth and viability due to the expression of multiple secondary metabolites, as well as competitive inhibition of *S. aureus* growth. Since *S. aureus* and other pathogens are fast-growing (i.e., r-selected) bacteria that can outcompete many potential probiotic strains, this study specifically selected endospore-forming, fast-growing *Bacillus* spp. that are well-adapted at competing for limited C substrates within the skin microbiome, but without the potential for virulence. Similar to the gut microbiome, the skin microbiome can play an important role in preventing skin infections primarily by inhibiting colonization and biofilm

formation. Iwase et al. (Iwase et al. 2010) demonstrated the role of serine protease secreted by *Staphylococcus epidermidis*, a skin commensal bacterium, inhibits biofilm formation and colonization of *S. aureus*. Therefore it is important to restore the balance of normal skin residents that can confer better protection against pathogenic microbes.

The mice microbiome studies indicated that the inoculation with AP183 and its metabolites significantly reduced the relative abundance of *S. aureus* and also increased in the relative abundance of other skin-associated bacterial taxa. These other residents of the mouse skin microbiome were not even detected in the absence of AP183 inoculation, and could conceivably play an important role in maintaining skin microbiome conditions that do not facilitate *S. aureus* infectivity. Although AP183 inoculation completely hinders symptoms of *S. aureus* Xen 29 infectivity at the wound site, it was still possible to detect *S. aureus* cells using both culture-based and culture-independent methods. Perhaps the *S. aureus* cells that remain in the wound site after exposure to AP183 and its metabolites are in a viable but quiescent state, such that they remain capable of growth when plated onto a rich culture medium, but may not be expressing virulence factors or be in a metabolically inactive state within the mouse.

The culture-dependent and metabolomic analysis of mouse wound revealed that strain AP183 can grow and produce potent secondary metabolites that inhibit MRSA infectivity in mice. However, by using LC-MS analytic methods it was only possible to detect the major metabolite surfactin in wound samples. It is reasonable to conclude that other metabolites such as iturins and bacillusin A might have been produced, but were not detected due to the short-half life of the compound or the inherent detection limits. Thus, future research to develop an efficient and sensitive analytic system for skin metabolome analysis will be of special interest.

Taken together, our findings indicated that AP183 spores and metabolites hinder *S. aureus* colonization *in vivo* through a novel mechanism of bacterial interference, which could lead to the development of innovative therapeutics to prevent *S. aureus* colonization and infection in skin.

Table 1. Antimicrobial activity of five different Bacillus spp.strains against fungal and bacterial pathogens (IC50 µg/mL).

Taxonomic affiliation	Strain	Source of isolation	*Candida albicans*	*Cryptococcus Neoformans*	*S.aureus*	MRSA	*E. coli*	*Pseudomonas aeruginosa*
			IC_{50} µg/mL					
B. subtilis group	AB01	Intestine of catfish	NA	25.24	NA	NA	10.45	34.83
B. amyloliquefaciens	AP143	Plant rhizosphere	NA	NA	10.5	13.2	NA	NA
B. amyloliquefaciens	AP183	Plant rhizosphere	NA	NA	9.83	12.6	49.49	NA
B. methylotrophicus	AP191	Plant rhizosphere	NA	14.1	11.9	12.27	21.38	NA
B. subtilis group	AP218	Plant rhizosphere	NA	5.11	9.51	7.76	NA	NA

NA indicates no activity.

Table 2. In vitro Antibacterial Activities of Bacillusin A (IC50/MIC/MBC, μg/mL)[a]

	S. aureus ATCC 29213	MRSA ATCC 33591[b]	MRSA EAMC30[c]	*E. faecalis* ATCC 51299	*E. faecalis* ATCC 29212	*E. faecium* ATCC 700221[d]
Bacillusin A	0.04/1.2/2.5	0.04/1.2/1.2	<0.02/0.6/0.6	0.2/0.6/1.2	0.2/0.6/2.5	0.1/0.6/-[e]
Vancomycin	0.7/1.6/50	0.8/1.6/1.6	0.2/0.4/0.4	3.4/6.2/-[f]	1.0/1.6/50	-/-/-[f]
Ciprofloxacin	<0.1/0.4/0.4	<0.1/0.4/0.8	29.4/100/-[f]	0.2/0.4/6.2	0.2/0.8/6.2	-/-/-[f]
Methicillin	0.4/3.1/12.5	-/-/-[f]	10.1/50/-[f]	14.2/50/50	15.3/25/50	-/-/-[f]

[a] IC_{50}: 50% growth inhibition. MIC: minimum inhibitory concentration. MBC: minimum bactericidal concentration. The highest test concentration for Bacillusin A was 20 μg/mL; the highest test concentrations for vancomycin, ciprofloxacin and methicillin were 100 μg/mL. [b] Methicillin-resistant *Staphylococcus aureus* strain. [c] Methicillin-resistant *Staphylococcus aureus* clinical isolate. [d] Vancomycin-resistant *Enterococcus faecium* strain. [e] Not active at 20 μg/mL. [f] Not active at 100 μg/mL (Ravu et al. 2015).

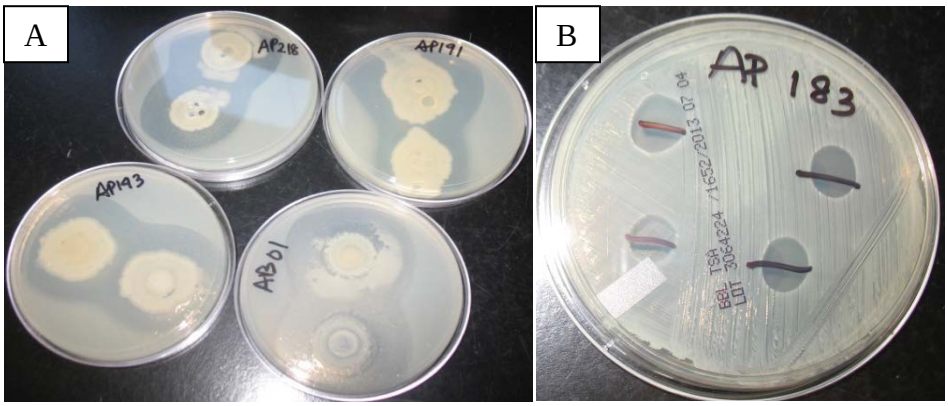


Figure 1. *In vitro* antibacterial activity of *Bacillus* spp. strains AP143, AP191, AP183, AP218 and AB01 against clinical MRSA strain 30 using soft agar overlay (panel A) and drop assays (panel B). Note that anti-MRSA activity of *B. amyloliquefaciens* strain AP183 only found in cell-free supernatant demonstrated by drop assay (panel B).

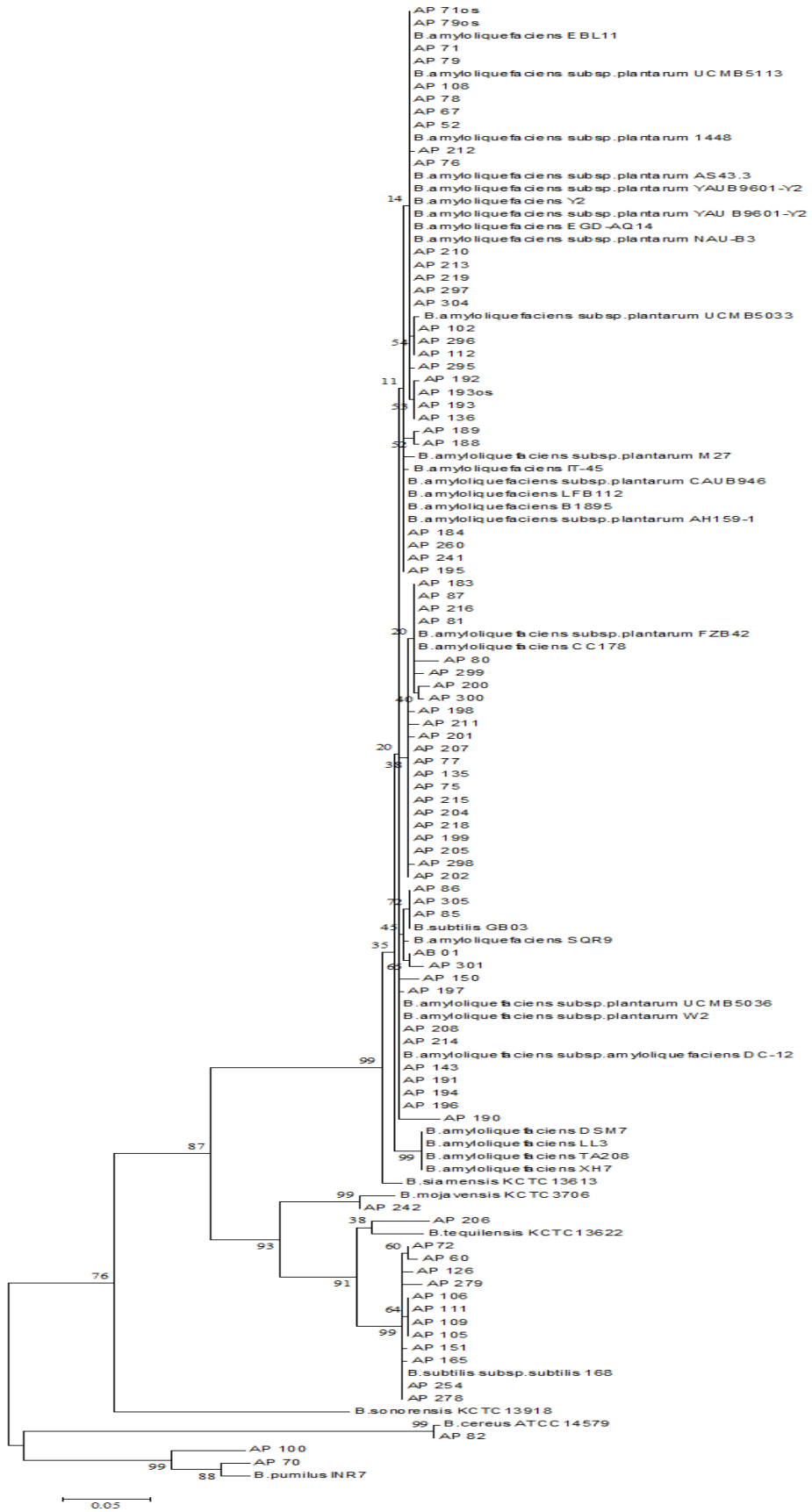

AP 71os
AP 79os
B.amyloliquefaciens EBL11
AP 71
AP 79
B.amyloliquefaciens subsp.plantarum UCMB5113
AP 108
AP 78
AP 67
AP 52
B.amyloliquefaciens subsp.plantarum 1448
AP 212
AP 76
B.amyloliquefaciens subsp.plantarum AS43.3
B.amyloliquefaciens subsp.plantarum YAUB9601-Y2
B.amyloliquefaciens Y2
B.amyloliquefaciens subsp.plantarum YAU B9601-Y2
B.amyloliquefaciens EGD-AQ14
B.amyloliquefaciens subsp.plantarum NAU-B3
AP 210
AP 213
AP 219
AP 297
AP 304
B.amyloliquefaciens subsp.plantarum UCMB5033
AP 102
AP 296
AP 112
AP 295
AP 192
AP 193os
AP 193
AP 136
AP 189
AP 188
B.amyloliquefaciens subsp.plantarum M27
B.amyloliquefaciens IT-45
B.amyloliquefaciens subsp.plantarum CAUB946
B.amyloliquefaciens LFB112
B.amyloliquefaciens B1895
B.amyloliquefaciens subsp.plantarum AH159-1
AP 184
AP 260
AP 241
AP 195
AP 183
AP 87
AP 216
AP 81
B.amyloliquefaciens subsp.plantarum FZB42
B.amyloliquefaciens CC178
AP 80
AP 299
AP 200
AP 300
AP 198
AP 211
AP 201
AP 207
AP 77
AP 135
AP 75
AP 215
AP 204
AP 218
AP 199
AP 205
AP 298
AP 202
AP 86
AP 305
AP 85
B.subtilis GB03
B.amyloliquefaciens SQR9
AB 01
AP 301
AP 150
AP 197
B.amyloliquefaciens subsp.plantarum UCMB5036
B.amyloliquefaciens subsp.plantarum W2
AP 208
AP 214
B.amyloliquefaciens subsp.amyloliquefaciens DC-12
AP 143
AP 191
AP 194
AP 196
AP 190
B.amyloliquefaciens DSM7
B.amyloliquefaciens LL3
B.amyloliquefaciens TA208
B.amyloliquefaciens XH7
B.siamensis KCTC 13613
B.mojavensis KCTC 3706
AP 242
AP 206
B.tequilensis KCTC13622
AP72
AP 60
AP 126
AP 279
AP 106
AP 111
AP 109
AP 105
AP 151
AP 165
B.subtilis subsp.subtilis 168
AP 254
AP 278
B.sonorensis KCTC 13918
B.cereus ATCC 14579
AP 82
AP 100
AP 70
B.pumilus INR7
0.05

Figure 2. Phylogenetic tree reconstructed based on *gyrB* gene sequences showing the evolutionary relationship of *B. amyloliquefaciens* subsp. *plantarum* strain AP183. The tree was inferred using the maximum likelihood method. Numbers at each branch nodes indicate bootstrap percentages based on the maximum likelihood method (1000 replicates) that was calculated using MEGA 6.0 software (Tamura et al., 2011). Bar indicates 5 substitutions per 100 positions.

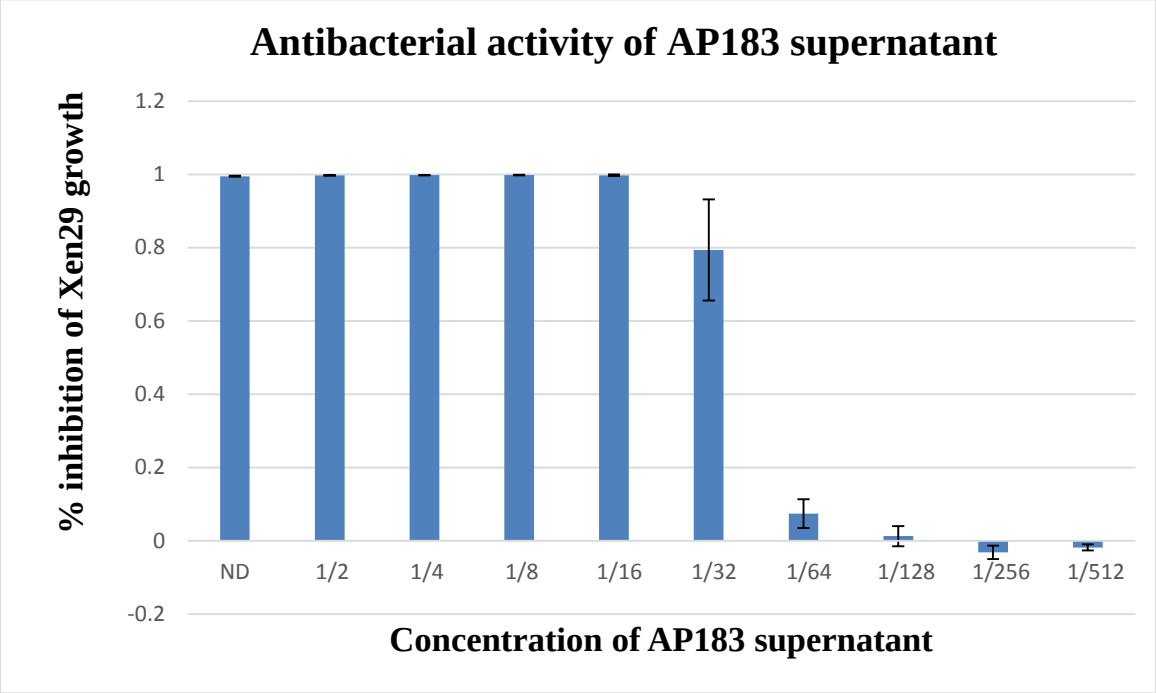


Figure 3. Antibacterial activity of *B. amyloliquefaciens* strain AP183 cell-free supernatant against bioluminescence *S. aureus* strain Xen29. AP183 metabolites showed potent antibacterial activity with IC_{50} of <1/32 of double dilution.

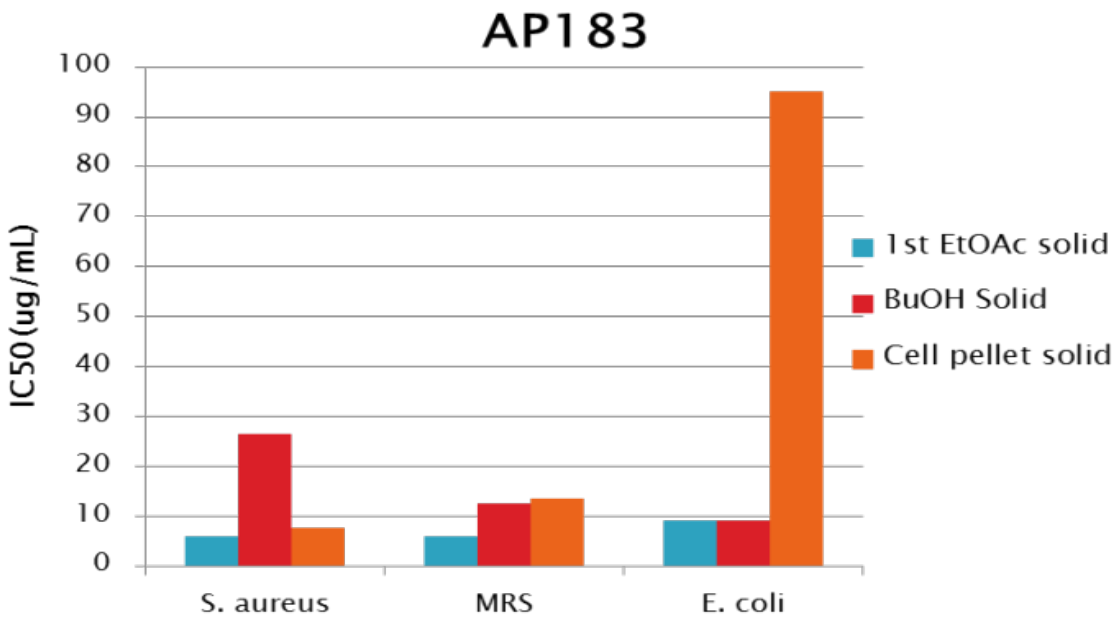


Figure 4. Organic solvent extraction methods for efficient recovery of anti-MRSA compound from AP183 supernatant and cell lysates. The ethyl acetate extract of the supernatant showed potent antibacterial activities against *S. aureus* ATCC 29213 and MRSA ATCC 33591. The methanol extract of the AP183 cell pellet also showed potent activities with IC_{50}s of 7.7 and 13.5 µg/mL, respectively.

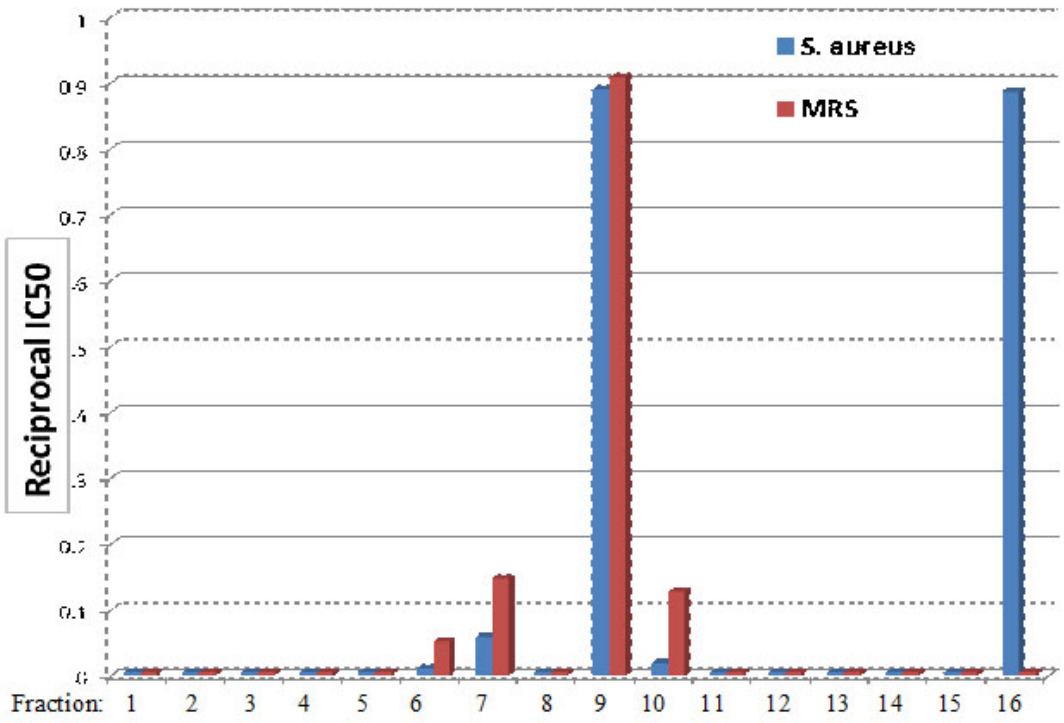
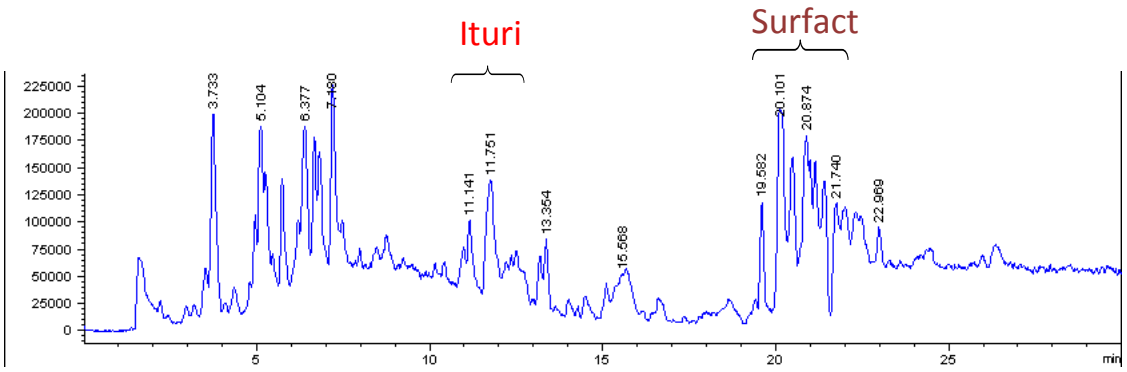


Figure 5. Antibacterial activity of reverse phase C-18 column fractions from AP183 ethyl acetate extract. The reversed-phase silica gel chromatography generated 16 fractions, with the 9th fraction being the most active with IC_{50} values of <1.1 µg/mL against *S. aureus* ATCC 29213 and MRSA ATCC 33591.

Panel A (+) ESI-MS chromatogram of EtOAc extract from



Panel B (+) ESI-MS chromatogram of MeOH extract

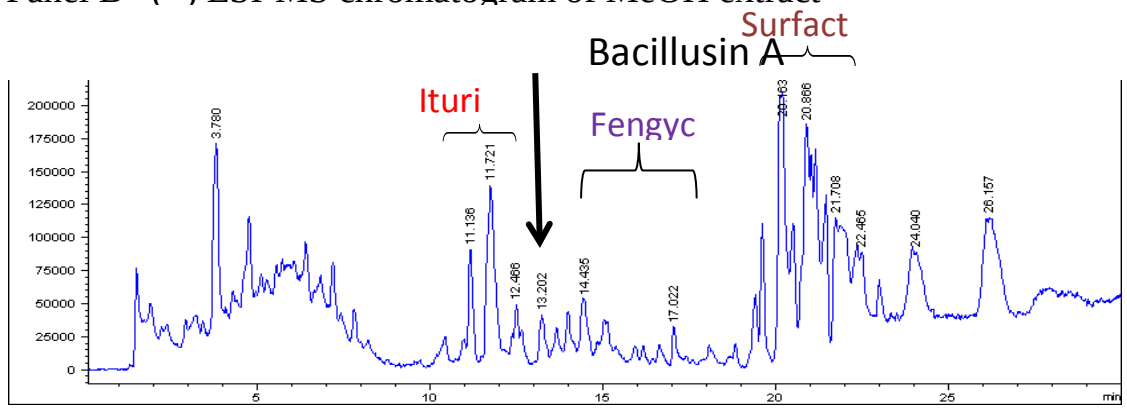


Figure 6. LC-MS Analysis of AP183 Extracts from culture supernatant and cell pellet. Surfactins, Iturins and Fengycins were major metabolites found in AP183 extracts. A novel compound, bacillusin A was detected by LC-MS analysis.

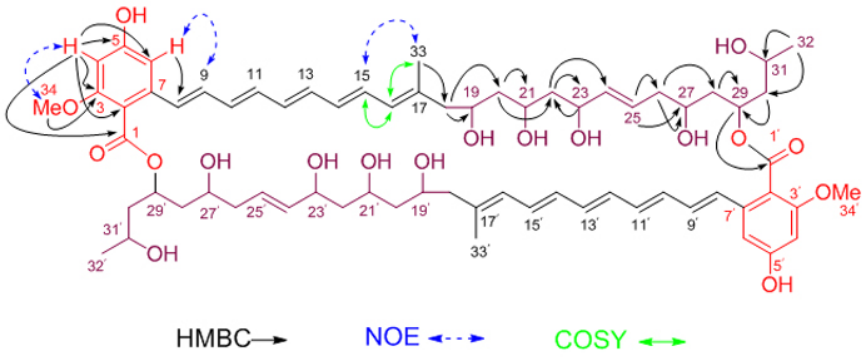


Figure 7. Structure of bacillusin A. 2D NMR correlations establishing the C-C connectivity of bacillusin A.

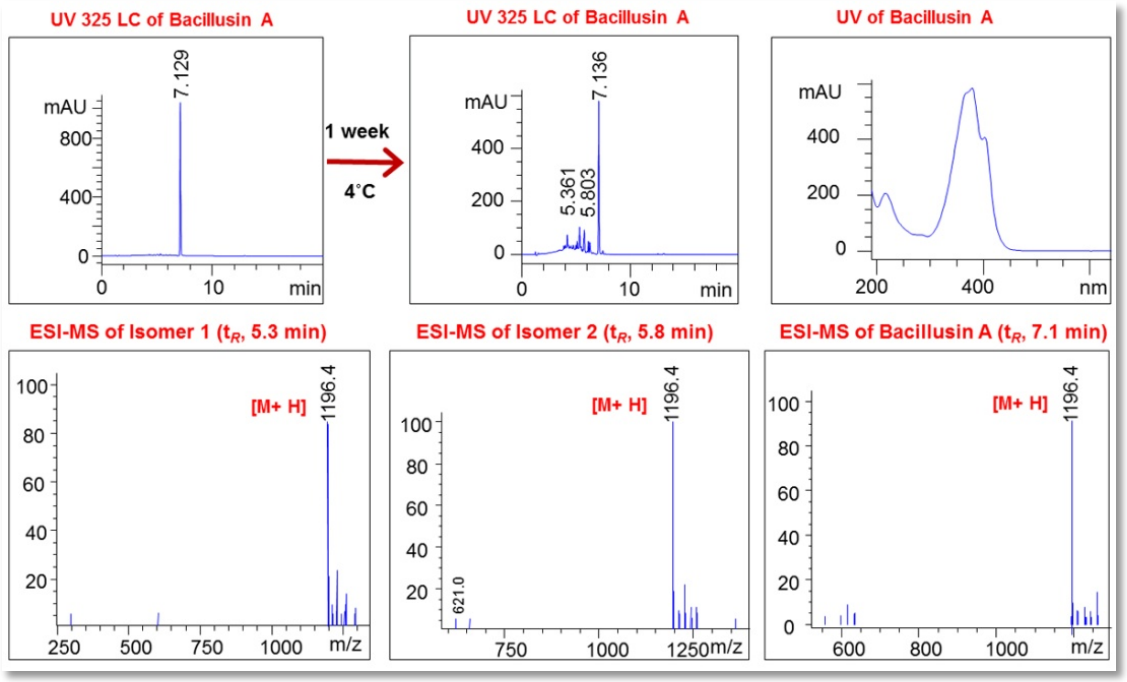


Figure 8. Stability of bacillusin A**.** LC-MS analysis showed photoisomerized peak for bacillusin A (t_R, 5.3 and 5.8 min) and degradation product present after 1 week at 4˚C.

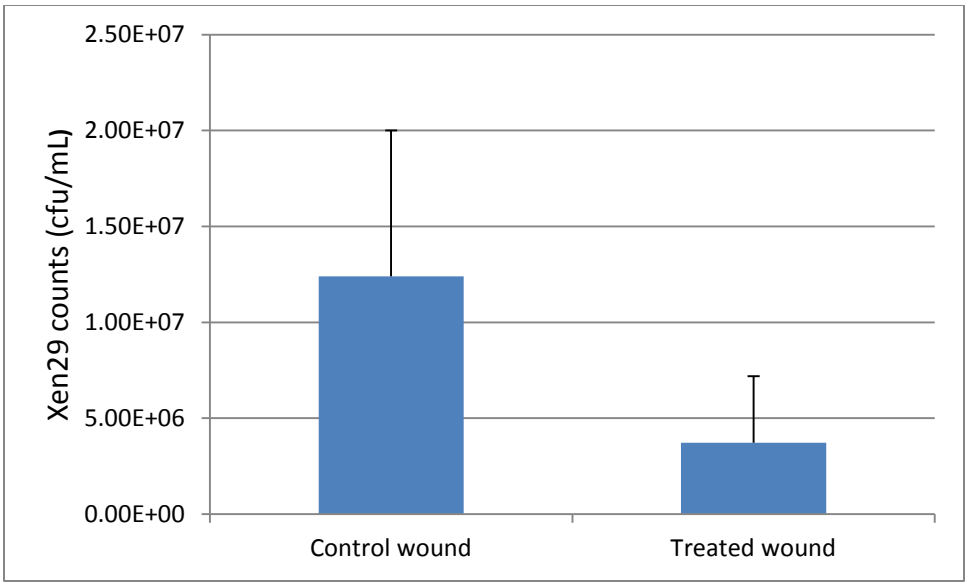


Figure 10. Viable Counts of Xen29 from tissue homogenates. Figure demonstrating reduction of Xen29 numbers as a result of AP183 spores application. Mice were simultaneously challenged with bioluminescent *S. aureus* strain Xen29 with (treated wound) and without AP183 spores (control wound).

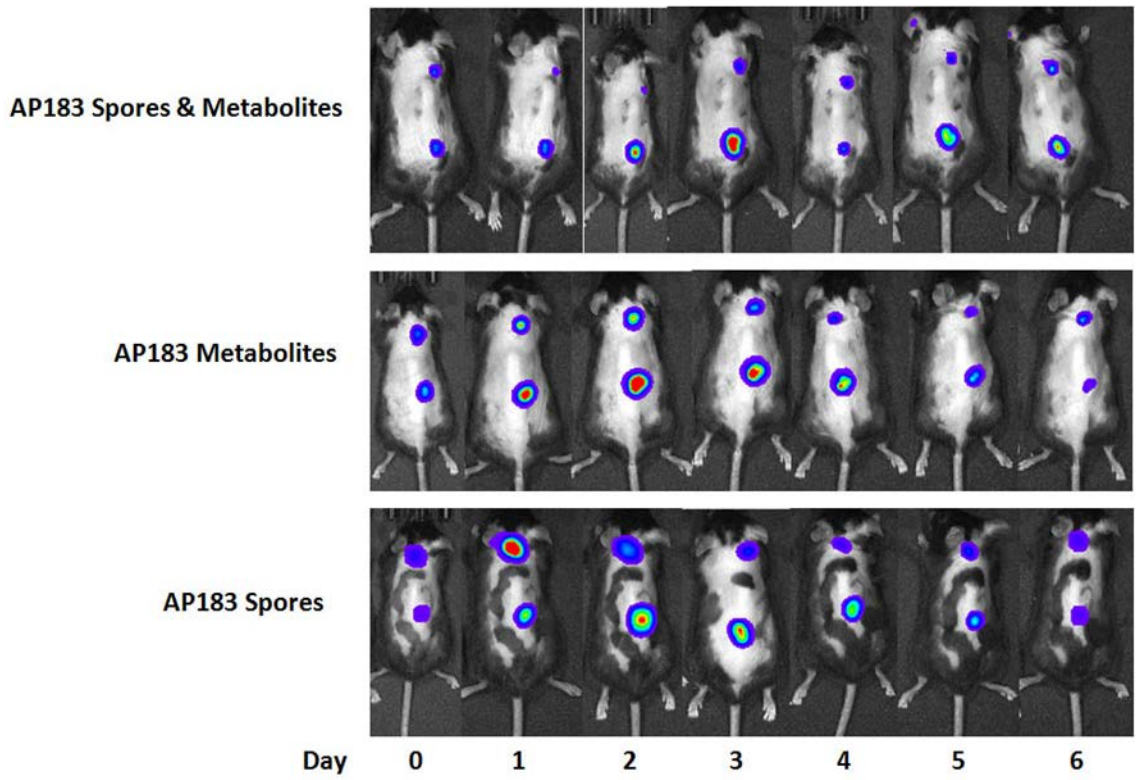


Figure 11 A. Temporal dynamics of skin *S. aureus* infections in a cutaneous wound model. Representative mice from the 3 groups are shown (i.e. AP183 Spores & metabolites, AP183 metabolites only and AP183 spores only). Subcuteneous injection of mice with ~1.0×107 S. aureus Xen29 was made in the upper (with ~1.0×108 CFU AP183 spores or metabolites and combination of AP183 spores and metabolites) and lower (without AP183 spores and metabolites) region of animal. Results demonstrated the reliability of the model and showed that Xen29 growth was inhibited by AP183 spores and metabolites.

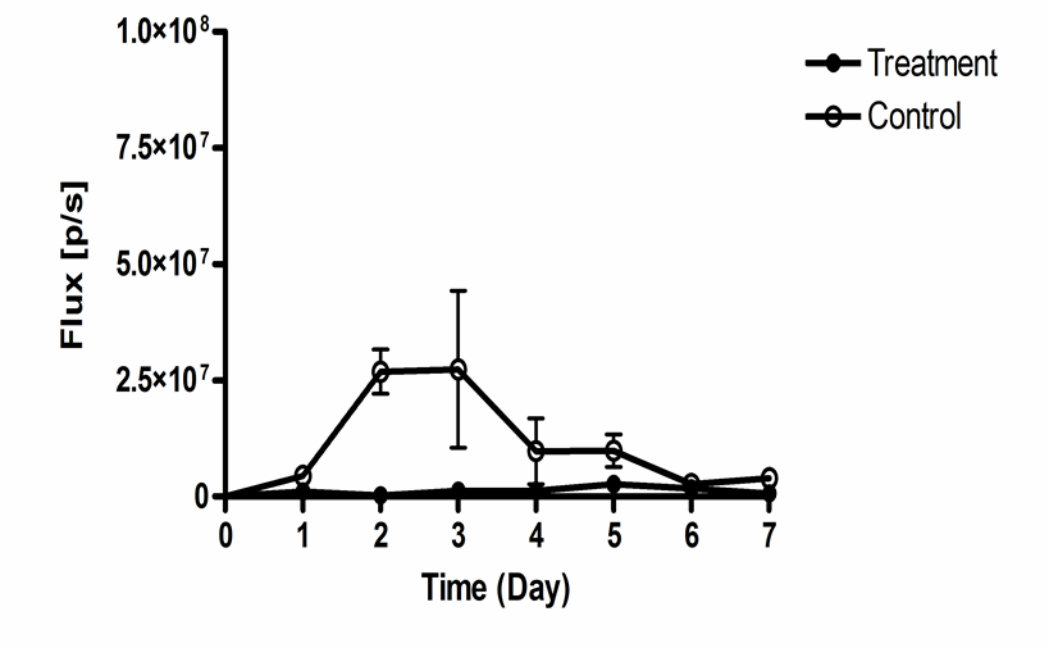


Figure 11 B. *S. aureus* strain Xen29 derived bioluminescence in treated (with AP183 spores and metabolites) vs untreated wounds.

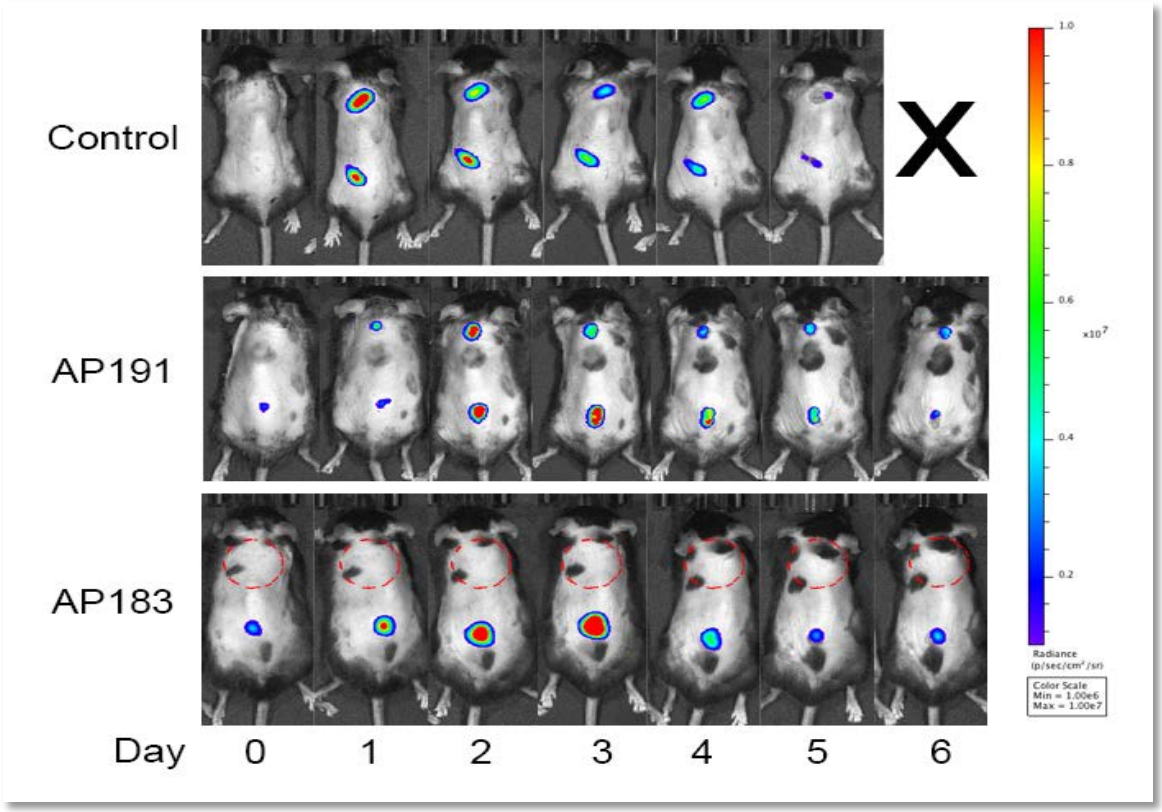


Figure 12 A. Temporal dynamics of skin MRSA infections in a cutaneous wound model. Representative mice from the 3 groups are shown (i.e. one control and two therapies AP191 and AP183). Subcuteneous injection of mice with ~1.0×10^7 *S. aureus* Xen29 was made in the upper (with ~1.0×10^8 CFU *Bacillus* spores and metabolites) and lower (without *Bacillus* spores and metabolites) region of animal. Results demonstrated the reliability of the model and showed that MRSA growth was completely inhibited by AP183 spores and metabolites.

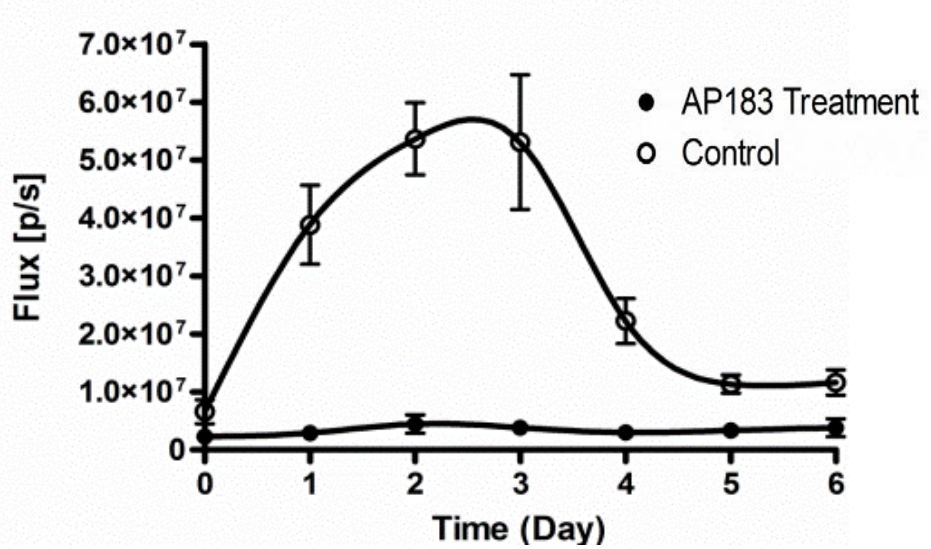


Figure 12 B. *S aureus* Xen29 derived bioluminescence in treated vs untreated wounds.

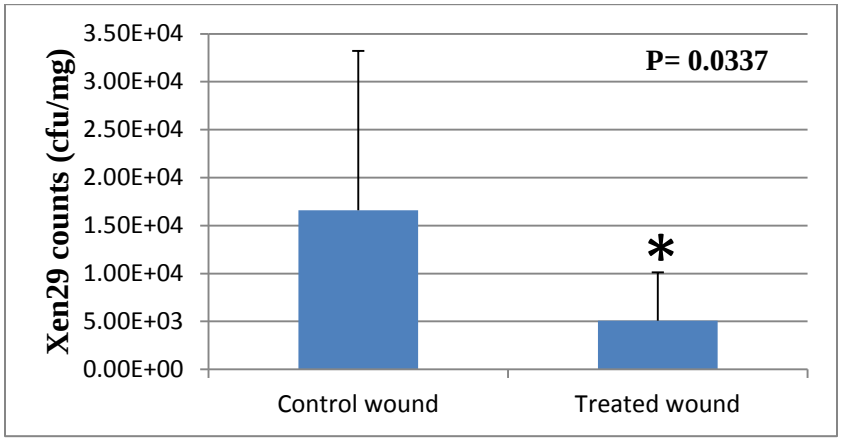


Figure 13. Viable Counts of Xen29 from tissue homogenates. Figure demonstrating a significant reduction ($P<0.05$) of Xen29 numbers as a result of AP183 and its metabolite application.

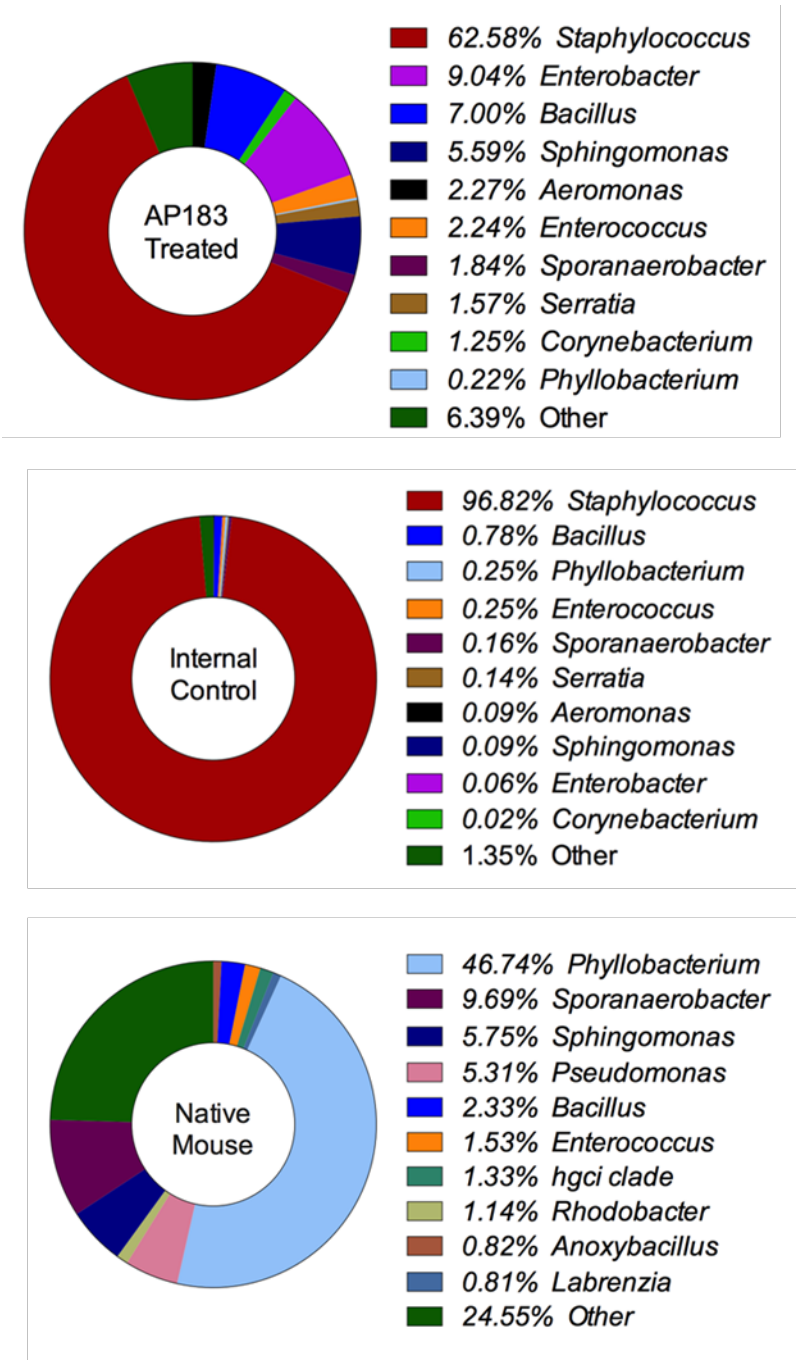


Figure 14. Microbial diversity analysis of the mouse skin microbiome of AP183 treated and untreated control wounds. Normal skin microbial diversity of native mouse was also analyzed.

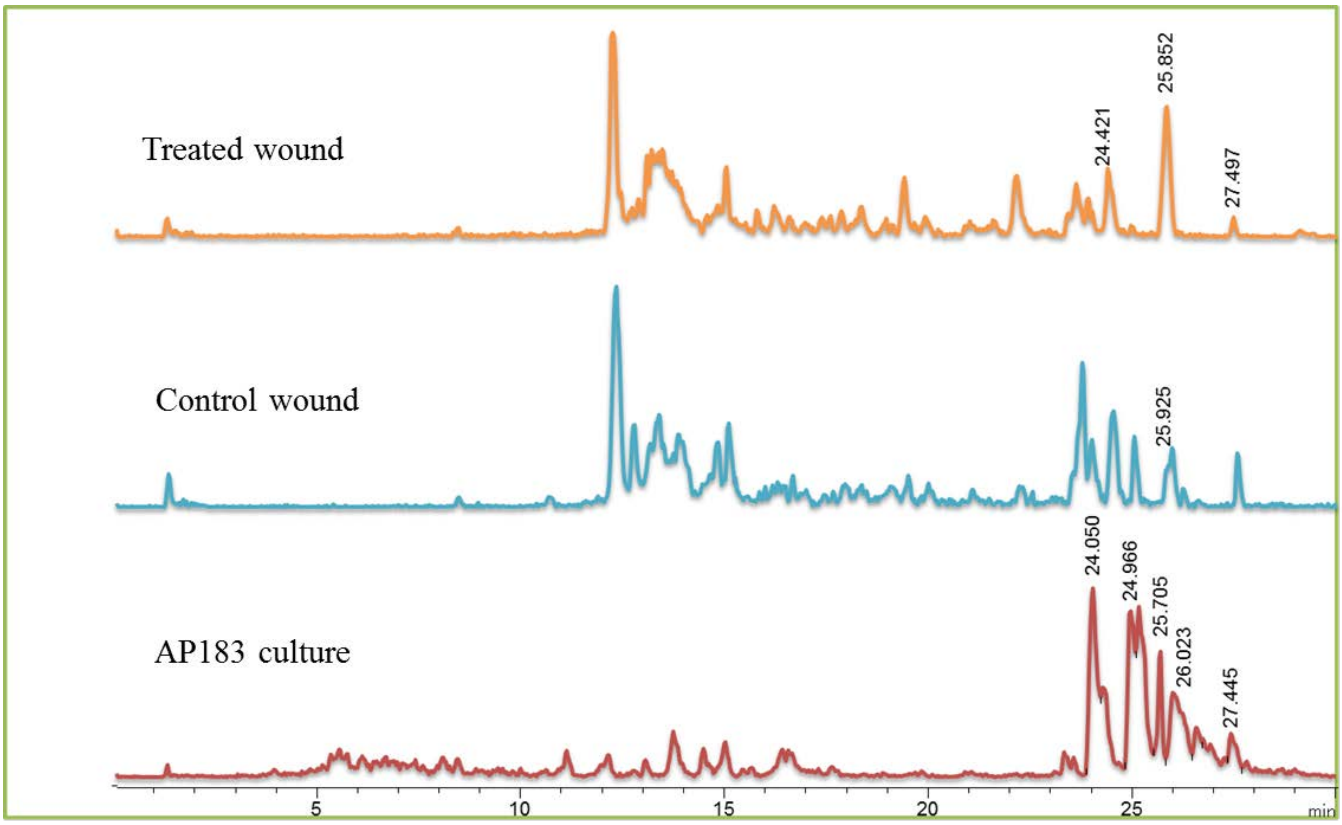


Figure 15. LC-MS analysis of mouse tissue homogenates. Only the major metabolite surfactin was detected at retention time of ~26 mins.

6. Reference

Ahmed, S. T., M. Islam, et al. (2014). "Effects of Bacillus amyloliquefaciens as a probiotic strain on growth performance, cecal microflora, and fecal noxious gas emissions of broiler chickens." Poult Sci **93**(8): 1963-1971.

Bello-Lopez, J. M., E. Fernandez-Rendon, et al. (2010). "In vivo transfer of plasmid pRAS1 between *Aeromonas salmonicida* and *Aeromonas hydrophila* in artificially infected Cyprinus carpio L." J Fish Dis **33**(3): 251-259.

Casula, G. and S. M. Cutting (2002). "Bacillus probiotics: spore germination in the gastrointestinal tract." Appl Environ Microbiol **68**(5): 2344-2352.

Cohen, A. L., C. Shuler, et al. (2007). "Methamphetamine use and methicillin-resistant Staphylococcus aureus skin infections." Emerg Infect Dis **13**(11): 1707-1713.

DeLeo, F. R. and H. F. Chambers (2009). "Reemergence of antibiotic-resistant Staphylococcus aureus in the genomics era." J Clin Invest **119**(9): 2464-2474.

DeSantis, T. Z., P. Hugenholtz, et al. (2006). "Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB." Appl Environ Microbiol **72**(7): 5069-5072.

Edelsberg, J., C. Taneja, et al. (2009). "Trends in US hospital admissions for skin and soft tissue infections." Emerg Infect Dis **15**(9): 1516-1518.

Evans, P. A., M. H. Huang, et al. (2012). "Total synthesis of marinomycin A using salicylate as a molecular switch to mediate dimerization." Nat Chem **4**(8): 680-684.

Gonzalez-Ortiz, G., L. Castillejos, et al. (2013). "Effects of dietary supplementation of Bacillus amyloliquefaciens CECT 5940 and Enterococcus faecium CECT 4515 in adult healthy dogs." Arch Anim Nutr **67**(5): 406-415.

Hamdache, A., A. Lamarti, et al. (2011). "Non-peptide metabolites from the genus Bacillus." J Nat Prod **74**(4): 893-899.

Hossain, M. J., G. C. Waldbieser, et al. (2013). "Implication of Lateral Genetic Transfer in the Emergence of *Aeromonas hydrophila* Isolates of Epidemic Outbreaks in Channel Catfish." PLoS ONE **8**(11): e80943.

Inoue, K., H. Tsutsui, et al. (2013). "Metabolic profiling of Alzheimer's disease brains." Sci Rep **3**(2364).

Iwase, T., Y. Uehara, et al. (2010). "Staphylococcus epidermidis Esp inhibits Staphylococcus aureus biofilm formation and nasal colonization." Nature **465**(7296): 346-349.

Kloepper, J. W. a. S. (1980). "Plant Growth-Promoting Rhizobacteria and Plant Growth Under Gnotobiotic Conditions." Ecology and Epidemiology: 642-644.

Krober, M., D. Wibberg, et al. (2014). "Effect of the strain Bacillus amyloliquefaciens FZB42 on the microbial community in the rhizosphere of lettuce under field conditions analyzed by whole metagenome sequencing." Front Microbiol **5**(252).

Kwon, H. C., C. A. Kauffman, et al. (2006). "Marinomycins A-D, antitumor-antibiotics of a new structure class from a marine actinomycete of the recently discovered genus "marinispora"." J Am Chem Soc **128**(5): 1622-1632.

Lee, H. A. and J. H. Kim (2012). "Isolation of Bacillus amyloliquefaciens Strains with Antifungal Activities from Meju." Prev Nutr Food Sci **17**(1): 64-70.

Li, B., Q. Li, et al. (2014). "Responses of beneficial Bacillus amyloliquefaciens SQR9 to different soilborne fungal pathogens through the alteration of antifungal compounds production." Front Microbiol **5**(636).

Maget-Dana, R. and F. Peypoux (1994). "Iturins, a special class of pore-forming lipopeptides:

biological and physicochemical properties." Toxicology **87**(1-3): 151-174.

Matilla, M. A., H. Stockmann, et al. (2012). "Bacterial biosynthetic gene clusters encoding the anti-cancer haterumalide class of molecules: biogenesis of the broad spectrum antifungal and anti-oomycete compound, oocydin A." J Biol Chem **287**(46): 39125-39138.

Milner, J. L., L. Silo-Suh, et al. (1996). "Production of kanosamine by Bacillus cereus UW85." Appl Environ Microbiol **62**(8): 3061-3065.

Monteiro, S. M. S., Clemente, J.J., Carrondo, M.J.T. and Cunha, A.E. (2014). "Enhanced Spore Production of Bacillus subtilis Grown in a Chemically Defined Medium." Advances in Microbiology **4**: 444-454.

Nasrin, S., M. J. Hossain, et al. (2015). "Draft Genome Sequence of Bacillus amyloliquefaciens AP183 with Antibacterial Activity against Methicillin-Resistant Staphylococcus aureus." Genome Announc **3**(2): 00162-00115.

Nicolaou, K. C., A. L. Nold, et al. (2006). "Total synthesis of marinomycins A-C." Angew Chem Int Ed Engl **45**(39): 6527-6532.

Ongena, M., E. Jourdan, et al. (2007). "Surfactin and fengycin lipopeptides of Bacillus subtilis as elicitors of induced systemic resistance in plants." Environ Microbiol **9**(4): 1084-1090.

Panizzi, P., M. Nahrendorf, et al. (2011). "In vivo detection of Staphylococcus aureus endocarditis by targeting pathogen-specific prothrombin activation." Nat Med **17**(9): 1142-1146.

Piel, J. (2010). "Biosynthesis of polyketides by trans-AT polyketide synthases." Nat Prod Rep **27**(7): 996-1047.

Ran, C., A. Carrias, et al. (2012). "Identification of *Bacillus* Strains for Biological Control of Catfish Pathogens." PLoS ONE **7**(9): e45793.

Ravu, R. R., M. R. Jacob, et al. (2015). "Bacillusin A, an Antibacterial Macrodiolide from Bacillus amyloliquefaciens AP183." J Nat Prod **10**: 10.

Silo-Suh, L. A., E. V. Stabb, et al. (1998). "Target range of zwittermicin A, an aminopolyol antibiotic from Bacillus cereus." Curr Microbiol **37**(1): 6-11.

Stein, T. (2005). "Bacillus subtilis antibiotics: structures, syntheses and specific functions." Mol Microbiol **56**(4): 845-857.

Stein, T., S. Borchert, et al. (2002). "Two different lantibiotic-like peptides originate from the ericin gene cluster of Bacillus subtilis A1/3." J Bacteriol **184**(6): 1703-1711.

Sumi, C. D., B. W. Yang, et al. (2014). "Antimicrobial peptides of the genus Bacillus: a new era for antibiotics." Canadian Journal of Microbiology **61**(2): 93-103.

Sun, Y. Z., H. L. Yang, et al. (2010). "Probiotic applications of two dominant gut Bacillus strains with antagonistic activity improved the growth performance and immune responses of grouper Epinephelus coioides." Fish Shellfish Immunol **29**(5): 803-809.

Volpon, L., F. Besson, et al. (1999). "NMR structure of active and inactive forms of the sterol-dependent antifungal antibiotic bacillomycin L." Eur J Biochem **264**(1): 200-210.

Volpon, L., F. Besson, et al. (2000). "NMR structure of antibiotics plipastatins A and B from Bacillus subtilis inhibitors of phospholipase A(2)." FEBS Lett **485**(1): 76-80.

Wang, D., Y. Liu, et al. (2008). "[Isolation and identification of surfactin producing Bacillus subtilis strain and its effect of surfactin on crude oil]." Wei Sheng Wu Xue Bao **48**(3): 304-311.

Wright, G. (2015). "Antibiotics: An irresistible newcomer." Nature **517**(7535): 442-444.

Zou, Y., H. Yin, et al. (2013). "A trans-acting ketoreductase in biosynthesis of a symmetric polyketide dimer SIA7248." Chembiochem **14**(6): 679-683.

Zou, Y., H. Yin, et al. (2013). "A trans-acting ketoreductase in biosynthesis of a symmetric polyketide dimer SIA7248." Chembiochem **14**(6): 679-683.

Chapter 4

Functional screening of a large-insert soil metagenomic library for the discovery of antibacterial compounds.

1. Abstract

The emergence of multidrug-resistant pathogens has increased the need for the discovery of novel antibiotics. Soil microbial communities are known to be a great resource for natural products but a majority of them have not been explored for their secondary metabolite synthesis because many of them are unculturable using currently available laboratory techniques. We cloned high molecular weight DNA from the microbial assemblage from the Cullars Rotation agricultural soil (a plot without N or P amendments) into a shuttle bacterial artificial chromosome (BAC) vector. The soil metagenomic library is comprised of 19,200 *E. coli* clones with an average insert size of 110 kb. The large insert size and broad host range of this library are designed to overcome some of the limitations commonly encountered in using functional metagenomics for natural product discovery. We screened all of the *E. coli* clones for inhibition of growth of methicillin-resistant *Staphyloccocus aureus* (MRSA) using a 96-well microtitre plate format. *In situ* lysis of the *E. coli* host enabled detection of both intra- and extracellular compounds, yielding a total of 28 anti-MRSA clones. Transformation of naïve *E. coli* with BAC DNA isolated from anti-MRSA clones confirmed the presence of their anti-MRSA activity. Sequencing and sub-cloning of these clones revealed genes predicted to be involved in various biosynthetic pathways as well as many genes with unknown functions. Interestingly, we observed that multiple clones (n=7) were capable of modifying chloramphenicol that was added to the *E. coli* culture medium, thereby resulting in modification of an existing antimicrobial scaffold. LC-MS analysis of the organic extract of the clones revealed three new

chloramphenicol derivatives. Chemically synthesized chloramphenicol derivatives tested separately did not show strong antibacterial activities against *E. coli* or *Pseudomonas aeruginosa*; however, they showed antibacterial activity against MRSA, *Mycobacterium intracellulare* and *M. tuberculosis* with MICs of 27.6, 12.5 and 50.0 μg/mL, respectively. These results demonstrate that large-insert soil metagenomic libraries can be screened using innovative functional screening methods to access previously undescribed genomic and biochemical diversity.

2. Introduction

Infectious diseases continue to be a significant global public health concern and have been a cause of concern with the emergence of hyper-virulent and multidrug resistant (MDR) bacterial pathogens (Jones et al. 2008). The rising prevalence of drug-resistant pathogens (Rossolini et al. 2014; Davies et al. 2010) threatens a return to the pre-antibiotic era in which infectious diseases are not going to be cured using any available treatment. The increasing antimicrobial resistance among *S. aureus* isolates and their ability to produce biofilms in tissues and medical devices limits treatment options (Archer et al. 2011). Vancomycin, a glycopeptide antibiotic, is the first choice of drug for the treatment of MRSA infections. However, its application is limited by the emergence of strains with reduced antimicrobial susceptibility (Howden et al. 2010) and the occurrence of vancomycin treatment failure and mortality in patients with methicillin-sensitive *S. aureus* (MSSA) bacteremia (Lodise et al. 2008; Soriano et al. 2008). The minimum inhibitory concentration (MIC) creep, the incremental vancomycin MIC, is a frequently observed phenomenon in MRSA infected patients with vancomycin treatment (Steinkraus et al. 2007). Until now, 12 vancomycin-resistant MRSA strains have been identified in the Unites States and comparative genomics of all 12 strains revealed that they are

belonged to the clonal cluster 5 (CC5) lineage (Kos et al. 2012). Wide spread use of vancomycin for the treatment of MRSA infection has led to the emergence of two types of glycopeptide-resistant (Tarai et al. 2013) *S. aureus* strains that include glycopeptide-intermediate-resistant *S. aureus* (GISA) and vancomycin-resistant *S. aureus* (VRSA). In addition to antibiotic resistance to vancomycin, side effects caused by vancomycin treatment are also another impediment to treat MRSA infections. Several new alternative antibiotics such as daptomycin, linezolid and tigecycline are in use to treat MRSA-infected patients (Micek 2007). However, their application efficacies are limited by the different varieties of infection caused by MRSA (Liu et al. 2011). The three newest drugs oritavancin, dalbavancin and tedizolid have recently been approved by FDA for treating MRSA infections (Network 2014). However, these new drugs can only be used for skin and skin structure infection and nosocomial pneumonia (NP). Since the applicability of these antibiotics are limited to only two types of MRSA infections, the need for discovering novel antimicrobial compounds is urgent to combat this MDR bacterial pathogen (Levy et al. 2004; Fischbach et al. 2009).

A vast number of antibiotics in use today to treat patients with infectious diseases are derived from soil bacteria or fungi (Newman et al. 2007). The metabolic and functional versatility of soil microorganisms makes the environment is a good source for the discovery of novel natural products, including antibiotics (Courtois et al. 2003). Classical methods for discovering antibiotics involve screening natural products (Reddy et al. 2014) or a chemically synthesized compound (Lawrence et al. 2011) against a target bacterial culture. More than 20 novel classes of antibiotics were developed in this way between 1930 and 1962 (Parsley et al. 2011; Yang et al. 2014) However, only two new classes of antibiotics have been marketed since then (King et al. 2009; Jenke-Kodama et al. 2009; Katoh et al. 2013). The past several decades

have experienced diminishing success rates for antibiotic discovery using a culture-based approach due in part to the >99% rate of antibiotic rediscovery (Piel, 2002)

New methods for drug discovery have been developed recently, including genome mining, novel culturing methods, and metagenomics (Wilkinson et al. 2007; Ling et al. 2015; Nikolouli et al. 2012)). Decoding the genomes of antibiotic producing microbes has revealed the presence of large numbers of new pathways (Bentley et al. 2002; Wilkinson et al. 2007; Nikolouli et al. 2012). Bioinformatic analysis of most antibiotic-producing microbial genomes indicates the presence of multiple pathways for secondary metabolite biosynthesis. However, these cryptic pathways are mostly silent and efforts to turn them on have succeeded, but not as a large scale platform (Lewis, 2013). New cultivation strategies using diffusion chambers to mimic the natural entrainment have been employed recently to cultivate as-yet uncultured bacteria (Ling et al. 2015). Using this technique, a new class of antibiotic, teixobactin, was discovered after screening 10,000 soil bacteria isolated using *in situ* cultivation methods (Ling et al. 2015). This approach, while successful in identifying novel cultured isolates that may synthesize novel compounds, is known to have a relatively high rate of failure for maintaining isolates in culture (about 20-30% can be maintained in laboratory culture) and these cultured isolates do not represent the full extant diversity of microbial life present in natural environments.

Another recent strategy for drug discovery is metagenomic analysis of uncultured microbes, allowing analysis of the collective genomes of all the resident organisms (Handelsman et al., 1998; Rondon et al., 2000). Metagenomic libraries are constructed by extracting environmental DNA (eDNA) directly from environmental sources such as soil (Sangwan et al. 2012), sediments (Havelsrud et al. 2011), activated sludge (Liaw et al. 2010), and hot thermal vent sediments (Wemheuer et al. 2013). Because of the ability of metagenomics

to incorporate the DNA from diverse sources, this offers a tool to discover and exploit a wide variety of natural products (e.g., enzymes, metabolites) from previously uncharacterized microorganisms that reside in natural environments.

Functional profiling of metagenomic libraries has recovered novel biocatalysts and bioactive secondary metabolites (O'Brien et al. 2014; Dougherty et al. 2012; Gillespie et al. 2002), many of which would be impossible to acquire by using culture-based methods. Metagenomics has led to the isolation of natural products with enzymatic (Dougherty et al. 2012), anticancer (Pettit 2004) and antimicrobial (Gillespie et al. 2002; Banik et. al. 2010) activities. One study isolated a small molecule, called indirubin, from a soil metagenomic library (MacNeil et al. 2001), and later it was found that indirubin and its derivatives inhibit tumor growth by antitumor angiogenesis (Zhang et al. 2011). Chow et al. (Chow et al. 2012) have isolated and characterized two novel lipolytic enzymes using a functional metagenomics approach. A functional metagenomics approach, while successful in some cases, depends upon successful transcription, translation and in some cases post-translational modifications for a natural product to have activity, and this may limit the potential for heterologous expression of cloned genes in a particular host such as *E. coli*. Given that many biosynthetic pathways, such as polyketide synthases (PKSs), require large genomic regions for biosynthesis, this also is a fundamental limitation for studies that clone or directly sequence smaller genomic fragments.

In this study a functional metagenomic approach was employed to identify recombinant clones that inhibited MRSA growth. A large-insert soil metagenomic library was constructed in an inducible-copy bacterial artificial chromosome (BAC) vector and these recombinant clones were heterologously expressed in *E. coli* strain DH10B and screened for clones that elaborate an antibacterial activity. There is a potential benefit of screening a larger-insert metagenomic library

for antimicrobial activity due to the higher probability that a recombinant clone will contain an intact biosynthetic pathway necessary for the synthesis of a new chemical entity (NCE). The use of BAC vectors for constructing metagenomic library has facilitated large-insert cloning of environmental DNA (Wang et al., 2013) and the expression of eDNA can be enhanced by using a copy-inducible plasmid. The copy-inducible plasmid contains an origin of replication (i.e., *oriV*) that is under the control of an arabinose-responsive promoter, therefore each recombinant clone is represented multiple times within each host cell (Wild et al., 2002; Kakirde et al., 2010). The innovations in this study include the use of high molecular weight metagenomic DNA from soil microbial communities for construction of large-insert BAC libraries and using an "*in situ* lysis" method to screen metagenomic libraries for antibacterial activities. The antimicrobial expressing metagenomic clones identified in this study may be promising candidates as therapeutants to control MDR pathogens.

3. Materials and methods

Metagenomic library construction

For constructing a large-insert containing metagenomic library, high molecular weight **(**HMW) environmental DNA was isolated from the Cullars Rotation soil (Auburn, AL), a agricultural plot that had not been amended with fertilizers for the past 100 years (Mitchell et al. 2012). The isolation and purification of soil HMW DNA was conducted by the protocol published previously (Liles et al. 2008) with some modifications. Briefly, random shearing approach was used to acquire the desired size range (>1 Mbp) of soil DNA fragments for cloning into a BAC vector. The sheared DNA was then blunt ended, ligated into the pSMART BAC-S vector (Figure 1) (Zhou et al., manuscript in preparation) and transformed into the host *E. coli* strain of choice. Clones were cryopreserved in 384 well plates at -80°C.

Metagenomic library screening for antibacterial activity

An *in situ* method was used for high-throughput screening of the metagenomic clones for activity against clinical MRSA strain. In this method, the BAC library containing *E. coli* cell growth and lysis, as well as MRSA growth and inhibition assay, were all performed in the same well of a 96-well plate. The library in a 384-well format was inoculated into 96-well plates by using a pin replicator. The *E. coli* clones were grown for 48 hours at 37°C in Luria-Bertani (LB) medium containing 12.5 µg/mL of chloramphenicol (Cm) and 0.01% arabinose to induce plasmid copy number, and then plates were frozen at -80°C followed by rapid thawing at 55°C, thereby lysing most *E. coli* cells. Each well was then inoculated with 100µL of 1:1000 diluted log-phase culture of a clinical MRSA strain 30 (East Alabama Medical Center, Opelika, AL). The medium also contained nalidixic acid (30µg/mL) to inhibit the growth of any remaining *E. coli* cells and the MRSA culture was grown for 24 hours at 37°C. Finally, 165 µL of the viability indicator solution of resazurin (0.02%) was added to each well and the plates were incubated at 37°C until color change was observed from blue to pink for the majority of wells (Martin et al. 2003). In the resazurin-based bioassays, fluorescence readings of reduced resazurin (resorufin) were recorded (530 nm excitation and 590 nm emission) using a fluorescent microtitre plate reader (BioTek, Winooski, VT) and used for calculating the % growth inhibition of the tester culture in comparison with the empty vector negative control.

Validation of recombinant clones with antibacterial activity

Each recombinant clone that inhibited the growth and/or viability of the MRSA strain 30 from the primary screening was re-tested to verify the antibacterial activity against MRSA. To validate anti-MRSA activity of putative positive clones, *E. coli* cultures from the original 384-well plate were grown for colony isolation on LB agar plates containing 12.5µg/mL Cm

without arabinose in order to maintain the vector at single copy. Six independent colonies from each positive clone were tested for anti-MRSA activity as described above. Clones that tested strongly positive in this secondary screen were re-tested twice. Those that tested positive in three consecutive assays were further characterized to identify the best lead candidates.

Preliminary characterization of active clones

A total of 28 clones were selected for further testing to characterize the active compounds produced by these clones. Supernatants from 28 anti-MRSA compound-producing clones were analyzed at the National Center for Natural Product Research (NCNPR, Oxford, MS) according to the standard protocol established by the NCNPR. Silica gel chromatography of an ethyl acetate extract of the supernatant of eight metagenomic clones resulted in the identification of eight different chloramphenicol derivatives (Cm derivatives). The structural elucidation of these compounds was achieved by NMR and MS analyses at the NCNPR.

DNA sequence generation and analysis

Each of the 28 antibiotic-expressing clones were selected for complete insert sequencing using either 454 pyrosequencing (454 Life Sciences, Branford, CT) or an Ion Torrent PGM (Life Technologies, Grand Island, NY). Large-scale BAC DNA was isolated from each respective clone using an alkaline lysis method (Sambrook and Russell, 2001) and extracted DNA was used to generate bar-coded shotgun subclone libraries for 454 pyrosequencing at the Lucigen Corp. (Middleton, WI). The 454 pyrosequencing and Ion Torrent PGM was conducted at the Lucigen Corp. (Middleton, WI) or at EnGenCore at the University of South Carolina (Columbia, SC) respectively, using a Genome Sequencer FLX system as per the manufacturer's instructions. The sequences were trimmed and assembled *de novo* into contiguous fragments (contigs) using the CLC genomics workbench (Cambridge, MA). The contig(s) that represented the complete (or

nearly complete) clone insert DNA was exported in FASTA format. The protein-coding Open Reading Frames (ORFs) within the inserts of the BAC DNAs were predicted using GeneMark (Lukashin and Borodovsky 1998) and the ORF sequences were compared against the GenBank nr/nt database using BLASTx and BLASTn search algorithms to predict the function of putative gene products.

Insert size determination for metagenomic clones

The insert size of clones was determined by complete insert sequencing and restriction digestion. A large-scale BAC DNA isolation was conducted for clones P6B5 and P335B14 according to methods described previously (Sambrook and Russell 2001). Purified BAC DNA was restriction digested with NotI (New England BioLabs Inc. MA) enzyme according to the manufacturer's instructions and resolved the RFLP pattern using pulsed field gel electrophoresis with a 1 to 15 second switch time at 6 V/cm on a CHEF gel (BioRad, Hercules, CA).

Subcloning of clones P6B5 and P35B14

Anti-MRSA clones P6B5 and P335B14 were characterized using sub-cloning to identify the ORF(s) responsible for anti-MRSA activity. The following strategy was used to sub-clone the clones P6B5 and P35B14 to identify the genetic regions responsible for anti-MRSA activities. A large-scale BAC DNA isolation was conducted for clones P6B5 and P335B14 according to methods described previously (Sambrook and Russell 2001). Approximately 20 µg of BAC DNA from each clone was sheared using a g-TUBE (Covaris, MA) to obtain fragmented DNA within the range of 4 to 8 kb. Fragmented DNA was separated using gel electrophoresis and DNA of the targeted size was excised for purification. The ends of the fragmented DNA obtained by shearing were repaired using a DNA Terminator kit (Lucigen, WI). End-repaired DNAs from both clones were ligated into the pSMART vector (Lucigen, WI) and then electroporated into

electrocompetent *E. coli* (*E. coli* 10G) cells harboring pGNS-BAC to provide Cm-resistance. Sub-clones were selected on LB agar supplemented with 200 µg/mL ampicillin and 12.5 µg/mL Cm. The transformants were picked robotically using a QPix2 (Molecular Devices, Sunnyvale, CA) in 96-well plates and were grown overnight with shaking at 200 rpm at 37^{o}C. All the cultures grown in shallow 96-well plates were transferred to deep-well plates and their anti-MRSA activity were determined according to the method described above. Sub-clones with anti-MRSA activity were grown for plasmid extraction and the ORF(s) contained within active sub-clones were identified by primer walking PCR (Table 1) followed by sequencing. The size of the insert within positive sub-clones was determined by PCR using vector-specific primers.

Cloning of *trfA* gene into pSMART BAC-S vector

The *trfA* gene from clone P6B5 was PCR amplified using custom designed primers AF-BstXI (5'-GCTTCGGATCCCAGTCACTGCGTCTT-3') and AR-BstXI (5'-ATGCATGCATGCCTGGTCGCCAGCAA-3') that included BstXI restriction sites at their 5′ end. PCR product of *trfA* gene and empty vector (pSMART BAC-S) were restriction digested with the BstXI (New England BioLabs Inc. MA) enzyme. Digested products were gel purified and concentrated with DNA clean and concentrator^{TM}-5.0 (Zymo Research, CA) according to the manufacturer's instructions. The purified *trfA* gene amplicon was ligated with the pSMART BAC-S vector using quick ligase (New England BioLabs Inc. MA) and the resulting ligation mixture was transformed into *E. coil* strains DH10B (genotype F- mcrA Δ(mrr-hsdRMS-mcrBC) endA1 recA1 Φ80dlacZΔM15 ΔlacX74 araD139 Δ(ara,leu)7697 galU galK rpsL nupG λ- tonA) and into "E. cloni 10G" (Lucigen Corp.., Middleton, WI) that is a derivative of DH10G that contains the arabinose-inducible *trfA* gene on the chromosome, by electroporation (1 mm gap cuvette, 1.8 kV, 600 Ohms, 10 µF). The clones were selected on LB agar plates containing 12.5

μg/mL of Cm and the presence of *trfA* was confirmed by PCR using vector and *trfA* gene specific primers (AF-BstxI and TRA-IntR; AR-BstxI and TRA-IntF). The phenotypic characteristics of transformants were evaluated with or without arabinose induction using different concentrations of Cm (12.5, 25, 30, 40, 50, 60, 70, 80, 90, 100, 120, 150 and 200 μg/ mL). The plasmid DNAs were extracted after normalizing the optical density of the overnight culture (OD_{600} 0.7) using an alkaline lysis method and the plasmid yields were compared with that of negative control by agarose gel electrophoresis.

4. Results

Library screening for identification of antibiotic-producing clones

Initial screening of the metagenomic library with approximately 19,200 clones in the identified a total of 136 clones with some degree of inhibition of MRSA viability (Figure 2). Multiple bioassays were conducted to isolate a set of clones that were consistently positive for anti-MRSA activity. After several rounds of validation experiments, 28 top candidates were selected for the next phase of biochemical and genetic characterization. It was revealed that growing clones in 96-well plates compared to growth in culture tubes gave comparable results. The BAC DNA from each clone was transform into a naïve *E. coli* strain by electroporation, and transformants from each clone were tested for anti-MRSA activity. All 28 re-transformed metagenomic clones showed significant activity in inhibiting MRSA growth, demonstrating that the cloned DNA in *E. coli* was necessary and sufficient to confer antimicrobial activity against MRSA.

DNA sequence analysis

The sequences of inserts from all 28 anti-MRSA expressing BAC clones were determined using bar-coded next generation sequencing. After trimming and assembly *de novo*, the BAC DNA insert sequences showed that the average inserts sizes are about 113 kb. Contigs that represented the entire clone insert with > 50× coverage were selected for analysis. The number of predicted ORFs in each of the clones was more than 100 ORFs for most clones. A large number of ORFs were predicted to encode proteins with unknown functions, which makes it difficult to predict the specific gene(s) required for expression of their anti-MRSA activities. The BLASTx analysis of the ORFs found within the metagenomic clones revealed a diverse array of protein-coding genes with homology to that of diverse bacterial taxa (Figure 3).

Determine the insert size of the metagenomic clone

The inserts size of the clones P6B5 and P35B14 were determined by BAC DNA extraction followed by restriction digestion and resolution pulsed-field gel electrophoresis (Figure 4). The RFLP pattern revealed that anti-MRSA clones P6B5 and P35B14 have approximate insert sizes of 2.4 kb and 45.0 kb respectively.

Characterization of Cm derivatives from metagenomic clones

LC-MS analysis revealed that seven positive metagenomic clones (P35B14, P22E10, P28I7, P37I10, P27M10, P6B5, and P5A4) that are active against MRSA had similar metabolic profiles (Figure 5). Cm (1) was identified in all seven clones with a retention time (t_R) of 6.37 min and similar isotopic patterns and UV spectra were observed for several compounds with longer t_R (e.g., compounds 2−9 in clones P6B5 and P35B14), indicating they were chlorine-containing Cm derivatives. Although Cm derivatives have been reported previously (Gross et al. 2002; Bizerra et al. 2011; El-Kersh et al. 1976), analogs producing some of these molecular ions were not present in current literature reports. Thus, clone P35B14 was selected for a scale-up

using fermentation and chemical isolation due to the presence of typical Cm derivatives as shown in the LC-MS chromatogram (Figure 5).

At the NCNPR, an ethyl acetate extract of the P35B14 clone fermentation broth was subjected to column chromatography on normal phase and reversed-phase silica gel to yield compounds 2−9, along with Cm (1). Compounds 2−6 were identified as 3-acetylchoramphenicol, 3-propanoyl chloramphenicol, 3-butanoylchloramphenicol, 1,3-diacetylchloramphenicol, and 1,3-dipropanoyl-chloramphenicol, respectively, by comparison of their NMR data and optical rotations with those reported in the literature. The structures of the new compounds 7−9 were established; compound 7 was isolated as a liquid and corresponded to the peak at t_R 11.39 min in the LC-MS chromatogram (Figure 5). Based on the NMR data analysis the structure of compound 7 was determined as 1-acetyl-3-propanoylchloramphenicol.

Compound 8 was also isolated as a colorless liquid and showed close similarities to compound **7** in terms of ^{1}H and ^{13}C NMR spectra. The ^{1}H and ^{13}C NMR data suggested that compound 8 had one acetyl group and one butanoyl group in the molecule and the structure of compound 8 was assigned as 1-acetyl-3-butanoylchloramphenicol.

Compound 9 had the least polarity among the eight compounds produced by this metagenomic clone. The ^{1}H NMR spectrum of compound 9 showed similarity to 1,3-dipropanoylchloramphenicol (6), and careful analysis of the ^{1}H NMR data concluded that a propanoyl and a butanoyl group were present in this molecule. 2D NMR correlations of compounds 7−9 were presented in Figure 6.

Using purified compounds as standards for LC-MS analysis, compounds 1−9 were confirmed to exhibit specific t_R in the seven metagenomic clones as shown in Figure 5. Monoacylated compounds 2−4 were major metabolites present in all seven clones, while

diacylated compounds 5−9 were observed in clones P35B14 and P6B5 that showed similar metabolic profiles. It was noted that compounds 6 and 8 were co-eluted at the same t_R 12.40 min. Compound 9 was a minor metabolite, present in clones P35B14, P6B5, and P5A4 was also co-eluted with an unidentified compound at t_R 13.19 min. The above analyses suggested that all seven of the metagenomic clones contain highly active acyltransferases that are able to acylate Cm.

***In vitro* antibacterial activities of Cm derivatives**

The seven Cm derivatives (2-7) isolated from metagenomic clones and the three synthetic analogs (10−12) were tested for antibacterial activity against, *E. coli*, and *Pseudomonas aeruginosa*. Results indicated that all Cm derivatives tested separately were inactive at a concentration of 200 µg/mL against these two pathogens, but chloramphenicol 3-acetate and tert-butyl dimethyl silyl chloramphenicol had activity against both methicillin sensitive and –resistant *S. aureus*. However, compounds 2, 4, and 10 showed strong antibacterial activity in an initial testing against *Mycobacterium intracellulare*. Compounds 1−7, 10−12, and 15−18 were tested for *in vitro* anti-*Mycobacterium* activity against *M. intracellulare* and the tuberculosis-causing strain *M. tuberculosis*. The results showed that *M. intracellulare* was more susceptible to all these compounds except compounds 5 and 7 with minimum inhibitory concentrations (MICs) ranging from 12.5 to 50.0 µg/mL. However, only compounds 10, 12, and 18 showed antibacterial activity against *M. tuberculosis* with MICs of 50.0 to 100.0 µg/mL (Table 2). It was also found that an aromatic-containing acyl substituent of compound 18 was the most active against *M. intracellulare* and *M. tuberculosis* with MICs of 12.5 and 50.0µg/mL respectively, in comparison with Cm (1) which was only active against *M. intracellulare* with an MIC of 25.0 µg/mL.

Subcloning of clones P6B5 and P35B14 for identifying gene(s) responsible for the antibacterial activity

The anti-MRSA compound(s)-producing clones P6B5 and P35B14 have approximate insert DNAs of 2.4 kb and 45.0 kb respectively, as estimated by restriction digestion followed by agarose gel analysis (Figure 4). Since the anti-MRSA activity of clones P6B5 and P335B14 were due to their cloned DNA inserts, therefore, the active compounds was expected to be encoded by an ORF(s) present within their inserts. To identify the ORF(s) responsible for the antibacterial activity, subcloning of BAC DNAs of clone P6B5 and P35B14 were performed in pSMART HC Amp vector (Lucigen Corp. WI). The subclones of P6B5 showed anti-MRSA activity but not the subclones of P35B14 and those active subclones were grown for plasmid extraction and sequencing to identify the ORF(s) required for anti-MRSA activity. The nucleotide sequence of this ORF was determined by primer walking and a single candidate gene, *trfA*, appeared to encode the antibacterial activity. This was apparently not the *trfA* gene present in the genome of the "*E. cloni*" host strain, as the nucleotide sequence of *trfA* gene showed 97% similarity to known *trfA* sequences of pGNS-BAC vector. In addition, the vector pSMART BAC-S did not have any *trfA* gene. These results collectively suggested that this gene may have been cloned from metagenomic DNA. The *trfA* gene encodes the TrfA replication protein and together with the *oriV* replication origin the TrfA and *oriV* are considered the mini-replicon for RK2 plasmids; therefore, when the origin is present on a plasmid (e.g., pSMART-BAC-S containing *oriV*) and with the *trfA* on the chromosome under control of an arabinose-inducible promoter, this allows the BAC copy number to be induced to approximately 50-100 copies per host chromosome (Westenberg et al. 2010). In this case, we hypothesized that the cloned *trfA* gene was not arabinose-inducible given that it was apparently cloned from an environmental source. To determine the effect of copy induction for Cm derivative generation, the *trfA* gene from clone P6B5 was PCR amplified and cloned into *E. coli* strains DH10B and "E. cloni 10G" (Figure 7).

The clone that contained *trfA* gene was confirmed by PCR (Figure 8) and screened for Cm resistance at a concentration ranging from 12.5 to 200 µg/mL. Results indicated that *E. coli* containing the empty vector were significantly resistant to the highest concentration (200µg/mL) of Cm as compared to the *trfA* containing clones (Figure 9) and the antibiotic resistance pattern was not dependent upon the addition of arabinose in the culture media. Arabinose induction of *trfA* was not expected in clone 9 (*trfA* containing clone in *E. coli* strain DH10B) since the *trfA* gene was cloned from metagenomic DNA.

5. Discussion

Function-based metagenomic approaches have enabled the discovery of active compounds without any previous knowledge of the DNA sequences encoding the biological functions, and have been successfully used to discover many classes of biocatalysts and bioactive secondary metabolites. Though a functional metagenomic approach has been successfully used for the isolation of antibiotics from soil metagenome (Gillespie et al. 2002) and for discovery of many other natural products, this approach also has unique challenges that can limit antibiotic discovery, such as cloning incomplete biosynthetic pathways, obtaining sufficient expression in a heterologous host, inefficient secretion of the antibiotic, and lack of detection of the antibiotic by a screening method. While many of these challenges are inherent in a functional metagenomic approach, in this study novel protocols were used to improve this approach for antimicrobial discovery. First, the metagenomic library used in this study was obtained using randomly sheared metagenomic DNA that resulted in large insert sizes (>110 kb) that increase the number of genes per clone and enhances the probability of cloning intact operons. Secondly, an *in situ* lysis method was used for high throughput library screening for clones expressing an

antibacterial activity that allowed for growth of the metagenomic cultures, lysis, and screening in the same 96-well plates, and detected both extra- and intracellular compounds.

One perceived advantage of using BAC vectors is the high stability of both the vector and the insert when maintained at a single or low copy and the ability to be induced to high copy number when required. The pSMART-BAC-S vector used in this study allows high-throughput conjugation-based transfer and stable maintenance of large-insert BAC clones into Gram-negative as well as Gram-positive hosts, with chromosomal integration or stable episomal maintenance for heterologous expression. As this was the first study in which this library was screened for heterologous expression, the original *E. coli* cloning host was first used for expression. Subsequent studies will investigate the expression of this library in other heterologous hosts. Present function-based analysis of metagenomic library revealed identifying soil-derived recombinant clones that showed growth inhibition of MRSA strain. The rate of identifying anti-MRSA clone from the library was 0.147% and some of the identified clones were capable of producing novel chemical entities, even when expressed in the heterogeneous host *E. coli* host that was not optimized for natural product expression. This study has not only characterized each of the compounds produced by anti-MRSA compound expressing clones, but there is clear evidence for a diversity of chemistry and cloned genome fragments among these 28 clones.

This study focuses on seven clones including P35B14, P22E10, P28I7, P37I0, P27M10, P6B5 and P5A4 that are involved in Cm modification. The supernatant from *E. coli* clone cultures contained multiple Cm derivatives, three of which are novel Cm derivatives. The mechanism by which these Cm derivatives are synthesized is believed to be by acylation of the Cm skeleton. This is supported by the genetic and biochemical data from these metagenomic

clones, indicating that a combination of metagenomic-encoded enzymes (i.e., acyltransferases) acted on the exogenously supplied Cm to produce these novel Cm derivatives. This "combinatorial metagenomics" method could be applied with other natural product scaffolds to generate novel derivatives of other chemical entities with enhanced or altered biological functions.

This study illustrates that each heterologous expressed metagenomic clone can contain a unique combination of genetic elements and biochemical products. Genetic characterization by subcloning of BAC DNA of clone P6B5 revealed the insert gene is *trfA*, with this single gene being necessary and sufficient for the antibacterial activity. The *trfA* gene product is known to be responsible for replication initiation and copy number control for RK2 plasmids (Perri et al. 1991); thus, having a copy of this gene in the metagenomic clone might be advantage for generating Cm derivatives. The TrfA ORF is highly conserved and encodes a putative single-stranded DNA binding protein (Ssb) that activates the origin of vegetative replication in diverse bacterial species (Wegrzyn et al. 2014). This single-stranded DNA binding property of TrfA might activate the expression of certain genes that are involved in modifying Cm for the production of variety of cm derivates. Out of seven clones that were involved in Cm modification, two of them (P6B5 and P35B14) contain *trfA* gene which was confirmed by sequencing. Sequence analysis from other five clones did not reveal the presence of *trfA* gene in their insets. This finding was not surprising since each BAC clone contains unique genetic elements; therefore, presence of similar gene in each clone is unlikely. Though all the seven clones were generating Cm derivatives but we found some variations in the production of Cm derivatives, these variations might be due to the presence of different gene(s) in each of this clone that regulate the generation of Cm derivatives. Further genetic characterization of these

clones need to be done to confirm the gene(s) that is responsible for generating biochemical compound.

A previous functional metagenomic study that identified Turbomycins A and B observed that these metabolites resulted from a combination of heterologous host chemistry (i.e., indole secreted by *E. coli*) with metagenomic clone-encoded chemistry (i.e., homogentisic acid pigment) resulting in novel natural products (Gillespie et al. 2002). Similar results were observed in this study that the metagenomic clone-derived chemistry acted on an exogenously supplied natural product scaffold, in this case Cm, resulting in novel Cm derivatives that had not been previously described. Given the toxicity associated with Cm that limits its clinical use, it would be of interest to evaluate the novel Cm derivatives discovered to see if they are similarly likely to induce any toxicity, or if they might have better potential for clinical application.

Table 1. List of oligonucleotides used in this study.

Primer ID	Sequences
SR4F	5'-GGAGAAGTACCGCAAGCTGTCG-3'
SR5R	5'-CGTCATAGTTCCTCGCGTGTCG-3'
SR4R	5'-TCTTGGTCGTCATAGTTCCTCG-3'
SL2F	5'-AAGATCGAGCGCGACAGCGT-3'
SR3F(R)	5'-AATCCGATCCGCACATGAGG-3'
SL1	5'-CAGTCCAGTTACGCTGGAGTC-3'
SR2	5'-GGTCAGGTATGATTTAAATGGTCAGT-3'
9_SR6F	5'-GGTGATCTTCACGTCCTTGTTG-3'
9_SR6R	5'-CTTGTTCGATATTGCGCCGTGG-3'
TRA-IntF	5'-GGAACTATGACGACCAAGAAGC-3'
TRA-IntR	5'-GCTCGATCTTGGCCGTAGCTT-3'

Table 2. Antibacterial Activity of Compounds 1–**7**, 10–12**,** and 15–18

Compound	IC_{50}[a]/MIC[b]/MFC[c] (μg/mL)			
	M. intracellulare ATCC 23068	*M. tuberculosis* ATCC 25177	MRSA	*S. aureus* ATCC 29213
1	10.2/25.0/25.0	Nd/–/–	18.6/Nd/Nd	<8/Nd/Nd
2	Nd/25.0/25.0	Nd/–/–	–/Nd/Nd	68.7/Nd/Nd
3	36/50.0/50.0	Nd/–/–	–/Nd/Nd	–/Nd/Nd
4	Nd/25.0/25.0	Nd/–/–	27.6/Nd/Nd	20.7/Nd/Nd
5	Nd/–/–	Nd/–/–	Nd/Nd/Nd	Nd/Nd/Nd
6	Nd/25.0/25.0	Nd/–/–	Nd/Nd/Nd	Nd/Nd/Nd
7	Nd/–/–	Nd/–/–	Nd/Nd/Nd	Nd/Nd/Nd
10	Nd/25.0/25.0	Nd/50.0/–	Nd/Nd/Nd	Nd/Nd/Nd
11	Nd/50.0/50.0	Nd/–/–	Nd/Nd/Nd	Nd/Nd/Nd
12	Nd/50.0/50.0	Nd/100.0/–	Nd/Nd/Nd	Nd/Nd/Nd
15	Nd/25.0/25.0	Nd/–/–	Nd/Nd/Nd	Nd/Nd/Nd
16	Nd/50.0/50.0	Nd/–/–	Nd/Nd/Nd	Nd/Nd/Nd
17	Nd/50.0/50.0	Nd/–/–	Nd/Nd/Nd	Nd/Nd/Nd
18	Nd/12.5/12.5	Nd/50.0/–	Nd/Nd/Nd	Nd/Nd/Nd

[a] 50% growth inhibition

[b] Minimum inhibitory concentration (the lowest concentration that allows no detectable growth).

[c]Minimum fungicidal concentration (the lowest concentration that kills the fungus).

Nd Not determined.

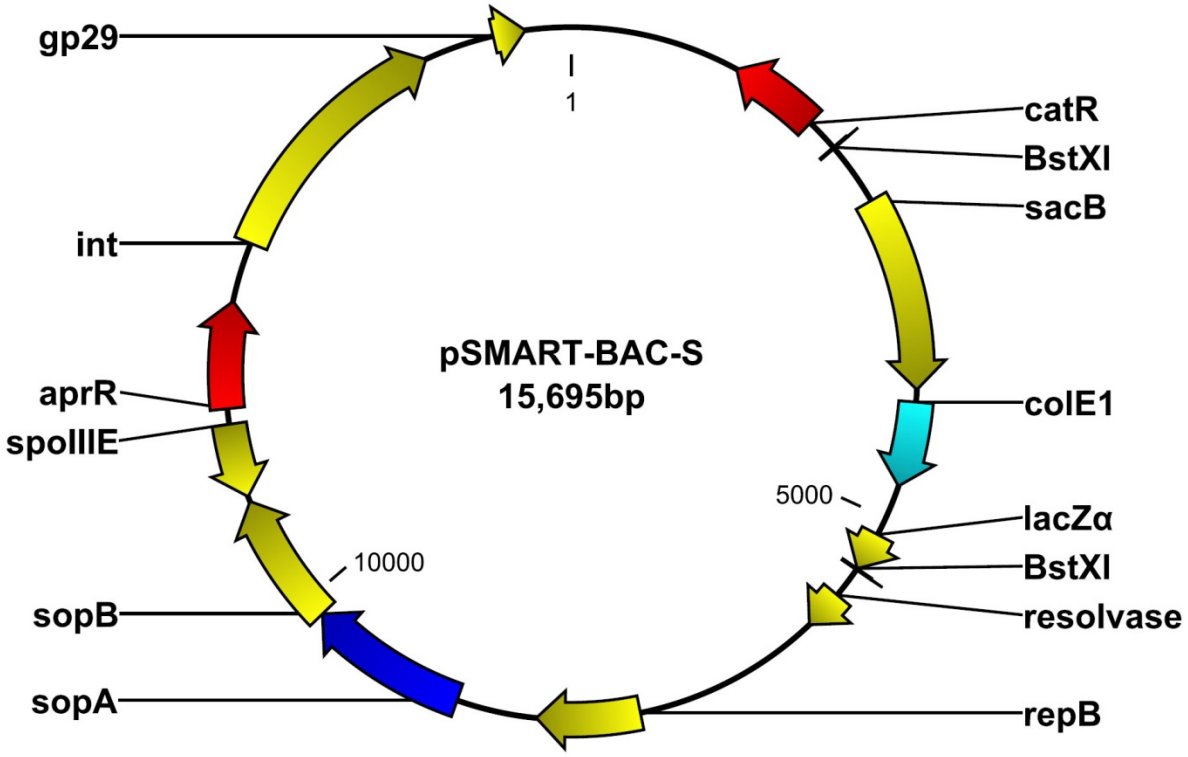


Figure 1. Schematic diagram of shuttle BAC vector pSMART BAC-S. The large-insert containing soil metagenomic library was constructed in this vector which provides high throughput conjugation for transferring large BACs into both Gram positive and Gram negative hosts, as well as integration and stable maintenance of the large BACs for heterologous expression.

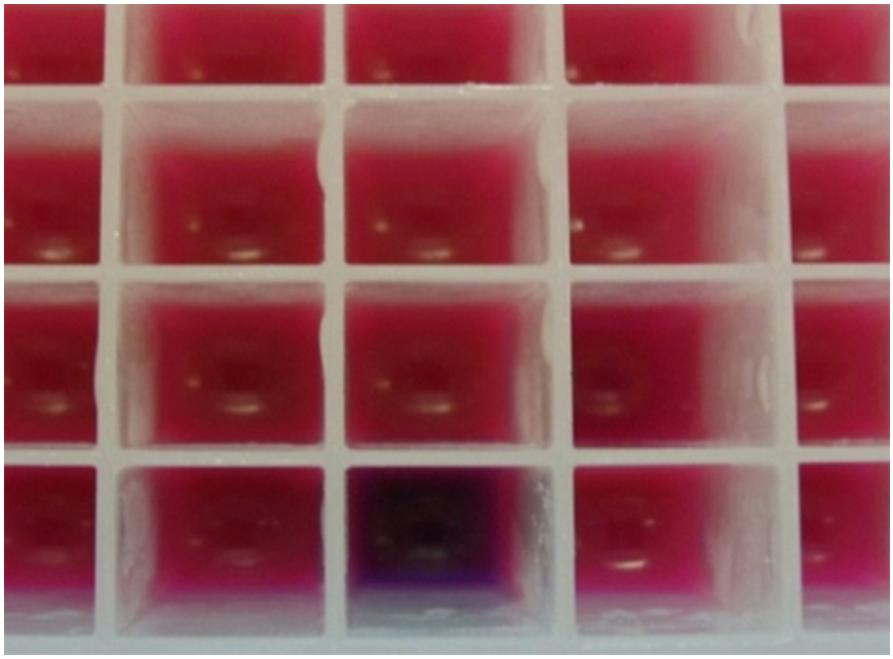
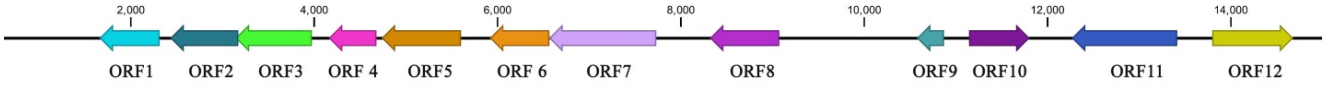


Figure 2. An example of metagenomic clone with anti-MRSA activity. The blue color indicated the lack of growth of MRSA strain 30 in the presence of BAC clone during the primary screen.



ORFs	Nucleotide position*	Description	Nearest neighbor	E-value	Greatest identity %	Accession numbers
ORF1	1660..2313	7-cyano-7-deazaguanine synthase	*Pseudoxanthomonas suwonensis*	4.4E-123	82.87	WP_024870324
ORF2	2432..3172	7-carboxy-7-deazaguanine synthase	*Rudaea cellulosilytica*	1.661E-107	77.25	WP_026293223
ORF3	3144..3974	tol-pal system protein	*Rudaea cellulosilytica*	2.4856E-77	63.51	WP_033409649
ORF4	4161..4679	membrane protein	*Mizugakiibacter sediminis*	7.002E-58	77.41	GAN44820
ORF5	4738..5601	translocation protein TolB	*Mizugakiibacter sediminis*	5.321E-157	74.91	GAN44821
ORF6	5918..6565	plasmid-partitioning protein	*Listeria monocytogenes*	1.875E-140	100	WP_039382837
ORF7	6565..7731	protein SopA	*Escherichia coli* O157:H7 str. Sakai	0	100	WP_000772446
ORF8	8319..9074	replication initiation protein	Escherichia coli K-12	0	100	WP_000852146
ORF9	10577..10870	resolvase, partial	*Acinetobacter venetianus*	2.0709E-57	100	WP_035311711
ORF10	11143..11802	chloramphenicol O-acetyltransferase	*Clostridium sp.*	4.001E-165	100	EDS05563
ORF11	12267..13415	TrfA	Cloning vector pGNS-BAC	0	100	ADO79617
ORF12	13797..14678	transcriptional regulator AraC	Broad host range vector pMLBAD	0	99.32	AAM63382

*** The number corresponds to the nucleotide positions in the largest contig of clone P6B5**

Figure 3. Schematic organization of ORFs located in the largest contigs of clone P6B5. ORF11 that encodes *trfA* was in a subclone after cloning into pSMART-HC-Amp.

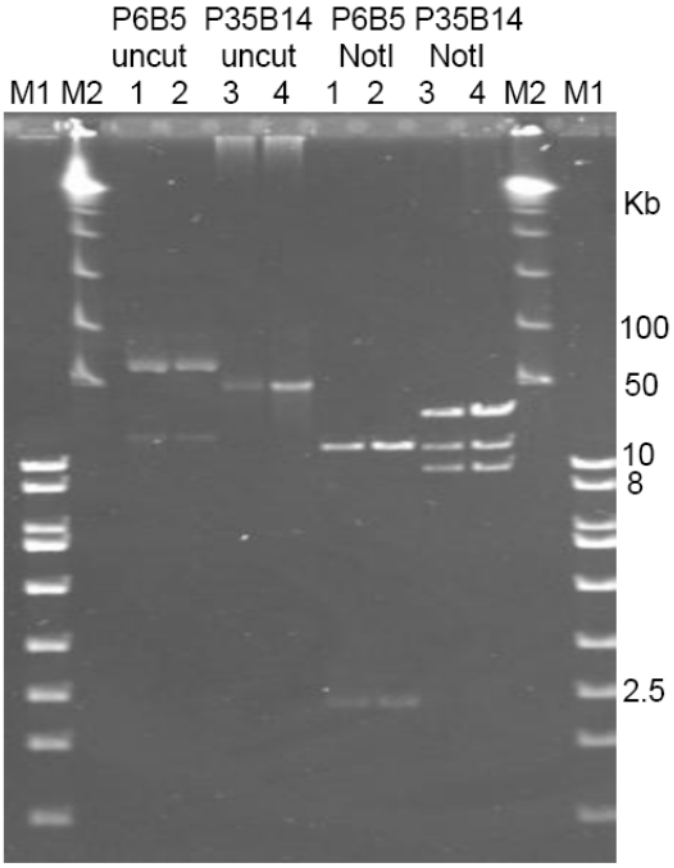
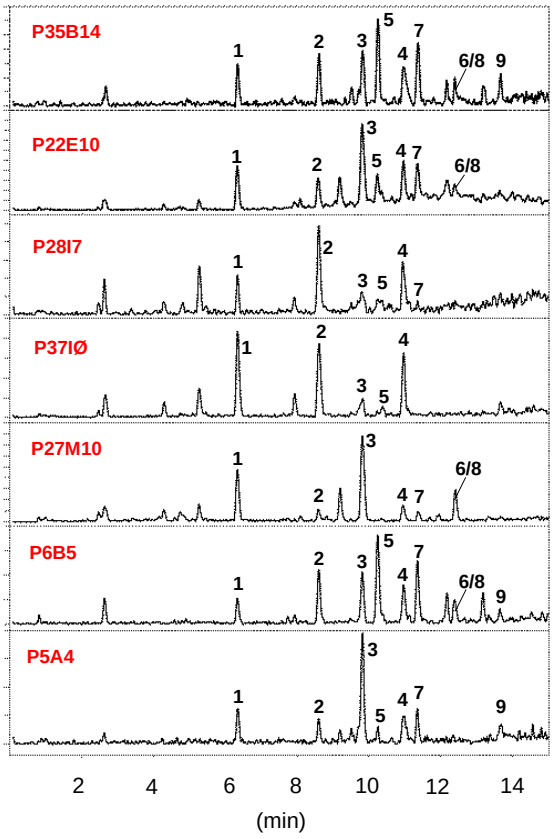
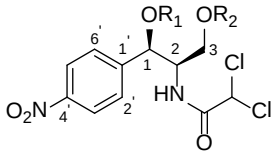


Figure 4. Size of the insert DNA in clones P6B5 and P35B14 was determined by restriction digestion followed by pulsed field gel electrophoresis using contour-clamped homogeneous electric field electrophoresis (CHEF). In CHEF gel, lanes M1and M2 contain 1kbp and Lambda ladders respectively. Each clone P6B5 and P35B14 was grown for 24 and 48 hours before extracting BAC DNAs (lane 1 and 3) and lane 2 and 4 respectively.





	R_1	R_2
1	H	H
2	H	$COCH_3$
3	H	$COCH_2CH_3$
4	H	$COCH_2CH_2CH_3$
5	$COCH_3$	$COCH_3$
6	$COCH_2CH_3$	$COCH_2CH_3$
7	$COCH_3$	$COCH_2CH_3$
8	$COCH_3$	$COCH_2CH_2CH_3$
9	$COCH_2CH_3$	$COCH_2CH_2CH_3$

Figure 5. LC-MS chromatograms of seven metagenomic clones showing Cm (1) and its derivatives (2−8). HPLC conditions: Acquity UPLC BEH C_{18} column (2.1 × 150 mm, 1.7 μm); gradient elution from 10% to 100% CH_3CN in H_2O.

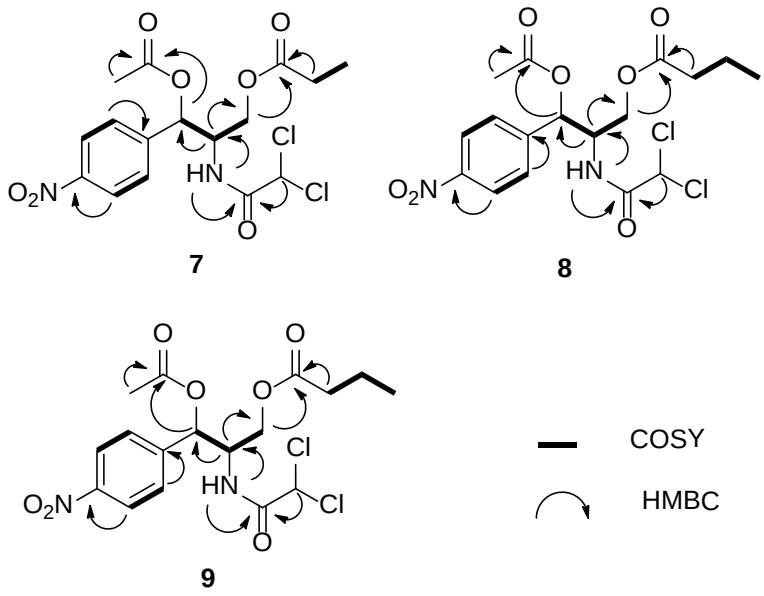


Figure 6. 2D NMR correlations of compounds 7−9.

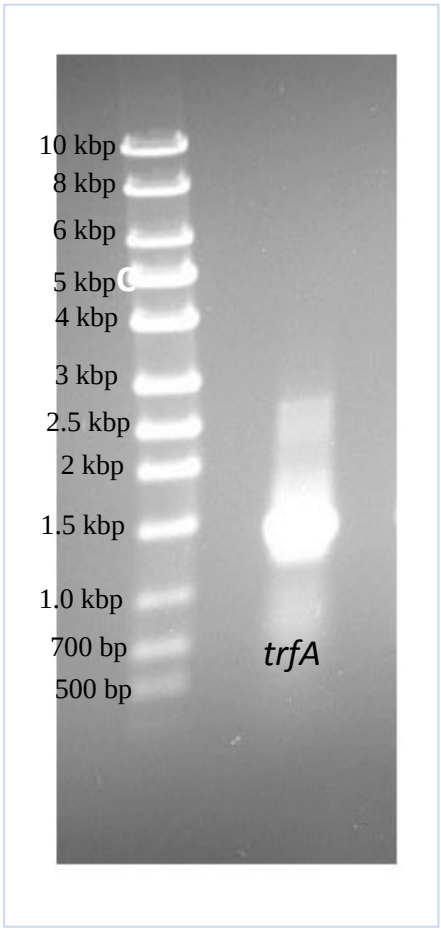


Figure 7. PCR amplification of *trfA* gene from clone P6B5. The size of the DNA band in agarose gel demonstrates correct size of the PCR product which 1,251 bp.

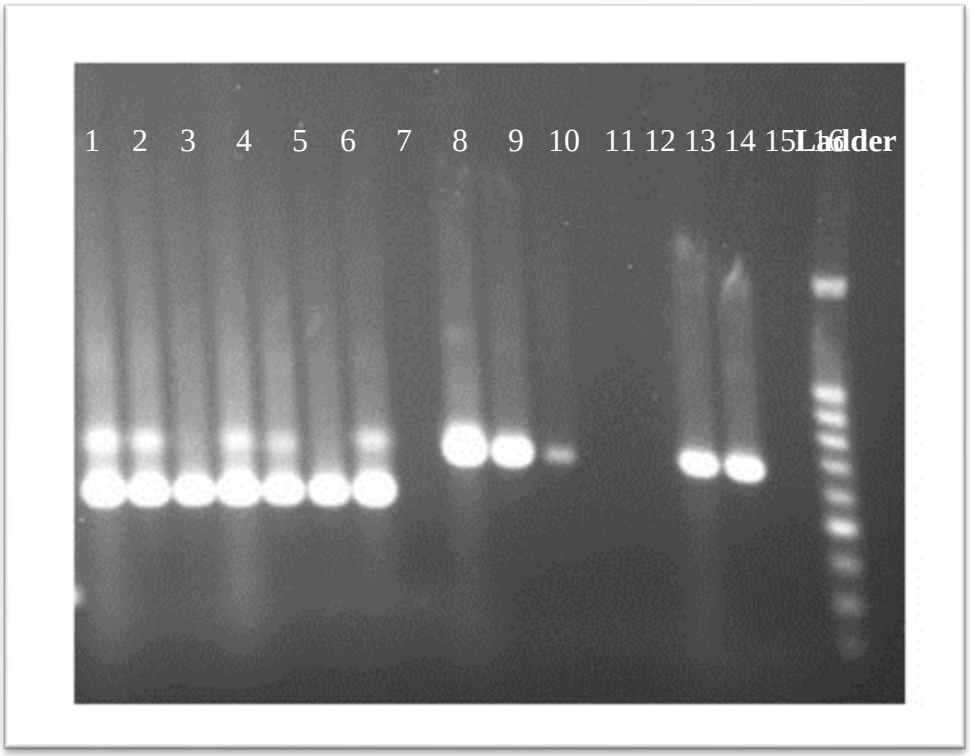


Figure 8. PCR screening of *trfA* containing subclones using vector and inserts specific primer sets. The presence of *trfA* gene in PCR positive clones was confirmed by sequencing.

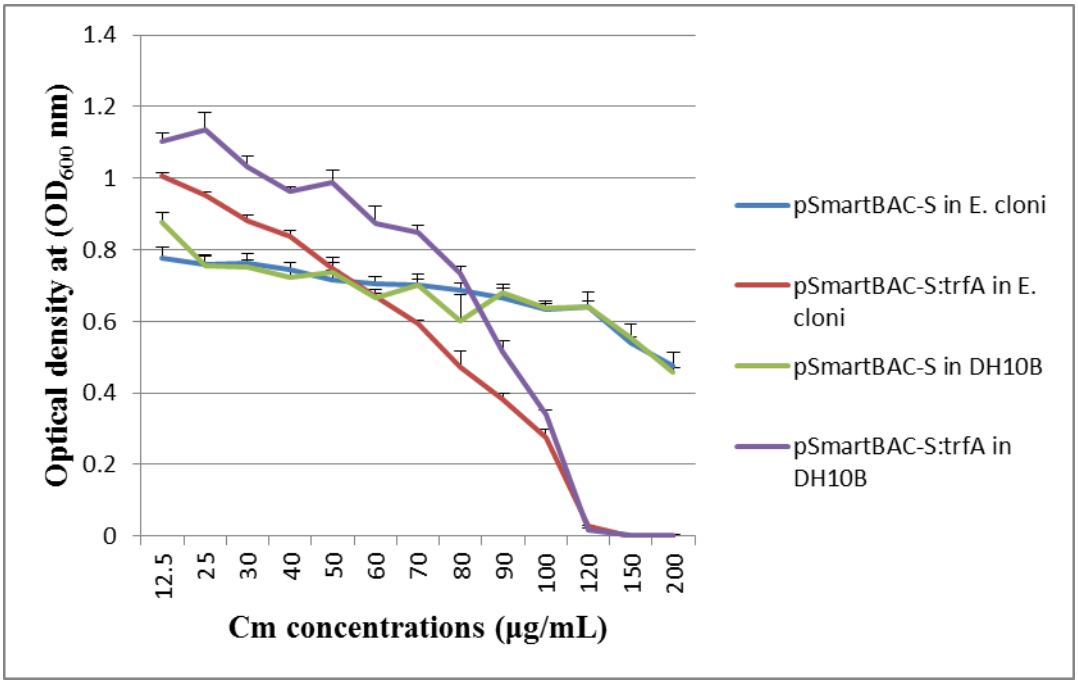


Figure 9. Antibiotic susceptibility of *trfA* containing clones. *trfA* gene was cloned into pSMART BAC-S vector and transformed into *E.coli* strains DH10B and E.cloni 10G. Antibiotic susceptibility of *trfA* clones was determined in the presence of Cm concentration ranging from 12.5 to 200 µg/mL. The empty vector controls showed significantly higher Cm resistance as compared to *trfA* containing clones.

6. Reference

Archer, N. K., M. J. Mazaitis, et al. (2011). "Staphylococcus aureus biofilms: properties, regulation, and roles in human disease." Virulence 2(5): 445-459.

Banik, J. J. and S. F. Brady (2010). "Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules." Curr Opin Microbiol 13(5): 603-609.

Bentley, S. D., K. F. Chater, et al. (2002). "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)." Nature 417(6885): 141-147.

Bizerra, A. M. C., T. G. C. Montenegro, et al. (2011). "Enzymatic regioselective production of chloramphenicol esters." Tetrahedron 67(16): 2858-2862.

Chow, J., F. Kovacic, et al. (2012). "The metagenome-derived enzymes LipS and LipT increase the diversity of known lipases." PLoS One 7(10): 24.

Courtois, S., C. M. Cappellano, et al. (2003). "Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products." Appl Environ Microbiol 69(1): 49-55.

Davies, J. and D. Davies (2010). "Origins and evolution of antibiotic resistance." Microbiology and Molecular Biology Reviews 74(3): 417-433.

Dougherty, M. J., P. D'Haeseleer, et al. (2012). "Glycoside hydrolases from a targeted compost metagenome, activity-screening and functional characterization." BMC Biotechnol 12(38): 1472-6750.

El-Kersh, T. A. and J. R. Plourde (1976). "Biotransformation of antibiotics. II. Investigation of the chloramphenicol acetyltransferase in Streptomyces griseus." J Antibiot 29(11): 1189-1198.

Fischbach, M. A. and C. T. Walsh (2009). "Antibiotics for emerging pathogens." Science 325(5944): 1089-1093.

Gillespie, D. E., S. F. Brady, et al. (2002). "Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA." Appl Environ Microbiol 68(9): 4301-4306.

Gross, F., E. A. Lewis, et al. (2002). "Isolation of 3' -O-acetylchloramphenicol: a possible intermediate in chloramphenicol biosynthesis." Bioorg Med Chem Lett 12(3): 283-286.

Handelsman, J., M. R. Rondon, et al. (1998). "Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products." Chem Biol 5(10): R245-249.

Havelsrud, O. E., T. H. Haverkamp, et al. (2011). "A metagenomic study of methanotrophic microorganisms in Coal Oil Point seep sediments." BMC Microbiol 11(221): 1471-2180.

Howden, B. P., J. K. Davies, et al. (2010). "Reduced vancomycin susceptibility in Staphylococcus aureus, including vancomycin-intermediate and heterogeneous vancomycin-intermediate strains: resistance mechanisms, laboratory detection, and clinical implications." Clin Microbiol Rev 23(1): 99-139.

Jenke-Kodama, H. and E. Dittmann (2009). "Evolution of metabolic diversity: insights from microbial polyketide synthases." Phytochemistry 70(15-16): 1858-1866.

Jones, K. E., N. G. Patel, et al. (2008). "Global trends in emerging infectious diseases." Nature 451(7181): 990-993.

Kakirde, K. S., L. C. Parsley, et al. (2010). "Size Does Matter: Application-driven Approaches for Soil Metagenomics." Soil Biol Biochem 42(11): 1911-1923.

Katoh, K. and D. M. Standley (2013). "MAFFT multiple sequence alignment software version 7: improvements in performance and usability." Mol Biol Evol 30(4): 772-780.

King, R. W., J. D. Bauer, et al. (2009). "An environmental DNA-derived type II polyketide biosynthetic pathway encodes the biosynthesis of the pentacyclic polyketide erdacin." Angew Chem Int Ed Engl 48(34): 6257-6261.

Kos, V. N., C. A. Desjardins, et al. (2012). "Comparative genomics of vancomycin-resistant Staphylococcus aureus strains and their positions within the clade most commonly associated with Methicillin-resistant S. aureus hospital-acquired infection in the United States." MBio 3(3): 00112-00112.

Lawrence, D. P., S. Kroken, et al. (2011). "Interkingdom gene transfer of a hybrid NPS/PKS from bacteria to filamentous Ascomycota." PLoS One 6(11): 29.

Levy, S. B. and B. Marshall (2004). "Antibacterial resistance worldwide: causes, challenges and responses." Nat Med 10(12 Suppl): S122-129.

Lewis, K. (2013). "Platforms for antibiotic discovery." Nat Rev Drug Discov 12(5): 371-387.

Liaw, R. B., M. P. Cheng, et al. (2010). "Use of metagenomic approaches to isolate lipolytic genes from activated sludge." Bioresour Technol 101(21): 8323-8329.

Liles, M. R., L. L. Williamson, et al. (2008). "Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms." Appl Environ Microbiol 74(10): 3302-3305.

Ling, L. L., T. Schneider, et al. (2015). "A new antibiotic kills pathogens without detectable resistance." Nature 517(7535): 455-459.

Liu, C., A. Bayer, et al. (2011). "Clinical practice guidelines by the infectious diseases society of america for the treatment of methicillin-resistant Staphylococcus aureus infections in adults and children." Clin Infect Dis 52(3): 4.

Lodise, T. P., J. Graves, et al. (2008). "Relationship between vancomycin MIC and failure

among patients with methicillin-resistant Staphylococcus aureus bacteremia treated with vancomycin." Antimicrob Agents Chemother 52(9): 3315-3320.

Lukashin, A. V. and M. Borodovsky (1998). "GeneMark.hmm: new solutions for gene finding." Nucleic Acids Res 26(4): 1107-1115.

MacNeil, I. A., C. L. Tiong, et al. (2001). "Expression and isolation of antimicrobial small molecules from soil DNA libraries." J Mol Microbiol Biotechnol 3(2): 301-308.

Martin, A., M. Camacho, et al. (2003). "Resazurin microtiter assay plate testing of Mycobacterium tuberculosis susceptibilities to second-line drugs: rapid, simple, and inexpensive method." Antimicrob Agents Chemother 47(11): 3616-3619.

Micek, S. T. (2007). "Alternatives to vancomycin for the treatment of methicillin-resistant Staphylococcus aureus infections." Clin Infect Dis 15(45): S184-190.

Mitchell, C. C. and D. Delaney (2012). "100 YEARS OF THE CULLARS ROTATION (c. 1911)." 2012 Beltwide Cotton Conferences,: 1335-1344.

NETWORK, D. A. S. O. "New MRSA drugs: tedizolid, dalbavancin, and oritavancin."

Newman, D. J. and G. M. Cragg (2007). "Natural products as sources of new drugs over the last 25 years." J Nat Prod 70(3): 461-477.

Nikolouli, K. and D. Mossialos (2012). "Bioactive compounds synthesized by non-ribosomal peptide synthetases and type-I polyketide synthases discovered through genome-mining and metagenomics." Biotechnol Lett 34(8): 1393-1403.

O'Brien, R. V., R. W. Davis, et al. (2014). "Computational identification and analysis of orphan assembly-line polyketide synthases." J Antibiot 67(1): 89-97.

Parsley, L. C., J. Linneman, et al. (2011). "Polyketide synthase pathways identified from a metagenomic library are derived from soil Acidobacteria." FEMS Microbiol Ecol 78(1):

176-187.

Perri, S., D. R. Helinski, et al. (1991). "Interactions of plasmid-encoded replication initiation proteins with the origin of DNA replication in the broad host range plasmid RK2." J Biol Chem 266(19): 12536-12543.

Pettit, R. K. (2004). "Soil DNA libraries for anticancer drug discovery." Cancer Chemother Pharmacol 54(1): 1-6.

Piel, J. (2002). "A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles." Proc Natl Acad Sci U S A 99(22): 14002-14007.

Reddy, B. V., A. Milshteyn, et al. (2014). "eSNaPD: a versatile, web-based bioinformatics platform for surveying and mining natural product biosynthetic diversity from metagenomes." Chem Biol 21(8): 1023-1033.

Rondon, M. R., P. R. August, et al. (2000). "Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms." Appl Environ Microbiol 66(6): 2541-2547.

Rossolini, G. M., F. Arena, et al. (2014). "Update on the antibiotic resistance crisis." Curr Opin Pharmacol 18: 56-60.

Sangwan, N., P. Lata, et al. (2012). "Comparative metagenomic analysis of soil microbial communities across three hexachlorocyclohexane contamination levels." PLoS One 7(9): 28.

Soriano, A., F. Marco, et al. (2008). "Influence of vancomycin minimum inhibitory concentration on the treatment of methicillin-resistant Staphylococcus aureus bacteremia." Clin Infect Dis 46(2): 193-200.

Steinkraus, G., R. White, et al. (2007). "Vancomycin MIC creep in non-vancomycin-

intermediate Staphylococcus aureus (VISA), vancomycin-susceptible clinical methicillin-resistant S. aureus (MRSA) blood isolates from 2001-05." J Antimicrob Chemother 60(4): 788-794.

Tarai, B., P. Das, et al. (2013). "Recurrent Challenges for Clinicians: Emergence of Methicillin-Resistant Staphylococcus aureus, Vancomycin Resistance, and Current Treatment Options." J Lab Physicians 5(2): 71-78.

Wang, X., Q. Liu, et al. (2013). "A BAC based physical map and genome survey of the rice false smut fungus Villosiclava virens." BMC Genomics 14(883): 1471-2164.

Wegrzyn, K., M. E. Fuentes-Perez, et al. (2014). "Sequence-specific interactions of Rep proteins with ssDNA in the AT-rich region of the plasmid replication origin." Nucleic Acids Res 42(12): 7807-7818.

Wemheuer, B., R. Taube, et al. (2013). "Microbial diversity and biochemical potential encoded by thermal spring metagenomes derived from the Kamchatka Peninsula." Archaea 136714(10): 27.

Westenberg, M., S. Bamps, et al. (2010). "Escherichia coli MW005: lambda Red-mediated recombineering and copy-number induction of oriV-equipped constructs in a single host." BMC Biotechnol 10(27): 1472-6750.

Wild, J., Z. Hradecna, et al. (2002). "Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones." Genome Res 12(9): 1434-1444.

Wilkinson, B. and J. Micklefield (2007). "Mining and engineering natural-product biosynthetic pathways." Nat Chem Biol 3(7): 379-386.

Yang, D., T. Mori, et al. (2014). "Expression, purification and crystallization of a fungal type III polyketide synthase that produces the csypyrones." Acta Crystallogr F Struct Biol

Commun 70(Pt 6): 730-733.

Zhang, X., Y. Song, et al. (2011). "Indirubin inhibits tumor growth by antitumor angiogenesis via blocking VEGFR2-mediated JAK/STAT3 signaling in endothelial cell." Int J Cancer 129(10): 2502-2511.

Chapter 5

Conclusion

Soil microbial communities are known to be a great resource for natural products but a majority of them have not been explored for their secondary metabolite synthesis because many of them are not readily cultured under laboratory conditions. In this study both culture-dependent and culture-independent approaches were used to discover antibiotics that inhibit the growth of methicillin-resistant clinical isolates of *Staphylococcus aureus* (MRSA).

Chapter 2. A novel cultivation approach using low-strength (1/200th) nutrient agar supplemented with soil extract and long incubation time was used for the isolation of soil bacteria that express antibacterial activity against MRSA. A collection of 548 unique bacterial and fungal isolates were isolated from soil using these novel cultivation techniques.Bacterial diversity analysis using 16S rRNA gene sequences of newly cultured isolates revealed that they represent diverse bacterial genera affiliated with the phyla *Acidobacteria, Actinobacteria, Bacteroidetes, Firmicutes,* and *Proteobacteria*. Two isolates, designated as A115 and F4, were found to inhibit the growth of pathogenic methicillin-resistant *Staphylococcus aureus* (MRSA). The isolate A115, member of the genus *Streptomyces,* produces a pink pigment after incubation for more than ten days. The isolate F4, identified as a *Nonomuraea* spp., produces a high molecular weight (>100kDa), heat stable reddish pigment associated with anti-MRSA activity. Genome sequencing using a combination of shotgun and mate-pair next-generation sequencing resulted in the complete assembled genome for each isolate, with the size of the A115 and F4 genomes at 8.6 Mb and 10.3 Mb, respectively. The %G+C contents of strains A115 and F4 were determined to be 71% and 70.4%, respectively. Phylogenetic analysis using multilocus sequence analysis with six housekeeping genes revealed that strain A115 was most closely related to

Streptomyces afghaniensis and *Streptomyces olindensis*; however, the low level of average nucleotide identity (ANI) values in comparing the A115 genome were 89.76% and 89.14% for *S. afghaniensis* and *S. olindensis*, respectively. These genomic results, combined with differentiation of strain A115 from other *Streptomyces* species by morphological and physiological characteristics, led to the conclusion that strain A115 represents a novel species of the genus *Streptomyces*, for which the name *Streptomyces alburnustigris* (Latin word for auburn tiger) sp. nov. is proposed. Phylogenetic analysis based on 16S rRNA gene sequence revealed that the closest phylogenetic relative of F4 strain was *Nonomuraea antimicrobica* YIM 61105. *In silico* analysis using anti-SMASH predicts that the A115 and F4 genomes encode many gene clusters for secondary metabolite biosynthesis, including the synthesis of terpene, aminoglycoside, thiopeptide, bacteriocin, oligosaccharide, phenazine, butyrolactone, siderophore, melanine and potentially other bioactive compounds produced by non-ribosomal peptide synthetase and polyketide synthetase pathways. Both *S. alburnustigris* A115 and *Nonomuraea* spp. strain F4 genomes are predicted to encode Type I, II, and III PKS pathways. The biochemical structure of the active anti-MRSA compounds are currently being characterized using liquid chromatography–mass spectrometry (LC/MS). An organic extract of A115 supernatant revealed strong activity against the fungal pathogens *Candida glabrata*, *C. krusei* and *Cryptococcus neoformans*. The antifungal compounds extracted from A115 supernatant appeared to be novel; however, none of the organic extracts were found to be active against MRSA and the compound(s) active against MRSA will require alternative extraction methods for their biochemical structural elucidation. This study identified novel bacterial isolates with anti-MRSA activity and demonstrates the utility of novel cultivation techniques in obtaining

previously uncultured and phylogenetically diverse soil microorganisms, some of which express potent bioactive secondary metabolites.

Chapter 3. In this study, a collection of bacterial rhizosphere isolates were screened to identify novel chemical compounds for MRSA control. Five *Bacillus* strains that expressed metabolites with anti-MRSA activity were identified. Among them, *Bacillus amyloliquefaciens* strain AP183 was found to produce a novel macrodiolide compound described herein as Bacillusin A that has potent anti-MRSA activity with a minimum inhibitory concentration of 0.6 µg/mL. Because Bacillusin A has a short half-life after extraction, we hypothesized that it may not persist within living tissue and would therefore be suitable for *in vivo* application. AP183 was tested *in vivo* as a skin probiotic to prevent MRSA infection using a mouse model. Mice were simultaneously challenged with bioluminescent *S. aureus* strain Xen29 with and without AP183 spores in two separate wounds. In additional experiments, we tested the effects of AP183 spores with and without accompanying secondary metabolites. After challenge, skin wound healing was monitored for one week and *S. aureus* growth was assessed by bioluminescent imaging. After one week, mice were sacrificed and wounds were homogenized and plated to determine culturable bacterial counts and to conduct a culture-independent skin microbiome analysis. Our *in vivo* studies showed that co-administration of secondary metabolites and AP183 spores resulted in a significant reduction in the number of *S. aureus* colonization compared to a negative control. Molecular phylogenetic analysis revealed a significant reduction in *S. aureus* relative abundance when AP183 was applied while the relative abundance of other bacterial taxa increased in the skin microbiome as a result of probiotic administration. In future work, the *in vivo* efficacy and safety for the application of strain AP183 and its active metabolites will be determined.

Chapter 4. A metagenomics approach was also applied in this study in order to discover antibacterial compounds. A metagenomic library was constructed using high molecular weight DNA from the microbial assemblage from the Cullars Rotation agricultural soil (a plot without N or P amendments) and cloned that DNA into a shuttle bacterial artificial chromosome (BAC) vector. The soil metagenomic library is comprised of 19,200 *E. coli* clones with an average insert size of 110 kb. *E. coli* clones containing metagenomic DNA were screened for inhibition of growth of MRSA using a 96-well microtitre plate format. *In situ* lysis of the *E. coli* host enabled detection of both intra- and extracellular compounds, yielding a total of 28 anti-MRSA clones. Transformation of naïve *E. coli* with BAC DNA isolated from anti-MRSA clones confirmed the presence of their anti-MRSA activity. Sequencing and sub-cloning of these clones revealed genes predicted to be involved in various biosynthetic pathways as well as many genes with unknown functions. This culture-independent approach discovered seven metagenomic clones with the capacity to modify chloramphenicol which was added to the *E. coli* culture medium, thereby resulting in modification of an existing antimicrobial scaffold. LC-MS analysis of the organic extract of the clones revealed three new chloramphenicol derivatives. Chemically synthesized chloramphenicol derivatives tested showed antibacterial activity against MRSA, *Mycobacterium intracellulare* and *M. tuberculosis* with MICs of 27.6, 12.5 and 50.0 μg/mL, respectively. These results demonstrate that large-insert soil metagenomic libraries can be screened using innovative functional screening methods to access previously undescribed genomic and biochemical diversity.The progress made in these studies toward generation of large-insert metagenomic libraries in shuttle BAC vectors will be applied in future for generation of larger-scale libraries that can encompass a greater diversity of soil microbial metagenomes and be expressed in multiple hosts. In particular, the use of specific *E. coli* and other

heterologous hosts engineered for expression of polyketide synthases as well as other biosynthetic pathways will take further advantage of these libraries for natural product discovery.

Appendix A

>A12
CAAGGNGNNGCTTANCATGCAGTTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGT
AACGCGTGGGAATCTACCCATCACTACGGAACAACTCCGGGAAACTGGAGCTAATACCGT
ATACGTCCTTCGGGAGAAAGATTTATCGGTGATGGATGAGCCCGCGTTGGATTAGCTAGT
TGGTGGGGTAATGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCC
ACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGAC
AATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAA
GCTCTTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTG
CCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCG
CACGTAGGCGGACTATTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTT
TGATACTGGTAGTCTCGAGTCCGAGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAAT
TCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGCTCGGTACTGACGCT
GAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAAC
GATGGAAGCTAGCCGTCGGCAAGTTTACTTGTCGGTGGCGCAGCTAACGCATTTAAGCTT
CCNGCCTGGGGAGTACGGTCGCAAGATTAAAACTCCAAAGGAATTGACGGGGGCCCGCAC
AAGCCGGTGGGAGCATGTGGTTTNANTTCGAAGCAACCGCGCAGAAACNTTACCCAGCCC
CNTTGACATCCCGGGTCGCGGTTTTCCAGAAAATGGAATCCCTTCANTTTCGGCTGGACC
NGGNNACAGGTNNTTGCATGGCTGTTCCCCAACTTCNTGTTCNTNAGAATNTTTGGGTTA
AAGTCCCCGCNANCGAAGCNCAACCCCTCCCCCTTNATTTGCCNTCNTTAANTTGGGCAN
NNNANGGGGCCTNCCGGNTANAAACCCAAAANGAAAGTTGGGAATAACCNCANTTCTNAT
NGGCCCTTTNNGGNTGNGCTNACCCCNNNCTTCNNTNNGGGGNGNNAATGGGGNCACNAA
ANNCCCCTGTNCTNNCTTNNTN
>A13
NATGCGCGNCTTANCATGCAAGTCGAACGATGAAGCCCTTCGGGGTGGATTAGTGGCGAA
CGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGT
CTAATACCGGATACCACTCCTGCCTGCATGGGCGGGGGTTGAAAGCTCCGGCGGTGAAGG
ATGAGCCCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCCACCAAGGCGACGACGGGT
AGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGA
GGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGACGGTA
CCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAA
GCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCGCGTCGGATGTGA
AAGCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCAGGCTAGAGTGTGGTAGGGG
AGATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCG
AAGGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGA
TTAGATACCCTGGTAGTCCACGCCGTAAACGTTGGGAACTAGGTGTTGGCGACATTCCAC
GTCGTCGGTGCCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGGCCGCAAGGCT
AAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGANCATGTGGCTTTATTCGAC
GCAACGCGAAAAAACCTTACCAAGGCTTTGACATNTACCGGAAAGCATTTANAGGATAGT
GCCCCCCCCTTGGTGGTCGGTATNACAGGTGGTNCATGGGCTGTCNTCAACTTCNTTTCN
TGANAATNTTGGGTTAAANTCCCGCAAACNAANCCCNAACCCTTTGTCCTTTGTTTGCCA
NCATGCCCTTCNGGGGNATGGGGANCTCNAAAGAAACCCCCGGGGTTCAATTTGAAAGAA
AGGNGGGGAANANCTTCAATTCNTCTTNCCCCCTTATTTTTTGG
>A18
NGNGNGNCTTANCATGCAGTCGAACGATGAAGCCCTTCNGGGTGGATTAGGTGGCGAACG
GGTGAGTAACACGCTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTC
TAATACCGGATACCACATCCTGCCTGCATGGGCGGGGGTTGAAAGCTCCGGCGGTGAAGG

ATGAGCCCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCCACCAAGGCNACTACGGGT
AGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAAACTCCTACNG
GAGGCAGCANTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCANCGACGCCGCGTGA
GGGATGACGGCCTTCNGGTTGTAAACCTCTTTCANCANGNAAGAANCGAAAGTGACGGTA
CCTGCANAANAAGCGCCNGCTAACTACNTGCCANCAGCCGCGGTAATACGTAGGGCGCAA
GCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCGCGTCNGATGTGA
AAGCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCAGGCTAGAGTGTGGTAGGGG
AGATCGNAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCATGAGGAACACCNGTGGCT
GAAGGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGNAAGCGTGGGGAGCGAACAGG
ATTAGATACCCTGGTAGTCCACGCCGTAAACGTTGGGAACTAGGTGTTGGCGACATTCNA
CNTCCTCGGTGCCGCAGCTAACGCATTAAGTTCCCCCNCCCTGGNGAGTACGGCCNCAAG
GCTAAAACTCAAANGGAATTGNCNGGGGGCCCGCACAAGCAGCNGAATNATGTGGCTTNA
ATTCTACNCNANCNCGAAGAAACCTTACCAAGGNNTNGACATATNCCGGNAAAGCATTTA
NAAAATNGTGCCCCCCCTTGTGGTCCGGTATNNAGGTGGTGGCATGGCTTNTCTNCANCC
TCNTGTCNNGANAATGTTNGGGTTAATTNCCNNNNAANAANTNCAANCCTTNTTNTTNGG
TTGGCCNGCCTGCCCTTTCGGGGTTATGGGNAATTCCANNNAAAACCCCNNGGGTCAAAT
CCNNAAAAAGGTNGGGAANAACNTCAATTNATATGNCCCNTNT
>A2
ATGCGCNGCCTACCATGCAAGTCGAACGCCCCGCAAGGGGAGTGGCGCACGGGTGAGTAA
CGCGTGGGAATATGCCCTTCGGTTCGGAATAACACAGGGAAACTTGTGCTAATACCGGAT
ACGATCTACGGATGAAAGATTCATCGCCGAAGGAGTAGCCCGCGTAGGATTAGCTAGTTG
GTGAGGTAATGGCTCACCAAGGCGACGATCCTTAGCTGGTCTGAGAGGATGATCAGCCAC
ACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAA
TGGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGC
TCTTTCGCCAGGGAAGATAATGACGGTACCTGGAAAAGAAGCCTCGGCTAACTCCGTGCC
AGCAGCCGCGGTAAGACGGAGGAGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGTG
CGCAGGTTGTGAGTTCAGTTGGATGTGAAAGCCCGGGGCTTAACCTCGGATGTGCATCCA
ATACTGGCTCGCTGGAGGTTGGAAGAGGAGAGTGGAATTCCCAGTGTAGAGGTGAAATTC
GTAGATATTGGGAAGAACACCAGTGGCGAAGGCGGCTCTCTGGTCCATACCTGACACTCA
TGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGA
TGTGTGCTAGACGTCGGGAGGCTTGCCTCTCGGTGTCGCAGCTAACGCGATAAGCACACC
GCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAAAGGAATTGACGGGGGCCCGCACAAG
CGGTGGAGCATGTGGNTTTAATTCGAAGCAACGCGCANAACCTTACCAGCCCCTTGACAT
GGGAAGTATGGACTANANAGATCTAGTTCTTCAGTTCGGCTGGCTTCCCACAGGTGCTGC
ATGGCTGTCNTCANCTCNTGTCGTGANATGTTTGGNTTAATTCCCNCAACGNNNGCAACC
CTTCNCCTTCATTTGCCATCNCTTCTAGGNGGGCCCNNTTAAAGAACTGCCNGNTNACAA
CCCGAAGAANGGGGGGGNATNACNTCAATTCCTNANGGCCCTTNANGGGTTGGGNTANCC
CNNNTGCTNNAATGGCNAANCCNNTGGGAAACNNAGGGANCNAN
>A25
ANGNGCGGCCTACCATGCAAGTCGAACGCCCCGCAAGGGGAGTGGCGCACGGGTGAGTAA
CACGTGGGAACCTACCTTCTGGTACGGAACAACCAAGGGAAACTTTGGCTAATACCGTAT
ACGACCTCCGGGTGAAAGATTTATCGCCGGAAGAGGGGCCCGCGTCCGATTAGGTAGTTG
GTGGGGTAAAGGCCTACCAAGCCGACGATCGGTAGCTGGTCTGAGAGGATGACCAGCCAC
ACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAA
TGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTGTAAAGC
TCTTTCACCCACGACGATGATGACGGTAGTGGGAGAAGAAGCCCCGGCTAACTTCGTGCC
AGCAGCCGCGGTAATACGAAGGGGGCAAGCGTTGTTCGGAATGACTGGGCGTAAAGGGCG
CGTAGGCGGTTCGTTGCGTCAGATGTGAAAGCCCCGGGCTCAACCTGGGAACTGCATTTG
ATACGGGCGGGCTTGAATCCAAGAGAGGGTGGTGGAATTCCCAGTGTAGAGGTGAAATTC
GTAGATATTGGGAAGAACACCAGTGGCGAAGGCGGCCACCTGGCTTGGTATTGACGCTGA
GGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGA
TGTGTGCTAGCCGTCGGGCAGCTTGCTGTTCGGTGGCGCAGCTAACGCGATAAGCACACC
GCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGGCCCGCACAAG
CGGGTGGAGCATGTGGTTTAATTCGAAGCAACNCGCANAACCTTACCAACCCTTGACATG

GGGAAGTATGGGCCTGGGAAAACCGGGTTCCTTCANTNCGGCTGGCTTCCNCACAGGTGC
TGCATGGCTGTCNTCANCTCGTGTCGTGANATTTTTGGTTAAGTNCCCGCAACGAGNGCN
ANCCTTCCTNTTTAGTTGCCCTCATTGANTTGGGCCCTCTGGAAATACTGCCCGGTAANA
ANCCCGGAAGAAAGCGGGGANNAACCTTAAATTCCNCTNGGCCCTTTNCGGGTTTGGGCT
ANCNCNTTGCTTNAATGGGGGGTAA
>A26
AGCGCNCCTACCATGCAGTCGAACGAGNACCTTCGGGATCTAGTGGCGCACGGGTGCGGT
AACGCGNTGGTGAATCTGCCCTTGGGTTCGGGATAACAGTTGGAAACGACTGCTAATACC
GGATGATGACTTCGGTCCAAAGATTTATCGCCCAGGGATGAGCCCGCGTCNGATTAGCTA
GTTGGTGAGGTAAAGGCTCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGATCAG
CCACACTGGGACTGAGACACGGCCCANACTCCTACGGGAGGCAGCAGTGGGGAATATTGG
ACAATGGGCGAAAGCCTGATCCANCNATGCCGCGTGAGTGATGAANGCCTTAGGGTTGTA
AAGCTCTTTTACCCGGGATGATAATGACAGTACCGGGAGAATAANCCCCGGCTAACTCCG
TGCCAGCANCCGCNGTAATACNGANGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAG
CGCACGTANGCGGCTTTGTAAGTTAGAGGTGAAAGCCCGGGGCTCAACTCCGGAACTGCC
TTTAANACTGCATCGCTAGAACGTCGGAGAGGTAAGTGGAATTCCGAGTGTAGAGGTGAA
ATTCGTANATATTCNGAAGAACACCAGTGGCNAAGGCGACTTACTGGACGACTGTTGACN
CTGANGTGCNAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACNCCGTAA
ACGATGATGACTANCTGTCGGGGCTCATGGAGTTTCGGTGGCGCAGCTAACGCGATAAGT
CATCCGCCTGGNGAGTACGGCCGCAAGGTTAAAACTCAAAGAAATTGACTGGGGCCTGCA
CAANCGGTGGAGCATGTGGTTTANTTCGAAGCAACGCGCAGAACCTTACCAGCGGTTTGA
CATGGTAGNACAGTTTCCAGAAATGGAATTCNTTCCCTTCGGGACCTACANCCCAGGTGC
TTGCATGGCTGTCNANCANCTCNTGTCNGGAAANTGTTGGGTTTAAGTTCCCCAACNGAG
CNCAACCCNNCGNCTTTTAGTTGCCTACNATTTNAGTTGGGGCNCTNNTAAAAAAAACCT
TCCCGNTTTAAAACCCCGGANGNAANGTGGGGGAANGACTTNAAANTCNNNTGGCCCNTT
ACNCCCG
>A28
ANGCGCNGCTTACCATGCAAGTCGAACGCGTGTAGCAATACACGAGTGGCGCACGGGTGA
GTAACGCGTGGATATCTGCCTTTTGGTTCGGAATAACCCCGGGAAACTGGGGCTAATACC
GGATGGTTCCTTCGGGATAAAGATTTATCGCCAAAAGATGAGTCCGCGTACGATTAGCTA
GTTGGTGAGGTAATGGCTCACCAAGGCGACGATCGTTAGCTGGTCTGAGAGGACGATCAG
CCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGG
ACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTGTA
AAGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAACTTCG
TGCCAGCAGCCGCGGTAAGACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAG
CGCGTGTAGGCGGTTGCCCAAGTCGGGTGTGAAAGCCTTGAGCTCAACTCAAGAAATGCA
CTCGGTACTGGGTGACTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTGGTGAA
ATACGTAGAGATTGGGAAGAACACCAGTGGCGAAGGCGGCTATCTGGACGGTTTCTGACG
CTAAGACGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAA
ACGATGGGTGCTAGACGTTGGCGAGCTTGCTCGTCAGTGTCGCAGCTAACGCGTTAAGCA
CCCCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCAC
AAGCGGTGGAGCATGTGGTTCAATTCGACGCAACGCGCANAACCTTACCAGCCCTTGACA
TGGGNACTCGCCNGGAGCAGANACGCTCCCCTTCGGTTCGGCCNGANTTCCNCACAGGTG
CTGCATGGCTGTCGTCAGCTCGTGTCGTGAAATTTTGGGTTAAGTCCCGCAACNGAGNGC
AACCCTCGTCTCCGTTGCCCTTCAGGTTATGCCTGGGCNNCTTNGNAAAACTGCCCGNTN
ACAACCCGGAAGAAAGGNGGGGATAACCTNCAANTNCCTNANGGCCCTTACGGGNTGGGC
TCC
>A29
ATGCGCGNCTATCATGCAAGTCGAGCGAATCAATGGGAGCTTGCTCCCTGAGATTAGCGG
CGGACGGGTGAGTAACACGTGGGCAACCTGCCTATAAGACTGGGATAACTTCGGGAAACC
GGAGCTAATACCGGATACGTTCTTTTCTCGCATGAGAGAAGATGGAAAGACGGTTTACGC
TGTCACTTATAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAATGGCTCACCAAG
GCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCC
AGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGC

AACGCCGCGTGAACGAAGAAGGCCTTCGGGTCGTAAAGTTCTGTTGTTAGGGAAGAACAA
GTACCAGAGTAACTGCTGGTACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACG
TGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAG
CGCGCGCAGGTGGTTCCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCA
TTGGAAACTGGGGAACTTGAGTGCAGAAGAGGAAAGTGGAATTCCAAGTGTAGCGGTGAA
ATGCGTAGAGATTTGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACA
CTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAA
ACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTTAGTGCTGCAGCTAACGCATTAA
GCACTCCGCCTGGGGNAGTACNGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCC
GCNCAAGCGGTGGGAGCATGTGGTTTTAATTCGAANCAACCGCGAAGAACCTTNACCANG
GTNTTGACATCCTCTGACAACCCTNAGAAAATAAGGNCTTTNCCCCTTCGGGGGAAAAAA
TTGACAGGTNGGTGCATGGTTGTTCTNNACCTCGTGTNCNGGAAATTTTNGGGTTTAATT
CCCNCAACCAAGNGCNACCCCTNGATTNTNATTTGCCNGCATTCANTTTGGGCNCTNTAA
GGGGACTGCCGG
>A3
ATCGCNGCCTATCATGCAGTTCGAACGGTAACTGGAGTAGCAATACTTCGGCTAGAGTGA
CGTAAGGGTGCGTAACACGTATGCAATCTGCCCTGTACAGGAGGATAGCTCCCCGAAAGG
GGAATTAACACTCCATAATATAGTTGGCCGGCATCGGTTGATTATTAAAACTGAGGTGGT
ACAGGATGAGCATGCGTCTGATTAGCTAGTTGGTAGTGTAATGGACTACCAAGGCGATGA
TCAGTAGGGGAACTGAGAGGTTGATCCCCCACACTGGCACTGAGATACGGGCCAGACTCC
TACGGGAGGCAGCAGTAGGGAATATTGGTCAATGGGTGAGAGCCTGAACCAGCCATGCCG
CGTGCAGGAAGAAGGCCTTCTGGGTTGTAAACTGCTTTTGCCAGGGGATAAAACGGGAGT
GCGCTCCTAATTGAAGGTACCTGGTGAATAAGCCACGGCTAACTACGTGCCAGCAGCCGC
GGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGG
TTCTTTAAGTCAGTGCTGAAATACTCTAGCTTAACTAGAGGGGTGGCATTGATACTGAGG
AACTTGAGTAGAGTCGAGGTAGGCGGAATTGACGGTGTAGCGGTGAAATGCTTAGATATC
GTCAAGAACACCGATAGTGTAGACAGCTTACTAGGCTTCAACTGACGCTGAGGCACGAAA
GTGTGGGGATCAAACAGGATTAGATACCCTGGTAGTCCACACTGTAAACGTTGATTACTC
GCTGTTGGCGATATACAGTCAGCGGCTTAGCGCAAGCGATAAGTAATCCACCTGGGGAGT
ACGCCGGCAACGGTGAAACTCAAAGGAATTGACNGGGGTTCCGCACAAGCGGTGGAGCAT
GTGGTTTAATTCGATGAATACGCNGAGGAAACCCTTACCTGGGCTANANTGCCCTTGATG
TCCCTCAAANACCAAGGAANTNCCNCAAGGACCAAGGANCAANGGGCTTGCATGGGCTTG
TCNTNNACCTCGTNNCCGTGAGGGNTTGGGTTTAANTCCCCCCAACNAGNCNCAACCCTT
TNTTTTTANTTNCCAACNNGNNCAACCTTGGGGACNCTNAAAAAACNGCCNNNCCCNANC
CNAAAAGAAANGGNGGGGAATGACCTNNANTNTTNATGGGCCCTTNANCCCNGGG
>A32
ATGCGCNGCTACCATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAAC
GCGTGGGAACGTACCCTTTACTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATG
TGCCCTTCGGGGGAAAGATTTATCGGTAAGGGATCGGCCCGCGTTGGATTAGCTAGTTGG
TGGGGTAAAGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACA
TTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAAT
GGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCT
CTTTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCA
GCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCAC
GTAGGCGGACATTTAAGTCAGGGGTGAAATCCCAGAGCTCAACTCTGGAACTGCCTTTGA
TACTGGGTGTCTGGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCG
TAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTACTGACGCTGAG
GTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGAT
GAATGTTAGCCGTCGGCAAGTTTACTTGTCGGTGGCGCAGCTAACGCATTAAACATTCCG
CCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCG
GTGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACNTTACCAGCCCTTGACATTCCG
GTGCTACATCCCAAANATGGGATGGTTTCCTTTCGGGANCNCCNANAANAGGTGCTGCNA
TGGCTGTTCNTCCANCTTCNTGTCNTGANATNTTGGGTTTAANTCCCCNANCAAGNGCNA
ACCCCTCNCCCTTTATTTNCCAACNTTTCNANTTGGGCNCTCTTANGGGNANTGCCCGNT

GATAANCCNAAAAGNAAGGGGGGAANANCNNAANTNCTCTNGGNCCCTTTCCGGGCTGGN
TTACCNCNNNCTNAANNGG
>A34
ATGCGCGNCNTATCATGCAGTCGAGCGAATCGACAGGTGCTTGCACCTGTTGGTTAGCGG
CGGACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTTCGGGAAACC
GGAGCTAATACCGGATAATCCTTTTCCTCTCATGAGGAAAAGCTGAAAGTCGGTTTACGC
TGACACTTACAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAG
GCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCC
AGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGC
AACGCCGCGTGAGCGATGAAGGCCTTCGGGTCGTAAAGCTCTGTTGTTAGGGAAGAACAA
GTACCGGAGTAACTGCCGGTACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACG
TGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAG
CGCGCGCAGGCGGTCCTTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCA
TTGGAAACTGGGGGACTTGAGTACAGAAGAGGAAAGCGGAATTCCACGTGTAGCGGTGAA
ATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGTCTGTAACTGACG
CTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAA
ACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTAGTGCTGCAGCTAACGCATTAAG
CACTCCGCCTGGGGAGTACGGCCGCAAGGGCTGAAACTCAAAGGAATTTGACGGGGGCCC
CGCACAAGCGGGTGGAGCATGTGGTTTTAATTCNAANGCAACNCNGAANAACCTTACCAN
GGTCTTGACATCCTCTGACNCTCCTANAAANTAGGATTTTNCCCCTTNCGGGGGGAAANA
ANTGACAGGTGGTGCAANGGTTGTCCTCANCCNCNTGTCCTGAANATGTTTGGGTTAAAN
TCCCCCCANCNANNCNCAACCCCTNGATTTTANTTNCCNACNATTCANTTNGGGCNCTTT
TAGGNTGACTNNCNGGTNA
>A35
GGCAGGNGCNGCTTAACATGCAGNTCGAGCGCCCCGCNAAGGNTAAGCCGNCAGTANGGN
TGAGGTANCNGCGATGGGCANCNTTACCCATCTTNCTACNGCAATNACCCNNGGAAACTT
GTGCCTNNTNCCNTATGNTGNCCTTCNGGGGAAAGATTTATCTGAAAAGGGATCNGCCCG
CNATTGGATNNGCTAGNTGGNGGGGTNAAGGCCTACCATGGCGACGATCCNTATANTGGT
CTGAGAGGATGATCACCCNCATTGGGACTGAAACNCNGNCCNNACTCCTNCNNGAGGCAN
CNGNGNGGAATATNGGACANTGGGCGCANGCCTGATCCAGCCATGNCGCNNGANTGATGA
AAGCCCTANNNTTGNANAGCTCTTTCACCGGNNAAGANAATGACGGTNTCCGNNGANNAN
GCCCCGGCNNACTTCGTGNCNGCNCCCCNGNTNATACGAANGGGGCNAGNGTTNNTCGGA
ATTNCTNGGNGNAAANCNCACGTANGNGNACATTTNANNCACGGNTGAAATCCCNNNNNT
CANCTCTGGAACTGCCTTTGATACTGGGTGTCTGNANTATGGNANACGTGANTGGAATTC
CGAGTGTNNNNGTGANNTNCGTACATNTTCGGAGGAANACCAGTGGCGAATGNNGCTCAC
TNGTCCANNNCTGTNCGCTGACGTGCNAAANCCNGGGGAGCAAANCGGATTANATAACCC
TGGTNNNGCCACGCCNGTAAACGAATGANTGTTAGNNCNTCCGNNATTTTTACTTCCNCN
GGTNGCNCCACCATTNNNGCANTTNTACATTTCCCCCCTGGTGGGAGNTACTGTCCGCNN
NATTNAAANNTTCCAANGGANTTTNACGNGGANGCCCCCNCNNNAAGNGNTTGCACNCNA
TNTNGGTTTTAATTCCTAAGGCNACNTCNACNGAACCTTNNCCCAACCCCTTNNNNATCC
CNGGGNNNNCNTNNTNCAAAANGATATTGNGTCTCNNTCCTTNNNNCNGCCATAANAAGN
NCCTTTNATGNNNTTCTTANCACCNTCAAANTCTCTCAANGNNNNGGTTNNANTCCCNNC
CNNCCANCTTNACCCCCNCCCTCTTTTCNNTGNCACGNNTTTTGCNTTTGGTNNCCCNAT
CTGNGNTNCGCCCNTTNTTTTTTTTANCCCCNNACNGGNNGTGNGNNCNGCTCNCNTNNAA
TTNAACTCCCNTCCTNCNGNANAAA
>A36
ANNCGCNGCCTATCATGCAAGTCGAGCGCACCTTCGGGTGAGCGGCGGACGGGTGAGTAA
CGCGTGGGAACGTGCCCTTTGGTTCGGAACAACACAGGGAAACTTGTGCTAATACCGAAT
GTGCCCTTCGGGGGAAAGATTTATCGCCATTGGAGCGGCCCGCGTTGGATTAGGTAGTTG
GTGGGGTAAAGGCCTACCAAGCCTACGATCCATAGCTGGTCTGAGAGGATGATCAGCCAC
ACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTGCGCAA
TGGGCGAAAGCCTGACGCAGCCATGCCGCGTGTATGATGAAGGTCTTAGGATTGTAAAAT
ACTTTCACCGGGGAAGATAATGACGGTACCCGGAGAAGAAGCCCCGGCTAACTTCGTGCC
AGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGGGCG

CGTAGGCGGATATTTAAGTCGGGGGTGAAAGCCCAGGGCTCAACCCTGGAATTGCCTTCG
ATACTGGATATCTTGAGTTCGGGAGAGGTGAGTGGAATGCCGAGTGTAGAGGTGAAATTC
GTAGATATTCGGCGGAACACCAGTGGCGAAGGCGACTCACTGGCCCGATACTGACGCTGA
GGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGA
TGAGTGCTAGTTGTCGGCATGCATGCATGTCGGTGACGCAGCTAACGCATTAAGCACTCC
GCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGC
GGTGGAGCATGTGGTTTATTTCGAAGCAACGCGCANAACCTTACCACCTTTTGACATGCC
CTGACCGCCGGAAAAATCCGGTTTTCCCCTTCNGGGAACNGGNACACAGGTGCTGCATGG
GCTGTCGTCANCCTCGTGTCGTGAAAATGTTGGGNTTAAGTCCCNCAACNAAGCNCAACC
CTNCNCCATTANTTGCCATCATTAANTTGGGCCACTCTAAATGGGAACCCCCCGNTGGTA
AACCCGNAAGAAAGGNGGGNATNANNTTCAAATTCCCCNTGGNCCCTTNACGGGGTTGGN
CTTN
>A37
ANGCGCNGCTACCATGCAGTCGAACGCGTGTAGCAATACACGAGTGGCGCACGGGTGAGT
AACGCGTGGATATCTGCCTTTTGGTTCGGAATAACCCCGGGAAACTGGGGCTAATACCGG
ATGGTTCCTTCGGGATAAAGATTTATCGCCAAAAGATGAGTCCGCGTACGATTAGCTAGT
TGGTGAGGTAATGGCTCACCAAGGCGACGATCGTTAGCTGGTCTGAGAGGACGATCAGCC
ACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGAC
AATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTGTAAA
GCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAACTTCGTG
CCAGCAGCCGCGGTAAGACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCG
AGTGTAGGTGGTTGTCCAAGTTGGATGTGAAAGCCTTGAGCTCAACTCAAGAAATGCATT
CAGGACTGGGCGGCTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTGGTGAAAT
ACGTAGAGATTGGGAAGAACACCAGTGGCGAAGGCGGCTATCTGGACGGTTTCTGACACT
AAGACTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAAC
GATGGGTGCTAGACGTTGGCGAGCTTGCTCGTCAGTGTCGCAGCTAACGCGTTAAGCACC
CCCGCCTGGGGAGTACGGCCGCCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCCNCC
CAANGCGGTGGAGCATGTGGTTCAATTCGACGCAACGCGCAAGAANCNTTTACCAGCCCC
TTGACATGGGACNTCNNCNGGGAGCAGAAAACNCTCCCTTCCGGTTCCGGCCGGGAAGTC
CCCCACAAGGGGCCTNCATGGGCNTGTCGATCAGCCTCNTTGTCTTGAANATGTTTGGGN
TTTNAAGTNCCCNCAAACGNANNGCCAAACCCTCGTTTNTCCNGTTNGCCATTCAGGNTT
NTNCCTGGNGCNNCTTTNGGAAAAAACNTGCCGGGNGAAAAACCCGGANGNAAGGTTGGG
GANTNACCCTCCANTTNCCCNG
>A38
TCGCNGNCTATCATGCAGTTCGAACGGTAAATATTGTAGCAATACAATATGAGAGTGACG
TACGGGTGCGTAACACGTATGCAACCTACCCAAAACTGGAGTATAGCTCGGGGAAACTCG
AATTAACCCTCCATAAGATCGTGGTGTGGCATCACACAGCGATAAAAACTCAGGTGGTTT
TGGATGGGCATGCGTCTGATTAGCTAGTTGGCGGGGTAACGGCCCACCAAGGCGATGATC
AGTAGGGGAACTGAGAGGTTGATCCCCCACACTGGCACTGAGATACGGGCCAGACTCCTA
CGGGAGGCAGCAGTAGGGAATATTGGTCAATGGATGCAAGTCTGAACCAGCCATGCCGCG
TGCAGGAAGAAGGCCTTCTGGGTTGTAAACTGCTTTTGCCAGGGGATAAAATGGGCGTGC
GCGCCTAATTGAAGGTACCTGGTGAATAAGCCACGGCTAACTACGTGCCAGCAGCCGCGG
TAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGTT
CTTTAAGTCAGTGGTGAAATACTCTAGCTCAACTAGAGAGGTGCCATTGATACTGAGGAA
CTTGAGTGAAGTCGAAGTAGGCGGAATTGACGGTGTAGCGGTGAAATGCTTAGATATCGT
CAAGAACACCGATAGTGAAGACAGCTTACTAGGCTTATACTGACGCTGAGGCACGAAAGT
GTGGGGATCAAACAGGATTAGATACCCTGGTAGTCCACACTGTAAACGTTGATTACTCGC
TGTTGGCGATACACTGCCAGCGGCCCAGCGCAAGCGATAAGTAATCCACCTGGGGAGTAC
GCCGGCAACGGTGAAACTCAAAGGNAATTGACGGGGGTCCGCACAAGCGGTGGANCCATG
TGGTTTAANTTCGATGATACCCCNAGGAACCTTACCTGGACTANAAATGCCCCTTGACCC
GATCCANAAGATGGATTAATTTCCCCAAGGNACAAGGANCCAAGGGTGCTTGCCATGGCC
CGTTCNTTCANNCTCGTGCCCTTGAGGGNNTTGGGGTTANGTCCCCGCAACNNAGGCNCA
AACCCCCCNNTTTTCTTAGNTTNNCCNNCCCCNTTTTTGGTGGGGGNACTNTT
>A4

ATGCGCNCCTACCATGCAGTTCGACGAGACCTTCGGGTCTAGTGGCGCACGGGTGCGTAA
CGCGTGGGAATCTGCCCTTGGGTTCGGGATAACAGTTGGAAACGACTGCTAATACCGGAT
GATGTCTTCGGACCAAAGATTTATCGCCCAGGGATGAGCCCGCGTCGGATTAGCTAGTTG
GTGAGGTAAAAGCTCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGATCAGCCAC
ACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAA
TGGGCGAAAGCCTGATCCAGCAATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGC
TCTTTTACCCGGGAAGATAATGACTGTACCGGGAGAATAAGCCCCGGCTAACTCCGTGCC
AGCAGCCGCGGTAATACGGAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCG
CGTAGGCGGCTTTGCAAGTTAGAGGTGAAAGCCCGGAGCTTAACTCCGGAACTGCCTTTA
AAACTGCATCGCTAGAATCGTGGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTC
GTAGATATTCGGAAGAACACCAGTGGCGAAGGCGACTCACTGGACACGTATTGACGCTGA
GGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGA
TGATGACTAGCTGTCGGGGCTCATGGAGTTTCGGTGGCGCAGCTAACGCGTTAAGTCATC
CGCCTGGGGAGTACGGCCGCAAGGTTAAANCTCAAAGAAATTGACGGGGGCCTGCACAAG
CGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCANAANCCTTACCAGCGTTTGACATG
GTAGGACGGTTTCCAGAGATGGATTCCCTTCCCTTACGGGGACCTACACACAGGTGCTGC
CATGGCTGTCGTCACCTCGTGGTCGTGANAATGTTGGGTTTAAATCCNCCAACGAGCGCA
ANCCCTCGNCTTTAAGTTGCCTACCCTTTTAAGTTGGGGCACCTCTAAANAAACCTGCCC
GNTGATAANCCNGANGAANGTNGGGGANTAACCNCAATTCCTCATGGCCCTTNACCCCCT
TGGGTTCNCNNGNNCTNNCAATGGGCGGTTANATG
>A40
ATCGCGGCTATCATGCAGTCGAACGGTAAGTAGTGTAGCAATACATTGCCTAGAGTGACG
TAAGGGTGCGTAACACGTATGCAATCTGCCCTGTACAGGAGTATAGCTCCCCGAAAGGGG
AATTAACCCTCCATAGTATAATTGAATGGCATCATTTGATTATTAAAACTGAGGTGGTAC
AGGATGAGCATGCGTCTGATTAGCTAGTTGGTAGTGTAATGGACTACCAAGGCGATGATC
AGTAGGGGAACTGAGAGGTTGATCCCCCACACTGGCACTGAGATACGGGCCAGACTCCTA
CGGGAGGCAGCAGTAGGGAATATTGGTCAATGGGTGAGAGCCTGAACCAGCCATGCCGCG
TGCAGGAAGAAGGCCTTCTGGGTTGTAAACTGCTTTTGCCAGGGGATAAAACGGGAGTGC
GCTCCTAATTGAAGGTACCTGGTGAATAAGCCACGGCTAACTACGTGCCAGCAGCCGCGG
TAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGCT
CCTTAAGTCAGTGCTGAAATACTCTAGCTTAACTAGAGGGGTGGCATTGATACTGAGGAG
CTTGAGTAGAGTCGAGGTAGGCGGAATTGACGGTGTAGCGGTGAAATGCTTAGATATCGT
CAAGAACACCGATAGTGTAGACAGCTTACTAGGCTTCAACTGACGCTGAGGCACGAAAGT
GTGGGGATCAAACAGGATTAGATACCCTGGTAGTCCACACTGTAAACGTTGATTACTCGC
TGTTGGCGATATACAGTCAGCGGCTTAGCGCAAGCGATAAGTAATCCACCTGGGGAGTAC
GCCGGCAACGGTGAAACTCAAAGGAATTGACGGGGGGTCCGCACAAGCGGTGGAGCATGT
GGTTTAATTCGATGATACNCGAGGAACCTTACCTGGGCTAGAATGNCCCTTGATGTCCTC
ANAGACCAAGGANTTTCCCCAAGGGACAAGGANCNANGGGTGCTNCCATGGCTTTTCNTC
CAGCTCNTGNCCNTGAGGGTGTTTGGGGTTNANTCCCCNCANCCAGGCCCAACCNTTTTT
TTTTANTTGCCCACNNGGTTATGCTGGGGACCTCTNAAAA
>A41
GCNGCTACCATGCAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAACGCGTG
GGAATCTACCCATCTCTACGGAACAACTCCGGGAAACTGGAGCTAATACCGTATACGTCC
TTCGGGAGAAAGATTTATCGGAGATGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGG
TAATGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACACTGGG
ACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCG
AAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTC
ACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGG
CGGATTGTTAAGTTAGGGGTGAAATCCCAGGGCTCAACCCTGGAACTGCCTTTAATACTG
GCAATCTCGAGTCCGAGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGAT
ATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGCTCGGTACTGACGCTGAGGTGCG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGGAAG
CTAGCCGTCGGCAAGTTTACTTGTCGGTGGCGCAGCTAACGCATTAAGCTTCCCGCCTGG

GGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGA
GCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATCCCGGTCGC
GGAATACNAAAAGATCGTATTCTTCAGTTCGGCTGNACCGGTGACAGGTGCTGCATGGGC
TGTCGTCAGCTCGTGTCNTGANAATGTTNGGGTTAAATCCCGCAACGAGCGCAACCCTCG
CCCNTGTTGCCATCATTAAGTTGGGCCCTCTANGGGGGACTGCCGGTNAAAACCCAAAAG
AAAGGNGGGGATTACCTCAAGTCCTCATGGC
>A42
AGNNCGCTNAANCATGCAGTCGAGCGGGCGTAGCAATACGTCAGCGGCAGACGGGTGAGT
AACGCGTGGGAACGTACCTTTTGGTTCGGAACAACACAGGGAAACTTGTGCTAATACCGG
ATAAGCCCTTACGGGGAAAGATTTATCGCCGAAAGATCGGCCCGCGTCTGATTAGCTAGT
TGGTGAGGTAATGGCTCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGATCAGCC
ACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGAC
AATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAA
GCTCTTTTGTGCGGGAAGATAATGACGGTACCGCAAGAATAAGCCCCGGCTAACTTCGTG
CCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATCACTGGGCGTAAAGGG
TGCGTAGGCGGGTCTTTAAGTCAGGGGTGAAATCCTGGAGCTCAACTCCAGAACTGCCTT
TGATACTGAAGATCTTGAGTTCGGGAGAGGTGAGTGGAACTGCGAGTGTAGAGGTGAAAT
TCGTAGATATTCGCAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCGATACTGACGCT
GAGGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAAC
GATGAATGCCAGCCGTTAGTGGGTTTACTCACTAGTGGCGCAGCTAACGCTTTAAGCATT
CCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAA
GCGGTGGAGCATGTGGNTTTAATTTCGACGCAACGCGCANAACCTTACCAGNCCCTTGAC
ATCCCGGTCGCGGACTCCANAGATGGAGTTCTTTCAGTTCGGCTGGGACCCGGAAAACAA
GTTGCTGCATGGGCTGTTCNTCCANCTCGTGNTCGTGNANATNTTGGGGTTTAANTCCCN
CAAACGANGNNCAACCCCCCGTCCTTTAGTTGCTACCATTTNANTNNACCNCTNTTAAGG
AAAACTGCCCGGTTANNAANCCNCCAAGGAAAGGTGGGGGATTNNCNTCAAATCCCCCNG
>A43
CATNCGCNGCTACCATGCAGTTCGAACGCGTGTAGCAATACACGAGTGGCGCACGGGTGA
GTAACACGTGGATTATCTGCCTTTTGGTTCGGAATAACCCCGGGAAACTGGGGCTAATAC
CGGATGGTTCCTTCGGGATAAAGATTTATCGCCAAAAGATGAGTCCGCGTCCGATTAGCT
AGTTGGTGGGGTAATGGCCCACCAAGGCAACGATCGGTAGCTGGTCTGAGAGGACGATCA
GCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTG
GACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTGT
AAAGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAACTTC
GTGCCAGCAGCCGCGGTAAGACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAA
GCGCGTGTAGGCGGTTGTCCGAGTCGGGTGTGAAAGCCTTGAGCTCAACTCAAGAAATGC
ACTCGATACTGGATGACTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTGGTGA
AATACGTAGAGATTGGGAAGAACACCAGTGGCGAAGGCGGCTATCTGGACGGTTTCTGAC
GCTAAGACGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTA
AACGATGGGTGCTAGACGTTGGCGAGCTTGCTCGTCAGTGTCGCAGCTAACGCGTTAAGC
ACCCCGCCTGGGGAGTACGGCCCGCAAGGTTGAAACTCAAAGGAATTTGACGGGGGCCCG
CACAAGCGGTGGAGCATGTGGTTCAATTCGACGCAACNCGCAAAACCTTTACCAGCCCTT
GACATGGGACTCGCCGGGGACCAGAGACGGTCCCTTTCGGTTTCGGCCCGGAANTCCGCA
CAGGNTGCTTGCATGGGCTGTCGTCAGCCTCNTTGTNNNTNANAATNTTGGGGTTTAANT
CCCNCNAACGAAGNGCCAACCCTCGTTCTCCCAGTTNCCCATCAGGTTNTGCTGGGCCNC
TTTTGAAAAAACTGCCCGGTNAAAANCCCNGAAAGAAAGTTGGGGGAATTACCTTNAANT
NCCTCATG
>A44
ANNGCGNCTNATCATGCAAGTCGAACGGGATTCAATCAGTGGCAACACTGAGGAAGATCT
AGTGGCGAACGGGTGAGTATCACGTGAGGAACCTATCCCGGTCTCTGGAATAACAGGTGG
AAACACCTGCTAATACCGGATGCCGTCACCGTCTCACATGGGATGGTGACGAAAGATTCA
TCGGATCGGGCGGGCCTCGCGGCCTATCAGCTAGTTGGTGGGGTAACGGCCTACCAAGGC
ATCGACGGGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAG
ACTCCTACGGGAGGCAGCAGTAGGGAATCTTGCGCAATGGGCGAAAGCCTGACGCAGCAA

CGCCGCGTGCGGGAAGAAGGCTCTCGGGTTGTAAACCGCTTTCAGTAGGAACGAATCTGA
CGGTACCTGCAGAAGAAGGTGCGGCCAACTACGTGCCAGCAGCCGCGGTGACACGTAGGC
ACCAAGCGTTGTCCGGATTTATTGGGCGTAAAGAGCTCGTAGGCGGTTCAGTAAGTCGGG
TGTGAAAACTCAGGGCTCAACCCTGAGCGGCCACCCGATACTGCTGTGACTTGAGTCCGG
TAGGGGAGTGGGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCG
GTGGCGAAGGCGCCACTCTGGGCCGAGACTGACGCTGAGGAGCGAAAGCGTGGGTAGCAA
ACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGTTGGGCACTAGGTGTGGGTTCCA
ACCAACGGAAATCCGTGCCGTCGCTAACGCATTAAGTGCCCCCGCCTGGGGAGTACGGCC
GCAAGGCTAAAACTCAAAGGAATTTGACNGGGGGCCCGCACAAGCAGCGGNANCGTGTTG
CTTAATTCCNATGATACNCGAAGAAACCTTTCCCTGGGGTTTNGACATGTACGGGAAAAG
CCGCAAAANATGCCGGTNTTNCNTTNGGGNTCNTACCACAGGTGGGTTGCATGGNCTNTC
NTNCNCCNCNTTTTCCTTGAAAATTTTTGGGGTTTAAGTTNCCCCCAAACGAGGNCNAAC
CCCTTTNTTCTTTTGTTTCCCANNCGGAATNATNNCCGGGGAACNTCTCCAGGAAAATTT
CCNGGGGGCCCACN
>A5
GGGAGGCGCGGCTTACCANNGCAAGTCGANCGGNCACCTTCGGGTGANTNGGTGAGCACG
ACGTNGAGTAANAACGCGGTGCAATNANACNACACTCTTCCTACGTNCAACAACNTCNCC
NGNNAACATGCGTATGCGTACATAACNACNGNANTANCCCNTCTTCGGAGGAAANAATTA
TTTTCAGTCAGCACTGATATGAATCGCCNCCNGTTGTCATGTATTATACGTATGGTNGGT
GGANAATGAGTGCGNATCCACACGGANGTACGAACTCACTCTAGTATGCGTNCGTGCATG
GAGGAATGAATTCATCCCCACCCACTTGTGTGNACGTGCATGGACNACGAGCCCNCACCA
CNTCCTNNCTACGNAGGANGGGCNAGCTGGTGNGAAATAATTGNTNCAACTGATGCNGGN
AAGACCTNNATNCATCCCCACNGATNGCCGNNGATGTGGNTGANNGANCCCCNATTGGNN
TGTTNATANNTCTTTTTNCGNCCGAAGAATAATGNCACTGTCCCGAGAAAAATACCCCCN
NNAACTTNTNGCNCCCACNCCCGCNAATACANAAGGGGCTNGCGNTGTTCTCAATTACTN
TGCGTAAAGAGCNCGNAGTACGNATATTTCNTTCATGGGGGGAAATANGCCNGGNTCATC
NACCGCTAACATNNTNTTTGTNTANTGCTGTATNGTCTAGTCCTTAGAGAAGGATGNNTG
GGAATTACTGAGTGTAGTAGGTGGAAATTNTTCAGTATNATTCTTCAGCGAACAACCACC
TGGTGGAAGNACGGCCTCACNTGCTGTNTCCTACTNCNCGACTGCAGATGCGNAANANAG
TGGTGGAGCAGCACAACGATTATTATACCACTGGGNGTNCCNCCGCCTNNANACAANTGN
TAANTTACCCCNNTCGTTGNAGATNTACTTNTCTGNNGGNGGCNCCTNANACATTTTNGA
NTTCNNNCCCTNCTGGAGTANNANTCGCCNCCATTNTNNNACTCCNCAANGANTTTTTCN
NGGGGGGCCCCCNACAANANGGNNGGGAGCANNGTGGTTTTTTNTATTTTAAAANAACCA
NCGNCNGAAANCCNTTTTCNNCCCCCCTTTTTCAATCNNGTNCCGNNGTTTTTCCCAACA
AAANGAGNANCCTTNNTTCTTNTGNNNTNGTGCCCCGGGAAACACAGGGGNNCCCCNTGG
GGTNNTTNNTCAACCATCTTNGTTCTNCGTAAAATATTTGTGGGTTTANAATCCCCCAAN
AAAGNNGCNAAACCCCTCTCCCCCCANTTNNNCCCNNTTNTTNNTATTGGGGNCANANNC
TNTGGGGGANNCCNCCNNTGTANTAAACCCNCACCNNAANANGGGGGGNGTNNAACNTNN
AANACCTCTTTGGGCCCTTTTNAGGGGGGGGTNNNANCCCACTNTTNNNANAAGTGGGGG
GAAAANTTNNCCACNN
>A58
ATGCGCGCTANCATGCAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAACGC
GTGGGAATCTACCCATCTCTACGGAATAACTCAGGGAAACTTGTGCTAATACCGTATACG
CCCTTTTGGGGAAAGATTTATCGGAGATGGATGAGCCCGCGTTGGATTAGCTAGTTGGTG
GGGTAAAGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATT
GGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGG
GCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTCT
TTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGC
AGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGT
AGGCGGATCGATCAGTCAGGGGTGAAATCCCAGAGCTCAACTCTGGAACTGCCTTTGATA
CTGTCGATCTAGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTA
GATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTACTGACGCTGAGGT
GCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGA
ATGTTAGCCGTCGGGCAGTTTACTGTTCGGTGGCGCAGCTAACGCATTAAACATTCCGCC

TGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGGAATTGACGGGGGCCCGCACAAGCGG
TGGAGCATGTGGTTTTAATTCGAAGCCAACGCGCCANNAACCTTACCAGCTCCTTGACAT
CCCGATCGCGGGAACANTGGAAGACATTGTCCCTTCAGTTTAGGCTGGGATCGGTGAACA
GGGTGCTNCCATGGCNTTTCCNNCNACCCTCGNGTNCNGGAAATTNTTTGGGTTTAANTT
CCCNCAAACGAGGCGCNANNCCCTTCGCCCCTTTANTTGCCCNNCCNTTCANTTTGGGNC
CNCTTTTAAGGGGGAACNTGCCCGGNTGAATAANCCCAAAANGNAAAGGGGGGGNATNAC
CTTC
>A6
GAGGGGGGCTAAATGCAGTCGAACGNNGACACNCTNANGNNGTGNATTATCTGTGNNAAN
GGAGGGAGCNCCNTTNTGGGCNANCTGCCCTTCATTCTGGGACAAGCNCTGGAAACNGGN
TNNAANNCCGGATATGACACACNACCGCATGGTCTGTGNGNGNAANGCTCCTNCGGAGAN
TGATGANCCCGCNGCCTATCAGCTTGNTGGNGNGGTGATGGCCTACCAANGCNACCACGG
GTANCCNGCCTGANAGGGCGACCGGNCACACTGGGACTGAGANNCGGNCCACACTCCTAC
NGGANGCTANCAGGNGGNGAATATTGCACAATGGGCGAAANGCTGATGCATCNTCGCCAC
NTGAGGGATGACNGCCTNCNGGTTGTNAACCTCTTNNAGCATGNNANAAGCGAAANNGAC
GGTACCTGCATAAAAAANCGTCGGCTANCTACNTGCCAGTCANCCGCGGTNATACTTAAN
NGNGCGAGCGTTGNCCNGAATTNTNTGGGNGTGAAAGAGCTCCTAANGCGGNCTTGTCAC
GNCNTNATGTGNAAGNCCCGNGTNTNTAACNCCTGGNTCTGCGTTCTANACNCNGCAGGC
TANAGTTCTNGTAGGGGAANATCGNCAATTCCATGNNTGTACCNNNNGCANATGCNCCAG
ATTTCACCNNGGANACANCNNNANTNNNCNNNGCNNTATNCTCTGGGTTCCNTANANTGT
ACNCNTNNANNGACCTTNTNNNCNANTNAGTCAATCNNAANCTTTCATTANNNAANNCCT
GGNTNGNCCCANNCNCNTNAATNGTTTNGTCNTCNANNGTTNTTNGGCNCANNNTTCCTA
CGTTCGANNCGNGGCCNNTNAGNAATNNTAANTTTACTTTCTCCNTNCCCNGTTNAANTC
NNGCCCNNNATCGCTTTNAAACTTCCNAGNNAATTCTCACTGNNGTNCACNNTCNAACNN
NNNNGNNACNTNGTTNTCTTANTTCTCNCCATCNCTNNTGATNANCNNTTANCNNNNCCC
NTNCNNTNNCTCTNGNAANATCCTTCNNNNACNGGGNCNCCCNTCTTTCGGNNCTNTCTN
NNNATNNGNNNCGNGNCCTTCGCTCNNCTCCNNTNCNTGNTNTATNTTTGGNCCANNCTT
NCTNTTTCNGAANCTAANCNNTNNNNTCCNNCANNNCTNNCANNCTTCTTANGTCGNTNN
NCGNATCCCTCCCTANTNNCTCGTNNGCNNANCNCNGTAAATTCATNTTATATNACNTTN
NTNTATNTGCTCTTTNTTNNCNCGTTTNGTTTCCNTATNACAG
>A60
ATGCGCNNCTANCATGCAGTCGACGATGATCCCAGCTTGCTGGGGGATTAGTGGCGAACG
GGTGAGTAACACGCTGAGTAACCTGCCCTTGACTCTGGGATAAGCCTGGGAAACTGGGTC
TAATACCGGATATGACTGTCTGACGCATGTCAGGTGGTGGAAAGCTTTTGTGGTTTTGGA
TGGACTCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCNACNACNGGTA
GCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAG
GGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAAGCGAAAGTGACGGTAC
CTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAG
CGTTATCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAA
AGACCGGGGCTCAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGA
GACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGA
AGGCAGGTCTCTGGGCTGTAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGAT
TAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACG
TTTTCCGCGCCGTAGCTAACGCATTAAGTGCCCCGCCCTGGGGAGTACGGCCGCNAGGCT
AAAACTCAAAGGAATTGACGGGGGCCCGCACAAAGCGGCGGNAGCATGCCGGATTANTTC
GATTGCAACCCGGAAGAACCNTTACCNAAAGGCTTTGAACATGAAACCNGNTNATACCCT
TGGGAAAACANGTTGCCCCCGCNTTGCCGGTNCNGNTTTACAAGGTGGGGCCATNGGGTT
GTTCCCTCCGCCNCGNGGTCGTGGAAAATTTTTGGGGTTTAANTTCCCCNCCAACCGAGN
NGCCAACCCCTTCGTTTTTNTGTTTNCCCANCCCCGTTTTTTGNCCGGGNAACNCNATA
>A9
NANGNGNNNCATAACATGCAAGTCGACGATGAAGCCCTTCGGGGTGGATTAGTGGCGAAC
GGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTC
TAATACCGGATACCACTCCTGCCTGCATGGGCGGGGGTTGAAAGCTCCGGCGGTGAAGGA

TGAGCCCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCCACCAAGGCGACGACGGGTA
GCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAG
GGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGACGGTAC
CTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAG
CGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCGCGTCGGATGTGAA
AGCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCAGGCTAGAGTGTGGTAGGGGA
GATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGA
AGGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGAT
TAGATACCCTGGTAGTCCACGCCGTAAACGTTGGGAACTAGGTGTTGGCGACATTCCACG
TCGTCNGTTGCCGCAGCTAACGCATTAANTTTCCCCNCCCTGGGGGANTTACGGCCGCCA
GGGCTAAAACTCAAAAGGAATTGACGGGGGCCCNGCACAAGCANCGGAGCATGTGGCTTA
AATCGAACGCAACNCGAAGAAACCTTTACCAAGGGCTTGAACATTNTACCGGGAAAGCAT
TAAAAATAAGTTCCCCCCCTTGTNNGNCGGTNNTACAGGGTGGTGCNNNGGCTTTCTTNC
ACNTCGNNGTNCCTGAAATNTTTNGGNTTAATTTCCCCCAACNANANGCAACCCCTTNTT
CTNTNGTTNTCACCCNNGCCTTTCCGGGGNNNTTTGGGANTNNACATNGAAACCCCCNGG
GNNCAAANNTNNANNAAAGGTGGGGAACAAACNTNAATTNTCTTTCCCCTTTTTNTTGGG
NTTCAACN
>CR-1.27F Trace of Y:\Nasrin\Sequencing from Lucigen\CR3\27F\CR-1.27F.ab1;
length: 1179; ambiguous 29; low quality 1179; medium quality 0; high
quality 0
ATGCGCGGCTATCATGCAGTCGAACGGCTCTTCGGAGCAGTGGCGGACGGGTTGAGTTAA
CGCGTGGGAACGTGCCCAAAGGTACGGAACAACTGAGGGAAACTTCAGCTAATACCGTAT
GTGCCCTTAGGGGGAAAGATTTATCGCCTTTGGAGCGGCCCGCGTTGGATTAGCTAGTTG
GTGGGGTAAAGGCCTACCAAGGCTACGATCCATAGCTGGTCTGAGAGGATGATCAGCCAC
ACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTGCGCAA
TGGGCGAAAGCCTGACGCAGCCATGCCGCGTGAATGATGAAGGTCTTAGGATTGTAAAAT
TCTTTCACCGGGGAAGATAATGACGGTACCCGGAGAAGAAGCCCCGGCTAACTTCGTGCC
AGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGGGCG
CGTAGGCGGATATCTAAGTCGGGGGTGAAAGCCCGGGGCTCAACCTCGGAATTGCCTTCG
ATACTGGGTATCTTGAGTACGGGAGAGGTGAGTGGAACTCCGAGTGTAGAGGTGAAATTC
GTAGATATTCGGAAGAACACCAGTGGCGAAGGCGACTCACTGGCCCGTTACTGACGCTGA
GGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGA
TGAGTGCTAGTTGTCGGCATGCATGCATGTCGGTGACGCAGCTAACGCATTAAGCACTCC
GCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAG
CGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAANAACCTTACCACCTTTTGACATGC
CCTGATCGCCACANAGATGTGGTTTTCCNTTCGGGGACAGGGACACAGGTGCTGCATGGC
TGTCGTCAGCTTCGTGTCNNGANATGTTGGGTTAANTNCCNCACCGAGCGCAACCCTCNC
ATTTAGTTGCCATCATTAANTTGGGCACTCTAATGGGACCNCCGGTGTTAACCCGAAGNA
ANGNGGGGATNACCTCAATTCTNTNGNCCTTTCGGGGTGGNCTNNCCCNTTNTTCATGGG
CACCNCAAGGNTTCNAACTTGCNAAGGGANCTTTCCCTA
>CR10
TGGNAGGNGCCGCATANNATGCAGCTCGAGCGGGGATTNNTTANATAGCTTGCTTCTAGT
AANCTAGCGGCGNCNCNTGGANGAGTAACACNTAGGCAACCGTGCCCNCAAGACAGGGCA
TAACTACCGGAAACGGTAGCTAATACCCGATACATCCTTTTCCTGCATGGGNGAAGGAGG
AAAGGCGGAGCAATCTGTCACTTGTGGATGGGCCTGCGGCGCATTANCTAGNTGGTGGGG
NAAAGGCCTACCAAGGCGACNATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGG
ACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCCGTNGGGAATCTTCCGCAATGGGCN
AAAGCCTGACNTGAGCAACGCCNCNTGAGTGNTGAANGTTTTCGNATCNTNAAGCNCTGT
TGNCNGGGAANAACGTCTNGCAGAGTAACTGCTNTAAGAGTGACGGTNCCTNATANNANA
TCCCCGGCTNACTACTTGTCCNNNAATCGNNGTTCTNCNTNCGGGNTCANNTGNTTNNTN
CGNNANTTATTTGNNGTGNTNNNTCNTCGNNANAGGNTGNNTNTTTAANCTCCTNGNGTT
NNANTTTNNCNANGCTTNTTNTTCCTGGTCTTCGNCTNGACATANTNTGNAGGCAGCTTT
GNNNTNCNNNTNTNACNANNNTNTNTATANCTCNNNNGTTGNNTAGCCGTNTTCTNAATT

NCNTNATNCTNTTTTCTNATTNNCTCTCCCNTNTNTCCCCATTGNTNTTTCCTNCTNTNN
NTCNCTCANNGCNCNTTCCTTNTNCTCNTCNTCCNCCTNTTTGCANTNNCANACNNTANG
CNCTNGTTNNTTCCCTCTNCTCTTATCNCCTTTCTTTTTCNTCTATNTTCNCNCTATATT
TNNCNTNNTCANTATCACCNCCCTCTNTTNCCNNTNCTTTCTNGTNACNTTANANNTNTC
ANTTNNTNNTGNNATCTNNNTTNTTTANTCNCCCGNNTNTTTTNTTTTCTACTTGCGTNA
TATTTNTNNTTNCCNNNTTNNGNCCTNTTCGNANACTCTTNNCTCACCTATCTCTNTTNT
NNCCTTTNCTCNTNTCCNTCTTNCTTTCTTTTNATTCNTNNCNNTTATNNNNCTNTTTTN
TTTTCCCTTNNNTTNNNNCTTCATTNNNTCTNCTNTTTTTTTATTCTNCCGNANNNNTCN
TTACTNCNTCNNCGTCTCGTCCTNTNCTACTTNNNTCTTTCTCTNTTANTATNNTTCTTN
TNTNTCTTNNTGTNCNNTTACTCNTACNACCTCTCNNNTNTCCTCNTTNCTNCTCTCTTC
GCNCTNTTTTTTACTTTCCNTG
>CR100
ATGCGCNCTTACCATGCAAGTCGAACGGCAGCATGAGGTGTAGCAATACACCTTGATGGC
GAGTGGCGGACGGGTGAGGAATGCATCGGAATCTGCCCAGTCGTGGGGGACAACGCAGGG
AAACTTGCGCTAATACCGCATACGACCTTCGGGTGAAAGCAGGGGATCTTCGGACCTTGC
GCGATTGGATGAGCCGATGCCGGATTAGCTAGTTGGCGGGGTAATGGCCCACCAAGGCGA
CGATCCGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGAC
TCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGCAACCCTGATCCAGCAATG
CCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCGGGAACGAAACATTGT
CGGCTAATACCCGGCAAGACTGACGGTACCCGAGGAATAAGCACCGGCTAACTCCGTGCC
AGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTG
CGCAGGCGGTTTGTTAAGTCTGCTGTGAAATCCCCGAGCTCAACTTGGGAATTGCAGTGG
ATACTGGCAAGCTGGAGTACGGTAGAGGAAGGTGGAATTCCCGGTGTAGCAGTGAAATGC
GTAGAGATCGGGAGGAACACCAGTGGCGAAGGCGGCCTTCTGGACCAGTACTGACGCTCA
TGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGA
TGCGAACTGGATGTTGGGTACATTACGGTACTCAGTGTCGAAGCTAACGCGTTAAGTTCG
CCGCCNGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGNANTTGACNGGGGGCCCGCAC
AAGCGGTGGAAGTATGTGGNTTTTANTTCGATGCAACGCGAANAACCTTTACCTGGCCCT
TGACATCTGTTCNAATCCNTGCAAANATNCGGGAATTGCCCCCANGGAAACNACAANACA
GGTGCCTNCANTGGCTTTCNTCCANCTNCNTGTCNTGAAATTTTTGGGTTTAAATNCCCC
CCAACCNANNGCAANCCCTTTGTTCCTTATTNCCANCCCCTTAANGGNTGGGNAATTTTT
GGGAAANCNCCCNGNNNAAAACCNAAANAAAGGTGGGGATAACNTCAAATTCTCAGGGNC
CTTTCCGNC
>CR101
ANGNGCNCTTANCATGCAGTCGAACGGCAGCATGAGGTGTAGCAATACACCTTGATGGCG
AGTGGCGGACGGGTGAGGAATGCATCGGAATCTGCCCAGTCGTGGGGGACAACGCAGGGA
AACTTGCGCTAATACCGCATACGACCTTCGGGTGAAAGCAGGGGATCTTCGGACCTTGCG
CGATTGGATGAGCCGATGCCGGATTAGCTAGTTGGCGGGGTAATGGCCCACCAAGGCGAC
GATCCGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACT
CCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGCAACCCTGATCCAGCAATGC
CGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCGGGAACGAAACATTGTC
GGCTAATACCCGGCAAGACTGACGGTACCCGAGGAATAAGCACCGGCTAACTCCGTGCCA
GCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGC
GCAGGCGGTTTGTTAAGTCTGCTGTGAAATCCCCGAGCTCAACTTGGGAATTGCAGTGGA
TACTGGCAAGCTGGAGTACGGTAGAGGAAGGTGGAATTCCCGGTGTAGCAGTGAAATGCG
TAGAGATCGGGAGGAACACCAGTGGCGAAGGCGGCCTTCTGGACCAGTACTGACGCTCAT
GCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGAT
GCGAACTGGATGTTGGGTACATTACGGTACTCAGTGTCGAAGCTAACGCGTTAAGTTCGC
CGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGGCCCGCACAA
GCGGTGGAGTATGTGGTTTAATTCGATGCAACGCGAAGANCCTTACCTGGCCTTGACATC
TGTCGAATCCTGCAAANATGCGGGAGTGCCCCANGGACNNANAAAACAGGTGCTGCATGG
CTTTCGTCANCTCGTGTCGTGAAAATNTTGGGTTNAANTCCCNCANCGAGCGCCACCCNT
TTTNCCTNATTTGCCANCCCGTAATGGGTGGGAACTCTNGGGAAACTGCCGGTNACAACC
CGNAGGAAGGGGGGGGANAACCTCAANTCCCNANGGCCNTTANGG

>CR102
TGCGCNCTTTANCATGCAAGTCGAACGGCAGCATGAGGTGTAGCAATACACCTTGATGGC
GAGTGGCGGACGGGTGAGGAATGCATCGGAATCTGCCCAGTCGTGGGGGACAACGCAGGG
AAACTTGCGCTAATACCGCATACGACCTTCGGGTGAAAGCAGGGGATCTTCGGACCTTGC
GCGATTGGATGAGCCGATGCCGGATTAGCTAGTTGGCGGGGTAATGGCCCACCAAGGCGA
CGATCCGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGAC
TCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGCAACCCTGATCCAGCAATG
CCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCGGGAACGAAACATTGT
CGGCTAATACCCGGCAAGACTGACGGTACCCGAGGAATAAGCACCGGCTAACTCCGTGCC
AGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTG
CGCAGGCGGTTTGTTAAGTCTGCTGTGAAATCCCCGAGCTCAACTTGGGAATTGCAGTGG
ATACTGGCAAGCTGGAGTACGGTAGAGGAAGGTGGAATTCCCGGTGTAGCAGTGAAATGC
GTAGAGATCGGGAGGAACACCAGTGGCGAAGGCGGCCTTCTGGACCAGTACTGACGCTCA
TGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGA
TGCGAACTGGATGTTGGGTACATTACGGTACTCAGTGTCGAAGCTAACGCGTTNAAGTTC
GCCGCCTGGGGAGTACNGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGGCCCGCAC
AAGCGGTGGAGTATGTGGNTTTATTTCGATGCANCCCGAANAACCTTACCTGGCCTTGAC
ATCTGTCGAATCNTGCAAANATNCNGGAGTGCCCCANGGAACGACAAGACAGGTGCTGCA
TGGCTNTCNTCACCTCGTGTCCTGAAAATNTTNGGGTTAATTCCCCCAACNAACNCAACC
CTTGTTCCTTATTTNCCANCACNTANTGNTGGGAANCCTTNGGGAAACTCCCGNTTACAA
ACCGNAGAAAGGGGGGGATAACTTAANTCCTCAGGNCCTTTNNGCCNGGGTTCN
>CR104
TGCGANCTTTNCATGCNAGTCGAACGGCAGCACGGGGGCAACCCTGGTGGCGAGTGGCGA
ACGGGTGAGTAATACATCGGAACGTGCCCAGTCGTGGGGGATAACTACGCGAAAGCGTAG
CTAATACCGCATACGATCCACGGATGAAAGCGGGGGACCGCAAGGCCTCGCGCGATTGGA
GCGGCCGATGGCAGATTAGGTAGTTGGTGGGGTAAAGGCCTACCAAGCCTGCGATCTGTA
GCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCCATTCCGCGTGCA
GGATGAAGGCCCTCGGGTTGTAAACTGCTTTTGTACGGAACGAAACGGCGACTTCTAATA
CAGGTCGCTAATGACGGTACCGTAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGC
GGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCGG
TGATGTAAGACAGGTGTGAAATCCCGGGGCTCAACCTCGGAACTGCATTTGTGACTGCAT
CGCTGGAGTGCGGCAGAGGGGGATGGAATTCCGCGTGTAGCAGTGAAATGCGTAGATATG
CGGAGGAACACCGATGGCGAAGGCAATCCCCTGGGCCTGCACTGACGCTCATGCACGAAA
GCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAACTG
GTTGTTGGTCCTTCACTGGATCAGTAACGAAGCTNACGCGTGAAGTTGACCGCCTGGGGA
GTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGNACCCGCACAAGCNGTTGGAT
GATGTGNTTTAATTCNATGCAACGCGAAAAACCTTACCCACCTTTNACATGCTTGGAACC
TCNCAAAAATTTGANGGGTGCCCNAAAAGGGAACCAGGACACAGGGTNCTTCNATGGCTT
TCCTTCANCTCNTGTCNTGAAATTTTTGGTTTAATTCCCCNACAAGNNCAACCNTTNGTC
ATNATTGCTANATTCNNTTGGGNCCCTNTAATGAAAACTTCCGGTAAAAACCCNAANAAA
GNGGGAATAACTNCAATCCCCANGGCCCTTNNAGGNGGGNNTCCCCCCTTNNAAA
>CR105
TGCGCNCCTACCATGCAAGTCGAACGGCAGCATGAGGTGTAGCAATACACCTTGATGGCG
AGTGGCGGACGGGTGAGGAATGCATCGGAATCTGCCCAGTCGTGGGGGACAACGCAGGGA
AACTTGCGCTAATACCGCATACGACCTTCGGGTGAAAGCAGGGGATCTTTGGACCTTGCG
CGATTGGATGAGCCGATGCCGGATTAGCTAGTTGGCGGGGTAATGGCCCACCAAGGCGAC
GATCCGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACT
CCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGCAACCCTGATCCAGCAATGC
CGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCGGGAACGAAACATTGTC
GGCTAATACCCGGCAAGACTGACGGTACCCGAGGAATAAGCACCGGCTAACTCCGTGCCA
GCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGC
GCAGGCGGTTTGTTAAGTCTGCTGTGAAATCCCCGAGCTCAACTTGGGAATTGCAGTGGA
TACTGGCAAGCTGGAGTACGGTAGAGGAAGGTGGAATTCCCGGTGTAGCAGTGAAATGCG

TAGAGATCGGGAGGAACACCAGTGGCGAAGGCGGCCTTCTGGACCAGTACTGACGCTCAT
GCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGAT
GCGAACTGGATGTTGGGTACATTACGGTACTCAGTGTCGAAGCTAACGCGTTAAAGTTCG
CCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAA
GCGGTGGAGTATGTGGTTTAATTCNATGCANCGCGAANAACCTTACCTGGCCTTGANATC
TGTCGAATCCTGCAAANATCCGGGAGTGCCNCAAGGNANCNANAANACGGGNCCTGCATG
GCTTTCGTCANCTNCGTNTCNTGAAATNTTGGGTTNAATTCCCNCAACCNAGGGCAACCC
TTGTCCCTNATTTNCCACCCCNTAATGGTNGGGNAACNTTGGGGAACTCCCCGGTNTAAA
ACCCGAAGAAAGGNGGGNATAACTNAAATTNCTTAGGGCCCTTNCGGCNNGGGTTCC
>CR106
NANCGCGNCTAACATGCAGTCGAGCGAGGCCCCACCTTCGGGTGGGTGTCCTAGCGGCGA
ACGGGTGAGTAACACGTGGGCAACCTGCCCCTAGCACTGGGATAACCCCGGGAAACCGGG
GCTAATACCGGATACGACCTCGAAGGGCATCCTCCGAGGTGGAAAGGGTTACTGGCTAGG
GATGGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTAACGGCCTACCAAGGCGACGACGGG
TAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTG
AGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCTCCGACGAAGCGAAAGTGACGGT
AGGAGCAGAAGAAGCACCGGCCAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCA
AGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTCTGTCGCGTCGGCTGTG
AAAACCTGGGGCTCAACTCCGGGCCTGCAGNCGATACGGGCAGACTATAATTCGGTNGGG
GAGACTGNAATTCCTGGTGTAGCNGTGAAATGCTCANATNTCANGAGGAACACCNGTGGC
GAANGCGGGTCTCTGGGGCCGATATTGNCTCTNANGAGCTAAAGCNTGGNNANCGAANAN
GATTAGATACCCCTGGTAATTCCACTNCCNTAAACTTTGNNCGCTAAGTGTGNGCTCCTT
TCCNCNNAATNTTGTGCCTNATTAACTNNATNANCNTCCCNTTTTGGGGAGTACTTNCTC
TAGGTCTAANAATTTANANTAATTTTTACNGGGGNTCCCCCNTNANTCNTGTCNNATNAT
TGTNTCTTAATNCTNTTTAACCNANTAAANCCTTTTCCTTNNNCTTTANTTTTATNGGAA
ANTCTTTTNNTNAANTATNNTCTCNNTAATTTCNTTTCNNTATNNGNTTTATTNNTTTTT
CTTNNTTCTCTTTNTTTNACATNTTTTNTTTTATTTCNNTNNNTNTTACNNNCCCNTNTN
NNANTNTTCTCNCATNTTNNNNTTTCTANTTCTTTATCTTTNTTNTGTNGTNTNTTCNAN
ANNANANNNGTNNATNTTNNTNNNTTATTTNTCNATTTCTTCANNCCTTTTATNNNTNNN
ATN
>CR12
ATGCGCNCCTACCATGCAAGTCGAACGAGACCTTCGGGTCTAGTGGCGCACGGGTGCGTA
ACGCGTGGGAATCTGCCCTCTGGTACGGAATAACTCAGGGAAACTTGAGCTAATACCGTA
TAATGACTTCGGTCCAAAGATTTATCGCCTGAGGATGAGCCCGCGTCGGATTAGCTAGTT
GGTGAGGTAAAAGCTCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGATCAGCCA
CACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACA
ATGGGCGAAAGCCTGATCCAGCAATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAG
CTCTTTTACCCGGGATGATAATGACAGTACCGGGAGAATAAGCCCCGGCTAACTCCGTGC
CAGCAGCCGCGGTAATACGGAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGC
ACGTAGGCGGCTTTGTAAGTTAGAGGTGAAAGCCCGGGGCTCAACTCCGGAATTGCCTTT
AAGACTGCATCGCTAGAATTGTGGAGAGGTAAGTGGAATTCCGAGTGTAGAGGTGAAATT
CGTAGATATTCGGAAGAACACCAGTGGCGAAGGCGACTTACTGGACACATATTGACGCTG
AGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTTAAAC
GATGATGACTAGCTGTCTGGGCGCTTAGCGTTCAGGTGGCGCAGCTTACGCGTTAAGTCA
TCCGCCTGGGGAGTACGGCCGCAANGTTAAAACTCAAAGAANTTGACNGGGGNCCTGCAC
AAGCGGTGGAGCATTGTGGTTTANTTCGAAGCAAACGCGCAAAACCTTACCANCCNTTTG
ACNTGGTAGGACGGTTTCCANANATGGATTCNTTCCCTTTCCGGGACCCTACCCCCAGTT
NCTGCAAGGNTTTTCTCNCCTCTTGTCGTNGAATTNTTGGGNTTAANTNCCCCNACCANN
CGCNACCCTCGTNTTTATTNNCTACCATTTANTTGGGCNCTTCTTAAAAAAACTNCCCGN
TAATAACCCGAAGNAAGGNTNGGNNTTANTTTAATTCNTCTGGCCCTTTATCCNCCTGGN
CTTCCCNCNTNGTTTNATGGGNGTTANNTNGGNNTTCNAA
>CR13
ATGCGCNCCTACCATGCAAGTTCGAACGAGACCTTCGGGGTCTAGTGGCGCACGGGTGCG

TAACGCGTGNGAATCTGCCCTCTGGTACGGAATAACTCAGGGAAACTTGAGCTAATACCG
TATAATGACTTCGGTCCAAAGATTTATCGCCTGAGGATGAGCCCGCGTCGGATTAGCTAG
TTGGTGAGGTAAAAGCTCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGATCAGC
CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGA
CAATGGGCGAAAGCCTGATCCAGCAATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAA
AGCTCTTTTACCCGGGATGATAATGACAGTACCGGGAGAATAAGCCCCGGCTAACTCCGT
GCCAGCAGCCGCGGTAATACGGAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGC
GCACGTAGGCGGCTTTGTAAGTTAGAGGTGAAAGCCCGGGGCTCAACTCCGGAATTGCCT
TTAAGACTGCATCGCTAGAATTGTGGAGAGGTAAGTGGAATTCCGAGTGTAGAGGTGAAA
TTCGTAGATATTCGGAAGAACACCAGTGGCGAAGGCGACTTACTGGACACATATTGACGC
TNAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAA
CGATGATGACTAGCTGTCTGGGCGCTTAGCGTTCAGGTGGCGCAGCTAACGCGTTAAGTC
ATCCGCCTGGGGGAGTACGGCCGCAAGGTTAAAAACTCAAAGAAATTGACGGGGGCCTGC
ACAAGCGGTGGAAGCATNTGGTTTAATTTNNAAGCACCGCGCAAAAACCTTACCAGCGTT
TTGACATGGNTAGGACGGTTTCCCAGAGATGGATTTCCTTNCCTTTACGGGGACCTACAC
ANAGGTGCCTGCATGGCTTTCNTCACTNNTTGTCGTGANATTTTGGGTTTANTTCCCNCA
AACNAGNGCAACCCTNNTTCTTTTGTTTCNTCCATTTTANTTGGGCACTNTAAANAACCT
CCCCGTTAATANCCNGAGGAAGGGGGGGATTANCTTAAATTCNTNNGGGCCCTTNNCCNT
TGGGGTTCNNNNTTNTACNATGGGGGGNANAANGGNTTNCAANCCCCCGGGGNTNNATTT
NTTTCNAAAACCNNNTNNTTN
>CR14
NATGNGCGNCTAACATGCAAGTCGAACGCTGAACTCCGCTTGCGGAGGGATGAGTGGCGA
ACGGGTGAGTAACACGTGGGCAACCTGCCCCCGGCTCTGGGATAACTCCAAGAAATTGGG
GCTAATACCGGATATTCACTGCTTCTCGCATGGGGGGTGGTGGAAAGTTTTTCGGCTGGG
GATGGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGATGGCCTACCAAGGCGACGACGGG
TAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
GGAGGCAGCAGTGGGGAATATTGCGCAATGGGCGGAAGCCTGACGCAGCGACGCCGCGTG
GGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGACGAAGCCTTTGGTGACGG
TACCTACAGAAGAAGCACCGGCCAACTACNTGCCAGCAGCCGCGGTAATACGTAGGGTGC
GAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCNTAGGCGGTCTGTCGCGTCNGCTGT
GAAAACCCGGAGCTCAACTCCGGGCCTGCAGTCGATACGGGCAGACTGGAGTTCGGCAGG
GGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCANGAGGAACACCGGTGG
CNAANGCGGNTCTCTGGGCCGATACTGACGCTGANGAGCNAAAGCGTGGGGAGCNAACAG
GATTAGATACCCTGGTTAGTCCACNCCCGTAAACGTCTGGGCGCTANGTTGTGGNGGGCC
ATTTCCACGGTTTTCCGTGCCNCANCTAACGCATTAAAGNGCCCCGNCCTGGGGAGTANC
NNGCCCCCAANGGCTTAAAACTCNAAAGAANTTGACNGGGNGTCCCNCNNNAAACCGGNC
GGAATCANTTTTGCTTAAATTNCTATGTNACTCCNAANAACCTTNCNTAAGGCNTTTGAN
CTTCCCGGAAAATNTNCCTNAAANTCNCGNGNTTCCNTTAAGGGTCNCTGCNCNAGGGGT
NCCATGGTTTTNNCCCCCNCTCNCNTTCTTTNAANTTNCTNTNTTNAANTNCCCCTAATT
AANTGCNANCNNNTTTTCTNTTTTACCCNTCNCTTTTGTGCGGGGACNCTTGGANNNNNN
TCNGTGNTTCAANCTTNNANAAAGGTTGGATTNNANNCAANTAACNCGTCCTTNNTTCTT
NTNTCCTCAACTTCTATNATNTGCTNGTNTAAAGCNTTCNCNANTTTGGNCNTTACA
>CR15
NGCCGCCTACNATGCAGTCGNACGGCAGCANGGGGGGAACCCCTANACNNNTGGATGGNN
NAGTGGCCNAATAANAGNANCGAANGCATCCCAATCNGCCGGNGAGAGGGGGACANCACA
CGGAAACTTGCCCTAANTNCCGCATACGANCNTCGGGTGANAGCGGGGGATCTTCCGACC
ATGCGNGAGGGGANGAGCCAATGCCNGATAGCGNGTTNGNGGGTTANNGGNCCACNNATG
CNACTATNCGNCGGTGGTNGGAGAGGATGACCNGCNACNCTGGGACNGANTCNCCGCCCA
NACTCCTCNGNAGGCAGCNTTGGNGAATANNGGACNATGNGGGCCACNCTGATCCNGCAT
TGCNNANATGTGCGNNNNAGGCCTCACNGGTCTTTNNTGCANNTTNNATCNGNCANCTNA
NNNNTNTGTCGNCCATTNTCCGNNANGATCTNANNGGTTCNATNNGNANCTTNNCNCTTT
CTNNCNNCNNCGCCCACGTNTTCNNCNNTANNCTCCAGANGGTNTTTTGNAANTAATATN
GNAATNTCTCGNGNAGTNACNNACANNGGNTNCNGGTNNNNANNCNTNTNAGTNCCNTTG
CNGAAACCNNTCTNNNCCTTTAATCTNGNNNAGTNNNCANTTGTNANNCCNTNNTCNAGA

NTCTNNANNTCNAGTTNNNCGTNNANNCACNCNNTTCCNTNGCNGTNANCNNTCNGNCGN
NTCNGNCNNANNNTCGTTTCNNNTNANTTNNNCTTNNTTNCTNCTNNGTATTNTNTNNNC
NCNNCTTTNTNNTNCNNNTCNNTNTACCNAGTTNACCTGTTNNTAGCNNNTTNANCNNTT
GGGNTATTNCTNNNTTNCTATNNACCCCTCANNACNTNTTGCATNNCACNNATNNTNNCN
GGNTCNTTTTANCGGNANNANNNTACACNNNANCNTTNCTNNCCTGTAAACTNNGTCCGG
CCTNANNNNTGTNGGGTACANCNTNTANNCGTTATCTTNTTNNNTNACCTNNNNCGCNTC
TCTCCNCNTATCTGNTNANNANTNTCCNTCCNNNTNTTCNTGNNCTNTNNAATTACCTNN
CCNCNANNCNTTNCTNTCNGNNNNTGCTTNTAGNNNTNATTCAANAGNNGCGTCNNTNNC
ACCCTNCNNANTCTNCCTTCTNCNTCGGGNCCTNTNCNCTTATNNTCNNCCTCCCNATGT
TATNTNTCNNCCANTNCNAATCACTCGNCCNCNTTTACCTTNGTNNCTCTCTNGCTTTTN
NANANTCCTACNNACTCTGTTCANNCAGTNNNNANTNNNTNNTTNNTGCNTTCNTNNNNG
ATNGTNTNNNCACTNTCNCNANNGNTTNTNNG
>CR18
ATGCGCTCCTTNCATGCAAGTCGGACGGCAGCGCGGGGGCAACCCTGGCGGCGAGTTGGC
GAACGGGTGAGTAATACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCG
GATTAATACCGCATACGATCTGAGGATGAAAGCGGGGGACCGCAAGGCCTCGCGCTCAAG
GAGCGGCCGATGGCGGATTAGCTGGTTGGTGGGGTAAAGGCCCACCAAGGCGACGATCCG
TAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
GGAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTG
TGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAAACTTCGTCCCTAA
TATGGATGGAGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGC
GGTGATGTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGC
ATTGCTGGAGTATGGCAGAGGGGGGTGGAATTCCACGTGTAGCANTGAAATGCGTAGAGA
TGTGGAGGAACACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCACGA
AAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAAC
TGGTTGTCGGGCCTTCATTGGCTTGGTAACNTAGCTNACGCGTGAAGTTGACCGCCTGGG
GAGTACGGTCGCAAGATTAAAACTCAAAGGAATTTGACGGGNACCCCCACAAGCGGTGGA
TGATGTGGATTAATTCNATGCAACNCNAAAAACCTTNNCTTCCCTTTGANATGGACNGAA
CCTNCNTTAANANTTTANGGTNCCCNAAANGGACCCTTCCCNCAAGNTCCTNCATGGNNT
NTNTTCACCTCNTGTTCNTAAANTTTTGGNTTAATTCCCNCAACNANCNCAACCCTTTNT
CCTTGNTTNCTNCCCANNAACNNTCCNGGAAAACTNCCGTTAANAACCNAANAAANGTGG
GGANTACNTNNATTCCNCNGNCCNTATGGNTNGGNTTCCNNCNTTANNNANNNNNNAAAN
AAAGNTTNTCNATCCCCNTNTTNAN
>CR19
NTGCGCTCCTTNCATGCAAGTCGGACGGCAGCGCGGGGGCAACCCTGGCGGCGAGTGGCG
AACGGGTGAGTAATACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCGG
ATTAATACCGCATACGATCTGAGGATGAAAGCGGGGGACCGCAAGGCCTCGCGCTCAAGG
AGCGGCCGATGGCGGATTAGCTGGTTGGTGGGGTAAAGGCCCACCAAGGCGACGATCCGT
AGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAAACTTCGTCCCTAAT
ATGGATGGAGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCG
GTGATGTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGCA
TTGCTGGAGTATGGCAGAGGGGGGTGGAATTCCACGTGTAGCAGTGAAATGCGTAGAGAT
GTGGAGGAACACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCACGAA
AGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATTTCAACT
GGTTGTCGGGCCTTCATTGGCTTGNTAACNTANCTNACGCNTGAAGTTGACCCGCCTGGG
GGAGTACGGTCGCAAGATTTAAAANTCAAAGGAATTGACGGGGNCCCCNCACAANCGTGG
ATGATNTGGATTAATTCNATCCAANCNCNAAAAACNTTNCCTTCCCNTTGACNTGGACGG
AACCTNNTNANNATTTTANGTTTCCCAAAAGGNACCTTCCCACNGTTCNTTCNTGNTTTN
TNCNNTTNTTTCNNGAAATTTTGGTTTATTNCCNCAANNNNNNANCCNTTNTTCTTTTTT
CNTCNCAAAACCTTCCNGTAATCCGTCNNTTCNACCTNAGAAAGTTGGNTTANTNTNTNT

NTCNTCCTTTTNGNNNGNTTNCNCNTCTNNANNTNNTTNAANNNTGTTTCNNNCCNTNGT
TTNCCNTTCNAAAA
>CR2
ANGGGGCCNCANATCATGCAGTCGAGCGGGGATTNTTTAGAAGCTTGCTTCTAANTAACC
TAGCGGCGGACGNGTGAGTAACACGTAGGCAACCTGCCCACAAGACAGGGATAACTACCG
GAAACGGTAGCTAATACCCGATACATCCTTTTCCTGCATGGGAGAAGGAGGAAAGGCGGA
GCAATCTGTCACTTGTGGATGGGCCTGCGGCGCATTAGCTAGTTGGTGGGGTAANGGCCT
ACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACA
CGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGGCGAAAGCCTGA
CGGAGCAACGCCGCGTGAGTGATGAAGGTTTTCGGATCGTAAAGCTCTGTTGCCAGGGAA
GAACGTCTTGTAGAGTAACTGCTATAAGAGTGACGGTACCTGAGAAGAAAGCCCCGGCTA
ACTACGTGCCAGCAGCCGCGGTTAATACGTAGGGGGCAAGCGTTTNTCCGGAATTATTGG
GCGTAAAGCGCGCGCAGGCGGNTCTTTAAGTCTGGTGTTTAATCCCGAGGCTCAACTTCG
GGTCGCACTGGAAACTGGNTGANCTTGAGTGCTTAANAGNAGAAGTGGAATTTCCNCGTT
GTATNCGNTGAAATGCTTTNNAGNTTGTTGGAGGTANCANCTNTGGGNGAATNTNTACTC
TTCTNNNNNTGTNTNCTGGACNCTNTNGGCTCTANNTNNTGTTGNGGNANCTTNTNTTNN
NTTTNGNTNNNCCNTGGTTAGTNTNNNCNTCNNTANNACTTTGTAATTCTNTTNTNTTAN
GGGTCTTTNCCAATANCNCNTGGNTNCTGAANGTTNNANANCNTCTAACCATTTCTGTTT
TGGGTNGATANTNNTCTCNNGNTCTNAAACTTANANGGTATTTTTCTTGNNANTCTNTCN
NGTGNATGGNATTTNTTGTTCTCTTNTNNNGNNTNNANNTNCANTNGNNCNNNTCCCTTG
TTTANGTACNCNTCNNNTNTATCGTTNTCTCATNATNCNNCCTATGTNCTTTCNCNTTNC
ATNNNAANTTTGTTTTGNTNTTTGTGNNNNAATTCNNNATNCNTTCTNTNNAGTTNNATG
TGTTANAAGCNGCACNTNTNCCTNTCTATTTNTTTTTTNTNTATATTNNGTNCTGATTTG
ATNNNTNGCTNNCTGNATNNNTGTGCNNAACATCCNNTTNNNAGGTTANNCTTNTTCTAA
TACTCTNTNCTCTCTNANCATTTGTTGTACNANCN
>CR21
ATGCGCCNCNTACCATGCAGTCGAACGCTGAACCCNCTTCGGTGGNNGATGAGTGGCGAA
CGGGTGAGTAACACGATGGGCAACCTGCCCCCGGCTTCTGGGATAACTCCNAGAAATTGG
GGCTAATACCGGATATGACCNCTGGNCGCATGGTCTGGTGGTGNAAANTTTTTCGGNTGG
GGATGGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTANTGGCCTACCAAGGCGACGACGG
GTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTGGGGAATATTGCNCAATGGGCGGAAGCCTGACGCAGCGACGCCGCGT
GGGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCANCTCCGACGAAGCNANAGTGACNG
TAGGANNANAAGAAGCACCGGNCAACTACNTGCCAACANCCGCGGTAATACGTAGGGTGC
AAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTTCNTCACGTCGGCTGT
GAAATCCCGTNAGCTCAACTCCCGGTCTGCATTCGATACGGGCGGACTTGAGTTCGGCAN
GGGATACTGGAATTCCTGGTGTAGCNNTGAAATGCGCAGATATCATGAGGAACNCCGGTG
GCGAAGGCGGGTCTCTGGGCCGATACTGACGCTNANNANCGAAANNGNGGNGAGCANAAC
AGGATTAGATACCCTGNTAGTCCACGCCGTNAACNTTGGGCGNTAGGTGTGGGGGCCATT
CCACGGTNTCCTTGCCGCAGCTAACNCATTNANCTCCCCGNNCTGGGGANTTACNGCCGC
CAANGNTAAAACTTCAAAGGANTTGACCNGGGGCCCCTNACANGNCGGCCGNACCNGNTT
GCTTAATCCNATNCNANNCTCTAAGAACNTTACCNTAGGNTTTGACNNTNCACCGGAANT
NTNGCAAANANTCCNGGGTTNNCNTTTGGCNTCTNGCCCNANGGGNTCCAATGTTTGTNT
TCCACTTCTTTTTCTTAAATTNTTGGGTANAATTCCCTNANCNNACTCACCCTTTTTTNC
TTTTTNCNANNCCNTTTTTGCNGGGNAATTTTTGTAANNCANCCGGGTNANNNCTNANGA
ANGTTGGTAATANCCAAATCTTTATGCTCCTTATNTCTNGGCTTTAAATNTNTCATANNN
NCCTNNNCNANGGCNNTNTTNNTTNNNNTNAACNA
>CR23
NTGCGCNCCTTNCATGCAAGTCGAACGGCAGCACGGGTGCTTGCACCTGGTGGCGAGTGG
CGAACGGGTGAGTAATACATCGGAACATGTCCTGTAGTGGGGGATAGCCCGGCGAAAGCC
GGATTAATACCGCATACGATCNACGGATGAAAGCGGGGGACCTTCGGGCCTCGCGCTATA
GGGTTGGCCGATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCA
GTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTGGGGAATTTTGGACAATGGGCGAAAGCCTGATCCAGCAATGCCGCGT

GTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTTGGTCCTA
ATATGGCCGGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGC
CGCGGTAATACGTAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGG
CGGTTTGCTAAGACCGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTGGTGACTG
GCAGGCTAGAGTATGGCAGAGGGGGGTAGAATTCCACGTGTAGCANTGAAATGCGTAGAG
ATGTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCACG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAA
CTAGTTGTTGGGGATTCATTTCCTTAGTAACGTANCTAACGCGTGAAGTTGACCGCCTGG
GGANTACGGTCGCAAGATTAAAACTCAAAGGAATTGACNGGGACCCGCACAAGCGGTGGA
TGATGTGGATTTANTTCNATGCCACGCGAAAAACCNTTACCTACCCTTNAATGGTTCGGA
ATCCTGCTTAAAAGTCNGGAGTNCTCNAAAAAAANCCGCCCNNANNTTCTTCANGGCTTT
TCTTANNNCNTTTCTGAAATTTTTGGNTTAATTCCCNCANCNANNCNACCCCTTTNCTTT
TTTNNTCCCNANAACCNTNTTAAGNAAANTNCCGTTAAAACCCGAANAANGNTNGGATTA
CTCCAATTCTCNNGGCCCTTTTNGNTNGGTTNCCCNNNTNNATGGTNNAAAA
>CR24
NTGCGCNCCTTNCATGCNAGTCGAACGGCAGCACGGGTGCTTGCACCTGGTGGCGAGTGG
CGAACGGGTGAGTAATACATCGGAACATGTCCTGTAGTGGGGGATAGCCCGGCGAAAGCC
GGATTAATACCGCATACGATCNACGGATGAAAGCGGGGGACCTTCGGGCCTCGCGCTATA
GGGTTGGCCGATGGCTGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCA
GTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTGGGGAATTTTGGACAATGGGCGAAAGCCTGATCCAGCAATGCCGCGT
GTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTTGGTCCTA
ATATGGCCGGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGC
CGCGGTAATACGTAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGG
CGGTTTGCTAAGACCGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTGGTGACTG
GCAGGCTAGAGTATGGCAGAGGGGGGTAGAATTCCACGTGTAGCANTGAAATGCGTAGAG
ATGTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCACG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAA
CTAGTTGTTGGGGATTCATTTCCTTAGTAACGTANCTAACGCGTGAAGTTGACCGCCTGG
GGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGACCCCACAANCGGTGGAT
GATGTGGGATTAATTCNATGCAACCCNNAAAACCNTTACCTACCCTTGACATGGTCGGAA
TCCNTNCTNANAGGCGGGATTGCTCNAAAAAAAAACNGNCCCANAGGTNCTNCANGGCTT
TCNTCNNCTTCNTTTCNTGAAATTTTGGGTTAATTCCCCCANCAANNCNANCCTTTTCTT
NNTTTCTACCCANAACCCTTTAAGGAAACTNCCGTTNNAANCCNAAAAANGGGGGNNTAN
NTNATNCCTNGGCCCTNTGGGNNGGGTTNCCCCNTNNANA
>CR25
CNNTNTTNNNNNNNNGNATNGGNCGCCNTNACATTTCAAGNTCGAACGGCAGCACAGCAG
NTAGCAATACTGTGGGATGGCGAGTGGCGGCACGGNTTGAGGAATACATCGGGACCTGCC
CAGACGTGGGGGATAACGTAGGGAAACTTACGCTAATACCGCATACGTCCTACGGGAGAA
AGCGGGGGATCGAAAGACCTCGCGCGGTTGGATGGACCGATGTTCGATTAGCTAGTTGGT
GAGGTAATGGCTCACCAAGGCGACGATCGATAGCTGGTCTGAGAGGATGATCAGCCACAC
TGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATG
GGCGCAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCAC
TTTTATCAGGAGCGAAATGCCATTGGTTAATACCCGGTGGAGCTGACGGTACCTGAGGAA
TAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGTGCAAGCGTTAATC
GGAATTACTGGGCGTAAAGCGTGCGTAGGCGGTTCTTTAAGTCTGCTGTGAAATCCCCGG
GCTCAACCTGGGAATGGCAGTGGATACTGGAGGGCTAGAGTGTGTCAGAGGATGGTGGAA
TTCCCGGTGTAGCGGTGAAATGCGTAGAGATCGGGAGGAACATCANTGGCNAAGGCGGCC
ATNTTGGGACAACNCTTGANNCTNAGGCACAAAAGCGTGNNNGAGCANNCNGGATTANAT
ACCCTGGTAGTTCCNCNNCCTTAAACTATGCGAACTGNNTGTTTGNTCCTNNACCTCGNA
ATATCCNATNTCCTNAAGCTAACCCCGTTTAATTNNCCCNCCNGGGTATTNCTGGCNCNC
ANNANTTANACTCATAGGNNTTGTCCNGGNTCCCTNCNCAATTNCGTNTNAATATGTTGT
TTTANTTTTTATCCNNCTCTNNNAAACNCCTTNCTTNTGTNTTTACTNTTTCTTNNATTT
CTNTAGAANATTCCGTNTTTCTNTTTNGGNAATNNNNNTNNCNTTTNCTTACNGTTTTTT

TGTTTNNNTTTCNTCTNCATTTTTCNNNNTNTTNCCNNNNCNTNNNCATCCCCTTTTTCT
TGATTCTNNCNNCTTTANGTCNTTNTTCTACTNNGTANTTNGTTNTTTTTNCGCGTNNAN
NTGTGCGNTCNTCNNNTTNTTATNTTNTTANTTNATNTGNCNTCTATTATNTTTTTNT
>CR26
CCCCGNAGGGNNNCNANNANNNAATCAGAGGGGCNCNCCGNGGCNANCCNNGNGGGGGGN
GACTGACGGNCCCNNNAAAAATAAANANCGGGCCCNGNGCGANCGNGGGGGGGNNCNCTN
AANGCNGANNANTNCGCCTTCCNTNCGAANCTNAACNGNTGNNTGTAGAGGNGANCGNNC
AGGNCANNNCCCNNCNNGGANGNGGANGANNNNNGNNTTANTNAGNATATGTAGGCGNNA
AGGNGCCACNANNNNCCNNNAACACNNCNGTTNNTCAGNGNNAGANNANNTGNCCTNNCC
CNNCGNNGGGNNTACGCNGNANNNACCCTCTNANCNNGNNTNNNANGGNTCNNGTNTANT
NANCACAGGGGNNNAACCNNNNCGCACCNTCNNNNNCNCCNNNGNNANNNTNGTANNNNN
TCNNNANNGTNTNCNNNCNNGTCCNTCTTNNTNNGNNNANCNANCNCNNNCNCNTANNTT
NCNNNGNATGNATCNTCTNCTNNCNGNCAATANNGTGTNNGANNCCTGCCNANTNACNCN
NTNNTNNNNTGNACGCGTTGGCGNNNNCNCNCNNGNTAANTNNGTNTNAGANNNNTCTNT
ANCNNTNNTTNCTTANANCNCTCAGCTTACNCNTCNTNNNCGNANNCATCNNCNACNCCT
CCNTCAAGACTNCCTNNNNNGNCNGNNCNNNCNNCNNNAANNCCNTGANNGGNTCCNNNC
NGTNNCNCGNAAATNGNNCNCNANNGTNCTNNCTGCAGANAGNTCNANACGCCCNNCACC
NNNTNATTTANTNNNNNGCNNNTTGTTCCNNNNNCNNTTATNNTANCNNNGTNCCTNNGG
NATNACACTNCNNGCCANNTNNCATTCTCNNGCNCANNCTNNTNNNGNNCTNTTCCTATN
NNNANNNANTNTTACACGGACGNNTNTTTNNNNCCNAATANTNCNANTNNTNAGGNNTNA
GGNTGNTNTNCCNTTNTCNNNTNCNNNATTNCCGNNNNNNNNCANNNGNAGNNCTNGNNCT
NGCCNCTNNNNNNNGAACCTNCAANAGANANNNTNNCCGANATNGCANANNNCNTCCTCC
GCNNNNANNNNCNTANGNNGCTNCGNANNTCANCNGNGGNNGNCTNNNNTNANTNCNCNN
CNNCANACNNTNCGCCNNGGCNGNTCTTTNNTTGCNTTTAANTCANNCNCNNNNGNNTAT
TGGNTNGANCCGNCANGCTCGNNCNNNNAANATNATATNNAACCCTCCTNCACTTANCTN
NNTNTNTANTNTCNATNAGNNCNCNGNAGACCNGNATGAACCNCANNTNNAGTNNNTCNT
NCTTNNCTATANTNCCGATATTTNCNNNNCTTTNCTNNNNTNNNAAGCTNNTCNCNCTGA
CTTATCGNATGANTNANGNCNGNCNNNACCCANCNNNNNANNTNATAGCT
>CR27
TTNNNTNNNNNNNNGNANCNGCGTCCTACCAANGCAAGNTCGAAACGGGCAGCACAGCAG
NTAGCCAATACTGNTGGGTGGCGAGNTGGCGGCACGGGTGAGGAATACATCGCGGTACCT
GCCCAGACGTGGGGGATAACGTAGGGAAACTTACGCTAATACCGCATACGTCCTACGGGA
GAAAGCGGGGGATCGAAAGACCTCGCGCGGTTGGATGGACCGATGTTCGATTAGCTAGTT
GGTGAGGTAATGGCTCACCAAGGCGACGATCGATAGCTGGTCTGAGAGGATGATCAGCCA
CACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACA
ATGGGCGCAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAG
CACTTTTATCAGGAGCGAAATGCCATTGGTTAATACCCGGTGGAGCTGACGGTACCTGAG
GAATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGTGCAAGCGTTA
ATCGGAATTACTGGGCGTAAAGCNTGCGTAGGCGGTTCTTTAAGTCTGCTGTGAAATCCC
CGGGCTCAACCTGGGAATGGCAGTGGATACTGGAGGGCTAGAGTGTGTCANAGGATGGTG
GAATTCCCGGTGTATCGNTGAAATGCNTATATATCGGGAGGAACATCNNTGGCNANGNGT
GNNCNTNTNGNNCAATNNTTTCTNTTTNNCTCGAATTNGTGTNGTATCCATANNNNNATT
TATTNNCCCCTNNTTNTCCNCTNCCTNNACNNTTNTNTNCTTNNTTGTTNNNTTCCTCAN
TTNCNTATNTTNNTTGNCNTNNNNTAANCTCTTTTTAATTNCCTCCNNCNNGGTGTGNTN
NTNNTCTTCTTTNTTTTNNACTTCTNNTNTTTTTTGNGNTTGTCCCNTTTCCANNNNTNT
NANTTTTTNNTTTNNNTTNNNTATTANNNNNNTTNNTATNNTTNTCNNTNNTTNANTTTN
TNNTANTTTNNCNCANTTNNNTATNNNCTCNTTTTTNNNTNNANTTTTTNTTANANNTCN
TTNTTTNNTTTTCNNTNNTTNNNTTNTNTTNNNTCNNTANTNTTTNNCTNNTNTNCTNNN
TTNTNTTTNNTNTNTTTCTTTTNTTTCNTTCTNNCNTNNTTTNNTNTNTTTNTCNTTTNT
CTNTTTTTTNTTTNTNTTNTTNNTTNNNNNTTTNTNTTTNNTNTTTCTNTCTNTTATNTN
NNTNNTTNTT
>CR29
NGCGCNCCTTNCATGCAAGTCGAACGGTAACAGGTTAAGCTGACGAGTGGCGAACGGGTT
GAGTAATGTATCGGAACGTGCCCAGTAGTGGGGGATAGCCCGGCGAAAGCCGGATTAATA

CCGCATACGACCTACGGGTGAAAGGGGGGGATCGCAAGACCTCTCGCTATTGGAGCGGCC
GATATCAGATTAGGTAGTTGGTGGGGTAAAGGCCCACCAAGCCGACGATCTGTAGCTGGT
CTGAGAGGACGACCAGCCACACTGGGACTGAGACACGNCCCAGACTCCTACGGGAGGCAG
CAGTGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGCGGGAAGA
AGGCCTTCGGGTTGTAAACCGCTTTTGTCAGGGAAGAAATCCTTTGGGCTAATACCCCGG
AGGGATGACGGTACCTGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAAT
ACGTAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCGGTTGTGT
AAGACAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTTGTGACTGCACAGCTCG
AGTGCGGCAGAGGGGGATGGAATTCCGCGTGTAGCAGTGAAATGCGTAGATATGCGGAGG
AACACCGATGGCGAAGGCAATCCCCTGGGCCTGCACTGACGCTCATGCACGAAAGCGTGG
GGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAACTGGTTGTT
GGGAAGGTTCCTTCTCAGTAACGTANCTAACGCGTGAAGTTGACCGCCTGGGGGAGTACG
GCCGCAAGGTTGAAACTCAAGGAATTTGACNGGGGACCNCACAANCGGTGGATGATGTGG
TTTAATTCNATGCAACNCGAAAANCCTTACCTCCCTTGACATGTCTGGAATCCTGAAAAA
ATTTGGGATGCNTCNAAAAAAANCAAAACACAGGTGCTNCATGGCCCNTCNTNNACTCNT
TCNTGAAATTTTNGGTTAANTCCCCCACCNACCCCAACCCTTTTCATTATTNNCTACAAA
GGGCNCTCTTATNAANNTGCCCNTTNACAANCCAANNAGGTGGGATNAATNNGGTNCTNN
NGNCNTTTNGGTNGGNTTACNCCNTNANAANGNCNNTNCAANGNTTNNCCCCCCGNGGGG
AACNNNNNCAAAAC
>CR3
NTGCGCTCCTTNCATGCAGTCGAACGGCAGCACGGGGGGTAACCCTGGTGGCGAGTGGCG
AACGGGTGAGTAATACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCGG
ATTAATACCGCATACGCTCGAGAGAGGAAAGCGGGGGATCTTCGGACCTCGCGCTCAAGG
GGCGGCCGATGGCGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCGT
AGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTCTGGGTTAAT
ACCTCGGGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCG
GTTCGCTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGGC
GAGCTAGAGTGTGGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAANTGCNTANANNT
NTGGAGGAATNCNNATGGCTAANGCAGCCCCCTGGGNTAACACTGACTCTNATTNCNCTA
AANNCGTGGGTGNGCNAACAGGNTTANATNCCCTGGTNNTCCACNCCCTAAACNATGTCA
ACTAATTTNTTGGGNATTTCNTTTTCNTTNNTTAACGTANCTANCTCTTNAANTTAACCN
CCTGNGTNAAGTACNGTCCTCANGATTNAANCTCAAATGNATTTNACNGNGAACCCTCAC
CNANCTGTNGNTTATTTTTNTTTANATNNTTTGCTANTNTNAAANNCTTTCTTANCCNTN
NNNNTTNANNTTATTCTTTCTTTTNCTTNTGAGTNTNTTNAATANNANACNTTTCGTNCN
GTTTTNATTTTNCTTTTNNNNTATATTGTCATTNNATTTNTTNGTTNTTTTTNTTANNCN
TGNCNTANCTTTTTTTTTNTTTTNTANAAGTANNNNNCATTATATNTTTNNTTATCCNCN
TTTTTNANNTNGTTNTTTNTTTTCCTCTNNNNCTCNNATNNTTCTTTTTNTTNNTNTNTA
NTNGNTNNNTNNNTNNTTTNTTCN
>CR33
ATGCGCNCCTACCATGCNAGTCGAACGAGACCTTCGGGTCTAGTGGCGCACGGGTGCGTA
ACGCGTGGGAATCTGCCCTCTGGTACGGAATAACTCAGGGAAACTTGAGCTAATACCGTA
TAATGACTTCGGTCCAAAGATTTATCGCCTGAGGATGAGCCCGCGTCGGATTAGCTAGTT
GGTGAGGTAAAAGCTCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGATCAGCCA
CACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACA
ATGGGCGAAAGCCTGATCCAGCAATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAG
CTCTTTTACCCGGGATGATAATGACAGTACCGGGAGAATAAGCCCCGGCTAACTCCGTGC
CAGCAGCCGCGGTAATACGGAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGC
ACGTAGGCGGCTTTGTAAGTTAGAGGTGAAAGCCCAGGGCTTAACCCTGGAATTGCCTTT
AAGACTGCATCGCTAGAATCGTGGAGAGGTAAGTGGAATTCCGAGTGTAGAGGTGAAATT
CGTAGATATTCGGAAGAACACCAGTGGCGAAGGCGACTTACTGGACACGTATTGACGCTG
AGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACG

ATGATGACTAGCTGTCGGGGCGCTTAGCGTTCCGGTGGCGCAGCTAACGCGTTAAGTCAT
CCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAAGAAATTGACGGGGGCCTGCACAA
GCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCAAAACNTTACCAGCGTTTTGACAT
GCCAGGACGGTTTCCANANATGGATTCCTTCCCTTACGGGACCTGGACCCAGNTGCTGCA
TGGCTTTCNTNANCTCGTGTCNTGNNATNTTNGGTTAAGTCCCNCNACCNNGNGCAACCC
TTNGNNTTTAGTTGCTACNATTTGGTTGGGCCCTCTAAAAAAAAACTNCCNGTNANAANCC
CGAAGAAAGGGGGGGANNANNTCNAATTCCCNNGGCCCTTACCCNCTGGGCTNACCCTTN
TNAANTGGCGGTAANAAGGGCCCNAANTNCCCNGNNNGNNNANCCCAAAANCCNTTCNNT
TTNAATTTTNNTNCACNCGNA
>CR34
CNTTNNNNNNNNNCGTCGCTCCTTNACANGCAGGTNGAACNGNTAACAGGCTTAAGCAT
GACNAGGTGGCGAACGGGTGAGNTAATGCTATCGGAACGTGCCCAGNTAGATGGGGGATA
GCCCGGCGAAAGCCGGATTAATACCNCATACGACCTACGGGTGAAAGGGGGGGATCGCAA
GGCCNTCTCTNTTTTATACCTTACCTATTTTNNGATTNTGGTTTTTGTCCTTTNCNNGGG
GAAAATGCCTNCTCTATGTTACAANNCTNANTGGATCTTTTAAGTNTCACCCTNCTCTCT
TGNNGNNNGNNACCNNTTNTNNACTTTCTACTGGANGTNGTANTTTGGNGNATTTTTTTN
AATTNNTGATCTCTCTCNTNCTTNCNATAGCNCCTATGNTCTCNTTTANTCGNCCTTCTG
NTTTNTNANNNNTCNNNCTTTTANTNNNNACGCCANCCATCNCCNCTTTCGAACNNCNCC
NGGTNNNCTCTNNACNNNNCNNCTTTTTTTNTATNATTCTCNCGTTTNTNNNTTTCCTTNC
CCNTTCTNCTTCNCGTNTTANTNCNTCNTNGCTNTCCTCNCTCTTTTTTTACTTGCATTT
CNTCTTTGTCTATNCNNNACTNTCNTTNTCCCTATNTTCATNTTCANNTTCTNNCNTCTT
TNCNATCCNATNTNNNTTTCATNCNCTNNTTTNCCTNATTCTCTNNANNCTTNTTTTNCN
CTTTCTNCNNGTCCNNTCCTNNCNATCTCCTCTANNNNCCCNCTNNTCNNCTCTTNTNCC
NNNNTNNNNTCCTCNTNTTTTTTCTTNNCNNTCTTCNTCGCTNCGGNTTNCNNTNCNTAT
TTNNNTNATNTTCTTTCANNCTCCTTNTTTCNTCNNTTNTCTNTCTCNCTCCGGTTTNGC
NATTTTCTTTTTGCATCTCTCACCTNNTNTCATTTCCATCTCNTTNNNCNTCNNTTNCTT
TGNCNTTNNTNCTANACCCNCTTCATNCCNCACCATCTNNNNANCCCCNNNCCTCATNTN
NCTTACGTCCACTTNNNTNNTNTNCNTTCTCACTNCCNGTATACNNACCTTNTTCNCTTC
TTTTCTNTNTCCTCNTCCNCNATCTCTTNTCNTCTNTTNNTNTTATNNTNTCTCTNCTCN
NTTCTTTTCTTNNNTATNTCCNNNCTCNNCCCCCANNTNNTNCGCTTNTTCCNACCNGNC
NTNNNACNNCTNTNCCTATANACTCTTTNNCCTATNTCCNTCNNCNTTNNCTATTTTTCN
TTNCNTTCTNNTTNNTNTTTCNTTNTNCNTANNNTTCTTGTTTTNATNCCTCCCCNCTTC
NTTTACCNTCTCTTCTTCCTCTCTCTTCCCNTNTTNCTCCTTTCTNTCATNCTTGNTAAT
CNNCTTTNNNCNNNNTCCTNTCNNGCGTNNATNNNTCNNTTTNNNNTCNTTNCTTNCCNN
CTNNTCNNCTTNCTNTT
>CR35
ATGCGCGNTCCTATCATGCAAGTCGAGCGAGGGTCTTCGGACCCTAGCGGCGGACGGGTG
AGTAACACGTAGGCAACCTGCCTGTAAGACTGGGATAACATAGGGAAACTTATGCTAATA
CCGGATAGGGTGTNTCCTCGCATGAGGAGATACGGAAAGATGGCGCAAGCTATCACTTAC
AGATGGGCCTGCGGCGCATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATGC
GTAGCCGACCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTAGGGAATTTTCCACAATGGACGAAAGTCTGATGGAGCAACGCCGCGT
GAACGATGAAGGTCTTCGGATTGTAAAGTTCTGTTGTTAGGGACGAAACAGTGCCGTTCG
AATAGGGCGGNACCTTGACGGTACCTAATTAGAAAGCCACGGCTAACTACGTGCCAGCAG
CCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGCGCGCAG
GCGGCTATGTAAGTCTGATGTTAAAGCCCGAGGCTCAACCTCGGTTCGCATTGGAAACTG
TGTAGCTTGAGTGCAGAAGAGGAAAGCGGTATTCCACGTGTAGCGGTGAAATGCGTAGAG
ATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGTCTGTAACTGACGCTGAGGCGCG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGAGTG
CTAGGTGTTGGGGGGTACCACCCTCAGTGCCGCAGCTAACGCAATAAGCACTCCGCCTGG
GGAGTACGCTCGCAAGAGTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGA
GCATGTGGTTTAATTCNAAGCAACNCGAANACCTTTACCAGGTCTTGACATCCCNNTAAC
GTCCTANANATAGGGGCTTCCCTNCNGGGCANCNGTTACAGGTGGTNCNTGGTTTTTCCT
CACTTCTGTTCNTGANATNTTTGGGTTTAATNCCCCCACCAGNNCCAACCCNTTTTTTTT

NTTNCCNNCATTCATTTTGGCCCCTTNAAAAAACCGCCCTCCANAAACCGNANAANGGNG
GGGATAACCTTCAAANNNNCANGCCCCTTTNACCNGGGNTCNNCCNTTNTNCATTGGNGG
TTCACNNGGNTTTTTNN
>CR36
ANNCGCGNCCTATCATGCNAGTCGAGCGAGGGTCTTCGGACCCTAGCGGCGGACGGGTGA
GTAACACGCTAGGCAACCTGCCTGTAAGACTGGGATAACATAGGGAAACTTATGCTAATA
CCGGATAGGGTGTNTCCTCGCATGAGGAGATACGGAAAGATGGCGCAAGCTATCACTTAC
AGATGGGCCTGCGGCGCATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATGC
GTAGCCGACCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTAGGGAATTTTCCACAATGGACGAAAGTCTGATGGAGCAACGCCGCGT
GAACGATGAAGGTCTTCGGATTGTAAAGTTCTGTTGTTAGGGACGAAACAGTGCCGTTCG
AATAGGGCGGNACCTTGACGGTACCTAATTAGAAAGCCACGGCTAACTACGTGCCAGCAG
CCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGCGCGCAG
GCGGCTATGTAAGTCTGATGTTAAAGCCCGAGGCTCAACCTCGGTTCGCATTGGAAACTG
TGTAGCTTGAGTGCAGAAGAGGAAAGCGGTATTCCACGTGTAGCGGTGAAATGCGTAGAG
ATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGTCTGTAACTGACGCTNAGGCGCG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGAGTG
CTAGGTGTTGGGGGGTTCCACCCTCANTGCCNCANCTNAAGCANTAAGCACTCCCCCCNG
GGGGANTCGCTCNNCNNNATTGAAAACTCAAAGGAATTTNACNGGGGGCCCNNCNCAANC
NGTGGAAGCNTGTGTNTTTAATTCNAANCCACCCCCNNANAACCTTNCCCNGGNNTTTGA
CNTNCCCNNTNANCNNCNTANNANATTNGNGTTTNCCTTTCCGGGNNNNCNNTTANNGGT
GGNTTNTGTTTNTNTTCNNNNCTNTTCNNTGGNATTTTNNGNTTNATTNCNNNANNNNNN
NNACCNTTTTTNTNTTTNCNNTTNNNNTTGNNNNTCTTNTNTNNNNTTNCCNTNNNTNNN
NNNNNNTNNNNNGNTNNNNNNNNNNTTNTNNNCCCANNNNTNNNTTNNTNNTCNTTTNTNN
NTNTNNNNTCACNNNTTTTT
>CR37
NTGCGCNCCTTNCATGCAAGTCGGACGGCAGCGCGGGGGCAACCCTGGCGGCGAGTGGCG
AACGGGTGAGTAATACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCGG
ATTAATACCGCATACGATCTGAGGATGAAAGCGGGGGACCGCAAGGCCTCGCGCTCAAGG
AGCGGCCGATGGCGGATTAGCTGGTTGGTGGGGTAAAGGCCCACCAAGGCGACGATCCGT
AGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAAACTTCGTCCCTAAT
ATGGATGGAGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCG
GTGATGTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGCA
TTGCTGGAGTATGGCAGAGGGGGGTGGAATTCCACGTGTAGCANTGAAATGCGTAGAGAT
GTGGAGGAACACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCACGAA
AGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAACT
GGTTGTCGGGCCTTCATTGGCTTGGTAACGTAGCTAACGCGTGAAGTTGACCGCCTGGGG
AGTACGGTCGCAAGATTAAAACTCAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGA
TGTGGATTTATTCNATGCAACNCNAAAAAACNTTACTTNCCCTTNACATGGAACNGAACC
TTCNATTAAANTTGAGGGTGCCCNAAAGGGAACCCTTCNCCCAGGTCNTTCCATGGCTTT
TCNTCANCTTNTGTCCTNAAATTTTTGGTTAANTCCCCCAACCNANNNCACCCTTNNTCC
CTGGTTNCTNCCCNANAACCCTCCCNGGAAACTNCCNNTNACAAACCCGANNAANGNNGG
GATTACTTCNANTCNCCNGGNCCCTTNNGGTTNGGNTTCCCNNTCNANAAGGTTNNAANA
AAGNTTNCNAC
>CR38
TGCGCNCCTTNCATGCNAGTCGGACGGCAGCGCGGGGGCAACCCTGGCGGCGAGTGGCGA
ACGGGTGAGTAATACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCGGA
TTAATACCGCATACGATCTGAGGATGAAAGCGGGGGACCGCAAGGCCTCGCGCTCAAGGA
GCGGCCGATGGCGGATTAGCTGGTTGGTGGGGTAAAGGCCCACCAAGGCGACGATCCGTA
GCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTGTG

TGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAAACTTCGTCCCTAATA
TGGATGGAGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGC
GGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCGG
TGATGTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGCAT
TGCTGGAGTATGGCAGAGGGGGGTGGAATTCCACGTGTAGCANTGAAATGCGTAGAGATG
TGGAGGAACACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCACGAAA
GCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAACTG
GTTGTCGGGCCTTCATTGGCTTGGTAACGTAGCTNACGCGTGAAGTTGACCGCCTGGGGA
GTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGACCCGCACAAGCNGTGGATGA
TGTGGATTAATTCGATGCACCCGAAAAACCTTACCTACCCTTNNACATGGACGAAACCTC
NATNANANTTGAGGGTGCCCNAAANGGANCCNTCNCNCAGTTNCTGCAGGCTTTCTCACT
CNTGTCNTGAATTTTTGGTTAANTCCCCCAACNAGCNCACCCTTTTNCTGGTTNTTCCNA
NNACNCTCCGGGAAACTGCCGTTAACAACCNGANAANGNGGGATTACNTNATTCCTCNGG
CCTTTTGGTNGGCTTCCNNCTTTACNNGGTTNGAAAAANNNTTCCAACCCCGGNGNACCC
TT
>CR4
NTGCGCTCCTTNCATGCAAGTTCGAACGGCAGCACGGGGGTAACCCTGGTGGCGAGTGGC
GAACGGGTGAGTAATACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCG
GATTAATACCGCATACGCTCGAGAGAGGAAAGCGGGGGATCTTCGGACCTCGCGCTCAAG
GGGCGGCCGATGGCGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCG
TAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
GGAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTG
TGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTCTGGGTTAA
TACCTCGGGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGC
GGTTCGCTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGG
CGAGCTAGAGTGTGGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGA
TGTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGCTAACACTGACGCTCATGCACGA
AAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAAC
TAGTTGTTGGGGATTCATTTCCTTAGTAACGTAGCTAACGCGTGAAGTTGACCGCCTGGG
GGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGA
TGATNTGGATTTAATTCGATGCAACGCGAAAAAACCTTACCTACCCNTTGACATGGANNG
AANCCCTCCTGAAAGGTGGGGGTTGCNCGAAANAAAACCTTCCCANAGGTTCNTGCATGG
CTTTTCTTCANCTCNTGTCNTGGAAATTTTGGGTTTAANTTCCCCCAACNAGCNCAACCN
TTTTTTTTTNATTTGTTTCNAAAGGNCNCTTTTAAAAAACTNCCGGTAAAAACCCGAAGA
AAGGNGGGATTACNTCAATTCCTCATGNCCCNTTTGGGTGGGNTTTCCCCCTNNNCAATT
TTNNAANAAAA
>CR45
TNNNCCGTTTTTNNTGGGGGNANCGCGTCCTACACANGCAGNTCGAACGGCAGCCACAGC
ANNTAGCAATACTGNTGGGTGGCGAGTGGCGGACGGGTGAGGAATACATCGGGACCTGCC
CAGACGTGGGGGATAACGTAGGGAAACTTACGCTAATACCGCATACGTCCTACGGGAGAA
AGCGGGGGATCGAAAGACCTCGCGCGGTTGGATGGACCGATGTTCGATTAGCTAGTTGGT
GAGGTAATGGCTCACCAAGGCGACGATCGATAGCTGGTCTGAGAGGATGATCAGCCACAC
TGGGACTGAGACACGGCCCANACTCCTACGGGAGGCAGCANTGGGGAATATTGGACAATG
GGCGCAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAANGCCTTCGGGTTGTNAANCAC
TTTTATNAGGANCGAAATGCCATTGGTTAATACCCGGTNTGAGCTNACTGTACCTGAGGA
ATNAGCACCGTGCTTACTTTCGTTGCCNANCTCCNCNGTTNTTCGAATGNTNCNNTCTCT
TTTTCNTATTTTANTGNGNNNTAATNNTNGCTCNNTGCNNGTTTNTTTNNNTTNCTTCTC
ANTAATCCNCCNNTACTNNATTCCGNTTTNTGGCTNTNGTATTCNTNTNTNNNTTTTNTG
TTNTTANNNTGTNATTTCTCTTTTTNNTTTTTTCTTTTNATTNTTTTTTTTTNANTTTTT
TTNATTGCNCNTTTTTTTTTTTNTTTGCTCTNTTTTTTNTNTNNNCNTTTTCTTNTTTNT
NTTNTTTCCGTTNNNNTNANTTTNCTNNTAGAATTTTTTCCNTNNTTNNTTNTTNTTNNT
TNTCANTNTNTTTNTTNTTNTNTTTTNTTTTCTTTTNTTTTTTTTTTTTTCCNTNNTNTN
GATNATCTATTNNCTCCTTTCTCNTTTNGCNGATCTTNTNNTNCTANTNGTNTTTNCTNN

NTNATTNNNTCNATTTGTTTTNTATNTTTNTTNTTTNTNCATNTATNTTNCNNNTTTCNA
NTNNNTTTNCNTTTNTNTCNNNNNCTNTNTTTNTNTTTTTNTNNTNTCNTTNTNTTNTTT
GCTTATNTTNATTCNTNNCNTNTTTNTTATTTNCTTTTCCCTTTTGNTCTTTNCTTTNTN
TNNNNTGTNTCTTTTCTANNTTNTTTNTNNCNTNTTTATTNTTATTTTTTGTNTANNAGT
NTNNTATNATCGTTTCTTNTNTTTTNNNTTCANCNTTNTATTTTTANTNTTTTCCTTNNN
TNTNNTNCNTTGTCTTTTCTCCNNCT
>CR46
GNNTTTGNNNNGGGGNANCGCNCCTACCANGCAAGTCGAACGGCAGCACAGCCAGTAGCA
ATACTGTGGGTGGCGAGTGGCGGACGGGTGAGGAATACATCGGGACCTGCCCAGACGTGG
GGGATAACGTAGGGAAACTTACGCTAATACCGCATACGTCCTACGGGAGAAAGCGGGGGA
TCGAAAGACCTCGCGCGGTTGGATGGACCGATGTTCGATTAGCTAGTTGGTGAGGTAATG
GCTCACCAAGGCGACGATCGATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGA
GACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGC
CTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCAG
GAGCGAAATGCCATTGGTTAATACCCGGTGGAGCTNACGGTACCTGAGGAATAAGCACCG
GCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGTGCAAGCNTTAATCGGAATTACT
GGGCGTAAAGCGTGCGTANGCNGTTCTTTANGTCTGCTGTGAAATCCCCNGGCTCAACCT
GGGAATGGCANTNNATACTGGNTGTCTAGATTGTGTTCANANNATGTTGNAAATTNCNGN
TNTATCNTTNAANNNTCTTNNATNTNTNNATTNATCATCNNTTTTNNTANGTNTTNCATT
TTTTNTNNANNTTNNTTCNTATATNTNNTNAATTNNTNNTTNTATTNNNTNTNNTTTTTN
TTTCTCTTNNNNNTNCNNTCCNTTTTTTNTNTNNTNNCTTTTTTTTTNTNTNTTTCTNNN
NANTTNTTTTTCNTTNNNNNNTTTTTNTTTTTTTTNTNNNNGTNTNTNTNCNTTNNNNNNN
ANNNNNTNNNNCTNATNNNTTCTTTTNNNTNNCNNNNCTANNNNNTNNTTCTTNCNNTCT
NTTTTTTNTTNNTNTTTNNTTNTTTNTNTNNNTNNNTNTNATTNTTNTTTCTNTNNTTNN
TATTNTNTNTNNNTNATTTNNTTTTTCTNTTTTNTNNTNTTTNTTTTTNTTTTTNNNTTC
TNTTNTCNTNTTTNNNNTTCNCTNNNTTTTNTNTTTTNTNTNNTTNTNAGNCTNTTTTNT
TNTTNTTNTTTNTANCTNCNTNTTNTNNTTNNTTNCTNNTNTTTTTNTNTTTNTNNNTNN
NTNTTTTCNNNTATNTTTNNNNNTNNTTNTTCNCNTTTNTTNTNNTTCTNTA
>CR48
NNTNTTNNNNNNNGNNNCGCNTCCTACCATNGCAAGNTCGAACGGCAGCACAGCAGTAGC
AATACTGTGGGTGGCGAGTGGCGGACGGGTGAGGAATACATCGGGACCTGCCCAGACGTG
GGGGATAACGTAGGGAAACTTACGCTAATACCGCATACGTCCTACGGGAGAAAGCGGGGG
ATCGAAAGACCTCGCGCGGTTGGATGGACCGATGTTCGATTAGCTAGTTGGTGAGGTAAT
GGCTCACCAAGGCGACGATCGATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTG
AGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAG
CCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCA
GGAGCGAAATGCCATTGGTTAATACCCGGTGGAGCTGACGGTACCTGAGGAATAAGCACC
GGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGTGCAAGCGTTAATCGGAATTAC
TGGGCGTAAAGCGTGCGTAGGCGGTTCTTTAAGTCTGCTGTGAAATCCCCGGGCTCAACC
TGGGAATGGCAGTGGATACTGGAGGGCTAGAGTGTGTCANAGGATGGTGGAATTCCCGGT
GTAGCGGTGAAATGCGTANAGATCGGGAAGGAACATCATTGGCNAANGCGGCCATCTTGN
ACAACACTGACNCTNAGGCACNAAAGCNTGGGGAGCAAACANGATTAGATACCCTGGTAN
TCCACNCCCTAAACNATNCGNAACTGGATGTTGGTCTCAACTCGNNANATCANTTTCNAA
ACCTNANCNTTAATTTCTCCNCCTGGGNANTNCNGTTNNCATTANTNNAANNTNAAGGAA
TTTNACGGGGNCCCNCNCANNCNNTNNGANTTTNTNGTTTNNTTNNTTNNACNNTANAAN
NTTTNNTNGNNCNTTTCTTNTNCTANNTTTTTCNANANTNTCCNNTANTTNCTTTCNNTA
NCNTNACACTNTTTNTNCNGNNTTNNNCNACTCTTTNTTNTNNNNTTNGTTNTTTTCCNN
ANNNTNNNAACNTTNTNNNTTTNTNTNNNNNTTNTNNNNNNCTNTNTTNTNTTTCCNTNT
NCNNCNNNTNANNTNNNNNNTTNTTTNTNTTNTTNCNNTTNNNNNNTNNNNNNTNTNTNN
NTTTTNNNNNTNTTTNNNNCNTNTNNNTCNNNNTNTNNN
>CR49
TTTNNNNNNNNGNNNCGCGTCCTACCANGCAAGNTCGAACGGCAGCACAGCAGTAGCAAT
ACTGTGGGTGGCGAGTGGCGGCACGGGTGAGGAATACATCGGGACCTGCCCAGACGTGGG
GGATAACGTAGGGAAACTTACGCTAATACCGCATACGTCCTACGGGAGAAAGCGGGGGAT

CGAAAGACCTCGCGCGGTTGGATGGACCGATGTTCGATTAGCTAGTTGGTGAGGTAATGG
CTCACCAAGGCGACGATCGATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAG
ACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCC
TGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCAGG
AGCGAAATGCCATTGGTTAATACCCGGTGGAGCTGACGGTACCTGAGGAATAAGCACCGG
CTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGTGCAAGCGTTAATCGGAATTACTG
GGCGTAAAGCGTGCGTAGGCGGTTCTTTAAGTCTGCTGTGAAATCCCCGGGCTCAACCTG
GGAATGGCAGTGGATACTGGAGGGCTAGAGTGTGTCANAGGATGGTGGAATTCCCGGTGT
AGCGGTGAAATGCGTANAGATCGGGAGGAACATCANTGGCGAANGCGGCCATCTNGGACA
ACACTGACNCTNANGCACNAAAGCGTGNGGAGCAAACANGATTANATACCCTGGTAGTCC
ACGCCCTAAACNATGCNAACTGNATGTTGGTCTCANCTCNNANATCANTGTCNAANCTNA
CCCGTTAANTTCNCCNCCTGGGGAGTACNGTCNCAANACTGAAACTCANAGGAATTTNAC
GGGGCCCCCACAANCNNTGNANTTNTTTGNTTTTAATTCNATNCNACNNNNAANAANCCT
TACNTGGNCCTTTNANTTGTTCNNNAATTTNTNNNANNNTNNNAANTTCCCTTCCNNGNA
TTGCNAACANNNNNTNCTNCNNTGNNTTNCTTNNNNTTCTTTNCNNNNATTTNTNNTTTA
NTTCCNNNNCNNNNNNNCCCTTNNTCTTNNTTTTCNNNNNNNTNNNNTNNNANNTNTNNN
NNNNTCCNNTTNNNNNNNCNNNNNANNTGGNNTTNNTTTNNNTTNNNNNNNCNNTNNNNN
NTNTNNNNNNNNNNNNNTNNNNNNNTNTNTNNNNNNNNNNNNNNTNNNNNT
>CR5
AGNGCGGCAATCATGCAGTCGAGCGGGGTTNNTTAGAAAGCTTGCTTCTAANTAACCTAG
CGGCGGCACNNGTGAGTAACACGTAGGCAACCTGCCCACAAGACAGGGATAACTACCGGA
AACGGTAGCTAATACCCGATACATCCTTTTCCTGCATGGGAGAAGGAGGAAAGGCGGAGC
AATCTGTCACTTGTGGATGGGCCTGCGGCGCATTAGCTAGTTGGTGGGGTAANGGCCTAC
CAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACG
GCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGGCGAAAGCCTGACG
GAGCAACGCCGCGTGAGTGATGAAGGTTTTCGGATCGTAAAGCTCTGTTGCCANGGAAGA
ACGTCTTGTAGAGTAACTGCTANAAGAGTGACGGTNCCTGAGAAGAAAGCCCCGGCTAAC
TACGTGNCAGCANCCGTGGTAATNCNTAGGGGGCAAGCGTTGTNCCGGAATNATTGGGCG
TAAAGNTGCGCGCAGGCGGCTCTTTAAGTCTGGTGTTTAATCCCGAGGCTCAACTTCTNG
GTCGCCACTGNAAACTGGAGAGCTTGAGTGCAGAANATGATAGTGNAATTCCACGTGTAG
CGGTTTAAATGCGTATAAGATNTGGNAGGAANACCAGTGNNCCAAAGGCTAACTNTCNTN
GGNTGTAACTGACNCTNNANGCGCGNAAGCCTNTGGGGAGCAAACCNGGTATTGNTACCC
CTGNGTAGTTCCNCNCCGTTAAACTATGAANNNTAGNTCGTTTNGGGGCTTTNCNNNTCC
CCTTTNGTTNCCNAAAGTTTACCNANTTATANNCNTTTCCTCTCNGNNGAANNNCNGNNN
CTCNNNGNANTGAANNCCNCAAAAGGGCANNTNNCCGGTNANNCCCNNNTCNNNGCTGNG
TGANNNTTTTTTGNNTTTAANTTNCNAAACTCCCCNNCAAGNNANTCTNNCNNGNNTTTT
NACCTCNCCNTCTTAACCTTTATCANNNTTTNNTATTTTCTCCNTCCTNNNNNNNTTGAG
NCNNNGGCGNTTTNTNNNTTTTTTTNCTNCTCGNTNCNNTNNTNTTGCCTTTNATGNTTN
CCNCNCCNNNTTATCTTTTNCTTTATATTGTNNTNTNGGTTNTTTNNNANNCTTNNNNCA
TNTTCTNTTTNNTANTNCCTATACATNCCCTTNNTCNTCTNNTATTNTACTANCCTACNC
CCNNTTNTATCTNCCTNNTTNNT
>CR51
NATGCGCGNTCCTATCATGCAAGTCGAGCGGGGATTNNTTANAAGCTTGCTTCTAAATAA
CCTAGCGGCGGACGGGTGAGTAACACGTAGGCAACCTGCCCACAAGACAGGGATAACTAC
CGGAAACGGTAGCTAATACCCGATACATCCTTTTCCTGCATGGGAGAAGGAGGAAAGGCG
GAGCAATCTGTCACTTGTGGATGGGCCTGCGGCGCATTAGCTAGTTGGTGGGGTAANGGC
CTACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGA
CACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGGCGAAAGCCT
GACGGAGCAACGCCGCGTGAGTGATGAAGGTTTTCGGATCGTAAAGCTCTGTTGCCAGGG
AAGAACGTCTTGTAGAGTAACTGCTATAAGAGTGACGGTACCTGAGAAGAAAGCCCCGGC
TAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTGTCCGGAATTATTGG
GCGTAAAGCGCGCGCAGGCGGCTCTTTAAGTCTGGTGTTTAATCCCGAGGCTCAACTTCG
GGTCGCACTGGAAACTGGAGAGCTTGAGTGCAGAAGAGGAGAGTGGAATTCCACGTGTAG
CGGTGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGACTCTCTGGGCTGTA

ACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCAC
GCCGTAAACGATGAATGCTAGGTGTTAGGGGTTTCGATACCCTTGGTGCCGAATTTAACA
CATTTAAGCATTCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGG
GGACCCCCCACAAGCAGTGGAGTNNTGTNGGTTTTNATTCNAAGCACCNCNAANAACCTT
ACCAGGTTCTTGACATTCCCTCTGACCGNNTTAAAAATNNNACTTTTCNTTCGGGANAAA
AGGAAACAGGTGGTGCANTGGNTTGTCNCNCACTNCTTGTCNTGNAANTTTTGGGGTTTA
NTTCCCCCAACNAGGGGCAACCCTTTTNTTTTATTTTCNNTCNGGTNAANCTNGGGCCTT
TTANNCAAATTGCCGGTGNNNAANNNGGAGGGAAGGGTGGGNATTAANCNCNAATATNNA
TCNCCCCTTTANCNNGGNNTCCCCNNNTNAAGGNCGGTTACCGGA
>CR52
TGCGCTCCTTNCATGCNAGTTCGAACGGCAGCACGGGGGTAACCCTGGTGGCGAGTGGCG
AACGGGTGAGTAATACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCGG
ATTAATACCGCATACGCTCGAGAGAGGAAAGCGGGGGATCTTCGGACCTCGCGCTCAAGG
GGCGGCCGATGGCGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCGT
AGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTCTGGGTTAAT
ACCTCGGGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCG
GTTCGCTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGGC
GAGCTAGAGTGTGGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGAT
GTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGCTAACACTGACGCTCATGCACGAA
AGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAACT
AGTTGTTGGGGATTCATTTCCTTAGTAACGTANCTAACGCGTGAAGTTGACCGCCTGGGG
AGTACGGTCGCAAGATTAAAACTCAAAAGGAATTGACGGGGACCCCGCACAAGCGGTGGA
TGATGTGGGATTAATTCGATGCAACCCGAAAAACNTTNCCTACCCTTTGANNTGGNCGGG
AACCCTTCTGANAGGTGGGGGTNTCTCNAAANAAAACCTTCCNACNAGNTCCTGNCTTGG
CTGTCCTCCACTNCTGTCCTNANANTNTNNGGTTTAATTCCCCCANCCANGCCCAACCCT
TTTTTTTNTTTTTTCNAAAGGCCTNTTTAAAAAACTGNCCGTGACNAACCCGAAAAAGTG
GGGGATTACNTCAANTCCCTTTGNCCCTTTTGGNANGGCTTCCCCTTNTATAANNGTCGN
AAAANAGGTNTTCCCCCC
>CR53
TGCGCANCCTNCCATGCAAGNTCGAACGGNNAANNNNNGTGTAGAATCTTGCNTTCTNNT
GCCANCGATTGAGCTNANCNGNGTGGANTANNACACTCGTGAACGTNGTNCCGTNCTAGC
TNCNGGATANCTAGTCGTAANNNTTAGACTTATACCTGCCCACGACCTNGTAGGGTNGAA
TGTTGNGGTACCGCANGGCCTCATGATNTCGCAGCGCTNTCATNTCTNATTAGNTAGTNT
NGTGTNNNNAAGGCCNCCNCAGGCTAACCATCCTNAGCTGGTCTCAGAGCNACNATNAGC
ACATGACTNCACACTGAAACACTCCTCCAGACTCCTACAGNAGGCNAATATTGCNNAATT
TTGGACAATGTGATANAGCCTNANCCNGTNATGCCATCANGTNNTNNGAATTGCTTNCNT
NTNTNAACNANTTTTATCCGAAAATAANTGTCNCTGGTTAATACCTGNCGNCTATCNCCG
TNCCNGNAGANTNNGGACNNCCTAANTNCCTGNCATNNTCCGNANTTATNNGNNGNNNNC
ANCTCTTAANCGGATTNNCTNGTCGTNNNTNANNCCCCGGNGCTTNNCCCCGANNCTNCT
GNNANNNCCGACNTNNTNTNGTTAATTNNNTTAGNTNCTNNTNCCTNNTNTTGNGTNAAA
ATGNGNNAANATTCCGNGTANCACNGTTNNNTNNGNNGAATTTNTGNGNNANTCCTNNCG
CTAAGGANCNCNNCNGTGANNACNNNANNGNNTTNGATNCNNTGNGTAGTCCACCCCANA
NNNTNNGATNCNCTTGGTGTGCCACAACNTTNNACGATGTCCGTACTTATTTGACATTNT
TTTTGCNCCTANCGTAGATTACGNCCGCAATTGCTCAAACTNGNNAAGNANTGNCCGGNN
GANNANACNNCCANNGAATTGACGTGGGNTNNCTTNCAACGCANCGGNANAAAATCTTTT
CAATTNCTTTGNCATACACCAAAAAATTCCCTNAACCCTTGGGNTCCCCCTTNANGGANC
GNTNAAANTGGTTGTTGNTGNNCTAANNNNAACTNTNCCCNGGNANTTTTTGGTTNATTN
CCCNCTCCNNNNCCAANCNTTNTTCCNTTNNTTCCCNCAAGNNCTTAANNTTTNNTTTTT
TTTGNTAAAANCNNCGNGTTNTAANNAAATAAGNTGNAAACCNNNNAAANTTTGGNNCCC
TTNNTTTTNGGGGCCCNTTNTTAAAGGNTTNCCNAANAAANGGNA
>CR54

NATGCGCNNCTTANCATGCNAGTCGAACGATGAACCACTTCGGTGGGGATTAGTGGCGAA
CGGGTGAGTAACACGTGGGCAATCTGCCCTGCACTCTGGGACAAGCCCTGGAAACGGGGT
CTAATACCGGATATGAGCTGCCCAGGCATCTGGGTGGCTGTAAAGCTCCGGCGGTGCAGG
ATGAGCCCGCGGCCTATCAGCTTGTTGGTGAGGTAACGGCTCACCAAGGCGACGACGGGT
AGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGA
GGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAGAGTGACGGTA
CCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAA
GCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCGCGTCGGTTGTGA
AAGCCCGGGGCTTAACCCCGGGTCTGCAGTCGATACGGGCAGGCTAGAGTTCGGTAGGGG
AGATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCG
AAGGCGGATCTCTGGGCCGATACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGA
TTAGATACCCTGGTAGTCCACGCCGTAAACGGTGGGCACTAGGTGTGGGCAACATTCCAC
GTTGTCCGTGCCGCAGCTAACGCATTAAGTGCCCCGCCTGGGGGAGTACGGCCGCAAGGC
TAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGCTTAATTCGA
CGCAACGCGAAGAACCTTACCAANGCTTTGACATACACCGGAAAANCCTGGNANACAGGG
TCCCCCTTTGTGGTCNGTGTACAGGTGGTTCATGGCTGTTCNTCANCTCCTGTCGTGANA
TTTTTGGGTTNAANTTCCCNCACCNAGCGCAACCCTTTGTTCCNTNTTTCCANCAAGNCC
CNTTNTGTNNTTGGGACTNCAAGGGAAAAACCNCCGGGTTCACTNCGAAGNAAGGGGGGA
CAACTTNANTTCATNTNGCCCNTTTTTTTGGGNCTGCCNCNTGTTAAATGNCCGNNNATT
AATNCNNACCNNGGGGGAGNAANTT
>CR55
TTNTNGNAANNNCATTNNNAAAANATGNCGTNGNGNGTCCNTACACATGNCAAGNTCGNA
ACGGGNCAGNCACAGNTTTCGNTAGNCAATTACGNANTGGGGGNTGGGCCCGNAGATGGC
GCGCACGTGTTTGAAGGAAATACACTCNGCGGTACCTTGGCCCAGACGNTGGGGGATAAC
GATAGGGAAACTTACGCTAAGTACCGCAAACGTTCCCTACGGGTANAAAGCGGGGNGATC
TTCNNGACCTCGNNNGNGTTGGNATNNCANNGNNGTTCTATTAGNTNGCTGGNGGNGTNA
TTGGCCCACCNNANCCNCCNTNNTNTAGCNTGGCNNNGAGNTCNCANCCTCCNCCTCGTG
NNNGACANNTATGACNCNNCCTNTNCNACNTTNTAGNGAAGAANTGCNNNGTACGTNNNN
NTCNNNCCNACNTGGNTNCCNTNCANNNNGTTCNAACANTTATTCNTCNNTTTCATTTAA
NNTNGTGGTCNCTNNTNTNTTTGNNTTNATNCNNTTCTNNTGCNNANNNCTCCNATGNCA
TNTGTNTTTCTGTTCTCNCTNCTTTCNGCGTCTNGNNNNNNTNGNGNNNCGNTTANTCNT
NTANAATNNTNNNTCTNNTCCNNTCNANNTNNTCNNTNNCTTATCTNTGCCGCANCNTNA
TNCTTAGTNTNNGACNNCNTGCTTNTNNTCNNNNNNTNTNNTTCNCCANNGCTGCTTCNN
NTGCTTTGTTTNCTNNGNNCTNAGNNNNTCTNNTANCNGNNNNNTTANANANNTGTNTNT
GTCTTNCNTNNGTNATNNTANANAATCNTTNNNGNTCTCNNNNNNNNNNNNCNAATAAGN
CGNTTTTNGNTTNNTGNNTCTTNNNCNGTCNGCCTTTNNNTCNNTNGTNCATTTCNGTAT
CCTCNTNTNTAGNGNCTNAACNCNCNTTNCTGNCNNNNCNNGNNNTTCCNTNNNGNNNTN
NATCCTTCCTNTANNTTGANCNGGNTNNTNNNTTCNANCNTCNCTTCAGNACNCANTTCN
NTNTNNTTNNCTCCNCATNNAATNGNNTACCATCTAATTNTTTNNTCTATATNNTTAAGC
NTNNCNNCNTTCNTNTCTNTNTCGNANTNANNNTNGNTNTNCCCGTGATATACNATTNNT
TTNNATNCTNTAGNNCTTATATCNNAANTCTTTANTNTTTNGNTNNTTNNTNTTNTTNCT
GCTCNNAGNTCAAGCGANATAACTNNNTAGNTNCTNTNACNNNTNGCTATCNNCTACNTA
TNCTTNCGCCTATANNTNTACTNGNACGTATTTNTNCCTGNNACANACNTGANANCNCTG
TNCNCATGNNNCTCNNTNNGNGNTNCCNGAGTNTANNNNCNAGTANGTNGTANNNTNANA
GTTNNANCNTCNNTTTTCCCCNNNTNNTCNNGCATTGACGACNNCNNTGNCCNCAGTNTN
NATNCNCNNCATCTTNCNAGCTCGCNNTTCNNCTNTTCGNNNTGCCAATTCANNNCNCAN
CTCTANNNTGGTTNCTATTTTANTNTNCCTNACCTTTTNTNAGTNNNTATNAGACTTNGN
CANNNTNANNTTNTNCCNATNNNTTNNGTGNTAANNCNACNTNACNNGTTNNCNTNTGNN
CNTTTNNGGCTTCTATCCNTNGNNNATGNTTNTNGTATTNTNNTTNNNNCTNATCCTNNG
NNGNTTTNTANNTTNCNCNTTCGANATCTNTCTTTCCTCNNTNCTTANCTNCNTTCCG
>CR56
NTTTNNNNNNGNANCGCNCCTCCATGCAGTCGAACGGCAGCACAGTCGTAGCAATACGAT
GGGTGGCGAGTGGCGGACGGGTGAGGAATACATCGGGACCTGCCCAGACGTGGGGGATAA

CGTAGGGAAACTTACGCTAATACCGCATACGTCCTACGGGAGAAAGCGGGGGATCTTCGG
ACCTCGCGCGGTTGGATGGACCGATGTTCGATTAGCTAGTTGGTGGGGTAATGGCTCACC
AAGGCGACGATCGATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGG
CCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCC
AGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCAGGAGCGAA
ATGCCATTGGTTAATACCCGGTGGAGCTGACGGTACCTGAGGAATAAGCACCGGCTAACT
TCGTGCCAGCAGCCGCGGTAATACGAAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTA
AAGCGTGCGTAGGCGGTTTTTTAAGTCTGCTGTGAAATCCCCGGGCTCAACCTGGGAATG
GCAGTGGAAACTGGAAAGCTAGAGTGTGTCAGAGGATGGTGGAATTCCCGGTGTAGCGGT
GAAATGCGTAGAGATCGGGAGGAACATCANTGGCGAAGGCGGCCATCTGGGACAACACTG
ACGCTGAGGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACNCCC
TAAACGATGCGAACTGGATGTTGGTCTCAACTCGNAGATCAGTGTCNAANCTAACGCGTT
AAGTTCCCCNCCTGGGGGAGTACGGTCNCAANACTGAAACTCAAAAGGAATTTGACNGGG
GNCCCGCACAACGGTGGAANTNTGTGNTTTAATTNCNATNCNACNCCNAAAAACCTTTNC
CTGGCCNTTGACTTTTCTTGGAATTTNCAAAAATCCNNNAANTTNCCTTTCGGAACNNAA
ACCNCANGTNCTNCATNGNNTTNNTCNNCTTNNTNTCNTAANTTTTGGNTTAANTCCCCN
ACNANCNCNCCTTTTNTCNTTNTTTCCNNNNNNTTTTGTTGGNANTTNTANNNANTTCNN
TTNAAACCCTANTANTNNNNNTATNTTNNTTTNNNNNCTTTNNCCNNGNTTCCCNNNNNT
NTNNNTNNNNNNANGNTTNNTTCCCNNNGTN
>CR57
TTTTTNNNNNNNGNANCGCGTCCTACCATGCAAGTCGAACGGCAGCACAGCAGTAGCAAT
ACTGTGGGTGGCGAGTGGCGGACGGGTGAGGAATACATCGGGACCTGCCCAGACGTGGGG
GATAACGTAGGGAAACTTACGCTAATACCGCATACGTCCTACGGGAGAAAGCGGGGGATC
GAAAGACCTCGCGCGGTTGGATGGACCGATGTTCGATTAGCTAGTTGGTGAGGTAATGGC
TCACCAAGGCGACGATCGATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGA
CACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCT
GATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCAGGA
GCGAAATGCCATTGGTTAATACCCGGTGGAGCTGACGGTACCTGAGGAATAAGCACCGGC
TAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGTGCAAGCGTTAATCGGAATTACTGG
GCGTAAAGCGTGCGTAGGCGGTTCTTTAAGTCTGCTGTGAAATCCCCGGGCTCAACCTGG
GAATGGCAGTGGATACTGGAGGGCTAGAGTGTGTCANAGGATGGTGGAATTCCCGGTGTA
GCGGTGAAATGCGTANAGATCGGGGAGGAACATCANTGGCGAANGCNGCCATCTGTGGAC
AACACTGACGCTTAGGCACNAAANCGTGGGGAGCAAACAGGATTANATACCCTGGGTAGT
CCNCNCCNTAAACNATTCNACTGGANTTTNNTTCTCAACTNGNAAATNNNTTGTCNAANN
TAANNCTTTANNTTCNCCCCCCNGNGGANTNCGGTCNNCNNTAACNNNNAACTCAAANAA
TTTGACTGGNNGCCCCCNCANNCNGNTGNNTATNTTGTTTTATTTTNTTTCNNNCNNTNN
NAAAACNNTTNCNNGNCCTTTNNNNTTNNNNNAATTTTTTNNTNAAATNCNNAANTNCCT
TNNNTGANTCNANTNCANNNTNTTNNTNNTTNNTNNANNCTNNTTTNNTNTNATTNNGNT
NNANTNCNTNNANNNCNNTTNTNTTNTTNNTTTTCNNNCNNTNNNGTTTNAATTTTNTNN
NTTNNNTTNNCANNCNNTANTGTTGNNNTNNNTTNNNNNNTNTNNNTCNTNNNNTCNNTT
NTNNTNTNTNTNTNNNTTTNNTTNNNTTNATCC
>CR6
TNNNNTTNGNANCGCNCCTACCATGCAGTCGAACGGCAGCACAGTCGTTAGCAATACGAT
GGGTGGCGAGTGGCGGACGGGTGAGGAATACATCGGGACCTGCCCAGACGTGGGGGATAA
CGTAGGGAAACTTACGCTAATACCGCATACGTCCTACGGGAGAAAGCGGGGGATCTTCGG
ACCTCGCGCGGTTGGATGGACCGATGTTCGATTAGCTAGTTGGTGGGGTAATGGCTCACC
AAGGCGACGATCGATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGG
CCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCC
AGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCAGGAGCGAA
ATGCCATTGGTTAATACCCGGTGGAGCTGACGGTACCTGAGGAATAAGCACCGGCTAACT
TCGTGCCAGCAGCCGCGGTAATACGAAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTA
AAGCGTGCGTAGGCGGTTTTTTAAGTCTGCTGTGAAATCCCCGGGCTCAACCTGGGAATG
GCAGTGGAAACTGGAAAGCTAGAGTGTGTCAGAGGATGGTGGAATTCCCGGTGTAGCGGT
GAAATGCNTAGAGATCGGGAGGAACATCNGTGGCNAAGGCGGCCATCTGGGACAACACTG

ACGCTGAGGCACGAAAGCGTGGGGAGCAAACAGGATTANATACCCTGGTAGTCCCACGCC
TTAAACGATGCGAAANTNGATNTTNGTCTCACCTCNGAAGATCATTTGTCGAATGCTAAC
NCCGTTAAGTTNTNCNCCCTGGGNANTNCGGTCNCNAGACTGANANNTCAAAAGGANTTN
ACCGGGGGCCCNCTNAAATCGGCGGGATTNTGTNNTTTTAANTTCNAANGCNACCNCNNA
NAACCTTTNCCTTGNCCCTTAATNTNTTCTGNANNTTTTNCCAATAATTNNAAANTNNCT
NNCNGTAANCCTNTAANNNNGGTNNTTTCATGTCTTTTTCTCTCTCNNTTTTTNTANNNT
TTTTGTTTNATTNCCTCCCNNTANCTNNNCCNTNTTNCTTTTTTCCTNCCCTTNTTTTTG
TGNATNCTTNANNANNTNTNAGTTNNATCNTTNATNANNTATNNNTTNATTNGTTNTTTT
TNCTCTNTNNGNCTNNNNCNNNTCNTT
>CR60
CGCGAGNGGGCGNCTTACACATGNCAAGTTCGAAGCGGGCANANGCCCCATATCGGCGNN
ATCTTCGGAGGCTGGCGAACGGGCTGCAACTAACACGATGNAANANCCCTGCCCCCTGTA
CNCATGNNGATAAGCCCATGNAAACTGGNNNTNATACCGTGATATNACCNCCGNCCNAAN
GGNCNAAGNGGTGGAAANTTANTTCGGTTGGGGATGNGCTCGCCNCCTATCACNCTTGTT
GGTGGGGTGATGGCCTACCANAGCGACGACAGGANACCTGNCTGNNNNGNCAACNNGANN
CACTGNGACTGAAANNCGNCNCAGACTCCNNNCNNANGNAACNNTGNGGNATATTANNNN
NTGNGNNGAAGCCTGACCNNANNNCGCCNNGNNGNGNATGACAGCNNTNNCNTTGTAANC
CTNTTTCNTNNTGNACGAAGCTAACTTGNCTNTACNNNNACAAGANGCNCCTGNTANCTN
CTTNCCAANANCCACNGNANTACTTAGGGCNCANGCNTTGTCCAGNANTATTGGNCNTAA
AGAGCTNCTAGAANTNNTNGTCGCNTCTGTCGCGAANGCCCACTGTCTTAACTGTGGGTC
TNNNTNNGNATACTNNCACACTAGAATGNANGTAGGANANTAANGGNATTACCCGTGTGT
ACACGAGTGAAAATGCGAATTATNTNGTGGTAGGANCNCNAAGTGNCTAAAAGGNTCGTT
CATCTAACTANCTGTACTGAAANCTNAGCTANTNGTNANNNCGTNCNCTAGACGANCNAT
GGATTTNGATACCCTGGTTANTCNACGCCGATAAACATTCNNNNGGTTANGGCCATGNNN
NTTTNTTCCATGTGAANTCCGGCATCNTTTNNTCTTTANCCNTTTAAGANTCCCCTTCNN
NNGGNTAATTNCGGTCANNNNNGCTNANACTNNNAAGGGAATTNATCGNGGTNCNCANAA
NCNAACNGGTNGTNGCNATGTTGCTTAATTTCAAACGNNNCGACTAAGAAGCTTTNCCAN
NGTTNNGACNTCGTNANNAANNANNCCNTGAAAATNCCNTGTTCCTTTNTTNGTNCNNAG
AACACGNNANTTCTNTTCNTNTANTTCNANNCCCNGTTNNCTGATATNTNTGGCTCAANT
CNNGCTANNNGNGNGAAANTCTNNTNATNTCNTNNCCATACCNTAANTTTTGGNCNNTNN
TTTNNTANNCCTTTGGTTNTTTNTNTAAGAAGNTTCTGTNCTTCNTCTATTCTNNTTTTA
CTNTCTTTTTTCTGTANNTTTNTTNGTTCTTTNNNANGNCCTTANTTATCTNNCNTTTGN
NTGGCCTATTNCT
>CR61
NTTTTTNNNNNNGNNNCGCTCCTACCATGCAGTCGAACGGCAGCACAGTCGTAGCAATAC
NATGGGTGGCGAGTGGCGGACGGGTGAGGAATACATCGGGACCTGCCCAGACGTGGGGGA
TAACGTAGGGAAACTTACGCTAATACCGCATACGTCCTACGGGAGAAAGCGGGGGATCTG
AAAGGACCTCGCGCGGTTGGATGGACTGATGTTCGATTAGCTAGTTGGTAGGGTAATGGC
CTACCAAGGCGACNATCGATAGCTGGTCTGAGAGGATGATCANCCACACTGGGACTGANA
CACGGCCCANACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCT
GATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCAGGA
GCGAAATGCCNTTGGTTAATACCCGGTGGAGCTGACGGTACCTGAGGAATAAGCACCGGC
TAACTTCGTGCCAGCANCCGCGGTAATACNAAGGGTGCAAGCGTTAATCNGAATTACTGG
GCGTAAAGCGTGCGTANGCGGTCTTTTAAGTCTGCTGTGAAATCCCCGGGCTCAACCTGG
GAATGGCAGTGNATACTGGGAGGCTAGAGTGTGTCAGAGGATGGTGGAATTCCCGGTGTA
GCGGTGAAATGCGTAGAGATCGNGAGGAACATCAGTGGCNAANGCGGCCATCTGGGACAA
CACTGACTCTNANGCACNAAAGCGTGGGGAGCAAACAGGATTATATACCCTGGTANTCCA
CNCCCTAAACGATGCNAACTGGATGTTGGTCTCACTCNNANATCAGTGTCNAAACTNACN
CNTTTAATTTCCCCCCCNGGGGAGTACGGTCNNCNANACTGAAACTCAAAGGNATTTNAC
NGGGGCCCNCACAANCGGTGGANTTTGTTGNTTTTAATTCNNTGCNACNCNAANAANCTT
NCCCTGGNCTTGAANTTTNTTGAANTTCTNCANAANNTNCNNANTTCCCTTTNNNNAACA
AAANNNNNGTTNCTNNTNGNTTTNNCTNNTCTTNNNNNANNTTTTGNTTANTCCCNCNAN
NNNNNNCNNTTTTTCTTTNTTTCCNTNNNTAAGNTNNNNNTNTTNNNAANTNCNNNNTNN
NACNNAANANNTGGNANNNTNNNNNNNNTNCCTTTNNNCNNGNNNNCNNNTTNTNNNNTN

TNTTNNNNTTTTNN
>CR62
NNNNNNNNNNNGNANCGCNTCCTACCATGCAAGNTCGAACGGCAGCACAGNTCGTAGCAA
TACGATGGGTGGCGAGTGGCGGACGGGTGAGGAATACATCGGGACCTGCCCAGACGTGGG
GGATAACGTAGGGAAACTTACGCTAATACCGCATACGTCCTACGGGAGAAAGCGGGGGAT
CTGAAAGGACCTCGCGCGGTTGGATGGACTGATGTTCGATTAGCTAGTTGGTAGGGTAAT
GGCCTACCAAGGCGACGATCGATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTG
AGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAG
CCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCA
GGAGCGAAATGCCATTGGTTAATACCCGGTGGAGCTGACGGTACCTGAGGAATAAGCACC
GGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGTGCAAGCGTTAATCGGAATTAC
TGGGCGTAAAGCGTGCGTAGGCGGTCTTTTAAGTCTGCTGTGAAATCCCCGGGCTCAACC
TGGGAATGGCAGTGGATACTGGGAGGCTAGAGTGTGTCAGAGGATGGTGGAATTCCCGGT
GTAGCGGTGAAATGCGTAGAGATCGGGAGGAACATCAGTGGCGAAGGCGGCCATCTGGGA
CAACACTGACGCTGAGGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGT
CCACGCCCTAAACGATGCGAACTGGATGTTGGTCTCAACTCGGAGATCAGTGTCGAAGCT
AACGCGTTAAGTTCGCCNCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGA
CGGGGGCCCGCACAANCGGTGGANTTATGTGGTTTAATTCNATGCAACGCGNAAAAACCT
TACCTGGNCCTTGANATGTNTGGAANTTCTGCAAAAAATTCCGAAAGTNCCCTTCGGAAA
CNNAACNNAGNTCNTNCNTGGNTTGTCNTCCACTTCNNNNCCTNNAAATTTTGGGTTNAA
TTCCCCCNACNAANCCCACCCTTTTTCTTTTTTNCCCCCCCTTAAGGNTGGAACNCTTAN
GNAACTCCCGGTTAAAAACCNNAGAANGNGGGTTACCTNAANTNTTNNGCCNTTNNNCNG
GNNTNNCCTTTTNANTTNT
>CR63
ANCGCGGCTATCATGCAAGTCGAACGGCTCTTCGGAGCAGTGGCGGACGGGTGAGTAACG
CGTGGGAACGTGCCCAAAGGTACGGAACAACTGAGGGAAACTTCAGCTAATACCGTATGT
GCCCTTAGGGGGAAAGATTTATCGCCTTTGGAGCGGCCCGCGTTGGATTAGCTAGTTGGT
GGGGTAAAGGCCTACCAAGGCTACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACAC
TGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTGCGCAATG
GGCGAAAGCCTGACGCAGCCATGCCGCGTGAATGATGAAGGTCTTAGGATTGTAAAATTC
TTTCACCGGGGAAGATAATGACGGTACCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAG
CAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGGGCGCG
TAGGCGGATATCTAAGTCGGGGGTGAAAGCCCGGGGCTCAACCTCGGAATTGCCTTCGAT
ACTGGGTATCTTGAGTACGGGAGAGGTGAGTGGAACTCCGAGTGTAGAGGTGAAATTCGT
AGATATTCGGAAGAACACCAGTGGCGAAGGCGACTCACTGGCCCGTTACTGACGCTGAGG
CGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATG
AGTGCTAGTTGTCGGCATGCATGCATGTCGGTGACGCAGCTAACGCATTAAGCACTCCGC
CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
TGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCACCTTTTGACATGCCCT
GATCGCCACAGAGATNTGGTTTTTCCCTTCNGGGACAGGGACACAGGTGCTGCATGGCTG
TCGTCAGCTCGTGTCGTGAGATGTTGGGNTTAAGTNCCNCAACGAGNGCAACCCTCNCCA
TTAGTTGCCATCATTAANTTGGGCNCTCTNATGGGACCCCCNGTGNTAANCCNGAAGAAG
GTGGGNATAACNTCANTCNTCAGGGCCCTTTCNGGGTNGGCTCCNNCGTGCTACAATGGC
AACACCNANGGTTNCAACNTNCAANGGGANCTATNCCNTAAANTCCTTNNNTTCGATTNC
CCT
>CR64
ANGCGCGCCCTATCATGCAAGTCGAACGGCTCTTCGGAGCAGTGGCGGACGGGTGAGTAA
CGCGTGGGAACGTGCCCAAAGGTACGGAACAACTGAGGGAAACTTCAGCTAATACCGTAT
GTGCCCTTAGGGGGAAAGATTTATCGCCTTTGGAGCGGCCCGCGTTGGATTAGCTAGTTG
GTGGGGTAAAGGCCTACCAAGGCTACGATCCATAGCTGGTCTGAGAGGATGATCAGCCAC
ACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTGCGCAA
TGGGCGAAAGCCTGACGCAGCCATGCCGCGTGAATGATGAAGGTCTTAGGATTGTAAAAT
TCTTTCACCGGGGAAGATAATGACGGTACCCGGAGAAGAAGCCCCGGCTAACTTCGTGCC
AGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGGGCG

CGTAGGCGGATATCTAAGTCGGGGGTGAAAGCCCGGGGCTCAACCTCGGAATTGCCTTCG
ATACTGGGTATCTTGAGTACGGGAGAGGTGAGTGGAACTCCGAGTGTAGAGGTGAAATTC
GTAGATATTCGGAAGAACACCAGTGGCGAAGGCGACTCACTGGCCCGTTACTGACGCTGA
GGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGA
TGAGTGCTAGTTGTCGGCATGCATGCATGTCGGTGACGCANCTAACGCATTAAGCACTCC
GCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGC
GGTGGAGCATGTGGTTTTATTCGAAGCACCGCGAANAACCTTACCACTTTTTGACATGCC
CNTGATCGCCACANAGATNTGGTTTTCCCTTTCNGGGACAGGGGACACAGGTGCTGCATG
GGCTNTNCNTCAGCTCGTGTCGTGAGATNTTGGGTTAAGTCCCGCAACGAGCGAACCCCT
CNCCATTATTTGCCATCATTTAANTTGGGCCNCTTTNATGGGACCCCCCGGTGTTAANCC
NGNANAAAGTTGGGANTTACTNNANNTCCNNTGGCCCCTTCGGGGGGGNCTACANCNTGC
TTNCATGGCGAANNCNNNGGTTTNCAAACCTTGCAAAGGGAGCTAATCCCTTAAANTCNT
NNATTNCGNATGCCCTCNCNNCCNGGNCCNN
>CR65
NTCCNNANNCNTTNTNCNNNGGGGNCANNGCGTCCCTACACANGCAAGNTCGAACGGCAG
CACAGCAGNTAGCAATACTGNTGGGTGGCGAGTTGGCGGACGGNTGAGGAATACATCGGG
ACCTGCCCAGACGNTGGGGGATAACGTAGGGAAACTTACGCTAATACCGCATACGTCCTA
CGGGAGAAAGCGGGGGATCGAAANACCTCGCGCGGTTGGATGGACCGATGTTCGATTAGC
TAGTTGGTGAGGTAATGGCCCACCAAGGCGACGATANATAGCTGGTCTGAGAGGATGATC
ATCCNCACTGGTNACTNNNATCNNNNCNTTACTNCTTACNTGCGNCANNGNTGNGGATGT
TTATNCCNTNGNNTNCANTACCTCTTAATNTTTGTTNTNTTTTTTTNNNTNNTCTNNTNN
GTTGTNTNNCNTNTNTTTCCNNTNNCCTCTTNCTANTTNTNTTTNANNNANNCGNNNTAT
CCTTTCNNTTNTNNTTNCTTNNTCTNTCNTNCNCTCTTTTNTNTTTTCNTNATCNTTTNC
NCANCNNTNNTTNNNTNATNGNNNNNATNATNNTNTNTCNATGCTNNTNTNCCTTNTTTG
NACTNNTCTTNNNTTTTTNCCCCCNTTNTTNTNCTTTNNNNNCTTCGNNTCNNNNTTNTT
TNTCCTCTANTNACTGTCNNTCNNCCNTCTTGNNCNTTNNANTNNNTNNCCNTTNNNCTN
TNNCGTCNNTCATTTGTNNAAANGTANNNANNTTNNTCTCNCTNTTTGGTCCCNNNTNNN
TTATNATTNTTNNTNNNCNNCTNTNCNTTNNNTNCANTNTNTGCNTTTTNCTNANTTATN
NGTCNCNTTTCNNTCCTTNTATTNTTTTTTNNTNTCTNTNCCCNNNTNNATCTNTCCNTN
TTTNNTTCNCCTCCNCTGTCTTTCCNTANNTTTATGNCGTNTNTTCAANNTTANTNGNNN
NTGCCTTNTCTTTNNTCTCNTTANNTCCCCNCTCNNTATNNCNTTNTCTTNTTNNACNTN
NTTTCTTNTTCNCNGTCGTTCCTCTNTTNNCCNTCANCTTTNNTNNTNNTACTNNNTTAN
NNTCNNTNTTCCTNNNTTNCCCNCNCCTCTTNTTCTNCCNTTNNTTCTNNTCNNNNANNT
NNNNATNCCACANTTANNNNTNCNNANNTTTTTTNTNTTNNATTNCNCCTNTTTTNANTN
TTTTTTNTACCNCGTCCATCTTGNTTNTTTNNCCCTCTNTTTCNCTCANNCNAATCCTTN
TTTTTTTCTNNNCNTNNNTNNNCTCNCNTNTNCTNTNCTANTTATNTGNNNTTCNNTTTT
TNNTNNAATCTNTNNTNTTANANCCCGACCATCTCTNCNTATNCATNGTTNCNTTCTTCT
NTNTTANNNTNNTTNNTACCACTTNCNN
>CR66
TTTTTNNNNNNNNGNANCGCNTCCTACCATNGCAAGNTCGAAACGGGCAGCCACAGCCAG
NTAGCAATACTGTGGGTGGCGAGTGGCGGACGGGTGAGGAATACATCGGGACCTGCCCAG
ACGTGGGGGATAACGTAGGGAAACTTACGCTAATACCGCATACGTCCTACGGGAGAAAGC
GGGGGATCGAAAGACCTCGCGCGGTTGGATGGACCGATGTTCGATTAGCTAGTTGGTGAG
GTAATGGCTCACCAAGGCGACGATCGATAGCTGGTCTGAGAGGATGATCAGCCACACTGG
GACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGC
GCAAGCCTGATCCAGCAATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTT
TATCAGGAGCGAAATGCCATTGGTTAATACCCGGTGGAGCTGACGGTACCTGAGGAATAA
GCACCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGTGCAAGCGTTAATCGGA
ATTACTGGGCGTAAAGCGTGCGTAGGCGGTTCTTTAAGTCTGCTGTGAAATCCCCGGGCT
CAACCTGGGAATGGCAGTGGATACTGGAGGGCTAGAGTGTGTCANAGGATGGTGGAATTC
CCGGTGTANCGGTGAAATGCGTANAGATCGNGAGGAACATCANTGGCGAANGCGGNCATC
TNGGACAACACTNACNCTTANGCACNAAAGCGTGGGGAGCAAACANGATTANATACCCTG
GTTNTCCACNCCCTAANCNATNCNNACTGNATGTTGGTCTCAACTCNNANATTNTTGTTC
AAANCTAANNCNTTTAGTTCNCCCCCNTGGNAGTTACNNTNNCAANNNTTAAACTCNNNN

GNATTTNNCCNGNGCCCNCNNNAACNNTNGNNTNTTTTNTTTTATTTNTTCTATCNNNNN
NNNCNTTTNNTNNCNTTNNCNTNNNNNNATTNNNNNNNNTNNNNNANTTNCNTTNNNNNT
TTNNANNNNNNNTNNNNNTNNNNNNTNNNTTNTTNTNNNNNNNTNTTTTNNNNTNTTTNN
NNNNTNNNNNNNTNCNTTNNTNTTTNTNNNNNNNTNNNNTNNTTTTTNNTNNNTNTNNNN
TNNNTNNNNNNNNNNNNNTNTNTNNTNNTNTTTNNNNNNCCNNNNNNNNNNTNNNNNNN
TNNTNNNTNTNNNNTNTNNNNTCNCNTNNNNNNTNNNT
>CR67
ATGCGCNGCTTACCATGCAAGTCGAACGCCCCGCAAGGGGAGTGGCAGACGGGTGAGTAA
CGCGTGGGAATCTACCCAGTTCTACGGAACAACTGAGGGAAACTTCAGCTAATACCGTAT
ACGCCCCAAGGGGGAAAGATTTATCGGAATTGGATGAGCCCGCGTTGGATTAGCTAGTTG
GTGAGGTAATGGCCCACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCAC
ACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAA
TGGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGC
TCTTTCAGTAGGGAAGATAATGACGGTACCTACAGAAGAAGCCCCGGCTAACTTCGTGCC
AGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCA
CGTAGGCGGATCGTTAAGTCGGGGGTGAAATCCTGGAGCTCAACTCCAGAACTGCCTTCG
ATACTGGCGATCTTGAGTCCGGAAGAGGTGAGTGGAACTCCTAGTGTAGAGGTGGAATTC
GTAGATATTAGGAAGAACACCAGTGGCGAAGGCGGCTCACTGGTCCGGTACTGACGCTGA
GGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACTA
TGAGAGCTAGCCGTTGGGGGGTTTACCCCTCAGTGGCGCAGCTAACGCATTAAGCTCTCC
GCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGC
GGTGGAGCATGTGGTTTTATTCGAAGCAACGCGAAGAACCTTACCAGCCCTTGACATGGC
AGGACGGTTTCCAGAGATGGATTCCTTCACTTCNGTTGACTTGCACACAGGTGCTGCATG
GCTGTCGTCAGCTCGTGTCGTGANATGTTGGGTTAAGTCCCNCAACGAGCCCANCCCTCC
CCTTTAGTNGCCATCNTTTAGTTGGGCNCTNTANAGGACTGCCNGTNATAAACCCGAAGA
AAGGTGGGGANTAACNTCAANTTTTCTTGGCCCTTACGGGTNGGNCTTANNCCCTNNTTC
ATGNNCGGTTACAAANGGGCANCTTTTCCCCNNGACATNNTAATCCCTAAAACCCCNNCA
NTTNGAATGCCCNCNCACCCCGGNCNTAANTNGNANC
>CR68
NANGCGCNGCTACCATGCAAGTCGAACGCCCCGCAAGGGGAGTGGCAGACGGGTGAGTAA
CGCGTGGGAATCTACCCAGTTCTACGGAACAACTGAGGGAAACTTCAGCTAATACCGTAT
ACGCCCCAAGGGGGAAAGATTTATCGGAATTGGATGAGCCCGCGTTGGATTAGCTAGTTG
GTGAGGTAATGGCCCACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCAC
ACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAA
TGGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGC
TCTTTCAGTAGGGAAGATAATGACGGTACCTACAGAAGAAGCCCCGGCTAACTTCGTGCC
AGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCA
CGTAGGCGGATCGTTAAGTCGGGGGTGAAATCCTGGAGCTCAACTCCAGAACTGCCTTCG
ATACTGGCGATCTTGAGTCCGGAAGAGGTGAGTGGAACTCCTAGTGTAGAGGTGGAATTC
GTAGATATTAGGAAGAACACCAGTGGCGAAGGCGGCTCACTGGTCCGGTACTGACGCTGA
GGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACTA
TGAGAGCTAGCCGTTGGGGGGTTTACCCCTCAGTGGCGCAGCTAACGCATTAAGCTCTCC
GCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGC
GGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAANAACCTTTACCAGCCCTTGACATGG
CAGGACGGTTTCCANANATGGATCCNTTCACTTTCGGTGACTTGCACNCAGGTNCTGCAT
GGCTGTCNTCAGCTTCGTGTCNTGAGATGTTGGGTTAAGTCCCNCAACNAAGCGCANCCT
TCGCNTTTATTGCCATCATTTNATTTGGGCNCTCTTNANGGACTGCCCGTTNATNACCCG
GAAGAAAGGTGGGGATTAACTTCAATTNTTCATGGNCCNTACNGGNTGGGCTACCCNNTN
TTNCANTGGCCGTTANNAAGGGNANCTTTCCCNAGGACTNCTANTCCCTAAAACCCNTNC
ANTTNGNATGCCCTNCCACCCG
>CR72
ATGCGCGNCCTANCATGCAAGTCGAGCGGACTTGATGAGGAGCTTGCTCCTCTGATGGTT
AGCGGCGGACGGGTGAGTAACACGTAGGCAACCTGCCTGCAAGACCGGGATAACTAGCGG
AAACGTTAGCTAATACCGGATAATTTATCGCTTTGCATGAAGCGGTAATGAAAGACGGAG

CAATCTGTCACTTGCAGATGGGCCTGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCA
CCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGAACGGCCACACTGGGACTGAGACAC
GGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGGCGAAAGCCTGAC
GGAGCAACGCCGCGTGAGTGATGAAGGTTTTCGGATCGTAAAGCTCTGTTGCCAGGGAAG
AACGACCGTTAGAGTAACTGCTAACGGAGTGACGGTACCTGAGAAGAAAGCCCCGGCTAA
CTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTGTCCGGAATTATTGGGCG
TAAAGCGCGCGCAGGCGGTCGCTTAAGTCTGGTGTTTAAGGCCAAGGCTCAACCTTGGTT
CGCACTGGAAACTGGGTGACTTGAGTGCAGAAGAGGAGAGTGGAATTCCACGTGTAGCGG
TGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGACTCTCTGGGCTGTAACT
GACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCC
GTAAACGATGAATGCTAGGTGTTAGGGGGTTTCGATACCCTTGGTGCCGAAGTTAACACA
TTAAGCATTCCGCCTGGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGG
ACCCGCACAAGCATGGAGTATGTNGGTTTTAATNCGAANCAACNCGAANANCNTTNCCAG
GTCTTTGACNTCCCTNTTANCGGACTANAGNTAGTTCNTTTCNTTCGGAACAAAGAAAAC
NGGTGNTGCATGGTNTNCCTCAACNTCTGTCTTGANATGTTGGNTTAANTTCCCCAACNN
GGCAACCCTNGATTTNNTTNNNNCCCTTTNNGGGGGCCCTCTAAATANTGCCGTTANAAC
CNGAAGAAGNGGGNATNACCTCAATCNNNTGCCCTTTNACCNGGTTCCCCNTNCNAANGN
CCNTTNNACGGGAANNANT
>CR80
GNNNGGNGNCTANCATGCAAGTTCGAGCGGAAAGGCCCCTTCGGGGGTACTCGAGCGGCG
AACGGGTGAGTAACACGATGAGCAACCTGCCCCTGACTCTGGGATAAGCCCGGGAAACTG
GGTCTAATACCGGATATGACCACGGGTCGCATGGCCTTGTGGTGGAAAGTTTTTCGGTTG
GGGATGGGCTCGCGGCCTATCAGCTTGTTGGTGGGGTGATGGCCTACCAAGGCGACGACG
GGTAACCGGCCTGAGAGGGCGACCGGTCACACTGGGACTGAGACACGGCCCAGACTCCTA
CGGGAGGCAGCAGTGGGGAATATTGCGCAATGGGCGGAAGCCTGACGCAGCGACGCCGCG
TGGGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGACGAAGCTAACGTGACG
GTACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCG
CAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGTCG
TGAAAGCCCACGGCTTAACTGTGGGTCTGCGGTGGATACGGGCAGACTAGAGGCAGGTAG
GGGAGAATGGAATTCCCGGTGTAGCGGTGAAATGCGCAGATATCGGGAGGAACACCGGTG
GCGAAGGCGGTTCTCTGGGCCTGTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACA
GGATTAGATACCCTGGTAGTCCACGCCGTAAAACNTTGGGCGCTAGGTGTGGGGTTCTTT
CCACNGATTTCCGCGNCCGTTAGCTAACGCATTAAGCGCCCCGCCCTGGGGAANTACNGG
CCGNCAAGGCTAAAAACTCAAANGAATTTGNACGGGGGCCCNCCACAAACNGGNNGGANC
NTGTTTNNTTAAATTCCNACCCAACCGCCAAAAAACCTTANCCAAGGNCTTGACATCCCC
CGNAAAAACTTCCCAAAAATTCGGGGGTCCTTTTTTGGGNCCCTTNANAGTTGGNNCANN
GCCTATCGTCANCTNTTTTNNTNNNAANTTTTGGGTTAATTCCCNAACNAANCNNACCCN
TNTTTCCANTTTNCCATCNCGTAATGGTTGGGGACNTNCTGTGNAANCCCCCTGGNTNAN
ATTNAAAAAAANTNGGGANTANTNCAATCTTNNNGCCCCCTTATTTNNTGGGCTCNAANN
NCTANANTNNCCNNTCNTAANNNTTTGNAATNCCGNAG
>CR81
TNNNGNTNNANCCACANNAATGCAGTCGACGGCAGCATGAGGTGTAGCAACTACACCTTG
ATGGCGAGTGGCNCANNGGTGAGGAATGCATCGCGAATCTGCCCAGTCGTGGGGGACAAC
GCAGGGAAACTTGCGCTAATACCGCATACGACCTTCGGGTGAAAGCAGGGGATCTTCGGA
CCTTGCGCGATTGGATGAGCCGATGCCGGATTAGCTAGTTGGCGGGGTAATGGCCCACCA
AGGCGACGATCCGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACNGC
CCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGCAACCCTGATCCA
GCANTGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCGGGAACGAAA
CATTGTCGGCCAATACCCGGCAAGACTGACGGTACCCGAGGAATAAGCACCGGCTAACTC
CGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGNAATTACTGGGCGTAA
AGCGTGCNCAGGCGGTTTGTTAACTCTGCTGTGAAATCCCCGAGCTCAACTTGGGAATTG
CAGTGGATACNGGCAAGCTGGANTACGGTAGAGGAAGGTGGAATTCCCNGTGTNGCAGTG
NATTGCGTAGAGATCGGGAGGAAACACCATTTGGCGAAGNCGGCCTTCTTGGACCATTAC
TNACGCTTCATGCACNAAAGCTNTGGGGGAGCAAACAAGGATTATNNTACCCTGNTTAGT

CCNACCCCCNTAANTCGATGGTNGAANTGGTATNTTTGGGGTACATTNCCGTTACTNCAT
NTNTCGAANNNAACNCCTNTCAANGTTTCGCCTNCTGGAGGTANTACNGTTCTNATATAN
CCGAAANCTNCAANGGNATTTANNGGNNNNCCCGTANCANNGCNNNTGGTATTATGTTTG
TCTTNNATTNTNNCCCNCCNNNCAANACCCTTTTCCTGTNCNTNANTTGNCNNNNNATCC
CTNNNTNANATNCCCGNNNTNCCTGANNNANACTNCAANNGCAGNNNTATNNCNATTTTT
CTGTCCCCCNNATTTTNNTTNCNNCTGTNANNNTAAANGCCNNGNTCCTCNTANNNTTTN
TTNANNNTTNTNGCCTATCTNCTTNNNGTTGNGNAATNTTNNNACATACTNNCNNCNCAA
TANNCTGTAATAGNCTNTGTNANCNNNNNCGNTNCTTNNTNCACCTGTATNTCCNTNNCT
CAA
>CR82
NTGCGCNCCTACCATGCNAGTCGAACGGCAGCATGAGGTGTAGCAATACACCTTGATGGC
GAGTGGCGGACGGGTGAGGAATGCATCGGAATCTGCCCAGTCGTGGGGGACAACGCAGGG
AAACTTGCGCTAATACCGCATACGACCTTCGGGTGAAAGCAGGGGATCTTCGGACCTTGC
GCGATTGGATGAGCCGATGCCGGATTAGCTAGTTGGCGGGGTAATGGCCCACCAAGGCGA
CGATCCGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGAC
TCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGCAACCCTGATCCAGCAATG
CCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCGGGAACGAAACATTGT
CGGCCAATACCCGGCAAGACTGACGGTACCCGAGGAATAAGCACCGGCTAACTCCGTGCC
AGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTG
CGCAGGCGGTTTGTTAAGTCTGCTGTGAAATCCCCGAGCTCAACTTGGGAATTGCAGTGG
ATACTGGCAAGCTGGAGTACGGTAGAGGAAGGTGGAATTCCCGGTGTAGCAGTGAAATGC
GTAGAGATCGGGAGGAACACCAGTGGCGAAGGCGGCCTTCTGGACCAGTACTGACGCTCA
TGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGA
TGCGAACTGGATGTTGGGTACATTACGGTACTCAGTGTCGAAGCTAACGCGTTAAGTTNC
GCCGCCTGGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCAC
AAGCGGTGGANTATGTGGTTTTAATTCGATGCAACNCCGAAAAACCTTACCTGGNCNTTG
ACATCTGTCGAATCCTGCAAAAATNCNGGAATNCCCCCAGGNACCANAAAANAGGTGCTT
GCANGGCTTTTCTCACCTCNTGTCCTGANATNTTNGGGTTNAATTCCCCCAACNANGCCC
AACCCCTTTTTCCCTANTTNCCCNCCNCTTAATGGGTGGNAACTNTNGGGGAAACTCCCN
GTTNAAAANCCGNANAAAGGNGGGATTANNTTCAATNCTTTNGGCCTTTNGGCCNGGGTT
ANNNCNTTNTANAATNTNGG
>CR83
ATGCGCNCCTACCATGCAAGTCGAACGGCAGCATGAGGTGTAGCAATACACCTTGATGGC
GAGTGGCGGACGGGTGAGGAATGCATCGGAATCTGCCCAGTCGTGGGGGACAACGCAGGG
AAACTTGCGCTAATACCGCATACGACCTTCGGGTGAAAGCAGGGGATCTTCGGACCTTGC
GCGATTGGATGAGCCGATGCCGGATTAGCTAGTTGGCGGGGTAATGGCCCACCAAGGCGA
CGATCCGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGAC
TCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGCAACCCTGATCCAGCAATG
CCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTATCGGGAACGAAACATTGT
CGGCCAATACCCGGCAAGACTGACGGTACCCGAGGAATAAGCACCGGCTAACTCCGTGCC
AGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTG
CGCAGGCGGTTTGTTAAGTCTGCTGTGAAATCCCCGAGCTCAACTTGGGAATTGCAGTGG
ATACTGGCAAGCTGGAGTACGGTAGAGGAAGGTGGAATTCCCGGTGTAGCAGTGAAATGC
GTAGAGATCGGGAGGAACACCAGTGGCGAAGGCGGCCTTCTGGACCAGTACTGACGCTCA
TGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGA
TGCGAACTGGATGTTGGGTACATTACGGTACTCAGTGTCGAAGCTAACGCGTTAAGTTCG
CCGCCTGGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGGCCCGCAC
AAGCGGTGGAGTATGTGGTTTTANTTCGATGCAACCCGAANAACCTTACNTGGCCTTTGA
CATCTGTCNAATCCTGCAAANATNCGGGANTNCCCNCAGGAACGACAANACAGGTGCTGC
ANGGGCTTTTCTTCNCTCNTGTCCTGANAATTTTGGGTTAANTNCCNCACCAAGNNCAAC
CCTTNTTCCTANTTGCCNNCCCNTAATGGTTGGNAACTNTTNGGNAAACNCCCGNTTANA
ACCCGNANGAAAGGTGGGGANTACNTTAAATTCCTTAGGGCCTTTACGNCCNGGNNTCCC
CCNTNNTAAATGGGTNGNNNA
>CR86

TGCGCNCCTTNCATGCAAGTCGAACGGCAGCACGGGGGCAACCCTGGTGGCGAGTGGCGA
ACGGGTGAGTAATACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCGGA
TTAATACCGCATACGCTCTACGGAGGAAAGCGGGGGATCTTCGGACCTCGCGCTCAAGGG
GCGGCCGATGGCGGATTAGCTAGTTGGTAGGGTAAAGGCCTACCAAGGCGACGATCCGTA
GCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTGTG
TGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTCTGGGTTAATA
CCTCGGGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGC
GGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCGG
TTTGCTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGGCA
AGCTAGAGTGTGGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATG
TGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGCCAACACTGACGCTCATGCACGAAA
GCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAACTA
GTTGTTGGGGATTCATTTCCTTAGTAACGAAGCTAACGCGTGAAGTTGACCGCCTGGGGA
GTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGA
TGTGGATTAATTCGATGCAACGCGAAAAAACCTTACCTACCCTTTGACTNGNACGGAAGN
CCCCTTAAAANGTNGGCTNTNNCTCAAAANAAAACCGTCCCACAGGTNCTGCATGGCTNT
CNTCACCTCNNNTCNTNANATTTTGGGTNAATTCCCCCACCNAGCCCAACCTTTNTTCTT
TNTTTCNNCCCAAGGGCCCNNTAAGAAAACTGCCCGTTTAAAACCCGNAAAAAGGTGGGG
AATTACNTCAATTCCTNTGGCCCTTNTGGGTNGGNTTCCNACTTNATAAATTGTTTGNAA
AA
>CR87
ATGCGCNCCTTNCATGCNAGTCGAACGGCAGCACGGGGGCAACCCTGGTGGCGAGTGGCG
AACGGGTGAGTAATACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCGG
ATTAATACCGCATACGCTCTACGGAGGAAAGCGGGGGATCTTCGGACCTCGCGCTCAAGG
GGCGGCCGATGGCGGATTAGCTAGTTGGTAGGGTAAAGGCCTACCAAGGCGACGATCCGT
AGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAATCCTCTGGGTTAAT
ACCTCGGGGGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCG
GTTTGCTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGGC
AAGCTAGAGTGTGGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGAT
GTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGCCAACACTGACGCTCATGCACGAA
AGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAACT
AGTTGTTGGGGATTCATTTCCTTAGTAACGAAGCTNACGCGTGAAGTTGACCGCCTGGGG
AGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACNGGGGACCCGCACAAGCGGTGGAT
GATGTGGATTAATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGNACGGAAGG
CCNCTAAAAGGTGGCTNTTCTTCAAAAAAAAACCNTCCCACAGTTCTTGCATGGGNTTTC
NTCACCTCTGTCTGNAAATNTTGGTTAATTCCCCANCNAGCNCAACCTTNTCCTNNTTNC
TNCCANGGCCNNTAAGAAACTNCCGTTACAAACCGAANAAGTGGGGATTACTNNATTCNN
NGGCCNTTTNGGTNGGGNTTNCCCNCTANANNGTNGGAANAAAGGTTCCANCCCCCG
>CR88
NGNGCNCNTANCATGCAGTCGGACGGCAGCGCGGGGGCAACCCTGGCGGCGAGTGGCGAA
CGGGTGAGTAATACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCGGAT
TAATACCGCATACGATCTGAGGATGAAAGCGGGGGACCGCAAGGCCTCGCGCTCAAGGAG
CGGCCGATGGCGGATTAGCTGGTTGGTGGGGTAAAGGCCCACCAAGGCGACGATCCGTAG
CTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGA
GGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTGTGT
GAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAAACTTCGTCCCTAATAT
GGATGGAGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGCG
GTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCGGT
GATGTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGCATT
GCTGGAGTATGGCAGAGGGGGGTGGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGT

GGAGGAACACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCACGAAAG
CGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAACTGG
TTGTCGGGCCTTCATTGGCTTGGTAACGTAGCTAACGCGTGAAGTTGACCGCCTGGGGAG
TACGGTCGCAAGATTAAAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGA
TGTGGATTAATTCNATGCAACCCGAAAAACCTTACCTACCCTTNACATGGAACGGAACNT
CNATNAAAAGTTAAGGGTTCCCCAAAAGGGAGCCNTCNCACAGGTGCTNCCATGGCTTTC
CTCCACCTCNTNTNCNTGNAAATTTTGGGGTTAAATTCCCCCAACAAGNNCAACCCTTTT
TCCTGGTTNCTTCCCANAACCCTNCNGGGAAAACTNCCGNTNACNAACCCGAAAAANGGN
GGNATAACCTNAATCCTTTGGCCCTTTGGGNTNGGGTTCCCCTTANNAANGTTNNNAAA
>CR89
ATGCGCTCCTTNCATGCAAGTCGGACGGCAGCGCGGGGGCAACCCTGGCGGCGAGTGGCG
AACGGGTGAGTAATACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCGG
ATTAATACCGCATACGATCTGAGGATGAAAGCGGGGGACCGCAAGGCCTCGCGCTCAAGG
AGCGGCCGATGGCGGATTAGCTGGTTGGTGGGGTAAAGGCCCACCAAGGCGACGATCCGT
AGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAAACTTCGTCCCTAAT
ATGGATGGAGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCG
GTGATGTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGCA
TTGCTGGAGTATGGCAGAGGGGGGTGGAATTCCACGTGTAGCAGTGAAATGCGTAGAGAT
GTGGAGGAACACCGATGGCGAAGGCAGCCCCCTGGGCCAATACTGACGCTCATGCACGAA
AGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAACT
GGTTGTCGGGCCCTTCATTGGCTTGGTAACGTAGCTAACGCGTGAAGTTGACCGCCTGGG
GGAGTACGGTCGCAAGATTAAAAACTTCAAGGAATTGACGGGGGACCCCGCACNAAGCGG
TTGGATGATGTGGATTAATTCGATGCACCNCGAAAAACCTTNCCNTACCCTTGACATGGA
CNGAACCTCCNNTTAAANTTTAGGGTTCCCNAAAAGGGANCCNTCCCCCAGTTCCTCNTG
GNCTTTCTCAACTNCTNTTNCTGAAATTTTGGGTTAANTCNCCNCACCAGNNNANCCTTN
TNCNTGTTNCTNCCNANANCNCNNCNGGGAACCTCCCGNTANAACCCGGANNAAGGGGGG
TTTACTNCNATCCTTNAGGCCCNTTTNGGTTNGGNTTNCCNCNTNNTNANNG
>CR90
ANGCGCNCTTACCATGCAAGTCGCACGGGCAGCTTCGGCTGTCAGTGGCGGACGGGTGAG
TAACGCGTAGGTATCTATCCTTGGGTGGGGGACAACCGTGGGAAACTACGGCTAATACCG
CATGATGTCTGAGGACCAAAGGCGCAAGTCGCCTGGGGAGGAGCCTGCGTACGATTAGCT
AGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCGTTAGCTGGTCTGAGAGGATGATCA
GCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTG
GACAATGGGCGCAAGCCTGATCCAGCAATGCCGCGTGGGTGAAGAAGGTCTTCGGATTGT
AAAGCCCTTTCGACGGGGACGATGATGACGGTACCCGTAGAAGAAGCCCCGGCTAACTTC
GTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATGACTGGGCGTAAA
GGGCGCGTAGGCGGATGGTACAGTCAGATGTGAAATTCCTGGGCTCAACCTGGGGGCTGC
ATTTGATACGTATTGTCTTGAGTCCGGAAGAGGGTGGTGGAATTCCCAGTGTAGAGGTGA
AATTCGTAGATATTGGGAAGAACACCGGTGGCGAAGGCGGCCACCTGGTCCGGAACTGAC
GCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTA
AACGATGTGCGCTGGATGTTGGGTGGCCTAGCCATTCAGTGTCGTAGCTAACGCGATAAG
CGCACCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGC
ACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCANAACNTTACCAGGGCTTGA
CATGGGANGGCCNCGTCCANANATGGGCGTTTCCCCAGGGACCCCTCTGCACAGGTGCTG
CATGGCTTTTCGTCACCTCGTGTCNTGANATTTTGGGTTTAANTCCCNCAACGAGGGCAA
CCCTTNCCTTTCATTTGCCANCCCGTTTGGGNGGGCCCCNNTTAANGAACTNCCCGNTNA
CAANCCNGAAGAAAGGGGGGAATAACCTTCAATTCCNCATGGCCCTTTNTTTCCCGGGCT
ACCNCCTGNCNAAATGGCNGTTNCATTNGNAAACNGGTTCCNAGGCCNAACC
>CR91
NANNGCTCTAANCATGCAAGTCGCACGGGCAGCTTCGGCTGTCAGTGGCGGACGGGTGAG
TAACGCGTAGGTATCTATCCTTGGGTGGGGGACAACCGTGGGAAACTACGGCTAATACCG

CATGATGTCTGAGGACCAAAGGCGCAAGTCGCCTGGGGAGGAGCCTGCGTACGATTAGCT
AGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCGTTAGCTGGTCTGAGAGGATGATCA
GCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTG
GACAATGGGCGCAAGCCTGATCCAGCAATGCCGCGTGGGTGAAGAAGGTCTTCGGATTGT
AAAGCCCTTTCGACGGGGACGATGATGACGGTACCCGTAGAAGAAGCCCCGGCTAACTTC
GTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATGACTGGGCGTAAA
GGGCGCGTAGGCGGATGGTACAGTCAGATGTGAAATTCCTGGGCTCAACCTGGGGGCTGC
ATTTGATACGTATTGTCTTGAGTCCGGAAGAGGGTGGTGGAATTCCCAGTGTAGAGGTGA
AATTCGTAGATATTGGGAAGAACACCGGTGGCGAAGGCGGCCACCTGGTCCGGAACTGAC
GCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTA
AACGATGTGCGCTGGATGTTGGGTGGCCTAGCCATTCAGTGTCGTAGCTAACGCGATAAG
CGCACCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGC
ACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCACCGCGCAGAACCTTACCAGGGCTTGA
CATGGGAGGGCCGCGTCCAAAGATGGGCGTTTCNCAAGGACCCTCTGCACAGGTGCTGCA
TGGCTGTCGTCAGCTCGTGTCNTGAAAATTTTGGGTTAATTCCGCAACNAGCGCAACCTT
CCCTTNANTTGCCAGCACNTTTGGGTGGGCNCTTTAAAGANCTGCCNGGTGACAANCCGG
AAGAAAGTGGGGATTACTTCAATCCCCCTGGCCTTTTNTCNGGGCTNNCCCNTNCTCAAT
GGCGTTACNTGGNAAACNGGTCCCNAGNCCCANCCAATTTAAA
>CR92
NTGCGCGNCCTACCATGCAAGTCGAACGAGAAAGTGGAGCAATCCATGAGTAAAGTTGGC
GCACGGGTGAGTAACACGTGACTAACCTACCCTTGAGTGGGGAATAACTTCGGGAAACCG
AGGCTAATACCGCATAACACCTACGGGTCAAAGGAGCAATTCGCTTAAGGAGGGGGTCGC
GGCAGATTAGTTAGTTGGCGGGGTAATGGCCCACCAAGACGGTGATCTGTATCCGGCCTG
AGAGGGCGCACGGACACACTGGAACTGAAACACGGTCCAGACTCCTACGGGAGGCAGCAG
TGGGGAATTTTGCGCAATGGGGGAAACCCTGACGCAGCAACGCCGCGTGGAGGATGAAGT
CCCTTGGGACGTAAACTCCTTTCGATCGGGACGATTATGACGGTACCGGAAGAAGAAGCC
CCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAGGGGGGCGAGCGTTGTTCGGAATT
ATTGGGCGTAAAGGGCGCGTAGGCGGTTTGGTAAGTCTTATGTGAAATCTTCGGGCTCAA
CTCGAAGTCTGCATGAGAAACTGCCGGGCTTGAGTGTGGGAGAGGTGAGTGGAATTCCTG
GTGTAGCGGTGAAATGCGTAGATATCAGGAGGAACACCTGTGGCGAAAGCGGCTCACTGG
ACCACAACTGACGCTGAGGCGCGAAAGCTAGGGGAGCAAACAGGATTAGATACCCTGGTA
GTCCTAGCCCTAAACGATGATTGCTTGGTGTGACGGGTACCCAATCCCGCCGTGCCGCAN
CTAACGCGTTAAGCAATCCGCCTGGGGAGTACGGTCGCAAGGCTGAAACTCAAAGGAATT
GACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGACGCANCGCGAANAACCT
TACCTGGGGCTCGAATGTAGTGGAATCCGGTANAATATCGGCNCCCANCAATGGGCCGCT
NTATAGGTGCTGCANGGCTGTCGTCACCTCGTGTCGTGANATGTTGGGTTAATTCCNCAC
NAGCGCAACCCTTGTTACNANTTGCTACATTTAGTTNACCCNCTGGTGAAACCCCNCNNA
NACNGGGAAGAAAGCGGGATNACCTCAATCTCANGCCTTTTTTCCGGGCTACCCCNTNCT
ANATGNCCGGTANAACCNTCCAAACCCNTNNNGGGGNCTTTTNNNAA
>CR93
NTTGCGCGNCCTACCATGCAAGTCGAACGAGAAAGTGGAGCAATCCATGAGTTAAAGTGG
CGCACGGGTGAGTAACACGTGACTAACCTACCCTTGAGTGGGGAATAACTTCGGGAAACC
GAGGCTAATACCGCATAACACCTACGGGTCAAAGGAGCAATTCGCTTAAGGAGGGGGTCG
CGGCAGATTAGTTAGTTGGCGGGGTAATGGCCCACCAAGACGGTGATCTGTATCCGGCCT
GAGAGGGCGCACGGACACACTGGAACTGAAACACGGTCCAGACTCCTACGGGAGGCAGCA
GTGGGGAATTTTGCGCAATGGGGGAAACCCTGACGCAGCAACGCCGCGTGGAGGATGAAG
TCCCTTGGGACGTAAACTCCTTTCGATCGGGACGATTATGACGGTACCGGAAGAAGAAGC
CCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAGGGGGGCGAGCGTTGTTCGGAAT
TATTGGGCGTAAAGGGCGCGTAGGCGGTTTGGTAAGTCTTATGTGAAATCTTCGGGCTCA
ACTCGAAGTCTGCATGAGAAACTGCCGGGCTTGAGTGTGGGAGAGGTGAGTGGAATTCCT
GGTGTAGCGGTGAAATGCGTAGATATCAGGAGGAACACCTGTGGCGAAAGCGGCTCACTG
GACCACAACTGACGCTGAGGCGCGAAAGCTAGGGGAGCAAACAGGATTAGATACCCTGGT
AGTCCTAGCCCTAAACGATGATTGCTTGGTGTGACGGGTACCCAATCCCGCCGTGCCGCA
NCTAACGCGTTAAGCAATCCGCCTGGGGAGTACGGTCGCAAGGCTGAAACTCAAAGGAAT

TGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGACNCACCNCGAANAACC
TTACCTGGGCTCGAAATGTAGTGGAATCCGGTAGAAAATATCNGCNNCCCAGCAATGGGC
CCNCTATATAGGTGCTGCATGGCTGTTCNTCNNCTCCGTGTCNNGAGATGTTGGGNTTAA
NTCCCNCNACNAGCNCANCCCTNGTTACCAGTTGCTACCATTTANTTNANCNCTCTGGTG
AAACCCCCTCGAANACGGGGAGAAAGGNGGGGAATNANCTTCAATNCTCTGGGCCTTTTT
TTCCNGGGNTCCNCCNTNNTNAANTGGNCCGGTAAAACCNGTNCNAAACCTNNNGGGGGA
NNTNTCGNA
>CR94
NCGCAGCCTATCATGCAGTCGAACGGCTCTTCGGAGCAGTGGCGGACGGGTGAGTAACGC
GTGGGAACGTGCCCAAAGGTACGGAACAACTGAGGGAAACTTCAGCTAATACCGTATGTG
CCCCTAGGGGGAAAGATTTATCGCCTTTGGAGCGGCCCGCGTTGGATTAGCTAGTTGGTG
GGGTAAAGGCCTACCAAGGCTACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACACT
GGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTGCGCAATGG
GCGAAAGCCTGACGCAGCCATGCCGCGTGAATGATGAAGGTCTTAGGATTGTAAAATTCT
TTCACCGGGGAAGATAATGACGGTACCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGC
AGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGGGCGCGT
AGGCGGATATTTAAGTCGGGGGTGAAATCCCGGGGCTCAACCTCGGAATTGCCTTCGATA
CTGGATATCTTGAGTTCGGGAGAGGTGAGTGGAATGCCGAGTGTAGAGGTGAAATTCGTA
GATATTCGGCGGAACACCAGTGGCGAAGGCGACTCACTGGCCCGATACTGACGCTGAGGC
GCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGA
GTGCTAGTTGTCGGCATGCATGCATGTCGGTGACGCANCTAACGCATTAAGCACTCCGCC
TGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGT
GGAGCATGTGGTTTTAATTCGAAGCAACGCGAANAACCTTACCACCTTTTGACATGCCCT
GATCGCCACANAGATGTGGTTTTCCCTTCGGGGACAGGGACACAGGTGCTGCATGGCTGT
CNTCAGCTCGTGTCGTGANATGTTGGGTTAANTNCCGCAACGAGNGCAACCCNCCCATTA
NTTGCCATCATTAATTTGGCACTCTATGGGACCCCCCGNTGGTAANCCCGAAGAAGGTGG
GGATACNTCANTTCCCAGGGCCCNTACGGGTGGGCTACCCCCTNCTACAATGGCACCANC
NAGGNTNCCAANCTCCAAGGGGAACTATCCNTAAAATNCNCNNATN
>CR95
ANCGCGGCTATCATGCAGTCGAACGGCTCTTCGGAGCAGTGGCGGACGGGTGAGTAACGC
GTGGGAACGTGCCCAAAGGTACGGAACAACTGAGGGAAACTTCAGCTAATACCGTATGTG
CCCCTAGGGGGAAAGATTTATCGCCTTTGGAGCGGCCCGCGTTGGATTAGCTAGTTGGTG
GGGTAAAGGCCTACCAAGGCTACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACACT
GGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTGCGCAATGG
GCGAAAGCCTGACGCAGCCATGCCGCGTGAATGATGAAGGTCTTAGGATTGTAAAATTCT
TTCACCGGGGAAGATAATGACGGTACCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGC
AGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGGGCGCGT
AGGCGGATATTTAAGTCGGGGGTGAAATCCCGGGGCTCAACCTCGGAATTGCCTTCGATA
CTGGATATCTTGAGTTCGGGAGAGGTGAGTGGAATGCCGAGTGTAGAGGTGAAATTCGTA
GATATTCGGCGGAACACCAGTGGCGAAGGCGACTCACTGGCCCGATACTGACGCTGAGGC
GCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGA
GTGCTAGTTGTCGGCATGCATGCATGTCGGTGACGCANCTAACGCATTAAGCACTCCGCC
TGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGT
GGAGCATGTGGTTTTAATTCGAAGCAACGCGAAGAACCTTACCACCTTTTGACATGCCCT
GATCNCCACANANATTTGGTTTTCCCTTCGGGGACAGGGACACAGGTGCTGCATGGCTTT
NCNTCAGCTCNTGTCGTGANATGTTGGGTTAATTCCCNCCAACGAGCGCNACCCTTCNCC
ATTNATTTGCCATCATTAAATTTGGGCACTCTANTGGGAACCCCCCGTTGNTAACCCGGA
AGAAAGTGGGGATTAACCTCAANTTCTCANGGCCCTTANGGGGGTGGGCTCCNCCNTNCT
TNAATGNCCNCCACNNAGGTTTCNAACCTNNCAAGGGANCTTATTCCNTAAANTCNTCTC
ATTCGNATNCC
>CR96
NATGCGCGNCTTANCATGCAAGTCGAGCGAGGCCCCACCTTCGGGTGGGTGTCCTAGCGG
CGAACGGGTGAGTAACACGTGGGCAACCTGCCCCTAGCACTGGGATAACCCCGGGAAACC
GGGGCTAATACCGGATACGACCTCGGAGGGCATCTTCCGAGGTGGAAAGGGTTACTGGCT

AGGGATGGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTAACGGCCCACCAAGGCTCCGAC
GGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCT
ACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGC
GTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCTCCGACGAAGCGAAAGTGAC
GGTAGGAGCAGAAGAAGCACCGGCCAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGT
GCAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCGCGTCGGCT
GTGAAAACCTGGGGCTCAACCCCGGGCCTGCAGCCGATACGGGCAAGCTAGAATTCGGTA
GGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGT
GGCGAAGGCGGGTCTCTGGGCCGATATTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAAC
AGGATTAGATACCCTGGTAGTCCACGCCGTAAACGTTGGGCGCTAGGTGTGGGCCACATT
CCACGTGGTCTGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGAGTACGGCCGCAA
GGCTAAAACTCAAAGGNAATTGACGGGGGNCCCGCACAANCGGCGGANCATGTGGCTTAA
TTCNATGCAACGCGAANAACCTTACCTGGGTTTTGACATGCANGGGAAATCTCNTAGAAA
TACNGGGTTCCGTAAGGGCCTTTGCANAGGTGGTGCATGGGCTGTCNTCANCTTCGTNTN
NTGANANTNTTGGGTTAANTCCCNNNAACNAGCGCAACCCTTTNTCCTTTTTNCCAGCGA
ATATTTTCGGGGNACTCNTAAGAAANCTCCNGGGGTNNACTNTGGAAGNAAGGGGGGGAA
TGAACTTAAATNTTNTTNCCCCTTTTTTCCAGGGGTNTCCNACTTGNTNAAANTGCCCGG
TAAAAANATTNCAAAC
>CR97
NANGCGCGNCTAACATGCAGTCGAGCGAGGCCCCACCTTCGGGTGGGTGTCCTAGCGGCG
AACGGGTGAGTAACACGTGGGCAACCTGCCCCTAGCACTGGGATAACCCCGGGAAACCGG
GGCTAATACCGGATACGACCTCGGAGGGCATCTTCCGAGGTGGAAAGGGTTACTGGCTAG
GGATGGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTAACGGCCCACCAAGGCTCCGACGG
GTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCANCGACGCCGCGT
GAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCTCCGACGAAGCGAAAGTGACNG
TAGGAGCAGAAGAAGCACCGGCCAACTACGTGCCANCAGCCGCGGTAATACGTAGGGTGC
AAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCGCGTCGGCTGT
GAAAACCTGGGGCTCAACCCCGGGCCTGCAGCCGATACGGGCAAGCTAGAATTCGGTAGG
GGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGG
CGAAGGCGGGTCTCTGGGCCGATATTGACGCTGAGGAGCGAAAGCGTGNGGAGCGAACAG
GATTAGATACCCTGGTAGTCCACGCCNTAAACGTTGGGCGCTAGGTGTGGGCCACATTCC
ACGTGGTCTGTGNCGCAGCTAACNCATTAAGCGCCCCGCCTGGGGAGTACGGCCGCAANG
NTCAAACTCAAAGGAATTGAACTNGGGCCCCGCACAAGCGGCGGAGCATGTGGCTTTAAT
TCGATGCAACCGCNNAANANCCTTTACCTGGGTTTTGANATGCNAGGNAAATNTTCTTTA
AANATACCGGGGTTCCCGTAAGGGCCTTTGCNCAGGTGGTTGCATGGCTTTNCTCANCCT
CNTTTCCTGAAANTTTGGGGTTAAATTCCCGGAANAAGNTNAACCCTTTGTCNCTTTTTT
TCCCACNGANNTATTTTCNGGGGAANNCTNAGNNAAANTNCCGGGTTCAACTCCAAAAGA
AGGTTGGGAATAANCTCAANTATTTNTNCCCCTTTTTTNCNNGGGTTNCACCANTTTTTN
AATNNGTCNGGA
>CR98
GGNATGNGCGGCCTNCCATGCAAGTCGAGCGAGAACCCACCTTTCGGGTGGGGGAAAGCG
GCGAACGGGTGANTNNCACGATGGGTAACCCACCCTTGGTACCGGGATAGCCCGGGGAAA
CCCGGATTAATACCGGATGGCCCAACAGCTCTTCGGGGCGGTTGGAAAAGGTAGCTTCGG
CCTCCGACCAAGGACGGGCCCGCGGCGGATTAGCTTGTTGGTGGGGTAATGGCCCACCAA
GGCGACGATCCGTAGCTGGTCTGAGAGGACGATCAGCCACACTGGGACTGAGACACGGCC
CANACTCCTACGGGAGGCAGCAGTGGGGAATCTTGCGCAATGGGCGAAAGCCTGACGCAG
CGACGCCGCGTGGGGGAAGAAGGCTCTCGGGTTGTAAACCCCTTTCAGGAAGGACGAANC
TACTCGGGTTAATAGCCCAGAGGGTGACGGTACTTCCAGAANAAGCCCCGGCTAACTACG
TGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAG
AGCGTGTAGGCGGCCAGGTAGGTCAGCTGTGAAAACTCGAGGCTCAACCTCGAGACGCCG
GTTGAAACCATCTGGCTAGAGTCCGGAAGAGGAGAGTGGAATTCCTGGTGTAGCGGTGAA
ATGCGCAGATATCAGGAAGAACACCCGTGGCTAAGGCGGCTCTCTAGTACGGTACTGACG
CTGAGACGCGAAAGCTTGGGGAGCGAACANGATTAGATACCCTGGGTAGTCCACGCCGTA

AACGATTGGGTGCTAGGTGTGGGGTNTNTCNACTCCNTCCGTGCCNAATCTAACGNTTTA
AGNNCCCCNCCTGGGGNATTNCGGCCGCNANGNTTAAANCNNCAAAGTAANTTGACCGGG
GGNCNNCCNAAANCAATCGAGANNCATGTGGNTTAATTTTCACCCCNACNCCNAAANAAC
NTTTCCCTAGGCTTTGACNTNCACNNGAATTNTTCTAGGNAAACNTTGGTNGCNNNTTCT
NGGGNTTNCTNGTANNANGTTTTTTCNTTGNCTNTTCNTCNACTCCCTNNCTNNAANATT
NGGGTTAATNCCCCCNAANAGNTNTNANCCCNANCNCTTTTTNGTCNNNTNTTAAANTNC
GGTTNCTCTTNGGTGNCTTTCTGGNCAACANTCTTAANGANGTNGGNTNTNNCCNNNTCT
NTTTGNNNTCTTG
>A100
TGCGCNGCNTACCATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAAC
GCGTGGGAATCTACCCATCTCTACGGAACAACTCCGGGAAACTGGAGCTAATACCGTATA
CGTCCTTCGGGAGAAAGATTTATCGGAGATGGATGAGCCCGCGTTGGATTAGCTAGTTGG
TGGGGTAATGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACA
CTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAAT
GGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCT
CTTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCA
GCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCAC
GTAGGCGGATTGTTAAGTTAGGGGTGAAATCCCAGGGCTCAACCCTGGAACTGCCTTTAA
TACTGGCAATCTCGAGTCCGAGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCG
TAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGCTCGGTACTGACGCTGAG
GTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGAT
GGAAGCTAGCCGTCGGCAAGTTTACTTGTCGGTGGCGCAGCTAACGCATTAAGCTTCCCG
CCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCG
GTGGAGCATGTGGTTTAATTCGAAGCAACGCGCANAACCTTACCAGCCCTTGACATCCCG
GTCGCGGATACNAGAGATCGTATTCTTCANTTCGGCTGGACCGGTGACAGGTGCTGCATG
GNCTGTCGTCAGCTCGTGTCGTGAGATGTTGGTTAAGTCCCGCAACGAGNGCACCCNTNN
CCCCTAGTTGCCATNATTAATTTGGCACTCTNGGGGGACTGCCNGTNANAACCCCAAAGN
AAGGNGGGGATNACTTCAANTCCCCANGGCCCTTNCGGGCTGGCTACCNCTTCTNNAATG
GTGGTNANNNTGGNNNCNAACCCCNNGGTCAACTATTTCNAAANCCNTTNNTTTNGATTC
CNCTNNNACNCNGGGCTTA
>A101
NNNNNANNGCGGCTACCATGCAAGTCGTACGATGANNCCCGTTAGGGGTGGATTAGTGGC
GAACGGGTGAGTAACACNNTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACG
GGGTCTAATACCGGATACCACTCCTGCCTGCATGGGCGGGGGTTGAAAGCTCCGGCGGTG
AAGGATGAGCCCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCCACCAAGGCGACGAC
GGGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCT
ACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGC
GTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGAC
GGTACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGC
GCAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCACGTCGGAT
GTGAAAGCCCGAGGCTTAACCTCGGGTCTGCATTCGATACGGGCTAGCTAGAGTGTGGTA
GGGGAGATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGT
GGCGAAGGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAAC
AGGATTAGATACCCTGGTAGTCCACGCCGTAAACGTTGGGAACTAGGTGTTGGCGACATT
CCACGTCGTTCGGTGCCGCAGCTAACNCATTAANTTTCCCCGCCTGGGGNAGTACGGCCG
CNAGGCTAAAACTCAAAGGAATTGANTNGGGGCCCGCNCAAGCGGCGGNANCTTGTGGCT
TTAAATTCTACCCCACCNCNNAATAACTTTACCAAAGGCTTTGNNANTTTNCCGGAAAAN
NATTTAAAAANTANTCNCCCCCCCTNTTGNTNCNTNNNNNNTGGTGGTNNCNTNGCTTTT
NCNTNNACTTCCTGTTTTGNAAATGTTNGTTTTAACCCCCCNANNTAANCNCAACCCTTT
TTNTNTTTTTNCCATAATNCCCTNNTNGGGTNANGGGNNCTCCNTTANACCCCNTGNNTN
CACTNCNTANNAAGTGTTTCNNNTCCCNTACCCTTNCCCCTTTTNTTGNTTTNCTNTTTT
TCCCCGCTTNNCCATNTCTTTTTCTTCTNNCCNTNCCTTNNTC
>A102
ANCGCNGCTTACCATGCNAGNTCGAACGCGAGATAGCCAATACACGAGTTGGCGCACGGG

TGAAGTAACGCGTGGATATCTGCCTTTTGGTTCGGAATAACCCCGGGAAACTGGGGCTAA
TACCGGATGGTTCCTTCGGGATAAAGATTTATCGCCAAAAGATGAGATCCGCGTACGATT
AGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCGTTAGCTGGTCTGAGAGGACG
ATCAGCCACACTGGGACTGAGACACGGCCCANACTCCTACGGGAGGCAGCAGTGGGGAAT
ATTGGACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGG
TTGTAAAGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAA
CTTCGTGCCAGCAGCCGCGGTAAGACGAANGGGGCTAGCGTTGTTCGGAATTACTGGGCG
TAAAGCGAGTGTAGGTGGTTGTCCAAGTTGGATGTGAAAGCCTTGAGCTCAACTCAAGAA
ATGCATTCAGGACTGGGCGGCTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTG
GTGAAATACGTAGAGATTGGGAAGAACACCANTGGCGAAGGCGGCTATCTGGACGGTTTC
TGACACTAAGACTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGC
CGTAAACGATGGGTGCTAGACGTTGGCGAGCTTGCTCGTCAGTGTCGCAGCTAACGCGTT
AAGCACCCCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGGC
CCGCACAAGCGGTGGAAGCATGTGGTTCAATTCGACGCAACGCGCAGAACCTTACCCAGC
CCCTTGACATGGGACTCGCCGGGAAGCANAANACNCTCCCNTTCGGTTCGGCCNGGAGTC
CGCACANGTNCCTNCNTGGGCTGTCCNNCACCTCCTGGTCCTTGANAATNTTTGGGTTAA
NTCCCCNNAACGANCGCAACCCTCTTCTCCANTTGCCATTCAGGTTNTNCNNGGGCNCTT
TNGAAAAACCTGCCNGNTNANAAACCCNNANNAAGGNGGGGATNACTCNAATCCNCNGGN
CCCTNCNGGNTGGGNTNNANCNTTTTNAAATNGNNNTNCAATGNNATCCATNNCNNNANC
CNNACCNATCNNAAAACCNTCNC
>A104
CANCGCNGCTACCATGCAAGTCGAGCGCCCCGCAAGGGGAAGCGGCAGACGGGTGAGTTA
ACGCGTGGGAATCTACCTAGCTCTACGGAATAGCTCCGGGAAACTGGAATTAATACCGTA
TACGCCCTTCGGGGGAAAGATTTATTGGAGTTAGATGAGCCCGCGTTGGATTAGCTAGTT
GGTGGGGTAATGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCA
CACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACA
ATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAG
CTCTTTCAACGGTGAAGATAATGACGGTAACCGTAGAAGAAGCCCCGGCTAACTTCGTGC
CAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGC
ACGTAGGCGGATACTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTT
GATACTGGGTATCTCGAGTTCGAGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATT
CGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGCTCGATACTGACGCTG
AGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACG
ATGGAAGCCAGCTGTCGGCAAGTTTACTTGTCGGTGGCGCAGCTAACGCATTAAGCTTCC
CGCCTGGGGAGTACGGTCGCAAGATTAAAAACTCAAAGGAATTGACGGGGGCCCGCACAA
GCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATC
CCGGTCGCGGATACNANAGATCGTATCCTTCANTTCGGCTGGACCGGTGACAGGTGCTGC
ATGGCTGTCGTCAGCTCGTGTCNTGAAATGTTGGGTTAANTCCCNCAACGAGCGCACCCT
CNCCTTTAGTTGCCTCATTCANTTGGGCCCTNTTAAAGGGACTGCCGTTATAANCCNGAN
GAAGGGGGGAATTACTNCAATCCCCAGGCCCTTACGGGCTGGCTACCACCTTNCTANATG
GTGNTANANNGGCAACNAAACCCCANGTCNANTTTNTCAAANCTTNNATTCNATNCNCCN
CNNCCNNTNNTAATTGAACCCNNTANCCCGAACN
>A109
ATNCGCGNCCTATCATGCAGTCGAGCGAATCGACAGGTGCTTGCACCTGTTTGGTTAGCG
GCGGACGGGTGAGTAACACGTGGGCAACCTGCCTGTAAGACTGGGATAACTTCGGGAAAC
CGGAGCTAATACCGGATAATCCTTTTCCTCTCATGAGGAAAAGCTGAAAGTCGGTTTACG
CTGACACTTACAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAA
GGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCC
CAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAG
CAACGCCGCGTGAGCGATGAAGGCCTTCGGGTCGTAAAGCTCTGTTGTTAGGGAAGAACA
AGTACCGGAGTAACTGCCGGTACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTAC
GTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAA
GCGCGCGCAGGCGGTCCTTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTC
ATTGGAAACTGGGGGACTTGAGTACAGAAGAGGAAAGCGGAATTCCACGTGTAGCGGTGA

AATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGTCTGTAACTGAC
GCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTA
AACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTAGTGCTGCAGCTAACGCATTAA
GCACTCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGGCCC
GCACAAGCGGTGGAGCATGTGGTTTTATTCGAAGCAACGCGAANAACCTTACCAGGTCTT
GACATCCTCTGACACTCCTAGAGATAGGATTTTCCCCTTCGGGGGANAAANTGACAGGTG
GTGCATGGTTGTCNTCAGCTNCGTTCCTGNAATTTTGGGTTAATTCCNCAACCAGCNCAA
CCTTGATTTTATTTCCAGCATTCANTTGGGCNTTTAGGNNACGCCGGTANAACCGAAGAA
AGGGGGGANAACTTCAANTCTCTNNCCCTTNTACCNGGCTACCNCTTTTAAATNNATGNT
NAAGGGCTNNAAANCCNAGTTTACNANTCCAAAACCTTTTNATTGATTN
>A110
ANCGCGGCCTACCATGCAAGTCGAACGGANCACTNCCTTNCTGGTGGATTACGTGGCNGA
ATANCANGATGTANCAACGTGCCGTTTNCTGTTCCTNNTNTNTNGGGACAAACCNNGTGA
NTACTNCCGTCNNATACCCTNNACNGCGNNNTNNNTNTCTNNGATGGATCTTGNNCGCTN
CTGATNAGANAGNTGNTNNNGNAATNTNTCACCTTGNTGACGATNTANTGCTNNNCTGAG
NCGANGANCNGNNNCCCNGNTNCTNAGACANCNNNCNANCTNCTNCCTGAGNCACCNNNN
NGANNNNTTNGACANNGNGCAATNGNNNNTTTTNCNCATNNCGCGANAGTCNTANACGCC
TNCGNCTTGTNAAGNNTTATTGTCNTNGAATNTNANNNCTNTTNCNGCAAGGAATAAANC
CCAAANTAAATTCTACCTACANAANCAAGCGTNCATGCTAAACTACGGNTACACNTTGCC
TCGGAATTACTGTNGNGTANAGGNNTTGTNCGCGAANNTNTNAACTNANNGNTNNNANNC
NNGGNTTNNNCCCTNGAATTNCNTNCCANANTGNTNNNCNTNAGTTCTGNATNCGNTNNT
GGANNTNTAAATNNNGANNTGAAATTNCTAAATNTNNGCANTAACACCAATGTCNAAAGA
TGTCAACTGGNNCNATACTGACGCTGANGCACNNANGCGTNNNNANNAAACNTGATTAGA
TACCCTGTTAGTCCACGCCNTAATNNATTAATNTCAGCCGTTNNGATGCATNCATCTCAN
NNTCGCAGCTAACGCTTTAAGCATTCCGCCTGNGNNAGTACGGTCGCATTATTANANCTC
NNGGAATTGTCNGGCGCCCNCACTAAANNTCGNANCATGTGANTTTATTTCCCAACCAAA
CCCGCAGAACCTTTTCCANCTNTTTACATGTTCCCGGTTTAGATCCCNCAANNAATGTCT
TTTCTTNANTTCGGNNTGGCCTTAAANNAAAGTGCCTCCCCTTGNTGGTCNTCATCNANT
NTCTTNANANCTTTGGNTNAANTNCCNNANNNAGATNATNCNTTNNCNNNNNTTACNNNN
TTCNTTTTGNAACTNTNNGGGNAACTNCCCCTTTTANCNNTNNGNAATNTNGGANTNCTC
NNGTTNTACTTGNCCTNNNGGGGGGNNACCCCNTTTTTATNCNNTTTTNTTTGGNNNNNA
NNNCTTTTTGCTNTCCNAAACCCCNCCNNTCNNTNNNCTTCNNNANNTNNTTCAGNNNNT
TNNNNNTTNTTNTNNTTNCNNNACCTTCNNCTTTNNNC
>A113
AAGCGGGGCTANCATGCAAGTCGAACGATGAAGCCCGTTAGGGGATGGATTAGATGGCGA
ACGGGATGAGTNACANNNTGGGCAATCTTGCCCTTCACTCTGGGACAAGCCCTGGAAACG
GGGTCTAATACCGGATAACACTTNNACATCTCCTGAGCTGGAGGTTAAAANCTCCNGCGN
TGAANGATGAGCCCNCGGCCTATCACCTTGTTGGTGAGGTAATGGCTCACNANGGCNACN
ACNGGTAGCCGGTCNGANANGGCNACCGGCCANANTGGGACTGAGACACNGTCNCAGNAC
TCCTACTGGANGNTAGCANTGGGGAATATTNCNCNCTNNTCGCAANGNTCNGATNNATCC
ACTTCGCNTGTAAGGATGANTNNACTTNGGGNTGNAAACCCCCTCCNNCACGGAAGANNC
NNNAGTNANNTTACCTGCNGNAGATNCGCCNNCNANANTACGTGCCAGCCCCCCTNNNTA
TCCCTAGGNTTTTNNTTNGTCTNTCNCNCCTTTTTNTNTNTTANNNGCTCGNTTNTNGTC
TGTTTCTCTNNNGTTNTNCCTTTCTTGNNNTTNAACCTTNGTTNNNCTAAACTTNCTCTT
NCCNTCTTTCCNNTTCCCTCNNGNTTANCNNNNTCANCCCCCCTTCTTTTNNTCCCTNNN
NCCCNTTCTATTCNCTCNCTCNCCCCCCGTTTTTGCTCCCNCNTTTCCTTNTTCTTCTCT
TTTTTCCTTNTNNTCTTCNTTTCTNCCCCTNTCCCTTNTTTTNNTTNNNGACNTNNTTNT
TNTTCCCGCCCCCCCTCCCCCTTTNNATCNCCNCCCCTACTNCTTCTCTNCCANATCTCN
ATANCNTTTTTCCTTCCACGTCCCCTTACNNNTCCCTTATCNTTCCACNCCCCCCCATNT
ATNCCTNCTCTTCNCNNNNTTTTTNTCNCCCTCTANCCCCNCGNCTTTTNTTNTCNCTCN
TCCNTACCGCCTCANTTNTCANNTNTGNACTTCNCTGTNTTNTTTCTTNCTACTCCTNNC
TCTTCNTCNCNCTCNTTTNCTTTTATTNTTTCCCACCACCCTTNCCNCCCTTATTCNNCT
CTTTTNCTTTTTTTTTNTTCCCNTACTNCCCGCCTTTCCTCCCTTCTNTCTANCCTNNTC
TTCATTNNNTNTCTCCTCTTANNCNNCCNCTACNTNNCCNCCTCCCTTTTTTCCTTCANT

TTANNNATTTNTNTTNTCANCTNCCNCTCCTCCTCCCCNCCCCTCNCCNTTTTNNGCNNT
TTANTTTCTTTCGTTTTTCNANNTGTTCNCTTTCNTCTNCCCCTTCGNCCCTTTTCACTT
TTTNTATATNTTTCCNCCCNTTNCCCCTATTTTTTTTTTTTTCCTTNCCCTCCCCNTNCNC
C
>A115
ANGCGCGGCATAACATGCAGGTCGAACGATGAAGCCCTTAGGGGTGGATTAGTGGCGAAC
GGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTC
TAATACCGGATAACACTTCCACTCTCCTGAGTGGAGGTTAAAAGCTCCGGCGGTGAAGGA
TGAGCCCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTA
GCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACNCCGCGTGAG
GGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCNAAAGTGACGGTAC
CTGCAGAANAAGCGCCGGCTAACTACGTGCCAGCANCCGCGGTAATACGTAGGGCGCNAG
CNTTNTCCCNAATTATTGGGCNTAAAGAGCTCGTAGGCGGCTTGTCACGTCNGTTGTGAA
ANCCCGGGGCTTAACCCCNGGTCTGCAGTCTNATNCGGGCANGNTANNANTTCNGNATGG
GANATNNCNAATNCCTTNTNTTTCCCNTTTAATGTTCATATNTCNCCTANCAACANTTTT
TNNTCNAAAGCTCGATNNTNCTTGTTTNACACTTTANCNTGTTCTNTCCCCNNNCCCTTT
TTNGNTAACNTCNATTTCTNTTTCCCCNGTCNCTCNCNCTTNTTNANNCCNNCGTGTTAC
TNNGTTNNNGGNNANANATNTTTCATNNTTTTCCCNTGTCCCCANTTAANCTCTTTTTAC
TNCCCCCCCTTTNTTTANTCCCTCCNTTCCATTTTTNCCNCTNTNANTNNATNCTTCTNT
NNTNCCCNNTCCTTNTCACCCCNNTTNTTNCGTCCTTTNTCCTNTTTTTNTCTNTANACC
CCTTTNCCTCCCCCTNCTTNTTTTCTTNTACCTCCTCGNTATCNCTNTTNNCTCTTTTAT
TCTTTTTNTTTTTTCCCCCTCNNCCCCTTTTATCCTCCNTTTTNTTTTACTTTCTCTTNT
NTNCNTTTCTNCCCNNTCTCCCNCTNCTNCANTTNTTTCCCCCCTTTTTTTTCTCTNTNC
TNTNCNTTCTTTATTNNTCTTTTCCNCTNCCTTCTCNTCTCNCTCCCCCNCCTATNTTTT
NNNTNTTTCTTCNNNNTCTCTATTNCCTTNTTTTACCCCACNCCTCCCCCTTCCTNCCNT
CTTNCCNTTTTNTNATTTCTNTNCCCCCNTTCTNCCCTTTTTTTTTTTTTTTCCTCNTCC
TCTCCCCCTCCCCT
>A116
NNNGGCAGNGGCGTCNTACCATNGCAGNTCGNTACCAAGGATAACCGNTACAACCGTGGC
ATACGTAGGTNGTAANACGGTGNGGAAANCATGCTNTNNCTGANCGTTGNCCCTTNNNGT
TGGGGNAATAACTTGNGTCNAAACGNGCATNCTAANTACCCGCNTTNCAACTCATGTNTN
GNNNCAAAGCNCGAGGGATCTGCAAAGGACCCTCTNAGCATATTGTGAGTCGGGCCCATA
GTNNCTATTNAGGCTANGTTGGATGGGGGTAAANGNCTCNCNAAGNCNNNCNNCTGTNAC
TGGNCCTAAAAGACGACCNGNCCNNCTNGGNCTNNAACNNGGGCCNAANTTCTNCCGGNG
GNNGCANTGGNGAATTTTGNANAATGGGNGCAACCNTNATNCNTNCATGNCGCNTGCGNN
ANNAACGGCCTTTCGGNTTGTAAACCTCTTTTGNCANGNAAGAAAAGACTCCTGNTNATC
CCNGGGGTTCATGACCGTACCTGNACAANNCTNACCCGNTANCTANGNGCCNGNAGCCNN
CGTNNTCCGTANGTTGCNNGCGTTAATAGNNCTTACNGGNCTTNNNGCGTGCNNNGTNNG
ATANNNAAGATTTATGTNAAATCNNCGNNCTCTACCTGNTAACTGNATTNGGGANTGNNT
GANTAGATTNCNGCNNATAGCGATGGAATTCNCAGNTNTCNNGTNNAANNCTTATATNTA
NGGANGATCNCNTNNTGGATTACNCANGCTCATGTGCCTGTGCTGNCNCTCNTGCACANA
ATTTTTGCCANCCANCAGGANTATATACCCTNGNNGTCCTAGCNCTTGCCANCTNNACTG
GTTGTCGGGGATGCTTTCTTCTCATTAANTTTCCCCACCNNNTNAAGTCNCNCCCCCTTN
GTTTAATACCNCNCCCNTAGTTTTNNTTTCCCCCGNACNTNACGGCGNACTNNCNACNCC
TTGTNNNNTNANATTTTTTAATTCCNTTNNNCNGCNCNNCCTTTACNTNNATNTNNTNNN
CNTTNNCTTCNTTCTTNTTTTNTTCTNTTNTTNTTTCCNTTTCTCGCTTTTTCCTCCNTT
NTTTCTNTCCTNNTCTTTNNTTTTTNAACCTCTTTTNCNCCTCTCNNTTNTCTCNCCCCT
TTTNTTCTNTNTNTNTNTTTCTNTCTTNTTTNTCNCCNTNCNTNNNACCNNCNTNNNCNC
NNTTTTNNNCTTNTTTTNCTCTCNCTTTTTCNNNTNTTTCCCCCTNCCCTNCCCCTCCNN
CNNNNNNTNNNTTTNTTNTCTTNTTCCCCCCNNTTTTCCNNTTTTTTTTTTTTTCCTCTN
CTCTNCCCTTCCCNTN
>A117
TNGGAGGCGNNGCTTACCANGCAGNTCGANACGCCCGCTCACNGGTATAGCGGCAGACGG

GTGTAGTTAACGCGTGGGCAACCTACCCNTCACTACGGAACAACCTCCGGGAAACTGNGA
GCTAATACCGTATACGCTCCTTCNGGAGAAAGATTTATCGGTGATGGATGAGCCCGCGTT
GGATTAGCTAGTTGGTGGGGTAATGGCCTNCCAAGGCGACAATCCATANCTGGNNTGAGA
GGATGATCANCCNCACTGGGACTGAGACACGGCCCANANTCCTACGGGAGGCAGCATNGG
GGAATATTGGACAATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCC
TAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGGTNACCGGAGAANGAANCCC
CGGCTAAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATN
ACTGGGCGTAAAGCGCACGNNGGCGGACTCTTAAGTCAGGGGTGAAATCCTGGNGCTCAA
CNCCGGAAACTGCCTTTTGATACTGAGTAGTCTCGAAGTCCGGAGAGAGGTGAGTGGAAT
TCCGAGTGTAGAGGTGAAATTCGTANATNTTCTCTANAACACNAGCTGGTCTAATNGCNG
CCTCACTGGCCTCNGTACTGACGCTGATNTGCNAAAAGTCCTNGGGGAAGCANNCNGGTT
TCNCNTACCCTTGGTAGTTCCACGCCNTGTAAACCTNNGGAAAGCTNCCCCNTCTNACAN
CTTTNNCTTTTTCTGGTGGCCGTATCTTAACNCATTTAAGCNTCCCCTCCCNGGGGGANT
TNCNGGTCGCANANATNNACANCTTTTTATTCANTTGNCGGGNNTTCNCCCCCCAANCCG
GTCGGAANATTTTGGTTCATTCCTAATCCNCCCTTCATACANCTTCCNCATNTCTTTTNA
TTTNNCTCNTCNNGNTCTNCCCNATNATNTCNTTTTTTTTACTTCCCCCNGTTGCCNNTN
ANAGTCCCTTTTTNNTCTTTTTTTTNCCCCTNCCCTTTTTTNCCTCGNTTCTNTNTNCAC
ANCNTTTNCTTNATCTTTCCCCTTTTTCACCTCTCTCTTTTCTCNCCCTNTTNTCTNCNN
CTCTTNTTTTCNTTTTTNTTNTCCTTCCTCCCCCCCCTCCNCCCTTNTCNCTTNTTTCAT
NNCCTNCTTTTCCCTNTTTTCCCTCCCCTCCCCCTCCTCCCTTTTTNTCTTTNTCTTTTT
TTTCCCCCCTTTNTCCCCTNTTTTTTTTTTTCCTCCNCCTCCCCCCTCCCNC
>A119
TAGGGGCTTACATGCAGTCGAGCGCCCCGCAAGGGNAGCGGCAGACGGGTGAGTAACGCG
TGGGAATCTACCCAGCTCTACGGAACAACTCCGGGAAACTGGAGCTAATACCGTATACGT
CCTTCGGGAGAAAGATTTATCGGAGTTGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGG
GGTAATGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACACTG
GGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGG
CGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTT
TCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCA
GCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTA
GGCGGACTATTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATAC
TGGTAGTCTCGAGTCCGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAG
ATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCGGTACTGACGCTGAGGTG
CGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGA
AGCTAGCCGTTGGCAAGTTTACTTGTCGGTGGCGCAGCTAACGCATTAAGCTTCCCGCCC
TGGGGAGTACGGTCGCAAGATTAAAAACTCAAANGNATTGACGGGGGCCCGCACAAGCGG
TGGGAGCATGTGGTTTTATTTCGAAGCAACGCGCANAACCTTTACCAGCCCCTTGACTTC
CTGGGTTCGCGGTTTCCGANGAGATCGGANTCCTTCAGTTNCGGCTGGACCNTTTANAAG
GTGCCTGCATGGCTTTTCTCACNTCNTNTTCTNGNGATNTTTGGGTTAATTCCCCCACCA
AGCGCAACCTNGCCCTTTATTGNCNTNTTTAANTTGGGCACTCTAAAGGGACTNCCCGGN
TANTACCCCAAAGGAANGGNGGGANTTACNTCAATTCCTCTGGCCCTTNCGGCTNGNNCN
CCCCCCNTCNCNAANGNTNTTNTTGGNCNTNNNTNNCCTCCATNCCCACTTNNTCCAAAT
CTTTTTNTTTNATTNCCNTTCCNTCNGTGG
>A120
ATGCGCNGCTTACCATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAA
CGCGTGGGAACGTACCCTTTACTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTAT
GTGCCCTTCGGGGGAAAGATTTATCGGTAAGGGATCGGCCCGCGTTGGATTAGCTAGTTG
GTGGGGTAAAGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCAC
ATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAA
TGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGC
TCTTTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGCCCCGGCTAACTTCGTGCC
AGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCA
CGTAGGCGGACATTTAAGTCAGGGGTGAAATCCCAGAGCTCAACTCTGGAACTGCCTTTG
ATACTGGGTGTCTGGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTC

GTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTACTGACGCTGA
GGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGA
TGAATGTTAGCCGTCGGCAAGTTTACTTGTCGGTGGCGCANCTAACGCATTAAACATTCC
GCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCCCCACAAG
CGGTGGAGCATGTGGTTTAATTCNAAGCANCNCNCANAACCTTACCAGCCCTTGACATCC
NGTGCTACATCCNAAAATGGATGGTTTTCCTTCNGGAACCCCNANACAGGTGCTNCATGN
NTTTCNTCANCTCNTNTTCNTGANATTTNGGGTTAANTCCCNCAACNAGCGCNACCNTCC
CCCNTNATTTNCANCNTTNNNTTGGGCNNNCTAAGGGANTNCCCNNTTTTAACCCTAANG
AANGNNGGGATATCTCAANTCCTCAGGCCCTTAGGGCTGGGNTCNNNCNNNCTNANNGNG
TNNNNNGGCNCNACCCCCNNNTTNANTNTTTNCAAANCTTTNNNTNCNNTTCNTTNNCNT
TNNNTNNTAANTTNNANCCCTNNACCCNNACCC
>A124
NNNNNNANNGCNGCTTACCATGCAAGTCGAGCGCCGGTAGCAATNAGGAGCGGCAGACGG
GTGAGTAACACGTGGGAACGTACCTTTTGGTTCGGAACAACACAGGGAAACTTGTGCTAA
TACCGGATAAGCCCTTACGGGGAAAGATTTATCGCCGAAAGATCGGCCCGCGTCTGATTA
GCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGA
TCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATA
TTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGT
TGTAAAGCTCTTTTGTGCGGGAAGATAATGACTGTACCGCAAGAATAAGCCCCGGCTAAC
TTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATCACTGGGCGT
AAAGCGCACGTAGGCGGACTCTTAAGTCGGTGGTGAAATCCTGGAGCTCAACTCCAGAAC
TGCCTTCGATACTGGGAGTCTCGAGTTCGGGAGAGGTGAGTGGAACTGCGAGTGTAGAGG
TGAAATTCGTAGATATTCGCAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCGATACT
GACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCC
GTAAACGATGAATGCTAGCCGTTGGTGGGTTTACCCTTCAGTGGCGCAGCTAACGCTTTA
AGCATTCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCC
GCACAAGCGGTGGAGCATGTGGTTTTAATTTCGACGCAACGCGCAGAACCTTTACCAGCT
CTTGACATGTCCAGGACCGGTTCGCAGAAATGCGGNTCTTTCCCTTCNGGGGCCTGGAAC
ACAGGTGCTGCATGGCTGTTCGTCACCTCNNTGTCGTGAGATGTTTGGGTTAATTCCCCC
ANCCGANCNCAACCCCCGTTCTTTAGTTGCTTNCCCNTTTTTTTAACCCCCNTAAGGAAC
TNCCGGTNANAACCCCCNAGAAAGGNGGGNATNANCTTAATTCCTCTANGGCCTTNCNGG
GCTGGNNNACCNCTGGNNAAATNGNNNNTANCAAAGGNTTNAAAGGTCCACCCTTCCAAA
TCNCAAACCCCNNCTNTNNNNTTTGNNTTTANCTCNCCCCNNAATTGANCTTTNTTTTNN
ANNCCTTTCGGTAAACTTCCGGC
>A126
AGNCGGGGGTCTTANACAANGCAAGNTCGGANCGANNGCAANCACATTTNNGGNTGAGGC
GATTAGATGGGCGNANGGTNATGGAGGTANACACAANGGTGCCANTCTGCCCCTTCACCT
CATGAGGAANAANANCTNGNGAANNAGTAGCTCGTAATACCNCCATNCCGACNCTNATCA
TNTNATTCGCGCAGTAAAGGNGTNGGGAANCGCCTCNCTGCATGTAGAATAGATTGAGTG
CNGGCTAACCTACTTCACCTTGNGTGGANGAGTCTAACAGGCTNCGANCTGAGNGGANTG
ANTGAGTACACCGTTGCTGACTGNGGACAACCGGCCACACTGCTAACTGAGACNCGCACC
NNGGACTNTATANGGACAGGTGGNGGNGANCCANTATNNNACNATGCGNCCGANGNCNNN
TTNNAGCCNNACCGNTNNANNGNTNNNTNNNTTCGGNTTATAANCNNCTTTCCGCANGAA
TAAACCCNANNTNACTTNACNTNNACAACAANNGCCANNTAANTNCGNNTNNCCTTCCTC
GGAATTACTNAGGGCANAAGNTTTGTCCGNAATTNTTTAACTTAAGGATNTCATNCTGGG
TCTGTCCTCNCGAANTNCAANTCCNNNTGNTTANCNNCAGTTCTGCATTCGATACGTGGC
AACNCTAGAGTGTAGGANAGTGGAAATATCGTAAATTTCTGNTGTAACNCNNANNTGCGA
AAGATATCTCACTGGAACCACCTACTGGACGANGGANGNATCTNATGGGTNATTACTGAA
CNCCTGATTAGNTACCCCGTGTGGGACNNAACANGATTAATTACCNTGNTNCNNNACTNN
GTTTACTCTNTNANNTNNGCTGNTGACGATTTNNCNCTTCCTCNNNGCCAGTACTNNCGC
NNNATTNTNACTCCCNGGANTTTTACGGTCGCNCGNNTAANNNTTANNANGAATTGNTTT
TNTTTCCACCNCNACNTNCNNAACCTTACNNTTTNNTTTCNACTTNTACCNNGAAAANTN
TTTACNAATGNTTTCCNCTTCCTTCNAAAACTNNNANANANGGTCNNCCCNTTNTGNTTN
NCTTTCNTNNTNTGNANTGTTTNCTTTANCCCTTTTCNTNNNTTNTNNNNTTCTTNCTTN

NNNANTNANNNCTNNTTTNTTNTNNCNCNNTTNNCTTNAGCNTTNGGAATTCNNNNNAAN
TNTNTNTTCTTNNNNANNANNNTNGNNCNNCNCNNNCCNTTNTNTTTNNTTNNTTNNNNN
TTTTNNTNNNNTTCNNNNNNTNCNNCTNNTNNNNTTNTNNNNTNTTTTTTNTTNNNNCTT
NCCCNTNTTTTTNTTTCTNTCCTTCCCTTNCNN
>A127
AGGGCAGCATANCATGCAAGTCGAACGATGAACCACTTAGGTGGGGATTAGTGGCGAACG
GGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCT
AATACCGGATATCACTTTCGCAGGCATCTGTGAGGGTTGAAAGCTCCGGCGGTGAAGGAT
GAGCCCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAG
CCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGA
GGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGG
GATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGACGGTACC
TGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGC
GTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCACGTCGGGTGTGAAA
GCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCTAGCTAGAGTGTGGTAGGGGAG
ATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAA
GGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGATT
AGATACCCTGGTAGTCCACGCCGTAAACGGTGGGAACTAGGTGTTGGCGACATTCCACGT
CGTCGGTGCCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGGAGTACNGCCGCAAGGCTA
AAACTCAAAGGAATTGACGGGGGNCCCGCACAAGCAGCGGANCNTGTGGCTTAATTCGAC
NAACNNCNAGAACCTTACCAANGCTTGACNTACACCCNAAAACGNTCAGANATGGTCGCC
CCCTTGTGGTCGGTGTACAGTGGTGCCATGNCTGTCGTCANCTNTTNTCNNNAATTTTNG
GGTTAAANTCCCNCANCAAGCCNNAACNTTTTTCCGNGTTTCCTANNATGCCCTTCTGNG
TTANTGGGNACNNCAAGGAAAACCCCGGGGTTNACCCNAAGAAGGNGGGNCCACTCCANT
NTTTTTCCCCTTATTTCTGGGTNCCCAATTNTNCAATGNCCGTCNATNACCNCNANCTNN
AGNTTANNATNNTTAAAATTTTTNTCCCCCNTTGGTTCCNTNTTCCTTTNTTCCTTNCTT
CTCCANTCNNTN
>A133
ATCGCGGCTATCATGCAGTCGAACGGTNAGTAGTGTAGCAATACATTGCCTAGAAGTGAC
GTAAGGGTGCGTTAACACGTATGCAATCTGCCCTGTACAGGAGTATAGCTCCCCGAAAGG
GGAATTAACCCTCCATAGTATAATTGAATGGCATCATTTGATTATTAAAACTGAGGTGGT
ACAGGATGAGCATGCGTCTGATTAGCTAGTTGGTAGTGTAATGGACTACCAAGGCGATGA
TCAGTAGGGGAACTGAGAGGTTGATCCCCCACACTGGCACTGAGATACGGGCCAGACTCC
TACGGGAGGCAGCAGTAGGGAATATTGGTCAATGGGTGAGAGCCTGAACCAGCCATGCCG
CGTGCAGGAAGAAGGCCTTCTGGGTTGTAAACTGCTTTTGCCAGGGGATAAAACGGGAGT
GCGCTCCTAATTGAAGGTACCTGGTGAATAAGCCACGGCTAACTACGTGCCAGCAGCCGC
GGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGG
CTCCTTAAGTCAGTGCTGAAATACTCTAGCTTAACTAGAGGGGTGGCATTGATACTGAGG
AGCTTGAGTAGAGTCGAGGTAGGCGGAATTGACGGTGTAGCGGTGAAATGCTTAGATATC
GTCAAGAACACCGATAGTGTAGACAGCTTACTAGGCTTCAACTGACGCTGAGGCACGAAA
GTGTGGGGATCAAACAGGATTAGATACCCTGGTAGTCCACACTGTAAACGTTGATTACTC
GCTGTTGGCGATATACAGTCAGCGGCTTAGCGCAAGCGATAAGTAATCCACCTGGGGAGT
ACGCCGGCNAACGGTGAAACTCAAAGGAATTGACGGGGGTCCGCACAAGCGGTGGAGCAT
GTGGTTTNATTCGATGATACNCGAGGGAACCTTACCTGGGCTAAAAATGCCCTTGATNTC
CCNCANAGACNAAGGAGTTNCTCCAAGNACAAGGAGCAAGGGTGCTNCATGGTTNTNNTC
ACTTCNNGCCGNNNAGTGTTGGGTTTAANTNCCCCCACCAGCNNCACCCTTTNTTTTNNT
TTCCANCNGTTTTTNNTGGGGACNTTAAAAAATGCCTCCCAANCAAAAAGAAGGNNGGGA
TNACTTCAANTNNTNTTGGCCCTTCCCCCAGGGCCTCNCCCTTCNNNANGNCTAACNNAN
NNTNAAATTGGNACATAACCCNTNNNAAAATCCCCCNATTCNAATNAGGNTTATNTCTGC
NGAATNNAACCNCTAACCCCC
>A138
NNNNNNGANGCGNGGCNTANNATGCAGTCGNACGATGAACCACTTAGGTGGGGATTAGTG
GCGAACGGGTGAGTAACACGTTGGGCAATCTGCCCTGCACTCTGGGACAAGCCCTGGAAA
CGGGGTCTAATACCGGATATNAGNNCCTCTNCTCATNNTTGGTTCTGNAAGCTCCNGCGG

TGAAGGATGAGCCCNCGGCCTATCANCTTGTTGGTGAGGTAANGGCTCACCAAGGCNACN
ACAGGTAGCCGGCCTGAGANGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTC
CTACNGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACNCC
GCGTGAGGGATGACNGCCTTCGGGTTGTAAACCTCTTTCANCANGGAAGAAGCGAAAGTG
ACGGTACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGG
GCGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCACGTCGA
TTGTGAAAGCTCGGGGCTTAACCCCGAGTCTGCAGTCNATACGGGCTAGCTAGAGTGTGG
TAGGGGAGATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCANGAGGAACACCG
GTGGCNAAGGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGA
ACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGGTGGGNACTANGTGTGNGCAACA
TTCCACGTTGTCCGTGCCGCAGCTAACGCATTAAGTNCCCCGCCTGGGGANTACGGNCCG
CAANGCTAAAACTCAAANGAANTGACGGGGGGNCCGCACAAGCAGCGGAAACATTTTGGC
TTTAATTCGACGCAACGCTTAAGAACCTTACCAAGGCTTGACATACACCNGGAAACGTTT
AANAGATGGGCNCCCCCTTGTGGTCNGTNTACAAGTTGTTGCATGGCTTNTCCGCCAATT
NCNTNCCTGANATGTTTGGGTTAATTCCCCNAACCATCGCAACCCTTTNTTCCGTTTTGN
CAAAAGGCCCTTTNNGGTGCTNGGGACTNCNGGGAAACNCCCCGGTNAANNGNANNAAGG
GGGGGANANNNNAATNTTNTNCCCTTTNTNTTTTGGTTTNAAACGGCTNAATGCCNNACA
ANNAACCNCTNCCNCNGGGGGAGCAATTNCAAAANNNNTNCNTNGANTGGGCTTNNNTTC
CCCTTANCCTGAGTCNTTTTACTCACTNC
>A140
GATGCGGGCTAAATGCAGTCGTAACAAGGGTNACCGATAGCTTGCTCTTATGAAGTTAGC
GGCGGACGGGTNAGTAACACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAA
CCGGGGCTAATACCGGATAACATTTTGCACNGCATGGTGCGAAATTGAAAGGCGGCTTCG
GCTGTCACTTATGGATGGACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCA
AGGCNACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGC
CCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGA
GCAACGCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAAC
AAGTGCTAGTTGAATAAGCTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACT
ACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTA
AAGCGCGCGCAGGTGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGG
TCATTGGAAACTGGGAGACTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGT
GAAATGCGTANAGATATGGAGGAACACCAGTGGCGAAGGCNACTTTCTGGTCTGTAACTG
ACACTGAGGCGCGAAAGCGTGGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCC
GTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTNAGTGCTGAAGTTAACGCA
TTAANCNNCTCCGNCCTTGGGGGAGTACCGGCCGCCAAGGCTTNAAACTCAAAGGAATTN
ACGGGGGCCCGCACAAGCGGTGGNAGCNTGTTGNTTTNATTCNAAGCCANCCCGTAAAAN
CCTTNCCAGGTTTTNNCCTTCTTCTGACACCCCTANAAATAGGCTTTCCCCNTNNGGGAA
CAAATTGCCCGGTGGGTGCATGGNTGTCTTCCGCNCCTGNCNNGGANATNTTGGNTTAGT
CCCCCAACAAGNCNACCCNTGANCTTNGTNCCCNTATAAATTCGGCCCNCCTACGTTTNC
TNCCGTTNNAANCTNGAGANNGTNGGGGTTTNATCCAANTTTTTNCCCNTTTNNNCCGGG
TTTNCCTTTTTNATTNGNNGTCCCCANNTTTCANCCTTATGTTTNTANTCCTAAANCCTN
CCCTANGNNNTG
>A142
TTTNNNTNNNGGNCGNGCTTAACATGCAGNTCGNANNANGAAACCGTTAGGGGGTGGATT
AGTGGCGAACGGGTGAGTAACACGATGGGCAATCTGCCCTGCACTCTGGGACAAGCCCTG
GAAACGGGGTCTAATACCGGATACTGACCANNTATGGGCATCCTTGATGGNGGAAAGCTC
CGNNGGCGCCNGATGAGCCCNCGCGCCTATCATCGNAGCTTNGTGNGGTAATGNTTTNNC
CAAGNCNACCACCNNGTTNTCCGGACTGANCATNNCNNCNTTTNCANNTNAGGNCTGNCA
CNCNGCTCGNCACNNNCTCCNTNNNNTNCCCCANTNGCCTTTTTNNCCTCTTCNNTCCCG
TTTCCGTTCTTANCNNGCCNTCCCNGCNCNTNNAAAANTTTTTNTCANCNNTTTNNNAAC
CCNCCNCCCCNTTCCCCNCAANTAAAAANTCTCNATGTTNTTNTCCNCNTCNCCCNCTTT
NNCCNCTNTCTAANNACGNNANACGNGNACNTTNTNTTNTTNCCCCCNANACTTNNTTNT
NCACNCCTCNCCCCNTTNTNTNTNTTTTTTTNNNTCNCNTNGCCGTNTNNCNTTTNTTTT
TNNNNTNNTNTATCCGNNTTNCCTTNCCNCCTNANNCNNTTCATNCNNNGTNCNNCTNTC

NTGNGACNNNTNNNNTTCCTATANTNCCNCNGNTNTNCTCNTGTCATCACCCCCNTTTNN
TTTTCCCNTCTTNNNNATTTCNTNNTCNNCNCCCNCNCCCCCCTTTATTTCTCCNCCGCN
TCTCNTAACCTTTTCTANTTCNCCTNTTTNTCTNNNNATCNTTCGCCCTACCCCCTTCNA
ATTNCGTGNTNNCNTGTCCTNTTTTCNCCNGCCNCCCNCCCNCCTTNTNNANNNNCTCNN
TTCNTNCNTCCCTCNGNTNCCNANTTTCNATCNTTNANNNNCNGTTAGTCCTNNCTNNNG
TTANGNNTCNCTCCGCCCCNTTTTTCNTCTCTTCCCNCNTCTTTTTNCCCCCCTTNNNCC
CNCCCCCNNNTTTTCCNNCNTTNCCTACTTCTNNNCTTATNCNNTCNCNCCNANCTNNCT
TTNTTNTTTTNNNCNNCCTTTTCNTCCCCCNTNCTNNTNTNNTCTNTNCNTCTCCCCNTN
NCNCNCTTTCTTCCCTTCTTTNCTTNTTTNTTATTCCCCCTTNNNCNCCTCTTCTCCTCC
CTTCTATCNTNNCNACTACTTNCNATCTTNCTNCCCNTGTTCNNCCCCTCNCTTCTNTTC
NCCCNTATCNCANTTTTCNNNANAATTCTNATNCNNNNANNCACCTCCTNCCTCCCCCNC
CNCCCCCNTTTCNNCNTTTNTNNNCNCTCNTTTTCCNCTNTTTCCCNGCCNCACCCCNNN
CCTCCCTCTTCTTCTTTTNTTTTTNNTCCCNCCTTNNCCCTTCTTTTTCTTTNTCCTCCT
TCCCCTCCCCA
>A143
NANGNGCGCCTTCCATGCAAGTCGAACGGTGACNCGGGNACAACCTGGCGACGAGTGGCG
AACGGGTGNCCTTTTAANNGGAACGTGCCCAGCTTGTGGGGGATAACTGCGTCGAATTAG
CAGCTAATACCGCATACGACCTGAGGGTGAAAGCGGGGGATCGCAAGACCTCGCGCAGTT
GGAGCGGCCGATATCAGATTAGGTAGTTGGTGGGGTAAAGGCTCACCAAGCCAACGATCT
GTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGANACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCCATGCCGCGT
GCGGGAAGAAGGCCTTCGGGTTGTAAACCGCTTTTGTCAGGGAAGAAAAGACTCCTGTTA
ATACCGGGGGTTCATGACNGNACCTGAANAATAAGCACCGGCTAACTACGTGCCAGCAGC
CGCGGTAATACGTAAGGTGCAAGCGTTAATCGNAATTACTGGGCNTAAAGCGTGCGCAGG
CNGTTATGCAAGACANATGTGAAATCCCCNGNCTCAACCTGGNAACTGCATTTGTGACTG
CATGGNTAGAGTACNGCAGATGGGGATGGAATTCCNCGNGTANCAGTGAAATGCNTANAT
ATGCGGAGGANCACCGATGGCTAACGCAATNCTCCTGNGCCTGTACTGACGCTNATGCAC
GAAATTCGTGGGGGAGCAAACANGATTANATACCCTGGTAGNNCACGCCCTAAACGATGT
CAACNGGTTGTTNGGGAGGGTTTCTTNCTCAATTAACGTANTCTNACGCGTGAANTNGAC
CGCCAGGGTAAACCNCCANNAATTTAAAACTCANCGGANTTGACGGGGACCCNNACNAAT
TANTNNANNACNNGGTTTAAANCCNTGCCAACGNGTNNAATCTTNCCTACCCCTTGACAT
GNCANGANTTTTCNTATANNTTTTNGNCTNCCNCCNNNNAANCCTCGGTNNCNNGNTGCN
NCNAGGNCGANGTTCGNNTCCNNTCTNGAAATGTCGGNTNANTCCCCCCCNNTNNNCCCN
CCCTTTTCTTTNTTNTCCCNAAAGNCTTCATCNACNGCGNNCNTCCACAANNATANGNAG
TNGNGNTNCTNCNTNTTCCCNTGNNCCNNGGGNTCNCNCCNNCTCTNNAGNNCCGNTNNN
NNNTTTNCNNCCNCCNCTGNNNCTANNNCNANTTCTNNTNCNCCCNNNTNCCTTCCCCC
>A147
CANGGGGTGCCTACCANNGCAAGNTCGAACGGGCAGCCACAGATANGAACCTTNGCCTNC
TTNTGGTGTGGNGAGTGCGCGACANGGTGNGAANNATANNTCGANNCTCTNCTTTTNTCG
TGGGNNATNACNTNGGNANNNTNACTCTTANNCCNCACACCNNCTANGGGNGAANNCGGG
ANNTNNTCNNNNCTTGCGCTNTNGNNNGNCCNTATCTCANANTNTCTNTNNNGNGGNNTA
ANNNCCCNCCNGNGNCANTANCCATNTCNNNTCNGAGNANANNNTCNCCCACTNTNNNNC
NGACACACTNNCCACACNCCNACGNGNNGCNNCTGTGGNNNATNTTNNANNNTGNGCNCN
NNNNNNNNCCNCCNANNCCGNGNGTGNNAAAANCNNNNTCGNNTTNNANNCCNNTTTTNN
TGNNAAAAATNNCCNCCNNNNTNANNNNTGNNNGNNANNACTGNCCCCANAANATANNCN
CCNNNNNNNTTNGTNCCNNCCCCCGCNNNTANACAANGGGNGCNCGNTTNTCTCAANNTN
NNTGNGNNTAANGCGNGNATNTGTNNNTNTTNANCTCTNTTGNNAANCCNNTGNTCTCCC
CNTGAANNCTGCNGTNAACNCTGCANNNCTAGNGTGTGNTAGAGNNTNGCANNTTTCCGN
GTGTNACNGTNANNTGNATAGATNTCGAGNANNANNTCNGTGGNAANCGCCTCNNNNTGG
NCNCACTCTNACTCTGNCGCACAAACGNGTGGNGNNNANACGANAATATANACNNTGGNT
CNNCACCCNCTANANTATGCACNNTNNATNTTGTGNGNNTTNNNGCACACTATNTCAANT
CTNACGCTTTNATTTTCCCCCCCCGGGGAGANTANTNTCNCAACNCTNACNCTCANNAGT
TTTCANGGGNNCCCCACACANGCNGGNNGAATTATGTTNTTTTATTTNTATACCACCGCA
AAAACCNTTCCNGGCCNTTNATATNNTNNAAAAANTNNCCAAAANAGANNTTNGNNCTTT

TNGAAAACTNAAANCANNNGNTCNNNNGGGTNTTTTCCTCCNCTCGTNNTNNNAANNTTT
GGGTTNANANCCCCACNNNAACCNCCCCTTTNTTTTTTTTTCCCACCCNNNNNGGGGAAA
NTTNANGNACCCCCCNTNAAACCCNNAAANGGGGGGNANTTNTTTTTGGCCCNTTNNNNG
GGCCCNNTTTNAANTTNNNGAAAANNTTNNNCCCNCNNNNNCCNCCCAAACCTTTTTNNC
GNNGGNNCCCNCCCCC
>A150
NANGCGNNCTTANCATGCAAGTCGNACGATGAACCACGTTNGTGTGGGGATTAGTGGCGA
ACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGG
TCTAATACCGGATACCACTCTCACGGGCATCTGTGAGGGTTGAAAGCTCCGGCGGTGAAG
GATGAGCCCGCGGCCTATCAGCTTGTTGGTGAGGTAACGGCTCACCAAGGCGACGACGGG
TAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTG
AGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGACGGT
ACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCA
AGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCACGTCGGGTGTG
AAAGCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCTAGCTAGAGTGTGGTAGGG
GAGATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGC
GAAGGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGG
ATTAGATACCCTGGTAGTCCACGCCGTAAACGGTGGGAACTAGGTGTTGGCGACATTCCA
CGTCGTCGGTGCCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGGCCGCAAGGC
TAAAACTCAAAGGAATTTGACNGGGGCCCNCACAAGCAGCNGGANCATGTGGCTTAATTT
CGACGCACCTCGAANNAACCTTACCAAGGCTTTGANCATTNCACCGNAAAACCCCTGGAG
AACAGGGTCCCCCCTTGTGGTCCNGTTTNACANGTGGGTGCATTGGCTGTTCTTCATNNC
TTNTCTTGAGATNTTTGGTTTNANTCCCCCNAACNAANCGTAANCNTTTTTTTTTTNTTT
CCACCATNTCCCTTNNGGGTNATGGGGAATTNANNAAAAACCTCCCGGGTTANCTNGGAN
GAAGGGGGGANNAATTNNAATTTTNTTNCCNTTTTTTTTTTGGCNTCNNACNCCTNAATG
NNCNNTNAAAAACTCCNNANCNTGTGGGGANGAATTTAAAACCCGNTTTCNCNNTATTTN
TTTTNNANCCCCCCNTANCCGNTTCCTAA
>A151
NAGGNGCNGCTTANCATGCAAGTCGAACGCGGNNTAGCAATACACGTAGTGGCGCACGGG
TGTAGTAACGCGTGGATATCTGCCTTTTGGTTCGGAATAACCCCGGGAAACTGGGGCTAA
TACCGGATGGTTCCTTCGGGATAAAGATTTATCGCCAAAAGATGAGTCCGCGTACGATTA
GCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCGTTAGCTGGTCTGAGAGGACGA
TCAGCCACACTGGGACTGAGACACGGCCCANACTCCTACGGGAGGCAGCAGTGGGGAATA
TTGGACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGT
TGTAAAGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAAC
TTCGTGCCAGCAGCCGCGGTAAGACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGT
AAAGCGAGTGTAGGTGGTTGTCCAAGTTGGATGTGAAAGCCTTGAGCTCAACTCAAGAAA
TGCATTCAGGACTGGGCGGCTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTGG
TGAAATACGTAGAGATTGGGAAGAACACCAGTGGCGAAGGCGGCTATCTGGACGGTTTCT
GACACTAAGACTCGNAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACNCC
GTAAACGATGGGTGCTAGACGTNGGCGAGCTTGCTCGTCAGTTGTCNCAGCTAACGCGTT
AAGCACCCCNCCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAANGANTTGACNGGGGC
CCGCACNANNNGTGGAGCATGTNGTTNCAATTNTNCCCAACGNCNCATAANCCTTTACNN
GCCCTTNANTNGGNACTNTNCGGGTTGNANAAACCCTTTCCTCCNGTTCTTCCTNTATTC
CTCNCNNGGGNCTTTATGGNTTTCNTNATTTNTNNTTTNTAATNTTNNNNTTTAACCCTT
TATCTNCCCCTNNCTTCNCCTCCCTCTTCCTNNTTTTNNCCCNGTNTNCTTCAACNANCT
TNCNTTCACNCNNCTNNNNCANACTGNNATTTTTNNTTNCTCTNTTNCCCATCNCTCCNC
NCCCCTCTCCCCCNTNTTCNNGNCTTTNTTTCTTTCCNTTTCNNNNTNTTNCCNANCNCN
CCNNTNNCTCNCTTTCANATCTTTTTNTATTTTTCCCCCTTTTCNCTTATTTNNTTANNT
TNTNCNNCTTTCCCNNCCNCN
>A156
NATGCGCGNCTTANCATGCAAGTCGAACGATGAACCACCTTCGGTGGGGATTAGTGGCGA
ACGGGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGG

TCTAATACCGGATACCACTCTCGCAGGCATCTGTGAGGGTTGAAAGCTCCGGCGGTGAAG
GATGAGCCCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGG
TAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTG
AGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGACGGT
ACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCA
AGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCACGTCGGGTGTG
AAAGCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCTAGCTAGAGTGTGGTAGGG
GAGATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGC
GAAGGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGG
ATTAGATACCCTGGTAGTCCACGCCGTAAACGGTGGGAACTAGGTGTTGGCGACATTCCA
CGTCGTCGGTGCCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGGCCGCAAGGC
TAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGCTTAATTCGA
CGCAACGCGAAGAACCTTTCCAAGGCTTGACATACACCGGAAANCGGCCANANATGGTNG
CCCCCTTGTGGTCGGTGTTACAGGTGGTGCATGGCTGTCNTCNAGCTNCTGTCNTGAAAT
GTTTGGGTTAANTCCCNCAACAANCCNACCCTTTGTTTTTTGTTNCCANCATGNCCTTTC
GGGGTGATNGGGACTCCCAAGAAAACCCCCGGGNTCACTTCGAAGGAAAGNNGGGANNAA
CTNNANTCTTCNNCCCCTNTTTTTTGGCTCCCCCCTGCTCAANTGGCGGTCAATNACTNC
NANACCTGGNGGNGANCNAATTCAAAAACCTTTNNATTTNGNTTGGGTNCCCCTCCCCCC
NTAANTCTANTTTTATTACCAAANCCATTTGCNNGTAAACT
>A159
ANCGCGNCTAACATGCAAGTCGAGCGGAAAGGCCCTTCGGGGTACTCGAGCGGCGAACGG
GTGAGTAACACGTGAGCAACCTGCCCCTGACTCTGGGATAAGCCCGGGAAACTGGGTCTA
ATACCGGATATGACCGCCCCTGGCATCGGGTGGTGGTGGAAAGTTTTTCGGTTGGGGATG
GGCTCGCGGCCTATCAGCTTGTTGGTGGGGTAGTGGCCTACCAAGGCGACGACGGGTAGC
CGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAG
GCAGCAGTGGGGAATATTGCGCAATGGGCGGAAGCCTGACGCAGCGACGCCGCGTGGGGG
ATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGACGAAGTTGACGTGTACCTGCAG
AAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTGT
CCGGAATTATTGGGCGTAAAGAGCTCGTAGGTGGCTGGTCGCGTCTGCCGTGAAAGCCCG
CAGCTTAACTGCGGGTCTGCGGTGGATACGGGCCGGCTAGAGGTAGGTAGGGGCAAGTGG
AATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGG
CTTGCTGGGCCTTACCTGACGCTGANGAGCGAAAGCGTGGGGAGCGAACAGGATTAGATA
CCCTGGTAGTCCACGCTGTAAACGTTGGGCGCTAGGTGTGGGGGTCTTCCACGATCTCCG
TGCCGGAGCTAACGCATTAAGCGCCCCGCCTGGGGAGTACGGCCGCCAGGCTAAANTCAA
AGGAATTGACGGGGGGCCCNCACAAGCGGCGGAGCATGTTNNTTAATTTCGACNCAACGC
GAANAACCTTANCCAAGTTTTGACATCCCCNGNNNCTCCCAGAAAATNGGGGNCTTNTTC
GGACTGGGTTAACAGGTGGTGCATGNTNGTCNTNACTNNTNTTCNTGANATGTNGNTTTA
ATTCCCTCAACNANTNCACCNTTTTTCCATTTNNCNACAAANCCNTTTTTGTGGTTGGGA
ATTCTTGGGGACTTCNNGGGTCATNCCAANAAGGTGGGNNTANCTCCATTNTTTTNCCCT
TTTTNTTTGGTCCNAANTNNTNANTNCCNGNCNAANGTTNTAANCTNNGNNAACANCCTA
AACCGTTNCTCCNAATTGGNNTTTATTTNNCTTTANTNANCCTTAANTCAATN
>A162
NATGCGCGNCTACCATGCAGTCGAACGATGAACCACTTCGGTGGGGATTAGTGGCGAACG
GGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCT
AATACCGGATACCACTCCTCAAGGCATCTTGGGGGGTTGAAAGCTCCGGCGGTGAAGGAT
GAGCCCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAG
CCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGA
GGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGG
GATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGACGGTACC
TGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGC
GTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCACGTCGGGTGTGAAA
GCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCTAGCTAGAGTGTGGTAGGGGAG
ATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAA

GGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGATT
AGATACCCTGGTAGTCCACGCCGTAAACGGTGGGAACTAGGTGTTGGCGACATTCCACGT
CGTCNGTGCCCCANCTAACGCATTAANTTCCCCGCCTGGGGAGTNCNGNCCGCAAGGCTA
AAACTCCAAAGGAATTGANTGGGGGCCCCNCACAANCACCGNANCATGTGGCTTAATTTN
NACNCAACNCGNAANAACCTTNACCAAGGNTTNACNTNCACNGNAAACGGCNAAANATGN
TNCCCCCCTTTTGGTCNTTNTAAANGTGGTTNCATGNCTTTNCTNCCTTCTNTNCTTAAA
TTTTGGGTTTAATTNCNNCAACAANNCCNACCTTTTTCTNTTTTCCCCNTTNCCTTTCGG
GTNNTGGGACTTCNAGNAAACCCCCNGGNTCATTNNAAAAAGGGGGAAAAANCNATTTTT
TTNCCNNTTTTTTTGGGTTTCACCTNCTNNANNGCCNTNCTTTNTTNNAAACTTTNNNNG
TANNANNCCCAACTCNNNCNTTTTTTTTTTTTCNTCCCCCTNANTTNATT
>A165
GANCGCNNCTACCATGCAAGTCGAGCGGAAAGGCCACTTCGGTGGTACTCGAGCGGCGAA
CGGGTGAGTAACACGTGAGTAATCTGCCCCTGGCTTTGGGATAGCCACCGGAAACGGTGA
TTAATACCGGATATGACGCGCTCTCGCATGGGGGTGTGTGGAAAGTTTTTCGGCCAGGGA
TGTGCTCGCGGCCTATCAGCTTGATGGTGAGGTAATGGCTCACCATGGCTTCGACGGGTA
GCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATATTGGACAATGGGCGGAAGCCTGATCCAGCAACGCCGCGTGAG
GGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTACCGACGAAGCGAAAGTGACGGTAG
GTACAGAAGAAGCACCGGCCAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAG
CGTTGTCCGGAATTATTGGGCGTAAAGGGCTCGTAGGCGGTTTGTCGCGTCGGGAGTGAA
AACGCCGTGCTTAACACGGCGCTTGCTTTCGATACGGGCAGACTAGAGGTATGCAGGGGA
GAACGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGA
AGGCGGTTCTCTGGGCATTACCTGACGCTGAGGAGCGAAAGTGTGGGGAGCGAACAGGAT
TAGATACCCTGGTAGTCCACACCGTAAACGTTGGGCGCTAGGTGTGGGATCCATTCCACG
GGTTCCGTGCCGCAGCTAACGCATTAAGCGCCCCGCCTGGGGAGTACGGCCGCAAAGGCT
NAAACTCAAAGGNAATTGACGGGGGCCCGCACAAGCGGCGGANCATGCGGATTAATTCGA
TGCAACGCGAANAACCTTACCTGGGTTTNNACATACACCTNCCCGCCTCAAANATGGGGN
TTCTTTTGGGGGTGTACAGTTGNTGCATGGCTNTCTTCNCCTCNTNTCNTGANANTTTGG
GTTNANTCCCNCNACNNAGNNCAACCCTCNTTTNTNGTTGCCNNCNCNTTATNGNNGGGA
NTCATAGGAAACTNNCCGGGTTNACNCGNAAGAAAGGGGGANTANNTNAATTATNCNNCC
CCTTNNTCCNGGNTTCCCNNTGNNTNANGGNCNGTAAAGGGCTTCCNTCCCNAGGGGAAC
NAATCCAAANCCGNNTCNNTNTNA
>A166
ATGCGCNGCTTACCATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAA
CGCGTGGGAACGTACCCTTTACTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTAT
GTGCCCTTCGGGGGAAAGATTTATCGGTAAGGGATCGGCCCGCGTTGGATTAGCTAGTTG
GTGGGGTAAAGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCAC
ATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAA
TGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGC
TCTTTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGCCCCGGCTAACTTCGTGCC
AGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCA
CGTAGGCGGACATTTAAGTCAGGGGTGAAATCCCAGAGCTCAACTCTGGAACTGCCTTTG
ATACTGGGTGTCTGGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTC
GTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTACTGACGCTGA
GGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGA
TGAATGTTAGCCGTCGGCAAGTTTACTTGTCGGTGGCGCANCTAACGCATTAAACATTCC
GCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGC
GGTGGANCATGTGGTTTAATTCAAGCAACNCNCANAACCTTACCAGCCCTTGACATCCGG
TGCTACATCCAGANATGGATGGTTTCCTTCNGGAACCCCCNANANAGGTGCTGCATGGNC
TNTCCTCNCCTCNNGTCNTGAGATNTTGGGTTTANTCCCCCNANCNANCGCCAACCCTNC
CCCNTTNATTGCCANCATTCANTTTGGNCNCTCTTANGGGNACTGCCNGTTAATAANCCC
AAAAAGAAAGNTGGGGANTANCTTCAATNCCNNNGGNCCTTACGGGCTGGNNTCNNCCTN
CCTNAATGGTNNTAAANNGGGCACCNNNCCCCNNNNNTNAACTNATTCCAAAACCNNTTC
NNTNCGGATNCCCCNCCANCCNNNNNNTTAATTGGAANCCCNANAANCCCGAACC

>A171
ATNGCGCCTNNCATGCAGTCGANCGGCAGCNCGGGTAGCNATCCTGGCGNCGAGNGNCGA
ACGGGTGAGCTAATACATCTNAACGATGTNNNNTCTTGGGGGATAACNCACCGAAANNAT
GCTGCGTANTACNGCCACTNCTATCATANGTGATGAAAGCATGGGATCGCAANACCTTGC
TTCGAATGTTAGCGGCCGATGGCAGATTNGNTNGTTGGTGAGGTNAAGGCTCACCTNGCC
TTCCNTNTGNNNCTGGTCTGANAGGACNACCAGCCNCACTGNGACTGANNCNCGGCCCAG
ACTCCTACCNTANGCAACNCTGGGGAATTTTGNANNNTGGGCNAAANCCTGATCNAACCN
TGCCNNGTGCNGGNTGAAAGCCTTNTNNTTGNAANCTGCTTTTGNNCTNANCNAAACNGN
CTTTTCTNNTNNNNAANGCTNATGACGGTACCNTANGAATAAGCACCGGCTAACTACGTG
CCAGCAGCCGCGGTAATACGTAGGGTGCAAGCGTTAATCGGANTTACTGGGNNTAAAGCG
TGCGCAGGCGGTAATGTAAGACNGTTGTGAAATCCCCGGGCTCAACCTGGGAACTGCATC
TGTTACTGNATTGCTNGANTACTGTAGAGGGNGATGGAATTNCTCNTGTAGCAGTGAAAT
GCNTAGATATGNCGNNGAACACCGATNGCGAAGGCAATNNNCTGNACNTGTANTGACGCT
CATGCACGAAAGCGTGGGGATCAAACAGGATTAGATACCCTGGTAGTCCACGCCCCTAAA
CGATGATNACTCGTTNATTNGCNCTNCACTGACTNNGTCCGANCCTAACCTTTNANATTC
CCCCTNTGGAAGNNCANCCCCCCNNNNTTAAANNCTCANGANNTTANNNGGGTCCCCCCA
AAAGCGNGGANATNTNTGTTTTTANNTTNNTNNTCCCCGCGGAAAACTTTTCNNNCCCTT
NNATATTTNGGGGAACTNGGGCTGAAAATTNNTNTTTTTGTTCCAAANANACCNTANCNA
CGGTNTTTTNTNTGTTTCCTNNTCTTTTCNNGGGTNTTTGGGNNANANTCNCNNNACNGG
CCNCCCCCTTNNTNTTNTNNTNATTTNANTNGGCACCCTTNNNTNATCCCCNNTNNNACC
NTNNNAAGGNGGNTTTNTNTNNTTCTTTTGCCTTTTTNCCGGNCNCTCCNTTTNCCTGGN
TNNAAANNGTTTNTNNTNNTNNGNNNTTTTCCCNATTCTTTTNTTTNTCTCGCCCCCCCC
CTNCNNA
>A175
NNCGCNGCTTACCATGCAGTCGAACGCGTGCTAGCAATACACGAGTGGCGCACGGGTGAG
TAACGCGTGGATATCTGCCTTTTGGTTCGGAATAACCCCGGGAAACTGGGGCTAATACCG
GATGGTTCCTTCGGGATAAAGATTTATCGCCAAAAGATGAGTCCGCGTACGATTAGCTAG
TTGGTGAGGTAATGGCTCACCAAGGCGACGATCGTTAGCTGGTCTGAGAGGACGATCAGC
CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGA
CAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTGTAA
AGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAACTTCGT
GCCAGCAGCCGCGGTAAGACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGC
GAGTGTAGGTGGTTGTCCAAGTTGGATGTGAAAGCCTTGAGCTCAACTCAAGAAATGCAT
TCAGGACTGGATGGCTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTGGTGAAA
TACGTAGAGATTGGGAAGAACACCAGTGGCGAAGGCGGCTATCTGGACGGTTTCTGACAC
TAAGACTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAA
CGATGGGTGCTAGACGTTGGCGAGCTTGCTCGTCAGTGTCGCAGCTAACGCGTTAAGCAC
CCCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACA
AGCGGTGGAGCATGTGGTTTCAATTNCGACGCAACGCGCAAAACNTTACCAGCCCTTGAC
ATGGGACTCGCCGGGGACCANGAGATGGTTCCCTTTCGGTTCGGCCGGAGTTCCGCACAG
GTGCTTCATGGCTTNTCGTTCAGCTCCGTGTCNTGAGATTTTTGGGTTAAGTTCCNNCAA
CGAGCCCAACCCTTCGTCTCCANTTGNCATTCANGTTNNTGCTGGGCNCTTTGGAGAAAC
TTCCGGTGANAACCCGGANGAAGNNGGGNATAACNTNANNTCCTNNGGCCCNTTCCGGGT
TGGGCTACNACNNNNTCAANGGNNGTTANATGGGAANANTGNCCNATCCGAACCCATCCT
AAAAACNNTTTTNTTGNATTNCCTTCCCNCNCNNGGCCNTAATTGANTCTTTNTCNAGNN
CNCCGCCCNTCGGANA
>A177
CGCGGCTACCATGCAAGTCGAACGCCCCGCAAGGGGAGTGGCGCACGGGTGAAGTAACGC
GTGGGAATATGCCCTTCGGTTCGGAATAACACAGGGAAACTTGTGCTAATACCGGATACG
ATCTACGGATGAAAGATTCATCGCCGAAGGAGTAGCCCGCGTAGGATTAGCTAGTTGGTG
AGGTAATGGCTCACCAAGGCGACGATCCTTAGCTGGTCTGAGAGGATGATCAGCCACACT
GGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGG
GCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTCT
TTCGCCAGGGAAGATAATGACGGTACCTGGAAAAGAAGCCTCGGCTAACTCCGTGCCAGC

AGCCGCGGTAAGACGGAGGAGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGTGCGC
AGGTTGTGAGTTCAGTTGGATGTGAAAGCCCGGGGCTTAACCTCGGATGTGCATCCAATA
CTGGCTCGCTGGAGGTTGGAAGAGGAGAGTGGAATTCCCAGTGTAGAGGTGAAATTCGTA
GATATTGGGAAGAACACCAGTGGCGAAGGCGGCTCTCTGGTCCATACCTGACACTCATGC
ACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGT
GTGCTAGACGTCGGGAGGCTTGCCTCTCGGTGTCGCAGCTAACGCGATAAGCACACCGCC
TGGGGAGTACGGCCGCAAGGTTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGT
GGAGCATGTGGTTTAATTCGAAGCAACGCGCANAACCTTACCAGCCCCTTGACATGGGAA
GTATGGACTAGAGAGATCTAGTTCTTCAGTTCGGCTGGCTTCCACACAAGGTGCTGCATG
GCTGTCNTCANCTCNTGTCNTGANATGTTGGGTTAATTCCCGCAACGAGNGCACCCCTCN
CCTTCANTTGCCATCNCTNCTAGGTNGGCCCNCTGAAGNAACTNCCNGTNACAACCCGNA
GAAGNNNGGGNATAACTTCAATTCNCNNGNCCTTNNNGGGCTGGNCNCCNCNNTNNTAAA
TGGCNACTNNANGGNAANNAANGANNTTCCNGANCCATCCCAAAANTNNTNCATTCNAAT
NCATTTNNACNNGCTCNNNAANCNNAATCTTTTATCTNGNATNNCCCCCCNNTAAACTTT
CCGNCCTTTCCC
>A179
ANGCGCNGCTACCATGCAGTCGAGCGGGCGTAGCAATACGTCAGCGGCAGACGGGTGAGT
AACGCGTGGGAACGTACCTTTTGGTTCGGAACAACACAGGGAAACTTGTGCTAATACCGG
ATAAGCCCTTACGGGGAAAGATTTATCGCCGAAAGATCGGCCCGCGTCTGATTAGCTAGT
TGGTGAGGTAATGGCTCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGATCAGCC
ACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGAC
AATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAA
GCTCTTTTGTGCGGGAAGATAATGACGGTACCGCAAGAATAAGCCCCGGCTAACTTCGTG
CCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATCACTGGGCGTAAAGGG
TGCGTAGGCGGGTCTTTAAGTCAGGGGTGAAATCCTGGAGCTCAACTCCAGAACTGCCTT
TGATACTGAAGATCTTGAGTTCGGGAGAGGTGAGTGGAACTGCGAGTGTAGAGGTGAAAT
TCGTAGATATTCGCAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCGATACTGACGCT
GAGGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAAC
GATGAATGCCAGCCGTTAGTGGGTTTACTCACTAGTGGCGCAGCTAACGCTTTAAGCATT
CCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGGCCCGCACA
AGCGGTGGAGCATGTGGTTTAATTCGACGCNACGCGCAGAACCTTACCAGCCCTTGACAT
CCCNGNTCGCGGACTCCAGANACGGAGTTCTNCAGTTTCGGCTTGGACCGGAAACAGGTG
CTGCATGGCTNTCGTCAGCTCTTGTTCTGGANATGTTTGGGTTAAATTCCCCCNAACNNG
CCCAACCCCCNTTCCTTNTTGCTACCNTTTTTTNNNCCCTCNTAAGAAACTCCCNGTTAN
AACCCCCCAGGAAGGTNGGGANTNNNTCAATTCCNNTGGTCCTTTNCGGGNTGGGNNTNC
ACCTTTNTNAANTGNCGNNAAANGGGNTTNNAGGGNTACCNTNCNANTNCAAAAACCCNN
CNTTTNGANTNGNTTTTNNTTNNNCCCCAAATTGAACCCCNTNTTTTNATNACCCCCCNG
TAACCTNCCGNCT
>A180
ANGNGCNGCTACCATGCAGTCGAGCGCCCCGNTAAGGGGAGCGGCAGACGGGTGAGTAAC
GCGTGGGGATGTACCCGAAGGTACGGAACAACTCCGGGAAACTGGAGCTAATACCGTATG
AGCCCGAGAGGGGAAAGATTTATCGCCTTTGGATCAACCCGCGTCAGATTAGCTAGTTGG
TAGGGTAATGGCCTACCAAGGCGACGATCTGTAGCTGGTCTGAGAGGATGATCAGCCACA
CTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAAT
GGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGACGGCCTTAGGGTTGTAAAGCT
CTTTCGACGGGGACGATAATGACGGTACCCGTAGAAGAAGCCCCGGCTAACTTCGTGCCA
GCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATCACTGGGCGTAAAGCGCAC
GTAGGCGGATTGTTAAGTCGGGGGTGAAATCCTGAGGCTCAACCTCAGAACTGCCTTCGA
TACTGGCGATCTTGAGTCCGGAAGAGGTTGGTGGAACAGCTAGTGTAGAGGTGAAATTCG
TAGATATTAGCTAGAACACCAGTGGCGAAGGCGGCCAACTGGTCCGGCACTGACGCTGAG
GTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGAT
GGATGCTAGCCGTTGGGGAGCTTGCTCTTCGGTGGCGCAGTCAACACATTAAGCATCCCG
CCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGGCCCGCACAAGC
GGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCANAACCTTACCAGCCCTTGACATCCC

GGTCGCGATCCTCANAGACGAGGGTCATCANTTCGGCTGGACCGGANACAGGTGCTGCAT
GGCTGTCGTCAGCTCGTGTCGTGAGATNTTGGNTTAATTCCNCAACGAGNGCAANCCTCG
CCCCCTANTNGCATCATTGAGTNGGGCCTCTAGGGGACTGCCGNTATAACCCCCAAGAAG
GTGGGATACCTTNAATCCTCTNGNCCTTACGGNTGGGCTNCNNCNTGCTACATGNCGGTN
NCATTGGATCCAANGGCCACCCTNANAANTTCNAANCCNNTNNATTTNATNNCCNTCCAC
NCNGGGCNNANGGGAATCCTTNAACNTAAACCCTTTTCGNAA
>A181
CATCGCNGCCTATCATGCAAGTCGAAACGGTAAGTAGTGGTAGCAATACATTGCCTAGAG
TGACGTAAGGGTGCGTAACACGTATGCAATCTGCCCTGTACAGGAGTATAGCTCCCCGAA
AGGGGAATTAACCCTCCATAGTATAATTGAATGGCATCATTTGATTATTAAAACTGAGGT
GGTACAGGATGAGCATGCGTCTGATTAGCTAGTTGGTAGTGTAATGGACTACCAAGGCGA
TGATCAGTAGGGGAACTGAGAGGTTGATCCCCCACACTGGCACTGAGATACGGGCCAGAC
TCCTACGGGAGGCAGCAGTAGGGAATATTGGTCAATGGGTGAGAGCCTGAACCAGCCATG
CCGCGTGCAGGAAGAAGGCCTTCTGGGTTGTAAACTGCTTTTGCCAGGGGATAAAACGGG
AGTGCGCTCCTAATTGAAGGTACCTGGTGAATAAGCCACGGCTAACTACGTGCCAGCAGC
CGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGTTTAAAGGGTGCGTAGG
CGGCTCCTTAAGTCAGTGCTGAAATACTCTAGCTTAACTAGAGGGGTGGCATTGATACTG
AGGAGCTTGAGTAGAGTCGAGGTAGGCGGAATTGACGGTGTAGCGGTGAAATGCTTAGAT
ATCGTCAAGAACACCGATAGTGTAGACAGCTTACTAGGCTTCAACTGACGCTGAGGCACG
AAAGTGTGGGGATCAAACAGGATTAGATACCCTGGTAGTCCACACTGTAAACGTTGATTA
CTCGCTGTTGGCGATATACAGTCAGCGGCTTAGCGCAAGCGATAAGTAATCCACCTGGGG
AGTACGCCGGCAACGGTGAAACTCAAAGGAATTGACGGGGGTCCGCACAANCGGTGGAGC
ATGTGGTTTAATTCGATGATACGCGAGGAACCTTACCTGGGCTAGAATGCCCCTTGATGT
CCTCNAAANACGAGGAGTCCCCCANGNACANGNAGCAAGGTGCTGCATGGCTGTCNTCAG
CTCGTGCCGTGAGGGGTTGGGTTAAGTCCCNCACCGAGCCCAACCCTTATTTTTANTGCC
AANAGGTTATGCTGGGGACTNTAAAAAAACTGCNNGCCCANCCAAAAGGAAAGNGGGGAT
ANCTCAATTCNTTTGGCCCTTNCCCCCNGGCTACCNNNTTCTANATGGNTTAANAAANGT
TNCAANTNGGNACNTAACCCATCCAAAANTTCTTTCANTTCGNATGNGNGNTCCCCCCCC
CTCCAAATG
>A182
NTCGCGNCCTATCATGCNAGTCGAACGGTAAGTAGTGTAGCAATACATTGCCTAGAGTGA
CGTAAGGGTGCGTAACACGTATGCAATCTGCCCTGTACAGGAGTATAGCTCCCCGAAAGG
GGAATTAACCCTCCATAGTATAATTGAATGGCATCATTTGATTATTAAAACTGAGGTGGT
ACAGGATGAGCATGCGTCTGATTAGCTAGTTGGTAGTGTAATGGACTACCAAGGCGATGA
TCAGTAGGGGAACTGAGAGGTTGATCCCCCACACTGGCACTGAGATACGGGCCAGACTCC
TACGGGAGGCAGCAGTAGGGAATATTGGTCAATGGGTGAGAGCCTGAACCAGCCATGCCG
CGTGCAGGAAGAAGGCCTTCTGGGTTGTAAACTGCTTTTGCCAGGGGATAAAACGGGAGT
GCGCTCCTAATTGAAGGTACCTGGTGAATAAGCCACGGCTAACTACGTGCCAGCAGCCGC
GGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGG
CTCCTTAAGTCAGTGCTGAAATACTCTAGCTTAACTAGAGGGGTGGCATTGATACTGAGG
AGCTTGAGTAGAGTCGAGGTAGGCGGAATTGACGGTGTAGCGGTGAAATGCTTAGATATC
GTCAAGAACACCGATAGTGTAGACAGCTTACTAGGCTTCAACTGACGCTGAGGCACGAAA
GTGTGGGGATCAAACAGGATTAGATACCCTGGTAGTCCACACTGTAAACGTTGATTACTC
GCTGTTGGCGATATACAGTCAGCGGCTTAGCGCAAGCGATAAGTAATCCACCTGGGGAGT
ACGCCGGCAACGGTGAAACTCAAAGGAATTGACGGGGGTCCGCACAAGCGGTGGAGCATG
TGGGTTTAATTCGATGATACGCGAGGAACCTTACCTGGGGCTAAAAATGCCCCTTGATGT
CCTTNAAAAACGAGGAGTTCCCCCAAGGGACNAAGNAGCNANGGTGCTGCATGGCTNTTC
GTCAGCTCNTGCCCTNGAGGTGTTNGGGNTTAANTCCCCCAACNAGCNCACCCNTTNNTT
TTTNNTTGCCAACAGGTTTTNCNTGGGGGACNNTAAAAANCTNCCTTCCCAACCAAAAGN
AAGGGAGGGATNANCTCAATTTTNCNTGGCCCTNACCCCCAGGGNTACCNNNTNNTAANN
GGGNTTAANAGGGTTNCAAATGNNNCCNNAACCCNNCNCAAAATTCCNTN
>A186
ATGCGCGNCTANCATGCAAGTCGAACGATGAACCACTTCGGTGGGGATTAGTGGCGAACG
GGTGAGTAACACGTGGGCAATCTGCCCTGCACTCTGGGACAAGCCCTGGAAACGGGGTCT

AATACCGGATATCACTTCCACTCGCATGGGTGGGGGTTGAAAGCTCCGGCGGTGCAGGAT
GAGCCCGCGGCCTATCAGCTTGTTGGTGAGGTAACGGCTCACCAAGGCGACGACGGGTAG
CCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGA
GGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGG
GATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGACGGTACC
TGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGC
GTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCACGTCGGGTGTGAAA
GCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCTAGCTAGAGTGTGGTAGGGGAG
ATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAA
GGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGATT
AGATACCCTGGTAGTCCACGCCGTAAACGGTGGGAACTAGGTGTTGGCGACATTCCACGT
CGTCGGTGCCGCANCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGGCCGCAAGGCTAA
AACTCAAAGGAATTGACGGGGGGCCCGCACAAGCANCGGANCATGTGGNTTAATTCNACG
CAACGCNAAGAACCTTACCAAGGCTTGACATACACCCGGAAAACGGCCANANATGGTCNC
CCCCNTTGTGGTTCGTNTNNANGTGTTGCATGGCTTTTCTNNNCTCCTGTCTGANATTTT
NGGGTTAANTTCCCNCAACCANNNCACCCNTTTTNNTNTNTTNCCNACTGNCCNTTNGGG
NTGATGGGNANTNNAGNAAACCCCCNGGTTCACTCNNAAGNAAGTNGGAACAANTNCAAT
TTTTATTGCCCTTTTTTTTTGGCNNCCCCCNTCTNNATNGNNGGTNANTGAANTTCAANC
CCNGNGTGGNNCNAATNTAAAAACCTTTNNNTTCGANTGGGTTCCAC
>A188
ANGCGCGGCCTAACATGCAGTCGAACGCCCCGNACGGNTNTAGCGGCATACTGGATGTAG
TAACNCGGTGTGGACCCNTGCCCNTTCGTGTTCGTGAATANCCTCATTGNGAAACTTGAG
CTTAATACCGTGATACNTCCTATAAGGAGAAAGATTTATTGNCNAAGGANCGGCCCGCNT
CCGATTAGCTAGTTGGTNAGGTAACGGCTCACCNAGGCTACNATCGGTAGCTGGTCTGAN
ANGATGATCAGCCACACTGGNACTGAGACACGGCCCACACTCCTACNGGAGGCAACATTG
GGGAATATTGNACAATGGGCGCAAGCCTGATCCATCCNTGCCNCNTGAGTGATGAANGCC
CTAGGGTTGTNAAGCTCTTTCTCCNGGGACTATAATGACGGTNCCTGGATAANAAGCCCC
NGCTAACTTCGTGCCANCNNCNNCNNTAATACCAAGGGGGCTANCNTTNTTCNGAATCAC
TGGNCGTAAAGCGCACNTAGGCGGACTGTTAAGTCGGGGGTGAAATCCTGGGGCTCAACC
CCACAACTGCCCTCGATACTGGCGGTCTTGANTGTGGAAGAGGTTGGTGATAACTCCGAG
TGTAGAGGTGAAATTCGTAGATATTCATTAAGAACNCCAGTGGNGAAGGCGGCCAACTGG
TCCACTACTGACGCTGAGGTGCGAAAGCGTGTGGAGCAAACATGATTATANACCCTGGNA
GTCCACGCCGTAAACTATNATAGCTACCCNTNGGTGTNCATGCACATNAGNTGGNGCACA
TAACTCGTTAANCTCTCCGCCTGNGGAGTACGGTCGCNANATTAAAACTCAANAGGAATT
GACGGGGGCCCCGCACAAGCGGTTGGAACCTGTGGTTTTAATNCGAAGCNACNCCCCAGA
ACCTTACCACCCTTTGNNATTTCCTTGGAAATTGAAAANAAANATNGATNNCNNTTTTTN
GGANCCNGGNAACCAGGTGCTTTNTGGCTTTCTTNANCTCTTGTCGGNANATGTTGGNTT
AAATTTCTNAACAAGGNGCAACCTNNCNNTTTATTTGCTNTCATCNNTTGGGNANTTTNG
GGGGATTTCNTNAAAATCNCCNNGNAAGGNNGGGATAACCTNAATCCTNTGGCCTNACGG
GNGGGTTCNNNTTTNANAATTNCTTTAANTGGACCAACCNNGGCCNATTNTTNCCAAACC
TTTTTTTTTNTTANCTCCCCTTNNNCTCTAAATCNANTTTNNNNNNCNANNAACCCCNGG
>A191
ANGNGCGGTCTTANCATGCNAGNTCGAGCGGAGNGGTATNAATACGCTNGGCCTCGTTAG
TANAGGTTGGACGTNNCGCACTGGGTNACGTNACNCTGTTGTGGNTTNCCNGCACACCTA
AGGACTGNCGTATNACTCTCNNGCNAACNNNNGGCCTANTACCGGGANAANATTTATTCN
CCGAANGATGCGCNCGTTGNNNGNNNNCCTTCGNNGTGTNACNTANCGNTNNACCCNGCG
CTCACAATNACTAGTTGGTNNGATAACATCANCCCNCNNCTTCGANCNGNAACCCGACCT
NAGACNCCNATCGGNCGCACTGGTGACTGAATATNGGACCATGACTCNTANCNCTGNTNC
CACNATGCCAATCTTNCTCATTNAAGCANTNTNTNNNTNAACATCTNNTCTTGAGTNATA
TAATGCNTTCTNCCTCGTAAAANTCTGTTNTTNTNNANTNANNNCTGCTACNTNAATAAT
NTGNNANCTNNTCGGTTNCTNNCCAANNACNNNCCGNTAACTNCGNGNCAGCAGCNCNTT
NAATNCGNNGNNNNNTNNNGTNNTTCNNATNNATNGNTNNNNNNGNTNCNGNAGNTNNTT
NCTTNAGTNTAANNTANNNGNNCTCCGCTNNNCCATGTAAGGTNNTTGGANANTGGNANA
CTNGANTGNNGAAAAGGAAANTNNANTTCCNTGTGTNGCNGTTAAATGCCTAAAGNTATG

GAAGAACACCAGTNTCGATAGCCNNTTTCTNNNCTGTAACTGNCACTGATGCCCGNAATN
NTGGGGANNACNCACNATTNGATACCCTGATAGTCNACGCCTTACACCATGAGTGCTNAG
TGTTCAAAAGGTTTNCNCNNTTTAATGCTNANGNTNAACGCATTAAGCNACTCCANCNTG
NGGANTACNNCCACNANGCNTGAAACNCNNTTNAATTGNCNTNGGGCTCCCACNAGCGGN
NTANCANGTNGNTTTNTTNCAANTNNCNCNGAACAACCNTACCNNGTNTTGAAGTCNTTC
NNCAACCCCTGAATAAGGGTTTNTCNTTAGGNNCCNAATNACNGNNNNCCCCTTNTTTTN
CTCNNCCCTTTNNNTTGAAATTTNNNTAAACCNCCCANNAAANCCNCCCTTNANNTNGGT
TTCNNNCTTAANTTGGNGCCCTTNAGGGNATGNCNTTANAAACCNTAAAANGGNGGGNAA
AANNNANNNNNNGNCCCCNTTNNCNNTTTNAANNTTNNTNNNNNNNTNNNNNNNTNNAN
CCCNNGNTNNNTNNTNNNANCCTT
>A193
CGCNCCTACCATGCNAGTCGAACGAGACCTTCGGGTCTAGTGGCGCACGGGTGCGTAACG
CGTGGGAATCTGCCCTTGGGTTCGGGATAACAGTTGGAAACGACTGCTAATACCGGATGA
TGTCTTCGGACCAAAGATTTATCGCCCAGGGATGAGCCCGCGTCGGATTAGCTAGTTGGT
GAGGTAAAAGCTCACCAAGGCGACGATCCGTAGCTGGTCTGAGAGGATGATCAGCCACAC
TGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATG
GGCGAAAGCCTGATCCAGCAATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTC
TTTTACCCGGGAAGATAATGACTGTACCGGGAGAATAAGCCCCGGCTAACTCCGTGCCAG
CAGCCGCGGTAATACGGAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCGCG
TAGGCGGCTTTGCAAGTTAGAGGTGAAAGCCCGGAGCTTAACTCCGGAACTGCCTTTAAA
ACTGCATCGCTAGAATCGTGGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGT
AGATATTCGGAAGAACACCAGTGGCGAAGGCGACTCACTGGACACGTATTGACGCTGAGG
TGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATG
ATGACTAGCTGTCGGGGCTCATGGAGTTTCGGTGGCGCAGCTAACGCGTTAAGTCATCCG
CCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAAGAAATTGACGGGGGCCTGCACAAGCG
GTGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCGTTTTGACATGGN
TAGGACGGTTTTCCAGAGATGGATTCCNTTCCCTTACNGGAACCTACACACAGGTGCTGC
ATGGCTGTCGTCAGCTCGTGTCGTGAGATNTTGGGTTAAGTCCCNNCAACGAGCGCAACC
CTCGTCTTTAGTTGCTACNATTTAGTTGGGCACTCTAANAAACTGCCGTNATAACCCGGA
GAAAGGTGGGAATNANTTCAANTCNTNNNGGCCNTNCCNNCTGGNTNACNANTGCTANAT
GGGNGNTANNTGGGCNCAAATTCCNGAANTAACAATCCCNAAAACNNNTCATTNGAATNT
TNTNCACTCANNNTNAAGGCGAATCNTTTAATCGGATCNNNNCCCNGNAAACTTCC
>A194
GGAANAGGAAGGNANNACAAGNCAAGGTCGCTAACAAGGGTTAACCGTTAAATATTTGGC
CTATNNTNGGNGAATAAAATGNNGGGCCCCNTTTANGGGGAANANTTNCGTGNATNNTNC
TTCTCNTGTGNAACNTNGNNTGNCTANTTTCTATNNNNNAACNTTCATNACTNGGGNCGG
AAANACCNCTTCCCCTCCGNNGGNTNGGTCTCNGNAGNCTCCCANANNATNNNNCTACGN
CTGGNGNACANAGTNNACCCCNCCCTCCGTNCAGNNNNNCNCATACCNTCNAGGTNTTTG
GAGGGTNANANNNNNNCGCCNCGANTTGNGGGGNCNANNANTNNNAATNNCCCNCCAANT
NNCCTCCTTTTNNNNCNAATCCNTANTGNNGCNNNNTGCCCNNGNNCTCNNNNTGNNCGA
TNNAGNNTACTGCTNTCTNACATNNAANCNNNTTNNGCGCNTCCGTTNCNTNNNTATTCC
CNCTCCNNTCCGCCGTCGTCAAGNNNTNNANACTNTANNCTNNGGNTNNAGTANTCNTAN
TGNNNNCCNNCNGCCTANTNCGCGGAANTNNNANCATCNTCTGTGTGCAANAAATTTTTT
CAGCNCCNNNGAGCACNNNNCNGCNTNNNATNCNGNNANANNGGGGTANGCNCGGAGCAT
NTTNNGTAGNTATNCNNCCGGTNNGCGNTATNNNANNNGNNTNNCAGTNACNGNGNCNNN
TATTNNGTTATTACNTTNCTNNNNGANNGNNNNNNTNTCGNNGTNGNCNNNTAANNNGNN
TNTGANAATATNGNCNNNTNNNNNTNNNTNNNNANNNNNTNNANTNNTNNTNNNTGANGT
CNGNANNTANGTTNTGNNNNCTCNCCNTNNNTTCNNGNNTGAGNGNNAGGGNANGTNNTN
NNGTTNNATCNNTAGNNGCCNCCGANTNANNNCNTNCGCNNTNGTNNGGCGNNTAAANNT
NNTNCNGNNGNTNNCNCTGCNCNNNCCGTNTTTATNNTNCNNANNTNTNNCNTNTCNATC
NGTTNACNNANAAGNNGANCATNNNNCGANNNTNTCGTNTNNCCNNTNNNCGNCNTNAGT
NTNNGNNNNNNNGNNAGNNCATNANTGNTTNNNCCNCNGCCGNTNNNGGNNGNATTANAT
GAAGGCGGGCTTATTNGGTNNNNNCNNGAANNTANGAGGTNNGNNATNAGAACNNNTNNG
NNATNGNGNNGNNTAGGNNNNNNNTNNGNNNGGNGNGGANCNCNTTANNTNAGNTTCGCT

TNNCNNTANNNNTTANACCNNNCATNANNGNNNNAGCNNNTTTTNATCGCAGTTNNNNTG
CNACGCNNATAANCNAGGACGNNNNTTCGNNTGANTNTATGTANNTNATANANNCNNANN
NGNCTCNCTNACNNTTCTTTTTACTNNNCNCGCNNCCNNTANNGTNGATTNNNCCNNNTN
NANNNATNAANNTNTTCNNNTTNATGCGNTANCATCNTACCCTTCTATCNNGNNGNGNNN
CTTNTNACNGTNNGNNNGCCGNNNTCCNGNGNNTCGGTTCGGTTCANANCGNNTNGTGGT
NANGTATTNACTATNAATAANACGGAATAAGNACGNACCTANGANNANNGNCCGGNNCTT
ATNGNNGGANTNNTNNNNTNNCTNNNTATTCTGTGANTNTNGCNNANCGNCNCCGCGCNN
GCGTCCTNATTATAAGTNNTATATATTGATAAANCNCNGNNCATANTNNCNNNCNTTNNT
TNTATTTATCNNTCNCNCCNCACGCNNNTATANNCCG
>A195
GNGGGGGNNNAANAGANTANNCCCNAANGTNGTANCTTCACGNGNTNGATAGTNGCGCCT
GNANNTGNGGNGATNAAAANNGTTGGGGNNNCCCCCCTTNNAANANGGGGTNNTCGCGGG
TATAACNNTNNNNANGNGNTAAGTTGNNGTTCTNANNNTCTNNTTNTNNTNCNTTNNCNG
CGGCGCGNNANANCCNACTTTCTNTTCGGNGGGGNAGNGNTNCTNCNGANCNNTCANNCN
TTTTCCCCNTNCGCNTNGNNGTNANNNTNNTCNCCNCNTCNGTNCTNNGGTNCTCCGACN
AGGCGCNCTNNTNACGANGGANANNNTCNCGCACCTATCTTGCNGNCNNTNNTCNNTTNN
TCCGCCTNNNNNCGTNNNTNTTNCNGGNTNNNNTNNGCGNCTNCCCCCNGACTCAGNAGN
NNTANCNGCGNCNTCGNTNGTTANNNANNTNNNNNNCGGGNTTGCTGTNNGNCCGGNCTN
NNNNCCGGGNNANAGTNANATANACNNNGGGANNGGAANNNCNCCNCNCGCGACCTTGNC
NNGGNANGTNANNTAACNTNNNNNANTANTNTANNNTCAGGNCCGANNCAGTNNTAATGC
NGTNGNTTNNNNNNAAANNGNGGGANNNNGNCGNACNNTTNTTTNAANNANNANCCGNCN
GGANNNNANTNGAATAAATNAGTNGNNNANGTCCNGCNGNGTATANNANANNTNTNANNA
NGNACNNANGGANANNCGAAANGATNTATNNTNGAANNNNTAAGTNNNCNNNNGCATAAA
ATNNNNCCGCTNNTGNNNATANCTAGAGTANGCGNAGNNAAGNACTAGATNNNNAGNNGG
ANGAANACGNCNGNGAGGGNCAANANAATNNCCANTAGNTGGANANGNGNAGTNNAATTA
NNAAAGNTNCNANTCCCGCNCNGCTCNTNTATTTTATACNNGCATTANNGGNNGTNATCT
TAANNTGGCGGTCNCCCANNNCCNNCNNACGTTNTAAAANNNTACNNNNTNGANCNANNT
GTCTGATNNTNTCGNAATAAGNNNNTNNAAAAACNATAAANATNTANTNCNCCNNNATNC
CNCCNTTNATTNNTNTNNACNGNNGNNGTANCGAATTANTTCGCACCNCNTCNNAGGNNC
ANGNANANAAANTAGNGCNCNATTNNNTCAGNCGAGGNGCACCNTTNATANNNNAACTAN
GTATNAAANTCNGNGTTNANAGTTNNGTACTNGGNNNATANCTANNNTNNNCAAGCCACA
TTTNTTNACNTTCGAGNTCNTCNGNCNCNGTNTANCGGCAGNGTGTATCGGNNCNNAGNN
TNTATTNNNAGNGANNCTNAGTTAGGNNCCNNTGTTTAGNNAAANACGTCGCNNANTGNT
AGNTTATGCAGNAAAAGNTNAAGATCATGCTAANNGANTNGCCNNTNNTNTCTTNTATTN
ANTGCCGNTNGNTNNTNACGGNGNGCANNGNGTGNNGATCTANNNNTTGTCANNTNANTC
CGNCNNNCNGCTNNAGNTNNCNNCCCCAANTNNAAAGTNNCNAGNTANNNNAAGTTNTAN
AGCNNGTATCNNNNGNCAGGNCTNTNGATNAGGNNGCNGTCTTGGCTNGGNAGGATAATN
NTNGGAANATNNTACCGGTCNGNCNACCNAGCGNNNCCNGCAATNAAGNGTNTATANNNN
NNCNAGTNTTCNNATGANATNGNGNGCGGNTGNNNCGTTANCNNANAGNNAACTTAANNT
ANTGNATCNGNNGCTNNGGCGTTGNATTATTANACNNCTNNGNNCCNNTNGCCCN
>A196
ATGGGCGCTTANCATGCAAGTCGAGCGCCCCGCANGGGNAGCGGCAGACGGGTGAGTAAC
GCGTGGGGATGTACCCGAAGGTACGGAACAACTCCGGGAAACTGGAGCTAATACCGTATG
AGCCCGAGAGGGGAAAGATTTATCGCCTTTGGATCAACCCGCGTCAGATTAGCTAGTTGG
TAGGGTAATGGCCTACCAAGGCGACGATCTGTAGCTGGTCTGAGAGGATGATCAGCCACA
CTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAAT
GGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGACGGCCTTAGGGTTGTAAAGCT
CTTTCGACGGGGACGATAATGACGGTACCCGTAGAAGAAGCCCCGGCTAACTTCGTGCCA
GCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATCACTGGGCGTAAAGCGCAC
GTAGGCGGATTGTTAAGTCGGGGGTGAAATCCTGAGGCTCAACCTCAGAACTGCCTTCGA
TACTGGCGATCTTGAGTCCGGAAGAGGTTGGTGGAACAGCTAGTGTAGAGGTGAAATTCG
TAGATATTAGCTAGAACACCAGTGGCGAAGGCGGCCAACTGGTCCGGCACTGACGCTGAG
GTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGAT
GGATGCTAGCCGTTGGGGAGCTTGCTCTTCGGTGGCGCAGTCAACACATTAAGCATCCCG

CCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGGCCCGCACAAGC
GGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCCTTTACCAGCCCCTTGACAT
CCCGGGTCGCGACCCTTCAGAGACGAAGGGTCTTCANTTTCGGCTGGACCGGAAAACAGG
TGCTGCTTGGCTGTCGTCAGCTNCGTGTCNTGAGATGTTGGGGTTAAGTTCCCCCAACGA
GCGCAACCCCTCNCCCCTTANTTGNCCATCATTCAGTTTGGCCNCNCTNNGGGGAACTGC
CNGTTAATAACCCCCANGAAAGTGGGGATNAACTCCAATNCTTCNTGNCCCTTTNNGGNC
TGGCTANNNCNGGCTTNAATGGCNGTNNAANTGGAATTCCAAAGGCCACCCNNNAAATTN
CNAANCCNTTTCATTNAATTGCCCTTTCANNNTNGTGNTTAAGGGGATNCCCNNTNNCCA
ANCCNC
>A197
GGGGGTTNNAAANGGCACGNCNCANGAGGCAATCACANAGTGANGTAGTTGCGCCNTNAA
NNGGNTGGNTATNACNTAGGGGGCCCCCCTTTTNAAGANGNNTTTTCCNGTANNNAACCN
TCAGNNGANNAACTTGGANTNNGANAAATNCNTNTTAATGAGGNNTTTCNNNCTGGGGCG
AAANNNNACNNTTCCNNCNNGCNNTGNNANNGGTNCCCGGCNACNNNGAGCTANTCCANT
TGGTGGNGCAANCTTCNCCCNNCCNTNNTNNANNATCCNNGANGCTGGGNTNTNNGNNNN
ATGACTNNCGCACNTTTTTNGCGANTNTANACNTNNCCNCCTNNATTNTCNNCTTTTCNA
GGNNNCCTACNTGNGGANCCCCNCCNCNATGGGNANCNNGNCNCANATCCANANCNNTTT
TCNNNNCTCNNTGCGTNNNATGCATCCNNNCCNGNNNATNGNAANATATNNGCCNNGAGA
NAGNNTATGNGGNCNGCNGCNCNCNGGAANANAGATNGCTNANNNNTNNATNNGTTTCNN
NTGANCCGCGACTGTTCGNGGGTNTANNNANAGAGGGGNANCNCNTGCCTGTTGNANANN
NATCNNCGTACGNTGGNTNTNTAGNAATGNNATTATTNGGCCGCNTGTNTANNNTTNTTN
NATNNTNCTTGGGAGGNGNNANNTCCNTNTATCANNTNTCGNCNNCTGCTTGAGACAANC
TGGATTNCNNTNAAANNNNNNNGCNNNNTTCNNGTTNAANTAANNNNAAAANNCGCNNAC
NNGNTGNAGNNANNTCTCNNTNTNNNNGTNGANAGGNTNNTACAGNNANTCGGAAGNNNC
CCTNATCATTNNATANNNNANTNNGANNNCTNTCTNNATANNNNCNCTCNCTNCNCNCGA
CTATTATATANNNNGNCAGNNNTNCCANCNNTAGANTNANNATCGGNNTNCTNGAACGAA
NANAGATNNTNGCNTNNNCGCTTATGNATANNTNNNGNANNGTTNGNNAATATTTCCNCN
CNNNCCCNNNNNCNTATNNANNANNGGCNGAGATNNGNNATTNNNNGNANNNTNTTTTNN
TATACNCTCNGANTGAGAACNNCNCGANTNGNGNATANNNTAGNNNNAGTNNNNTTNNAN
TNGAGNNNCGCNCNTNNTGTTNGTAACNCNATGNANCACNGATGTTNNCANNGNNNGTAC
GNGCGCNTNNGCGTTNATTANTGNGNGNCTTANNNNTNGCNNCCGTGGCAGNTNNGTAAN
TGCANCTNNNAANNNNNTTNNTATNTCNANTCTGCTACNCGATAAACCTCCCGNAANNNC
TNNCTTNAGNTTTNTTCCNNNNANCNNCNTNCGGNNGNNAAGNNANCTACNTANATCNAT
GTGTNATTNTNCCGCNAGNNANGCNGNTNNNCTNCNCTTCNAANNTNGAGNGNCNNTATA
NTAANTTNNTTCCTCTATNTNGNACNCCTGCNCAANATTNNCCTCNNNTANGANNCNNAN
NGTGTAAANTNTGAANGAGNNTNTNGGNNACTTCTNTCNNCNCNNNNCNNCAATCTTNTA
GAANANNNNNNNNTNGTNNTANNNTNNATGANANCGTAAGTGNANCGGGNACGCNANNTT
TNNAACTTATCAATNATANGNCNCCNCCTTNCGCCCTANTNNATTATAATCCNCAGCNNT
NNGTNAGNNCGGC
>A198
GAAGGGGGGGTATAACATGGCAAGCNGNNNANAAGNGANAACCTGTTATGTTGCGCGNAT
TNGGNTGTNNAANANGNAANGGCNCNCCCCTTTNAANNGGGNGATTTNNCGGNTNNNANN
GNNTNGTGGGCANGGANANGTTGCNNANTNCNTGGANAACGNAGNTTNGCNNANNTNCNC
ANATNCCGTTCCCNNGGGNAGGNNGGNGGTCNTCGCGCCCNCNGNNNGTNGCGCNNCGNN
ANAGNCACNNGTCNCCCCTCCCANTNNGNGTANCNTNTANCGTGNNTTNGTCGNNGNNNN
NNNNGCCNNATCTTGGCCNNNAANNCGNTNAACCGGNNATNNCNNNTNAGCCGGAGAACN
NNNNCCNCNCNCCCNCTNNNAGGNNGNCNTGATNNNCACNCTCCTNTNTNACANAANNCN
GNGCCNNNTTCCCTNNANTNGNNNCNNANCNNCCTCCNNAATANTGNNATNGACNGCNNN
NNNCNNNANGNTNNNGNCNNCCNCCCTTTCGCNTNCTTNAATANNATTTNGNNTNNNGTA
GNANNCGCGNNCCCCGCGTNANANACGNNNNAANANCNNTNANGANNNGAGGNNNNNGGA
NGCCNNNTTNANTTNANNNTCCCGNNCTNNANNTNATANNNACANCCANNNGNACNGCNN
NNANNNNANTGNTNNNAGNNAGNCNANGNANCANGAGNNNTTTTNNAAANTGNGNTTACN
NNCACGNCNAGAACNANGNCGATTACNTANANNTNNNNAGTGGNGTACNTANNGNNTCGC
NTNGNACCNTAATACNNNCNNNGGTNNGCGGATTTCNNGCAANTNNNNNGCNGGNAGGNT

CTTCGNGNNNACANNNNCNCNCCTNAAACNNGAANCACCNCNTTNGCGANTNNCGGANNN
NGANCNNNCNNNNNNCCCGANCGGTTNTNAAATNNCGNNNGNTANTCNANTNTNAGNNGN
NCNNNAAGNCNANTNTGCNNCGATCNCATANATNTNNCCCNTTCCCGCTNTNNTTGTNGN
GATGCNGCCNNNNCNNAGNANATNTTCCANNNCNCNTNNCCTCNNCCCNTANCTTNTCNN
NNNGNNNCGCTAANNNTTGCNGAANNGCTNANTNGTAANNNNCNGACNNNNNANANGNAN
CNCNNGGGTNNANNAGCGNNTNNCGAATNATNCAATGGNGTNANNGNGCGCANATAATNN
TNACNTTCNTANAGNGGANCNGTNAANCACNNAGTNTCCANNNGTCCCGGNTNNTTTATN
AGNNACNTTAGCCNTGNNNCNCNNNAAAAANGNNTCNGNNAGTCNNNCNNAGATTTTNAN
NNTATNGANTNNGNNAGTNNCCTCNCCNNACTTNTGTANTCTATNTGCACNNTNANNGAN
NNNGGNGNANTGNGATANCGGNNANNTCNATNNTANNNNTNNCNGCCTNTGCCCGNCNAN
TTCACNNNNGCNNNCNNNATGTACNGGNCAGTCNTAGNNANNNTNATANNNNNNTAANNN
NNCAACNGCGCTTNGNCNANCNTCNCGNTNNANNGCTTCANGANTATGGAATANGNAAAC
NAATTTAGNNGCTNGCTNCCTNCNNCNANCNCNCCGNNTNTATGCCTNGNNTGNACACNN
GNNTTTNNCCTNGTGTTCNNNCNANNANGCCGNNCNGCNNGCAGTTNGNNNNNGTNTTTN
NTNNNNNCGTTNGGCCNGATNNTTNNTNNTTCCGCNNNCCNACCNNCGCNGC
>A199
AGGNGCNGCTACCATGCAAGTCGAGCGCCCCGCAAGGGNAGCGGCAGACGGGTGAGTAAC
GCGTGGGAATCTACCCTTTTCTACGGAATAACGCAGGGAAACTTGTGCTAATACCGTATA
AGCCCTTCGGGGGAAAGATTTATCGGGAAAGGATGAGCCCGCGTTGGATTAGCTAGTTGG
TGGGGTAAAGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACA
TTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAAT
GGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCT
CTTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCA
GCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCAC
GTAGGCGGACATTTAAGTCAGGGGTGAAATCCCAGAGCTCAACTCTGGAACTGCCTTTGA
TACTGGGTGTCTAGAGTATGGAAGAGGTGAGTGGAATTCCCNAGTGTANAGGTGNAATTC
TTNCNTCTTNNTANGANNCTCCNTGNCTTANGGCGGCTNCCCCNNTTTNTTNCTGACTNT
CANGTGCTCATNCGNGTNNACCNAANCNTANCCTACCCCTTGTNTTNTCCACCCNNTCTN
ANTCTNCCNCNCTACCCTTTTGGTCCTNNTNNTGTTTCNGTGGNTCTNTTTACNTNTTNT
TNNNTTTCNCCNNTTTACTTNNGTCNTCACCNTCNNNTCTNNNACNCTTTTTNCNTTTTT
CCCCCTNCNCCNCNCNTATNTTCCCCNCNTCCTTCCTNNCCCTTNCTNCTNNACNTCTTN
CTTNTTTTNNTTCCCCNTTTCCTCTCCCCTCNTTCTTTTTTTTTTCCCNNCCNTTTNCNC
CTNTNNNACNNTTNTTCTTTTTTTTTTTCCCCTNNCCCNTCTTTCTTCCTTTCTNTTTNT
TCCCNCTCTTCTTTCTTTNTACCCCTTTTCCNCCCCCCCTTTTTTNCCCCCTCATNTTTN
CNCCTTNNCTTTNTTNTNNNNTTTTNCCCTNCGTCCACCNCCTNCCCCNCNCTTTNTCCT
NCTTTNTTNCTCTCCCTTNTCTNTTTTCCCCNCCNTCTNCCNCTTNCCCCNNCNTNNCTT
TTTTTTTTTTTNNTCCCCCCNTTNTNCCNCTTTTTTTNTCTTTTCCNTNTCCNTCCCCTC
CCTCC
>A200
GGGGGTTANNNTGAAGCCNAGNNNNGCCNGNTNAANTTNGCACNCNNGNGGGAGNANANN
TNGGCACCCCTTTANGNCANNTTTGTGCNTANNAGCNTGTTNNNGNNGGNANAANNCNTT
TTNAGGGNGAANTTCCTAGGNCGCAGGTCCNACTTTTNTACCGNGGTAANGTTCGCGCNC
CNNNNANANTATCNNGCGNTNTNAGGGTANACGTTGTCCCNCNNCACANNNNNTNNGCGN
CGNNNGTNNTNTNTGNCGNNAGNANCAANCGTCGCNNNTNTCCNANGNANNNGTTNNTCC
NNNTTTATNCANNNNTANATNGGGANCCNTNNNCNNCGNGNNCCGCNAACCNCCNANNAA
GNANNNGGTCNATNNTATTNAGNCNNTNNTCGCNNAGTTNCNNNNTATTCNCACNNACGC
CGCGNGATAGNNNNNNATNTNGCGNNTAATAGNNAAAGNNCNGNCCCTNACTCNNCGAGA
NNNNNTNTTNNCANNTTTNTTANNGCNANCNANNNGNTNGGCGAGTNNACCGAANANAGN
GGGTNCCGACNNNGNNNTNTTTATNATNCCGGNCCANNTNATGAAANNTNTGANNNCCGC
NCNNTGNTTAATTAATGAGGAGNNNGAAANANTNTGAGTANCNGNNATNNNTCCGTNGNG
NNCNCCCNNTNTAAGCNNNAGANNNCTNNAATTTNCGNANACATANTNCNGNGANTCTNC
NNANGANCNNGANNGGGNGTNGTTANANNNTCACTTNCCGCNCATTNNTNCATNACGTAT
TCGNANTNNCNANATGTNCGGACTCGCTCCCCCNNTNTTGTNNNCNNTNNNACCNNTCNN
GANNTNANTACGNNNNNNNATCNNGNATNANTCNNNAAGNNNTTATTTGANNTNNNCGNNT

NNNNNNNNTTCNNCACCCNACCGGNCGNNTATNANAGNNCGNNGAGTNTNNTAGGGGANN
TNCTNATAANNNGGGANNANANCGTCNCCNNNNNANGNNGANTNCNGNNANTAGTCCNGN
GGNANCCCTTTNNTNNGNTGCNCNCNGNNANAGNTATNNCCGNNNTCNNNNGCNCGAGGN
TNATTNANNNNTTNTNGAAAANCCTNTNATNNNNANCNGCNNCNTNTCNANNNTTNNNNT
NNNTNANNGNNTTNNNCGCNNANANNNTCTGNNTTTTTANANNNTCGACCNTNGGNTGGA
TAGCNNCTTNGNNTCTTNNNTCATCTTCCAGACANNNCNTATNANCCTCNCTTNNNTNCG
TNGNANNGCATCNNNNANTTANANNGCATTTGNCAGCNNCGNTGNNNNNTCNANGGTTAN
NGTCTNNNTTCNATACTNTANTNGNNNNNGGNANCCCCTCNTCNNNCCNNNNCGNCCNCN
TNTAGNTNNAATAAANGCNNCTNTNGTTCNNCNGANNNNCCCGNCNNNNCCGNTNCNNAC
TNNTNGAANNNTNATATTNTTANANGCNNCCTTTGNCCNCATGNNANTCTTTNGCACAGC
CNGNNCGNTACCNCC
>A201
NNCCCCNATNGNNAGGGNGGTTTTGGGTTTATATANATANTGANAAANNGNGCTGNCCCT
TNCANGGGAGNANNTTCTNCNGGACGANATAANTNTNTTGTNCCGGGGCATGNTACNCGC
NGNTTNGANTAGCNAGTTGNTGAGNCAANNACCGCACTTCANGNTNACGACTCNNTAGNT
GGNTTGAGAGGATNNTAACNCCACTACNTCGNNCTNANAANAATTTCCNNNAATTCCTTA
ANTNAGGNTANCGNATNGGGAANTNNNGNNNNANTNNNNNGANCGNTCCCANNTNTNCAN
CNCTCTGGNCNTTNCNNANATNCNNCNNCNNCGCGNAGNAATNANTTACCCNNNANNAGC
AATNNACNGNCCNNTCNCNAANNNNNACNCNNGCNAANTNGTTGNNCACCTANCCNTCGA
GTNAATNNCNGAGANGGGNNNNNNNNCNNTNNNNNNNTNCNGNNCTNNTTAATTCGCTNN
GNANGCGCGCNTTNTTNNAGNTTNNCNGGNCGNGAGTNTGGGAGNTTTAAGACCNGGAAC
TGTNCTNTCNNANCTCGCNGTCNAGNNCATTNTGCGTNTTAANAGNTNANGGTNANTNNT
NCNCTGTNANNAGGNTNTAANTCTNNNNCNTANCCNCNNTTNANTAANNNNGATTNGNNN
NNCTNNCNCTTCACNNGNCCNCCNTTNTNNNNNTGCCNGNANCNNCCAGNNCNTTTANNG
NNTTAAAANTNNNATNANANCCCNTNCNCTTNNTTTNNANCNNCNGTCTTCNNTNNTCNC
CGCGNCGNCCNNNNTNTNNGNAGGNTNTNTNNTTGCAGCGAGAGTTGNGTNATTATANNC
NCNTTGCGTTCGNGNANTATNNGCTNCCNNNGGGTCGNCNGCCCATNNTANNTNGCNCTN
GGGNTCTTTNNNGNCCNANGNNNNGCNNGCTTNAAACACCGCCNTGTCTCAGACGCANNN
NTNTGNANNNGCTGTNTNNTNNTNNANATTANTCNGTNTGCGAGNCNGTCTTTNTNANTT
ATAGNNNCAGNNTAACNNTNNAANANANNTTTATCTCTNCNTTATNCCCNGGNCCNNNNT
TNCTNCNCTNNTACANTNNGNNNTTNTNTANTTANNANNNTNTNNGCNCTNCNTTCNTTG
CNNGNTNTNACGNNTGTGNAGNTTNTAANTTANTNNACGNNACGTNGNGCCCCCNTCNCC
NGNNATTNGTCNTNANANNNGNTANNTTNCNTGAGNNGGNNGAGCGNACCCNNGNNANNC
TGNTTAATTTTNTNTTTTATTNNNGNTNNNNACNNAGNTATNNTTATCCCGACCNTTNNT
ANNGCG
>A202
GGGAGGGGGGGCATAAATGGCAAGTCGAAANAGGNTAACCGGTACNAGTTGGCGCACTNG
TGCTTTAACGNGTGCACCCNTNANGGGNTNCAATCCGNGNAATNNNNTCTTGGGACNAGA
CCTGCGTAANACNGGATGANATCTGCGNNNNCAAANCACTGANCGTCCGGGGATGNTNCC
GNAATCCGATTGTTCGAGTNGGNNGNGNNANCGCNCCCTCAGGCTNCTANCGNNNNNNGG
NNTTACCTAANNNACNACCGNTTATGNGGCTNNANNNNNNCCCCNNTTNCTNCNGGATGG
AGCATATGGGNNATACTGCTCCNTNGAGGNAGCANCNGNTNANTATAGNNTANTGGGNGN
ANNTATNAANCNANTNGCNGCGNAGANTCGTTTGACCGNNNTTATGATGNCNNCCCCTTT
NCAACAATNNANNAAGCTAAANTNANGNNACCNNCCATANGAAATNGCCNGNNNNNTACN
GCGCNNNAAGNCNNNTGCCNATNCGTANNNGNGNANGNNNNCNCCCTNNANTTATNGGGA
GNCNAANAGCTCGCTAGGTNNNNTGCTTGANCCGNCTNTGAAACNNCGGCGNTAGATNCN
NNGGNGNGTNGAANGNGANGNCNTCCGCTNGNGNGTCGNTNCNCNNACAACCCGCANTNA
NNTGANCCCNGTGGGNGNNACTGCNGNANCACTTNCGCNNNNTAANNTANCAGNNGGNNA
CNANNNANTATCNTTTGGNNCTAAAACNTGATCTATANNCNNNCCGNNAGTTTTTNTNNN
GNCGANNCNGNNNTTCGCCCNNCNCNNNGCNGTNTNCNGANGGNNNNCNGNNTGGGNNCT
ANNNCNTNNGNNTATTATTNCCNCNCCNNCNNNGTNTTGGCAGTTAANNNGTNNNCACNN
GNNTNTNATTNGGTTNANANNCNTTTCTCCNGGGNAAANNNGCTCNNTTNTATTNNNNTN
GGNNTNNGACNNAATTANCCNGANNGANGATGTTNTNANTGTANAGTANNNTACNNAAAN
NNCNCCTTNTNGTTNNGTNNNTNNCCGNGNAGTTGTGANGTNATAANTGANNNTTACTTA

TNNNTNATTCCCNGAAGGCNTANNNNCCNCCCCNTNTAAAAATNGNTCNNNANTNTANTT
NGACNNNNTTNANATNCNTCANTCCNCNGAACNCGNTTNTCCTCNACCTNATAGAGNATT
ATGAANNTNGTCGNNTCGTCTCCNGCCCNCCNGNGNANAANNTNNAANTNANNCTGCTCA
ANTTNGNATAANNCGCGAGAGTCAGCNCCTANANGNNTTTTTCTTTTTTTTTNNGTCANG
CNNTTANNNNNGNANNTTTTNTATTCCTCNCCCCTCNNCNCCNGC
>A203
GTGGNTANCANNNTGAANGGCCNGNAAGNGGGNNNNTNCCCTGGGTNGNTAGTTGCGNCN
NACNGGGGNNGGANNNNANNTGTNGGCCCCCTTTNNNNGNGGAANNATTNTCNGGNNNAA
CACACTTNNCGNTNGGAAANTTGGGCTNGCNNNNNATNNNNNAAGGTACGNNACTGCNNG
GAGNCAANTNACNNTTNCNNTCGCNGGGGAANAGGNTACNCNCGCCNTNCTCCGCACTNT
ANCNCAGGTTGNCTGNGNNTANNTACTCCCCCGCTNANNNNANNNNNNTACTCNTCTNNN
NNTGGGTCTTANANGCNNNCATGNATTCACGCTCCATTNNTTAGCGCACTANATNANNTT
NTACNNCTNAATTNNCTNCNTTTTNAGGCGAGCCANCTTNNGNCGAACACCTCCNNCCNG
TNNCTGNGNANGNTCNNCTNTANNANNTTAANNNATNTNNNNCTCGNCANTTCNGTNNAN
ATANCNCNNNTTCNGCGNANCGCAATAANTCNNGTNNCTNNNCNTNTNAANNNANTTACG
GNNNCCNCNCCCTTNTCANNNAAATAANATNTTGTCCGAGNNAGNNATNNTTNACCCANN
CNANTCCNTTTCGGANGTNTNCNAGAANNGGGGGTNCAGNTGGNCNNNTTTCNAATNNAT
NCNCCNGTNGACGANGCTNNACNGCCNGNNNNANGCNGGNGNNTTNNATNNNANNTGANN
GGGATGGANGNNNNNNGCGATNGNNGTAAANNNNNCNNGNAANGTTNCCCGNNNTGCNAN
AANTNGCAGGNNNNNCGTNNGGNTNANNGNGNCANGNNNGNGGTNCNCNNGTAGNNTNAT
ACCAGNTGNNCNGCNNNNNTGNGNTNACTCATNNCNTNCCCGNNCNTTNTTTNNANNNNC
NNNTTNGGNCGTTTCGTGCTCANNGNCNNCNCTGCGCNNNCNGTTTTNNNAACNNNNCGC
NNNNCGATCNNANNANACGANNNANGCNNTNTANNNNCGTTNNNANCNAAACNNNTTNTC
CNNNNTTGTTTNNNNNAAGCNNGCNNTCATGNTNTNTCNGCCCNNCNTCNGCNNGNGTAA
TNNAANNCCACNNCGTNTNTGCNTNGCGGGTNNNNTNGNNTCAGNAAAANNTAANNNNCC
CGCTNGGNAAGGANNTANGTTNNCCAGNGNTNNANCGACCGCATANNATNTNNNNNATNG
NNCNGNNGNTTTNGCNGCGNTNNNCGGNNGCCANCGTTTTGTATNGNNTNCTNANNTGNA
AGCTNTNATGAANCTCNNANCTACCGTGTTNTNATTATTNNATTTTATNNANNNCGCCNA
NCCGNNNGCNNANNNNNNTTNNTTNNACNNANTNNNNCNNNNGGGCTGTNNTNCNNTTTN
NCNTAGTTTNNNTTTGCGNTGGAAANGNCNNTTATCCTCCCTTNNATNANANNGNNTACN
TCNGAANNATTNTTANNNGTNNNTCNNGGTAGNCGNTTTANNANNANGCCGNTCATNNCA
NNNGNNNAGNGTGTATAAATATCGANNNTCTANCGANTCGGTNNCTCCCGNGNGNCGCCT
NNNTNNNNGATNNGNNTNGTGNANCNGTGCTTTGACCGGTCATANCCNCGCNGAATCCGC
NGCNCGNGNTNATATTNNNNNNNNNNNTNANNTANNTNCCCGCNGGTCNAAGCGNTATCN
NNTTATTNTTTCCGCNNCCTGCCNCNNTNCCNTAN
>A204
GGGGNNTAAAATTNNAAGNCCNGAGNNGCCNNNCCAGNGANNTAGTTGGCAAANTNGGGN
GNANNNNNNNGGGNGCNACCCCTTTNANANGGGTTCNNNAAAGACCTATNGTNANAANTN
TGCTGNNANAATTNTTTTNTGNACGNACTNCTCGGGNAANTANACCNTNCCCTCNNGGGA
GGNTCTCCGNGCNCNNANNATNNCANGTNCGNTNGGNNNNNGANCCNACCNACNNNTNNG
NNNCNCGTCGNNNGNGNNTTTNGNGANTNATCGGCNCNTTTGGGCNGATNACGTTTNNCC
GNTNTTNTCTNCNNNTCGGCTGCTNNTTCNGCNNCCCCNCACNNNNGNCCAGAANGTCGA
NTTAATNCNTGCTTNNCANCGCTGCNNNNTANNCCCCNTCCCCGNNANANNNTCTNTACA
CGNNNTANNAATTANCCNCGNNCCCCTCANANAANAAANTCNANNGCNNANTTCTNNGAG
NNCAGNNANNNTNGNGNGGAAATNGNANAAGAGTGGGNNNNCCGTCNNNNNTNAAATTNA
CCGNNNGNGNNATAANNNCCANNATNGNNCNNNATANATANNATGGGNNGGGAGGCAGAG
TTCNNGNAGGNNTTTTNNNANNTNNAANNNNCNANTCNNTNNNNNGTNNNNNTTNNTNAT
NNCNGGNNNTTTAGTTTNGGCNANNAGNCNNCGNNTNNGGNANTNANANNANCTCATCNN
CCGNGNNNTNAATNCNANTACNGGGNNCTANTCNNNGCNACAGANCCCNNTTNNNCNATC
CNNTNANTCNNCGNANANCTNNNTGNGNTAGNACNACNATCNTCNCCCNACCCGCNANNT
NNTNNCANNNCNNTTNANNNAANCACCNCCNNNNNNNNCNANTNCACNACNGNNTGCNNN
TNNNNGGNGNTNNATAACNANATNNNTACTTCGNGTTNAGNNNNNTAGTNGCANNCATNC
NCNCNACNGCGTNTATNAATNCGTNGNTGCNNNTTTTGNNCCNTTANNNGCNCNNNTATN
NNCNNNNTNCCTNCNNNGAAGTTTNTCNNACNTCCTCTTTTNNNGNGAATNGATNNGGNG

NNTCNNCGCNNNGCTNANTTTTNNTNNANNNNNANNNNNNCGNTNAAGNNNGCTGNNNAT
TNNNNNNNANTCCNNNNNGNCCTTNTCNTCNNTNNNNATCAANAGNNGNTNNCANNNTAT
NTGANCCATNTACNACCNCGNCNTNNANAGCCNCNNTATGNNCNTGNTATAANNNTNATA
AACTNAGNANCCGNCCNACCTCNGCCNNNCCACACTNTANTAGCGGNTNNNANNANNNNT
TNNTGGCANNTAGGCNNAGNCGNTCGCGANCNCNCNGNTAGATANNTNATTNTTNATCCC
GCNNATCNGCCNTANCTTNTTATATTCCTCANCNCNCNCNNNCNTCC
>A205
GGGNGGNATNAAAAGNGNNAANGGCNCNNGGGTANGCGACCGNTGGTNNTNGTTGGGCGN
NACNNGGGNGNNNNNANNCNNGGGNGGGACNCCNTTNNAANGANAGGGTNNCGGGGTANT
AANCCTTANGGTGTNNANTTTAGGNGANTNATANCNTNNTTNANNGNAACNNTNCTNNCA
NGGAGGNAAAATANGCCNATCCTCGTNNCGNAGGAGTGANTNCCCGNNGTCATNGCATAT
ATTCCANCNTCGTNCTNGTGTAAAAGNNGCGNNCNCNACGGNCNANNGCNANNCGTATAN
GCTGCTCNTTNNNNNGTNNTAATAACGTNNNTNNTTGGCNANTNATTTNNCNTNACCACT
ANNTACCANNNATCGAGNGANCCCNGNCGNGNANNCCTGGACACATAAGNANNTGAGGAG
TCTGATNTNATTGCCNNTNTTCNGCNGNNNTTNCGTNTANNAGCNNNCNNCNGCCGTCGA
AANNANCNTNTCNCNNTCNNNTGANAANTTATGNNCANGGCCNNTCGCNANGAAATATAA
AATCTNNNNNCNTANANATTTATNANCCGNNCGNNNNCNTGNNGNGNGCNANTCCNNCAA
AAGGNGNGNNNACGGTTCCCGATTTATNTANNAGACCCNNNTNNNTGTTTANAANATCTA
TNNGNANGNNCGNGCNNTTNNTATTTNANATTGTGNGGGGGTGNGNGAAANNNNACNAAN
NNGTATTGNGCNTNTAGTTAANTNNANNGCGATCTCNNAANTNGCANNCAGGACGAGNTT
TNTNTACNANNCNAGTNAGANAAGANNNNTNNNGCNGTNNNGANTCATCCANNNGNGGNN
NAGGTNGNGGNNNNCNACNNNTNCNTANCGNGCCCCNTTTNNTNTNNTNTTCNNCNANCT
NGNNNNNTNTCTNGANCNNNNGCCNCCNCTNCGCCNCCCNTTNTANCCTNNCNNNCNNTA
NTCCTATCTTNNTNTTTCGTTATTNNGNGNNNNNNNACNATNNNANCATTTCNGCCNNAA
NCGNCGTNTNNNTNCTTTTNNNGNCGCGNNNAGNGANANTTCNCCCNGNCCGCNCTCANG
GNNNNTAATNAANNCGNNNNAGCGCNTNAGTGNGNTAGANAAGNANATTNAACNTNNANN
NNNTGTACGNNGTTNNGNNNGNNNNNNANNTNNNNCGNTGACTCNNNCGNCGCTTNTNAN
TGNTGACGCNTTGACNCGNCNTTTATNNTNNNCCANNTNGGNNANGAGTGTTTNTAAGCC
CANCCTCNNTGTNNANGGNNNNCTNAANNAAGGACCATCCTAGNGTGNTTNTNNNANTNT
TTGNNCGGTGGNTANNACCNCGNTNNNNATCTATTNTNTNNNACCCGNCNNNGGNACCNA
GNGANTNGNTACTANNCNNTANNNNTTNNTATATAACGNCANNNTGCNCCTANNTNCGTC
CCNTTTTATTTNCNTNNNTCNNTTTTNGNTTTTAANCCCGTTATCNGCNGNCTANTNTTC
NGTTNANNCTNCTNGTCCTNCGCNGGTANNCTTTTNNNANTNNNNNTNNNGGNGTCNATA
NCTCGAGCCCTNNNGNNCGTTATTTCCGCGTANNTNGNGACGANTATNNNCNAGTAATCN
CNNGNTCTACCCNGTANTCCNNTAATANNTTTATGAANNTNTTTTTCGCGNNTANNCCCT
AANTATTNTATTNCTNNNCGNCNTCCNNCCCNN
>A206
AAANGGCGGCATAAATGGCAAGTTGTTAACAAGGAAACCGTAAAGGGTTGNATTANTGGC
AGAACGGGNTGCAGCCCNCNTGAAGGAGCNATCCTGNTTANGGNATCNTTGGAGATAAGA
TTNTCGTTAAATTNGNGNNNCNANTTATCCNCNTANNCNGANCCATNNNNTNNNTAGANG
GNTNGGTNGCNNCNCNNNTCTTTNGNGNGCAANNGAACCCGCTNCCCNNNTNNCNNGTNN
NGGNNNGAGGGACNAAGNCCCNCCGCGNNNANNNCNNTNTCCNNTTNNTTCNTNGNNGCN
CTTTNTTCCNNATCTCCNGNGAGGTNNNACGCTCGCCCNCCACNCNNTTNAGCGGTANGG
NANCTGNTTGTNNNCNTCTAGNANNCCNGGGCTNNTNNNGNNCNNTNNGCNGACANNNGA
NTTANNNTNANATNTNNGTTGTTNTTGTGGNGNCCNGCTNCCCNTTTNCNNATNNGTAAN
AANNAAAANNGGCNNNCCCCANTTGNNNNGNTNNCCNGNGNCNGNANTCCCNNTNCNTNN
ATAAGTTTTCCNGCTNNTGNAANTANATTNGGAATNNNGGCNGGNTGTANATTNTNNTAG
AGNGCGNCNNANNNNTNNATNNNNGNTAAGNNNTNAGGGNTGACNNNCNNTNCNTAGNAC
ATNTNNGNCTTNGATNNNNGNNCANNTNTTNNNNTANNCGNCGCANAATNGGANNGNAAG
NAGANNNNATTCNTTNTCTNCGGGGNNGANNNNTTTTCNNNTTATANNNNNTTAGATATC
CNANNNNNCNCNCCCGCATTNAANNNNNNTNNANTNNTCCNNANCNNTNTCTNATTNNNC
GANTNNAACGNNGANTNNNTANNNNCNTNNNCATNATTTNTNAGTNANNNAGNGTGTTAA
NTNNNCNCGGANGCNGAGNGTTTNTTGANTGGNGGNTNNTCCTNGNGGNNNANGNNNNNN
ANANTNATTTTAGATCNGTAANANANNNTNTGANGTTNCGANNNCGTGGGNCNCCTTNTT

TNNTTTCTCTNCNTNNNCCTTTTNGTNACNCNACNNNNNCTCCNTTTNTCNCCCNCTTGA
CCTNNNGGTCAGTTNTTNCTACTNGGTCTTAAANTNNNNGTTNNNTNNNCNCNTTNCNCN
CNGNNNNTCNNTNCNNTNNNNCNGGATNNNTTTCNCTNATANNNTCTNTTNTATNCTNTG
NNNATNNCCNNNTNCCCCNNTTTNNTCCNTTNTATANTAAGNCNGNGNTCTAAANNNTNN
ANCCTTATNCNNGGGGNCTGNCATTNANCNTCANANGNGANGNTNATNTTTATGTGAGNN
AAAANNGCAACGTCTNCNCNGCNGCNGGCNTTAATNTNNATTTGANGCNCNNGCTTNGCN
TGNANTNCCGCNTTANCANNANTNCGTNTAATCGNNNTGATNTNANGCNNCGTTNNNNNT
NTTNTNATTTCTNNNNCCTNCCNNACGCC
>A207
NATGCGCGNCTANCATGCAGTCGAACGATGAACCACTTNGGTGGGGATTAGTGGCGAACG
GGTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCT
AATACCGGATACCACTCTNCNAGGCATCTNNGAGGGTTGAAAGCTCCGGCGGTGAAGGAT
GAGCCCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAG
CCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGA
GGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGG
GATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGACGGTACC
TGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGC
GTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGTCACGTCGGGTGTGAAA
GCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCTAGCTAGAGTGTGGTAGGGGAG
ATCGGAATTCCTGGTGTAGCGGTGAAATGCGCANATATCAGGAGGAACACCGGTGGCGAA
GGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCNAACAGGATT
AGATACCCTGGTAGTCCACGTCGTAAACGGTGGGAACTAGGTGTTGGCGACATTCCACGT
CGTCGGTGNCCGCAGCTAACTTATTAAGTTCNCCGCCTGGGGAGTACGGCCGCCAAGGNT
AAAACCTCAAAGGATTTGACGGGGGCCCCNCNCAANCATCGGNANCATTTTGGCTTAATT
TCNACNCAACNCCNAANAACNTTTACCAAGGNTTNGANATNATANCGGAAANNTGNANAA
TNTGGTTCNCCCCCTTTNTGGTNNGTNTTCCNTTTGTTCATGGCTTTNNTTCNCTTCNTN
NCNNTAAATTTTGGGTTTANTTCNCNCANCTATNCNAAACCTTTTTTTTTTTTTNTCCTTT
GCCCCTTCGNGTTANNGGNGNNTCCCCTNNANATCCCGGGNTTNTCTCTNNANANNGTGG
NNANANCNCANNNNTTCCNCCCCCNTTTNTTTNTTTTACTTTTTNANTGCCGTTAATGNC
CTTTTTTTTNTNTGGNNCCCTCTNNNATCCCCC
>A208
AGGGGNGGAAAATNATGCCGGGCCNGTTTAACAAGGGCNAACNTGTTATNAGCGNGTNNC
TGNGCGGANNGACGNNNGATGNCGACNNCCCCNTNTTAGNNNNNGCNTTTNNNAGCNANG
CNNANNCGNTANNGCGGGGTGGGNNTNNCGNACTCNATTNTAAGNGTNGATTCANNCNNG
NCNCGNTGANTCCNCCCNNCANCTANCNTNNNNNGNAGTNGGNGANNCNNNNCCGGACTC
NNCNCGTACNTTTNTGNNCGGGTNGTGACNCTTNANTTANNGNNACTCNANTNNNGNANC
NNNANTACCTNACNNNCCTGNNANANNTTTTCTTCGCGNNTNNTNCNNTTNNNNCNNCTN
TNTNTGNNNNATCTGANNANNNAANTGCGCTNGNCCNNCACTNCGTANGCGGTNNGCTCT
CNGNNAATTNACNATATNCTATCNGTCGTTGCNTTCNACNNNCAGCGNTCNNGCNCGTNT
NNANNNTTNGTGANGGANGNAANNGGTNCTNCNNNCCNNANCTTACNNGANACGNGNTGA
TNGATNNANANANTANNANTNGGNCGTNAGATNNGNANNTGNNNNGTGNATGNCCNNNGA
GNCAANNATNGNCNCNTNGNNTNTNGATAATNNTGNTCCCNGTATTTAGTANTNTNTGTT
GGNNNTGTGGCATCGNGCCNNNNGNNNNATNNTNANNNNANNGNAGCTNNGATGNNNTNN
NGGCTATANGTAAAAGAAATGTNANGAAGNNAGCGGNGNTCNNGNGNNNAAGGNGACGNN
NGTGTGNATNNTNNNGGNANAGTANTTNTNNTNNNNNTGNTCAGNNTTTCNNNNTANTCN
GCTGGAGAAAANTNANNNTGGTNNTANCCCNNNGTTNNNTTTANGNNGNTNTACNCGNNT
ANCNNNCNANTNNCGGNCTCNCGCCNCNTCTTNTATTNGATCTNNNCNAAANCCNNCNNN
TGNNTNNATNTGANGNNTNNANATCGATAANATCNANNNCCGNTNNGNCCNATANTANTN
NNTANNTNNANTATNTNTNTNTNNCGCCANGCCNTCCGCNANCGTTTATAANNNCCGCNC
CCNNTATNAGANNTNCGGCGANNTTTTGNNATNAGTGNGNAGAGANNANNNGNANNNNGA
TCTNTNNNNATAAGNCTGNGNTCGNNATCCGCTTNTTNTTGANTCNNCTNCNNCTNCNCN
NTTNATNNCANATTNCCGGGCNNNNCCTTTNATACNNCANTGTTGCGNNTNAAGGNNCAG
CTNAGNNGCNAGCCGTNGGACNCATATNAANAATNTNGGNTNNAGNNNCNNGTAANNGCG
NNNGNNNTTTNCTGTANNTNNNGNNACCNTCNNNTGCTNTNNGANNNATNTAATNGTGAT

TNNTAGCTTATAANGTNNNAANCCGCNNGNNNCNNGNANTTATCNTCNGGNNGTATGGCN
TNGTNNAANTTNTAAGNNTANANNCCNCTNGNTNNNTGCNANGGCNGAATAGTTNNACNA
NNATTANNGCANNTCNNTNANNTATANNTNNAGCTNNTATNGNGNGNCNGTCANTCNGCN
GCANACGCNNCNNNCATCNGNNNCTTNCAGNTNAAGANCTNNGAGCNGTNTATNNCGCGN
NNNNNANCGCNTTGCNGCNNCNATNAGTNNTNNTANNTANNGATATNNNGCGCTTNNNCC
NNNTGATNTTANTNATANCNGNATNGGATCGCNNCTAACGNCG
>A209
AGGGNGCTAAAATGCAGTCGAACGATGAAGCCCTTAGGGGTGGATTAGGTGGCGAACGGG
TGAGTAACANGCTGGGCAATCTGCCCTGCACTCTGGGACAAGCCCTGGAAACGGGGTCTA
ATACCGGATACTGACCATCNTTGGGCATCCTTGATGGNGGAAAGCTCCGGCGGTGCAGGA
TGAGNCCGCGGCCTATCAGCTAGTTGGTGAGGTAATGGCTCATCAAGGCNACGACNGGTA
GTTGGCCTGATAGGNNGACCNGNTANACTGTNACTGAGACNCTGCCCNNACCCCTACCGN
TANGCNNCNCTNNTGANTNTTGCNTNTCGNCCNCAANGCTNTNATCCCNNCNNCCCCCCC
AGTGANTNGATTANGNTCTTCGATNTTCCNANCCTCTTCCCCTCACNNTAANAAGNTCAN
NNNNTCTCTCCCNTTCCTTANTNTATTTCCCTNTNNANNTNCCCTNCCAGTATTTTTNTN
CNCCCTNTNTTATNTTTNCTNTNTTNNCCCCCTATANTTNTNTNNGNNGTNNNANNACCT
NNNTANCNCTTCTNTNCTNNNTTCTCCCNATCTAATNCTCNNGTTTCTNNANTNNNCACT
CNCNTTCTTANCATTNTANNTTCNTATTTCCCTCCGCNTTCNTNTTCTNTCCCCNCATTT
TNTTNTCNCTTTNCCTNNTNCTTNTCCNNNCNATCNCTCCCCCTTTTTTTNNACTCCNGT
TTCCTTCTNTTTTTTTNTTCCNTNTCTTNNNCTTCTTNCTTTCCCCCTNNCCCNTNNATT
NTTNTNTCTCCNCNCTNACTNTANTTCCCCCCCNCCCTNTCACTTTATTNNCCCCCCCNT
ATTTNTTCCCCNNTATTNNTTTNTNTNCCNCTNTCNCNATTNCTCCTTTNTTTTNTTNTN
CCNTNTNTNTACCTTNNCCCCCNTTATTNTCCTTCNCTTCTCCCCNATTTTTNCCCTTTN
NCNCNNCCCNTTTTTNCTCCNCCTCCNTCCTCCCCCNCTTNNTNACTNCACCTTTTTTTT
TTANTTTTTCTCCACTCTTTNCCTCCCCCNCCTATNTTTCNTTCTCCCCTCNCCNNACCA
CNCTATANATCCNCTNTTNCNTNTTTTTTTTTCCCNTNNNCCCCTNTTCNCCTCCCTTNT
NTTTTTNCCCTCTNTTTNNNANTANNNACCNCTTNTTCCTCTCACCCTNNTACATTCNCA
CCNCTTCTTCATCTATCTNATGNTTTTTATNTTATNNTNNNCANANTNCTCCTTCNCCCC
NNCNCCTNATTNTCTCCNNTNNTNTCCTNCTNNNCCCCTNAAANCCCANCTCTCCANNTA
CCTCCTCCTTTCANTTTTNTTTTANTATNTGNCCCTTATNCCGCCCTCTNTTTTTTTATC
CTNTCCCNTCANCTNCCCCN
>A210
GGNNGGGGCGCATAAAAGGCAGGCGTAACAAGGTTAACCGGTAANAGTTCGCTTTTNTTG
NTGNNGTGNATNNNCGCCCCNTTANAGGANNNGCNTTTCCGGNTANGNCGNNATGNTNGA
NGGNATAGTTGGNNGNANNTGNATTATCNNGNCNTNCTNCNNCNCNCGNNNTACNCGCTN
NNGCGTCGNNTCGGNNTNNNNTCTNTCCNCCATCNTCGCNNNCTNTCCNCCCNGCGNANN
ANTNTACAGGCNTTCCNNCCNNCCGCANNCGNNTNNNCNTCNNNNNNNAACNATNTCTGG
CANCNNAAGGNNNCNNNGGTNTNATCNCNNGTNTANNNCGCNANTTTCCAGTAANNTANG
NNACNTANTNCNCNTACNNTACTTNCNCTCGNCNNCTCCNGCGAANNCGNANNNNCGNCN
CNCGNTNNTNTNACNNNNTTNNTACCNGCNCNTTCNCNTCNNATANGNCGTANCGNCCTC
GGANNGANNTNAGGATNNNCNCNNNNANNGATAAANNNNGGNNCCNCNGTNNNGCNGANG
TCNNAANTNNTTTNGCANNNTTAGNANNAGCCNCTNCGCNTNCTGNNGCGNANTNTTCCG
NCNGNAGGCAGNTNCGNNGNNGCGTNTTNTTNNAAATANTCGGGNACNNANNANANNNAG
ANNGAGACTNNACGNCGCNTGTNAATNTGTTTTAGNNGNANNNTNNNGGGGNGCNGTCTN
NCNTNNNNCANTNTNNGGNNNGNAATGAATGATGNCCGGTNNTCNNNNGTNNTNANNNGN
CNNTGNCTNNNGNNTNANGCGANGGGANATANGNGGATGAAGGCGNTTGNCNANNNNNCG
CGNCNTACTNNGNGGGCNNNGTTAATCGNCANNTTCANNNCNCGCNNCNTNNNNTNGNNT
TNNNNNGNGANNCCGCGANNGNGACGTAATTTCNNACNNCGGCCTCTTCGCGCNNANNNA
TGTANCNNAGCNNANGNTNGNNNNCNNATAGGTGTAGGTTCTANTATCTNGNCGATCTNN
ANACGGNCNATTNCGATNNANCNNCNANNGNGNNGTTANNATNTANTTGNNGNGGANCNN
NGTGNNNNNAGTATNNTNGNCGNNCNGNNAGGGGANNNNANNTNNTGCNNNNGCGANCGN
NGTGNACNTAGGGGNNNNNTGAGNAGATAGNGTAGNATNNNTNGNNTNATNCGGNGNAGN
ANTGAGTNNTTCANGCTNNAACANTANNGNGNGGCGNNTCNTNTTNNNAANNCGCNTNGA
GNANTNAGTAATTGNNNTGNNCNTANTANCNNGNNGGAATGCTTTAGTGTNTGTGNTACN

NANNTANNGACCGANACNGNAAATNNANAANAAACNNAGGCATTATNNTGANANTGNGTN
ATCNTNTGNGTANGCGNCGCGCCGTNTGNNTNTGTTAATATNGNNCAGTNTCCNCNNNAT
AGNGNTGNAATANTNTCNNNTNNCTNCTNNTTNNNNNNATNGCGCNNNNNAAGNCGNTNA
NTNCGGTCNCGNTNTNATAGTANNCGANGNNTNTNANAGTNTNANGNGCCGNCNNTGATN
NGNNGATCGGANGGNTNGCTANGGGCGCNATAGNNNNNNNNGGNNANANNGTNTNCNATN
ACAATAANNNNNNACNNNCACACGNCGNNCCNGCTNCTNNTTNATGNCAGATANTANGCA
ANAGNNGTGNGCGTAGTATNCGCNGGNCNCGGCGNNCNGNCAGCNGNTANGNNNNTCGAN
TNGANTCNCNCANTANTGCNNNNTNTTATATATTATACGCAAGCGCTNNCCGGTANCNGC
G
>A211
GGGGGGGGTGNAGAANTGGNACCNGCCNGNNTNTATATNGNACNCNGNTGANNAAGAANC
GACNAACCNGTGNGCNGNAAGNANNNNTGTANNNAGNGNCCCCTCNTNTAANNGGGGCGT
TAATGGNCNTATGNCCNATTAATNNATGTGAAGGGNTAANGGNNTGNNCAGTAANGGNAT
ANATCCGCANNNTNTTNTNCCTNNCGNGCNNTGNGNANNCNNCCTTTTNTCNTANAGANN
CGNGCANTCGCNTTTCTGNCGACNNCNCNCTCNGCGNNNNNANTCTGNACGCTTTCGNGT
NTTNNNNNGNTGCGNGCNNCGNCNNCGCNGCNTTANCNNTCNANGNANNGAGTTNNNNGC
NNANNNTNTNAGTAANTNNGTACGNNATNGNNNANGNNNAAACNACNTCNGNTANTNGTN
ANNCCCNGNTAGANATANGCGCAATNTNNACCTCCTTATANNNNGCNCNANNANTTTCNN
CGGNNNTANNNNGANACNANCNGNCCGTACNGNNCNNNGNNNNNAACCGNTTATAGNCNN
TTNNAGGNNTTANNGNNACNTNNANATTNNCCGNNTCTTTNGTTCNANGGNACGTGTCGN
CGNNNCNNNATNTNNTNNCANAAGCTNCGCGANCNNNNNNNNAGGGNTNGTNNNNTANAG
NGTANNGNGNAANAAATATTATAATACGGNGNGGNGCNCCCCTGTANNCNNCGNNNCCGA
NGNATTNGGNNCATTTAANGNTTCNNANTNAGCNGTTCNAACGNGGANTANNANGCTNGN
NTNTANCGNNGGAANNCANAAGCCTGANTNGNANCNGNNNTNATNCCCGNGNNGACNCCG
TNGATANAANNTNNNATTTCCNCGCGNGNCGTGNGNNTNNNANNNGANGNGATGTNNTGT
GAATGCGCCGCANNGATTCTATANTNNATTATGNANGATNNAGANNGNANCANANGTGTC
GCAGCNCTNTGNCANNNTANATGNGTNATTCTNCNCTGNNANCNNGNGANGAATAAAGGN
NNGTNGGAGTNTNGGTAAGNNGCANGCTNGNGCANNGNGNGGGTANGAGGTNNCNGNGGA
NCATAAGNGTNNTGNTNGANANTNAGGCGAGGNNAGGTANGGTNCNNNANNANNGNNNCN
NCGAGANANGNNNNNCNGANNTAAACNTNTACNCGNCCCCGTANTNTATNNTATAACGNN
NGGGNTNAGNNNANATATNGTTGTCGNNGAANNGNACGTANTCNGGNCGCACATNAAANA
AGAGNCATGAAGNANANACNGNTAGGNTCGAGNATCTATANTANTGGNTNANATAATANN
CGNAANTNTATCNANTGNNGCGNCTANGNCGCGNCTNNNNAATTNTNTNTNATGCNGCAN
TTACTNNNGGNATTNNTGTANNATNCGCCGCGANANGGAGGTNCGCNNGATTTNGNTGTA
NNNACNNCGNNACNNNCANNATGANNTANGGTGGCNCGTANGNTCAAGANAANCTNGNGN
NGGCAGGATAANATACNAANNGTGCANGAATNTNATATGGTNNATGNAGAAAACGCTAGN
NNATNANNNGNAACCAGGGGNACNCACTGNTATAANATGNTCNACGTACTGNNNNCCGNG
CATATNNGNGATNGGCCANTAAGCNCNCTGGGNCANATTTNAATAAGNGNGGNANGNTNN
NGGTACGCGTNTGNTGNNGNNACANNGATGTNGNANCGCGACGNTNAGTTATNTGATAAT
TTNTNTAGNNGNCGNGCATANNGTNACCGGCGNTNACGATNNCANTTANTTTGNNGTGNG
NNGTGNNGANNNATATAGCGCGTNNAANNCNGANNCNNATNTNNGNNACTNNATAAANTN
TNNNNNCGNNNNNTNCNNNCGCNACNATNACCTACGAGNTNTNNTACNNAACNNGTNNGT
NCTNGCNTCNTNAGNTNNTAGCCCNCAGNNTGTCGANAGGCTTACCACNGNNAGACTNAG
ANNNCGGNNTCNNNGCTNGCGNAANNATGCGNTNNCTATNATNAGNNGTTTNANACGTNT
ACCGTGNGGACCGTNNNNGTNCGCTNCGTAAATANNTNGAAANATGNAGNACGGCNGATA
NNATANCANGAAGAAATNCGNNNNNGCANCNNGCNTNGGNTGCTCANCNAANGGCNTNCN
TGTTAANNAATCGCNNCGNTTGNNNNNATTNANTTTNTNTATNNNNNNCNCNNCTCGCCN
NNCGCCNN
>A212
GGGAGGNGACGGNTANNACATGGCAGGNTCGTTAACAAGGGGATAACCGGGTTAATNTGG
GCTTNTCNTTGTGGGATTAANGTNNNGNNCCCCCNTNANGNNNNGGGGGTTTGGNNATTT
TCNTTNTNTNTGGAGTGNTTGCTTGNCNCTCTTNTNGTNGGGTNNTTCTCTTGGAGGGTT
TTCNNCNTTNTNTTNNGGNGANNGCTTNNNNNCTCCTTCTATTTTTTTGGCNGTTGTTGT
GNTNCNCNTTCCTCNCTCTNNTNGTGNTCTNTTNNTNTNNGGATTTTACAGTGCGTGANT

TTGNTGNNNTTCCGNCNTNNTNCGTTTTTCCCGTTTTTNNTNTTTCTNGTNNNGTNCTAC
GCCAGNCCCCCNCNCCGTNANGNTTCTNGTCTTNTTNTTTTNANNNTTCTGTNCCTTTNG
NTTTTTATTNNAGTTCCNNCTCNTNTAGANNANTNTATACTTNCTNNTNNGNTNNCGGNN
GNGNNNTTTTGNNNCNGANNGAATTNANNTNNTTCTTTTCTTTTNCGNTNTCNNTTTNNT
TTGAGTANTTTNTCCTCGCNTAGNANTNTNNCNGTNNCCCTNTTTTTTTTTTTTTCNCGN
ANTNNTNTNTNNGATTNNCTATNNTTNNNTCCGCGATATTNNTATTATTTCTTTGNNNGN
TNTTNTNATNGCATTTTTTNTNNNNNNGTTTTNTTTNTGTTTTNNNACANGGTNNTTTTN
CTNTNCANCATNTTNTCCTTNNNTNNTNTTTTTNNGTGGGTCNNTAATANCNNNNGNCAG
NTNCTATNANTTTNCNNCANNTTNNTTNTNTTGTTTNTTCGTATGTNTGTTTCCNGCGGN
TTTCNNNTCTGATGNTGNTCGTNNTGGNNNTANNNTTANTGTTNNNNNNNCTCTTTCGGN
NNCNGCNNTTTNATTNNNCTTNNTNNGTCNTTGCTNATAGTTTCTTTTNTTNGGTGTTTN
AGTCGTTNTTTTCTNTTCGGCGNTTNGGNCTTTNTTTATATTTTNCNNCGNNNNNTCAGN
ANATTCGNGGCCGNNNGTCTCTNCNGGTNTNATNTTCTNCCCGCCNTNTTTNATTGGTNN
NAGTNTGTTNTATTNNTNTGNGTTNTTTATNGTTNTTGATTNTCTTTTAGTTGTGTGTAN
NTTCGCGNNNCGNATTTTNNTNTTCNNCTTTNTGCTNNTATTNCCTNGNNTTTNNNNGCN
NTGGTTTTTTATTNNGTTTTNCGGTANTTGNTNANTTNTTTTTNNGGNTNTTTTCTTTTN
TTTTTTTTTCTTTNNNTTTNNTNTGNCNCTNTNTTTTNCTTTTGTTTCGGNTCGGGTTTN
CTNTCTNTGGTGNNGNTTNTTNTTTGTTACTNTGTACNGCTTGTNNNNGCGNTNTNGTCN
TCCNTTTTATNTANTNNNTNNNTTGTAGATANTTTCANGNCTNGTNTNAACNTNCNCCNT
ATCNNTNNTNNNTNTATNTNTATNTTNGNCTGAGNTTTTTTTTATNTNTGTANGNNNNTA
CTATNTNNGGNNTTNNTCCNTTTTTTGCNGTTTATTTNNACTNTTTTGNNTTTTTTNTNN
NNNTTGTGGCCTTNCCAGTCCTTTTGNTNTTTTTTTTTTTGTNNCCGCNTGNNCNNCNTT
NTTATTTTTTTNTCTTCNATNGCNTNCCCTC
>A213
NGCGCGCTAACATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAACGC
GTGGGAACGTACCCTTTGCTACGGAATAGCTCCGGGAAACTGGAATTAATACCGTATGTG
CCCTTCGGGGGAAAGATTTATCGGCAAAGGATCGGCCCGCGTTGGATTAGCTAGTTGGTG
GGGTAAAGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACATT
GGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGG
GCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCT
TTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGC
AGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGT
AGGCGGGTATTTAAGTCAGGGGTGAAATCCCAGAGCTCAACTCTGGAACTGCCTTTGATA
CTGGGTACCTGGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTA
GATATTCGGAGGAACACCAGTGGCGAAGGCNGCTCACTGGTCCATTACTGACNCTNAGGT
GCNAAAGCGTGTGNAGCAAACAGGATTANATACCCTGGTAGTCCACNCCGNAAACGATGA
NTGTTAGCCCTCNNGCAGTTTACTGTTTCNNTGGCGCAACCTAACCCCNTTANACATTCC
GNCCTGGGGNGTACNGTTTCNNNNATTCNAACNCATNNNNAATTGTCCGGGGGNCCCTNC
AAACGGTNTANCTATGTNGGTTTTANTTNTAANCTNCNTNCGTNNNANTTCTTTCCCACN
CCTTNNACNTCCCCNATCGCTTTTTTNTCNNCANANTGTTAACNTTTTTCTTTCGNCTGG
TCTNGTAAATCANGCGCCTTCCNNNNTTTTTCCTCCGTTNTTTNTTCCGTNANNTTTTNN
GCGTTAATTCCCTTTTCCNNTCCAATCCNNTCTCNCTCTTTTTNNNNTNTTTTTTNTTGN
TATTTTNAGNGNAANCCCCTCGNTCNNNCNCCNCNNCCNAGNNTGTTANGNNNTTNNNTC
CCCTCTTNTTTCTTTNGGNCGNGCCNCCNTACCTCTTNCNNGTNTNTTACCCTTNTNTTT
NTTGANNCCCCCCCNTTANCCCTCTNTCTTATNTCTTTTCCNCTTCCCTTCTCCCTTCCC
NC
>A214
ANNGGCNGCNTACCANGCNAGNTCGAGCGCCGCTANCAGTACGGAGCGGCAGACGGGTGA
GATAACACGTGGGCAACGCTACCTTTTTGGCTTCGGAACAACACAGGGAAAACTTGTGCT
AATACCGGATAAGCCCTTACGGGGAAAGATTTATCGCCNAAAGATCGGCCCGCGTCTGAT
TAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCNACNATCAGTANCTGGTCTGAGAGGAT
GATCAGCCACACTGGGACTGANACACGGCCCANACTCCTACGGGAGGCAGCAGTGGGGAA
TATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGG
GTTGTAAAGCTCTTTTGTGCGGGAAGATAATGACTGTACCGCAAGAATAANCCCCGGCTA

ACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGNAATCACTGGGC
GTAAAGCGCACGTAGGCGGACTCTTAAGTCGGTGGTGAAATCCTGGANCTCAACTCCAGA
ACTGCCTTCNATACTGGGAGTCTCGAGTTCGGGAGAGGTGAGTGGAACTGCGAGTGTAGA
GGTGAAATTCGTAGATATTCGCAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCGATA
CTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACG
CCGTAAACGATGAATGCTAGCCGTTGGTGGGTTTACCCTTCAGTGGCGCAGCTAACGCTT
TAAGCATTCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGC
CCGCACAAGCGGTGGAGCATGTGGTTTAATTCGACGCAACGCGCANAACCTTTACCAGCT
CTTGACATGTCCAGGGACCGGTCGCANAGATGCGGTCTTCCCTTTCGGGGNCTGGNAACA
CAGGTGCTGCATGGCTGTCNTCAGCTCGTGTCGTGAAAATTTTGGGNTTAAGTCCCNCAA
CGAGCGCCANCCCCCGTNCTTTANTTTCNTACCATTTTNTTTGAGCNCTCTAAGGAAAAC
NTNCCNGTNNATNAACCCCCCANGAANGGNGGGNNTAACCTNTAANNTCTNNTTGGCCCT
TNCGGGCTGGCNTCNCNNTNGCTANANTGGNGGTTAAAAAAGGATTCAAAGGCCNAANCT
TTCNAAATTNNAAAANCCNNNTNNTTTG
>A215
ATGCGCGNCTANCATGCAGTCGAACGATGAACCACTTNGGTGGGGATTAGTGGCGAACGG
GTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTA
ATACCGGATACGACACTCTCGGGCATCCGATGAGTGTGGAAAGCTCCGGCGGTGAAGGAT
GAGCCCGCGGCCTATCAGCTTGTTGGTGAGGTAACGGCTCACCAAGGCGACGACGGGTAG
CCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGA
GGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGG
GATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGACGGTACC
TGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCGAGC
GTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTCTGTCGCGTCGGATGTGAAA
GCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCAGACTAGAGTGTGGTAGGGGAG
ATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCGAA
GGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGATT
AGATACCCTGGTAGTCCACGCCGTAAACGGTGGGAACTAGGTGTTGGCGACATTCCACGT
CGTCGGTGCCCGCANCTAACGCATTAAGTTCCCCGCCCTGGGGAGTACGGNCGCAAGGNT
TAAACTNCAAAGGAATTGANCGGGGGNCCCNCCACAAANCANCGGAGNATGTGGCTTTAT
TNGACNCCAACGCGAAAGAACCTTTCCAAGGCTTGACATTCANCCGGAAACGNGCCANAA
GANTGGTCNCCCCCCTTGTNGTNCTGTGTTANAGGNTNGNGCATGGCTTTNCTTCANTNC
NTNTCCNTAANAGTTNGGTTTAANTCCCTCAACNACCCAACCCTTTTTTTTNTNTTTCCA
ANATNCNCCTTCNGGGNTNNGGGGGANTNNAATGAAAATTTNNTGGGTCCTNNTTTAANN
ANGNNTGGAANACCTNTAANNTTTNNTCCCTTTNTNTTTTGNTTTNNCNTTCCCCTNTCN
GGCNATTTTTCTTTTAATCCTTAGCNTGACCATTNNTAA
>A216
NNGGGGCGGGANCNACGANAAGCCNGTNACAAGGCTAACCGGTAATTANNGNNGCNATNG
NGTGGGATGTAACCACGCNNNCNNNNTTNAGNGGCGTATCNCGCTGANCCGNTCATTGGN
GGNGNAGANTCNCGAGTGNNTATACNGNATGATANGGNTANCCGCNNANGCGNNCACTAG
CGGANAAACNTGGNNATCCGNCAGCNNNTNGCNTACTTAGGATANCGTGTACCNNNTCNN
TNNNNNANTCGGTCNNGGNNNCNNTANTCCCTTCCACGCATCNNNATANATCCCNNGTTN
TTCCAGNTACNACATTTNTCCNTNNNTGCNCNNCCNAATTCNNCNCCGNNCNNNCNCNTT
TCNATNGGNGTCGTGNAANANANNTNNNGGNCNCNCNTGCATANNATCNNTNCANNNNCC
ACACNTNGTAANTGGNNCGTTTGNTNTGNTGTANNNNNNNNNNNCNCTTCGCNNTNNNCN
TGNTNNNNNNGATGNNNTCNTTCGCNCTANCGNCTACTGNANNAGNNNATCCTGNNNANG
NCAGTTNNCATTGCGCNNTNTNTNNTNTTCCNNGNNNTGATGNNTCTTTANGANTNNTCG
CGCNNNTANTNNTTTATNCNTNNNGGCCGTNNCNNGTTTNTGGNTTNNAATNTNNTGANN
NGNGNCGTCCCNNTGCNCNNGGNCTNNGNNCGTNCNACTNNTCNNAANCTNNNGNTNNAN
NNNNCTTNCAAGNGNTCCNGGGNANGNGCNGNNTTAAAGNNNTCANGNNNCNCGNNTATT
NNNNTNNCACCTTATCNCNCNTANCTNATNNCNCCNCNCCCACCNCATTNNNTCTNCNGC
ANTTCTACNNCCGNNTNTCGNTTTCNCCNANTAANTCGTNGTNNTCNTNCCCNGNNGCGC
GNNNNATNANNNTGACNCGGGNGNNGTNNNATGNCGCCCNGCNNNNTNTCGNGTANATNN
NGCAGCGNAGATNGATTGNNGNNTNCNNGTTANNGACNTCNNTTANGTNNCGCCGGNNNG

ANTGNGNCTNANNTTTNGNGNCNNTCAGNNNCCNNTNTTTNANNACNCTTNNCGGCCGNT
TNNNNCAGNGTNCAGGCNGGGTNNNTTGNGANNNGNTTCCTTCCCACNTNCNTNGNNANN
NGNNACNTTNTTANTNNATTTNTATNTNCGNNNCATTNGTCNGNNTNCGNTNNNNNGTTG
NTNANCCNTNCNTNTTTACNCNNAATNGNNCGNGCTNNNNTNTTNGAATNNTGNCCNTNC
NNACCNTNGNTCTCCTCGCTTNTNTNTNTCAGNNAANNNTGTTTTGNTTANATCCGNGTT
TTCGNCCGTGCNCCGTTTGCTTNTANCGCNNTTGTNCGANGANNTNNNTTTATNATNNGA
TNGATNCCNANCGTCNTCNNNNNNGGNTCCTNNATTNGCTNTTNATNTTCTCCAGTTNTG
NNNTNTNANCCNCGTCCNNNNCNCGNTCCNNNNNNGTTANNTCTTNTNATATNAGNTCNC
CNNGTTNNCCTCATNNNTTTTTTTAGNTNGNCNTCCNNTANCCNN
>A217
ATGCGCNCCTTNCATGCNAGTCGAACGGCAGCGCGGGAGCAATCCTGGCGGCGAGTGGCG
AACGGGTGAGTAATACATCGGAACGTGCCCAATCGTGGGGGATAACGCAGCGAAAGCTGT
GCTAATACCGCATACGATCTACGGATGAAAGCAGGGGACCGCAAGGCCTTGCGCGAATGG
AGCGGCCGATGGCAGATTAGGTAGTTGGTGGGGTAAAGGCTCACCAAGCCAACGATCTGT
AGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGC
AGGATGAAGGCCTTCGGGTTGTAAACTGCTTTTGTACGGAACGAAAAGGTTTCCTCTAAT
AAAGGGAGCTCATGACGGTACCGTAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCG
GTGATGTAAGACAGTTGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCTGTGACTGCA
TCGCTGGAGTACGGCAGAGGGGGATGGAATTCCGCGTGTAGCAGTGAAATGCGTAGATAT
GCGGAGGAACACCGATGGCGAAGGCAATCCCCTGGGCCTGTACTGACGCTCATGCACGAA
AGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAACT
GGTTGTTGGGTCTTCACTGACTCANTAACGAAGCTAACGCGTGAAGTTGACCGCCTGGGG
AGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATG
ATGTGGTTTAATTCNATGCAACNCGAAAAACCTTACCCACCTTTTGACATGTACGGAATT
TNCCCAAANATGGCTTANTGCTCGAAAAANANCCGTAACACAGNTGCTGCATGGCTTTTC
NTCACCTTCTNTCNGAAATGTTTGGGTTTANTCCCCCAACCANNCCCAACCCTTGTCATT
ANTTGCTTCNTTCANTTGGGCCNTCTATTAAANTNCCCGNTNAAAACCNGAAGAAAGTGG
GGATGACTTNAANTCTTNTTGGCCTTNTNNGGNGGGNTNCNNCCTTNTTNATNGNTGGNT
AAAAGGTTNCCACCCCC
>A218
GAGGGGGGNGCTATAANACGGCCAGGGTCGATAACAAGGGTTAACCGNTAACATTNTTTA
NGNTNTGTGTNTGNCANACGCNNNNNNNCCCTNTTNNANNGNNNGNANATNCTCGTTCTN
GTTGGNGGGNNNNNTNNAGNNATGAANNTNTACCGCNTTNAGNNCNGGCNGGGAAGAACC
CCCTATNGTNCGNCATNGGNGGNNTNTCGNCANNGAGNGNACTNNGCCTACNGCNGANAT
AANNGNTCACGNCTCCCNTTCCGCTACNTTCNCACCNNNCAAATNACCNNGGTANNTTGC
GNCNCATNANNNNNGCNGNNNNCCTTNNCCCCTAANNTANCNANNNTCNNGCNNNGNTNN
TGCCNCTCNCNCGGCNGNTCNNNTNCNTAGGNNNCTNCGTTANTTANCATNTNGGNNCCT
NNNNCTNNTNTACGCTCGNCCCNNNGNAGAANATAAAANNAANNTTNTNNANAANTANCG
CGGGNNNNGGNCTTNTCCCGGNNNNATATNNNGNACAGANATTATNNGACCNNCNTGNNT
TANNCNNATNAACCGNNANANGNGATNNNNNNTCNNTTAAAAANAATNNNNGNAGAANNT
GNANAAAAGNTANNNNGCCGCNNCTATNNTNATATNNNANNANNCGGNGGNNGGANAANN
NNGNNANGNTNNATAGCACNAAANGAGNNAACNGNGNNCCNNNANGGGGNCANTAAGTGA
GAANNAGGAGTNANNGNNNNNNNNTNGCGGNTNGNGNNANNGCTCANTNGNGCNAGNCTAG
NNANNANNCTNNTCNCTNCNCNNCTCNNNNATAATTCGNCGANAANNGANANATCGNNNN
NNTNCNGNNCGNNANGCGCANATNANAGANNGNNATNGGNANACNATCTANNANNANGNT
ANANCGGANNNNANNATCGTTGNCTTNANNTNNCNNCTNNNNCNTCGNNAATTTTNANGA
AGTGNNACTGTNATGNNAANTNNNNNNCNCNNGNNGTANCGNACTANANANNNCGCNNAG
AGCTAANTTNANNAAGNCTGGGATTNNNNTAAGTNTNNANNNNGACTGGCNNCATANGNA
GGNNGTCANNGGANNGACNNNANGNNGCNNCNTTNTNAAGTNTAGNCNNNNGAANAATAT
NNNTNNNNAGGTNANGNGGGTGNGATTNTTNANTGGCNNGTGGCTGNNCANGTGAGGNNA
NGNNGANTAGNNGCGANTNTTGAGNANANATNNNANACNNCNCATANCNTCCNGGNNGGT
TTNTNNNANTTANNCNCNGCGNGCNNACCNNNGNAANANGNCGNGANANTGCNTNTANNA

ANNNCNGTNNTNNNNGCNNNTAANCGTANCNNTNAGANNAANCNGNAGGGTCANCNNGCT
NNGNCNGCGNNNNGGNNNNNGGCTNNNNATNNGCNGGCGTNTCNGCNNNNGNGTGNTGGN
NTNTGTNNANANGNANCCGGCTCGATATNACNNNGTNGGNCNTCGCGTATTAANGNNGNA
NNTATNTAGNANNTNNGTCNACCNAGANNNGCGNNNGNTNNNNGGTNNNNGCAGTNNNAN
AATNNTNNNTNNTAANTACNCCCGNTCGNGCNCNTGNNATATNTTGNNNNCTTNCNCNAC
GCGCTAGCCNC
>A219
NGNGNTTNNNNTTNNCGNCCCNCTNNTNNCNTGAANGAGTGGGGANATNGAAAGGGAGTN
NATGCGNNGGGAGNNNCCCTNTNANANNGNANCGTNTGATGTCNTTTGNTNAGAGGGTTG
NNTTGNGGTNNNATTTTACNNACNGNTCACCCGCNNGGGTCNNNCCTCNTTNTCNTCCGT
NNGANGGANCNGCGCNGCCNNNNANNCNTTNCACNGNCGTGTGGNTAGACGCANANNCCG
CNNNCATANACNNNAACTCCAACGNGCCNNCTATNNNACNTANCCTNCNNNNNGCNNNNA
NTGTTCNNNGGNTNGNACCNTTTNACNCCCNNNNNNNTNTTNTATACNNNNCTACCGTAC
NGNCGCCACTCCCCNNCNNNTNCTTNATNNNNANNCGGGNCCCTNTNTANTTNTNNNTTN
TAANNNNNCNCNCNTTGCCTNTANCTNNTNTGNNTATNNCGANATANANNTTAGACNATT
AGNNNNTAANNANGGNTTATGNNNANNNGCCCNNNACNGGGANNTNNGTATACNTCANAN
TNANAGANTATNTNACCNNCAGCCGNTNTTNTCNGNNGNTNTACNNANNANAAGGNNCNN
TGNNNGNGANNAGANAAAANANNTNGNNGNNGNAGAGAGTTNGNTAAGNAANNAATANAN
NNNCCNNNTTGTNTNTNTNAAANGCNNNCNGNNGNNGNAGGNNGGNNGANNAAAAAGGAN
NGTNCTTANAGNNNANNCNATGNGAANNAAATCAGNGNTNAGNGCNNNTNGNNNGGNNNA
GANGNNANNCNNNNGANNNAAGTNGTCCGNGANNNAANGNGCGNAAATTCNCNNTCGATC
GNNNNGCTTNNTACTATATNNNNGNTTNCNNTNTCTATCGANGNTCANTCCGCGNCNTCG
CNCGCCGATTAATAAGNTNAGGCTGTANACGATNTNNGNANTNCTTNNNCTGCNNNTNTN
NGCGTNNANTCNTANNNNCGNAACNCNCGTTAATNNGNTAAGTNGNAGNNTNNNNTAANT
CCCNCCGGNGGCCNNANNGNTTNANANNTNCGGCNNTNNTTAGGAAGNAANTTNNANTGT
AAGTNATNACNNAANNNTNNCCNNNTAGNNGGTCNGANNTNNATTACAAGTGNNGNNCAG
ANGCNNGATTNTTNNNTTCNGNNANGCNGGNCGTNTTTAANNGCANNTATTCNCGCACNN
NTTTTNNATNANNNNAGTTGAGGANGCNNCGTNNGNNTANNTGNNTNNGTCANNCGNNNG
CNNNTNNNNNNTNAANGCGGGGGATATGNGANNGCGNANGTTNNNNNNATNNTTTAGGGC
NATNGNNNNNNNGGNCNTANAANACCANGATNNNTNAANTAAAGGANNGNGNACNGCGCN
TANNAACCTCGCGTTNNTNAANNNGGTNGTGTANAANANGNAANAGCGCATNATGNGCNA
CGACGGAANGCNNNTGCGNNANNNTAGTTNNGNNNAAGNNTNTATNTNANGNTNNTAGAC
AGTNCGATANCCNCCNNTGGCNGCTNGCANNTAGGATGGNTNNNGNACGCNTCNTNTATN
GNNATGNTNTNCGGNGGNTAGAGCGNGGNGCGTGCTATNTNNNANNGTTANTANTTAAAA
NANGCCGNTGTTNGGGCCNTAGTTNTATNNTTANNNCNACNNCTCGCNNTACCNGNG
>A220
ATGGGCNCCGTATCATGCAAGTCGAGCGCACCTTCGGGTNGAGCGGCGGACGGGTGAGTA
ACGCGTGGGNATATGCCCTTTGGTACGGAACAACTGAGGGAAACTTCAGCTAATACCGTA
TGTGCCCTTCGGGGGAAAGATTTATCGCCATTGGAGTAGCCCGCGTTGGATTAGCTAGTT
GGTGGGGTAAAGGCCCACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCA
CACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTGCGCA
ATGGGCGAAAGCCTGACGCAGCCATGCCGCGTGAATGATGAAGGTCTTAGGATTGTAAAA
TTCTTTCACCGGGGAAGATAATGACGGTACCCGGAGAAGAAGCCCCGGCTAACTTCGTGC
CAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGGGC
GCGTAGGCGGACAGTTTAGTCAGAGGTGAAAGCCCAGGGCTCAACCTTGGAATTGCCTTT
GATACTGGCTGTCTTGAGTTCGGGAGAGGTGAGTGGAATGCCGAGTGTAGAGGTGAAATT
CGTAGATATTCGGCGGAACACCAGTGGCGAAGGCGACTCACTGGCCCGATACTGACGCTG
AGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACG
ATGAGTGCTAGTTGTCGGCATGCATGCATGTCGGTGACGCAGCTAACGCATTAAGCACTC
CGCCTGGGGAGTACGGTCGCAAGNATTAAAACTCAAAGGAATTGACGGGGGGCCCGCACA
AGCCNGTGGAGCATGTGGTTTTATTNCGAAGCAACCGCGCANAAACCTTTACCNACCTTT
TTGACATGCCCTTGATCGCCCGGAAAANTCCGGGTTTTTCCCTTCGGGNACAGGGACACA
NGNGCTGCNTGNTTGTCNTCNNNNNTCTTGTNNNTGAAANTGTTTGGGTTAAGTNCCNCA
ACNAAGCCCAACCCCNTNCCNTTTANTTTCCCTTCATTAAATTNGGNCCCTNTNNTGGNA

CCCCCGTGGGCAACCCNGAAGGAAANNGGGGATATAANTTNCANCCCNCTTTNCCCTTTN
GNGGGGGGGCTACCCCCNTTTCNAANGGGCANNNCTTAGGTTTTCCACCCCCCCCCNGNG
ACCNTTCCTANAAATTTTCTTCAACCNANCTNCTCNCTCC
>A221
GTGGGGGTNNAAANTNGAANGGNGNCNNNTGTGNGNAACCTGTGTNAATAGNGNTGGNGN
NNACNTGGNATGAGTANATGATATNGNAGNNCCGCCCCCCTNATANGAGNGCNNTACNGN
NATNATNNCTNTACGGGNGNCANATTAAGNNNTNTNNANNNNCTNTANTATCGGNNNNTA
NACNNNCNTTGCNAGGNNGATATCACCCNCTCNTGTNNNANCNNNNNGGANGNNGNTNCN
CNCGNANNCGCCTACNNANTNTNANNACGCNNNACGNTTTGTGNGNATGNGNGGGNCANC
TNCNCANCNANTGNNCCTAGTNANCGTATCNNNNACNNGTCTNTNTNTACGCNTTCGGTA
NTCNACNNGNNNGNNTATCNNTCNTNGNNNTNTATNNCNNATTNNTNCCCCCNNNTTATA
TCNTNNNTATTCNNNGCNNCNNCNNCGTNNCNNCNNNCNCCNNNNGNNANNCGNTAGNAC
CNNATCNNNNNTCNNNTNNCTATNTTNNCNCCNTANTANNTNTCGCGNCGTTCNCGAANN
TNTTNCCCTNGNNTCGCNCGCGNNNNNNNNATAAAANANTTNNAGNGNTNNNNCANNNAN
NTAAATANNNNCNCNNCNCNCTTGTNCGNGGACAGNCNAANNNTATTNGNGTNTGTTNAN
ANTNANGNNNNNGNTGCGNGCTCNGTANTCCNTNNATNTCGNNNANCNAANCGTTTNNCC
CANNCNNTNATTNNANANNGATAGNNNAACANNTNNNAATNNAANGNAAAGGTGNNGCNC
NCGCTNNNGTANNNTNATTGNCNANGNNTGANCANNGCTTNNNTNANATNTAGAAANNNG
GTGANGTNAAANNGCCNNGGNNATCCNNTGNNNCANNAGNNGTNCCTTANGNNATNAGCG
GGNGTGNNNNNAANNAAAGNNNNTNAAAAANATTCCNGGGAGGAATGGNNNGNNNNNATC
GGTANNACGNNCGCCCNNNCNNNTATNTGNNTAACNCNGNNNTAGGGAGNNTNNANTNTA
GNNGNAGNNGACNNNNCCCNCGCCNTTTATTATGNANTCNNGATNTNTCCNGACNAANAA
TNNTCNNATNNNNGCTTTTTCAAGACGTANTAGTTCANANCANTCCNTNNCCNGCNNAAA
NTANNTTANNAGANGANNANTNNGANTANATATNGAGCGNCANNNCACNCNCCNCACNGN
NNATANAATACNACNNNTNNNNTTNAGNTNGNNANNNGNNNNTNANATNTAGNGGNATTN
AAGATNNCGNNCNATATNCNNTGNNGNAGTNGNGAAATNAAGNACNGTNCGCGCANGNCN
ANATNNTTATANANTCGANTNTANGTAGGNNNGTTANANCAGAGAGCGTNACNNGNGCNT
GNGCNTTTNTTANACNNCGNANNANNGGTANNANCCTATNNGTTATGTNCAAACCNCCNA
ANGNANNATANNCTATNCNTANAGTGNNAATNCNCTNNCNNNCTNNCNGNCGNANGTAAT
NANNANATGCNNNNCGAGGTNTGGGNGNGGATNANNCGNNGNTAATNTNANTAACTATAT
ANNCNNNTGAANNCGCNANNTNNCNNCNNCNTANTNTAGNNGNNGNNNNNNNTANTCNGA
TAGAAANNGCANTAATTNAGGCGANCGNNGNTTAGTATNAGCGNGCGNNTATNACNNGNN
TGTANGTTNNTTANTNTTGATNAGNNTNAANNCCNACGGTCCNTCTCGCGNNTACCNGGG
CGTTAGTATNCCCCNTTNGACNGNNNCNCCGNNANTACNNNNTNTGTNTNCGNCNCGTNG
TTCNNGGNTGNNTAGANNNNNANGTNTTTTANNATNNANNNTACACGCNGGTNNNCCGNA
TACTNANNATGTAATCGNTNCNNNNTCGNCNNGNCCGC
>A222
GAGGGGAGCGTAAATGCAGTCGNAACAAGGTTAACCGNAAGGTTCNNNGNTGNATTAAAC
GGANAGGGGNNNCCCTTNAANANGGCGGNTNNNGTNAAGNAGCAGCTNAANNNACNNATN
GAANNNCNGCTNTTGCNNNCGNNANGAATCCCCNNNNNACNNNNCANGGNNGCNNNCNNG
NANCACNCNCNNTTTNNCGCTACCTATNGNATACCGNNNTCNGCNGGTTANATANANNNC
NNCNCNTCCCNCNCNACTNTNACNCGCTNNNGTACNNNTCGNAANTTNNATTCCACGNTT
NNATCNNATNNNCCCGCNTNTNTATCNACNNATCCNNNAANGNCNATNANCCCNNNNCCN
GNNCGCNCNNNNACNNTNGCGCNACAGCTCNTTNNTTACATNNNNCNNNCCGACNNNCGN
TNATTATGNNNNCANAGGNCGANANATNTNNNNGAGNNNAAANANGNGTNAGNNNNGNCG
CNNGCGCNNGTCNCGAGGNGNNANGAATANNNAATATCTGNGCGNGTNCNGNNGNNNNCC
NNTTTNGCTNGGGNNNNGNNGCGGNTCGNNANTNAAGNAGNNCGGGNNNGNANTTTTAAT
GTGNNNATCTCCCCGNNTTTATNGATTTAAAANTGGNGCNNNCCCNNNTNTNNTNNNATN
ACTNNTNTCNNAGNNGAANGNNANNNNNNGNANNAAAANNNGTTCTNNANGNANNCNNAN
ATCANTNNGCNNNCNNNTTNNNGACANNTNCGGGNNTNNTTCGNTTATCNNNCCCNNNNN
NTTTNTTATANGCNNNTTCANNNTACNNATCCTCNCNACTCGCNCCNNTTTNTCTTNTCA
ACNNGNNNCCNTNCNCNATTCANTTCGNNTNTNANCNNGNNNNCTNTCNNGNTTCCNGGT
TANNNTNNTAGNNNANNGANNNCNNACNTCNCCNCCCGCGCANNCNGTTANGNNGNNGNN
NANTTANTCNNTNGNCTNNTTNNNATACCATTNNTAGANGNNNGNGTANTNTANNNTGTA

GAGNTATANANNTCNNCCGCGCNNNTTNNNGACNCTTNNNNNNNGTTNTTNCCNCNAANN
ANGATCGANNNTTTTTTNCNGCTANNGCCNTGNAANNAGNTTTGAANAANCNANTNANTT
AANNAANNATATGTATNCNNNGGAAAGNGCCNCNCNANNTCNNTNANNANGNCAGNNGGN
GANACGCNGATANAGCGATTATACNATTTNNANTATACCNTNCCNNNNNNTGTNTNNTCN
CNTGTANCTTNNNGCNNNTNAANNAANNACGANNNNATNTNNGGCNNGNACGCNTNTGTN
NANNGCNNNNNGGNNGCGNNTNAGANTAAGGNATNGANTNANNCANCGCNNNNACNGNNN
GGNCNGNGTNANTTNGNGTAGGANNNTCGNNACNNTNGNGNNGNTTTNCGNACGNANACC
CCNNCCGGNNNNCTTGNNNTTNAGANNNNGATACCNNNCNTANNCCGNATNTNTNATNTT
TTCCTNNNCNCCTNCCCCNCCCGC
>A223
GNGNAGCGNNNGTANTTNNCNCACNNCCANNTTNTNGNANNGGCCCNNNNNTACNNTGGG
TAGGNGNGGTGGNCGAANNNNGGGGACGNANTTNNAANGANNGGGAGGGNNCCCCCCTTT
NTAANNNGNGTGNNGANNANGTAACGNCGCANGNAGGGATATNTTNTNGNCATNCNGNTN
TANGNNGANNNTNNTNGCTTCNCANNAGNGCNNCANNANTACNCTTACTCAAGCAGAAAN
NTAATNNGAGNCCCGACCNCCTTTTNAACCCNGTCCTTTNCTCTGNNTNNNNNGTCNNNG
CNTCCACNANTCNAGTTCTCCANCNCAATGGNAGNANNNTNCNCGNTCNGCACANTNNCT
CGTNGTANNGTTCNGCNCNTGNNNCNNNTTNTANGCCNCCNNTNTTNNNCAANATNNTCN
CNCNNNGCGTANNTCANGCTCCNNCGNNNNNNNNCTNNGANCATGNNNNNNNANNGTGNN
TANCATNTTNCNTATTACTNCNCCTNCNTTGNCNNTANGNTGGNNCGAACGGCGACGTAG
NGGAAANGGAATANGNNTGANAAGNNNNTAANNCNCNGCGCNNNNNNATNCCGNNNNNAC
TTGTATNANNAANTTCAGTNCGCGCCNANNTTNNTTGNCCNNTANGGGNCNNNNGTGNNA
NCCGANNNTGNATCNTTANANTCNCCNNTTNGNTNTTGAANGATANTAACNGCCCNGCTT
TNNTTTNANGTTGNGNANCTNCNNNGATANTTANTGNAAGTAAAANNGGTNNNNGNNNGG
CAACCCANTCNTCGANANANTNGGGTGTNTTTNANAANNTNGGNNANTGTTNCTTGNNAG
CATNNNAAAAAANACTGAANGANNNTGCTANNTTNGTCNATTTTCGTNNCACCCNNTCTA
AAANNNATACNNCACTNCCNANCNTNCGANANCNNGCCNNCCGNNTCNCCCCCNANNTTT
GNNANTNACNNTNNTCCNNNNCAGTNTNCNTATACNNNNTNNNGGCGNANNNACNNNGNC
NNTNTCGNCNGTNNANNATNNNGANNNANGANGAGNCTCTGTNANNCNCCGNCNCNCANN
NNANTATAAATGNGCNNGNANGTNGTACTGNTNCTANATAGNTNTCNTAGNTTNNNTGNG
NGNGGTTAACCNNANAANGTTAGNGTGCGTCCCCCCNNCNTNTAANNNNNCNCTGNGTNC
NNGTTTNCNNNNACTTNNCGCGCNNNCGNNATTANANGGTTTCCTNNGNACTNAATTGAT
NNTCANNACNGANGTTTCAAAAAAAATANGTNNNGCGATAAGCNCGCCGTNNGNTNTTTG
TATNNTGGNTNNNGNTTANNCGNNGAAANGNCNNNNTTTATNGTNNTGTTTTNCGCNNCC
CNCCCTCTNNTCGTCNCGTTTTAGANAGATNTCNGANTNTNNNNTTNANNCCNNTNNNTG
CTANGCGNGNNTTNCNTTCCNGCCGNAANATCNAGAGTNNNATAATGNTATGTNGTTNNN
ATATTNACGNGTCGCTANNNCGGCCCNNANGNCTNGNTNANTTGGNNAAGTNTNNTCNGN
GNCNNCTANNTGNNCGCGTAGTTNCNCNNCNNNTGGTCNAGAGGNGCGATTCGTGTTANN
GNTNNCANNTTATTNNATCCGGCNNTGTACTNCNCNCTTGTTTNTNTTGNCGNNNNGNCN
TNNCGCCNACCCNC
>A224
NCGCNGCTACCATGCAAGTCGAACGCCCCGCAAGGGGAGTGGCAGACGGGTGAAGTAACG
CGTGGGAATCTACCCTTTTCTACGGAACAACTGAGGGAAACTTCAGCTAATACCGTATAC
GGCCGAGAGGCGAAAGATTTATCGGAGAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGT
GGGGTAAAGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACAC
TGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATG
GGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTC
TTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAG
CAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCACG
TAGGCGGACTATTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGAT
ACTGGTAGTCTTGAGTTCGAGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGT
AGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGCTCGATACTGACGCTGAGG
TGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACTATG
AGAGCTAGCCGTCGGGCAGTATACTGTTCGGTGGCGCAGCTAACGCATTAAGCTCTCCGC
CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG

TGGAGCATGTGGTTTAATTCGAAGCAACGCGCANAACCTTACCAGCCCTTGACATCCCGA
TCGCGGTTAGTGGAGACACTATCCTTCANTTAGGCTGGATCGGTGACAGGTGCTGCATGG
CTGTCGTCAGCTCNTGTCGTGAGATGTTNGGNTAAATNCCCGCAACGAGNGCAANCCCTC
NCCCTTANTNGCCNTCATTCANTTGGGCCNCTCTTAAGGGGNACTGCCCGTNNATAACCC
CAAAAGNAAGGTGGGGATNANCTTNAATCCTCNNGGCCCTTACNGGNCTGGGTTNCCCCN
TNNTTNAATGGTNGTNACNNTGGNCANCAAACCCCCNGNTCAACTATTTCC
>A225
ATGCGCNGCTTACCATGCAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAAC
GCGTGGGAATCTACCTAGCTCTACGGAATAGCTCCGGGAAACTGGAATTAATACCGTATA
CGTCCTTCGGGAGAAAGATTTATCGGAGTTAGATGAGCCCGCGTTGGATTAGCTAGTTGG
TGGGGTAATGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACA
CTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAAT
GGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCT
CTTTCAACGGTGAAGATAATGACGGTAACCGTAGAAGAAGCCCCGGCTAACTTCGTGCCA
GCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCAC
GTAGGCGGATACTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGA
TACTGGGTATCTCGAGTCCGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCG
TAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCGGTACTGACGCTGAG
GTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGAT
GGAAGCCAGCTGTCGGCAAGTTTACTTGTCGGTGGCGCAGCTAACGCATTAAGCTTCCCG
CCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCG
GGTGGAGCATGTGGTTTAATTCGAAGCACCGCGCANAANCCTTACCAGCCCTTTGACATC
CCGGGTCGCGGTTACGAGAGATCGTATCCTTTCAGTTCGGCTGGNACCGGTGACAGGTGC
TGCATGGCTGTCGTCAGCTCGTGTCGTGAAATGTTGGGTTAAGTCCCGCAANCANGCGCA
ACCNTCNCCNTNANTTGCCATCATTTAGGTTGGGGCNTNTAANGGGACTGCCNGTTANAA
ACCCGNAAGAAAGGGGGGNNTNACTTCAAANTCCTCATGGCCCTTTNCGGGCTGGGCTAC
NCCNTTNTAAAATGGNNGGTNAAATTGGCCANCAAACCCCCANGNTCAACNTAATTTCCA
AAACCC
>A226
GGNNGGGAGGGNNAAACACGNCNAGTTNGNTAACAAGGTGTAACCGTTAGGCNTNNCNGA
NTGTGTGNANANTNACGCGNGANNGCNNTGTNANTNCNNANTNNNCGTTTNACGATNNCT
TTGCGAGGTTNGNTNGGNNATCGGAATNTCCCAAGNTTTNNCCNCNGCGNNGNNNCNACT
NNNANTNNNGCNATGNGTCNNGGGNNNGCNCCGNTNTTACCCNNCNTGTATNNNTTNNNA
NTNGGNCCCTACNNTNTCGACNNAGCTCNGNCNAGNACNTTNACGNACGGNCNGNNCANT
NNTNTNNNNCGCTTTNTACTGANNCCCCTTTTNTTTTATNTNCTACGNGCTACACNCNNC
CCNCNCCNNCNCNATNACNNNATNNNCGCACGNNAATTNCNNNNTNGACACCNCTTCCNN
CANNTGGCGCAAAGNNNNNNNATANANGTTAANANACGNCTNAGNAGNNAAACCNNNGCC
GCNNAGGNNGNTCTCATNGNNTAAAANNNNGNTNNTTACGGNGNCNCCGGNNCNNAANNN
NNTTATCNNNNANGAGACAGANNNGNNNCNNNTANGNTAAAAAAANGGNTNNANANANTN
TTANCGNGNNNANTNCCCGGNGNATTATTNTNATNATGGNNNCNAAGNGNNNCGTAGTNG
TTGNNNNNANNGTNCTNNNNNCNNTNCCGNTTTNNTGGNNGNGANGNNNNGGGNNNGTAN
AGGNGTTAGTNNNTNATNNANNGNNGANATNTNNTNANATNCGCNNNGGNNTTCCCANNN
CNNNCNNGNNCNNTNTANTATATNCANTANANNTNTNATCNNTNNNNCNCNCNNNCGGCG
ANANNNNGNGNNNAGNNGAACTCCNNNCNNANAANNNNATNNNNTCATACNNCTTNTNNA
NANNCNCNGNNNNGCATTNNNTTTNNTANATGTTNCTTANCGTGNTTCCCGNCNTNNCNN
NNGCCNANTNAAGGCGCTNAGTGNTACGACNGGGGNGNGNTNNANNNCNANNANTANAGA
NNCNNNTATNAGNCNNNNNGNCNAAAGAGTNCNNCGNCCNNNCNNANANNATTNCGCTAN
NNNNATNNNTATGATNNNCNANNNCGNNNGAGGGNTTATGNNNGCNNGNNTNNNNNNNGA
CNGGNNGNNNCGNCNNATGCACCTAGTNAGNTANAGNAANAGAAGACGNANNNNANCGCG
CCNCGGNGNANNCNNNTATTTATGCNCNNGCNNCNNTTNGNGCNGNCTNGCTGNNTNNTT
NNNTNNANNANTTCCCNGNGTNGACNCCTATANTCCTACGNTNNNATANAGGCCGCGNNC
TTNNNNTAATANNNAAGNNNANTTTCNNNNANCGCGNAANNTNNNNCNNNNCANNNNNNC
AGNGANATAAATTATATNGNAANATNAAACGTNCCTCNNCCCCCNNTTCCCGCNTNTTTA
CCTATNTNNNTCNNCAATNNNNGTCTNANTCNAGGGCNGGGNCCGNTNNCNCNGNNNTNA

TNANNGATNNNTNNTGCGCNNTTNGNCCNNTNNTNTTTNNTNTCCTNGANCNNNCACTNC
NTC
>A227
NGGAGGGGNGGTNTTTNCCNCNNCCNNTNNNNTNGTCNCTTTTGNATNTNGCGGGNNTGA
TTGGTNGNNNGTNGTTATNNAAGGNGGGGNNCCAATTTANNNGANNNGGGGNTTAAGNGA
NNNGATTGNGGGGNAGANGANGCNNNNNCNANTTNTTNGNNNNGATATNNNCCANNNANN
ACNTCNCNNCNCNNCACANCGANNAGNNGTNNAGTAAGNNNANCCACNNCNCGGGTNATC
GCNCNTTCNATTNTNNATNNCNNNNNTGCANNCCNTACTNNCNGCNNCNNACNGCATCGT
TNGGNATAGTNATACCGNTCNTANCANANNCNNATTANTATTCGNCNCTTTTNTNNTATN
AACNCCCNTANTNTCNANATNTCGNTGTNGGNACNGCNGCCTCNNCCCCNGNCTNGTNTG
NAATGGNNACTCNNNGAGNTTANANNTTNTNCTANNCCNCGATCNNTNNNNTAANNNNGN
AACGCCNTCGNNNTAANNGANANANTGACTNNGNCNATNANCNANNNGGNNNTANCGNNG
ANNANANTTNNANTAAANGNNNCNTNGGCCNCNNCNNCNACNTGNCNTGNTTNCGNANNN
NGNAANNCCGTNCCNTAATNTAAGNTTTGCNNCAANGANTNTTAANNNAGTNNATCCCNG
CCGATANATTTTTTGNAGCNGNGNNNNANCNGTTANGNTNNCNNNAANNNANANANNNTG
CGCCGTNCGNNANNNNANNGNAGTTTNNGNANNCNNAGGTNNNGGGTCTTNGNNCANNGA
ANANNCGNANGNNNCNGNNNNTNTCTGCTTNCNANCCNCCGNGAATTTTNTTGACNATNN
NNCTTCGNNNNNCCCCACNCGTCGGNCNNNTNTGTTTCNNCNGGNNTNNNCNACNNATCN
NNGNTTATGCNTNTAAATCTATGANNCTNACCNNNGTNCNCTTNNNTNNNTNAGGNNGAG
NTNNGTNNGNTGCGCGGGCNACTGTNGNNNAANATACNTNCNGGNGNGANGANTNANNNN
NCGTNAATNNANTNGNTNGNNTGTTTGANNNNGANNGGNNATNNTNNNNATTANTANCCN
ATNCCNNNNGTTNATAANAACGNNNAGAAGCCNTNNTAANCNNCNNTACNTCNCNACCNT
TTTNNANCGNANTNCCTTCCTACNNANGGTNCGNNGACTGNTNGACATNGNATNGNTNNN
AANNNGGNANNGTTAANNGCGNGNNAGGNNNNNTATTTTNACANCNGNNACNNNNAGNGG
NGAAAANNNACNNNATANNTTNNGNAAATTGNNNNNNGGGCANNNNANCCTCCCCNTTTA
TNTGGNCGGNGNCTTTAGNNATTTNANCGNNTTTNGNANGACGCNATANNCTNNNNNGCG
NGATGNNCTNNTGCGATAAAGAAANGNNNGNGANATACACGANTCNTGNCACNGNGGGAN
NCNGCNNNTTNTTNGATNTNNTNNANNANNTTTNGGNNNGTNTTCNGNTNGNGTTNNCGT
NGGCNGGAGTNAATANGTTNTTNTTTNNNTNCCNNCTGTNNNGNNTTAANTTATTNATCT
GNTCNGNNCNNTACCNAN
>A228
NAGGNAGGNNNNGNGNANNCCCNNNCCNGGTNGNGTTTNCCCNGNGTATGANGNANANGC
GATNTCAGNANGTGGGNNGGANATNNNAANANGNANGGANCCNNTTNAAGANGNNGACNG
NGGGNNAGNNGGAANTCTANTGGNGGGNNTNGNGTTGNNNATANANTNTTATGNGANNTN
NNTCGTANNNNNTCNGACTCNNNCGTCNTCNNNTCNGCNTGNNGGAAAANNTNNGNNNCG
NNNCACNNNTCNNTNGCNNGNGNNTNANTNNTANNCNGCNCGTNCCGCCCNNACCANTCN
ANNNNNNCNNCTGNGNNNNCGTCNNNTNNNTACNNTNCGNANNNANACTNTNGATATANT
NNCCCNNNTTNTNANCNANNTNNCNCCTNTNNATTTCANNANAAGNTACGNNNGNACTAN
NCCNACGCNNGCNNACNNCTTATATANNTNCCNANANTCNCNTNNNTATNTNAGATTTTA
CGTACNCCCCGTTNCCTTCNTACNANANNCNNANCNNCTCGACGGTGTTGGANNNTGTGN
CTNCNTGTCGTAGNNTNTNNNNGCCNCGCNNNCNTTTACGNGNANCTTNANAATAGNNTN
AANAAGNATNACAANNGCGNNCNCNNANTNNNANGNNANNTNNTNNGGTNANNNACNAGA
ATCCNCNTTCGGNTNNAATNTNANATNTCGGNNANNTANNAATNTNAATAGAGTNNNNCG
NNCGNNNCGNTTNANTNTTTGTCNNTANANGCTNNNNGACNNNNATTNNNTNNNNNANAA
ANTNANGTANATGNGTCAACNGGNTTTCNNNNGNNNNCANAGGNTTCNTNTGNAGNTANA
GTGGANCNTGGTATNTATTAGCGNANNNNAAGNTTGNNAANANNNCCNNTGATTTTTCAG
GNNANCANGNGNCNNCNNTTTNTTNTTNCGANGNTGNNNNNTNTCNNANNNNNGNNNTCN
NTNNANCCAANNNTNNNNNNNANNNANGNNNNCCNGCNNAANNNCTTNNTNNGGTTCANA
ANNCGNAGNNCNNNNATCCNNNNNNNNGCTNNANTTNTCNNNTNTNGNCCNANNNCNNNA
TNNTTNCCCNCGCCGNNANNGNCNGNTATNTGNNNAGCCCNCANGTAGNNCNCNTNCCNN
TGNNCNNNTTANNATANGANNNTANTANNCCNTGNNNNNGGNTNNNNCGNNTNGNGCANG
ANANNNNNCTNNCNGCGNCCTTATTNTNGTTNCNCNNCTGTANGCTATNATATANNNCNN
AAACNCGNCGNNNANTTATTTAGCCTNNTNNNTTANTNGTNANGGNTAGTNAATACNNAG
NAGGNTAATTNNNNNTNAANGNTNNNGAAGTNANNACNCCGANATNCNTTANNTACTGCG

NTANANNNTNNNNGNNCAGANTCNNNNNCNNNANTTNNNNCATTTTCCGCNNNNNGNNCT
NTTNTCNTCGCGTTTNATGTNGTNGGNTGGNCTNTAGANNNNTTNNNNCGCGNTNNCGAN
NGCGACNNTNNAANNNGCNCCCATTNTNNNGGANANNANANNTGNNATANTNAANANNAG
ANCGAANNTCNNCNAAGGCNNNCNCANGNTAANGNTCGAGTTANGGNNCCNNCTATTTNN
GCNGNATACCNCGNCNNNCNCNNCNATCCCGGNNCNGNANGGNNNTANTTTNANTGNGCN
NGTTNNNGCGTATNTTATANGNTTNCTANCANNNGTNNNTACNGCG
>A229
NNNGGATNNTNTACNGCCNTNGTACGCCANNANTTNNCGNNNGTTACNAGNTGTNNGANN
ANCGNNNNGCCCTANAAGANGNNGGGTTGNGGGATGTTNGNGNANGANNNGTTGTNTNTT
NCNGNGNNCTCGNCNGNGNNATGNCNNNNCTGGGGGGTNAGTANNCCNNNGCNCGCGNTA
TCNCNNCNTCNGCTTGGCGCCTCNCNCNGCCNNNNCNCNNCGCACTNTTCAGNCTNNNCG
GACGNGNNNNNGCNTGGNAATTTNCCCCGTNATCNNTTNANNCCTATNANTCNTGTNTNG
NNGNGGTAGACGCGCTCCCNGCNNNAAGCNTTNNNNANCCGCNCNCTCNNNNNNGCATNT
ACTANCCNCTNNCTCCNNTACNNGNANCNCCTCTNGNGGGANTATAATTNGGNNNNANNN
ANNNGCGGCCCNGGNTCCCCNCTAANANTNCNNTAAANANNNNNANNGNCNGGANGNNNN
NTNNCCAANTACNTGCNGNANANANGCNNTCGNNGATNNNNNTTCNGCNNNNNNATAAAT
GNTNTNNNGNCGNGCGGTNANNNATTCNNGNAGCCGNAACGGTTNANANANTNNATAGAN
NANGNGATATNNNNGGTANCTANNCNAGGNCGGTNTNNGAATAANGGNACTCNTCGANAG
NAGTNAAATCTGTAANTGGNGCNGANNNCTTTCAAGCNCNGATTTTTTNGNGATNNNNGG
NNCGNNNGNNCGNGNNCNCCGCGCGTTNANNGNCNNCNAATCNTNNAANNTCGTTTTGNT
NTAACNCNTNNATCATTCCCNNANCCCNNTNTATNTNNGNNNGNNTANTGNNTNNCCNNN
NCNNCNNCNCNCNATNTANNNGCGCNGNNNNNNTCTTNGNGCGCNGGTTTGTNNCNNNTG
TNNTNNANNGGANANGANNGNATNNGNNGTANTCNGTCGNNGNGCGTANNTTNNTCGGNC
NNGGCAAATNNCCCNNCNNNNCACCNNTNGCGTTTTNNTNNTCGNAACGCTANNGNNANA
TTAAGGNTATCCATCGTTNTTNNNNAAAANANTATANGANCGATNNCGCTCCCTACNTNT
TNNNAACANGCANNNAGCNNGGGGGNTNNNCNNTCANNNTNGTANNGAAACGCNAGGNGC
CATTNAGNCNCTTTTTNGNANTNTNTGGAGNANATTANNAANCNNTNGNCAACNGCTNGT
GNGANNNACGNGTGNTATNNNNCNNGNAATNNTTATTNNNCNTNTTNNNCNCTNNTCNNC
CGNNCNNCNCGGNTTNTGAGNNTGTNANANNCNNNTGATNTATCGATTTTATATNNGNCN
NCNAGCNCGANNNNNCNCGNGTACTGNGTNANNNGTATTTNTANCNNCGCCGTTAGNNCC
CNTANGTAANATATATTNCNNNNNNGTCGTCNCNTAGCNNN
>A230
TNNGNATCGCNGCCTANCATGCAAGTCGAGGGGCAGCAGGGGTNNGCAATACCCGCTGGC
GACCGGNAAACNGGTGCNGAACACNTACGCAACTTTCCTCTNNGCGGGAGATAGCCCGCG
TAGAAATCCGGATTAATANCCCATANGATTGTAGATCGCGCATCGATCAGCAATTAAAGA
TTTATCACTTANAGATGGGCGTGCGCTCTGATTAGGTAGTTGGTGAGGTAACGGCTCACC
ANNCCATCGATCAGTANCTGGCTGTGANAGCGCGACCACTCACNCGGGCACTGANACACG
GGCCCAACTCCNACNGGAGGCAGCCCTNAATGAATATTGGNCAATGTGACGCAATGTCTG
AACCTCGCCATGCCGCGTGNAGGATGAAGGTCNTCATGGATTNCTAAACNTCGNTTTAAC
NGGGAAGAAATGNTTNCTTNCTAGCGATCGCTGACGGTACCAGATGAATNAGCNCCGGNT
AANTTTGTGCCANCTACCGCGNGTAATACGTGAGGGTGCAAGCGTTNTCCGGATTCACTG
GGTTTNAAAGGGTGCGTAGGCNGAACNNAANGTCCGNTGTGCAATNTCCNAGCTTAACTN
GNAAATTGCNGTGGATACTATTGTCCCTTNAATATTGCNGNCGTTNNGNNCANTATGTNA
TCTCCTCNTNCCCTGCNNTTNTTTNTTNCNTTNTCNTTTCNCCNCCNCCNCCTTTNNNNN
NANNNNTCTNNNATCTNTTNNNNTCNTTTTTAGNNTNNGGGNCNCTCNNNCNNTNCTCCT
CTCCCCCCTTTTCNTCNCTNCCNCNTTTTTNNNCNTCCCCCNCNTTTTTNNTCCTCCNTC
CNCTNTNTNTTCNNATCGTTTTTTTTTAATNNTTCTNCCTNACNCNCCCTCTTTCTTTNT
TTTTCNNTNCNCCTTACNCCTTTCCCTCNNTNTTTTTTTTNTTTTNCCCTCCNCCCTNTT
NCTCTCTTTTTTTTCTNTNTCTTNCNTNTTCNCNCCCNTNTCCTCNTCTCCTTTNNTCCG
NCNTTTCTCTCTCNCNCNGNCTTTATTTNTNCTTTTNTCCNTCCTCCACCCCCCNCCNTC
CTTTTTTTCTTCTTNTNTNTCNCNTTTNTCNCNTTTNNCCCCCCTNTCCTCCTACNNCTC
TCNTNTNTNTTTNTATTTTTTCCGCNNTTTCCCCTTTNGNTNTATTNTACNTCNCCTCTC
CCCTTNCCCC
>A231

GTGANGACNTANNNCACNAANNNCGCGCCNNTTCNNCNANTTTGCNNTTTATNNNNNNGN
ANCGNGTGNTNCNCNTNNGTANANNGNTTNTANNGCGGNGTGGNNTNGANANCCCCCTTT
NTNGGGGGGGTNGGNNGATAACATNTGNAGGGAAAAGAANNGNNNNNNTNNAANATATCC
NGNANTNATAANCCNTCANNCCNNNACTCGGNCNTCNAANCNNANCNGTNGNNCNCNACN
AGGCGGGNCCGNGAGCNNNTTTATNCNCNNTNCNTATATNNNGTAGNNGAGGNNACCGCN
ANCNNNNNTNGNNNTNCNCNNNNNNNCNGNNCACGTNTACCNTNNGTTNNNNCANGNNNG
TATANNTNCNNNGNTNNNNNCNCTNTNTTNCCACNTNAAANTNNNGNTTAANGNGNNCNN
NGGCTANGTANNCNCNCNCNNCGCCNNNNNCNTTGTANCGNTTCCGCNANNNGAGTNNNT
ANTNTNCANGTAGNANNGNCTCNNNCNNNTNGNCANNNANTTANATNANCANNTNNGCCC
NGNGNAANNCATAAANGATANNNGAAGANAGGGAGGGANNTAAANNCNCCCNCNNGCNCN
GCTAGAANNANNNACATAAGNAANNATNNGNANCANANNAGATGNNNNNAGCTNNACNAN
GNNANTGTNCNNNNNGCNATATNNNAGNNNTGANGGNANNNANGCGNNANNNGAGTANNT
GTAANNNTATGNNNCNCNNANNNNGNNATACGATGCNTGANCNGNTGNTCNTANNCNCGG
CNNACTTNAATNGNTTANNNNNATTGGNGANCCNNANTAGCGCCNCNNNANGNGTTGGNG
TNNNNTNGGAGGNNAGNTNANCGNNAGNNANNGTTGGTCGTNNNNANTNATGGAAANNAT
GTCNNNTCNTATNAGTNANNAANNTGAGNNCATTNGANTNTTGAGNGNNGTNANNAANTC
CTTACNTGAAGNNANNGGATNTNNANTCANTATTATCANTNCGGNCGNCCNNNNNNNTTT
NTNGNCNNNNTTANNNNTCNATNAGTATNNTNNANGCAGACCTCNNNGCNCNCCNTTNNA
TTNNTATACGANNNTNGNANTNNTNTCGCTNACNNTCTTATNACNNCNNNCAANANNCGT
GNGTNATNNTGNNACGCGCNNNTCNGCNCNGTTATNNTATGATNAGNAGNCGAGNGCGAG
GCNNNATTACNANCGCNANNCGTCTGNGGCNGCCCGNTANANTTNNGTAGGCGNNNAGGN
TAGAGANGNGCGTGNAAATNGNTNANNTATGNNTNNNGTNNACNANANNNGNNGGTGNAT
NNCNCGANTANGGNACNNANAGNNGNCGCGTNCTGCCACCGCTCGTNTTNATNCATGNNT
CNTCGNNNGGNNATNNNTANNGCCNNNCNGNGGAGGNTANTGNNGTTANTNTNGGNANGN
NGNTNNNNATNCNANNNNTNANTATANCAGGGNNNGGAGNCNGGTGNNNTANTTNAANTA
GNANNGCAGNGNGAANNGTGNNGNCTAGNNTTCAATTTNNNNNNNTCNNGGNANTTTTNA
NNANTANNAAGNNNAGNNNNNCNNTATNNAAAATACCGCNGGNTCACNGGTCTNNACCNC
CCCTTCTANNNAAGNCGGNAANTTNNACTTNATTAANGGNGCGTNNTCTANGANNANCAC
NATGNTTCANACGCGTNNNTNNGCGCNNCNTNNGGANNTNNAGACNTATNGTAAGCGCGG
TGNNNGACTCACCNGAGCNTGCNGNNCNGTNGTNGTCGNGAANTTANANTGANGCNNNNN
TTATCGNNTGANGTATAGCGGNCCTNTAGGGCNNTGCNNNNCNTCTTNANGANTNACTAN
TNANGAANCNGCGCNNTACCTCNNCTATNTTATNTATATANNNGNCAACGCTCGCNCNAA
NCNCG
>A232
ANNAGNNGNNGGNTTNNCCCNGNNCCANTANANNATGCCCNTNTAAAGGTGNGAGTGGNC
AGACNNGCGAGNNNNNNAACGCGGCGAGGNGCAACCCCNNNNNNTCNGCGGNAGNCNANC
ANNTTNCAGGGGNAAATNNGCNAATNNTTAANACCGNACGATACGGCGGCCGNANGANGT
NNANNNNTNNANTNNANNNNGNGCANTAGAGTAAATGANANCNTNCCGNTNTNCACCCNN
CCTTANCTNTGTCTGCGGCGTCANANNGGNCCACTNCNCGACNGNCCGNNANGNNATCNT
NAACTANNTTCTCGCNTTCACGCANGCNTNGCNNNNANNTCANTTCGNCNNGCNTGTTNA
NCNTTGTTNCCGCCTTTNTTNCGNNNNTATNNCANACNNTNGNANNACNCNNCACGNNNN
GNCGNNNNNNTTANANCTTNNNGNGAGTNNANNNNNCNTTNGNANTNTTANNNTNNCCCG
NNNCCGCANGNNNTTNNNNGATNNNCGCCNCCNNGNNAGNATAANTANANGNTTNTATNN
TNAATNNNAGCNNANANTCCCCNTGTTGCNGNNGNANANANNNTGTNNNANNNANGNTNC
NGAAGCCCNNACTNNTCNTGTANTNCANGNTTNNGTCNNNNGGNCGGNNNACGGTNANNG
TNNNNNTTATNNTNGGCCCNTTTTNNTTTNTTNTTCTTGNTNTANNCNNNCACGNNATTA
NNTTTAGAGANNNCGANCTNATANNCNNTTTCNGNGTGTNNTTANATNCANNTANAATGN
TNNANNACCGNNNTCCTATGNNAANGNNNGNCAANNATGNANNNNGCGGNTNANTGTTTC
NNTAANGNGNCGNANAATTCNNNGATAANNTGNNNNGNNTTATCNNNNGTNNNACCGNNN
NTGATANGTNTTCNCAANTNNCCNAGTANCGANNTNNANNNCNCATNCACGCGACATTAT
NNANNNTCNNACANAATCNNATNTCGNTTCTATTTACNNNTATTNAGAGNNNNNATANCN
NNNCGCNCGTTGNCNGNTTNTNATATGGTTANGCNATGNNNACTCATNANTCNCCNGCNC
NTNANTNGGNNCNNAATAACNNNGCGNGNNCNNTNANANNNGNCGGCATNGTTNGNAANT
NTNANAGATNNATNGAANNNCNGGTANNANANCTNTNANCATTAGNATNNNNTCNGNCCG

CCCCTNTTNNNNTTCNCNGGCNNTNCCCATTNTAANCCCNTNNNNCNNCNCGNNNNTTTA
NTTNNGGGANTNNNGNACGAANTGNNNNTNNGAAAGANGANGNTACNTAANCAATGAAAA
NANTANTTCGGAGGAGNNNNGTANGNGCTNNGATNNNNTTTTNTNNANCTCTGNCNNTNT
GNNGCAGNTTAAATGGNNNTGNTACANNTNCNNNNTAAGCNNNGNGACCCCTNTCNNNNT
CNCCTTTNANTGTANGNCGNCNCNTNCGNNNNTNTANTNCNGCGTNTNNNNCNNNGACGN
AAGTTTNGCNNCTGTTCNNNNNNNCNATATNNATTCNATANNNCNNTNNNCANNTNGCNN
NCTCCTNNCANNNNTNNTNTATGTTNNGNNNAANACAGNNCTTTANCGNNNTNNNCNCNA
NANGCACNGGGNGCGGNGNNNNNCTGTATTNNTNATNANNNTCNCGCGCNTTNNCCNTTN
TNANTGNGNTNNNTGGGNCANCGNNNNCNCG
>A233
TTNTGGGNNAGGGCNNNCCGNNCNTNTNAGATCGCCCAGNTNTAAGGNGTNGTGGAANNN
AGGGGGGNGTAGAAGCGCGGGGGTGGNNANNNTTTTCTNNTATGNGTANGNCGNACANNT
TGCAGCGNTAAATNGTNTAAACNNATTACCNCNCGNANNCGNCCNTAGGANGNNCNCNAT
GATTTANCNCGCNANNNNAGTAANTGNGNCNNACCNTTNNANACCCNTACNTNNNTNGNC
GGGGGTANCCNNAACCNTCNCGTAGNCGNACNCATNCNGTTAGTTGTCNATNGCGCNCCT
TCTANCTNTTAAANTCGCCNNNTGTNNCCNATGTNCCGNCNNTNNTCTTCACAATTNACN
CTCNCNCNCTAAGNNCANNCCTNCNNGCGNGTNCNNCTTNGNACNAANTTGCNNCNGANA
GAGNAATNANATTNTNACGTNNCCCGNGGCCTTNNNATANGNCTNAAACNGCCNCNAAGT
TGGATTCANANNNNCNCNGTCTTTATTTNANNCNCTGNCNCNNNTTNCCCTTCNGNACAG
TGTNNTANNAANNACAATGGNNNCNNACGTTNNNNTGNNAAGTAANGCTCCTGGTGNNAN
TNCCGAANGGAGNGGGTNNNNANTNGCNCCNCNTGAATATTTTANTGCGNGAANTCNCNG
NAATTANATAATNNNNNNATCNANCAATCANGNTNNNGGNGNAGTNNAAGNNNGCTGANN
GNACNANANGNNNTCCNNGAANANNNAANNCCTNNTTAGTNNNCNGGNNCNGGTTNGNNG
AGNCTTNCNANANANCGGGGNNCANCNGNGNGAGGTATCNNGCNGTCTAANNNGNNNNNN
ANNTTTTTAGGNATATANNGNCNTATCGANNTNNAACCCNCTCACGCCNAAGTNTNNNNN
NNNCNTCTNTNNNCNANCNATANTNCNNANNGNNNNTTNNGTCNTACANATCNTNTCGCC
NNNNGGNAGAANATATGNAGNNNNGTTANAANTCNTTAGTTNNCCCCGCNNCNCNNCCCN
NTTANAANNAANNGNAGNGNNNAAGTTNNTCGANNNNTNGGTANCANANCNTAATNAGGG
ATGNNGNANCATNNGATNAGGNNCNTNNGANTAGCNTCNCCNCGGNACNTATTNGNCTNA
CNCTNGTNNNCGNNCNTTTAATCCNNCCATTCCGNNCGNGNNNNTTTNCATNCNGCNACT
CNTNCNNTCGNAGNNGNNNTATGCGNNGANNNNNCCGGTNGNANANTNATNAAGAANNNN
AACCGNNNGNNGACTGNCGCCCNGACGTTTCTNNNNTTCNGAGCGANNNNNCGTTGNCCG
CNCTAGNAANCNNGANTTAANCNNNTTNTAAATCTTCNGCNNGNCNNNCCTTNATAACCN
CNNTTNTATNNTCGNNGCGNNCNTACNNTNNGTNGNANNCCCGTNTTNGNACTGNGTCTC
NNNNTTTCNCNNNNCATATNTNTCAGNAGCATGAGTAGTTTNNTATNATCNNANTGAAAC
GCGCGNCNCNCTNCCCNGCGNNNNCNCGNGNTTNTANNNNTNCTNTNATCTNCNACCNCG
TTNGTCGGTGTATTAGCCNNGTGNNTNGTCCCGNNNNACCTGCCNCGTTNGAGCNTGTTT
CNTNATTTGNTGTCGCCGCNCGTTNTGGCCNNANNNTATTATTNNTATNCNNGNGAGNNT
NNGCCNTTANCCAC
>A234
GCGCNCTTTTNCATGCAGTCGAACGGCAGCGCGGGGCAACCTGGCGGCGAGTGGCGAACG
GGTGAGTAATATATCGGAACGTACCCAGAAGTGGGGGATAACGATANCGAAAGTTACGCT
AATACCGCATACGATCTACNGATGAAAGTGGGGGACCTTCGGGCCTCATGCTTGTGGAGC
GGCCGATATCTGATTAGCTAGTTGGTAGGGTAAAGGCCTACCAAGGCGACGATCAGTAGC
TGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAG
GCAGCAGTGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCAATGCCGCGTGAGTG
AAGAAGGCCTTCGGGTTGTAAAGCTCTTTTGTCAGGGAAGAAACGGNAATTTCTAATACC
TNTTNCTAATGACGGTACCTGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGCGG
TAATACGTANGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCGGTT
TTGTAAGTCTGACGTGAAAGCCCCGGGCTCAACCTGGGAATTGCGTTGGAGACTGCAAGG
CTTGAATCTGGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTG
GAGGAACACCGATGGCGAAGGCAGCCCCCTGGGTCAAGATTGACGCTCATGCACGAAAGC
GTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCTACTAGT
TGTCGGGTTTTAATTAACTTGGTAACGCAGCTAACGCGTGAAGTAGACCGCCCTGGGGGA

GTACGGTCGCAAGATTTAAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGNAT
GATGTGGATTAATTTCGATGCAACNCGAAAAACCTTTACCTACCCTTGACATGGAAGGAA
TCCTTGANAGATTTGAGGAATGCCCCNAANGGGAANCTTCNCACANGGTGCTGCATGGCT
TTNCNTCANCTCCTGTTCNTGNAAATTTTTGGGTTAANTNCCCCNACNNANNNNCANCCC
NTTGNCATTANTTNCCNCNTTNCNNTNGNNCCCTTNAATNNAAACTNCCCGGTNACAAAC
CCGGANNAAAGGNGGGGNATAACNTNNANTC
>A235
GAGGGGANGCNAANACACGGCAAGTTCGTTAACAAGGNTAACCGGTAAGTGTCNCCTGTG
NGNGCNAGTAGCNNNGCNGGGCCCCNTTNANGNGGNGNGTNAGNGAGNCNCTGNAGGGNG
AANTGNNAANCTTNANNNNCNNCGTNNATCGCCCNAGTAGGGTNNGNNATANCTTNNCGN
GCACNNANTANTGNGTNNGACCNTNTNANNNCCCNCNTNNCTNNNGCGGANTTNNNCGCT
CCCTANTCTATGCGANCCATCCNGTATNNTGNNANCGACCNANNNGGNAATATCNCNGAN
TNNCTTGNANCCNCTTNNTTCCANNNTNCNTNNNNGCTNNTGCCCNACCNCNNNNANCNT
NTNCATTNCGNGNCNCNNNNNNANNNNACGTCCCNCTNCNTANNTTGAGNNANGCGCNNG
ACATNNANAAACNNNNNTANANNNNNCNGNCCCNANNNNNCANANNGNGATAANNGNNNT
NGGNCNANGNNGTTTCGCNTTNNNNCNCTAGNNANCCCTNGNGTGNNNTNNNTCNNCAAT
TGNANTGTNGCNGGANNCCCCCGATAATTTNNNTNTCGACCAATNCNGNTGNGTTNNATA
NNNCTCTGGNTNGACCTAANCTTNNCTNNNCCTTCNTTGTNACACGNNNNNTATCGNGAC
CACCACNAAGGGCNNNATNGCANTNTTCNNTTCNNCNCCGCTTGNNTTTAGCANTNGNTT
TCNNNTCNNCNNCAGCNCCACNNNGNNNGNAAGNGTATCTTCGCGNTCTTTNNCCNTNNG
TCTNNCTNATCNCGNGGCTTNTTANNNNGANNNAGTNTANANCCNNGCNGCGGAANNNNC
CNNNNGACNCACGNGNNTNNNANCNNNATGCTTNCCTNANTNANNNNNNTCTTACNCNCN
CCGGTTNNTACNTNGGACNANTTNCNCNNNNNNCNCNNCTTNAGNNGGNTNCNTGGNGTG
NNTAGANANNNCCNGTTNNTNNNNANNTTTCNCNNAACGCNNGTNCNTCTTTTNGNCGCT
NNCNNTTNNTNNNTNGTTNGTTNTNCNTNNAGANNTTCNCTNACCCNATTNCNTANCGTN
TATTNNNGNNGNNNNANNNAGTNNNTNGNNTTATCNTCNGCGCNNNNCANTCTCTNTNNT
NNGNNNGNTNTNTTTNNNTACTTANTNNCNNTCNTNCACNNCCNGCNNNCNCNNAANNNN
NNTANANCGANGNANNTNTNGAGNNTATNNCNNNNNTTCNCNGGAGCTNNNNGNNTNNNA
TNATTNTTTACNTANCCCNNATGTCNCNCANTTATTNTNNTCNNNNCCNNNCNCNCCCNG
>A236
GGNGNGNTNNAGTTTGCCGCCCNGTTANNTTNCCTGATNNNTTTACGGNANAAAANANGA
GGGAGNAAATNGCGCAGGNNGGGAAACCCCTTNTTNANNGGAGGGNNANNNNNNTCNTTT
GGGGGGTTACAGTTGGNNGATNTNTTTNTNNNNNCTNTTANGNCNGTANNNNGGTCGGNN
NNNCTACCTANNGCNNNNNTNAGNTANTNNNCGNTCCCNNCNNTNTGTTTTGCCCNNCCN
TNTTTCNNNAGTNGCANTNGNCNNCCNNNCTAANCAACNNTNTNCNNCANCCNTTNGNAN
CGTTTTTGNANACTNNGNGGNNAGTCNACTNNNGNNATATNTTNGNNCNTNNCTTTTANC
TATGNTTCNGCCNTTNNTCTNNNNTATTNNGNGNCCGAGGCAGNNNNNCNCGAGATNCGN
NGNNNGANANNNNNANNGNGNGNNATCCNCNANGGGNNNNAAANATNAGNNNTNTANNGN
NCTCANCACGNNATNCCGNATAANANNTCNGCTNNCNTAACGTNNNACGCCAACACNNNA
TGANNATGNNAATTTANNTTNNNAANGANTNANTNNCCGCGNNNNNGCCNATNTNNNNCC
NNAGTANGAANGNNNNANANTNNGNTGTTAACTTGNANGGNNNGNGAANNTGGANCTGTN
NGCACAANCTNANCNNNGACTGAANNNATAAAGNNCGCNANGGNNANNNNGANAANTATG
NNGNNCNANGATNANATGNTNTATNNANTNANNGNATCGNCGCGNNANTTAAAANATANT
ATNTNTNCNCACNCNNNNTCGCNNCATNNGNNANGAAGNNTANTNNNCNANTTGNANTAG
NGTGTNNNGCCTNNTCNGNNAAGAGNAGAAANNNTCTNGNGGTTAANANNNCNNTACTAG
NNTCGNTNNNNGANTANCNAANGNGANNGCNGAGNTNTNGCGGANNNNNNTNATAACGNNN
NNNGACNATCCNNCNNNCTTTTNTNNTGTTTNNNNTGANNANNGCANTNNANCGNTTNCN
NAANNGCTGNNCTTCGNCNGNCNNTTATNATAGGCANTNTATNTGNAGCNTACNNACGNN
TAANTAATNNNNNNTNNANNNANNNCTNANTNATANTTNACGCGCNNTTGNCNGNNTNAN
ATTTNGNNTNNGNGNNNATANNTNNGNCTANNNNTCNCCGNCGNNGNGCTNANNCNGTAT
TANANAGTNCGGCNNGCGNGNTTCAATTNANTGGCNTATNANTNTAGATAANCGGNANTG
CNAGGNTTNNNGCGNNAAANNTTAATCNAGTCGTACCTGANAGNNGTNCAGTCNNCGCNG
NCNCNAAATTGTNACNNNNCNCNNNGNTNAGGTACNGNGNATNCNGGAGAGGTANNGGAC
GCNANCNGNNGATNGANNNANNGCGCNNCGTGANGGCNAACCNGGNNGACACATANAGCG

AGNGCNNNGGACGCNTGGCTTTAGATTNAANTNNNNNTNCAGNCNGGAGGGGTANAGCGG
CNGNNGTNNAATTGTTNANTNTNGANNNGNNTNGTCGNTNTTANNNAGTAAANAAGGNAN
ATGNGTANTAATNNANANANTNGNNGTCNTNNGGCNGNGTAANCATCGCGNTTTNTTNNA
GANNTNGNNGTTNNNNNNTCTTTTNANNNNATTATNGAGNCNATNNCCANNTAAAGNNGG
GGNNTNCANNTCGNNNNNGNANTNGTGNNTTTATNANAGNNNNNNCCNGAAGCNTANTNN
NNANCNGCCNNGNCCGAGTTNNNNAAGNNAAGNCGTNTANTATTNCGGGTTATGNGCGGG
NGNNGNATCCNNGNAGNACNACGGNTNGNGNGTNTTNNGATANNACNNTANGCGTCNCTT
NATNNCCTNTAGNTATNTATAATNCGGTGTAGNCGNCGNCATNATGNNGCG
>A237
GAGGCGNGCTAAATGCAGTCGTACAGGTAACCGTAGGGGTGGATTAGTGGCGAACGGGTG
AGNNNCACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTCTAATA
CCGGATACCACTCCTGCCTGCATGGGCGGGGGTTGAAAGCTCCGGCGGTGAAGGATGAGC
CCGCGNCCTATCAGCTTGTTGGTGGGGTAATGGNCCACCAAGGCGACGACGGGTAGCCGG
CCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCA
NCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCNACGCCGCGTGAGGGATG
ACNGCCTTCGGGTTGTAAACCTCTTTCANCAGGGAAGAAGCGAANGTGACGGTACCTGCA
GAAGAAGCGCCGGCTAACTACGTGNCAGCAGCCGNGGTAANACGTANGGNGCANGCGTTG
TCCGGAATTATTGGGCGTAAAGAGCTCGCAGGCGCGCTTGTCACGTCNGATGTGAAAGTC
CGAGGCTTAACCTCGGGTCTGCATTCNATANTNGTNANCTNGAGTGTGGTANGGGAGATC
GCAANCCCNTCCCCNCTTTGTNTCNNANNCTCCAACANNTTCTANANNATTNANCTTCTT
TNTANCCTNTCCCTTTTTTTNTTCCCTTCTTNNTNTTANTCCCCCCNNCCTCTNCCTTTN
TNTNNCNCTNCNTNTNATTCTNCCNNNATTTTTTGCTTNTTCTTTCCTGTATNNCNCTTN
ANTNATTNNCTTCGCCCCCNTTTTTNTTCTCTNNCNTNCCTTNNCGCCCNTTACNCCNNC
NTTTTTTCTCCCTTCCNTNCNNCTNNTTTNNTTTTTNCCNTNTTTTTTTATTTTNTNNAN
ANTNCATTTCTNCNCCNTNCNTTTTTTCTTTTTACCCCTCGTCTTTCTCCCTTTNTTTCT
TTTTTCTTTTTNTNTTTTCCACTNNCCCCTTTTTCCTCTNTTTTATNATNNCCTTCNCTN
TNTTATNTCTCCTCCTTNTCNTCCATCNCNNTNNNNTTCGNCCCNATTCNNNTNTCNNNN
ACTTTNTNTTTCTATTNNACCCCTCACCTCTNNCCCCCTCCCCCTCGTTTTTTCTCATTT
TTTNNTCTCCNCCTTTATCNTTTTTGTCCCCTNCTNTACNCTATNCTCCNNATTTTANTT
TTTTCNTTTTNTTCNCCNCCTTNTCCCCTTTNTNNNTATTTNCNTCTCCCGTCCCCCTTC
CTCC
>A69
ANGCGCNGCCTTACCATGCAAGTCGAACGCGTGATAGCAATACACGAGGTGGCGCACGGG
TGAGTAACGCGTGGATATCTGCCTTTTGGTTCGGAATAACCCCGGGAAACTGGGGCTAAT
ACCGGATGGTTCCTTCGGGATAAAGATTTATCGCCAAAAGATGAGTCCGCGTACGATTAG
CTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCGTTAGCTGGTCTGAGAGGACGAT
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATAT
TGGACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTT
GTAAAGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAACT
TCGTGCCAGCAGCCGCGGTAAGACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTA
AAGCGAGTGTAGGTGGTTGTCCAAGTTGGATGTGAAAGCCTTGAGCTCAACTCAAGAAAT
GCATTCAGGACTGGGCGGCTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTGGT
GAAATACGTAGAGATTGGGAAGAACACCAGTGGCGAAGGCGGCTATCTGGACGGTTTCTG
ACACTAAGACTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCG
TAAACGATGGGTGCTAGACGTTGGCGAGCTTGCTCGTCAGTGTCGCAGCTAACGCGTTAA
GCACCCCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCG
CACAAGCGGTGGAGCATGTGGTTCAATTCNACNCAACGCGCANAACCTTACCANCCCTTG
ACATGGGACTCGCCGGGAGCAGAAACNCTCCCCTTCNGTTCGGCCNGANTCCCCACAGGT
NCTNCATGGCTTTCNTCACTNCTGTCNTGAAATNTTNGGGTTAATTCCNCAACAGCGCAA
CCCTCNTCTCCNTTNCCATCAGTTNTCCTGGCNCTTTGGAAAACTNCCNGTNANNANCCN
NAGAAGGNGGGNATACTTAANTCNCCNGGCCCTTNNGGNCTGGNTANNANNNNTAAATGC
GNNAAATGNATNNNNNNNNNCCNNACCCNTCNTAAAAC
>A72
NNTNNNNNCNTTGGAGGGGGAGCCTTAACATGGCAGCTGNACGGGGCNAGGGNACAACCT

GGTCGCACGAGNTGGCGAACGGGGGANCNANTGNCATCGNAACGCNGNTCCAGTTTGCTG
GGGGATAACGTGCTNGAAAGANCANCTAATACCNCNTACNACCTGAGGGTGAAAGCGGGG
GATCGCANGACCTCTNGCANTTGGAGCGGCCGATATCANATCANGTAGTTGGTGNGGTAA
ANGTNTNCNAAGNCGACNATCTGTANCTNGTNTGAGAGGANNACNAGTCACACTGGGNCT
GAGACNCGGCNCAGACTCCTACTGNANGTAGCACNGNGNATTTNCNGNNCTATNNGNGCA
TNTTNTGATNCNTCCANGNCGTNNGCGNTANNAAGGCCTTCNGGCCCGNCAACGGTCNCC
TTGNANGAGNAATAANNGANTACTCCCANNNNCTNTGGGCTTAAGGACNNACACCNCGNG
GACCTTTNTTTNTCTCNCCATNACNNNTNTTNGNNTNGNCNCCNNNCNATANATTNTGGN
NGNCACNCNGNCCCTNNNTGANTCTTTTCGNGGNGTAAANTCTCNNANNCCNCTCCTCTT
AANNCNGGNATANNNCNCTNANGTTCCNTCGGNNANTCNACTGGNAATANCCNNNNNGNT
NTCGTNTTCTCTCCNCCCTNNTTATTNTTCNCTNTNNCNNNNTTNACNNGNTCNNACNCN
CTCTCGCCCCTCTTTTTNNNNNCNNNCCNTCNCCTANNNTCGATTNTANTACACCNNCNN
CNTCNTNGNTTCNTNNCGCCNCATCCCCCNTTATTTTNNTNANGNCNNCNGTTTGATNNT
NNNTTCCNCCCCCNNTCTNNCCNTTTTTNANNNGCACNNCNTCNCNANNTNCTCNNCNAG
NNNNTNCNNCTNCNCTCNTTTNTNCTNTCNCCCTNCNNTTNTNTNNTCCNTTTANGNCNG
CNCNCNNCGCCTTNNTTNNCTNCNCTANCGNNCNTNTTATNTTNCCNCNNNTCTNNCNGC
NNCCNTNTTTTCCACCNCCTCCCTTTCNTCNCCTNTTNCNTCNTNGCTACTNTNTTGTNN
TATTNTTNATGTCTGGNNTGGNCTTNACTCCNCCTNCNTTNNTTTNGTNTNTACCCCTNC
NCGTNTCNNNCTNNNNNCCACGTATTTTACTTTTAANNNNTCCGCCANNCCTCNTCNTAT
CCTCCCCTCTGTCNTNACNNCNNCTNTNTTACTTTNTTCCCNNTTTTTCCTCNCNCNCTN
TNANTTCNCCCNTTATATNCGNCCNNTNNCTTTTCTNNNANNTATATNGCCNNCNTCTCC
CNCCCTNCNNCNCCNTNTTNTCCNNTTTTCTTNCTCTCNGTATTTNNCTTNTTCCCCNTN
CTNTCCNNCGTACCTCCCTTATNTNNTTNTCTCTTTTTATATANCCCCGTTTCCNCCTNT
NTTTTTNNTCCTTNCCTNNCNNTTNCNNC
>A77
ANGCGCNGCCTTACCATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTA
ACGCGTGGGAATCTACCCAGCTCTACGGAACAACTCCGGGAAACTGGAGCTAATACCGTA
TACGTCCTTCGGGAGAAAGATTTATCGGAGTTGGATGAGCCCGCGTTGGATTAGCTAGTT
GGTGGGGTAATGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCA
CACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACA
ATGGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAG
CTCTTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGC
CAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGC
ACGTAGGCGGACTATTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTT
GATACTGGTAGTCTCGAGTCCGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATT
CGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCGGTACTGACGCTG
AGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACG
ATGGAAGCTAGCCGTTGGCAAGTTTACTTGTCGGTGGCGCAGCTAACGCATTAAGCTTCC
CGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAG
CGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCANAACCTTACCAGCCCTTGACATCC
CGGTCGCGGACTCGAGAGATCGAGTCCTTCAGTTCGGCTGGACCGGTGACAGGTGCTGCA
TGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGNTTAANTTCCCGCAACGAGCGCAACC
CTCNCCCTTTAGTTGCCATCATTTAAGTTGGGNCACTCTAAGGGGACTGCCGGTNATAAN
CCNAAAGGAAAGNTGGGAATNACTTCAATTCCTCNTGGCCNTNACNGGCTGGGCTACCCC
CTTCTTACAATGGGTGNTNANATTGGNCACNAGGCCCCCAGGCCCAATTNTTCCCCAAAA
CCNTTTTNATTCGNATTCCCTTTCANCCNG
>A78
AAGGNAGCNTANCATGCAAGTTCGTACANANANNNCTCTATCGGAGTGACTCGAGTGGCG
AACGGGTGAGTAACACGTGGGTGATCTGCCCTGCACTTTGGGATAAGCCTGGGAAACTGG
GTCTAATACCGAATACACCCTGCTGGTCGCATGGCCTGGTGGGGGAAAGCTTTTGCGGTG
TGGGATGGGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGATGGCCTACCAAGGCGACGAC
GGGTAGCCGGCCTGAGAGGGTGTCCGGCCACACTGGGACTGAGATACGGCCCAGACTCCT
ACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGC
GTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAATAGGGACGAAGCGTAAGTGAC

GGTACCTATAGAAGAAGCACCGGCCAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGT
GCGAGCGTTGTCCGGAATTACTGGGCGTAAAGAGCTCGTAGGTGGTTTGTCGCGTTGTTC
GTGAAAACTCACAGCTTAACTGTGGGCGTGCGGGCGATACGGGCAGACTGGAGTACTGCA
GGGGAGACTGGAATTCCTGGTGTAGCGGTGGAATGCNCAGATATCAGGAGGAACACCNGT
GGCCAAGGCGGNTCTCTGGGCAGTAACTGACGCTGAGGAGCGAAANCNTGGGGCAGCGAA
CAGGATTAGATACCCTGNTANTCCACGTCCTTAAACGGTGGGTACTANGTGTGGGTTTCC
CTTCCTTGGGATCCGCGTCCTTNNCTNANCGCATTAATTTTCCCCNCCNNTGGGGNAGTA
CNGTNCCCAANGNTTAAAACTNCCAAATTNATTTNACGGGGTGNCCCCNNNTAATTCGGC
NGGACCATGTNGCATTAATTTCNTTTTCAAACNCNAAAAAACCTTACCCGNTTTTTCANT
NTNTATAAAATTAACTCCCCAAAAANCTNTTTTCCTCCCTTCTTTCTTGNTNCCACTCGN
NNCNNTNCNCTCTTATCCCCCCCCTNCCTNNAANANNNNNGGTTTANTNCCCANTNNNNT
TTNTNTCTTTTTCCCTTCCTCCCCCCCCTCCNCGCGTNTTCTGNATTTNNTTCTNCTGNT
TTTCCNTATAANNNNCNCNTACCCCTNACCTCTCTTTNCCCTTNTTCNTTTNTTCNCCCC
TTTNNCCNTATTTTNATNTTTTCCTCCTCCCTCCCCCTCCCCC
>A79
AGGNGCNGNCNNANCANGCAAGNTCGAGCGCCCCGTCANNGGTATAGCGGCAGAACNCGG
AAGTAGTAACACGGTGGGNAACGATACCCTTTCANGTTCGTGAACAACCAAGGGGAAACA
TTTGGCTAATACCNAATACTNTCCNTAAGGAGAAAGATTTATCGCTGAAGGATCGGCCCG
CGTCTGATTAGCTAGTTGGTGGGGTAATGGCCCACCAAGGCGACGATCANTAGCTGGTCT
GAGAGGATGATCAGCCTCACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCA
GTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCANCCATGCCGCGTGGATGATGAAN
GCCCTAGGGTTGTAAAGTCCTTTCNGCGGGGAAGATAATGACGGTNCCCNCANAAGAAGC
CCCGGCTAACTTCGTGCCAGCANCCGCGGTAATACNAAGGGGGCTAGCNTTGCTCGGAAT
CACTGGNCGTAAAGCGCACNTAGGCGGACTCTTAAGTCGGTGGTGAAATCCTGGAGCTCA
ACTCCNGAACTGCCTTCGATACTGAGAGTCTCNAGTCCGGGAGANGTGAGTGNAACTGCG
AGTGTAGAGGTGAAATTCGTAGATATTCNNAAGAACACCAGTGGCGAANGCNGCTCACTG
GNCCNGTACTGACGCTGAGGTGCGAAAGCGTGNGGAGCAAACNGGATTANATACCCTGGT
AGTCCACGCCGTAAACGATGGATGCTAGCCNTTGGCAGGCTTGCCTGTCAGTGGCGCACC
TAACGCATTAAGCATCCCGCCTGGGGAGTACCGTCGCAAGATTAAAACTCAAAGGAATTG
ACGGGGGCCCCGCACNAGCGGGTGGACCATNTGGTTTTAATTCCGAAGCACCGCNCCANA
ACCTTTACCAACCTTTGANCATGTCCCNTTATGGGCCACCNAAAGATGGAGCCCCTTTCA
GTTCGGCNTGGCNGGAAACACAANGNGCTNCNTGGNCTNCTCTTCCAGCTTNGTGNTCNT
NAAGATTTTGGGGTTNATTCCCCTCAACGANNGCCANCCCTTCCCCNTTATTTGCCATCA
TTTCAATTGGGNNCTNTTAAGGGGNNTTCCTNTTANTNACCCCCCCAGGAANGTTGGGAN
NAANTTCNNTTCCNTCANTNTCNTNAANNTTGGGGCTTCCNNCNNNTTNCATGGGNGTTA
CNNTTNNNTTTNNNTATCCCNCCCTCCCAANCTCAANAATCTTTTCCTCTNCNATNCCCN
NNCCNC
>A80
ANNCGNGCTTACCATGCAGTCGAGCGCCCCGCACGGNNAGCGGCAGACGGGTGAGTAACG
CGTGGGAATCTACCCAGCTCTACGGAACAACTCCGGGAAACTGGAGCTAATACCGTATAC
GTCCTTCGGGAGAAAGATTTATCGGAGTTGGATGAGCCCGCGTTGGATTAGCTAGTTGGT
GGGGTAATGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACAC
TGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATG
GGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTC
TTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAG
CAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACG
TAGGCGGACTATTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGAT
ACTGGTAGTCTCGAGTCCGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGT
AGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCGGTACTGACGCTGAGG
TGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATG
GAAGCTAGCCGTTGGCAAGTTTACTTGTCGGTGGCGCAGCTAACGCATTAAGCTTCCCGC
CTGGGGAGTACNGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGGCCCGCACAAAGC
GGTGGAGCATGTGGTTTTAATTCNAACAACGCGCANAACCTTACCAGCCCTTGACATCCC
GGTTCGCGGAACNCGAGAGATCGAGNTCCTTCANTTNCGGCTTGGACCGGNTGACAGGTG

CTTCCATGNNTTTCNTCACTCGTNTCCTGAGAATTTTGGGTTAANTCCCGNAACAAGCGC
NANCCNTTCCCCCTTATTTCCNTCNTTAAATTGGGGCNNTTAAGGGGAACTNCCCGGTNT
TAACCCCAAAAGGAAAGGNGGGGANTNNCTTAAATCCNNTTGNNCTTTNNGGTTTGGCTN
CCCTTNNCTAAATGGNGGNTAANNGNGNCANNNGCCCTCNGCCTTNTAACCCCAAACNTT
NTNNNTCGANTGCCCCTNCCCC
>A85
AGGNGCNGCTACCATGCAAATCGAACGGGCACTTCGGTGCTAGTGGCAGACGGGTGAGTA
ACACGTGGGAACGTACCTTTCGGTTCGGAATAATTCAGGGAAACTTGGACTAATACCGGA
TACGCCCTTCGGGGGAAAGATTTATCGCCGATAGATCGGCCCGCGTCTGATTAGCTAGTT
GGTGAGGTAATGGCTCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGATCAGCCA
CATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACA
ATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAG
CTCTTTTGTCCGGGAAGATAATGACTGTACCGGAAGAATAAGCCCCGGCTAACTTCGTGC
CAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATCACTGGGCGTAAAGGGC
GCGTAGGCGGACTCTTAAGTCGGGGGTGAAAGCCCAGGGCTCAACCCTGGAATTGCCTTC
GATACTGAGAGTCTTGAGTTCGGAAGAGGTTGGTGGAACTGCGAGTGTAGAGGTGAAATT
CGTAGATATTCGCAAGAACACCAGTGGCGAAGGCGGCCAACTGGTCCGATACTGACGCTG
AGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACG
ATGAATGCCAGCCGTTGGGGTGCATGCACCTCAGTGGCGCAGCTAACGCTTTAAGCATTC
CGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGGCCCGCACAA
GCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGCANAACCTTACCAGCTTTTTGACAT
GTCCGGTTTGATCGACAGAAATGTCTTTTCTTCAGTTCGGCTGGCCGNAACACANGTGCT
TGCATGGCTGTCGTCNANCTCNTGTCTTGANATGTTGGGNTTAANTCCCCGCAACGAGCN
CAACCCCTNNCCCCNTANTTNCNATCTTTCANTTNGGAAACTCTAGGGGGANTGCCNGTT
NNAAACCCCCAGAAAAGTGGGGATNAACTTCAANTTCTCANGGCCTTNANGGCTGGGNTA
ACCCCTNTTTCATTGGCGGTAAAATGGNNACCAAAGGGGNNACCTTNANTTNTCCCAAAN
CCCTTN
>A86
ATGCGCNGCTACCATGCAAGTCGAACGCGGNGTAGCAATACACGAGTGGCGCACGGGTGA
GTAACACGTGGATTATCTGCCTTTTGGTTCGGAATAACCCCGGGAAACTGGGGCTAATAC
CGGATGGTTCCTTCGGGATAAAGATTTATCGCCAAAAGATGAGTCCGCGTCCGATTAGCT
AGTTGGTGGGGTAATGGCCCACCAAGGCAACGATCGGTAGCTGGTCTGAGAGGACGATCA
GCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTG
GACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTGT
AAAGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAACTTC
GTGCCAGCAGCCGCGGTAAGACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAA
GCGCGTGTAGGCGGTTGTCCGAGTCGGGTGTGAAAGCCTTGAGCTCAACTCAAGAAATGC
ACTCGATACTGGATGACTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTGGTGA
AATACGTAGAGATTGGGAAGAACACCAGTGGCGAAGGCGGCTATCTGGACGGTTTCTGAC
GCTAAGACGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTA
AACGATGGGTGCTAGACGTTGGCGAGCTTGCTCGTCAGTGTCGCAGCTTAACGCGTTTAA
GNCACCCCGNCTGGGGNAGTACGGCCGCCANGTTNAAACCTCAAAGGAATTGACGGGGGC
CCNCACAAGCGGTGGNAGCATGTTGGTTTCAATTTCNACGCAAACGCGCANAAACCTTAC
CAGCCCTTTGACNTGGGNACTCTCCGGGAACCAGNANACGNTNCCTTTCCGGTTTCGNCC
CGNANTCNCCCNNAGTTNNCTCCTTGGTTTNTCNTCCNCTCNTNTTNCGTNANATNTTNG
GTTAAATCTCCCTNAANCTANNCCCAACCNTTTGTTTTCCATTTTCCCNNCNGGTTATGC
CNGGNCNCTTTNTAAAAACCTCCTGGTNNNAANCCTTNAAGCAAGGTGGGGNTCACNTCA
NCTTCCNNGGCCNTNCGTGGTTGGTNTNNNCCCCCCNNAATTNCCTNTTTATTTTTTNCC
TTNNGNCCCCNCCCNNC
>A88
ANGCGCNGCTACCATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAAC
GCGTGGGNCTCTACCCATCTCTACGGAACAACTCCGGGAAACTGGAGCTAATACCGTATA
CGTCCTTCGGGAGAAAGATTTATCGGAGATGGATGAGCCCGCGTTGGATTAGCTAGTTGG
TGGGGTAATGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACA

CTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAAT
GGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCT
CTTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCA
GCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCAC
GTAGGCGGATTGTTAAGTTAGGGGTGAAATCCCAGGGCTCAACCCTGGAACTGCCTTTAA
TACTGGCAATCTCGAGTCCGAGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCG
TAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGCTCGGTACTGACGCTGAG
GTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGAT
GGAAGCTAGCCGTCGGCAAGTTTACTTGTCGGTGGCGCAGCTAACGCATTAAGCTTCCCG
CCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCG
GTGGAGCATGTGGTTTAATTCNAAGCAACGCGCANAACCTTACCAGCCCTTGACATCCCG
GTCGCGGATACNAGAGATCGTATTCTTCANTTCGGCTGGACCGGTGACAGGTGCTGCATG
GCTGTCNTCANCCTCNTGTCCTGANATGTTTGGTTTAATCCCCNCAACNAGCGCAACCCT
TCNCCCNTANTTGCCATCATTAANTTGGGNACTCTAGGGGGACTGCCNGTNATAANCCCA
AAANNAANGTNGGGATNANNTTAANTCCTTNNNGNCCTTTNCGGGNTTGGCTNCCNCNTN
CTNCAATGNGGNNNNATNGGNNNNNAACCCNAGGTCAANTNATNNCCNAAANCCNTTNNT
TNNNATTCCTTTCANTNNGGNCNNAA
>A89
TNNNTNGGNATGNGCNCCNTACCATGCNAGNTCGTACNGGNGACNACNCGGTACAACCTG
GCGACGAGNTGGCGAACGGGTGAGCTAATGTATCGGAACGTGCCCAGTTGTGGGGGATAA
CTGCTCGAAAGAGCAGCTAATACCGCATACGACCTGAGGGTGAAAGCGGGGGATCGCAAG
ACCTCGCGCAATTGGAGCGGCCGATATCAGATTAGGTAGTTGGTGGGGTAAAGGCCTACC
AAGCCGACGATCTGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGG
CCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCC
AGCCATGCCGCGTGCGGGAAGAAGGCCTTCGGGTTGTAAACCGCTTTTGTCAGGGAAGAA
AAGACTCCTACTAATACTGGGGGTTCATGACGGTACCTGAAGAATAAGCACCGGCTAACT
ACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTA
AAGCGTGCGCAGGCGGTTATGCAAGACAGATGTGAAATCCCCGGGCTCAACCTGGGAACT
GCATTTGTGACTGCATGGCTAGAGTACGGTAGAGGGGGATGGAATTCCGCGTGNAGCAGT
NAAATGCGTANATNTGCGGAGGAACACCNATGGCGAAGGCAATCCCCTGGACNTGTACTG
ACGCTCATGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCC
TAAACCNATGTCAACTGGTTGTTGGGAGGGTTTCTTCTNCCCAACTTANNTCANNCCCNT
GAATTTGACCNCCTGNGGGAGTACGGTNCCAACGGTTTGNNNCTTCAAGGGAATTTANCG
GCGACNCCCTCAANNNGNTNGAATATNNTGTTNACTTCCNTNTNNCTTCTAAAAACCTTN
CTNCCCTTNANCTTCCNAGTATTTTTNTNATNCNTTAACCCTTTTNNTNCCTTCTTNTTC
CTTNNTTTCNTTTTNTTCCTTTNTCCTCCTCTTNATNNTTTTCGTNCNTTTTCNCCTCTT
NNCGATCTCTTTTTTTNTTTTTCCNAANGNTCTCCCCCTNCCCNNCTTTTTACNNCNTCT
NANNCGTCCTNTTCCCTNATTTCCCNTNCNCTCCTCTTNCGCCCTTNTNTCNTTTNTTTT
TTNTTNCCCCNTTTACCCCNCTTNTTTTTTNTTTCTATCTCCCTNCCNTCCCNN
>A90
TTTTTTNNNNGNNTGCGCNCCTTNCATGCAAGTCGNACNGGNGACNACNCGGTACAACCT
GGCGACGAGTGGCGAACGGGTGAGCTAATGTATCGGAACGTGCCCAGNTTGTGGGGGATA
ACTGCTCGAAAGAGCAGCTAATACCGCATACGACCTGAGGGTGAAAGCGGGGGATCGCAA
GACCTCGCGCAATTGGAGCGGCCGATATCAGATTAGGTAGTTGGTGGGGTAAAGGCCTAC
CAAGCCGACGATCTGTAGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACG
GCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATC
CAGCCATGCCGCGTGCGGGAAGAAGGCCTTCGGGTTGTAAACCGCTTTTGTCAGGGAAGA
AAAGACTCCTACTAATACTGGGGGTTCATGACGGTACCTGAAGAATAAGCACCGGCTAAC
TACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCAAGCGTTAATCGGAATTACTGGGCGT
AAAGCGTGCGCAGGCGGTTATGCAAGACAGATGTGAAATCCCCGGGCTCAACCTGGGAAC
TGCATTTGTGACTGCATGGCTAGAGTACGGTAGAGGGGGATGGAATTCCGCGTGCNGCAN
CGNAATGCTTAGATATNCNTAGNAACNCCTATNNTTAACGCTATCCCCNGNNACCTGTAN
TNACGCTCATGCACGAAATCNTGGGGANCAAACNGGTATTTAGATACNCTGGTTGTTCCC
CNNTCCTTNANCGATNTNCAANTTNNTTGTTTGGTATGGGNTTTCTTTCCTNTNNCCCTT

TTTTNATTCNTTNNANNTNNTCTTCCNNNNNNNAGNCNCCNCCCTTTTTGTCTCANNTCN
CNTAATTCNATTTNTCCGGGACCCCCTACNNNNNTGTTANTTTTNNNNCGTNTTNTCTTC
CCCCNCCNNTNTTNNNNNNTTCTNCTCTNCCNTTCNCCCTTTCCCNCANTTTTTNCTNTT
TATTTTTTCCCTCTCTCCCCTNTTTCNTCCNTTCTNTCTTCCCTTCTCTTCTTCNTTTTC
TCCCCTTNTTCACCNNTCNCTTCTTTCTCCCCNTTCTTCCTCCCTTTTTCTTTATTTTNN
NCNACNNCCTCTCCTNTCTCCCCCCNNCCCNCCNTNTTTTCNAATTTNNNTCTTTNACTT
TTCNCTCTTNTTTCCCCTNCCNTCCCNTNCCTCTCTTCTTNNCCTTTCTTTTTTTTTTTN
CCCCCCTTTCCCTCTTTTTTTTTTTTNTCCTCNTCCNTNCCCTNCCTTC
>A98
TCGCAGCTACCATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAACGC
GTGGGAATCTACCCATCTCTACGGAACAACTCCGGGAAACTGGAGCTAATACCGTATACG
TCCTTCGGGAGAAAGATTTATCGGAGATGGATGAGCCCGCGTTGGATTAGCTAGTTGGTG
GGGTAATGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACACT
GGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGG
GCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCT
TTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGC
AGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGT
AGGCGGATTGTTAAGTTAGGGGTGAAATCCCAGGGCTCAACCCTGGAACTGCCTTTAATA
CTGGCAATCTCGAGTCCGAGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTA
GATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGCTCGGTACTGACGCTGAGGT
GCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGG
AAGCTAGCCGTCGGCAAGTTTACTTGTCGGTGGCGCAGCTAACGCATTAAGCTTCCCGCC
TGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGT
GGAGCATGTGGTTTAATTCGAAGCAACGCGCANAACCTTACCAGCCCTTGACATCCCGGT
CGCGGATACNAGAGATCGTATTCTTCANTTCGGCTGGACCGGTGACAGGTGCTGCATGGC
TGTCNTCAGCTCNTGTCGTGANATTTTGGGTTTAANTCCCGCAACGAGNGCAACCCTCNC
CCCTAGTTNCCCATCATTAANTTTGGCACTCTAGGGGGACTTNCCGGTNATAAACCCAAA
GGAAGGTGGGGATAACNTCAANTCCNNATGNCCNTTANGGGNTNGGGCTCNCCCNTNCTA
CAATGGNGGNANANTGGGCACCNAANCCCCNNGTTTAANNTTATTNCNAAACCCCTNNNN
NTTCGAATNNCNNTTNNNCTCNNGGCCTAAATTGNAACCCTN
>A99
ATTCGGCNGTCCTTANCATNNCAAGGTTCGNAACGGNCCACGCCNGGGGGNGACATGTCC
ATGANCTGGGCTGAAGTTNACGANACGTGGGGTAAATCTTACTCACAGTCNTCGTAACCG
GTNACNACACACTCAGNNGGAGAGACNTTGACACCTAACTACNAGGNATGANGCCNACAT
ACGCGGGAGGAAAGAATNTTACTCGAGTAAANTGGCAGTGAAGCACCCGCNANTGCGNTT
TNGCCNNGANTGGNGNGGNNANNGNCTANNTAGGNAANTATNNATNTNTNGGCTCACCAG
GATNNTCAANCTCTCTGTGGACTNAAGACACACGACCAGACCTCCTACGGGACTGCAGAC
ACTGGGCCAATATTCCTACAATGAGGCGAAAGCCTGATANTTTTCATNCCATNGTGCAGC
TGATGNTGNTCCTTACCGGTTTGCTAAATGCTCTGTTCAAAGGCCTTNGATGTTGNCAAC
TGCCTTTTGNACGAAANCCCAAGACTAATTTCTCTGACCTAACAGACCGGCGTCTATNTA
CGGANNCNGGTNCTANATATTNTNCGGGATTTACTAGCGGTNNNAANGCGCCANCGTATN
ATACATCAGTGTATAGCTCAGCGGTTTGATAATCAATNTANTNTCAACTACCAGCAANTG
CCTTCGATACATGGTAAATACTTGTTGTCCAAATCCACGTGAGTGGANTTCCTACTGNAT
NTGTGNCAGTCATTGCATGTATATGCAAGCAACACCGGGTATGCGAATTCCGGCGTGTAT
GCATTCGAATATCTGATCGCATANGCTGCAGAAAACCACNGATGNGCNAAACGCAAATNA
CCANACNCCTGGCTCTTCACACCTCATGCACCTAANGAGAGCTNGAGCATNCNGNAATTT
ATNNCCNTGGTNGCTCCACNTCCTAAANTATGTCNNCTGNCTGTTGAGTNTTCNCTNACT
CTTTAAACTAACTNANTTGTNCAGNTGACCGCCTGGCGAGTNCAGCNGCTGGGTTAATTC
TCANCNGCAATTNACAGNCTTNCCNCNCNTNGACTTGATNATGTNGNTTTANTTCAANCC
CAANTTNTTAAANNTTTANCACCTTTTNACATGTNCTGAATTTNTCNNANNNNCTTTNTT
NTTNAAANNTANNTTTAANCCCNNTNCNTNNNGNNNCNNCTCCCCTNNTNTCCNTAAATT
TTTNGTTAATCNCNNNNNAATNNCNNCCNTTNNCCTTNATTNNNNNNTTATTTGNCCNCT
TTTNNNNCCNCCNTNCNAACCNAAAAANTTNNATNNNNNNAANCNNTNNNCCTTNTNNNG
GNTTCCATTCTAAAANNNNNNNNNNNGTTNCCNCCCNNNNGNNNTTACCNNAACCTTNNT

NCN
>A238
GGGGNGNTNANAAANCANCAGAATGAAAAAAGGNAAGNANCGNNAAGNAAANTTGNGTNN
TNANTANNTGNNCNAGANAGNGGNGCCNCCCNTTTTAAGNNGGGGCCNACNGCGNNCACC
CNTNCNCCGNANGGNNGCNTAANCNGNTCTCTGCCAATNTGTTNNGNNCCCNNNCATAGN
ANATNTANNNCNCAANNNACCNCNCGNCNNCGNNNNTACCCNTNTNCNNCGCCGCCNCCC
NTANCNGCTCNTNNANNANTNCNGNNCGAAGNNAGTANNNNCCAATTTTACNTATTCNTN
CNCNNNTNNGNTANNCGNCNNANNGCTANAGTNANNNTACNACGNCNANGNGNNGNANGN
NATTNAGCGTNCNNNTNNTTGNNGTNCGNCANNNTNGNACACAGCCNGTGNNNNTTNNAN
NCNNAGTNAANCGNGATTNTAGCCAGTNAATGAANTAGCAGCCNANNNGANCNGATNAAN
NGCCCNNTTTNNCNNAGNTCTTNNAANANNGNGANNNNCNNNNCNNANNANNCNNGNTNG
NGNNGCATNGTCTTAANGNNGNANNNANCANCCTAGTTTAAGNGCNATCGAAANNTANAN
TGAGANCCNNGGGAGNTGNNNNNANAACTNTNNNNNTTTTNNNCANNCANGCNNNGGGNA
NNGGNNNANANAGNTTNNAGNCATGNNACNCNNNTNNGAANCNCANNANTCANCGNAGGC
NNCNNNTNNANNNGTTAGNNANCNGGANNAGANNACGACACANGGATTANTGNNNNTCAT
NGTGNACNANANNATCANNAAGNNCANGNTNNGAATGNGAAGNNGNNANNCGNGANNNGN
NGGGNNACNGAAGGTNAGNNGANCAANGGCANNGANNCACGGATNAGNTATNNANATGAT
GTNGAGATNNANGTGNAAGTGNNNGNNGANGGNACNANAANTTTANNANCCNGACNNNTG
CCATAATNNTAANANNCNCNNGNGANNNAANNAAAGAGGNNNTNNACGAGGANGGNTNNA
NGANTANNNNNAATANANNGGGNGGGNACAATGCNGNNGTNAAGNTANAAGGNAANNANN
AAANNAANGNGAGNNNAGAGGGAGNTTNNAANAGGGTACANNNGANGGGANAACANNNGA
NAAGANAGCGANTNGCNGGCTNNANGTNTNNGNNANNCNANTCTTAANTTGTATTNNANN
NTATTNCANNCNNNCAANNTNTNNTANNACNGTTNAGNNNGNTNNNANANGNAGNNGNNT
ATANTNGTCTAAGNTAAAGGNNGANANNNGNNNNGNNNNNNTANANCANTNNNNCCNNAA
NAAGNNNNNNNGGTNNTNATTATATNATACCGNNCNGNNTNACANANTNGTATANGGACG
GNGNCCATNTGGTANGACTGNNANANNCNNNCNNANGNTNNTTNANTACNACNNNCNNTT
AGNTNNNGNCCNNNCNAGNNNCNTNTACGCCCANATNANGGTTAATANANTNNGANNCGC
TCTCNTANCNANANAAATNNNCCCGCCNNNANTANTTNNTANTAATGNANNNANNNNANG
ANCCNNAGNANANNCCATACGGGNCNTCTNTAGCNCNGTNANTNNTTGNNCATNNCNNNG
TANCGACAAGNAAANTANGNCANNANAAAAAGATNANTTNANCNCGTCNATGGNGACNGG
CANGNAANNANNNANGNNCNGTNTNNTAGNNGATAAGAAAAGGNAANGAGGGACNNTACN
GACCGNANNAGNNNATAGACAGNCNAANCTNTNTAANANNANTNCTNNGNCCNANAANCG
TTNTGATNNTCTNTCCGGCCNANGAACTNATNNATGACNANCNNCNNNNGNAGTTACAGA
CGACNANAAGAGAAAAGNNNTANAGNANCGANNGANTANNAGNNNNGCNNCCC
>A239
ATGCGNGNCTANCATGCAAGTCGAACGATGAACCACTTAGGTGGGGATTAGTGGCGAACG
GGTTGAGTAACACGTGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAACGGGGTC
TAATACCGGATACGACACTCTCGGGCATCCGATGAGTGTGGAAAGCTCCGGCGGTGAAGG
ATGAGCCCGCGGCCTATCAGCTTGTTGGTGAGGTAACGGCTCACCAAGGCGACGACGGGT
AGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGA
GGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGCAGGGAAGAAGCGAAAGTGACGGTA
CCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCGA
GCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTCTGTCGCGTCGGATGTGA
AAGCCCGGGGCTTAACCCCGGGTCTGCATTCGATACGGGCAGACTAGAGTGTGGTAGGGG
AGATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGGTGGCG
AAGGCGGATCTCTGGGCCATTACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGA
TTAGATACCCTGGTAGTCCACGCCGTAAACGGTGGGAACTAGGTGTTGGCGACATTCCAC
GTCGTCGGTGNCGCAGCTAACGCATTAAGTTCCCCGCCTGGGGAGTACGGCCGCAAGGCT
AAAAACTCAAAGGAATTGACGGGGGCCCNCACAAGNAGCGGANCATGTNGGCTTAATTTC
NACGCNAACCGCGAAGAACCCTTACCAAGGCTTGACATACACCGGAAAACGGCCANAAGA
TNGTCNCCCCCCTTNTGGTCGGNTGTNNAGGTGGTNNCATGGCTNTNCTNAACNTCNNGN
CNNGANATGTTGGGNTTAAGTCCCNCAANNANNNCAACCNTTTTTTTTNTTTTTCCAACA
TNCCTTTTGGGGTNATGGGGATNTCCANGNAATTNCCNGGNTCATTCTNANNAAGNNGGN

NANANCTNNAANTCNTNTNNCCCCTTTTNTTTGGGNTNTNNNCTNTTTNAAGGNNGNNAN
ATTAANTTTCNAACCCNTNGNNAAANAATTNTAAAA
>A240
NAGGGNGCTTANCATGCAGTCGAGCGCCGTAGCANTANGGAGCGGCAGACGGGTGAGTAA
CACGTGCGAACGTACCTTTTGGTTCGGAACAACTGAGGGAAACTTCAGCTAATACCGGAT
AAGCCCTTACGGGGAAAGATTTATCGCCGAAAGATCGGCCCGCGTCTGATTAGCTAGTTG
GTGAGGTAACGGCTCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGATCAGCCAC
ACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAA
TGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGC
TCTTTTGTGCGGGAAGATAATGACTGTACCNNAAGAATAAGCCCCGGCTAACTTCGTGCC
AGCANCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATCNCTGGGCGTAAAGCGCA
CGTAGGCGGACTCTTAAGTCGGTGGTGAAATCCTGGAGCTCAACTCCNGAACTGCCTTCG
ATACTGGGAGTNTNGANTTCGGGAGAGTTGAGTGNANCTNCNAGTGTTTAGNTTNAAATT
CNGNNAATATNCTCCNANTANNNTNNNGTGGTNGTANGGCGGGCTCNCNTGNCCNCNCNT
ATNTTATTTGTTATNTNNTCCNAATNTNNCNGNNNNCCACCCCCNTTTNNCCNNNCCCTT
TNTTTTTTCCCTTCCCCCTTCAANTTATTNANTNNNCNCCTNNCTNCCTTNCNNNTTTTT
TTNTTTTTNCCGNCNCCCTTCTCTCCTTNTCCNCGCCNTTCCCCCCCTNTGTTTTTTTNT
TACNTNCNTNNTNNNNNNTATCCCNNTNTCTTNCCCCNTTTTCCCTNCCCCTTTTTTTTT
CCCCNNCNTTTNTTTNCCCTTCCCCCCCTNNCCNCCCTTTTTCTNTTTTNTCTCCCCTCT
CTTTNTTTCTTTNNCCCCCTCCCCTTTNCCTCTTTCTNTTCNCCCCTCNCCTTNTNTTCT
CTTCTCTCCNCCTNNTANTTNTATTTNCNCNNCCCNNTNCNCNCNNTTNCTTTTTATTTA
TTTTTNTTTTNTTCCCNNNTTNCCNCCCCCCTATCTACNCTCCTCCCNTTCTTNCCTCNN
CCTTNTNCNCNTNCNNNNCTNNCCTACCCTAATNGTCTNTCGTCTCATNTNCTCTTTTNT
TTTNCTNTCCNCCTACTNNTTNTNATTTTCATTTNNCCNCCCCNTCTCCNTTTTTCANCT
CNCCCTTTCTNTTCTTCNCCCCATTNNTTTTTNTTNTCTNGTTCNTTTAATCTNCTNCCC
TC
>A241
GAAGGGNGGNAAAATGCAGTCGTAACAAGGNAACCGTAAGTTTGTNTTTTCCTNGCGCNG
NNTCGNNNCCCCNTTAGGGGGACNTGNNGACNTGANTNNTATNGNGNTTACNNCNNGCTN
GCNGNTGCCCCNCNNNNTGTNTNCNNTTNCCCCCGCCCNGATNCNNNNTGCTANNNGNNA
NNNCGACTTTGANTNNANCCNCNNGNNNAGCCNNNANTNCCANAACAATNNNNGTANANC
ACCAANNNNNNCTTANNNACCNNCNNCCCNGACCATTCNGCANNNTTTTACNACGCNNANN
GACGGCCNCTCGTTTTCTCCCNNTTTNTTNNTTTNATCTNNNTNGNNNAGCGGGNNNNAN
CNNGCNAGTTANGGNANNNTNGNNTCNGCNCNGNTCCNNCNCTNTGNTTNCCNGCTNAGT
GNANNAAGCTGANNCAGNCTGNTAANATCTNACACNCCCACNACGACTNTNNGTGTTANN
AANNNCNNGNCTTNATGNATCTGNNNGTGAGCNNNNAAATTTANTTATNNCNTNNTNGCN
CCNCNTNTNANANNTNNTAANTNANNGTCGNNGNNNANTNANNTGNAGNNAAACANGNNA
NCGNGTANNGNNNGAGNNNGCGNNANGATNCNGNNGNAGNAAANNTNATGNANTTGCGNT
TGGNNNNNANGNNGNNAGANGCNCANGANTGTNNAGNAGNAANAANTAGNCACATANNCA
GAGNNNGGNNNGANNNGNNNANNNAGGATNGANGNAGTNTNNNNGGTTTGGTNNGNNNAN
TNNANCNNANNNGACNNAGAACATTTNNTNTTAAANNNTNANNANGTAGAGCNNNNTGAC
GANANGNNNNNAGANNNTATATTNNNTNCNTNCGCCGTNCANNNTTNTCNNNNNCNCNGA
NNNCNANNCNNNTNTNTTAATNCGNCCGTCGTANNTNANCNNANNACTGNNNNNNNGCAG
TTNTTTNNNTATGNTANNGGAANNACAAGACGTATGTNCGNCNNGANNCNTNNGGTCNAC
NGNNTTAANNNCACNANNTCGTTNCNNNCCCNNCCGNNCCNCNACAGTTNNNGNNNANCN
GCNCCNTNNNTTNNTNNTAACCGACNNGGATNTNNCATTNANTCGGANNACCNCTTNTNT
TNCNTTNNNNTANCAGGAANATANTNNNAACNNNCNACCTACNCGATNTTCANTNTNNAG
TNATTANNTNGNATNGCNNNAGNCGNANNNGGGNCGCAGNGGNNANANACNTAANTANNN
NGCCGACGTCANTNCNGTCCANATGANATAANACNGNNNNNGNNATATTAAGTTNNCATC
CNGNNTAGAATAAANTAANNNTNNCTNCNNANANNGATAANTNNACNAGCNANNTAANAN
NAANNCNAGANAANATAGNGTAGGNCNNNTGGTGNANANAATAACNNCC
>A242
AGGGGGGGGCCCAACAATGNCAAGNTCGGNAACAAGGGNTAACCGNTAATAATTTTTTTT
TTGANNGCGGGGNNNTANTNGNGGNCCCCCNTTNNNANNGNGGNATCGNGNCNNTNATNA

NNNATNACGGCNTNGAGTGNTCNCNTNCACNNCTNTGNNTCNCGCNNCNTATANGTTAAC
CNTCATATTNCNCGNNNCNGTCTTCCCNNTTNNTNCNNNNNCCNNTTCCCNCNANTNANC
NGNNAAGNNACNNNANNCNNNACTNNACGTGNCNNAAGTATNNNNTGCCNCGCTNNNNGN
AGNCNNNNAATANTGAANNNGNGNNGGNCNCATNANNNNNNNGTTNNTTNNGNNGCCGNN
NANCNCAANCNNCTCGTNTAATACNNNGNNNNNTNTTTATCTNGATNNNTNNNNNCGCNG
GGCNGCTTCTTTCAGGCGNNNTTTNNNNNCNCTNTTNNCTANNNNNANGNANCGNNCTGN
GCNNATTANCNNTATCTGANTNGANTGNGANNNNNGCCAANAGNNNNCCTCNNNNGNAGT
TNNNNGNNNCNNNCGCGCNGNNANANTTNTNNGTNNNGNANCNNTNCNCCNCNTTNGCTN
NANNATNANNATCGTCGGATNGANGCNCNGGNANTNCNNGTTNNNNCGNAATNNNAANTN
ANNGNNGNNNCNGTATNCNNGNGACNTNAGTATNAANCGNNNGNTNNNTATGAAGGCNNC
GNANNAAAAGAGAGNCGNGNNNGATNNGNAGNGNNAGTGGGAANTNCNTNAANNNNNGNC
TGTNGGNNACNCNTNNATATNNNTATAANNTNTGGGNANCTNNGTTANANCNAGCGNNNN
NANGNANTGAGGTAAGCANNATNNNTGNGNNNCATTNGNCNNANNAGNANNTNGTTTATT
CGGNAGGTAGNAATNATGGNNNAGNANTAGANGAATGTANNNGNGANANTAAATNGGNTA
NNNGAAAGNANCNAGANNCCAACGNGAGATGATATGNGATAGNGNCATGANTGAGTAANN
NNTTNNANNNANTTANANNCCNNTNNACNNNTAANNNNGAGNNNTTTTTNTTTANTATNN
NTANANCANNTNCTGCTTTNTNTNACANTTTANANTNTGNGNNTGCNNNCNNNNNTNNNN
ATNTNCNNNGNAGNNNGNGNNAGAANGNGANANGAGNNANNNAANNANNAGNGANTTTCA
TNTNNTATNATTTATAGNTGGGAAANAGGGANCGACNNANGATATGNNGNGNANACACGN
TTATNNTTCCCNNGGTCTTGTNTANNNCCNTNATNTTATTAATNTNCCNNNGGCCNGNNA
NNNCNNNANNAANTCNANTAANNNNGANCNCNCGTNCTTNATATNNTNATNNTNCNGNGG
ANCGTNNTNACTNCNGTATACNTNTNNCNNACNNTCAGNGNTNNNANAGCTAACTANCNG
TNAGTCNNNTTNNGAGNTNACNNACGCACNCAACAAAANTAANTCGNNAAAACAAANNTA
ANNNNNNAACGNNNCAGCNCGNAGNAGGAGCNANGNNGCNNNTCCGNGATGAANNNCGNA
CCCANNAGNGCGNGCNNNNNGANNNNTAGNCATGNAAGCTAACCANNNCANNNANNNGNT
NNANANGGCCTNCTNNNNNGAANGTCATNNAAGGTCAGTACNGCGCANCNACNNATATGN
NNAGGGNNGANTANNCTATAGNNCCGTCCNNNANNNTTNNGANGNCGTNNGTNCGGGGNN
NGANCATANATNNGAGCGCNNGG
>A243
AGGGGGGGCCTACAAAGGCAAGNTCGATAACAAGGGATAACCGNTAANNNNTTTTTTGNT
GTGGCGGCANCAAAGGNNNATGTCCCNTATNAANGGCGGNANNATNCGCGGTTAACNNNT
AGTCNNNTNATGNGCNGGNGTCNANTCNTNNGNNNNCNCNCCTCNACNCATNNCNCTNNN
NNCCNCGNNNCGCNTATCNGNTTTNNTNNGCNCNCGANATTNNNNCGTTNTGNANAAGNG
AAGNACTNCANCNANGACTNGTNACNNNTCNNNNNNNTTNGNTNNNNCNGTNNCNNTNCN
NNTNCCTGACANNNNNNCGNGCGNGNGCGNNANTAGNANTNTNNNTTTGCNAGTNNGTNG
AAANANGCCGCTNNNCNANNANNCCCNNGNTATANTTTTCTNCGTNTTTCNNNNNTCCNN
CGCCCGGNNTTTNNNNCANTNTNTANCNCAGCGTTTNNNNTCGTAGACGNCNNCCNCNGG
GCTNTTTNCNNNTNNNNNCGNGTTCNANGCTNNANGGCNNCNNATNCCNTGGNCGNTANT
AGCGCANNGCNNNGCNNCTCGTGNNNTTNTGTNTNANAGCNTCCNNCNCCTNGNCNANTN
GNTNNTTGTGCCNNNNNNANNCCTTNNTGNANNNTATANCCNNCNNNGNGNNNTGGNTNN
NTTNGNNTNCNNCGNGAGCGNGANANNNNNNNTNNGTNNNNNTNNNANNTGGAAGAGGNA
GNNNAGNNGGNGCAGGAAAGNNGGGGNNNAGTGNNGNTGGNANNGGNGGGNNNNGNCNCN
GANGNGCCTGNGTTNGGTNNNANNAGTNGGGNGNNCTNAGANNTNAAGGAACGNGANTCT
NTGATGANATNNCNNNGGNNTGAGACNNNTNCNNNCNNNAGGGANTGNANNNANNNTGNA
NTANCNNTGGGAGNANAAAGAAAGNANAAAGAGGGNAGANNNCNTNNACATNTNGTNTCT
TANNGNNNGNCNNNCGNNGNTGCCGNGCACANNATTANNTNGGCNTNGTAGNNNATGGAN
AGACNANGTTNGNACNAGTGTTNNGGNNGNNANACNTTNNGNNANNCCGANNTNGTNNNN
NNTTNNTCTAACNCNTNNGATNNCTNTGNTNNNNTCTNNNGANTNTANCCGNNGGGNCAN
CNNANNNNATGGTNTNAACNTNANNNNNGATATNGNTACGNNCNTTANACATNNGNNTCC
GCNAGCAGNTTCGTNTNTNTANTNTNNAGNANCTNACNGNATNACANCAGGANATNACAN
NTGNANGTNNAGNCNNGTTNTGTACGNNGCCGNNCNNNGTTGTNGNTCAACNCNTNGATT
TNGGTTNNGCCNNGCGNCNTGCCTANNNGCGNANTNTTGTNGATNNNNNNGCNNCGCCTC
TNNNNTAANTTNTGNNNANCCNCNCCNCNTGCTTNTNGNACCTCGNTCAGNNGNATCNNA
NNCGANNTGNTGCNNGTTANNNNNGGAANNGNNANCNNNNACNTTANGNGNCNCCACNTN

GTNNNGNGCAGAANAGTNNANTNCANAANTNNTANNNATAGNNNCNCGNGNNGNNCGCNC
NNNGCGNNCNNTANCCGTCGNNGCATGATTACATAGNTNGAGNNTNNNACTCNGCGTGCG
NCNNCNTNNCGNCCGCNGAATGNAGNNNNCNNNGAANANTNGNNCGTTNAANNTNCGNCT
CCGCNNNCNTCANNNNTATTNTGCNTTANCNGNNNCCANANCNNNTNNNGGTNNANGNNN
CCCNATNNNTNNNCNACGNCGGCGGNAAANCGNTANNGNGCNGNTGGNANTAANAGNANG
NCNGCN
>A244
NAGGGNAGGCGTANAATGGNANGTCGNGNACAAGGGANAACCNGGTANAANTNTTTGGTN
GCGCGNGGGGNNGNGTTAANNCCCCCTTNAAAAGGGGGNTANATGNNGNNNATGNANTNN
ACACNATGNNNNNGNGNNNTGCTTGNTACNANNNNTGTNTATNTCCTCTATTTCNNCNCN
NCCNCNNTTTTNCNNTTTCTTCGCCCNGNNATCCNCCNTNTTNNGACCNNGTNANNNGAG
NCGANTNNTNGNNTCNATTGNTNANAGATCTNNNNNAAAANNGGGTNNNCNNGNCNNNGN
NCCNACGCGNANACCCGTTGNTGTNTNCNNGCNGGTAGNAANCNNGNANGNTANGCCGCN
NNTGATANNNTNGNTATTTTTATTNACGNCNCNGNNCCNCTNCNNNTTNTNNTCNGNANT
TNCCGAGGNCCNTCGGNGCCNTTNNTNNANGNTNTNNTTTTCTTAGATTGCCNTNANCNN
NNANNCNNGTNTGNNNNCNNCACAGNTANNTTGNGTAATNNANCTAGCACCNNTTGTGTG
ATTANTANACATTNATCANNNNNTNCTTGTNNNCNNNNNNNNAGAAGNGTTNNNTANCNC
NAANANNGAGNNCGNGNCNAANNAANNGTNANNNTGNNANNNCNGGNGGGTNGNNGTNNA
NGCNNANNNANANNNCGGNNNNAGNANACTTNNTTNTGTNNANNNGNTNNGNGAANACAN
NNTCGTNNCNNNAGTGNAAATANNTATCTTNNNTNGNNNNNNCNNNNGGNNNANCATANN
NANNCANTNGNGNNGGNTGCNGNGANTANTNTNGAGNNTNAATNTNCTNGTTCNGTNNNN
NNNNGANCTCGCNNNGTGANNNNANATNNNNNNNTAGNGCNNANNAAGGTTNCGCNCGNT
NNTATNACGTCNNNTNAGNNCGNCCNCNNTNNNTTNCNNNNTGNNCANGTNTCGCNTTTG
TNNNANAANCGGANGATNCCANNGNNNGTTGTGTNTNGTGTAANGNCCGNCGNNNNNNCG
NTAANANNANNNCGNCNANGTNTNGNTTTTTANTANNANTANACNNNGNGNNNAATNNNC
GGNGNNGNCTNNNAGGTATNTANCANNCCATTATNTANGANANANCATNTTTTNGCCCNC
CCNGCNNAGNNGCNGNCTTNNANNTNNANANANTNNTTNNNNATNGNTTNTAACCNCNNG
NGTTNNGTNNTGGTCTNCCAGANCTNGTCNNNTNTTTNNCGGAGGNNNNGNNCNAGANNA
NGCNAGNNGAAAGAANNAGAAGNGANNATANNGNNAAGATGGAGCCNNNNNNNGNAGNCC
NGTGNNGTNATCGNNNTNNAGAGANANANNTNCCGGNGCNCAANCGACGCNNCNNATACC
TNAGACGNANNTANNCNANATATAANNCNATCNANAATNGNGNAAATATNNNNGCATANN
NNNACCNNAANTGNGGAACACAGNNNNNATAACACGCGCATNAGANTAANGNGGTNANNA
GNNGANNATNNNACNCG
>A245
GAAACGGCGGGCNAAATNNATGNCAAGTTTCGATAACNAGGGGTTAACCGNTAANNATTT
NATNNGNCGACGGNCGGAGANGCGACTGNTGNTNCCCTTTTTNNANGNANAACAACCNNT
AAANNTNTTTGNNNNANTNGCNNNNCNCATGGNNTGNNTNACCNNNNCGANTACNCNNTT
ACNNCNNTTCGCNNCCCTNCNNNATCTCGNNNTTCCTTCTCCGCNCNTGTNCCNCTNTTT
ATNNNGGCGANCNNNNANGNNGNAAGANAGATTNTCNGCNTGTTCNTCGATGCGNNGGTA
AANNCNNANAGNANNCNCGNNGNNNGGACGNGTTCAGCNNGNTTCNNNTGNCCNCCNNNN
NGANNNCCNNCNNGNTACTANGCNCGANATGTNTTAGNTNCATNTANTACCGCNCCNCTN
NGNCTTTNCNACANTTNTANNNGNAGTCTTNANNCTNTAGCNNCNNCCCNTNTCGNGTAN
TNAGTGTNNCCNTTAACTNNATNNANCACNANTGNACNNGNGGTGNANNGNTTGNACCNC
NNCNNAANCNNNNTNCNTGCTNGGNGNATCNCNNNNCAGCCNNNGNCNNTNGTNTNNTGN
NNNCNNATNNNNGNGNATNNNNGCNTATNTNNCCCNTGNNNTNNACTCNGGNTNGTTNNA
GATGNGNATNNAGNGNNCGTNGATGNNNNNNANNNNTNCNACANNNGNANANTACGCNNA
NGCANACNNCNGNNGNCNACNNGGNGNNCNCNCAGCTNAGANTNGNGNCNGNTTTGGCTN
NNGNNTNNTNTCNNGNTNTNNGANTTGTNNNACAATNNNNATNGAGNGANTNAANNTAAT
GTNANNNNATNNNANNNTNANANCGNNNNNAGNGATTAGCNGGNNNNGAGNATGGANTNA
AAGNNNTNNNTNTANCNNGTAAANNNACNAGTAANCCATCAANANNANNNGATGATNNNT
NTNNTNNNAAGCACTNGGCNNNNGNNAAGANANNACTTGANTNGATNGCNANNNGGNNAN
NNCNNGAAANCGANTGTNANGCAGNNTCNNTNNNGCGCNAGNGGTNTANNTCCGACCNNN
GNNNTTTTTTNNNNTCNCTGTNCNNNCGNNACANATTTNGTNTNGACAGCNCGANNCNCN
NNTNNANANNGTNAGAGNCTGTNTATNNNTATATANCGNCACNGNNGNGTGTANTNTCNA

CTATNGCNTTNNAAGANGANCGCANNTCCNAATTTNTAATTTNNTACGTGTGATNNANNG
NCANNANATANCNGGNTNATNNNNGGGACNNNTNNTNNNGTGCGCNGTNNTNTNTNTTNN
NCCCCTCNGTATNANNNNNNGAGNCCGNNGNTANGNGGGNGCNNTTAATTNATNNNNANN
AANCGCGNNCNCAANCTNTGGATNAANCCGNNGNGNNGNTNNANTTANCCTNTCACAACG
NNNNTGNGGNTNCGGNNNCGANCNAGANGGCANATGTTNGANTNTNCNGCNGANACGGGN
AGGNNNGGGNANNNTTTNTATNANNNTNGTAGCTCGAGNNCGNNTGTCANTTCGNNANNN
AACNGNAAGNTNAANNGNGANNTANTCNACNNNNCCGCNACNNNCNNGGAANANAGNNAN
NTNNANCAANTAGNNAANNAGANAGNGNNCGGANACACANNTGATNACNANNGNTCNCTN
GTGNNAANCNNTGNGTGNNCAGCGNANATNNGTANNNACNCGCNGNNANAANATNGTNNC
CTCNTANNGGNAATAGANNANGCNCN
>A246
NNGGAAGGGGCGNCAAACAAAGNCAAGGTCGGAACAAGGGTAACCGTTAANATATCNGTA
ATTGCAGACGCGGNGAAATNACGGCCCCTTTTNGGAGGGNGGCTACACNCCGCCCNCGTN
CAAGGCCANTGTAAAGCCCGGGNGGGTACCTGCAAGCTCGTGNTAANCNCCAAGGCAGCC
GCCCTAGGGCAAANCNNGGCNANTGNGGCTAANTTCNTTAACAAGGATANCCGTANTNNT
TCTCNTGNTAGTNGNGNNNAAAANGGTNCNANCCCNCCNCCNNACCNTTGNCCCCNTTNT
NGTNTGNCNNTNNNGNGNNCGCNCNTTNTANGCCTNGGANGGAGCNGGANGNCCGNGCCG
CTTNTNGNNNCNNNTNCTNCTNTTTTCNANGGCGNGGNCNANCNCTTNTCNTGGCCNGNT
NCTTNNNTNCGCTATGGTAGCNCNCGNGGNATNNNGNCNNCCNCNNTCCNNNCTTTANGT
TCTCACCTNACTCNGCGNNCTGNNATTNNNNNNTNCNNNGCNCCCNTNNTGCNTNTCTNN
GACTNCGCCGGAGACNNNNGNATTNCNNNNTCNNCNTNNANNGTNTNTNCTGNCCAGCCN
NNNTCGCNNGAANGNATGNNNNNGGGNNAGNNGTCAANNCNTGCGNGACTANTGNNANGT
NGGCANNGNNANTGGNNGNTAGTTGNNNATGGTGTTGGNGNAGCGNGCNNNNNNNAAGNN
ANNNAGCATTATANCNAGAAGGTTGGNTGNTANTTNACCGNCTNNGGNNCNTATNAGGGG
GCANNNGNGTNTNNTTAANNATNANNNGAGNCCTGNGNTGTNCGGCTCNGGTGGNTTTTG
TATNNNCNCGTNNACCGCNANTTTTGNCNNNGNGNGTNAGGTAGNNNCATNNNCNNGNGG
NGANTCGATNTNTNNAAGNGGCTGGAGGANNNCCAANNTTTTANNCCNNNACTCCCTTNT
TNTTATTNTNCCTNCCGCNNCNCNTNATNTTATTNNGCGTNNACNNCNNNNNNTNCNCNG
NCACNCNNCNTTANATNNNTATTTNNNNGNCCNNTCGCNCCNNCGCCNTNATANCTCCNN
CCNTCNGCNNNNGTCGTGTTTTATTTNNNTNCCNCNNGAGTNGAANANNCNNANNCNNTN
TTNNCGGNTGTGCCAGTTNTTNNCGNCCCNCTNNANNTTNTTACCCCCTCATNTANTTNC
NCCCCCCCCCNNNCCCCACNCNTTANTANTNTNCACTNNNCCCNCTCTTTATTTTTAGAN
CCNNCGNACNCNTNTAACCCCNTCNCTCCNNNANCACNCCNTNTTATNGACTNANCGGNC
CCTCNNAGATAAAANNGCNAGNCCAGNAGNCGCTGTATACNNCNAATTNTTNAGAANGTT
CANNGACGCNNGGCGGNNANANCNTCNGCCGCCGANTATNNCNGANNNNNACACCNANGC
GATANNTNNATCCGNANTNGGAGTNTCNNNNCNNNNCCTTTNANANNCCNNGCGCNCANA
NNTNTANGTNTATCGNNCCNCCCGCCANTCGGTANNGNNGCNNNGATGTGTNNAATTCGC
NNCNANGTANAATNNNAGANNNGCNCNGNCNNTNTAGNTNCCNNC
>A247
GGANAGGGGGACNGNNNATGCAAGGTCGGTAACAAGGGATAACCGATNACATATCAGGCA
TTTGTCTGCGGNGGAAATACGCCCCNTTTGGACCTTTGTTACACACCGCCCCGTTCACGC
TCACGAAAAGCTCGCGGTAACACCCGAAGCCGGTGGCCCAACCCCTTGNGGGAGGGAGCT
GTNGAAGGTGGGACTGGCGATTGGGACGAAGTCCTATAAAGGTAACCGTATTTCTCCCGC
NNGNTCNTTNNTCCCCCACCCNNCCNTTTTNCCTTTNTTNTNCGCGNNGNCCCCCGNCTC
NTNNCGNCTTTTNNTGNNTCTNTNTNTCCCCCCCCNTNNTTCNCCTTTNCCGACNTTTTC
TNNNTNCCCCTCCGNTCGCTTNTGCTTNCCNTNTGCTNTNNNNCNNCGNNTTAGCNACTG
TTGTTCTGTNCNCNNGCCNNNNNTCTNTCTTNTTTTNAGTNCCACCCTGNTTGTTNTTNT
TTTCCCGTNTNCCCNCTGCTGNCATTNTTCCNTCNNNGNGNTCNNCTGNTTTCTTNCCNG
TNCCATTCTNTATTNGTNTTNTCNCNCCNTNNNNGTCCCANATTTTTATNNGCANTANNN
GTTGNNNCGNTNNTANNTCTTCTNANGTNCGNGCNTCNGCTGGGATTNNTTNTTCTNNCN
TTNTTNGNNATNNCCTGCGNNNNGACNTCNTTNNGTGNNTGANNNCNGTTNCTCNANTNN
TAATCNCTTNNNNNNTNCNANCNNATNTTTNACGNTTTCNAGAACNTTCAGNNNATTNAN
CGNNCNTNTNNTGTTNTNTTGTCTNNCTANTCNNNNACCNGNTNCNGCGGCCAGTTTTCT
NAANNNNTTNNNCGNNTNNNNGNNGGNNNTNCACGTCNCCCTACNNCNNNNNNNTTTTCN

CCACCCTACTTTTTTTTTTTCTTGTNCCNCCNCNCNCCNTTTTTTTNCNATNACTNTCTTN
TNGNACTCTNNCNTTTTTNTTNTTTNTNNNCNNCCCTCNCCTTNCTNCCNNNTCGNANNC
NCCCNCNCNNCNNTTCCTTTTTTTTTTTTTTNACGNCNANGANNTNTTTCATACNNTTNNC
CGCTNGACCGCCTTTGTNNNTCNGCCNTTNTCNTATNCCCCGTNANTTTNTTTCCCCNTN
NCTCGCNAGCCCGCTNGNNTANTNATTTNACNNCGNNNCNTTTTGTATATTTTTNNCTCC
CCCCCTTNTTCTCTTTNNTNTNNNACTCCNCNTNTCATTCNCTNNTNCANCCCNTNGNNN
TNNATTNNCTNCCCCANTCNCGANTTTNNCTNTCNATTTNTNCTATGTGCNCTCTCTGGC
CCCNCNGNNNNANCNNTCGCTNTNNTTATTCNTCNTTGCTTNNNGNCNNNNNCNTTNNNA
NNCTTTTANNANNCCNGTNAGTNCCNTTTTTNTGCCCCGNCCAACNTGNTNTANNTTGTN
TCTCCTNCNNNCTNCNTNTNCNGCGNCGCTTTTTTNNTTCNCNCGNTNNTTCATTGANNN
NNNNCCNANCTGAGNATCCCCN
>A248
ANNNNGGGGGCANNNACANNCANGGTCGGTAACAAGGGTAACCNGATAAAANTTTTTNTN
TTTTTGCNGTGGCNCAANAGAGNNCCCNTTTNAANNNNGANNGNGNNNCNGACGCCANNN
NNATTTNGGGGGNNTNACGNNGCCCCGNCNANGNANNGNCCCNCACNTTCNNTNATNCNC
NNNTNCCCCCCNCGCNNACNTTCCCGTNTNCANNCNCNCNNTTNCCCCCCTNNTNTCCGC
CNCCANNGNCNCCCGCNCCNTNNNCNNTTAAGCGTTNNTATNTGTTTTGCGNNGNNCNNN
NANNNNCGTANCCCGCCTGGNANATNCCGNATNAGNNCAGTNTCATNTNCNNNGNNNAGT
ANTGCNCNNCTNNNTNNTNTNTNNCCGGAGANTTNNCANCAACCCTGGTNAANAGNCCCN
CCGCNNCNCTTTNNGGCAGTNNTGNCNNNGTNNTNTNCGCCGNNNCCACNCNCNGGNANC
NTCNTCNTGTTCTCTNTNTNGANNTATCNATNCNGCNTAGNANNGCCTNNNNCNACTGTC
TNCNACNGNCNCNNNNNANANTTGTTNTTTTNATNCNTANACNNCCNNTGTCNTGTNNCA
NTNTTGANCNNTNNATCCTNNNNCTNNTCGTAGTAGNNCCTNCTCCNNCNCNGNNNGNNN
TATNNCGTTCCNNGTNGCNAANNANNNGTNTNNTNNTAGNNNTNCNANNGNNTNNACCAG
AGNNNGNNCNTGGTNNAATAGAAAGTNNCGGGTNNTNCNNNNTCTNCCAGNNNATNNNTA
CNCAGNTNGTNTNGTGNNTANTNGNTGNNGGCNTNTTGGCCNGNNAGANNNNNTCNNTAA
GTTNTCNGAAGNTTNANNTNNNNNANANANNGCAGNAANNNNGAAAGNNTNNTANCNNGC
NNAGGGTNTNAGNNATANCNGTAAATNNCNNNNNCNGTCNTTTNTNNTGTTTANGNTAGT
NAGNNGNNAGTGNCCNNGGCAANNGANTNGTCNNANTNANCGNGNTTNCNNNGTNTCTAN
GGCGTCCNNCACCNNNNNNNGTNNNNANGCGCNNTNATTNNATTTTTACCNTTNCCCCNT
NNCATCTTTATTANCGNNNNCNNCTNATNTANCNGNCNNCNTTTTNNATATCTNTTNNCN
CCNGACCCCATAGCCTANNNCCTNTNNTNNTCCNCAACGTNAGGACNTGCCTGTNTTTAT
ATNANTANCCGNNCGACANTCNTATAGCNNNCCGNCNGTNCANNCCANNNTATNCTACGG
NGCNCNCNTTNTNTTATCCCNCCNNCNTATTAGTGTNNCCNCNNNANCNTCCCTATACCC
NNANNTTTTTNANCNATNNANACNNNNNAGNTATTNTTTTATNNNCGCNGCGCCGCNTTN
NCTCTAATGNGCNNCACGTACACTNGCTNNNNTTTCNNTTTNNCNGNNTCGTNTNNTNCN
NANCNNAATCANTANNNNNCGAANCGTCNGCCATGTNAGCGCNNTATNNATGNATTTNNT
ANTNANTGNNNNNGANNCTNNCNNCANTGNAGCATGCNAGCNGNCNANNNTNAGCNANNG
AGNTNNNNANNNTCGCNNTANNGNTCGTCCGNGNATATANGGNAAGCGAGTNCNTCTCNC
GCGNNNNATNTANTNNNNCCNNNNCNCNGNCTNTNNNNNCTNNNANNNNCGNNGCGNNCG
TGANTNTATTNTNCANNCANCTNGTATNATNNNANNGCCGNGCGCATNTANNNTNTGTTA
GCNGCCCTNCCAGNANAAAANGANCNCGNN
>A249
AAGGGGANGGCCAAACACANGGCAAGTNTCGNTAACAAGGGGTAACCGNTAATATTCTTN
TGTNNNNNANANCGGNCGACNCANAGGCCCCCCNNTTNAANGNGNNNNGCANACANGNGN
CCTGTNTNTNGNNACNNGNNATAGNNNCCCCCTNNNNANNTGGCNCNTTTNGNNTTCCCC
GNGNTNCCNCNTCNNCNCTNTCCNNTTCTANNCGCNNTANTCAANCTTTTTNCNNCNGNN
NNTNNNNNCGCNNCNNCNTNTTAGNACANCTGGTTTTNCTNGNCGGCGAANNTTNANNAN
ATTANANCCNTTCNGGTNGCCNNNNAGNCNNNNNNTATTNNGGACCNNNTANGNNCGGNA
NAACTTACTCGNNNNATNGNATTATNGCCNTTTATTCNNCCCCNNNGCNCCTGTTTNGGG
NNNNTTAACGNTGACNNNGTGNNNAGNNGCNCCNNNGCCCCGNGCGATNTNGANNNCACG
GTCNNTNNTTNGANNNANAGCTACGCNGAGCCACGTNTGTNNGTTNTANNNNCCNCGCNT
NNTNCANTTTATNANTAACNCTAACTNCGNGTNCGNANNNTCATGATTTANANGTNTTAA
CANCTGNNATCGANTTTNNCCCTNNGATTTNGTNGNNCTNTTTNTATCNNCGAGTGNNNA

TTNNNTNCGNNNGNTTANTAAAGCAANNNNNAANNAGNANACGTCGNANNGGGTANANGN
GNNNNANGNGCNNNNCNANGAAANNGNNNGGCNNNTNAATNNGGANGCNTTGNNGGNGNT
NGGNNNNNNNNAGGACGNNGCANGAGTCNANCNGAAAGGTNAGATNGAGTGTNNGCNACN
TTNTCGATNNNNNCNAAAGGANNCNANGGNTTTNTCNTNGTGTCNNNNNANNANCNNANT
AAAAGTNNNNCTTNATGTANNNTANTGTNGGTAGGNNTNNNNNACAGNNNTNGNNNGNNA
AANTGNNNTGGNAGNGANNTANGCNGCTTNANACAGNANATGNNCNANAGANTTACNANC
ATNNTTTNNATNNNNCANNNNTTTNNTTNTNNNNTNTCGANCNNNCAANTNTTTTNNNAA
NTGAGAANNNTATNANNCNTNNGCATNTTTNNTTTTNNTNATCNCCCNNGACCGATTCNA
GCNCGNNANNANTNCCNCGCNACANNNGTCTTTGTTTANTATNTTTANNNGNNACNTACC
NTANNAGANGNCNNTNGAGATNNNNNNNGTNNTNNNTTNNNCNNNATNTTNTNNCGGNNG
NATTCNTNAAGNCACGCGNCNNCTCNGCCNNNGNNANTNANNATTNNTTATNTANNNCNN
TGNNNNNNTATTNANACAGCNGCNGNNGNNTTAANACNGNTNATNGANNNGCNAGGNNAG
NNNGNGTNCGANNTNNNGGANCGNTNNCAGANAAAANTNTANNCGNGNNGAAGCANNNGA
CNNTGCNGNTGAAGNTTNTGAANAATGTATACCCCNCNCNNNCNNGCCAGCTNTNNANAG
NNGNACANNNNNNTNGACNGCAGNANANCGGNCCTNCGGNANTAANCANCNNTTNANNNT
ANTCGTNNCATTTNTCCAGTNAANANNCGNANCGCNNAGGAGATNGANNATAAANGTTGN
NCAGCCCCTGNGANGCAAAATACGNGNAGNGATNNTNNATNNAANNGCGNCNTNANGGTN
TNTGNANNNTNGGTNNCNTTGAAGTCACTAGCTCC
>A250
ANGAGGGAGGNGTNAAAATGGCAAGGTNCGTTAACAAGGGTTAACCGANTAAGTATANTN
NTTTTNNTTANNNGGCANAAANNTNNGNNCCCCNTTTANNNGGGANNNNCGANANCTTNG
NNCNCAAATAACNATGTANNNGNCTTTGGCTGNCCCNNTCTNCNCATNATGCGCGCCTCC
NNNATNACNTNNNTCCCNNNNCNTCGCNNCCNTCNGCNTATCTNNCGCGTATTNGNNCCC
NCACNNNNNATNNCCNGATATNTTGTCNANCNTGNNTTTACCATNNNCNNCGCGNNNNTN
NNGNCGAACNATNNANTNCGNTTCGNCCNNNAACTATNNGNATNNGGNGTCTGNCGNCNT
ANNCCGCTNAGNCNNTNGAGNGTAGGNNNACNTANNGGNGCGNAGNGGNNNANNGCTGNN
NNAGCNCNGTNNGCTNGCCNANGAGCNATNAANNTNTNANGGANNNTGNTNTTNGCGNCG
CGCACNGNNATCNNNTNGNGNCNNGNTTTNNCNAGCNGNNTNNTCATCNANNGNNGNCCC
AANNGCTNTNNCTNGTTCGNNGGGNCTNCNCANTCGGGNNNTTNGCNNCNNATGTNNNGA
GGAGGGTGGANNTNNAGAACACNNNNCACNNGCNNTCTTNANATCNTNTTTTAAANCTNG
NCTNGNGCNNGNCGGNGTTCCNNNTGNAAGGAGGNANNNCCGACNTCTTNGNANTTNNTT
CNNTNNNTCNTCACNANNNNNANTNNGNTANNGACTTGAANGNTTNNNNTNNGNGNNGNN
TAGTACTNGAGGNANNANNAANNCNAGNGTCGGGTGNAGGNANNNTATGGGAANNTGNGN
GNNTCTTGNAGNACNGGGATANTNANTNNGANNNGAGNNGTNTNNNATNTGNNCTGNTTN
NNNNAACANNNNNNNATCGNNCNNCGNGGGNCGATTAAAGNGGTAANNCTGNGNNGNTAT
CNGNAGNTCATNANTNNTGTNNNAANNGGACGATNNAATAGTNNGNNGTNTNNTNNAGAN
ANGTAGNNNNNNTANTTGNNTANGNGNGTNNCNGGNANGAGATGGTNAGNNANNACNNNN
NTAAGNCGTGNCNGATNTTNGNNGGNNTNGNANNGNNNTNTNANANTNGTATANNANNGA
TNATNTNTANAGANNGNNTANTNAGGGGNCNGACCTNNTTTATNNATNGNNNTNNCNNTC
TNACGCNNTTCTTNGTCAAGCNNGNNGNNNNGGNNCCNGNNGAGNTNTNTNAAANATANA
GAGNGNTCAGNANACGNGNGACNAGNNNATGNGNGGAGNNNNAGGTACAGNNTATNANTA
NGNGAGGNANNNCGANNTNNGNTGGAGAGAANAGANANTTTTNGGNCNCCNNTTNNTNNN
NGNNCNTATNNTNCNTCAGGGNGGNGNNCNNNGNNNNCCANNTTNANTAATTNACACGNN
CNNNGNNTTNTTNGTTTAANNGNNNGCTCTNTNNGTGNATCGGTGANNCNAANNACNNTN
TNGTTTTTNTATNANCAAGNGNCTAGNATANNAGNTTNGNGCNCCCCTGAGCGCAACTNT
TACCCNNGANGTAGNNGAGTATAGTNGTGNNNTGTGNGANTANNCAGGTTGGNTNNCGNG
GNNGCANCGANCATNGNCNNNCANTNNTATCNGNCGGNGNCTNGTNNACNGANNNCNNTT
NTGACNNGGNNGNCNCATNNCGGNTNANAATGNNNNCTCNACGNNNANGAAGNNGNTNNG
TTATNTACNATNCNNTNCCGNCNACNGNATGATTGNNATNNANNCCAGNNNTAANATGAN
GNCNACNCTNANTTGTACGNANNANGACNACGNNAACCAGGACNNGTGTGTAGCGNCC
>A251
GAGGGGGAGGCNTACAATGGCAAGGTCGTAACAAGGGTAACCGCAATANTTTTTTNTTNG
TNTTGCNCNAATANNANCCCNCTTNAANNANGGNCNAGAGANNCNNNNCANTTGGGCNNC
TTTNCTCCGCNGNGCNCNNTNTNGCNCAACNCNTTNNTNTTTNCCCNNTNNNCCCCTGCC

CCGNNTTNTACCNNTNTNNNNNACNGCCNCNNNANNACTTNATTNTNNNCNANAGACGCN
CAACCGNCAGANNCNNTCNCNTTNTNNNGNTNCCNNGTANNAATCGCNTNNNNNCGCCCC
NNCNGNCGACNGNTNNNCGACTNTTGNANAACAGNGNNCGNCNNANAGNGGCACTCGATN
ANNGNCCGNNNNNTTNNTATTCTGNNNTTNTNGGCCNGCCCCTCNCTATNNNGCGCANTT
ATTANCNNANNNATNNCNANNTANCCCCCGNNGNCNNNNCGTTTTCGTGTNATAGNTAAC
TAAGNNNTNANNANNGNNNNGNCGGGNNNANNNCNNACNCANNGNNNANNNANNNTNTTA
TGTTNTTGCTNTANTCCCNGTNNGNGGTTNTNTNTNNACGNTTAGNNCANNTNCTTGNGN
NGNNNNGCTNATTTCNTNNNGCNCCTNGNTNTNNNANANTNTGNNAAANGATGNNNNNNG
TGTATCANNANGNNNGNNNGNTAGAAAANCTAGGTGCNNNGANNGNGGGAGGAANNNNGN
GTNGNGNNANGACGNNTTGNTGANANGNNATNNNTCNGCNNNNNNNNTNNNNGNANAANN
NGGNANGGCNANNNNCATNGGNNANAAANAANCNNGATNTNNNANAGNGNGTNNAGNNGN
NGANAATNNNGNANNGGGAANGANGANNNANAGGANGNNNANAAGGNTAACANTGNGTNN
NGNNGTTNAGAGGGANNGNAANNATCAGACGNATTNNTGNANNNTNTANNTGNANCNNCN
NAGTNACAGCGNTNAANNGGANNCTNTCNANNGNACGTTNNCCCCNNNCCANTCTTTNTT
TTTTTCNNNGNNCNNCNTNNNNGNTTTTTNNGANNAGCGANNGNTGNNCNANNGGAGAAN
TTNTANGTNTNATNGNGNCANATNNNNNANNTANCNNANNNNTACNGCNGGNNCNCAGTN
ACGTTNTNTTANNTATNNAGGTNGNNGNNATCCAACNNAANCCGTNGGGNCNGANNTNNN
TTNTNCNNCNCNCTATTTTANNACNACTNGAATTTTTTNNCCCCNNNCCCNCTCNGANAG
GCNAATNNNANTNAATACACNGNNCTATNNNNNNANGNATTNCNCTNCNANGATANTNNT
NGTNCGCNANANAGCGGNTTTTGTTNGNTNNNCGACNCCANNNANNTANTNNNGTNCANN
CNGNCNNNNAGAANANNGCNTATNAATNNNNTTGNATNAAANANNNGGGGTCNNCNAGNG
TCNTGATAGGNNNCGANANTNTTANNACNNNNCNTNNGCCCANGAACNGAGANNCNAAAA
AGATCNANCAAAAATNTNTTGANNNNTNNNNNCNNTNNATNCAGTTNNAANAATNATTGC
CNTNGCNAGCAGANAGATGNNANCGTCNAGNGANTATNAGNCANCNCTNNANAGATTANN
GACATNNNGCTNGAAANTAANANNNCC
>A252
GNGGATNNGCNNAAAAGATAANTCNAAAAAAGGNANAAGGNAAAGATGGTTNAATGGCGN
NNAATNGNTGANACGGNCCCCCNTTAANNNGAGNNGGAGNNCNGGNNTATNNAATTNCGN
NNCTTGGNNNACNCANNCNCNNNGNGNANCCCGNNTNNNTNNNCCACNTTTACCNCCNNC
CGNAATTANCNNNTTCNTCGNGCNNANNCNTACCNNNTTTTANNNCNCCNNNGCTNNGGN
AGANGNNCGTATANNNGANGNNCNNNTCTTNNGNGGGNGGNAAAAAANNGNAGCATNGGN
NNCCCCACGNCNCGTTNGTNANGGNTCTTTGCNGGNGCNNNTAANGGNNNCNAGNANATC
GNCNGCNANATGAANTNTNNNNTCTNTATGNCCCCCCNNNCNNNTTNCCNCNTNTNANAC
NNCCTGTTNCCNNNNNTGTNANNNCANACGANATCNNTTTTNNTNNGTANNNTTNTCNTT
NTANNCGNNNATTCNNGCCGANGANGAGNNNNNNANGNCNCNNNTNGGGNAGNNNATTNG
NCNNCCNANCTAANCTTNCNNATNTNNANANTTNCCTNGNNNNNGNAGNNTTATNCTNAN
NNCNNNNCNNNNANNNNNNTNNATTCCNNTNCAGGTAGNANTCGCNGNNTGNGTNNTAAA
NGAANNNGNANNNGCANAGGNNANNNNANTNNNGNNGNANACCGNANCTTGANANANNAC
ACGGNNNNAANNAAGCNNTTNANNNTANNGNTTAGNANNGTCANNAAANCNNCNCTGNTT
GGATNTGGTNTAGNNNCNATTNNNNGTNGATCCATNGNAAGGANANTANNTTTNGGNGAC
CTNAAAGANNAAGGANTNACNCNTAAAGTNTNNANAGANNNNNGACAGANANNCCACNNC
NNANANGCAGNNGGNNTNNNNNNTANGANNANGGGNTCAAGNNGTNCANGAAAGGTNNAN
NCNNNNNATATNCANCAGNGNGACNTATATGTTNNATANNNNNANCANNGNCAGTTTNGT
NGANGNTCNANNTNTAANNGACNTAAGGNTGTATATANNNGTTANGGCAATNNNNTANTN
NNNTAAGNGGNNANCNNNNCANAAGNTNGTNTTTANNNTNATNTGNGTTAGNCNNNANAG
NNATNANNTGGNGTNNGAGNTTTGTACNNNCNANTTNTTTNNTACNNNNNAGNGNANTAC
NCGTGCNNGCANCNNNNNNNCACAAAATAGCNATGCNNNCGNTCTNTNNNNNGNTNTANA
AGGCNGNCNNNNNTTCNNAGANANCTNNGNAGCCNANANGNANANGCNNCGNANCNGNNN
GNANNTAANNTATNAANACNGAAACNNNNANATGCGTANNNNAANAACGNCNATTNNTGN
NNTNNCACNNNNCNNTGNGCTNGGCTNGNNGAANNTAGTNANAAGNNAGGCGNGNNNACN
ANCGGNNANNATTAGANNNTNNNCNANTAGGNTTNANNTNTNCCCTNCNCNANTNGATGN
TANATTATGANTNNCNTNCCCGGANNGTATNTACANCGGNNAANAGNNNTNNGNAGCNNG
TNAANANCNGGTNNCACANGCNTNNTNGANAANAGNAAGCNGC
>A253

GNNNGGANNGNGGGGGNCCNACANAAGGCCANGNTCGNTAACAAGGGGTAACCCGGATTA
ATNATTTTTTTNTNNTGGCGNCNGNGNGNNGGGCGGNNCNCCCCCTTTANNNNGGGGGAG
ATNNCGCGNNGTNNNGTATTTGNGTCNNTATTNNNNNANGCNCNNANTNCATNNNTNNCG
CCNGTCNTGTNCTTNTNNNNTNTTTCTTACGCGCGGCCNNCCNTNTNTCCNGNNTTNTNC
GGCCTNTNCNNTTCCCNCGTNTTNNTNNNACNGCAGCGCGNNTGAANNNNGACGATNANN
GCTNTNGNNCNAAAGTNGTGGAACGNNNGGTNNTNTGCNNNNTTNATTNNNNGGNGCCNT
NCNNNTNNCNNTACNTNGTGCTNAGTNNNTAANNNCCNNGNCGNTACNNNNGGCCNCNCT
GNNNNCANNNGNCNTGTGNNNTNTNTNCNNNTNGTTTNNNTNNNNGCGNCCCGCTGACTT
TCNGNCNCNTTNTGNNACCNNNNGGTTTTCNNTANGTNNCNCCCNTNGCGNCCNGCNNTN
TTGNGNGCNNCNNTGNNTCGNNNTTNTNGCGNNCATNNTCTCGNCNNCCGNNTGNCNTNN
TTCNCCCCNCNGNGTCNCGGNNNNTNNCNTNTNTTNNTNNCNTNNTNTGCCCCCNGTACN
CGGGNNTANNNTTNTCATNTTNGCNTCNNNCNNNTNCCNTTGTTCGGNACNNGNNGCNTN
NNCNNNGTTNGATANNGCTNGGNCNTNTCTNGTCTCTNGGGGTNNNAGNGCCGGNGNNGN
TNANNAGAGCNNTNNCNGNCANGTTGCNNCTGGNATNNGCGCAAGAANTGNANANNTGNN
ANCNNNNNTGGTNTTTACNTGTGGNCNNGGGNTTNNNGNTNCCCTNTNCNNNGCATCCTG
AANTGNTCAGTNATCNNTANTNNTTNCGTCNNGCTATNTNAGGANGCNACGNTTTTATGG
NNNGGNNANGTCNNGNGNNTTANNAAGGCGGNNCGTGTNNNNGTGCNGNGNTNNANGNNC
TNNCCNGNCNCTNNNCCTNNNAGNNTANNTNNNTNNNNACTGNNNNGTNANAGTNTTNAN
TNGGNGCNGNNNANCGGAGNCTNTGTNNGCNGNGNGCGCNGTTTTANNNTNNNCTTTACA
NCNTNNGNACGTNTTGTNNNGANGCGCGCCGCTTANNNNGCNNNNNTCTCTGNTTNTNAN
NNGTTGGNGNNNNGANCNANTTGTNTGNCATGTATNNACCNNTNGNNCGCCANNAGNNGN
NNGNNNTNNATANNTNNTACCNNNGGCAGCNNGTCGGAANNNNTACGCNCGTACGNGGAN
GNNGCGNCTNTTTANNCNGTTNCNNGTNATTTTNTNNANNNNNCNNCGTTTATTTNCNTC
GCNCNCGCNCCNCTGNNTANNNGANNANGTNCNANNANTNATNNTNGNGGCANGNCTATN
NNNNNGTAGNNCNCCGNNNCNGTTATNNNCTCGTTGTCTNGTCNCNGACNNNCGNTGTAN
ATTNNGTNTNTGNGCGTGCNGAGTCCNNGNANNNNNTTTATGNCNGCCATNTCANGCGGN
AGCNNANNNNNNTTCAGAGGTATNTGCNTGTATGAGTGCGCNCGGCGGTCANTCNNCNNG
TNTCGTGCTNCTNNCCNNTGNAATACNGTNTNCNNAGANTGNNNCGGGGCNNCGTANCGC
GNGTNTNNTGGGGTGNAGGNCGNNNGNCNCATTTTNTGATGANNNNGGCNANNTGCATGT
NGGNATNATNCTNGGGCNCGGACNCNANNGNNATAANTCNCNGNGNCGNNNGNTTATCGN
TCCGCNANNCNANATNNNNNCNNNNGTAANNNTCTNNGANNNNCNATCCCCN
>A254
GANGGGGGGGCTAAACATGCAAGNTCGGNAACAAGGGTTAACCGNAAGATANCNTTNAAT
GGAGNNAGATGAAATACNGNCCCNTTNAANGGNTTGCCGNCACNCGCACCGNTCANAGGN
NNGNANNGCNCGNGGGTGACNNTGNAANTCGGTTAANGNNGCANAAGAGACCNCANCCNC
GGGAAAAATTGGGTAACGCGGNCGCGTAANTNGNCAACAGNCGGCNGNCGCAAGGNAAAN
ANNCGCGAAANANNNACCNNNGTAGCTTGGTTTNNNACTNNTGNNCGNCTNNNCTAATAC
ATNTTTNTNNANACTNAAATATCNNATNNCATTCNNNGGGGGGNAGCGNNGCGNGGGGAG
ANAANCNNCACNAGNTNNGGGGCGNGGTTTNAGANTNTGNNANCAACNNCCNGCGAAGNN
NNTCNTGANATGNTNNTNAGTAGTNTAANNACANCTTTCNNANCNNCCNNTNTTTNTNAC
CNNCCNCTNNCNNTNTTGNTNANTAACNAANCNNCNCTTNNTTNNTTCNTTNTTNGCCNN
NTNNGCCCNCTTGNCTTTNTCCCNNNCATCGNGNNATNTNTTTCNTANNANTNCAANTGN
NTTNNTTTTGTNNANGAGTTNNCNNNNAGAANNGANNNTACGGNAAGTNNATANGAANNA
AGAANANAAGGTTGNNCNNAANGNGGNNANNNGNNANTTNNNNNTNNGCGCNNTNNNCTG
NGNANCNTNCGTCNACNATCNGNTNGTTGCTNNNCCNCCNATTNGNGAGNAACTNNGCTN
TGNTTNTNNNNNANNNATTANGNACNNAANNCNCANNCNNNNNNGTANTGTTNANCTANN
NCCCNNANNNNNGCCGGCANANNTTGTANNGTCTTACNNNNATNNGAGGNTTTNNNAAGC
NNTNNCNCNNNNATATNCNCNNNACNCANATTTNNTNTTANCTTCCCCCNTCNNCTTTNT
TANNGNTNAACCCNGTANANNCNNNCNNNNTTNTTTTTTTTTTNTCNNNCANNGTTANTCN
CGNTTAANTCCNTCCGNACGNTTCTTTTTATNTTTTTANNTGCNANNTACNNNGGNCNAN
NTGNGATNNTAGCGTTNTNNCTNCCCCTTTTTTNNNCCNNTNNTTTNTANNNCCNCCACC
CCTNCNCCCNCCNNTANTNCTTTAGNGNNNNNTNTTATTANNATTTACCCGCNANCNNTA
TTNNNNNNNNNTNCNTCNNANCTNTNANTACNTNNACAACNCNNNNNNNCNTNNTATTNN
NCANCCNNCCNCCNCCANANAAANCTNNNNTNTATNNCNAANTNNNCCCGTCGCNNCGNN

NGNANTTGACTNCCCTCCTCCNAATAANAGNANNCACNTAATGCNGNNCNNATGGCNTNC
NCNTNNTAATTACANNTNCCACTCNTGNCNNTNTCTANTNTGGNAGTCNNNACNANGNNN
TATNNNTNCTTTTCGNNCTCNNCTNCNNANTTTNNNANNCNCCGNANNTNTACNNNNGGC
NCGNNAACTTANTCGATNNCCNGNNGNTNNNNNNGCACNNCGCGC
>A255
GCAANGNNGCNANNANGCANGTCGTAACAAGGGTAACCCGGTAAAANNTTTTNNTNTTTT
NNGCNGCGGAAAAGCCCCCTNNNAAGGNGANCCAGGCNCNAACGCACTNGNGGNCCNCNG
NNGCGNCCCNTTTNAGNAAGNACGGCNCCCNTTTAGGGGAGGGNAGACGCCCGCCCNANG
TCGGGGTGCGNNNCCCCGCCCCCNGNTAGCNCNCCCCNCGNCCGCGCCNNCCCGAANCAA
GCNCGGNNNGNGNGNANGTACTNTCNGNTNNNAGCCNNCANGCNTAAGCATNATCACCTT
NTTNAGGGCNCGACTNAATNTAGGGCGGGNTAGNGCGGCGAGGNNANNGGGGGGNGGGAG
ANGCGNGCGGANTGTGCCGGNNNNTAAGNNNGAGTGNANTACNGCCCGCNNANNANGNGC
TTGTGGNGGGNCGNTTTNCNANNAAGACNNCAANCNNANNNTNGNNACNATGGANCNGAC
CNNACACGCACCTNNNNATTAGCGGAGANGCNNNGATGGNGGNGTAGTTNNATCNGCCCG
CCGTCCTNCCTGNTACCCGGACNGNNNNNGNNGGGCGGCNANANCNCGCNNCNGCCGGGG
GTNTGCNCNCANNACNNTNGCGANCAGANGNGNNGGNGNCGNNANAANGAANGGANGNNG
NANGAGNGCAGCNGGTANNNCNNNCCGCANCGGANCTNGTNNGNNACCNGNNANGTNNCA
NAGGNCACGNGNAGNCCNNNGCAATGCNGGCTNNTAGCGCCNNGTGGNNTAGNNCGNNGC
GNNNNNNNNTNTANGNCTGTCAGCNNNTNGAGCNNCGNCNNNNNGTNGGATAGNNNCGCN
GCCGNNNAGCNCCGNCNGNGNGNGAGNGCNANANTGNGGNNANNNCTTNGGGCCCACCNG
NCGCGCNCNNTNNCNNNNCGNCCTTTGTNTANANNCTTNCCCCNCNCNNCTTNNNNNCNC
TCGNCNCTTNGNACCCGCGCNNGNNNNTCNNTNTNNGGCCCCNCGNGCGCCCCCTANAAT
CNAGGCGCCACNNGCGTCGGNNTNNNTTNNTACCCGNNTANACNNCCACANNNCGGNCCN
NGCNGCNNTANTNAGGNCNGTNNATNTTNGTCCCCNTCCTNNNNCTNCCNCGCNCCCCNN
TCNCCNCNCNCTNTANTNTNTNCNCACNCCTCTCTNTANNNGNGNNCNCCNCNCNGCTAT
GNGCNNCNGNCNNCGNNNNNNNNNNNGANNNCNTCNCGNNCCCNCGCNTNAGGNTNNANCC
CNNANCCGCNCAANTANGNTGTGGTTNNNNAGNGTNNNAGCNGCNCGNNCNGGNNNGATN
NNNGGNGCNCNCNNANNNNCCNANGNANGTCANCNNANNANNAANGNANANANGAGCNCG
NACNGNCNNNTNAGNCACNCCNCCCNAGACNNATGNGCNTNNCNCCCNGCNCCCNCNCNA
NNNACGCNNANCNNGTNAGNTCTTNCCNGAGTNTNGNNNNGAGNNNGCGNCNCNAAANNA
NNCCNN
>A256
GGGANNGGNCGGAATACACAGGNCAAGTTCGGNNACAAGGGNTAACCGATAAATTNTTTG
TNTNGTANTNCNGGCNGAAATNCGCCCCCCTTTNAANGNGGGGACTACANACGGGCCCCN
TTTCNAGTNCNNTGNATAAGCTCGGGGNTNNTNNCAGTGGACGTNCCGTGTTANNNCGCC
NCGAGTAGCCCGCGCCCACNGNGTTACNNGTTNGGGCANACGNGNNGGCNTCAATATCNG
NNCNCNNCGGCNANNCCCNACANTATNCTNTCGNNNCTNTTNTNNNCGGGNTATTTNNGT
TTNTNCNNNNCNCGNCANCGCTNNACNNANNANNTTNNTGNNTNGCTANNAAANNNNCNG
GTNCNANNNAGCNGTTGNGANCNNGGGGCNCGAATTGAGGCNCCTNGNGCTGTATCGGNC
NGNTTNTNCACNNNTGNNNTNGNCCNGGGACNCNTNNCNATGTNCTGCGNCNNCTNTTNG
TNNNNTTGCGNNTTGTNCGCNCCTACTNANNNNTNNGGAACNCNNNCCCNTCNNNNCCTT
TNTTTNTTCNCNNNCCTCNGTNNTTNTNTTNATNNTTNNNNNNNTNGANCCCTGCNTNCC
TTTNNNNCNTNNNNGNCNGTGNGTNNTTNTGNTTNNCNNTCNANGTNTCAATTTNTNTTN
NTCGAGCCCNTNNNCCGCNNNAGANTGGTNCNGGANGGATGAANNANNNNTNCAGAACNC
ATNTGGTGNNNNNTANNTNATGNGNGNNANANNTTTNATNATCTTNTNTACCTCGCANCT
CNTGNCCCTCNCNGANTGTTANATAGTNGNACCTCNGNNTNCANCGCANTTANANNGNAG
GAGNCNNGTGNNNNNTCTANNNNNTNTNANCAGNCNCGTTNTNGNNCNNTNTGNNTNTAG
TNGNTGATNNTNNNCNNGNANTTNNGCCNNCCGGNNTATNNTANTNNGANNNNGNGGTGG
GATNNNGTANNCCNCNCCNNAGGCCNCTNNNTNTCCCNNCCNGGNTTTTTNTTAGTTNCC
NTCCCCCCTCCTCNTNTCTTTNNCNCGCNNNNCNTTAGNTCCCCNNCCNNNTTNTTAATA
TATTTCNCCGCGANTCNCTCTNTTCGCNNCTTANNCTANNNNNNNANNNNNGTTNNNGTT
TATTATTTNATACGNNCCNCNATNCCNTNTANNANCNATCACACNGTTNNNCCCCCTATT
TTCCANCNTNNTTTCTTTTTTAACCTCCNATNTTTAGTTACACGCCCNCCCCNNCTANCN
CNNCNNGTNTNTTNACTTTTCGNTCCGACCNTTCNTTANTNTNNTTNTTACCNCCCCCCT

CNTTNTCNTCNTATNGNNNNACNTANNTNCNGNTNNCTTNTNNNNTCTGCNCCNCNCGNC
NAANNTNGNTNTATCCGTANNNNCTNNNCGGNGANTAATNNNCNNNGAATNTNACTATAT
NTTCGCNCGNGACGNCCNNAGCGTNACCTNTANNAANNGTCCCGNANNNNANNCNGNGNA
NNTNNGCGNGCCCCTNNNNNCCNCCCTATNGCTCTNTNNNCCNCATNTTNGNCTTNNTAT
ATGCCNCTTCNGCANCACTTANAGTTTANNNNANCNTTCGCNTNNNCGNTCNGNCATTTT
TCACCCNCGNGTANTNTAGTCNANCNCCGCTATACTGNANTNTTAGACGNNNGGAGNNTC
ANNTATANGCCCCC
>A257
TGGGNAGNNGGGNNGCCTTNNACAGNGCAAGGTCGNTAACAAGGGNTAACCCNGATANAT
TANNTNTTTNTTNTTNTGNGGTCGGNCTNCNNTCNNNNCCCCCCTTTTTNNNGGGNGGGC
NCCCCTNGTNGNCCCCCTTATTCNNCTTTNNTTCCNTTTTCCNTNNCCCNNNNTTCTTCT
TNTNNCCCCCCCCCTTTCTCNTTTNGNCNCCTNTCTACTCCCCNGCCCCCTNTNNTTCGC
CTNTTCGCTCNCCCCNCCCNNCCTNTCCCCTCTTTNTTNNCCAGCNNNNCNCCNNNNTNC
TGTNGACCNCTTTNGNCTATTNCCCATTNTNCTGTTGNTNTCCCCCNCCCTCTTCTTCCN
NTTTTCNTTTCNNCGTNCCNTNCNCNCTCTCNCCTTANCCNNCNCGTCTTNTTTATGTNN
CGTCNNNNCTATNTANTGNCANNCCNCTCTTTTTCTNCCCCCNTTTANTTTCTCTNTGCC
CTTGCTNNTTTTTTTNNCNNCGGCGNTCCTTNTTTCCCNNCCAGTTTTTCTCNTTCNCTT
TTNCCTTCGTGTCTCGTNGCCNCNCCCNNGNTGGTTTTTNNCTNGNNCNCTTGTTCTTTC
NTTTCTTTACCNNCTTNTTNCTCNNNTTNNTTTTTTCTTTTTTTGTNGCCNTNCCNCGNG
TCGCNTTTNATTGTNNNTATNTTNCNNCTNCCCCCCCCCNTNGTCTTGNGTNTGNTNTNT
TNNTCCNNANCNTCGCCCCCGTTNCTTTTGCTTTTTTCNCCNCNTNGTCTNTGGTNNNCT
NTGTTTTTTCTATNCTNCNNNGCCNCTNTCNNTTTNTTNTNGNNCCTTCACNTCGTCCNT
NTTCNNNTNCCCGGCTNNGTTGNNGTNNNGNNGTCTNTNNNCTGNNNGNNGGTCCNNTCC
NNGCCTTCNCCNTTNGNNTTTNCTNGANGTCNGNTCTNTTTTTGNTCCTTNCNNTTTGTN
NCCTCNNNCTNATNNTNANNGCTCNCGTTCTGTNATNTNNTNCCGNNATTNCTNTCGTTC
NTNTCNNTCNNCTTTTNCNANTGTGCCCNCTGAGNCCTGTCTNANTTTTNNACCNNCTTT
TGCCCNNNTNTNGNNNNTNTNACCCTNTGCNTGTCNTNNTNTNTNTTTCTCGGGCACNAC
GTNTTTNCCCGCNNTNANNCGGGNNNGTTNATTTGTCNCCGNNTNNTNTNTNANTTNTNN
TCNCNCCTCCATCCNGCCCTGCCNTTTTCCTNNGCGNCNTCNTTTTTNTTTTTTTCTCNT
CTTCCCCTCCTCCTGCTATTTTTTTNCGCCTTCNGCNTTTTNTTNCGCNTCCTCTNTTNT
TTTTTTANTTTGTTTCCNNCCTCCTCCCTCTCCTCCNNNCCTNCTCCTTCTCNNNCNNCN
CCANTTTGCTTTANTTNTATTTTTTTGCGTTTNTNGNTNCTGTNCNNNTNNCCGGTCGCN
TTGNNNNANTNTNATTTCCCCCCCCCTTTTTTTTTTTNNTCCCNTCCTTGTANTTTTCAC
CCTCCCCNGCCCCNNTNNNCCCTNCNATTTNTNCTTATTCTCCGCGCCCNCTCTTNCTNT
TTTTTTTANCCCCTCCCCTCTNCTTTTTNCNNNTATCTTCTNGNATNCNNNACCTCTGTC
TTTCTCNTTTNTCCNACCNCCCNNCANNTTTCTNNTTGTCTCNCTGCGCNCCANANCNTC
GACCTNAGACNGNTTTCNANTNGTGTTAGTNTTGNCTCNGCNNNNNCTNNTNNCTNTCCN
NGTNTTCNNCTNNCGNNNCTCTNGTCCTTCTTNTCGNGTCNCCTNCATCTCCCCTCCCCT
CTNTNNTANNCTCCCGCGTCTCTTTCTAATNCCACTCGCATNCNTNTGTCTATTANNTNC
NCCCNCGCACACCNTNCANNTGTTTNNGTGNNCTTNTCNCNTCCCCCCCNTNTGTCTTCA
TTTNGNCCCNGCNNCTATNTTCTNCGGTCCCNNCNNTNTTTCTGTGGTNTNGNNNCCNNN
TTCCNATNNTNANTTCTCCTCC
>A258
ATGCGCGGCCTACCATGCAGTCGAACGGCAGCACANATAAGCTTGCTCCTTGGGTGGCGA
GTGGCGGACGGGTGAGGAATACGTCGGAATCTGCCTATTTGTGGGGGATAACGTAGGGAA
ACTTACGCTAATACCGCATACGACCTACGGGTGAAAGTGGGGGACCGCAAGGCCTCACGC
AGATAGATGAGCCGACGTCGGATTAGCTAGTTGGCGGGGTAAAGGCCCACCAAGGCGACG
ATCCGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGAACTGAGACACGGTCCAGACTC
CTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCC
GCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAAAGCTTGAG
GTTAATAACCTCGAGTCATGACGGTACCGGAAGAATAAGCACCGGCTAACTTCGTGCCAG
CAGCCGCGGTAATACGAAGGGTGCAAGCGTTACTCGGAATTACTGGGCGTAAAGCGTGCG
TAGGTGGTTTGTTAAGTCTGATGTGAAAGCCCTGGGCTCAACCTGGGAATGGCATTGGAA
ACTGGCTTACTAGAGTGCGGTAGAGGGTAGCGGAATTCCCGGTGTAGCAGTGAAATGCGT

AGATATCGGGAGGAACATCCGTGGCGAAGGCGGCTACCTGGACCAGCACTGACACTGAGG
CACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATG
CGAACTGGATGTTGGGGGCAACTTGGCCCTCAGTATCGAAGCTAACGCGTTAAGTTCGCC
GCCTGGGAAGTACGGTCGCNAGACTGAAACTCAAAGGAATTGACNGGGGGCCCGCACAAG
CGGTGGAGTATGTGGTTTAATTCGATGCANCNCGCANAACCTTTCCTGGNCTTTNACATC
CACGGAACTTTTCCANANANGGNATTGGTNCCTTCGGGAACCCNTNAAAAAGTTGCTNCA
TTGGCTTTCCTCACCTCCNNGTCCTNAAATTTTGGGTTAATTCCCCCACCNANNCCAACC
CTTNNCTTNATTNNCNNCCANTTATTGTTGGNANCTTTAGGAACCNNNGGNTANAAACCN
GAGNAGGGNGGGGATTANNTTAATNTCNNGCCCTTNCGGCCNGGNTNCNCNTTTTANTTN
GTTGGGACAANGGGTTTAANCCNNNGGGNNCCCNTTCCANACCNTNTTTNCCGTN
>A259
GGGGGNCAAAANANNGGNANGTTNAAACNGTGGNTAACANAGNGAAAAATANNATAGAAN
ATNNNTGAGCGCAAAANNTGGGCCCCCCCNTTAAGGGGGNAGAANNNNTNCTCGNCNCGN
GTNTTGNNCTTNCNNGANTNTNNCGGNNCCGCNCTANTGTNGANNGNNCGTCCNNTNTNN
TNTTTTCTNNCNTTNTNNGNGGCGNNCNNAAANTNCNCGTTGNCNTANCGCACGNCCTAN
CNCNNCTNTGGANAGNCNGCNNCNACGNGNNCGACNNANCANNNTTNGCNATNCNCGNGN
TNTTNNNNNGCGNCNNCTAGGAATNGGNNTANNNTAAANNANNTNNNGNGNGTNANGCAA
TTAAANCAANCGATNNCTNATTCGNNGNNGCGTNNNTAANNCNTNNNCNACTNNNTNANT
ANATTGNCNGCNNNNANAANGTTNTANGNGNTCAGNNAGTATACNGNNCCGNCCCGCNTA
CTNTTNNGGCNNCCNTANTAAGNGGANNNATNTGNNCNNANTNAANNCNGCGNAATNNNN
TGNTNCCTTNTNCAGAGTNTANNTCNCTNACGTATANNNATAANGGNTTNTANANANGCA
NNNTNTNCTATGNGTANNTNNNGNGCGNNNGCCCAGANNAANNTNTGTTTTNTTANCCNN
TGCCCACCNNGTANGGNTNNTTNNTTNTTTTACNNTNTTTGNGNGNNNNCTATTAATTNT
NANCGTNAGAGTANGNGANTNTATNNNNGGNGTGANNAGTANNACAGATAATNGNNNGTT
NNGTTNNAAGCGNGAGTNNGNNGNGNANANCAANNGNATNNNNGGANANNNNNGAGNAGN
ANNNNTGNNNNTNAGAAANATNAGANGATNNGAANNNCAAGTCGNNAGNAGAGNTGGAAT
NNNGNTNTGTGNNCGNNNNAGNATANNNGAGGCTNCATTNGTTGNNNACNGNNGANNGGG
NCGTNGNTNNATNGGCACNANNGNGAGAAGTACNGAGGAAGGTATTNGNNNNNNNTCNGN
NNANANAGNAGNNGNNNAGNNTGGGGATNAGANAANGNNNNNTNNNNNTTTGTGNTAGNG
NCACTGNNNTNNGNNNCNGNCNACGATAANNTAGAGNNAAAGGNNGACNNANGTNNANCN
NNNTNNCCNCNANACTNNGCGCNCANGNNTTTTANNCNACNTNCCTNTTTATTTTNGTNC
CTANGNGNGCCTCNNATTTTNTTTTTNNNANNNGCNNCGNNTANTCNTANNAGNTGGNNT
NNTTNNATNTTTTTTTCGNNNCTNNGTNTTTNTNNNNCCAANTATAGATTNTCGNNGGNG
AATGNNTGAGNTTTNNNNNNTNTNNNGTTGNNNNAAAAGNAANCNAANNNTTCTNGNAGC
TATCGNANNGTTNTTGCCTNACCCANANTNTATNTANTGCCNNCNANGAGNTANNGTNNG
NCCACGNCAGNNCCTNTNNNNCGGCAANNATNTGTNTTNATTNTACNTCNTCNNTANNTN
TNNNTTATNATACANNCNNCCNNAGCCNATTTNGNTCAGNTGANCNNNNNANNNCNAGTA
GTATNANATNTNNNTNGNNNANNGNGAGNAGGNNNATACNTNNAGATANAGCCCGAANGN
CGACGGNNGAAGGGNTTAGAGATANTANTATNTANNGTGNNGCGAGAAGCCNANATNGAN
GCNNNANNNGGTNNNCGNACANTNATNAGAGNAGNNNATNGANGGGNGGANANAANANGN
NCGCAANTGAGAGNATCGCTCGGCNNNGNATGCATNTNNTATCNNGTGNCNANCAGNNGA
TANGAANNCNATANNNNNNNTGCGNGNGACNGAANAGTNGAGGCNNNGNNGGNNNNGNAN
NGTACACACGGAAGNNNNNTATATATANGNCNNNCCTCNANAATANANCCNNNN
>A260
GNGAGGGGGCCGGGCCTACNATGGCAAGNTCGGTAACAAGGNTAACCGCTAAGGATGTTG
NGGGGNTNGGTGGCCNANACGGTGCNCCCNNTTNGNGNGNNNCGAACNNCGCCCNNNANT
GNGTGCGGNANATACCNGCCCCNCTANGGNGNANTCCNCAACTATCCNNNNNNNNCNAAT
NANNCNNNNTCNACAAATCCGANTGNNTCCNACCCGNANCCNCGTNNNTNNCNNCTCGGG
GNNCNNCNCNCANNCNNNTANTATTTNNNGGNCTNATNCGTTACGCGGNGAANAANCGNT
ATAANACAGNGNGCANCCNNAAGCNGNAAANTNCNNTNTTNTTGTNTCNNGNNTAATNNN
CCNNNCNCTNATNATGCNCNGTATNNTTNGTCCGNCTATGATAGNACCCCNCNNTGTNGN
CGCCCNTNTTNGGAACGNTNNNTTNANCGNNCCGNCGANGTCTGTGCNATTGNNTNTNCN
TTNTNCGTNNCGNNACNNNTATTGNNNNNGGCNTTNTNTANTNNCCCNGCNNTANNNNNN
ATTGNTTTTTNNNGNNNGNCGACNNNNANCATNTTNTNTNTTTNNNTNGNNGCTNCTNCT

GTCNTTNTCNNCNCTCCTNGGNCNNNNNATTNNGNTNCCNTGCAANTATNTATTNTTTCN
TNNACANANCNGACCGCCANNANNANGNNTNNNGCNATGNNTNGACNGCANTGGNNGAGN
GANNTGAGTGNNGANNANAANATNTGGGATCTNTNNCTGAANTNGTNAANTGNNNGNNAN
CNNTNNNGNANNNCACANTTTTTCAGTNNCGTGNGCCTNANNNCNTNANNCCTANTTGAA
ANAAAAGANTGGNTTNNTAGTCGTGNTNNANNATAANANANNCCNTNTNNNTNGNNANNN
GTTNGNNATNATTTNCTTNCCNANCNACGNGTNACCTNNNNATNTTGANGNNNNNANNNG
NNNAAAAGNANANNNATGAGNNCNGGGGTGCNCNGCAANCATTNTCGNCACNTCANNTNTC
TTTNTTTTGNCTATCCNCNCNNNCTNATTTTNTTNGCNNTATNANNNCNTCANTCCNGNA
GCANNGNTTAATNTTNNNNTTTACGNCCCTGTCCNANNNTNNGCNCCATAANCNTTCTCG
AAATCNCNGNNTNGTNATTTTNNTTTTNTNTTANTNCCGTACNNANCACNAANTANNNCC
TNTCGNATNNCNACNCTATTTNCNTNGCCCCTTATNTTTTTCCANNCTCANTTNTNTTCC
NCNCNCNCNCNNNNNCCNCANNNNNNNGTATNCANTCNGNGNTNTAATNTANTTANNCCC
CNNNCNNTNTNCTCNTTAGNCTGCCANCNNCNGTNNTCTTANGCTNNNCCGNCCCCGANA
ANNAANNCNNTCNNNGCNGNNNTTCGACNCCANTANNACTAAANTTANTNTATATTANNG
NCTCNNATNCNNCANGNNNATTNATTNNGTGNGACNNNNTCATCTNTNCANNCNTTCNNC
NCGNGCGCNNGNCNACCCGNTAANATAANCGANNATANNCCNNNTATATAGNNNCCNGCN
ACACGATNANNTTTTTGTANACTCGCCCNAACGNTAANNNGCCNNCCNTTNNTTANNNCC
NNCGNNAAGNNANACNATGGAGCNTNCNTNAAAANNAATAGNCCN
>A261
NTTTGGNGAGGGNGGCTTAACATGGTANGCGNGGCAGGAAAANNGATATGNTGNTTNGAA
NNGGCGAAGGGGNGNACCCCTTTAAAANNGAANNGNGNCNNCNGTTTGCCNGGGGATNNA
CCNGGCGNTNTAAGAGGGNNCTCANNANNCNTNANCCCGATNNAGCCGNCNNNAAAGCCN
TTNTTGAGCCCCCNCTATCCCTTNTTANNCNGCNNGACNGNNCNAAATTNNAATTAATCT
TANTNGNACGGGGGNAAAGGGATCAACANCNCCANGCGTCCNTACNNTCGTTTNTTCNCC
CGACNATCCGCCCTCNNTTTGNNNNNTGTGNAGANGNNGGAGATNCNACNCNNGCTGTGT
ACGCNNTTTACGGANATATNCNANNTNCGCCGNGCANNCGATNNTATNNNNGTNANNCNN
NNNANANNCANNGTCNCAACGTNGTTNCNNNANCNNNCAAACANAANTTTNTATCNNNCT
NANNNNCCTGTCNNNTTTNTGTGTTNGNCNNTNACNNCNGTTTNCNTTANNNNCGNNAGT
NGGNNANAGNATNAGTNGNAGNTGNTCAANANATTGTNNANNGTAACAAAGTGTTGCAGN
GNCNNACAATNCGNNANGTGNNNNATCAGGNNNCTGNNTANGNNNACTNNTANANGGTAN
ANNNNANNNTNCCTNNNNNNNTGAAAGTNTNTNNNAGNACCNNTGNCAGAGNGNNGNCNA
NNTNNTAAGTANCGNNTGGNTNANATTAGGTNGCNNANTGNGTTCANNCANNCNNNTTNN
NNNNNATACATGNANAGANANGNATATNNNGGGGTNNATNGNATNATNTGATGTCAGTCN
ANCNNNNNNNNNNNCNNGCNNNTAAGATTTAANNNTAAANNNNCANATNNACTGATNGTT
ATCNGNNNCNNNCANCGANCANANTTTTCGNCCNNNNACCTTTATNATTATTANCTTTCC
CCCNCNCCNNTTATGTNNAACCCNACNNCNCNTTNGAACNCCANTCCNATCTATTNNTTN
TTATNNCCCNNCCTCTCANNTCCANCTAANGNTAACNGNCTTGCANCGTTCNTGNTNGTN
TTTTNTNNTCGCTCNTATNCNNNAANNNNNGNTGCNNTANNGCACNNATNNNCTTNNNNN
TTTNNTTTATNCNNCCNNGTTNTCTNNCCCCNCGCNCCCCCNANCCNNCGCCAATTTNAA
TAATNNNNTCNCNCTCTCTTATANNTNATNNAACNCCGCNNNNCTNTCCTCNNNNGNTCT
GNGNANNGNCCCNNNNNNNTNNTCTTCNNNNNANANCANNANANNTATANTNNNANACNC
ANCATAACCANAGANAGNNCTATGNATATNTNANTTTAATNGTNCCTCTGCTNAGAACNG
NNNCTGTNNTAAGCNCGCANCGAANNNTNNGTNNNACNNTCGACGGANNGCNATNAAGNN
NNCCNAGATNNCATNNAGCANTCCTTGNNTCNNNTATANTAAGTTTCCACNTNAATNCTN
NAANANATNTCTTNGCCGCNACTNNAAAAAATCNATNNCGANAGCATGNNNNCTNTNNGT
ACNANNNGCACCNNTANAAATTNANAAGTCCCGNNTNTCNAAANNGAAANAGCAAC
>A262
ATGNGCNGCCTACCATGCAAGTCGAACGCATCCGCANGGGNAGTGGCGCACGGGTGAGTA
ACACGTGGGAACCTACCTTCTGGTACGGAACAACCAAGGGAAACTTTGGCTAATACCGTT
ATACGACCTCCGGGTGAAAGATTTATCGCCGGAAGAGGGGCCCGCGTCCGATTAGGTAGT
TGGTGGGGTAATGGCCTACCAAGCCGACGATCGGTAGCTGGTCTGAGAGGATGACCAGCC
ACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGAC
AATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTGTAAA
GCTCTTTCACCCACGACGATGATGACGGTAGTGGGAGAAGAAGCCCCGGCTAACTTCGTG

CCAGCAGCCGCGGTAATACGAAGGGGGCAAGCGTTGTTCGGAATGACTGGGCGTAAAGGG
CGCGTAGGCGGTTCGTTGCGTCAGATGTGAAAGCCCCGGGCTCAACCTGGGAACTGCATT
TGATACGGGCGGGCTTGAATCCAAGAGAGGGTGGTGGAATTCCCAGTGTAGAGGTGAAAT
TCGTAGATATTGGGAAGAACACCAGTGGCGAAGGCGGCCACCTGGCTTGGTATTGACGCT
GAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAAC
GATGTGTGCTAGCCGTCGGGCAGCTTGCTGTTCGGTGGCGCAGCTAACGCGATAAGCACA
CCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAA
GCGGTGGAGCATGTGGTTTATTTCGAAGCAACGCGCANAACCTTACCAACCCTNGACATG
GGAAGTATGGGCCTGGGAAAACCNGGGTCCTTCANTTCGGCTGGCTTCCACACAGGTGCT
GCATGGCTGTCNTCAGCTCGTGTCNTGANATGTTGGGTTAAGTCCCNCAACNAGCGCAAC
CCCTCCTTTNNNTTGCCNTCATTGANTTGGGCACTNTGGAAATNCTGCCNGTNACAACCC
GGAANAAGGCGGGNATNACTTCANTCCTCATGGCCCTTANGGTTGGGTTACCNCCTNCTA
CAATGGCNGTTAAANTNGGCNNCCAAGGGNCACNNNGAACTTATCCCCAAANCCNCTCNN
TTNGAATNNNCTTTCAANNCGGGGCCNNAATTTGAATCCTNNAANCC
>A263
GNCNGTAAANNGCGGATTNTCGAANGGGANCCANGNAAANAGTGNTTTTTATAGNATGNA
GNTNACANNNNNCCCCCTTTAANNGGNGGCCNGNGNNGGNANNNTAANNACTNGCNNNGT
GTGCNGNNGCCNCTCNNTNTTCNCCGNGCTNTNCNTTNCNAAANTCGNNNTCGGCGCCNT
CNNGTAGNNNCCCCCCNCTNCCGACATTNNCNANCCNACCNNGCNGCCNGACNNGGCGGA
TCGCNCCTNTANNNAGCGNGGNGGTNGGTNTTAACNANGATNACCAGNGAANANAAAGGT
TTTTNNNCNNNNANNATANNGACATTTTAANGGGGNNGNGNGNAGANGGCGGNAATGGGG
CGNAGAGGGGTTAGGNNGNTANTCNNCCNTTNTNAGGGCANGNNCGAGGGGATTNNNGNG
NNCGTNCACCTTAANTNANANNATCTANAGCTNCCANNNTCTNANCNCNGCGNNCACNCA
CNGNTNANTAAGCAAACNCNNNGNAANTNNANNTTNNTTTNCTTATTAAGNCTNNNNNNT
NNNNNNNNANNNNNNCTNNNATNTTNNTTNNATATAGAANGGNANGGTNANTGNTTNNCA
GCNANNAGNNNNGAGCGGGANAGGTNNNNNGNGNNGGGANANNTGTGGAGCAACTNNGNG
TANCTTANNANNGNAGGTGNGTNTGNNATNTNNNNTTNANNNCTACACANNNNCNGNANC
TNAGNAGANNCAAAAGNGCATNGTNATATCNCANGANGAAANNNNATTGAANAGTGGNTN
GNNANNATTAACNNNANAAATGNNNNNTGCNNANCTTGTATNGTAATCNNNNCNNTGNGN
GNCNGCTNAATTTAATNTNGAGCAAGNATTGATNGTNANNNANNGANNCNNGCGCNNGAT
TTTNNGNNNNNNCNTTTNTTTTGTCNTNNNACGNNACNACTTTTTTNNANNNANNNCNTN
ANNTNGCCNGCNNNTTTTTTTTNNTTANCANNCCNTGNNTACNGCNTATAANNCNCNNNN
ANAGNTTTCNTTTTTNTNNTNNNCCNGNNNNGTTTNCAAGTANAACNTNGNGGNTTANNN
NAATTTATNGNNTGGGCNTNANTTNNTNNCCCTNNNTTATNTNCGCCACNNCCNCNCANN
NCGNNCGNAGTAANTATNNGACCNNNANTCTNTGNNNNNNGANNNNNCNCGCTTNTCCTC
TNNNTNCNNCANCAACCTCCNCCNNNCATTACCNAACCNAGCANNNNNGANTACCNANCC
ANANNNNNNANATNNANCCTATNNAATTNTAANAANNNGNCGGATACNNNCNNGGNNCTN
NTNNTGNANNNTNGNGNNNAGNANAANNGTANGNNNGNCGGANNATCCNAGNNNTATAAA
ANNNCANNCTCCTTNNGNGNGAANANTNGNNNANGANAANATCANANNANTTNNTTACCG
TCNCNNNNAGNNAAATTTAACCCNANCNTANANANTCATTNANGGCTANTGNNTCTNTNN
NNNCGCNNNNTGCNGNANATGNATACNCGN
>A264
CNNNGACANNAATNGNNANNTAAANNTTNNNNNNNNTNACCGNGTTGACNCGGGNNAAAA
GNNTGTGGTAACANTAGCTANGGNNNGCNNNTGGNCCCCTTTAANNNAGGNNACNCNGCN
GACNCNNTTATANACTTACNCNCNNNANNCACCCACCNNNTATNNNTATCGNCCNCCNCA
NTCNTGTNNNCNNTNACNCNCNNACCCCNACGNNCNNGNNCCATNNCNGCCCNNCCCGNC
CTTATTGCCGCCNCCNCGNNGGGNNNNCCGGNGTNTGCAGNCGNGNGAGTTTNTTNCCCC
NNTAGNNNGCNTTTTTTANGTANNTATCANANCGCATTCGANGTTTATTNNNCNGCANNN
NAAGNNANNGTNATNNAANTTAAGGGGGGTANAGGCGGGGGNGNGNGCANGGGCTGTTTG
NCNCGGTTAGNTNCGNTATATCTCGNCNGTGTNGNGNTNNGTGNNAGGTCCTGCAAAGTA
TNATTNANANAANCATNNCTGCNCGANGATNAACNNNCCNNNNNGCNNNANNGTNNATTA
NCCNNNCTNNNCGAACNGNNNTATTNTNGGNANNNNCNGNNNTCNTNNNATTNGNANCGC
CCCTNCGTNANTGNGANANCNTTACNNTGGCNNTANNANNTNNATAGTNTTANAAGCNNT
NAACTACTNAAGAGNGGNTANTNNGAGANNGNNGGNTGNTAATGANGTAANANNNNNGNN

GGTTGNAAAANTGNNGNNTNATNNCTNTANTCTNNTGNNNNCNNCTNCANGANACNNCCN
ACATTNTNNTAGNNNNNGTNGTANANANGNNCNCANAGGGNNNNAAGGNACAATGNNANT
NNCATNNTNNCTTANAGNATTNANTNCACGNNAGTGNTCCNNTNANTTNTANNTGTNNTT
NNNTNNNAGACCGNGANNNNAGCNGNGAANNGNGGNNNTATACTANNNAAATNGTNNTAN
NCANTTAAACCCGNGNNGNTNNNNCANAATTTTCCNNNGGANCNATTTTTTTNNTNCTCT
TCCNCNCTCCNNTGTNNNTTNNANNGAANNCCTTGNCNCCCTGCCGNTTTANTTGTNTNT
TNCANCNTGTNCNTTTNNTACCNCNGNATANNCNCNNNNNNCGNNCNTTGANNNTNANNT
NANNNGCNCCCCNCTNNANAAAGNNCANNGGCNNNNNNTTGAACCNNNTATCNTNCNNCN
TNTTTTTTTCCNNNNNNATATTTTTTNNGNNCACNCNCNNGNATNNCCCGCGNANNAAAN
TNTANNCNCCGNCGCNCNNATNTTATNANTCCGCNAGAGGGANNNTTAGTNNNACNCTAA
NAGNNACGNAGNNANATNNNTTCGNCCTGCNNNANCNNCTNANNCNNNNCTCTNCCNCAA
ACNCGCCANNTNGGTTACNGNTTANANNNATNTNTNGGGNGAANANGNGCCNNNAACTNN
NNNNNNAANNTNANGNAAAACNANNNATNAANNAGGGNNNTTCCNAANANAGNNAGNAAC
NGNNNCNCAAANATGANNTGTNATTNNNGTNCTNNCNNAATNNTGTTNAAATNNNGTCNA
ACGGTCCCNNACNCATAANGNCNACCGTNCATNNNANACTGTNGCCCACANAANAACAAN
TNNATNNGNCNTTCNNCAAGNATGNNATNCNC
>A265
NGGGNAAAANCGNAANTCNAACGGGAAGNCCGAAAAAATGNGNGNNNNNTGNGAGGNAAA
CNAACCCCTNNAANNNATNNNGGAGNTGTNGCNCCANNNNANNNGCNNATNTGGTNNCAC
NACAANATTAACCGCCCCATCTCNTNANCNAAGAGGCNCGCNGCNCCGNCNNNATTNCNC
CCCCCGGNATCCGCACTNTNNNCCGGNCNAGNCCNGCCNCCCNCANANANNGTCNCGNAA
TANANTACGCGTCCTANATNCCGTTTNNNANNNNGCANNNNNGACNCGNTTNANTCNNTN
TTTTNNNANANNATGAANCNNNCAGNTATNGTNGGGGAGAGTAANNAGNGGGNGAAAAGA
GGGGGCGGGGGTGTTNANGCGCNGANNNGANGGTNNTNNANNCCCGCCNCNNNTCTNNNT
AGTNAGNTCTCGTNNNTNATAGNNNACNATATNGNNANNTNNNGANNGNNNNANCCGNNC
NACNCNCTNNATNTNNTTTCANNGAACTCNCNTGNNNTNNTNNNGTNNAACNTTNNACCN
CNTNTTGCNTTTNNGCCNNAGNNTNGNANANNTGNNCNNNCANAACNNNTCNNTTGATTT
NGNNAAGACAANANCGCNNNNACNAANGNAGNGGNTANNGNTCGANATNATNNGCCGNTN
TNGNCNGNGNAGANNATNTNGNTNGTGCTATANTCTGNTAANCCACGNANNANGNAGNNG
ANAGACGAACACCGNTTNATGNATNAGCAGNNNNGNCGNANNTAAGNGNNTCNCNAGGGA
ANTAGANNTNGAAACNNNTNANGNGGNTNNTNGTNNTGTGATNNNTANNNANNCNNTTNG
CGCNAACNNGTTTNTGAGTGTTNTNNTNTCNNNATNNTNTNGACNCANCNTCNGNGNTCC
NTNTCCCCCNGNCNTTTTTCTTTTTACNNGTGCNNNNCANNGNTTTNTTNNCGNNANNNN
ACNNNAANCCNTGACNTNTATNNTTNNTTNANAGCTCNTCNNTTNTNAGNCTTANNNNAT
NTNNNCGNNNGTNTATTNTTNNANTTNNTNCNNCNNNNTNCNAANNNGGTGTNANNNTGG
CCACNNTTANGNNGCCNNCTTTTNTNTTNNNANNATTNNNTTTNNNCCACCCNCCNNCNN
CNGNGCNCTNTTNNACTNTTNTNANNCNCNACNANTTNANTTTACCNCGNGCGNCNANGC
CTCNTAAGTCANNNANNCNGGGCTCNTATCNAGATNNCNNNNCCNCGANGATANNNNTAN
GNGCGGGCANCNCNAACNTANACGCTAAANTTTAGANANNAGTNAGTACNNNGGNNANAC
GNNGANNNNCNNNNCNATNAGNAGANNCNTAAGNGNNANTAGTNNAACTNNCCCATTCNA
CNGGACANNATNACAAANAAGNNAATCCNTANCTCNATCTNNANNNNNNNNACNAAACAA
TTATAGNTNNGNAGANGGNNGNANCGCACCCAAGNGCNTNTNTNCCANCANCCNNTNTNA
GNNTNNATGGCAANTGGTNNAAGNTAAGATANNNCTNNGTNNNAGNNCATNANANTTACN
NNGCC
>A266
GGGGAAANNNNNGCNNGNACGNAACGGGGANAANNGNANNANNGCGGTATANTNTNTGNG
GTTNNANTGTCNCCCCCCTTNAAAGGNGNGNAANGNNCTNNGGCGACANATANNCNTANN
AGACTNNGGNGCGCCNCTANNCGNANGNNANACACNNATNCGAANNNNGNNGCNCNCNCN
GCCNCGGCGTNCNCNGNNCNNGNCNCGCNCNANTCNCACNTTTNTCNNNCCCACCNCCCG
NNNNNNNANGATNNTNGTNTCTACNTNANTTTGTTNTCGNCTTCTCNCNNTTTNNAAGNC
NCCNATANCATNGCNNNTNNGANCNTNNTNTANANGNNCGATNANNATNCCNNNTGNATA
CANCGCGNGAGGANGTTTGCNGTNNTATTGNCCGNNANCNNCTATANCGNCNGNNNANCN
CACTNNTTNTTANTATCCCGCNCCGCANGNNATNNAGNNGNTNTGNNNTNTTANNNAANN
CTNTNNANNNNCGNNAAGNTAGTGNGCGANNNACNACNGNCCTNTNTTNTTNNNANANNN

CNGTNGTNNTNANTANTNTNNCTNTNTANAACNATTNNGTNTNNNCCGNNNGNGCNTGNA
NGTTNNTNGTNGNNCNNANNNATGNNNNNTTNTNTTGAACNCNNNCTAGCCANGCAAANN
TGANANGAGGNTGNNGAGTGNTGGNNANNNNNANGNGNANNANNNAANAGNNANNNGATT
ATNNTATGTNNGATAGNANGGGNNNGNCAGNAGNNNANNCANNNNNAAAANNAGTTGNGG
GGATNNNNNATNGNNAAGTGNNACGGNTTNNATATTNNTAGANANGAAGNGAAATGNNGT
ACNNGATTATNNNANNGNNAACAAANNNGNCNNNACNNGGCCAGGANNTTAANNCNAGNG
GNNGGNAGNANAATNTNANNGGACAAGNNNNGNNNNTNTANNCNNNANNNNTTATTTTNN
ANCNNNNAGGNNNNNANTTNTNNANGNNTNGCNCNNNANNTNNGNCNNANNTTTTTTTNT
TATCANNGCNCGTTNGCTCANNTAANGNNCNNNCNNGCANNNNNCGTNNTNNTTATTANN
ACGTCGCNTNNGNNNACANNAGCCGNNNTCTNTTGCNANCTGTNNNCNNNGCNNNTAANA
GANNNTNANACTANANNTANTAACCCNNCNNGCCNCCCTCTCNNNCNCNNNNNNNACTAN
TANANACNGNGTCTCTNNGNTNATGTNCNCNNCNNGNGCNNTNNNTCTTATCNNTCNCNC
NCCGNTNNNNTATNCTNNNCGCANNNGNANNGNANNANNTNNTANCNANCCGNNNNNCGC
ANAGTGTGNCTANGANATNTANANTTANAGNGCGAGNGTNTNNNANGGANGNNGCCTTNG
CTGNTNTNAAATAGNNCANATANNNANGAGCNANGNGCNTANGGNANAGNNCNNNANAAA
GNNTNACCGNTTTAGNNNTGNNANGNTGNGGNNNGCNAAGNNNTNNNTATCATNTNGCNN
GNANNNNNANATTANNTAGCTNCTANNNTNGAATGTAATTNGANCNGAAGNANAATNGNA
NCNAACCNANGGNNCANANANTANGNCNN
>A267
GGGNGGANTAAAANGGGGANGTTTNAANGGGNANAANCNGGNGNAAANAGNNANANTNAN
TNGGNGGNNNAAAANNGCCCCCCNTTANGNGGNNTGACGCNNNANCNNGCNCNNGTTATG
AATATGCGCAGCTAGGNTGGCNACCTNNTNGNGNNNGTNNCCNCCANTATTTNCGNNNAN
NCNCCATTAANNNGCGGNCCCCTNGNGTCCCNTNTTACNNCCCNCCCGANNNTCCGCCAT
TNGNNCCAACCCGNNNCNGNAGGNNCNNNGNNNACNANNCNNCNATNGTNAGTCNNGCGG
GNNNNTAGGNTATAAAGNNANNANNGNNCANNNCATTANAGANNTTANTTTANNNGNGCC
NAANNANTACANACTNATNTNTATTNCCNTTGNNNNNNTNACGGTNGGNANANNGTCCCG
AANGGCNATATCGGCNCATTNTAGCGNTCGATTTTGNCTNACGGNCGGGTTCNACNGNNN
NNGTAAAGGTCTCTNGGTTACANANNATNACNNATGTGCANGGTNANTGAANTTNNTNGN
CGNGCCNNNNGNNGNTNNNTTNTNTNNTNNNNNNNACNNNCGGGTGTCNTNNNTTTTTNT
TNTGGNTAGTNCNNNNGTGTNANNNTTANGNGNNAGNCANNGGANAGNNGTTGTANGNNG
ANNGTNGTNNNNNGGTNTTTTNGNTNNGNNAANATTAACNNGCCCNAANNGNGAAGANNN
GNNANANAGANNGGGNGTNNGTGTNGNANNTTNAANNNTGGTTNNTNAANANNNTNATAA
GANNNANNTNNNNNNNNNAATNANNTNNANATNNNNAGNANNAATNANGNAAAATAAGTA
GANNGATGTNTNCGGANGNTNGGANNNNAAGNNTNTNTCNNNNAANNTGANCGTNTNATN
CAAGTANANAGNGNNNNNGGGNAAGNNGGTNGGAANNNNATNNNNTCNNNANNNNGCNGT
NNGNGNTNANCATAGANNNNAGNGACNTGNAANANGGTANCNNTAGNNNNAGNCGGNNAN
AANNNNCCAGNNTTNTGCNCNGGNGTNNNTTTANTGNTTGGCTTANCNTCCTACNTCTTT
TNTTTACGNGANGNGCCNNANNTNCGNTCGGNNTTTAANNTTTNTNTTNANGNCNNTNNN
ATNNNNANANATANNATAANNCGGNNANGNGTTANTANGTTNTAAANATACNNNACANGN
ANAAAAGANTNGTNNAGNCGTGCTAAGAANNNGATTTNNGCNNNNNNNGNNTNATTTANT
CNNCCNTCAATNTNTNNNCCNGNCNCGCNCNTNGNCNNCGATATTTNNTAATNTTGNNNA
CNCNNCNTCTNNATTAAAANNACCCTCCGGNCCNTATTNTNCNTTNTANGNGGNAATNNG
GCNNNANNNAGAGCGNNGGCGAACNCGTAANANGNNNGAANNNAANNCCNTNTNAANNNG
NANACACAGCGATTANNTNTNTNNATNTNGTTCCGACGNTNNNCNNGACCCGNNANANNN
NGANTCGAAGATGNAAACNNTANAANTCNANCNCCGNNCATANNATGCANCNAANATTNN
CNNNNGANAGAANNAGNNGNACANAAAAGANGCTNNGAAAGAAAAANAANNAAATCAGNC
NCNCNNNAAAAAATTATTNACNCNNNCGNTANAANTNTCANNNGNNANAGTAAAGTNTAG
ATNGTNNNGTAGGNTCATGTANNTGGAGG
>A268
GGGGGAAANACANNNNCGTNGTGCGNAACGCGGNAAAACNGGGAANNAAATAATNGGATA
TNATNGAGGNGAAGAAAANNNNGCCCCTNTAAAANNNGNTANNNGAGCGACNGNNCNNNA
NANTAGTAGNGAANTAAACGNGNCCTNNNNANNNNNNNNCCCNCCANNTTNNNATGTNNC
CNNCNANCNANGCNCGTNCCCNAANNACCTGTTTNTANCCNCCNNCCNNTCCNTCNTTNN
TNNNNGTNNNAGNCNNGGCCNNAGANGTNTTNGATTATNANNNNANTTNTNGNTNGNCNG

CCNCNNNNNGNANATCTNNCANNNCCTNNTNACCTNNCCNGNTNAAGTCTNCNNTTNTGN
NCNCNGTTNNAANNNANNCANACAANTTNNTGCNNNNTNANTANANTNNNNGNTGCANGA
ATNGGAGGNCCNCGNNNAGTTTCNGGNAGTNTNNGCGAACNTTATNTGAAAGCNNNNCCN
NNGCNNNNNNNCCTNNAGANANGNNTTNNGTNNGNTNTANNNNNTGCNNNTATCNGNNNG
NNNNANNTATNNAGCGNNNNACCANNNNAAGCTTTTTTNTNTNANGNGACCGANGNATAT
ANNGTTNTNTNGNNGGNNNCTNTNGCCNNGGGCTTCNNTNNNNTNGGGNNNCTNGGANAN
NNNNTATNAGTNNANACNTGANNGATTNNNNNTNTNNANTNTCNNNAGNCCANANTNGAN
GGNACAACGNCNNGNTGNTANGGGANGGNGNANNNNNGAGAAGTGNNNANCTGNNNGNTN
NAGNTGGTATANGTACNGTATGCNNTTTGTCNGNATNNNCNTNTCTAANCTNNTNNANNN
NAGNANTNGGAANTNANTGGGGNGNGAAATTTANNTNGNNNNTGCTACNNNAANANANAA
NANGNAANTNANNTTNNNNAANTTNNTNTANANGNANGCACAGAANNTANNNAGTGGNTT
TNGATGNGNNNNNTANANTCANTANNNGCNCCNNNGTNNANCNACGGNGAGNNNAAAGGT
NNCNAGTNNGANAAANGNAGNNAATTNNNNCNNANACGTTNNANCCTTNCTTNTANNCNC
CGNCNNTNNTTANTTNNTAANTNNNANNCACTNNANNTTTTTTAANCNTGTANANANCNA
NNNTAGCTNNNGNATTTNTNTATATGTTGTTNNCGCGCANNTGNGTANNNTNGNAATAAN
ANGGANCNANNNGCNGNACTTNTTNTGATATNTTTNNTACGTGCNNCGANNGCGANNNNC
NANCNNTCTCTGNNANCAGNCNTANTTTNGTCTCCCNCTCTTTTGTTNTANCGNCCNANN
TTTNNNTAGCCGNNGCNGNNNCCNANNNNCNTAGNTNTNGTAANAATNCNNNNCCNCCCN
GTTNATAANATNGNNTNCNCGCNNNCGGNNNTNGNNGANGNCNGATNGAAATNNGNNNCT
CTGNTGNNNTTTGGCNANGNNGNAACANANANNCNATTAAACNAANGNTCNNAACGANNN
NNTANTNNTNATAGNNANATNAANAATATGAGNTNNCNCATNTCCNATNNNGCTTNGTGN
AANNCANGNNTNANTNANNGNGTNGANAANNNNCGNGCGNCGGCNATANGGACNCACNAN
NTGATANTATANACANGCTNNANTNNCNGTANANTNTANNTCNCNNNNAATNNATGTATA
ANNTNCACTANCGCNNACCNTNCCNGNAAANGNNGNCNNNGNNCATTAANNTNCNANNCN
NGNGGATTAANNTTCTTATNNNNACCNNATGNNNAAANNAAGGCNG
>A269
NGNGCNNCTTNCCATTGCNAGNTCGAAACGNGCTAACGACGGGGGACAACCGTGGCCGAC
GANGTGGNGAACGGTGTGGAGATAATGCTATCNNGAACGCTGCCCAGATTGTGGGGGATA
ACTTGCCTCGAANGAGCANACTAATACCGCATACNACCATGTANAGGTGAAAGCGGGGGA
TCGCAAGACCTCTNGCANATTGNAGCGGCCNATATCANATTAGGTANTTGGTGGGGTAAA
NGCTCACCAAGCCNACNATCTGTAGCTGGTCTNANANGACGACCAGCCNCACTGGGACTG
ANACACGGCCCAAACTCCTACNGGAGGCAGCAGTGGNGAATTTTGNANAATGGGGGCAAC
CCTGATCCNTCCATGCCGCGTGCGNNAANAANGCCTTCGGGTTGTAAACCNCTTTTGTCA
NGNAANAAAAGACTCCTGTTAATACCGGGGGTTCATGACNGTACCTGAANAATAANCACC
GGCTAACTACGTGCCANCAGCCGCGGTAATACGTAGGGTGCAAGCGTTAATCGNAATTAC
TGGNCGTAAAGCGTGCNCAGGCNGTTATGCAAGACANATGTNAAATCCCCGGGCTCAACC
TGNGAACTGCATTTGTGACTGCNTGGCTAGAGTACGGCANANGGGGATGGAATTCCNCGT
GTANCAGTGAAATGCNTANATATNCGGANGAACACCNATGGCGAAGGCAATCCCCTGGGC
CTGTACTGACGCTCATGCACNAAAGCGTGGGGAGCAAACAGGATTANATACCCTGGTAGT
CCACGCCCTAAACGATGTCAACTGGTTGTTGGGAGGGTTTCTTCTCANTAACGTACCTAA
CGCGTGAAGTTGACCNCCTGGGGAGTACGGNCCCCAAGGTTNAAACTCAAAGGAATTGAC
GGGGGACCCNCACAANCNGTGGATGATGTGGTTTAATTCATGCACGCNNAAAACCCTTAC
CTACCCTTGACATTCCNNGAATCTTGCANAANTTTTAAAANTNCTCNAANAAAACCTGGA
NNNANGTNCTGCNTGNCCNTCNTNNCTTTTNTNTGANATGTTNGGTTAANTNCCCAACNA
NCCNACCCTTTTNNTTNTTNTANAAAGGNANNTNATNAANTCCNGTNNAANCCNANGAAG
GGGGNNNANTCNGGTCNNTTGNCCNTTNGGNNGGNTNCCNNTNTTTAATGCCGGNAAAAG
NTTCNNCCCNNGGGNNTTTTCCNAANCNNNTTTTTCN
>A270
AGCGCGCTACCATGCAGTCGAGCGCCCCGCANGGNNAGCGGCAGACGGGTGAGTAACGCG
TGGGAATCTACCCATCTCTACGGAATAACTCAGGGAAACTTGTGCTAATACCGTATACGC
CCTTCGGGGGAAAGATTTATCGGAGATGGATGAGCCCGCGTTGGATTANCTAGTTGGTGG
GGTAAAGGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCANCCACATTG
GGACTGANACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGG
CGCAAGCCTGATCCAGCCATGCCNCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTCTT

TCACCGGAGAAGATAATGACGGTNTCCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCA
GCCGCGGTAATACGAANGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTA
GGCGGATCGATCAGTCAGGGGTGAAATCCCAGAGCTCAACTCTGGAACTGCCTTTGATAC
TGTCGATCTAGAGTATGGAAGAGGTGAGTGGAATTCCNAGTGTANAGGTGAAATTCGTAG
ATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTNCATTACTGACNCTGAGGTG
CGAAAGCGTGNGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAA
TGTTAGCCGTCNGCAAGTTTACTTGTCGGTGGCGCAGCTAACNCATTAAACATTCCGNCT
GGNGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACNGGNGCCCGCACANNCGGTG
GAGCATGTGGTTTAATTCNANGCAACNCGCAGAACCTTACCAGCTCTTGACATCCCGATC
NCGGACAGTGGANACATTNTCCTTCNNTTTAGGNTGNATCNGTGACAGGTGCTNCATGGC
TGTCTTCACCTCGTGTCNNGAGATGTTNGGGTTAANTCCCCNACCNAGNGCAACCCCTCC
CCTTTGTTGCCNNCNTTNANTTNGGCCCTCTAAGGGACTNCCGTGATAACCCAAANGNAA
GGGGGGATNANTNNANTCTTNTGGCCTTANGGTGGGNTACNNCTTNTNAATGGGNGGNAN
ANGGGNNNNAANCCCNNNNCANTTANTTTAAAACCTCCCNTTNGATNCCTTNNCNCCCGG
GCNGANTNNANCNTNNNNNCNANNNCNCCNGGAAA
>A271
NNANGGNAGCTTAAACATGCAAGTCGNACGATGAAGCCCGTTNGGGGTGGATTAGTGGCG
AACGGGTGAGTAACACGTGGGCAATCTGCCCTGCACTCTGGGACAAGCCCTGGAAACGGG
GTCTAATACCGGATACTGACCATCTTGGGCATCCTTGATGGTGGAAAGCTCCGGCGGTGC
AGGATGAGCCCGCGGCCTATCAGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGACG
GGTAGCCGGCCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCAGACNCCTA
CNGGAGGCANTTANTGGGGTAATATNGNACAATGNGCGAAAGCCTGATGCACNTTNCNCC
TTCTTTGNNGGATNACGGCNTTCNTNNTTNNAAACCCCTCTNCNNTTTTTTNACAANCCT
CNTTTNTTTTGTNNNTNCTTAANANNCTTNCCTGTTAACNACGNNGNCCTTCNTTTCTTC
NGTAATTCNTTTNTTTTNCTCANCCCTTTTNCCCTCCAAATTATTTTTTNCCTTNNATAT
TNTTTTATNTCCNNCTTTNTCTTCTTTCTNTNTNTTNANATCTNCTCTTTTTNNCTAANC
NCCTTNGTTCTCCTAATTTTNATNNNCTNTCNTNTNTCNNACCCTCNTTTTNNNGCCATN
ATTTTNTATNTTTNTTCTACNTCNNCCCCNCTNNTTTTNCCNTTTNTTCTTTTCTTTTNC
TTCANCNNCCNCTTGTCCCCNNCGNNNTTTTTTTCTTTTCTNCTATTCNCNNATTTCCCC
ATTCATANCTCCNNNCTTTCTCCANCCTCCTTTTNTTTTTTACTTTCCCCCCTCTTTTTA
TTTTCNNTTCCNCCTTTTTCCTNCCCNTTTGTTTTTTTNTTNNCNCTTNCTCTNNNTCNN
TTTCTNACTCNTCCTCGTNACCTTTTCTTTTTTNTTCTTNTTACTTTCCTNTTTNCTNNT
CTTCCTNTNTTTCTTNTNNNNTTTTTCTTCCNCTTTTTTTTTTNNCTCCTTNTTTTTCCC
CCCNCNCNTCTNACNCCNCTCTTTTATCTTTTNNCTCCCTNCTCTTNCTTTTTTTNCCCC
CCCCANCTTTCNNCTTTNTCNTCCTCCTCNCTNTTTTTNTCTNTTCTCNCCANCTTCTTT
NTTTTNCCCCCCTCCTNCCNCACGNNTANNCTTANNTTTNTTTTTTNTTTTNNCCCNCTC
CNCCNCNCCCTTTATTACNTTCCCNCNTTTNNNCTCTGTTTNTCNCCTCNCNATTNTCCA
TACNTTTTTTCCTTNCCNCCCTTTTNTCTNTTTTANCTTNCCCCNCTCCTNTTNNTATGT
CNCCNCCTCTCNNTNCCTTCTTNCCCNTACNNTTTTTTTCNNTTCCACAANTCTTNNATT
TATCCTCTTNCANNNNATTNTNCCCCC
>A272
GGGNGTAAAAAAGGAANGANGNTAANGNANNGGGANAANCGGAAAAGAAAATATNAGTNA
AAGTGNNCGTGTTGGTGNACGGNGCCCCCCCNTAAAAANNGNGGATTTCCGCGCGNNAGN
GNCANNNTNTANTCNTGNNNNAGCNTNNNTTNGTGTNNACATNNTNTAGTGTNTNNCNNN
NNCNCNNCCTGTTTNCNTTGNNTNNCNAATANTNATNTNNCNTNCGNCNAGCTGNNCCNG
TATNTGCNTNCANCNGCCCCGTNNCGNCACNTTTNNTACGCNNACCNNCNCGNCTGNGCG
NCNNANGCGTGTTAGCTTNACCACAGNGTTCTANNGATAAGNGNGCCGTACTNNNNGCGN
TAANATTACAANNTCNAANAGCAATNAGNAATTAANGTNTTNNGNTNNTACNAANANCGA
NGNNAGNAGANNAANGATNGGTGAAGTANGGAGNGNTAGATAGCGGGNGNGGAGGNANAA
NTAAGANNGNGNNCGGGNTTCAGTNGCCNNCTNTNNTANCGTGCTNGTATTGCTNTTANN
TNCAGGCCNCGCCNNNANAGCTNTTTAGTTNTTNCTTNTTNNNNNTNNATNTNNNNCCCT
CTTNNCCGNNTTNNTNGTTTNTNTTTNAGCNNNCGNCNANCNNANNGCTTNTTTTATNTA
TANCANNANACNNNNTGTNTNGATGTNATNTTTTGNNCNNNNTTNNANNNCGTNCTTNNA
NTATTCNTACNNNCCGNANGCGGCNTGNANNTTATNGGNNTCNCNNNTNNAGNNTNNGNT

TGTNTNTNTTGATTNNANNTNATTTCCTCCNNCNAANAGTTNNNTTTGNGATNTNNAGNT
TGNGTGNNGNGNGTNGACAGCNTATANTGNNNNTNNNGTATAANGANGANAGTGATTCAT
GGNNTNTNGTTGTTNGNTANTANGNGTAAANTAANANNTCAGNAGGCNANGANGGGNGGC
TCGTNTNTTCAGNNNCGTTNGAGNANNNANAGNGCATANTGTNNAANCNNNGANNAATNG
NATGNAATTAAATNCGTNNAAANGNGNTTGGANNGGGTNTAAAGCNATANNAGNATTAAC
CNTTAGCNNNTAANTGANTTTNAGTTTGNAGGAGAGTNANNNGNGTTANANNAANCNTNN
CGGNNAANATTTTTCGNNAGCAANNCTGTTNTTNTTNTNTANNCNNNNCGCACCGNNCTT
GTTTNTTATGCNNNTTNGNNNNTCTANNTNCCTANCCNNTANNTTTTTNTATNTATNANT
NGTNNGGGTNGTTNNNTNNNATTANTNGTCAGAACNANNNNTTTANNNTTTTNTTCTTTN
TAGTTNTNCTTTNNGTNTAGANCNNNAATNTNNGANTNTATNNTNNTTTNNGNNTNNNNC
NTTGTTTAATNNATANTTCNNTNTATTTTNAGCCANGNNCNCCNCANAAGANANNNGTTT
ATNTATATTNNAAACCCNNTNTCGTNNNTNNTANNNTAGGNCGNCNNNNTNGNTCTATTT
GNTNTACNGAAANANATGNGNNATTGCTTTNNCNNANCNTANGTCNNTNNTTNNACNGNN
ANNANTTNANGANAGCGANTNGNTANNNTATAAGNATTNNTTCANAGNGANAGNNNNNTN
ANTCGTTGGTTAGGGNGANNATNNATTNANGANACNAAAANNGGAAGGNGNGNAANNNAG
TCATGAGNNACANNTNNACNAAATATCGGCNTAATAANGTTCTCNNANANAAAACGAATA
NTNNGNTNGAANNNNANNNAAATNNNTGGTNGCNANANNTTTTANNNTNAGCCANAANAN
ATTNANTANANNCNNATGGNNNATATNNNTNCGTNCG
>A273
GAGGNGGNGGCCCTANACAGGCCAAGGNTCGTTAACAAGGGTAACCGATAANATTTNGTG
TGTNTNTATTGANGAGCNNNNAGNGANGNNCCCCNNNNAAANNANNATAANGGCGCGTGG
GNNAGCGCATNNANGNNNANNACGCCNATNNCGGGNCNNNANNNNNNNGTANTNGCGCCC
CGTTNANGGNTNCCCGNANNNNCCNNNCNTCNCGTNGTTGCCCCTTTACTNCNCCNGCNC
TNNCGNACTTNTNNCNAANCNNNGNCNNGTNNNGNGCGGATTNCATNNCNGCTAGNNGNT
NGNTGTCTCGNAANNTNTNTTGNNNTAANNNNGNGNCANNNTNAATNNCTCTNTCNTATT
TNGAAGNACCGGNNAACACTANCAACNGCNAAACGGCNGGGANTCNAAGGGNCCANGGCN
GNGANGNCNNNGCGGGNCGNTANTCGGNNNNNTTTAANAACCNCNNTTACCTNCNNTGNC
NGCCCGNGGCANNGCNTATTNNNNGNNNCNTNTTANNNTTTCNTTNNCGCCTCTNNATCC
CNNANTTNNTCATNNNNTTNCANCNNNCGCNCNGNNANNNCTTTNNTNTTTCNCCCNGCN
TCNCAGNCTGNGTGNTTNTNNTTNTNCNNNNNNNNCGNCCTCCTTNCCNTNTNNAANCCG
NACGNANAGNANNTNATNNNNTNCGTNCNNNNTGCNCCTNGGNNNTNNGGTTTNCAANTN
CNGNANATGNNNGNCTANNAAAAAGTNANAGTNNTNATTNGGTGANGNNACAGGAGTNNN
NTTCCTTCCTNNNGNCAGGNNNNTNNTANGGNTNNGNGGTNTANTNGTTNGTNNTGNTTN
TANACGNNNACANNNGNNNNANNATAGGTGANNGTNNNAGACCNAGCNGTGNNAGTCTNT
TGGGGGNNCANNATGGNNANTAANGANCGGGCGTNANNGGNTGNTNGNNNAAGNGATGGG
NTGNNAAGCNTTGTAGNCCCNCAGTTCGTCTNTNATTCNTTNTNANTNNNAGCGNNGNCC
TTTANGCCNATAGCNNTNTGGTNNNTAANNNCATNNTGANGAGGNCGNANNNNGNGGNNN
CCNCGCCNGTNNNNNNNGTNNNNTNTNCCGNGNNNGTGCTNTTTTTTTTNTTNGNGTCTCN
NNAGCGNNNNNNNGTTNGNTTNAANGCNNNCCGGNNNGTNANTCGNNNNGNANGTTNTTT
ATTTNTTTTTTNANNNTNCNNCNTGNNNNTTNNTNNCGTTNAATCTNCTGNANCGNTAGA
CGCATTGNTAGNNTTNTATTTATTTAAAGCNNNNNGAAGNNNNGTCANNTCNCGTCGGNN
TNNAACNCATATTATCNCCNNNNCTNTATGTTNATACCCNTTNCTNATTTGAATNCCGAC
GNNGACNCNNGCGGCCCGCTNTNNANNNNNTACNCNCNCCGCNNCNNTNTTNNTNTANNA
NNNCGCACNCNNCCNNNTNGNNNTCNGTCNCGTCANNGNCNCATNNTTNTNCTTCNCTAC
AGANAAACGNNTNTNCAANTGTNGGCNNNNNAAGGCGNAGGTANTNNTCTNNNNNTGNTN
ATAANNTTNCGCCCGAGGNNTGCNCAGAANCNNTGGNTNNGTNAAGTNNTAANAGTCNNN
NCGNNAATNNCANGNGNGCTGNAGNNGCCNACNGNGANGTNNGNGANATTNGNGNNNANA
NNNNCCANGCNNNNGAGAGATNANNAGTCGATNCGCTCNNGNNNNGNGANAANCGACATN
CGAANTGATNNNNNGGCNGGCANATAATGGNTGTNTCNNGCTANGNCNATNTANATAGGA
GCG
>A274
GNGGGGGGNCCANNAGGCAAGGTCGTAACAAGGGTAAACCGTAAAATTTNNNTTTTTGGN
NGGTAACNACTGANCCCTTTNAAANNNNNTNANNGCANAGNNNANATATNCNGNNAGCCC
NTNNGGNGGNCCANTGNTNTTCTNCCTNNNNNATCGTNTCNNGNGNNTANCNCGCNCNNA

CGTCNTNTNNTCCCCNNTNCNTCNNNNCTTNNNCCCCCCNNGNNNCNNNCGGNGGNAAAC
NNTACGNATAGNATAGGNANNNTCGNAGNANCGCNTNAATTNNCGAGGNANCACTTATCA
TTATACAGNTNNTTTANAACNNGNANNAAAANNGNACGGTGANGAGANGANNTGAGAAAG
CGTCCCNCNAANNCNNCACGCNAGCNGGGNNNTTAGGNGNCGNNTNACNNNAGNATNNGT
NCNGTNNCNCNGCCCNNNNGCNTNNNNACNGTNGCGTNNNTTNNATNCNTTANNNNNNCN
TNGNCGNTGTNTNNTGTNTNNTNCNANNCCNCGNCCACCNNCCCTTTATTTATTANNCNA
CNCNNGGGAGANTNTTTNATTTATTANNGTNNNNCCGNNGNTGATGTANTNTTTAAGACN
AANNNNAGAGAGNAGTNGTAGCTTACANATAANCNTGNGNCNTTNGTNNTGNCGNNCANG
ANATNNAGGGNCGGAAANTNGNGATTGATAANNTNNNTNATGGNGTATTTNGNTGNNAGC
AGNNTGGGNNGNTNATNAANGAAGNTNANGAGTTNCTNTNTTATGTTACCGNANNANNGA
ANANNANGNAGGNATNGAAAANAGANGNGGNGAGNGTNNGTNCACNCNAGNCNTANTGGG
NAGNNNGANTNTGNNTANTAATNCNNCANGACANNANGNNNNGNNNGGCNANTNTNTANC
NATTNTNGNNNTCNCGCGNGACANANNNANAGNTCNGNCAANATTNNNNGGNTAGTGAGT
ATGANNGNNNGNNTNNATACNAGAANGAANGAGGCANAACGTATGNCGAAGGANNNCTTN
TTTTTTNTTTCNCNATGCACACCNNNNCNGNTNATCTTAAGCNNANGNANNGATNANCTG
NCGNGNNCGTTNTATANTNAATATGTNNCNGCANTCGCTCNCATNNCNCNNANAANNGNC
NANNNGTCNGGCNGNNNTATTTNATNNTNNNNANNGNCGAANGGNNGCGNNAGANGANNC
NNNGGNNNTNGAGNNNCTANTTCCCNCNANAGTTNTTNATTANNGNNNAAGATANTTCCC
NNNCGCCNCAGANACCNAAAAGTNNNNNTTAAANCANTGNCGANTNTNNNATTNNNNCCC
CCNTGNNNATNCNNNNTTNGNCAGANAAAAGCACANNNNNANGNTNGNNNAACGCANNAN
NTTACCTTNNACTCNNNGTNTAGCNNGNGGANTNCNANNANATANTGNNTANGANCCAGG
GAGCGNCANGNANNTNNNTCCGNANGNNANTANTANNNNTNAANATNNACGCGNNNGANA
CNTGAGNNCNNNTGANGANANATNATAACNANNGCCTGNATNNNNNCNNNGCCGAAGANA
NGTNNNNCNAACANNNNCNNNNCNNANAANTNNNANNGAANCNCGGCNNCANTATNNNAC
NCTNGGGNATAGNTGGNANAGNNTGTCNGNNNCNAATAANANANCGG
>A275
NTNGGNNNGGNGCGGCANACACAANGCNAGNTCGTTAACAAGGGTTAACCGTTNATATTT
NGTCTANGTGTNCGGNACGNAAACCNGCCCCANGTAGCCNTGNTACGNGCCNTGCCGATN
AAGGGACGAANCATNTGGGGGNACCGTNATGATNCCNCTNTNCGCNTGNNGAGNNNCCGG
NNCGAAAGCTGGCGNCTNGGTCNNAACNTATTNACGAGGAGACCGCAGCCTATCAGNTAN
NNNGGNTAGGTNATGGCTCNTCCNTAGNCNACCNCACGTNGCNGGNTTNTNNNTGACNAT
CNGNCCNCCTGTNNGGGNGANAAGANACGGACANTTACGCNAGGGNNCGCGGNTGNATAG
NATATNCGGCGCGNGAANNCTGAATNNNNTCCNGCNGTNTNNCGCNTTTGAGNCTTNTGN
TNANGACTGCNNNNCTTTTTTAANCAACCAGNNNTTTTNTTTNNNTCNNGNCNCAGTTAN
CANTNNNNCNNTCCNATCNNNTNCTCGCNNNCNTNNNNTNTTGNACCNCCCTTANCCCGA
ATTNTTNNNCGGTGNNNNNNNACTCNGNGGNNNTGTNNCGGNTGCTNTGTCNNANNNCCC
TGGGCCNTNGCTNCGGGNCCNCNGGATNGNTACTTNNCNTTNTANCGNTTTGGTACNNNN
TNGNTTNTNTATTTCTGCTTNAANCNNTTATAANNNNNTTATNTNTNTGNNGNNNACCTT
NNCCCCTANNTGNNNNNATGNNTTCGAGANANNNTAGCCGACNATCCCCANGTTTGNCCC
ACCCTTTTNTTTATANTTNCCCCCTCTCTTTTTTCCCTNCNNNNTTCCTCNGGTNCCTNA
TNTTTTTTTNTCCNCATTCNTTNCTCNNNTTNTGGTCGNTCCGTCANCTTTTTTTTTTNTT
CTTCTCTNCATANNNNACGCCTCCNNGGNATCGANCCTTTNANCTGNNTTTTTTTTTTAC
CCNTCATTTNTTTCGCCCCCCCCCCTTCCCCCTCCNTTTTTNTTTNNNTCCTCTNNCTTT
TTNTTTTCCCCNCCCCCGTTTTCCTCTNCTNCTCNTTNTCNCCTTGGTTNTTCTTTCNAC
TCNGTNCNCCTTTNTTTTNTCTNCNCTANCNCCCTNNACCCNTTCATTTTTTTCTTAATC
NTTCCTCCNCTCCNCCCCNTNCCTTCCCCGCTNATTTTNNCNCGCTGNATNNCTCCCNNC
TNCTNGCTCTTCNCTTTTTNNTNNGGCCCCNTTNNNTTTTTTNNTCCCTCNCCNNCTNTT
NNNGTTTTTNTCTTNNNNTCCCCNTCTCCATGTCTNTCNCTCCCTTCTTTCTNNNGTCNN
CNTACTTTCTTTNTANCCTCCCTCCTTTTTTNCNNTCNCGCNC
>A276
GANAGGGAGGGCANACANAGGCAAGGTCGGTAACAAGGGTTAACCGGTAAAATTATCCNT
TTTTGNNNGCNGCAGNAATACGCCCCCTTTNGGGNGGNNGNTACANACNGNCCNTGNTCA
NNGNNCACNTAAAGNCCNGGTNNACNANNCGTNAGCCCNTGTTGTGCNCAACANTTNTCG
CGCNGAGANGAGCTTNNCCTCAACCGCGGGTACTGGTTGATTGGCNACNAAGNNCCCNCN

NANGGTANCCGGTATTNTTTCCGGTTTTNANGTNNTNNCANNANANCCNTNGNTTANACT
NCNTTTNTNANNNNNNNNANCNANNCACAATATAGGCNGNNGNGNAGGAGGGGGGNGNGA
NNGGNGCGNGGTGTNTGCNCGGNTNNCGNCANNGNNCGAANNCCGGGGAACTCTNGCTTT
TGCNGNCCNNNTNNNCTTNAATNNNNNCAANNNNANCCCTNGNGNNTTANTNACNCNCCG
NCCGNNCNNTNTNTTTTNTNANCNNANGCANNCNTTGCTNGANNNTNNTTAANNNNATCC
NNNACGTTNCGTTNNNCNNGCNGCTCCGCCNTGTNNTNTNTNACATNTNNANNNGGTTNG
TTTTTNATNNCCAGGTNAGCANCCANNANGTGTNNNGNGNGATNNANNAGGTCNNNGNNG
GGGCNNNTGCGGNNNATNNGAGNCNNANCTNNATNAGTNTNNTNNTGNANTNGACTCAAN
ANGNCGGAAAGNANAATNANNNNNNNATNGNNNNGGNNGTNTGGNNNANNTGAGANTNNG
NNNAGNGNTNGNNNTGNNNTANNNNNNGTGANNANANNANANGNGCNTNGGNNTGATNAN
TANNNANNGNNNNNNNNNTNGNNGNGGNCNNATNTNTGTTNNCTNTNNTNGATCTNTANN
CGNNATTGNANNNNGCNTTNAGAAGNGNNNTNNTGNCNGNNTNNNNTNNTGNTGTNTTTN
NCTTGCCANGTNTCTATGTTTTTANACCNNNNNGCCAANANGGCGNACNCCNTTAATNTT
NNTTTTATGCGNNNACTNNANTNCCNCNANNNNNTACTNNNNCNNACCNTTCTTTTTTNT
ATNNTTNACNACGCANNAAACAANNNNGAANCCTCNNGGNNTTAGANANNNNTTACNCNC
NNCCGCTTTNTTTTNANNNCANNTNNNANTTNTCNNCCANNCCNCNCTNNGCCCNANNNA
TNNTTNTCNNTTNNGAACGCCNNNCGTNNATNNNTAANNCNCNNNNNACGCNNATNCTNG
NTCNCNGCNNTCACCAGGCNTNTNCTTNCNCTTNCNNNCNNANACCAATATNCACTNTAC
NNANGNGCTAGCNNCTCNCANAGNNCTNTANATATAATANNNAATGTNAGCCNNANTNGN
ANNNGCANTNTNNNNGNATNGCANCNCNTAAANNNNCNANNNGNTACGNGCANNNNGATA
GTAGNNCNNNANAANNTANNNTNNANGTTNGCGTNANTTNNTNCCANGNNNNNNNAANAA
NNTATNNTNNTGTGACGCCNTGCACCANNCTTAGNNTTNGTAATNACACCNNTATATTAT
AANANCNGCGTCANNNNGNACTNNNAACGTGTNGNNGTACAAGANNAGCNGTTCCACAC
>A277
AGGNGGCGGGCCTACACAAGGCAAGNTCGGTAACAAGGGATAACCGNTTATATTGNGCGG
NTTNTGTNNNNNGNAANNCGAGNGNGNCCCCNTTTNAANGNGGNCANGNCNTGCGGCAGC
CANCTNNNNGTGACAANTTTCNNGGCCAATNNGGGGGNACTCCNCCCCCTCGCATTNCNG
AANTNAATCCTCGGCGCCCCCGNCCNANTCGTGNCCCACCNNNNNCCACNTNTGCCCCCC
CGCNNGANGCCNCCNNGCGAACCCATACAGCGTGTGTNTNTNCTGNGCTNNTTTGNGTTT
CNTCAANGNCNTCACTTACNNNNNCAGACNNGCATTGNAGAACNNNTTGANCNGGTCAAC
ANTANGGGNNNGAGNACAGACGNCGGAAAAAGCCGGACNNGGNGGTGGCNNGTGTGAGGA
ATANAGTTAGNCCGCGGNGNNNATTCNTTTGCTGNCCTNCTTNCNNTTNTNNAGAANNNA
ANCNCGGTTCTTCNTNCCCCNCCAANCCTCNTGTTNTTTAANNTNACNTCACGTGCNAAN
CTNANTTGNCTCCTCNCNCCTATGANTTTTNCCGNNNNACNNNCGNNGCTNATNTNNCCT
GGCNGNNATTGNTNTGATNCANNNNTNTGCCNGANANCTTGNTNANTTGATNGGGGGNNN
CNNNAGNANNCCTCNATNNNNGGNTNTTTNNGNCNNNATTNTTTNAGNNTANTTNATNAT
TNCNCNGNATNTNTCNNANCANCACTANGTTTTNTANNANTTNNNNTNNNNGNNNNNGNN
CTTNTANNATCATANNNTNNNNCTGNNGTAANNTNTANAANAAATNTCNCAACNGTTTGC
NCTCNNNNANNNGCNTATTATCNTTGANTNANTCANAGGNNNTNNCCANCCGANCNTGTN
AGGNNTGNNGTNNATANANTTNTNTGNGGNTTTTGACANCTNNTATNNCANNNAGGGTTT
TCTCTNNNNGTCCTTTTTTTTTNTNTGCNNTTCACNCNNCACTTTNTTTANCNATNTNCCN
CTTTNGACCTNCNNNTNTCNTTTTGNTTATATAACNNTNNNNNTCNCTNCNNNTTANATC
TTCNNNNCNTNNNNNTTTGCTNNTATATATTTTATCNTNANNNGTTNCCAANNNTNNGNG
TTNCNCNTGCNCACCCGTTTGTNGCNTCCCATCNTCTTTTNNANCCGCATCNTNTTNTNT
TCACNTCNCGCCCCCTATCNCACCNCTTNTNAGGNCTTGTNTNCACCCCCTCGCTTTNNT
NNTNTTCNCCNCNNNACNNTTTTNCNATTTNTCNNNNNCNGNAGNNCNTGNTTTCAAGTN
NTCCNCNNGCNTNNNCTNTATNTTTAGCNNTNCGCNNANNANNCTNATNGNACCGNNTNT
TTNGCTANNNTGTNATGNNNTCANCCNCCTNGCATCTTGNCCTGCNNNCTCNTNTATTCC
NNCTCTTNNNCNNCNNCCCNCCTTTCGTGCCNGNCNANAAANNANTNCCNNCANNTNNTA
NGTATNTANTNTTACNCCCNANGNTNNTTNTATNNTCGNCGNCNANNCNNTTTNNNGNNC
NNCCTATCTNGNNTTNNNTNCGTAAAATTTTNNNANCCNACGCCTGTTTTNNNATNNNCC
G
>A278
GNGCNNAAAAANGNGNAGTTTTTNGCAGNNANGGAAANAGGAGAAAAANNATNAAGNGAA

NAGTGTATGACGTNAAATGCGGAAGANCCCCNTTNNAAAGGNTATTCANCNCTNCGCGNT
ACTANTNATNNGCGCGNAANNNNACNTGGNGGTNNGCAACNCTTTGNNAANTNNTGNCCN
CCCGTNNTGNTNNNCCCCGCGAATATNCCCGCGNCNACNNANNTNNCNNCTTTGNNTNCN
NNAGCCNCNNNCNCANATTANTCAGTTCCGCNGAGCGAGAGGNNNANGGGAGGTGTNACC
TANTCNNNCAGGGATNAGNTNGCCNTGCCNNTCCNGGNNCNGTGTAGTNTGGAACCNNNT
NNNTATNNCAATTNNAGTACGNTATGTTTTGNAANNANGGGNTATCNATNGNNCNANAAC
CNATGGAGCNGNGNCNGGCTATCATGAGNCCANGNNGAAGTNAANCGGGNNNNGNCGAGT
CTNTCNGNGNGNTTTTTANCNNNCANCTNTNCNGTCTNTATGCAGNNCGCNGGACGNANT
ANGNGTTGNGNATTGTTCTNNNTAGNNGCCCCCTANTTCTNNCCNTATTTGTNTTTNTAN
CNGCGNNNTNNNNNCCNNCNNTTTNTGTNTTANCCNCAACCCCTGTNANNNATTNTTNTN
TTNTNTCNNNNTNACCNCCNGNTTNTNNTTTTTCACNGNGACNGGCANTGNNTTTTTTTG
NNTANNATTTNCNTGTATNTGTTGTTNTNANNNNNNNGNTGGNGNCGGNNNNNANTNNNN
NNGNNCTNNGNGAGTNNGNGNGANTNNGCNACCNGGTNNATGNNGNTCGNAGGTNANNTT
GNATNNGTNNGTATNNTNGTTANCNANCNTNNNGNGNTNCCNNCNTNACNTATTTNNGNT
GNGTGGANNTGCTTNNANNTCANTTATAATNAAGNTNNNAGNTTTTTTANNTNAATCNAN
ANNNATNNNGANCNGNCTGNTTCNNNNCNTGNNTGTNTTTTNGTTNTGCTTGNTCCACNA
CCCTTNNNNNCGTNCGAANANNATGNGAGCACTGNGNNNGTGNGGNNATNCNGTTCCNNN
ANNNCGNNGNCNATNTTNNTTTTTNNNNCTNNCCGTATATGTTNNATCCTTATCNCANCN
TNNNAATATTCTTATTCNNGGTGGAGNTNTNTNCTNNNTTNNCNNNNGTTTTTNTTNNNT
TTGNACANCNNNTNNCNGTNTNGCANNCNNTATCCGTNNGCCANCTANANCACTATCNTT
TTATTTAGNNTTTNCNTTGNGAAGTAAGNGNTCNATANTACCTTTCNTTNTNNNCCCNGN
NTNTTNNNNTNCANTNNGTTTNTTATTACNCNTNCTNCNTNNTTGTNAGNNNNNNTGANA
GANNNANCNNCCTNNTNTANCNNTNGNTTNAANTNCGCTCTNNNTTGTATTNNNTNNNNG
CCANCCNANNCTCTATNCTCNTTNTTNTNNCNACNTNCNNTNGTTTTTTGGTTNTGCNAN
CNCCNNNTNNNTGGNCNNTGAGNGNCCCNTAGACNNCGCANNTTCNGCNNTNAGNNTNTT
TANNTNNNTNTCCGCANTATGCTNNCNACCNNNNNTTTTNTTCNCNNGNTGCATANTNAT
NTGCTGCATNTNCCTNACNNNNNCNNGNTNNGTNNTANTNTANNANNATATNCANTGGAC
ACNNNGACTNGNNGTNNNGGNNANGTATNTGTACCTNNTNNAAAAGTCANNATGGTANTC
TTCGGCGCGCCGCNANAACNNAGATTANNTNGACNANNCCCATTGCTANNNNNATGNGTC
CCGCNCCNGANTNNNTATNNTTGNATNCNCANGNTNNTCTGANGNCNTATATNGTCGCNT
GN
>A279
GGNANAATAAGGGNAANGTTTTAAGACNNCGGNAANGGGGNNAAAGATTNNATNAANNGT
GTTGCNGTTNCGCNNNANGNNCCCANNTAAANNGNATNNCGNTGCGNCAANNAANNTTCN
GNGNNAANNNCTTNNGNCGNCNCNACTTCNGTATTANGNNCACCGNNNTTACTGGGCNCC
NAATGACNCGCGTCNACNNGNCGTCNCGATACTCCCNCCGCACTNCNNNNTTNTANNCGN
NCGCCNGCNNNNCANNCGNNNAACCTANCGNCAGAGTAAGNATGNNGNCTNCGGNCNATN
AATAAGGNAGNAAANATNAACATAATATNNTTTNTNGTAAACGANNANAANAAACAANAN
AGNGCGGGANTAANGNGNNGGGGAAAGAGNANCCGCCGCTCTTNCGNGNGTTNNCNGAGN
TANTTCCNCCCGCNGCNGANCNTNGNTGTNTCNTGTTNATATCNNGNNCCTNTNNCCTNN
TTNNAGANTTAACNNANCGNCTCTGNCCTTTTNTGTAGCCNCNNNNNTTAANTTTTTNTT
NTTANCTTATNCCCCCNCTTCNANNNTNAANNCGNNCGGNGNGGTTGTGNNTCGTNNNNT
ACNCNTTNGNTTTNTNTNGCATCNCNAANTGTCGCNGGNAGATTATTTNGAACCNGAGNN
NTGCGNANNANATANACGGNNNTCNNNNNNAANANAGAGNNTGANNTTTCCTNNGTGTTG
NNNNNCNNNTNTGANGGACNNGGCNNNNGGGNCTNAANNNNNNNNTNATTGNGNGTANGN
TNNGGNTGGGGANNGNGTNNGTNNANANTNNATGNGCNTNNNNCCACNTNCNTNNNTNTT
AGTGTTNTATGATTTNTTGNTANCAANNANTGTGACAGNANNANNAGTTNTGAGTNTANT
CNGANNGNTNAGNGGGNGGTNAACCGNANCNATNNGGNAAGTCTTTTTNACCNCGNNNNN
TTTNTGTNNNTNANTNNCCCNACNNCGNNTNNTTTTTATNNTNNNCGNCTTTCNNNCGTN
CGNNCTTTTNTNGTTGTTTNNCNCCNNTNAGTGNTCTCGCGNNNNTCATNTGNNACGNGT
NNAANGTCTCGTTTNTTTNTTNNTTTNCTTCNCTCNAAATNTNAANCAANCNCTNCGGGA
TGNCTNNNNTNTGNGCNTCNNCNNTTNTTATANCCNANTATNTATCNTTANNCCNCCNTN
ANNNCNCNCCCNCTCTTATANNCNTNNANTNNCCNCTCNCNTNANNATTTNNNNCCGCCT
CNNTTTTNNTCNATTNNNCGCNCNNANNCCNNNCCTANGNGTNNNGCANCNNCTNANCNN

TANATTNNNNNNCNCNNANNANNTCNGAANNACNNNTATTCATTAANNANTNNNTCATGN
ACNCNNCNCNNNNCAANGTNCGTTNNCTCCTNCANCNANNGNNNCGNAAGANNNTGNAGN
CNGGGNTAGGACNTNNAATAGANAAANGNTNCCAATANCNNTATTTTAATGCNANCCCGT
NTAAANNCTNNNACGNTCTANCCNTNNCANNNANNGTNNATNANNNGAACCNNATNTANN
ANNACNCCGCANTANANANGNAAGNNANNNAAGCAAAANAGNCANNNGN
>A280
ANGCGCGGCCTACCATGCAAGNTCGAAACNAGGGTAACCGNTAAAGAGGGGCACGACGGG
NGAGANAACGCGGGCCCTTTNTGNNNTTCTACTTCNGTNACAACACCGGGAACAACGTTG
CGNCATTNTNCCNTGATACCCTNCCGAGGAGGANAAANATTTNTCGCNNAGAGANCNCCC
CGCGNCAGANTNNCTAGGNGGGGAGGAAAGGNCTCACCNNNGNCACTANCANTANNNGGT
CNGAGANGANGATCNNCCANNNTGNNACNGANACACGGCCCACACNCCNACGGGAGGCAN
CAGNGGNNAATANTNNACANNGGGCNCNNGCCTGANCCNCCCNTGCCGCGNGAGNGANNA
NNGCCNTNGGGNTGNAAANCTCTNTCCNNCGNNAANATNANNACTGNCCCGNAANAAAAN
CCCCCNGNTANCNTCGNNCCNNCNCCCGCGNANANACAAAGGGNGNTNGCNTTGCTCANA
ANCNCTGNGNGNANAGNGCGCNTNCGCNNATTTNTAANTCGGGGGANAAANCCNGGNNCT
CNNCCNCANAAGTGNCTTNNACNCTGTGTNTCTNGANNCCNNGAGANGNNAGTGNAACNG
CGAGTGTAGAGGNNAAANTCNTANATATTCANNAGAACACCAGTGGNNANCGNNNCTCAN
TGNCCCGNCTCNNACNCNGAGGCGNAAAAGCGTGGNGAGCANACNNNATTANANNCCCTG
GTNNNCCACCCCNTANACTATNAATGCCACCCNTTGGNNNNCTNNCNGNTCAGTGGCNCA
TCTAACNCTANAAACTANNCCCTCTGGNGAGTACTNTCCCNANATTAANNCTCANNAGNA
TTGACGGGCCCCCACACANGCNGTNGANTATGTGNTTTNTTTCANANCNACTCNCANAAN
CTNNCCACCNTTTGACCNNTCCCGNANNNACCCGNGAAANCTGNCTTCTTNCATTNCCGC
NNNCGNGAAANAAGGGCNCCAACNGGGGTNTCCNCCNCTTCGGNGTNGAGAAGNTNTTGG
TNNAAACNCCCAACAAAGGGCAAANCCCCNCCCCTNTTAGNCCCCATTTTTNGGGGCCCC
TNANGGGACCCCNNNNTAAATCCCTNAANAAANNGGGGGNTANCCCCCACTCCCCTCGNC
NTCNNNGGGGGGNTCCCNCCGNTNCNATNNNGGNNCACTNGNCCNAANCCCNNNNNNATA
TTTCCCAAACCCCTCTCCTGATNTCCCTCTCCCTNTCTNTAAGGGGNNTTTTATATNAAT
NTNACCCCCGTAAATCTCCNGGTCTTNACCCCC
>A281
GAGGNGGAGGCTAAAATGGCAAGGTCGTTAACAAGGGTAACCGAAAGATANTNTTTGAAA
GGGTGNTGTANACATNCAAANGCCANTNANNANNANANNTCNCCNGCNACGAACNCNNTG
NANNNNAANNACCCTATNCNNGCCNNNACTCGATNTTAATCNCNCNCNATNNCGAGNGCC
CNCGNNTNTCGCNCGNCNCCGNCTTNNCCTNTACNCNNNCGNNCGNANCGGNGCNTATTN
GCGCNCGGNCGNCNGGCACNNNCNTNNGTNTNNNNGNAGNNNNNAANNANNNNNCGTCNN
NNGGTTNTNGTACNGNTCANANANNTNAGNNTATTANTCNANTTTTNNNAAGAGCNGGAA
CANGNNCACGNNNNTNAAGCNGCTNANCAGANANGGNAGNAAANTAGAGNGCNGCNGGTT
TCGGCCNTTTNCNNAGCTTNNATTCNTTNCCGCACGNNNTNTCNTTNCCTTNNNTTNTNN
TNTATNNNCNCTGTTCTCCGTTANTGNNTTTNGNANNNCCNNNTNACCCCCNTTTGNTNA
NCGNCNNCCNTNGNTTTTTGTNTNTATTNNNNTNNNNCCACCNATNNTNCTTTNNANGGA
CNNGCNNCAGTTTNNANNGNATACCATNGNCATNAGNTTGGTNNANNNANNNNGGANCAN
NNCNAANNNNTTNNNGNNATGNGNANNGNAGNGTGCGGGATGCGGGCNNNGNGATTNNNN
NGCNGTANANTGTCTTNNNANNGNANGTNNNCNNNGTNNNCAGNTGNNTNCNNNTNNNNN
GNGNNGANNTGNNNNNNNANANNNGTGNANTNANAAAGNTNNNNANAGTGTNNNTAAAGA
NTNNANGNAANNTNANNTANACGGCGNNNGAGTGTNNNNCNNNANNNGNGCANTNNGNAA
ACTNTNANCTTGGNNANNNNNNNNNNGNAATGNANNNNGNANANNNNANNNNTNGNCNNN
NNGTTTCCNNCTCNCNNGTTTNNTTTNTNCNNNTNNCAACNTCCTGNTTNNTTNGCNNNN
AANANNANNNNNCATNCCCNNTCNTANNTNNTTTTTTNCACNNANCANTNNCTTNTNNGT
TANNNTNCNTANANAAGCTNTTNNNTNNATATTNNNTACTNANNNNAATCCNANTNGNNN
TGCNTGGGTNGNGGNCACTTNNCGCCANCGCACTTATTTTTTTANTNACNTNNGATGGNT
NCCCNNNCCGNGNCNNANCCCCCANTATNNNCNTNNNNNNCTCGCCNGTCNTNATNTATN
NTNAANNGCNCTCATNTNNNNCTTTNNCTGGNNAACNGGCNGNGATTAGANTNGCCANNN
TTGNNNCNNANNAATAACACCGCNTATACAAAANATACNCTATNNATNTNTTTTNATGNA
GCGNGNGGNCCNNGNNAGTTNTNNCNNNTCGANCANTNNAGNNNNANGTNNNTNCTNNCN
GGNNCGGNNCGNGCNGAATANANNAGANNNNNGNTCGNGTCAATNGANANNNNTNNGAAA

GANNAANAGTAGNATNTTNAGNNCGCGNTCANTNTANTNTNGGNCCTNTCNGATATATNG
AAAATACCCGACNAATGNAGANATANANGCTNNANNNNAAACNNCAATGAAAGNCAC
>A282
GGNNGANAAAAANGGGGGANGTTGAATGGGGGTNNGCGNGAAAGNTANTNTGCNGTNAGT
GGTNNTGNAGCNNANAANGGCCCCCNNTTNAAAANNAAGNNACCGNNCAGCNNCANNNNA
TNTNNGNGANGNNAAACTNTGGNCGNCNNNACTATNACNGTTAGANCNNNCTANTNNNNA
TGNNANNATTATCCGCNNCGNNNTGGNACNGNTTGGGGCANCGNGNCNANNCCNCNTNNT
ANNNCCCCNCGCCCCNACCNNCNNNATNACTNTAAGNNGTGTATNGGACCGCGNTNNTCT
NCNTTCTGNANCNNNGNTTACNGANAGCATATTANTNNNNNNTTTTTTNANACAGNNAAA
NAANNACTCAGAGTNTAGGGGGCGANCANNGCGGNGGNCGANATAGAANNNNCCCGACNG
NTTNGNNGCANNTAGNNACNATTNTATNGTCCNTAGNGCTNANNNCANNNNGNCNTGTNN
ANGGGTCTNTTAANACTNNANTTNAANGCAANTTAANNCCTNGTNNAATNNNGTAAANNA
NNNNANNGANCTNNCTTTTGNATTTNNNNNANAANCANGGTNTTNNNNAATNTNCNTTTT
TGGGCTNNTCCNCNANCTNCTTNTCNTTTTNNACNCNNTGNNNGGNNNNTNGTGNTNGNT
TGNNCNNTGTCAANTCNCTCTGTNNNTNTNNTACGAANNAATNAACANNANAGANNGTTT
GTGNTGTTCNACNACTNCANGGNNNNNTNNTGNNAAANCNTNNNCTNNCGNTATAAATNA
NNNTTGTGNTGTNANAGTGGGNTATNTAACNGNANATCTTANNTNGANAGNTANTTAGTA
NAGNNACTATGNNTNNTNAGTTNACGNNCATTTGCTTNGNNANNNGNNNTTGGNNATTTT
TNGGNNNNNTNNAGCNCGNCATNAGTATNNNANNTANNNNTANAGATTNNTTGTANNNNT
NAGCNNGGCTTNNNNNNCTNNNANTGNGTGNTNNTNNNTNANANNGATAANTGTNNNAGN
NNACACNANCGNNNNCNNANANNGTTNTGNNCCANNAANCTTTTTNNTTANTTNACNATC
GNNCNTNNNATGTTTGNTNNGNTCGNNNCNNCNGTNANNNNGGTCNNCNANTNNNTNTAT
TTNTNTNCANNNCGGNNCNNTCNNCTTNAGANANAATNACNNCNNNGNNNCNAGTANTAG
NNNTTNTTTATNNAGNACNTANNNTNNNATANNAANTANNAATCNGNNNNGNTTNAGACN
NGNNATNTTCNTGNCNAGTNATTTTTANGTNCNCNCATNATNNTTTGTTTTNACGNCCNN
CGTCNNTANACNNCGNNCTNTTTNTNNTTNTNNANNNNNCNNANTCTNNTANGANTNTAG
GCCGNCTGGGNGTATTCNTCATGGATTNTNANATCNNNNGNATNGGTATCAATTAGNGNA
CNCCNANANTNTTTANNNATTAGNNNNGNGNNAAACGACAGTAGNNNNTAAANNAATNTN
NNATNANAANTTCCTCTCANCGNACGAGGNGAGGNTNNTNTTGCTGTGCGAGANNNCANG
CCTNNAATNAANNNCNAGTCANNANNCTNNCGNGNTCNTTAAANAAGNNAGGANANGGCN
ATGANNNNNCCNNNTCGCNCAACNAATGGATAATGTCTNTANNCNGGGCGNAGNTAAGTA
GTTCANNCTANGNNTTCANNTTANGACNGNACNGNNAANNTNNTANNGNGTATGTTGNGG
AAACATATANNGCANCN
>A283
GGNNNANAAAANGNGNNGTANNTNANNCAAGGGCGAANGTGGCGNANAAAAGNTGNNNAA
ATTNANTATNTGAATNAACANCNAANCCCCCCTCATANNAANAGTNNNNNNCGGNNNATA
CGNNNCGCCTNNTTTAANNGNTANNNANNCGGTNNATNNCGNCCCATNCNANTATAGTGT
ANNCCNTNNGCTTATTTTNTTGTGCNANNGNNNCTTNNNCGCGGTCGCNNNACNCTNNCN
GGNTATNCTNNCCCCCCGNNTCTNTCCGNCCTTTTTNTANANCNCNCNCCTACCGNANGN
CNNGGGNGNTNNTANCAATANGNGNAGNATTCNNGTTTCCNNCCNTNCCTTGCGCGTTAN
TCNTTAGNCAANNNTAAAACANNGANTATNAANATNNNTNTANTTGAGNTAGNNCNGANA
AGCANTGAGGGACGCANTNTTNGNGANNGNGNTGAANNNNGTGNNNGNNNGAGCNGAGAN
TTGGGCGCGCAATGGCGCACTGATCCGGNTCNGGTNGNATGNNNCNTTTTNNTNTAAGGG
GAGACNNNNNAGGGAAGTTGTCGTNTTTGNNGTCTCTCNGATNNANCGTNTAGNNTNCNN
NCTTATNTNNNCCCNNNGATGNNNTANTNACAACCGNNGCGNNCNANGCACTCNTTGTGT
NTTCTATNCACNANNCCNCCGCGTTTNGNGATTTNANTAGNNTTTNNNNNNCNCCAANGN
AGTTAANTTTTGATGNCGAANNNTNAGGANTNGNNGTTNATNGCTNTAAGNNTGCNNNTN
GGANTNNTTNNTTTGATTNACTCANNATNNGNAAGNGCNAANTGNGTANGNGAGNNNCNT
TGTANNTGTNANGANTGNTGNNCNGACAGNGNNGNGGTGGTGNTNNATAAANNNGGAGGG
GTNGTACTTAANGNTTNTNNTNANAGNGCAATNGTGNGNGNTGAANANGGNNACGTNGGN
ANCANACACANNTTGGGTGNGGNNNGNNCTNNNAANNANGAGNGNNAAGAAGTGGACTNG
GGGNTNTANGGGTANAGGAATNNNTNTNNCAAANNTGNGTGACNNAGAAGNNNTNAGNNT
NNATANACNAGCNGAGGNANCNCNTTTNANGNCACACNNANANNNGTTTTTATNCTNNNA
NANGAGNNTGATNNNNGNAATAGNNAGANANAATTNGNAGCGNNATTGTGCGGNNNAGAN

NATGTTTTTATTATNNTNANNGTTNAGNCCCTCCNACTTTTNTTNGCACGNNNNNNNGCN
GTAGNGTGCCNCCNCNCCTTNTTTTATTGTANNNTAGNTCNCNCTCNNCTTTNCTNTCNG
AATNNNNNTTNNTCGGNNGAGAAGNTTTGNTGTNTTNNTATTNTNTANNGNGCNNNNTAN
CCNNANCGACNATNGANNNTNTTCACTTCNANCNNGNNNGTANAGNGNGGNNANTATTGN
TTTATNCCNNCGNNANNNTATGTTTTNGNNNTNCCNGAANNCTGTCAGNCNNTGNGNTNT
NTTTTNANATTCNACACNTNNCTGTNTTTANTTNNTATAAANCNNNNCGNNNCTCTTNAC
GTNNNNTATNNNTNGTANNGANCNNNGTTTACTTTTNNCNTTTCGCCAANANCCNTNAGN
CTANAACTGATNNCACAANGCNGNNANNACGGCGCAANTAGTNACTTAAATTAANGANAA
TANTTNGTANNNANGAGGANNAANAAGCAANATAGNTANGNGTGNCANGGNCNATNCTCN
NACTTCNNNGNCCNTGAANTNNGGANGGNNGANTTNTANNAATNTNCNTNNCNNTNNNCC
TGNNNANNAGTNCCTGNCGANGAANANNNNTANAANATGTNANCNNCAANGAGNAANTAT
TANNAGGNGTACCACTGNANGTTCNNCGNNNCNNAAAAGANCANNNNCNGATCGTTAANC
NNAANTCANNNAGNNN
>A284
NNGGGANGGGGGGGGCCTACACAAGGGCAAGGTCGGTAACAAGGGATAACCGGATAATAN
TGTTTNTNTTTTGNNGGGGNGCGNANAANTNCGNCCCNTNTNNAGAGNANNGNGCGTACC
NCCGGGNCCNTTCTNANTTTTNNACNGNTTTGGTCGGCNCNACTTCNCNATGTTNNCNCC
NCGATNGTTTCNANNNCTTATNNCCGCGNNCNCANTGGTTGCCNTNTTCCNCNCCGNNNN
CGCTTNCCTGCCNTNTGNTCCNCNGCNCNCACGGCNCGGCCNCCGNGANNGNGCCTATAC
CCGTGTATGATTGTTNNNCCTGGGTTCTCNTNGNCNTTTNTNACCNGNGCNTNACTCNNT
CGCNGTTCNNCATCNTTATNTTTATCCGCTANCCNATTTGCTCNANNGCCANCTCANAAT
NAGAGCGGNCTTAGTGTANCATCTANNNCGNGGAGAGGANGAGNCNNCCGCGNTNTNTTA
CGGGCCGGNNATNTCCNCNTTNGAGTNTCTTCGGNNCNCCCNANNCGNGGTGGNNTTTTN
NNGGTCNCNNTCGATTCNTANTGATCGGACGNCATCNGNNTANCNCTCNNNNNTNTTANC
GTNAACNGCNNNNCCANNNANNAANNNTTNTTGTNNTTNACGGCNTNCGNACNNCNTNNT
NGTNTTTNTTNTGNGTTTNNNCTTTTTTNCNTCNGTNNTTNTNGTNTTCTACCCNCNNNC
TNNNNNNNGANTNTNTGTNNANNGANATNNCATGGNNNNNTATNCANNGNTNTTGTNCTC
CANCCCGNNCTNGACACGCNGAANNAATGGNNTTNTAGGANNTGTGCTNTNANNNNTGGN
NNNNCTGCNTCNNNCCTTGNNCNNNNNGNTGACGNTGNCANGGNTNGNTGGNNNATGGNG
GTTGNGGTNCGTNGTNGANNAACNCACGNTAANNNTNGANGNNAGATTAANNTTGGTTNG
NNNTATNGTGNGCNTGNTTGTNTCTCNNCGNCCTAGTNGNNNNCNNGNNNCNNGGTGNAT
NNTANNNGNTTGNGGNGNNGACGTNGANNANTGNNNCTTNNNGGTNNGCNANNANTGTNN
NNGTNNGNCGTTNNNNCAGNNNNANNNGCANANNNNGCNNANNNGGACNNGNNNNTNNCA
TNCTNGNNTNNNNGGNNGATNNANGNAGNNGTNATCNNCTCGNCCACNTNNNNCCNTTNG
CGCTATTCGNGCNCNNCNCACNCTTTTTCTGTTTTTTTNCATATGNCCCNCCCANCTANN
TTTTTTNTNTNGGAATNNTGTNNTTTNTCNTCCGTCCCNCNNNNTTTNNTTNATGTTNNA
CNCCNCGTCNNTANNCTTCNCGGCTNNNANCTNCNNGCNNNAATAGNNNTTNNTNTNTNN
TTGNTNNTAANGTGNGANGAATAGGCGCANAACGNACCNCNNNGNNGTTTNNCGAGCNTT
TTNCTCNGCNCNCCCNTTATGTTNNTANCCNNNNNTCANTTTANTTTTAANCNNCNCCCG
NCNGTGTCNCTCCCTNNTATTTTTNAANTTNNANNNNGCGCGCTCTTTANNNTTGGTAGC
NACCNGCNCCANTNTTCTTCNNTGTNTTGTGCNCGAATCGNCNGTGTNCNTNNTNTTNNC
NAANCNNNCGNNNNTNCNAGTNTGCNCACGANTNTNGCCNCGGCTANNACNNGAAGNGTT
ANNGTTCTNGGTCNGACNNGNTNTGNTGNNTCCANNGNANNGNGNGNGNGANCATCNNGC
GGAAAATNNCNNGNGGCGNNNNCGAACATNNGGANNTGTNGTAANCGCGNCCATCNNNCC
TNGAATTTAGNGNNNNCNCNNCTCTNNNNCTTNGNNNAGAGCGNGTNCNNNTACNTANTT
TNNGCAGNAGACNNANNTNNTANAANTNTCCNGNNNNANNNTACTGTNANCNTTANNANC
GCAATATATGNCTNNGGCNN
>A285
TTNGGNNTGNGCGGCCTAACANGCAGGTCGTAACAAGGGTAACCCGTAATGGGNGGATTA
GTGGCGAACGGGTTGAGCCCANNTTNNTGGGCAATCNTGCCCTGCACTCTGGGACAAGCC
CTGGAAACGGGGTCTAATACCGGATACTGACCATCTTGGGCATCCTTGATGGTGGAAAGC
TCCGGCGGTGCAGGATGAGCCCGCGGCCTATCAGCTAGTTGGTGAGGTAATGGCTCACCA
AGGCGACGACGGGTAGNCGGNCTGAGAGGGCGACCGGCCACACTGGGACTGAGACACGGC
CCAGACNCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCA

GCGACNCCGCGTGAGGGATGACGGCCTTCNGGTTGTAAACCTCTTTCANCAGGGAAGAAG
CGAAAGTGACGGTACCTGCANAANAAGCGCCNGCTAACTACNTGCCAGCAGCCGCGGTAA
TACGTAGGGCGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGCTTGT
CACGTCGGTTGTGAAAGCCCGGGGCTTAACCCCGGGTCTGCAGTCGATACGGGCAGGCTA
GAGTTCGGTAGGGGAGATCGGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAG
GAACACCGGTGGCGAAGGCGGATNCTCTGGGCCGATACTGACNNTNAGNAGCGAANTTTT
TNNGNNGCGAACANNATTNNATACCCTGNGTAGCCCCACNCNCTTNNNATGGTTGGGCCA
CTATGTNNTGGGCAACNTTCCCNGTTTGTTCTTNTTNTGTTNTNTNNACCCCNTTAANNG
CCCCCNCCTGGTGGANTNTGTTCCGNCCNAGGTNTNNNATTNCNNANTGTATTTTANCCN
GGGNNCCCGNACNACCCTCTGTAANTTTNNGNNCTCTNNTNTNNANTTCTANNCCCCAAC
NATCCTTTNCTANNGCCTTNANCNTNCANCCNNATNCTTTCTNCNGANCCCNNGCANCNT
NTTTGNGCTCNNGCTAACCGCNNGGTTNCANNTNTTTTTTTTTATTTNTTTTNCCTGNCN
TNNNNNCNGNTTATCNTCCGCTTCTTNCTNTNNTCCCTANNNNGNNTTNCCCCNTNNCTA
CNTCTGNNNTNANCNNANTNCTCCANNNNCCCTCTTNNTCTCCNCCCCANCNACTTTNGT
CTATNTCTTCCCTCCCTCTCNTATTNTTTCCCCGTCCCCTTTCTTTCATNTCCNNCCTAT
NNCTNTNATTTTTCCNCCTNNCCNNTCTTCTANTTCNNNCC
>A286
AGGNGGGGCTAAAATGGCAAGGTCGTAACAAGGGTAACCGTAAAATNNTGTTTTTTTTGN
CGNACNGNGNANNCCCTTTTNAANGGNGAACTGGGNAACCCNNATACGGNGACNNGGNNC
TGANANTTTTGTNAGNNGNNACCGNTNTTGGNGAGGACACCNCTCCGNGATGCCNGTCCC
CCNNNCCCCCNCTNTNNCNCGNCCACCCNACCGCNNNANANCNNNGTGGNNACGGATATA
TCNGTANNAGNNACNNCNCANGATTNCCNCATAAGTAANNANNGAANAGGCGACCAGNGG
ACNGNGAACAAGNNNNACNAANANGCCNCGNANCATTCNGNNGNTNNGGCNTGTNNCNCN
GGNNTNCNANNTNNNTGNNNANTANTCTGTAATACNNNATTNCCTTTTNTNNTNNNNNNN
NNNNCCNCCNNCCTNTNTTNNCCNNCNGNTNTATNNGNGTNTNTCCTTNANCGCTCACAC
TTTNGNGTCGNTACCGNGNGTNTTATNGCNTTNNGATCATNTTNTGANTNANNNAATGGN
GCACCGNAGNNAGNNGAGNAGNTNGNNAGATNNNNNANANNNGAGTGTAGNCNGNGNTAT
AGTNNGGNNAAGTANNTNNNNTATNNAACNGNCANANTAANNTAGNANANNTACCTNNCT
AGCAAACTNTNNNGGNNCGNGAGNNNNNTTGCATNAGCNGGACGACANNTGGNNGAAATG
NNGNNGGTNNNNNGGANNNNCGTNGNNNGTNCNGTNNGNGNGNNGTNNNNAGNTTGNATN
NNACACNNAGCNNCNANNTTTCCNCNNTCCNCTTTNTTAANTNNCNGNANACTTTTTANN
TNNCNTTNNTCCNNNNTTNTTATNNTNTNCNNGGGNCANNNTANTCNTNNTCGCNNNGNN
GNAGANNNTTTTTATTTTNTNTACGNNCNGNACNNNNNNGNNGGNNNTNACCNNTTTNTN
NTCCNCTTNTTTTTTTANCCNNNNTNNTTCACNNNCCCCNTNNCCCNGCNNNATTNNNNTN
NTCNNCNTCTNNCTTCNNTNCNGNNNCCTATNCCNNNNNTCNNNCNAACCNTTGNTCGCT
TCNNACNCTNGAANTAANTNNNNNNNGNANNANCAGGTANCTNAATTTNNANTATNGNCG
CGGNNNCNCNGGCCTGNNNNNNCGANNAAAANACANCNGANAGCNGNCNCATAGNGNCGT
AGTNATANGGNNNANNGAAATNNCCCNATNAANCNCNACNNTNANAAAGTNNTCTNANCN
CGCNNTGCCANNNNANAAANTACAGNGNGCCNNNTNGTATNACAACCNCGGATATNTGAN
NTNNAAGATNGCGGTNNCNANNAAAGNGNANNC
>A287
GAGGGGGGGGCCCTACACANGCAAGNTCGTTAACAAGGGTAACCGTNAAATGGGTNTTTT
ANTTGGNCNGAACGGGGTGAAGCCCCCTTTNANAGGTNNANCCTGNTCCANTCNNNATGN
GGTACAAGCTNNTGGGAAACGTGGGTNTTAACGACCNCGATAGTTACCCCGTAGTCGCNN
CACCCCGCCNTCNGGNTCCGANCNNCCCCCCNNCCGCGTANNACNGACGCAGCCCTNNNN
NNNNNATANGTCCNGNNCNGGTGANGCNNCNANCTCATTCAATGCNACNTACGGGTATTN
TTNCTTTATNGGGGGANCNGANACACNGGNANGGAGAANGAGGGNAGACGAANTACNGNA
NGCNNTTNGGGGNNANNTNNTATACTGGTCNANNNCCTGNTGTNNNNTACTTTCNNCTTN
NNGGATNNNCCNCCCTNCNNNTTTNTANNANCNGCNCTTGANTTTGTNAANAANCNCNTT
TNGANNTGTNNNNTGCNNNTNNGCNCTCTTTCGTGTNTANCNNNCTNCAGCCNGTTNGNN
CNTNCGCGTANATTACNTTNTGTGGTTGNTAAANNGNNNTNNNAGNTCCNANGNNNNNTT
NNNTTNNNTANAAGNAGNCTNCAGTAANTGACCNTNTNNNGGNCNNNGATNNANGNNNGT
NNCANATTNCAATGNTGNGTNNNTNAACCNCCCAGGANTACGNGTNATTTTTNNNNTATC
NNNNGNCTTNNNTTTTNCCAGCNNTGGGNTNGAGNGGNAAGAATNCTANGNAANTTNNNG

CNGTCATGNNCGCCNAGTTATTNNACCCNCANNNNTNGATNTNNTNNANTAATGNNCNCC
CCNCANTTGNNCCGCCGCNGNTTNTTTNTTTNNTTGGNGGTNATACTTTNGNTCAGTGNT
CTNCNCCNNCCTTTCCNAANANNCCTTTTCNNCNCGATCNCTNNTTTTTNTTNNCNNAAN
NCNCCCCCNNCTTTTCTNANCCCGTNCANNNCTNNCGNCAGNNCNTNNTTTGTTTTTTTT
TTTNNCCCCNCCCTCNCCTCNCCNNTANACTGCNCNNNACNCGCNNCNTCNTTTTTTTTT
TTTTANANCGNCGGNTAANTACCANACNGNGNGGGNTGNNNAAGCNTTTTACCCNNNNCN
GANTGGTTNGTGCCCCCGTCNTNTNTCTACNCGCANGCCCNCCCCTNTNGCGCCTNCGTN
TNTNTNTTTNNTNNTGNTNCCGCGCNNATNTTTNTATNAAGNCGNNNANCNCTANANCGN
GNNNGCAGNNNNCNACNNCCNCTAATNNTAAANCNNCTCNGNNNNNNCCNCANNGCNATA
TNGNNNANNNNNNAGNNNACCNANANTNGAGNNNTATNNNNNAATTGGNGCNNGNNCNNC
NAGNNATGCCNNANTTNNCGANCAGCNGNTAAACATNGCNNNNTAANCNNGNCCCNNNGN
TAAGNNCGTNTNGTNACNNTCATNNNNNNNTNNTCTNNNATGGNNAGNCAGTNNNNCAGN
NAANGTTNNNATTTTNNNCCNCCCNCNNNNANNNTANCTCGCCNACGNNTNGTGTANNGC
TANCCACANNNNNGNGCTGNGCACATCCNNNCTACANGNNNAAGTANCGNN
>A288
GCNCGTAAAAACGGCCAANTTATGTAAAAGNGGNNGTNNACNGGNAAAACNNNNANGAAT
NANATNTGNACNNTTGAAGNCGNNAANNCCCCNNTTNNNANNNNNTNCTNCGNCCGNGAN
CNTACGNGGTNNNACNGNGTAGGNNNNTNAAGNNNCCANTNTGNTCNNNATTCANGGCNN
NNTCGTTNCNAGTANCNGNGTANTAAGCNNNCGGCACNNCGNTACCCTTNTNCGNTCNNC
CCCCCCCTTCCGNNCTTTTNNTTACCCACCNNCNGNCNCGTNTGCGAANNGNTTAAACCT
GTNCNGCAAAGGCNAGGTTNANANCGCANTANNCNNGGNTACNGGNTNNAGTTNNTNCAN
GTTATCCANNTGNTTANTCNTTANCTTNNAGGAGCANNGANGANGGAGGAGGGNACGGGG
ANTNTANNCGNANATAGTNANAGANNNNCNANGNTNCNTATGTACANCGNNNNCGGNCNN
TTCNNNGGNNTGNATCNNNCCGTTTNTGNCTNCTNGTTTNCTNNAANNNGNNTNNTTNNN
NNNGCNCNCGTTGNTTTNCNTANTNANTGACGNCNTNTATNAANNCTTNTNTGTATTTNT
TTTAANCNNCCTNTNCNACTNTGAATNNTTTTTTTTAATTNNANTANACCNGAGAGNGNN
NNTTATNNTTANTNTNNNNCNNNTTTACANCACGNTNCTNTGGNNNNNTNNNGANCCAGN
ANNNCNNGNGTGNTNGTTTAATTNCNNNNTAGNNNTGTATNNTNTATNGTGTTCNTNNAC
NNNGATTTGNNGTATAGANAAGGNNGTGTGANGANAATNGNNGNNCGGTANGNGTNNNAC
TGNNNACGNATGNGGAGNNTANNGCNNANAANTNGTNTTGNNANTNNTNNATTTTNTNTT
GANAAGGGANANAGNGTNNGNNNNTNGTNCNTTCTTTTNAGTNNCATNAANTNTNACATN
CNGTTTNNNNTCCANTTAANGNNANAGGCNGANAAGNTNNATTNNTTCNNGTNAAGTTTG
NGCAACNAAANNTGTNGANNNGANAGNAAATANNTCTGNGNTGNNAGNGGNCNCNCANNT
TGTNCCAACGNGACGANTNTATNNNNNCTATNNNNAGNNNCTTNTNGNCTNAANTTNNNC
CNNGTANCTNNCNAGAAGANGNTTTTNNGCGGCGNTGCNTNTNTTTTTTNTTTCCCNTAN
CNNNNANTCNGANCTCTTANTTTTNNACNGTNANCCNNCTGNTCATNCNTTTCNNCNCTN
TTTTTTTNTTCTNNTTTACNCCNCCCCANANGNGCTNNGNGTNTNNTATNTGCCNCTGNG
GNACNACNNNTNCNNTTNTTATNATTNNTTACCTCGTGGCNNTNTCNGATTNNAGCANAN
AGNGNGNNGNCTCNCGNNCACGTTTNTNTAAGCNNCNCACNNTTTNTTNTNTACNCNACA
NACATTNTTTNTTANANCGNCNNCNACNCANTGGNCNNNNNNNNTGTNGTNNNTNTTANN
CNTCCNCCCCTCTNATGNNTNTTGNTTNNNTCNNNCNNNTGTNTNGNNNTANTTCATTGC
CNCNNGNTNNNNANNGNCATTTCTNCTNANNNNANACCNGTNCTNANNNNNTTGTTNNNN
CNNACNNNNCNNAANNNCNCNAGTTCCGNNGNANAANNGNTTNGTNTNTTNATGNCGTGN
GCNCAGNGAANNNNTGNATNNTNCGANTCGNTNCGATTTNTGCCTNNANATNANATNCCG
CNNTANNNACNNTNNNNNAGCNCTNGTAAGCGNANGNCNNTNCNATCGNNCGTGTNANAT
ATGCCCCNCNGNGANAANANNTANCANNAANNNTATCNCGTCGCGNNNNGTGNAATTNTC
NCNNNNGANNCNNTCTGNNTGAANNCNCNGCNGGNNGANNNNTNGTANNGCNGCNNGGNA
NNCNAGAANNNNANANGCG
>A289
NAGGGGAGNCNGNNANACNGGCAAGGTTGTTAAGCAGGGGGGGAAAGCGGGGGAAACNNT
CGTGATTTGATNTNANTAANGCNAACGNGANNNATCCCCCNTNNTATNNNNANANTTGNC
TGCTACAANCTNGNNTNTANGCNNATAGCNGNNGNNNGNNTNGCCNGNACTCGNTATNCT
GNCAGGCTNNTCTTNTTNGCNNANTNNNGCCGNNNANCNGNNCGNTTTCNNNNNCCNNNN
CATNGNTCTTANTCATCGNCCCCNANCNCCTGCACNCGNNTNNACGAATCNCNNNNCNCN

NNTACNCCGNNCNATTCNGNTCATCNACANTGNGTNATCGTNNCGNNTNANGNANNNATN
NNNANAANACGNAANAAGNNAGNGNAGGACGAAANAGCNGNGGANNNGATGGGNGNANNN
GNGANANTANNGNGNNNNCGCGATCTTTCNGANNGTNNTNAAGNCNNNGNANTNCCTTCN
GNNNNNNGCCNCCCNCTANGCATGTNNGNNGGACNCTNNTCTCTNTANNCTGANGNCGNC
NTCTNGTCACTTTNNTNGANNTTTTNTNACANNCNCGNNCNATNNNCNNTTTTTTNNTCT
NCCANACCTCCACNGGTNTTAANTNNNTNNTTCCNNCNTTNNGNCNACTTNNCCTTTTTA
NNNTCNCNNNNANGNNGTNNTTTTTTNNNNNCAGTNNAAAATGANAGGTTTTTTTGATNN
AAGGNNATGANAGANNNNANAGNNTGNTGAGAGNNTCACAGNANNANAGGANGGNGNGTC
TANNNTGTANCANATGNTGAGAGANANNNAGAAGGNNGATNTTNNTCANNNTTCGGTGNT
NTNANGGAGNGTNTNTCTNGNTCANNNNCCTNNANTNNTTTAGGNNNCAGNGNNCGTNNN
TNGTNACGTGACTNNTNNAGNNNAGNGAAGTNGTNTANNTNNGTNGANGNNGANNTANNT
NNACATGCNNCTTGGNNTANCANGNNATNNNTNTNCGTGNNTNANNNTCNNNGCCNGCCN
CNTTCNNCNNGNNAATATGGGNTTCTNNTANNNGTCACNGNTNNNNGGANNGNTNTTNNA
GCCANACCTNNNNNNCNNNNNATTTNNNCGNNGGNNANNNNTTTNTTGTTTANCTNNNNA
CGNNCCCNTTNNTNTNGTTCTCTNCAGCNCCNTTGNNANCCTCNCACNNNTTTTTGGTNC
NTNGANCNNNNCGCTGTGCNNCGGCATNANNNNNGCGCACNANNNNGANNNTNTTATTNT
TTNTTTNNNNACTNGCNGNTNNTNNGAAANACNGGGNNCTNGGNGTTGAAANGNTTTNNG
CGGNNGNNCGTTTTTNTTNTCNNACCTTTNTTTGTTNTACCNTCCCCNNNGNCCNNTNCG
GNCNCATNTTTTTTANTCANTANNTNCGGCCANTCTGTTNTNTANNTNTNNCNCNNGCTA
CCNCTGTANNCTANTTNTNTAGTCANNCANCNNGGTNNATCTTTGCTCTTTATCGAACCN
AAAATGTTTCNNNTTACTCTGACGCGNNTTANAAACNTANTGNCACTATCAGTATNTNNN
TNNNNTTNACNNTCGGNCNCCCANAGAGAGTNTTGNTTCNNTNNANTCTGNNAGNNNTNN
ATNCACCNGTGNNANCNNNANTNANNNGTTNAGNGTATANCAGNCTATNCCNCCNNATTN
AGNNNCGTCGCNNAATNNATGANAANAATNTCNNNNGCCGNNNNCTNNNCNATNTGTNCN
NNCNNNCGCTCTNANTGTCNNTCGCCACNANNTATNANNCTATAGNNNTNCGTATTNNNN
NGANCAGGTCGNGCN
>A290
ANGNGNGNNCCTATCATGCAAGTCGAGCGAATGAAGAGGTAGCTTGCTCCTCTGATTTAG
CGGCGGACGGGTGAGTAACACGTGGGTAATCTGCCTGTAAGACGGGGATAACTCCGGGAA
ACCGGGGCTAATACCGGATAATAAGAANAAACGCATGTTTCNTTTTTGAAAGTCGGTTTC
GGCTGACACTTACAGATGAGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCACC
AAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGG
CCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCGGCAATGGGCGAAAGCCTGACCG
AGCAACGCCGCGTGAGCGATGAAGGCCTTCGGGTCGTAAAGCTCTGTTGTTAGAGAAGAA
CAAGTACGAGAGTAACTGCTCGTACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACT
ACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTA
AAGCGCGCGCAGGCGGTCTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGG
TCATTGGAAACTGGGAGACTTGAGTGCAGGAGAGAAAAGTGGAATTCCACGTGTAGCGGT
GAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGCCTGTAACTG
ACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCG
TAAACGATGAGTGCTAGGTGTTGGGGGGTTCCACCCTCAGTGCTGAAGTTAACACATTAA
GCACTCCGCCTGGGGGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCC
GCACAAGCAGTGGAGCATGTGNTTTTATTCGAAGCAACGCGAANAACNTTACCAGGTCTN
GACATCCTCTGACNACTCTAGAAATAAACTTTTNCCCTTCGGGGGACANAGTGACAGGGN
GGNGCATGGTTNTTCNTCCACTCNTGTCNNGANATGTTGGGTTAATTCCCCCANNNANNG
CACCCCTTGACCTTANTTTCCANCATNCNNTGGGCACNTTANGNGACTGCCGGNAANANC
CGAGGAAGGGGGGGGANNACCTNAAATTTTNTNNCCCTTNNACCGGGGTNANACNNGCTCA
AGGATGGTNAAAGNTTTNNAACCCCGGNCCACCATTCCCAAAACCTTNCTTTNAATTNGG
GNNCCNCCCC
>A291
AANGNGCNCCTANCATGCAAGTCGAACGGGACGGNTAGCAANACCGTTTAGTGGCGGACG
GGTGCGTTAACGCGTGGGAACCTGCCCTGAAGTTCGGAATAACTGCGGGAAACTGCAGCT
AATACCGGATGTGGCCTGTGGGCCAAAGGGGAAACTCGCTTCAGGAGGGGCCCGCGTCCG
ATTAGCTAGTTGGCGGGGTAACGGCCCACCAAGGCGATGATCGGTAGCTGGTCTGAGAGG

ACGACCAGCCACACTGGAACTGAGATACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG
AATATTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTGAGTGATGAAGGCCTTC
GGGTTGTAAAGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGC
TAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGG
GCGTAAAGGGCGCGTAGGCGGCCTTGTGAGTCGAGTGTGAAAGCCCTGGGCTTAACCCGG
GAAGCGCGCTTGATACTGCAGGGCTTGAGTGTGGGAGAGGTTGGTGGAATTCCCAGTGTA
GAGGTGAAATTCGTAGATATTGGGAAGAACACCGGTGGCGAAGGCGACCAACTGGACCAC
AACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCA
CGCCGTAAACGATGTGTGCTAGACGCTGGAGGACCTAGTCTTTCGGTGTCGCAGCTAACG
CTTTAAGCACACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAAGGAATTGACGGG
GGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAANAACCTTACCAN
CCCTTGACATGGGGAGTATGGGACGGAAANATCTGTCCCNTTCANTTCGGCTGGCTCCCA
CACAGGTGCTGCATGGCTNTCTCACCTCTGTCNTGANATNTTGGGTTAATNCCCCANCAN
NNCAACCCTNACCTTNATTNCCANCGGTTCGNCCGGNNCTNTNAAGAACTCCCGNTACAA
CCNNAGAANGCGGGATNANTCAATNCTCNNGNCCTTNNGGNTGGNTANNNNTTCTNANTG
CGTAANTGGANNAANCCCAACGNNCCNATTTCAAACCCNTCATTNATTNCCTTTCANTCG
GNNTNAAGNGAATCNTTANTCCNATNCNNCCCNNAANCTN
>A292
CANGCGCNCCTACCATGCAAGTCGAACGAGACCTTCGGGTCTAGTGGCGCACGGGTGCGT
AACGCGTGGGAATCTGCCCTTGGGTTCGGAATAACTTCGGGAAACTGAAGCTAATACCGG
ATGATGACTTCGGTCCAAAGATTTATCGCCCAGGGATGAGCCCGCGTAGGATTAGCTTGT
TGGTGAGGTAAAGGCTCACCAAGGCGACGATCCTTAGCTGGTCTGAGAGGATGATCAGCC
ACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGAC
AATGGGCGAAAGCCTGATCCAGCAATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAA
GCTCTTTTACCCGGGATGATAATGACAGTACCGGGAGAATAAGCCCCGGCTAACTCCGTG
CCAGCAGCCGCGGTAATACGGAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCG
CGCGTAGGCGGCTTTGTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCCT
TGAGACTGCAAGGCTAGAATCCAGGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAAT
TCGTAGATATTCGGAAGAACACCAGTGGCGAAGGCGGCTCACTGGACTGGTATTGACGCT
GAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAAC
GATGATAACTAGCTGTCAGGGCTCTTGGAGCTTTGGTGGCGCAGCTAACGCATTAAGTTA
TCCGCCTGGGGAGTACGGCCGCAAGGTTTAAAACTCAAAGAAATTGACGGGGGCCTGCAC
AAGCGGTGGAGCATGTGGTTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCGTTTTGA
CATCCTTGTCGCGGATCGCANANATGCTTTTCCTTCAGTTCGGCTGGACAAGTGACAGGT
GCTGCATGGCTGTCGTCAGCTCGTGTCGTGANATGTTTGGGTTAAGTCCCNCANCGAGCG
CAACCCTCGTCTTTAGTTGCCATCATTTANTTGGGCACTNTAAAAAAACCCCNGNNATAA
NCCGGAGAAAGGNGGGATAACNTCAATTCTCNTGGNCCTTNCCCNCTNGGNTTCCCCCTG
CTTAAATGGNGTNACNAGGGCTCNANCCCCAGGGTAATTATTCCAAACCNTTCATTNGAA
TTTTTTCNACTNNNANNNTNAGGNNATCCTTNTATCCGANCACATCCCGNN
>A293
GNGGGNGGGACNANAAGGNCAAGGTGCTAAAAGNGANAANNGATAANTGGTGTTAANGNG
NGNGCAANGCGNTGAGNCNCCCCCCTTTTAANNNGAGNTGNGTCNTGACNTGNTGTGGGA
ANNAANTNTGGTNGCAAACTTATNNNGTNTAGACCNNCNTTCTNGTGTCCNNANNACNGT
GGTNANGCNTNGCCNNNTCCNCCACCGNTNNTNGACNNATTTNANNGCCCGCGGCNGGNN
NANAGCGCGTACCGGCCACAGGGNANGATACATGNGAGTNTGNANTNANGAGGNCTNNNC
ACTTAACATNGNATACNTANANTTAGAAGACGANGTACATGANCNCACNNTGGGCAGGAG
AGAANCNNNNGNAGGNATNCCGAGGNANNNGTTTCGAGNTTTTTNNNCNACNNACCTCTN
TTCCNNNNNGCGNACNANTNTGNNTGTTNCTGNNTTNTATNNTANANNCCTCTTTCNNNN
TTNTTCNNATTTTNTNNNCCNNNCCGNGCNNCNCCCCTTNTTNTTNNNNGCANCCCGCNG
NATATTNNTNTGTTGTTTNCCNATTACCCCCNTNTTNGATNTCNNNNNNNNNGNCCTAAN
TNNTTTTTANTNCNCTTAACNANTGNTNNANTTNTANATNAGANNNTNNGACNNAANNAN
AGTGNTGTNGTNNNGNTGNGNGGNGNNGANNGANTACNAGNGTTTGNGTTGCTNCACATG
NGTGTAGTNTNTNTGGATTGNTGCGGNANNNANGGTAANANGNTNNGTGANANNGANNTT
NTCCTGNTTTNTGATCNNCTANTGTANGNGNANNATGNANNGNTNANNNTTAGCNNCNNG

NNNAAAANCNNCTNNCNGNNTNGTNTTGAGTNGTTTNTNNNNNNNGCGANANNGNNAANA
NATNCNNCNNTGTTTNNANNCGGNAAGNGGTNGGAGGGNTNNANNNNNGANNGTTTNNGC
NNAACTNTTCCNGGNTNNNTTTTTTTTTTTTTNATNNGNNCACTANNNATTTNTTTACTA
NCTNGCNNNTACNNGGTNTNNNTNTNTTNTTNNTTNGNCCGCANTGCNNNTCTACNNGCG
GCNNNNNCCACNNANCACGAGNTTCCTTTNTANNTNNTTANTNNNGATTAATANNNGTCN
NGCNCTCGNGTNNNNNNNTATTTTNCNGNCGNANATTTNGTATCCCNNTNCAATTTTACC
CNNNCNNCNACTATNANCCCCNTGTTTNATTCTNTANNNTCCCTCNANNNTCNNNTNNTA
TNNCGCNCNTNTTTTTTNNTNGNNCGCNGNNCNANTNNNCNTANTNGTAATTGCTTNTTN
GNCNNNNNNANNCNTCNNTGNNNTACATNACGCANNNGCNNNCACANNTANTCTNTGNNT
TNTTNCNNTNTANCGCNNNANNCNCANCNACGANNANTNNATTNNGNCGNNNAATAAAGC
TNANNCNTTGNCNAGCGCGTNAGGTANGNATCGNNGAANNANTTNGGCANTANGGANNAN
NTGTTNNCNCNCATCTTNNTCAATNTATATCGTNNNCNCCCNTTGTCAATATTCAGCNCN
CCNNTCTNATANCNTNCGNGCAAGNGANACGGATANCNTGCNNACTNNANNTTCNNTCNG
NCTG
>A294
GGGGGTAAANTNNNGAGATATNANCANGGGAGNANGGGGAAAAATGAANTAGACGTTATT
GNNGAGANGTAAACCCCCCNTTAANNNGGNNNANCAGGAGNGAGANGTGNNAANGTTANC
CCNGGNAGANNNNNCNNNGTGTNNAGNNNNNNTGNGNTAAGTGCCNGTAATTNCNCNTNC
NCNTATNCNNNNNGCGCNCNNNGCNCCTCCNTANTTNTAANCNNGCGNNGANGGGNNGAA
NAGTNANATCTNTNCGTNGTTCTGCGTTNGACGCATTCTAGGGGTTCNANGNGCGNGCGA
NNNCNCCCACTTNNNCGCGCATTGTTGGNNCNCNGNAGTANNACNANCTCNCANGNNCNN
CGNNTANNGATNNNANGTNANAANGGGGCCNGCNNCGNCANATCGNNGNGNTACANCCCN
NTCNNNNTTCNTNNGGNGCCNGCCATAGANGTGTNNCTNCGTNNGTNTTNGGNNNCGGCT
NTNGTCNTNTNTTTTTTTCTTTTNCNNNACNCNNCNGCACTNNCCTTNTATNTNTCNCCC
AGNCCACCNNNTTNNTTNNTNNTTTATANNTCTATNCACNNTCNANNCCNTNATNNNNAC
NGGNNNNNAGNGGGTTNNCTTGANGNNAANGATNAAAANNNNNNNGTNNAGNNNANAGAN
TGACANNTNCAGNGNTCGGGGANAGNNNTGNANGNNCNAAAGAGAAGGTNAAAGATATNN
NGANNNGNNCNAGAAAGGCNNGNGAGCAGNGGANGAATANCNGGNNNNNTGCNCGANANG
ANNCNNCNGACTTGAGTATNTNNANANGANNACNGTTANNGGCATNTNNCNNGNNGNGNG
NATNATGNAGNCAACAGTNTAAAGCATGANANGANANANNGNGNTNNAGTNTNNAANAGT
NNNTNTNNNNANGGCCANNNNNANNGNNGNNNNNGNTGTGGNNNNAGNNGNNTANNGNGA
GNTGANNNNNGNTNANNNGGNNNNNANGACATGNNATANCGGCTGANNNCCAATTTTTTT
TATTNGNCTNANNCGCCNNNNNNCTTTTNTTNGNCNNTCANNNNTTNTCATNCGNCCNNN
ANGTTTNTTTCNNNTNTANNNGCNNNNGGTNTCATACCGTGTNNAANCTNNTNGCNGANC
NNNGNTTTGCTGANTTANTTTAGTNTGGCCNNANGNNNTACGANNGTCNNAANNGTGNGN
ACNTGNGAGNNTTTATAGCNNNCACNATTNTTATGTTNCACACNTNNNTTNNTTNCNNNN
NGNNCCGGCCATNNNGNTANATTATNATTTTATCANNTCCCCCTCTNGTNTANATTTATA
ANNCNGNCNCCNNNTATNNCNANGNTNCNNNNNAGCANAAGACNCTGTNNTNTCTTCCCG
AANANGTACCTAGNTNAATTNGTCAACGCGCGNGANNANNNNNNNGAGCTNTTNATNNTN
NTANTNAGTGTGNGNCNNGCGCCNCCNGNNNGCGANNCTGTNTNNCCAGGANTCGAATNA
NCNNAGANGCNNNATANGNTNANGCCAGANTGANAGNNNNAGNNNATGNANNTGNNNGNG
TCNATTGNNNNNAAGNCNCNNCGANTGTCCNNGNNANTNNNTNGNATGNATGNNCNATGN
CACNCANGCAAGCAGNATNATGACGNNNANGCNGATNTCNTANGNNGTCGCCCNGACNAN
NNNNNNNNNCGNGTNCNCTANTGNNGANCNATNANNTATACGTGTNTNNNC
>A295
GNGGGGGNGGGCANAANATGGCAAGNTCGNTNAACAAAGGGGTAACCGATAANNTTTCCT
TTNCNTTTNGTNNGTNGGANCNCTTANGGCCCCCNTTAANNNNGGGCGGANGNNNANCGN
GNANTATNNANAGNNCGGNGTNGTGGNCGNCNANTTANNNNNNTNNGGGCNCNNCNTNNT
ATNNCNGNNNNTNNANCCNCTNCGNNNCGNANCNCNTTACCTCNCGCCNNGCCNTTCNCN
CCTTNTNNCCNNNNCGANNGNANNAANTATNTANTNTNNNNANNGNNGNNGTACCNCNNC
NTGNNTNGANATNNGANGNNNNATNACCGTTCNGNTTNNTGNACGCNTATNGTNGTANNC
GTGNNANNNCNNNNNCAANNTANCGACCGGNTAGAAANNTNNAANGGANNNNGNTNAGNN
GGNNNCNNGCTCTTTCNCGGATNTTTATNNGGANNNNTTNCNNCTNTNNCNNGCCGCNCN
TGCNTATTGNNNATANANTTTGTGNTNCTNTNCNNNTCTTCNCCNTTTTTTTTNTTTNTN

CCNNCNCANNCANNNNNCNNNNTTNNTNTNCGCNANCNCCGNNAGNTNGNNNANGNTCGG
GATNCCCCNNGTTNNNNTTNTNGCGCGCTNGGCNGGNTTATGGNANNNGNCGGATNTANN
NTNNNGNTGNNNACANANNGGNAAAACGNNGGGGTGNAGNGNNANGAAGNNTNNNNNNNN
GTAANNACANTGCNCGAATCNTNAAGNANGATGNNNTNTGNNGNNTNCTTNTNGNCGGTG
NNTNAGNNNANNNAAGNTTNNNNAGANGGNGNNATGNTANNANATGGGCGATNNGGNGNN
NNTTTGGANGNNNGTNNGCGGTAGNNGGANNNNNNNTTNGCTNTGTNCNNNATNNTTTNT
GNNNANNNNACNNGNANCGGGGGNNCGNATGTCGGANNAGAGNNNNNNTGAGANGGANNC
NGGGGNCNTNGNGNGAAGGNTNNCGNNAGNCTNNNNTTTTTTTNTTGNTTANCNCCNNCT
NTTATGTTNANGNNTANGNGTTNCNTGCGCNCCCCTNTATTNTNATATTTNNTNGNCNGG
TCNCTNTNTCNNACTNANNCTGCCNNGNNGAANGTNTCNTNTNTTATNTANNTNACCNNN
CAGNNNNGNANNNNNNANGNNNGNNGNAGAANANNTNATGGNGTCNNCCGTTNGNNTGNT
ANTNACTCNNNATANNANGCNNNNNGNCNCNNNNAACCNCGCNNNTNTTNTTNTNANNNC
NCCCTNTNTCTTGTTNNNANCCNNNNNANNATNNNNNTANNNNAGNGCNGNNAAAGCCNG
NNNGTTTGGNCTNNNCGACNCGNAACNNGTTTAGTCNTTCGCTNGNGGNGANNCNGNAGA
ATAAGGNNAAAGTNGNANNTNCTNTNNAGGNGGNNGNNGCNGNNCTCNTNNNNTGCNNGN
CGNGNTNATNGNGNCAGNNCTGTNNNCCCTNCNGNTGNANCAGAGNCTATNTGNGNTGNA
NNNGNNAAGTNTNCATNNTGANACTCNTCNGNNGCNATAGANNNAGTNNNAGTCNGNCGN
CANNCNNNNANAGAATANAGTCNCGGGGGCGGTGTNATANAGAGAGGNANGCGGNAGAAN
TANGACTGGAGANGTGNGNNCGANNAACNNGTNAGNANCNNCC
>A296
GNNGNGCGGCTAAACANGCAAGGTCGTAACAAGGGTAACCGTNAGGTGGGGATTAGTGGT
CGAACGGGCTGAGNCATTTGANGGGCAATCTGCCCTTCACTCTGGGACAAGCCCTGGAAA
CGGGGTCTAATACCGGATACCACTCCTCAAGGCANCTTGGGGGGTTGAAAGCTCCGGCGG
TGAAGGANGANCCCGCNGCCNATCANNCNTGTTGGTGAGGNAATGGCTCAGTAAGGCGAC
NACNGTTANTTNTAGNCTTTATANGNNAGANANNACANANCGGNAACTGANAGAGNGTCG
NNGACNAATACNGGACGCATTNTNGNGGNATGTAGCATAATNGGCNNNANCCTCTNTTNT
NCNNTNCNNNCNTCTTGNTGCACTCNNNGCCTCTCTNNTTNNAACNCCNNCNCCCNCTTT
TNTTTNANNCNGCNACNNNTTTTTNTTTATTTGNCNNTAANCNNCTNTTNCTTTTTGCCT
CACNTCNCCATNTTTTTCNNGNCNTTGCNTNTCNNNTNNTTNNNTCACCNATTTNNCGNC
AANANANGTAGGNTTANTNGAGGTTNNAGTTAGTACNNTATNAATNGCTANNNNNNNTTN
TNTNNNNTTTGTTNNNNANCNCCNCTCNCNNNCAACGNTNNATTNNTATNNANGNTTNCN
TNTTTTCCCACGCNNTNNGNCNTNNNNNTNCTCTNTTNCTTTTTTTNCNTTAANNCNNNN
TCNTNTNNAATTTANNNNNNTNTNTNTNCNNNCCCNCCCTTTCCCNCTCCNNNNTNNTTA
NNTGNACTTAGGGNGTTAGCATCNTTTCACCCNTTCNCNGCNNNNNCTTTCGTCGNNCNC
NTTTTTTTTTTTTGCCTNNCCCNNCNCAATTTTTTTTNNNCTTNCNCCTNNTNTCCNNGT
NNTTTTTNTTTTGTTTTCCNGNCCCNCTTNNTNCGCCTTNTANNNNNNNAATCAGTCTNT
NTTTNTTTTTTTTTTCNNCNCCACTAAANANAGCNNCCCATGTNCTTNNAANNTTTTNTN
CTGCCCANNTTTTTTTANNNANCTATNTTATCTTANCCTNCCCCCCCCNNCNNCCNACCT
NTTNNNTTTACACNCNGCCCTCTNTTTTTNTTNNCCCCCCNCTCTTTTGCTCTTGTCTGC
NNCGTCNANCATTATTTNCNCTTNTCCNGCCNCTNCTTCNTGNNNTTNNCNACNCCNGTN
CCCNNTATCNGCTTTTANTNNTNNNTTATGNTCCNTNTCTCGCCGNAANTNNGTTGNANC
TNNAAAAANGNNANNNANNNNCNNANCCCTATTGCCNNNCTNTTANTGNNNANCNCACNN
NTCNNCAATANNANNNCCTCNNATNANNTATTANTATNNTCTTGNCTCNCCTGTCTNNNA
TTTCNNCNCACCCNTTCTNNNTTCACCTCNACNNNTAANNNTATCCAGTNCGTTTNNNGN
NAATCCCAC
>A297
TGGGANGGGGNGGGNCCTNACAAAGGTCAAGGNTCGGNTAACAAGGGGNTAACCCGGNTA
ATANANTTNTTNNNCTTTTGGTTGNACGTGNCCTNCNGCTNNNCCCCCCTTTTTNGNANN
NNANNGTTCNCCGGGCAANGTGCNNTCTGNNNTNTNCTGTNNNNCANCGTNNNNCTGGGC
NNGNCCTTNTNCTTTNCCNNCGACCCNCCGTCNTCCTTGNNANNCGNTAAGNNNNGNGNT
NCCNCCNACCTGTCCCGCTTTACCTGCCANTGGCGNNCTTTCNCTCTTTTNTNCCGNNAC
CTNCTCNCTCNGCGNCNNCGCGTTTTNCCTNTNCCTCCNGNNCTGTNTNCGTTGNNNCTC
CATTNCTTNNNCNNNTTGNTTTNTTTNNATNTAGGGTNTCNCATTTTGNCCTCCNTTTTC
NTTAGGANCNNGCGGANGNAACANCGGNGNGACTNANGANNGTGTCNCCGCTTTTNAATT

ANANNCNTNNGNNATNANNNNGNGCNCTNGNNCNNTCTATTCGGGCNNCNNATTNTCCTN
CNNCNCNTTCCTTCTTTNNCGNCCCNCNCNNNTGTTCGTGNTCNGTNGNCNCGNTTATTG
TTTATNGTNGTNCCGCGCTTCNNNNGCCCNTTTCTTTNTNGTTTTGTTAAGNNCNGCNGT
NGNTNCNNTCACNNNGTGTTTTTTTTTGTTTNCGNCNNNCCCCACNCGCNGCGTTATNGNN
NTTTNCTNTTTCTGCCGCTTGTNTCNCCTNCCTTNTTNNNNNTTTTGTTTCACTNNTTCT
AGCNGNNCANNTNNTCTGNTGTCCGTNNNNTGTTNCGCCTGTATNATTGTNNTTTTTGTC
NTANNACCNTNCNNNTTNNANNNCGNTGCNNCANGGNNNNTTTNNNGNNCNNNTNGNGNC
GAGNGGAGNTACNTTNNNTGNTNAATAATANTNNNTTNTGAGGTNTNTNNCNACAANANG
TNGTGANTNCTGCTGCCTNTTTNNTNTNTGTCNCTTTNNNGTNGTTGGNNATCATATANN
NTGNGTTTNTTNNNNAGNAACTTNTTCCGNTNNTTNTATTGNCACTCNCCNNTTTGTANT
TTNNNACAGANNGNCANNNNNATANTTNNTCTNTANGGCNCCTGNANTGCCCCCANNTTT
ACCNNNNNNNTGANNTGTNTGNCCNNNTNTGGTGTATCTATAGCNTNNANNGNTCCGTTT
ATCGCCTNNNCNCGTATNGGTTTTTATNCTCNTNTNTACNNTGNTGTNGNNCNTGACTNG
TTNGNNCNGACGNNANNAATNCNNCTNGCNTTTTCTNTNCNNCNGNCCNTTTTTNTCTTT
TTNGNNCTNTCNCCCNNNTCNNNNTTTGATTTTNNCNGTTANNACANCTTTTGNTCCCNT
TNANANNTTTTTTTTTCTNTTCTTTACCNNCCGCTNCATCNTNTTCCGCNGTTNTCTGTG
TCCNNNNGNNNNANTTTNTTATGNTTTNTTGNTNGTTACNTCNNTANNTCACNANCGANN
TNCACCGNTGNCGCNTNTGNTNGTNCTTTTTTNTNCCTTNNACGGNTNTTNTCTTNTNTGC
GCNCCNATNNTTTTNTTNTNCACNCACTCCGCCCCNTTTNNNCCGCACCNTNTTTTNTNN
TANTNNNCNNCGTCNCNCTNTTTTTCTATTTATNNANCCCNNCCCCGCACNTTNNGCGNN
TGTTTCNNGNNGNGNANNGCCNNTTNATATGNATCTTNCTACTANNTNNNGTGNCACNTT
TNCTNCTTGTCGCANACCCNTCNNNNNNGTCAANATCNTTGCGCNTTGATCGNATANNAN
ANTAATGNTGNNNNNGNGNCACTCNATCTNCNGCGTNTTACNTACNTGNTCTCGNCTACT
TTNTTCNCGNCNGCNATGTNNTCCGTNNTGCANCNTATCTNTNACGNANCTTNTTATATA
NNATTTACGCANTCNTTNGNNTNNNNNGTCNTNTTNTNNCTCCTTNCCCCACANCTATNA
NGNATNGTNNTTNTTTNCCNTCCNNCGTAGCTACNTTNTTNTCANGGCNTCNGNNTNTNN
TGGTNGTTNGCCACGCGACNAAAGTAGANTGATANCGCGANGTNNNNTCNCATNTATTNT
CTGCNCTNCN
>A298
ANNAGGNANGGCNTACACAAGGGCAAGGTCGNTAACAAGGGATAACCGGTAATAATTTTT
GTTTTTNTTGGGNTCTGCCCTGACGCACCCCTCNTTANNGANGAAATNNAGCGGCTANNA
NAGNTNANGACACNTNGGNNCGTANGNTNNCCNANNTAGGNATCAGGCANCANNTTNTTN
GCCAAATANGCNGNCGNGCNCTCNGTNTTCNCCCCNCCCTTGNCACTTNTATCGGGCCNG
GCCNCNGCCNNCGNNNACANTCACTGTGTAGNNNANNCNGGACTTTGGNTCANNNAGNAA
GNNGAAGTACAATATTNACTTATANANAAGNAGNANAGNGAAGAACNGAGNNGNGGGNNG
NGAAAGGGGNCGGCGGAATAAAGAGNACCNGNGNTTTCGANGANTTANGNNNCNGTANNN
CTNNNCCNCCGNNGCNGNNNATGNNNNTNNNANNTCTNGNGTTNGANATCNCCTCTTCTN
CATTTNTTGNNNATNNAANNCCCNCNNATNNCCTTTNTTTNNNANCANCCCNCNNNTNNN
NTNTNTNNNNTTNCCTTNGNNCNCTNCTTNNGNNTCNNGNGNNNTTGANNGTNNTNTTCT
ANCACGTAAANNANNTTNGNTTTNANANNCNNNTNNCNCCAGNGCAGACCGTGNAANNNC
NNTNATNGGTGNNNGAAGGANTNNNNGGGTNGGTNNNNGACANNGANTANATTANTNTNN
NNTGNTTGTGANCTACNGTTCNNNNANAGNNNCCTATNGNAGTANNAANNTTNNNTNNCN
NTNNCGCAANTNNNNNNAGNGCGAAANTTTGNNNTNNTNATANNNNNTNAGNACGGNAGT
NGNTNGTGCTNGTAGNTNANGGNNGNTTNNTNNGAGNNCNNGNNNNNCCCGCTNAAGANN
GNATAGATNGGTNTCGNNANATTGNNNNTANTNANNNCCNCNNATCGNCNNNCTTTTTNN
NNGNCNTNNTNTTTNTTTTTTANCTTCNANCNCNTCNATTTGTTTNNNNGCTNGCNANCN
TTCATCGGAGNNCANTTTTGTNTTTNTTTNNACNCCANNNNCTNCCTANCGCGTATTNAT
GCCGCNNCGNANNGCNTTCTNNTGNTNNNNTTTNACNNNNNNCTATTNCGANNNGCNGNG
NNTNNGNNNNNNNACANTNTATNCNNTCNNATATTTNNTTATACCNANNTNNTTTTNTTN
NANGCNNCNCNACNNTNTNCNCNGCTATNTTNNATNTNNTATTTGCNGGGCTNTTTACTT
ANANGNANCCGCNNACGCTTNTTNNTNTTANTGNAGNCGTCANNCCTTNNCTTTCNCNTN
NCCAAGNCCCNGNAGANATACNNATNNGCANANNGNATNAGCGNNCACNTNTAGTCTNTT
TNNNNTAATTTNGNTNACCCNTGNCTCCGCACCATNNCNTGTTNANNNNNNCGGANNATA
NNTGCTAGATNANNTAACCGTGNNCNTNGGACGNATGNNNANTANNNATAANGCAAGNTN

NACTGTCANNANTNGCNCCGNNGATANNANGTNNGNNNNATCTTNCNTNCCCGTAGNAGA
TATTNCNNCNGCGNGNTCNGTGNCNNANCCCGNCNNNTTNNGATNTNNTTNGTNGNNNAA
AATNTNNNNNC
>A299
GGCCTANAATGCAAGTCGNAACAAGGGAACCGAAAAAGNNTTGGTGGAGGACNGNNAGNC
GNTAAANNCCCCTTTAAAGNANGTCNGCCTGNAAGATCCGNCAAGTAACGTGCNGNTAAN
NNTGNCACACGTGANNATNCGGAGGNTGGNCCTTGTTGCNGTNATNACNNGGNNAGAANN
TNNNCNTTTCCTGCNNGCGGNNNNCGCNCTTTNTANCTNNCCCNNCNCCNGCCNGGCGNA
ACNGTCCCAGNCANAGANACATNCATNCNTGGATNNTGNATNTCGANAGNACGATTCNGC
TCTNATNTGNCAAAAGNAGTAAAGGGGNNNAAGNATTCCCNCGNGNAGGCANNCAGAGAA
GANANNNGCCGGATTACGGGNTTNAACNCTGATACTAGNNANNCCCCCGAGNTNTNANNA
ACGTCTTGTTTNTTNGCCNAAANNCCNTTTTGANTTTNNNCGCANNCATCNCNCTTTTTN
NANNCCAGNCNNGTTTNNNNNTTNTTTTTNCTTNNCCNCCTANTTNGCTNTGTACTANNG
NNANGGTGGTTTAGNNNNCNGTNNNNTTAATTTTTTTGGNNNTGNCAAANGGCNNNACGA
AAGGTAGGNNNAATGNNTNAGATANGANANANGAAANCNCTCTNGNGNAGAATNANGTNN
CNGTNAAGTNANGTTTTGNANNTTNNNTCGTGGNAANGATANNNTNGANAAAGAGTNGGA
TGNANATCANTNAGNNTNNCNACNTGGCTATAAATTANTNTNNNGNNNCGACNCNAATAN
ANNNNAGTTNATTCTTNTNNTTTCTNNNTACATCNATTNACNNNCCANNTNTTTTTAGNN
NNTANTNNTNNGANNCTTNNNCCNNGCCNAACNNNNNACATTTNCTNCGNNCCNTTTTTT
NTNNCTNTNCCNNNTCNNNTTTTNATCNCNTCNNNNTNTNNTNNCNNTNTTTTTNTTTTT
TTGNTCNCACTNTCNTCNCTNANNTNANACTGCGCCNNNGCCNTTTCGTTTNTNTTNTTT
TGNCCNCNACANNGNTACCNTNNCNTNTCNCANAGCANNNNTNTCTNNTNNNNNNTTTAT
TTTTACCNNATATNTTTATATATCCCTCCNNGNCNCNNCCNCNNNTTATTTNTATNTATC
TNCCCCCTCTCTTATCTTANNNTNACCTANTNCCCGTNGACNTCNTGTCNCNCCCNNCNC
CCGNNTATGTTTNNNCTTNCCNNCNCTANNAANATNCCNTTNCNGTTNCCNNANACNGAA
NTANANCNCTTCAATATANNGNCATNNNCGCGCTANNNNCTCNNNGCTTNNTTNNCTCNT
NCAANNTNNNCTTNNATCTCNCCCANNNCGCANGTACNCACCANANTATNTNANCCNNNA
NATTNNNNCTNNTATNNNGGCTCCGCNAAATTANNNTNNAANNNACCGCCNCNNCCNTNN
CNNANNNTNNACNNACNTCTNTNTTNTNTCACCCNCTNTAGTANCNTNTATGCNTGNCNT
CNNGNANAGGNNNNACGNNC
>A300
GNCTAANAANNNNGANTGTAAGNNGCNGNNACGGNAACAANTGNNTGAAGNNGCGCCANA
NGCGNNAGNGNCCCCTTNNAAAGTCAGCNGANNATNCANATANTAACTTGNANNNATGNA
GCAAACTNNCNNCNNGACGCTGGTCTTGNNNCNAAAANGGGCAACACCTCNCGCGNTTNN
GANCGCNCNNCNNNCTTNNNNGGNCCGNGNNNGAACNGGTTACTNTCACAGCGNNNGNNC
ANGAACNNGGATCTGNNNACGANACAGANGATNANNNCTTTNTTAGANACGNAGAACACG
AGAACAACNANNGGCGGGGNGTNAANGCANGCNGGAGATANANCNACCNANACTNTCGAN
NNGNAANACCANAANCCNNNGTACCCCNCNCNGACGNNNTNAATGTANCGGNTTGCTTGG
TGACAGCATNNTTCTCCNNTTTNNTGNNNTNNANNCNNNCNNACNCCACNCNTTTTTATT
AACNANCCNCCNTTNNNTNNTTTNTGTANCCNCTGTNCCCCCCTCTTNACNTTTACNNGN
ANGTCCGNNGTAGTNGTGTNTGNCCNTNACCTTAATATNNTTNTNNNCACACAAGNCGNG
NCNAAAGGNATATAANAAGGANGAGTCGATGAAAATNTCNGCNCCTNGTTATNCNCCTNN
AGTNGTGTCTTGTTGTGTANTCANNGTNGNNCAAAGATNANNNGNAGAGACTNNNGTGNA
TGGTNNNNCTTNTNNNGNNANGNGTGCAGATTNNTTNTNNNNNANAATNNCANGAANTAA
TTTGNCNCGGNNGTNNNNNATGNTATNTTNTNNCANCCNTNAACNANNNCGCNTNTCTTN
NGTTTNTNNATNTTCGNTTGCNGNTAAANANNNNCNNNANGCNNCNTTTNGCCNGCTCNC
CNTTTTTTTTNTNNCNTTNNCCCATNCNATTTTTNNTCCGTANGGNNTNNCTNNNGNTNT
NNNNTNNTTTNNTTTTNTNNANNCNTCCNTNTNTTNCCCANNTNNTTNNNNGNTANNNNT
TTAGTTTTTATNTTTNTTNNNCCNNNCNANCCNACNATGNNCCNNNCGNNGNGAANCNTT
NTNNNNGGGNNNNTNATNTNATTTNCCNCTTTCTNNTNTCTNCACNTNNCNCCACCTCNT
CTCCACANTTTTTTTANTATCNCTCCCGCCTCATNTCATTNATAAACCGACCNCCCTATT
NTTNNATCNCNCNCNNCACACNTNNNANNTNGCNNTTTCCTCNCNTACNANANANNTTTN
NNTCCNNCNNNNCACGCANNNTNCTCTATNTTTATNTNTNNATGTTCGNTNTTNNTCCAN
ACAANNTNTNNTANNTNGNCNANNTNCTAGCCANAANNNTTNTCCCNCCNNNCNGTNGCC

ACCNTACNTACNTNTGNGGNNCCATNNCNGNTNNAATATTNCCTNACCNNNNAAATTTNN
ATGTGTCTNTCCGCTCGCCCNNCNNTTATTTCAACNNCGCNATTTTTAATANNACNCCAC
ANNNNTNNCNTGATGCGNNTTNNTAATNANNATNCCGCN
>A301
GCGCCGCTTACCATGCAAGTCGCACGAGAAAAGGGGCAACCTGAGTAAAGTGGCGCACGG
GTGAGTAACACNATGGATCATCCACCTCTTTGTGGGGAATAACCTGGGGAAACCCGAGCT
AATACCGCATAAGCCCGAGAGGGGAAAGCAGCAATGCGCAGAGAGAGGAGTCCGCGGTCG
ATTAGCTAGTTGGTAGGGTAAAGGCCTACCAAGGCTATGATCGATAGCCGGCCTGAGAGG
GCACACGGCCACACTGGCACTGAAACACGGGCCAGACTCCTACGGGAGGCAGCAGTGGGG
AATCTTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGGGTGATGAAGTCCTTC
GGGATGTAAAGCCCTTTCGACAGGAACGATAATGACGGTACCTGGAGAAGAAGCTGCGGC
TAACTACGTGCCAGCAGCCGCGGTAATACGTAGGCAGCAAGCGTTGTTCGGAGTTACTGG
GCGTAAAGAGTATGTAGGCGGTTGGGTAAGTTTGGTGTGAAATCTCCCGGCTTAACTGGG
AGGGTGCGCTAAAGACTGCCTGGCTCGAGTGTGGGAGAGGAAAGCGGAATTCCTGGTGTA
GCGGTGAAATGCGTAGATATCANGAGGAACACCTGCGGTGTAGACGGCTTTCTGGACCAT
AACTGACNCTGAGATACNAAAGCGTGGGTAGCAAACANGATTAGATACCCTGGTAGTCCA
CGCCCTAAACNATGCATACTTGGTGTTGGCCATTCANTTGGTCANTGCCNGANCTAACNC
NTTAANTATGCCNCCTGGNGANTACNGTCNCANGNTNAAACTCANANGAATTNACNGGGG
CCCNCNCANNCNGTGGNANNNTGTGNTTCAATTCNACNCNNCNNNNANAACCTTACCTGG
NNNTNNNANNNNNATTCAANNNNNNTNNAACNTNCTNCTCCNCNNNNANNTCTTNTCNNN
NNNTNNNTGNTTNNNNNNNTNNNNNNNNNNNNNTTTNNTTNANTNCNNNNNNNNNNNNN
CCTTNNTNNNNTTNNNNNNNTTNNTNNNNNNTNNNNNNNNNNNNNNNNNNNTTNNNNN
ANNNNNNNNNNNNNNNNNNNNNNNNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNN
>A302
CCNNTTAAAAAAGNACCNGCNGGCNTCCGGAATAGNNANCNGCGTGAANATGGNGGNCTN
ATNCACGCAACNGGNTTNNACNCGTANAAAGANNNNTNTGGCCANTNAGAGGAGGCNNTN
CGCTTGNNCTNNCCCNGANCGGANGGGGNNATANNNTCNNTNNAGGTACACGATGGTNAT
TAATGGCCNGAAAGGACGAACAATCNNACTGGGNCCGCGANACGGGCCAGATCCCAANGG
GAGGTNGTANTNTACCAACACCGNACNNNGNGGTGAACCCCCTGTNCCCTTNCATGCCAN
CGNGANTGANNNNNNNCATTTCTTACCNCTTANGCTTTTTTTANNTNGCNCNACCNCNAC
TTTNTTCNNCNCNACNNNNTCNNNGGNNTNNNNNNNNTNCCANNTNCTNNNGNCTNGANN
GTCGCGGNNNATNGAAGTNNNNGTNTTNNNCTNNNNAGGNNAAANAAACNCNCNNNTNTN
ANTANCCCANANCNNCNNGAAAAANGNNATGGCNGNNNNATNTCGCNNANCGTGCNCNTT
NANCAATNNNCNCNACTANTGNGTTATNANNNAAGAGNTGGCNNATCNTNNNNNCGCGGN
ATGTNNNCAANANGANGCTNCNNAAGNNNGNANNNANNNAANTANNNNTGNNNNNGNAAT
NNAANNNNNAGAACNNTNTTTNGAANNNCANCANGNAGTNNAAATNCNNNGNNNNNAANN
ANNAGCTTTCCCAGNGACNGTTTTNTATANNCNTCNCNCNCCNCTTNCTTACNTNTNNCN
TNNNAGNCNTCCNNTGTTATTTNANTTAANCNCGCCCCNTATNNGGCNNANNNNCNGNNC
NAAGTTTCTNNNNTANTNTTNGNCNNANGNGCAANANNGCNGANGCNTNCGANCNTTTNA
GGTNNNCNTNTTTTANACNNCAATANNNNNGNNNCNNACCCCGCNTNNANCGANNNNAAG
ANTNTANACACGCANTCAATAANAAAANNNGCCACGNTCTTNNNNTCNNNANCNCCCCAC
GNNCNTGNCTTCTCTTCCNGNNGGNNTANGAAGANTNNCNNCCCNTNCGGGNNNANANAC
NNAGNNATATCNTTNTNCNNNACANNNNAGNNNNNAANTNNGTGNNCNNAGATAACAGNG
GGNANNAGNCAGANNCANGNNNAANAANTNNTNAANGGGATANNNNGNAAATAGGGTAAG
NGANAAANGCNAGANTGTTGCAGTNNANGNCANNTANNNNGANNANNGGNNNNNCAAAAC
ANGNCAANANTATTANTNANGGNTCNNNGCNANANAGTCNNN
>A303
GGGNGGATNAAANNAGGNNTATNTNGNAAAAGGNGNGANNGNAAAAGAANANGTTGNNGN
NGTACNGGCGCTGAAGCTAAANGGCCCCCNCTTTTAAANNGNNTNNNNNNGANNCNCNTG
TAAACAANTTANCGNGCNNAAAANCCGAGCNGTNNTANATANCANGCAACNGNTATNANT
TNTCNCGNTAATAGANNANCCCCANTCNTCCCNNTNNGGNNGCAGCNCTTTGCGNNTNTT
TATTATANACNCANAANNGGGANGNGNGNTAAACGNNATNNNAATCTCAGGNNGCACGNA

TNNANNGNGTNAATGAACCNGCANAGGNNNCGAATTANCTCNTTNNTTGNCNNNCGCNGT
ACNACGCNCGACTCANNAGGTNCAGGAGGGGTAGTNNAGGNTNTTGTGTGCCAACCACCC
NNCCACNNACNGGNTNNTANGCCCCGNTGTANGCNNCTTNTNGATNGNCNNCGGGTTGAA
NNTGNANGNAAAAGGNTCTNTTNNNNTNCCCGCNCNNNNNCCTTTNTTNTNNNTTTTNTN
GCCTNCNCACNNNNNNCACTTNNTTNNAGTCGNNNNGNNCNCNAANNCNCNNNCNNGNAA
TANNTNCNCNNNTANNCTCNACNTGNNNNTTNNGNNNTANAGNANCGGAGCGCGNGGANG
NATANGANGANTNNTANTCTNTGACNNNCNCNGTTNGGAAANGNCGANNATGTNGTNGCA
GNCNGNCGCNGTATGTTNAGCCCNNACGGNNGAGGAANNTANNCANACNNAATGNNGCCN
GNNNCNAANTTNTNNTNGANNNCANCGNNCNNNNNAANNATACGCGNTCATNANTAGNGN
NNGANAGNGGATTCGANAAGTANCCTGNNGTGNNCGTNCCNCTNAGNNNGNNNNTTNNNC
TGAAGNNNTNAANCNGGANANANNANNGAGCGNATATAAANANTNTGGANGGGNTNNCAA
NAGGNANGNNCNAANCCCGNNGAACNGNNGTTTNGGANAGCNATAANNNATNNGGGNAAT
GTTATGACNNGTNCTNNNTGANNANCANTNNANCCNANNACNCNTNTTTNNTTNTANNGC
NTTACNCNGNNNCNANNTATGTNANCAGNGTGGGCGCTNAAAANNTNANTCNTNNGTNAT
TNTNATNTATANNCGNNGTCCNTNNCNTNCGCNNTNNNNCATCNNCGGCNACNCNNNTTG
CTGNTTNNTAATNNNNTANNGNCGNAANAANNCTNGAACAANNNTNAANNGGTNNNNCNA
NTTTNNNNGNNNCCNCNNTNTNTTATCCNGNCTAANATTGCTTNNNCGGACCNGNNGNCN
ANNCCNCCGTGTNANNCAATGATNNCANNGCCCNCACTNTCNANTAATACAGGNNCCGNN
GCNTTNAAANTCNTANNNAGTNCNANAGGANNNNNNGTAGAGNCNTGACGAANNGGGNAN
NNNCATAANGNNANNNNAGTCTNNGANNCGTCAGGTNNGCGTCGATNTAGGGNCANAGGT
TNNNGAGGCNGAANCAGGGCAANGNAAANNANNNNAGATAGCGNGNGANNTTGATACNNG
NCGGAGNGNGATANGNAGCAANANCNATCGNANNNGANTNCGNNAATAACAGNGNGNGAA
NCGNANAGTTACNGNGGNGNGNAANNTTAANCTGNCNNTTNNANTCANNGGNGNTNCGAN
GNGNNAAGCCTNACNGTNNACNGANATNGTANGCTNNNAAAACGANAAGNATNAGTGNTN
GTGGNGTGAAGTGGNANANAGNNNATNAN
>A304
GGGGGGNCTNANAAAGGGNCAANTNTNGTTAAANANGGGTATAAACCGAATAATANATTN
ANNTGNTNGCGTGNNANNGNTGACAGGGGCGAAGANCGCCCCCTTTTTNNNNGGATNTNA
ACTNTNNNNNANACCNGTCGTNNTCNNNTTNCGNNGCNCTNTANNANGCNACCCNCNTNG
TNNNNTNNANNNANANCTNCNTNTTNGTTTTNNGNANNANTNTNNNNGCCATCGCCCNTG
ANTTNCNCAGTTTNCNCANGGNCNGCGCTTGTCCCNCTNATTATTCNNNNCGNCCATGCN
TCNGNNNGNCGGGCGATTNTNAGNNTNNCGAGNNTNTGNGNANCNACCNGGAAANNAGNG
GNNNAAGAAAGNANAGCANANNCNGTANNTCNAATGAGCTNNGCTATCTTTNAGGNACGN
CNGCNNTNCTGNNNNCGNNCNCNGTTNANNNGCCGGCTGANGGTATCNGGTNGGNTTNTA
NNNTTTNCNNCCNNNCCGGNCNCTTNCNGNNNGTNNTTNACCCCGGCCNTANTGNACGTC
NNTNNNGTNNGNCANGCAGAAGNTGNAGTGAGTNTCNNNNTANTTNCNTTCNNTNNNCCG
CCNNGTATCTNCCTNTTANTTNNNATNTATTCNCCTAAGNGCCCNNNCCNNCCTTTGTNN
TATNNNCNCCAGGNCANCTGANNNGANNNTGGGANNTTGATNNNAGCTTGGNCNGCGGTG
TTNTNNTNTANTAANNNACANANTNNCATNGNNTGNNGNNNCTTNNNANNNNGNGNNNNA
NNAGGTNNNNTGNANTAAACAAANCNNCATTTNANTGTNCNTTGGTTTGCTNNTNNAGNN
ACNCNGGAANGCNGCATNTNNNTNAACNNCGGNNCNNNNNNGNGNNNNTCNTNGANTCNG
CNGGAAGGGANCANGGNNNNTTANAANNNATANGNNGCAANNNNANNTNNGTNTTATNNN
AGGAAGGANATGATANNGNNGNAACAGTACNNTGGNGGCNGCNGACNGANTNTNTANCGG
NTTAANGGNNNNGNGGGGNGANAANAATNGGCNGGNNAAAACNNGTTGAANTGTAGNANA
NNGGNGCNGGNNACCNGGNGNCNNNNGNGNAANGNNNTNGTANNGATANCCTCANNGNCN
AGGAAGNGCAGNATNTNGAGAAGNNNTATAGNGGNGNNGAACNNTNGTGNGNGGNGNNNG
NTTTCTTATNANANNCTTTNNGGGNNNCGNGNTTTTTNTTAAACNNGTNNANNGGTNAAA
NNCGANNCGNNNANNNTTNTATNNTTNGTTTCACGCNNNNTCNNGTNNNNTNCGNNATNG
AAGCNTATNNAGNNNATNANCCNNNTCNTNTNTTNAANATNNTNTANCGNCAGGNGGTAN
NGNAAGGNCANNAGNNNAGATANATNTGNACNGGANTATATNGNNNCGCCGNCCANTATT
TNNTNNNGNGNNNCNCTTNATNTNTNNGCCNTNACGCANNGNGTANCGNNGCNTNGATGT
NNATCNTTNTTGGNCNACCTNCCNTCTTNTATNNNNTNNTACTNGCGGCAGNGGCTTTNG
CTTAGTTGTGGGGNAGNAGGNNTNANNGTANNNGNAAACGTNACCNACGCGGCNNCCTNN
NTANGAGATTANNNGNCCCTNTCNNGCAGNAANGTANNTNNNNTANAGNNTANNNNNTAC

ATNNTNGAACNNNGAAGANNNNGGTNTANGACTGCGCGATAGTGNCNGNGGGANNAGAAA
NGCGGGNAGGAAGANGGNAGTGNNGGNGTAGANANANGANTNTNANNGAANGANGGAAGG
AANTGGATGCGNNNANNAAGTGGNNGCGGANNAGAAGCANTGNGTGNGAATGGATGNANN
NNTGNAGCGAGNTGNGGNNNGNNAGNGCNGNNGCCNTATNTAAGAGCNAGTNNCNCANAN
CAAGACNGNNANGTAAGTGATAGCGNAGCNCATACAATANNAGNCATAGNT
>A305
GGGGCGGGGTANACANGGCGAAGGTTCGTNAACNANGGGGNAACCAGGATANNTGNNGNG
NTATTAGGTANAGGCGAAGANGATAAAAGCCCCCCCATTAANNANANNNNNNCGCGNAAN
TNNNCNTGTAAACGTNTNACNNGNNNTNNNNNGGGNCTTNNGTNANCNGCCACCCCNTNT
TTTNTNTCCCNCNTTGNCNNCCATCCACNGTCTTNCCNNANTACTCCCCCANNNNNTACN
NCTTTNTNNTANCGANTGNNATGNACNNNNGNTTNANCTNTANGCGCGANGNNNAGCAGN
NNGTACANGNAGAGAAGGAGCNANNCANCNGNACGCNNNNNNCNNCGCNANTNNNCNCGC
GGNGTAGNACTNCNTCTNAATTATANGCCGGANANTGNTAGATTTNCTNTNNTTNTNANN
GCNNGNNNNANGCACTNTCNGNCGTTAGGCCNCNNNNTNNTNGCTNNGTGTCGTGCCNNG
CGNTNTTNTTNAGTGNNCNTNTNNNNTGTGGTNNCCNCCNNGTNCGNCTTTNNNTNNNTN
NNNNTNNCNTNNNGNNNTATNCGNCTTNTNTTATNCGNNCANNNTANGNGNNNNNTGNNN
GNNNNNNNNNNNGNNNACANTACNNCNANNNGNTTTCTNNNNNNGANATCAGNANCTGTT
GNNGANNNNACATGNGNCATATNAAGANNNGNNGANNAATNNAAANNAANGGNNNNATCN
NNNNCANGTCGGGANNCNNNAANGATGNGNNNACGNNNAGGGATAATAACNCNNGGNNNA
NNGNACNNGNGACGANTNNGGCTCCAGNGNNAGAGNNNGCNAANCNATNTAGNGNGNNAN
ANTTCGNGCGNNNNNTNANAGGGATAGNNNTGNGANANGANAGTNANNANNNNCGCAANN
ANNNNATAGNNNNNNNNNGAANNNAGNNNANNACAATCGGTNNANNNAAANTNCNNGAAN
ANGGNGAGNNNNNGGNNANNAGNANNTGGAACGCAACANNANTATTNTNANAATGGNCNG
NNANAAGNGCGCTANANTAANNNGCNGNNGNNNNNNGCANNTGCANNNNAGNCCGGGAGA
GCNNTTTNTNTNTTTNNCGNNNNCNNCCTTNCANCANTNNTNNACNNCNNNCGNCNNTGN
ACATCCCNGGTCNGNTTNAGNTTATATNATTNNNCGCNAGCGNAGNGTGNGNTAGNACTN
AAAANGNTANAGAAGACACNTNTCCTTNTNNATANTNTATTCCTCCNGNATNACCAACNG
NANGANNNGGGGTNGAGACCGGGTNTNTGGNNGCNCCTCTGGTTTNTTTNNNNNCNNTNT
TTANNTNNNACCCACCNNCNANNNNATCCGCCGNNGTAATNAGATATANNCACNNCGCGC
TNATTGNGNATNNNANNACGCGNAGNNNCNTTTCNTAGNTGNCGGANAGCAGGNANNAGA
NGGTTGGANTTNNNGTGNCCGNANCNNNNAAGNANTGNANNNGNNNAGNGANNTNNNCAG
AAGNNGNNANNATANANTNANTGTNNNNNNNNACGCGCANNCGCNNACNTGTCNGGCGATC
GCCGGNAANAANAGTGNNGNCAGNCNGNNNGNCGCNNTGNGNANANAAGNCAGGAGANNG
NGCCNTGTGNGNGCNNANTNNNCGNNCGAGGANGAANCNTNGNNGNTNGANNNANCAANC
NAGANNGATNNNGNNANCGNNTNAGNANGNTANNNGCGNNANNACGNNNAAACNNGNNNG
NNNNAAACGNAGGTTCNA
>A306
NNNNCCNATTTNNNNNNNNGANNTCGCNGCTTCCNANGCAAGGTNGAACGCGTNGTAGCA
ATACACGAGTTGGCGCACGGGNTGAAGTAACGCGCTGGTATATCTGCCTTTTGGTTCGGA
ATAACACCGGGAAACAGGGGCTAATACCGGATGGTTCCTTCNGGATAAANATTTNTCGNN
NAAANATGAGTCCTCTTANNATTACCCNGTTGNTNANNTAATNNCTCCCCNNNNNNNCNC
CNNCNANCNNGACACANCNNNCNNCCANNCNCNCTACTNNNCCCNGTCCCTTCCTCNNTT
TNTCCTTTTTTTNTTTCTTATTTTTCNCCCNCCCCANTCGCTTTTNCCCCNTTCCCCTTT
CTGCCGCGATTNNTTNNTTTCTTTNTTTCNTCCCCTCTTTCCTTTTTTTTTTTTCCCNCC
NCCTCCNCCTTTTTTTNCCCGCCCCCTGNTTNTTNTTAGTTTNTTTCCCNNTNCTNCNAT
CTCCTCNCNTCCNNGCGTTTTANNCGTCCTNTTAGNNTNTCTNCNGCNCGNNTTGNCNCC
GCNNTTCTTNNCTATCNNCNNNCGNNTTNTTCNNCNCCTNNNCCNNCNTCNTTTTCAGTN
NCNCTTCNCCCTNCTTTTCCCCCCNCCNTNNTTNTANNTCGNTNTATTCTANTTGGTGCC
GCCCGNCNGTTNTGNCNTNTTTNNCNCTNNCTNCCTNNNNTTNNTNTGATTTCCTTNCCN
TNNCCNCGCCTCCCCCGTCNTTTTANNTNTTNTCNCGGTNTTCNTCNTTCCACCAGCGTC
NGNNCTNCNTNTCCNNNTCNCCTTTTTTNTTTNCTTCCCGCTCCNCTTTTTTTNNCTNGC
CNTTTCTCCNTCCCCTTTTTCTTTTTTTTTNCCCCCNCCTTCTCCCCTTTTNNTNCCGNCT
CANTCCNNTCTTTTTTTTTTTTTCNTCCCCTNTCNCNTCNNCCTTTCNNTCNNTNCCNCT
TTATCCTNCGCCCTTTTTNNTNNCNCTNTTTTNNTTTNCCCCCCCCCCCTNTCCCCCTCT

TTTTTTTTTTTTCCCTCCCCTCTNTNTTTTATNCCCCCCCNCTTTCCTCTTTCNNTNCCN
ATCCCTTTTTTTGNCTNCCCCCCTCNTCCGNCTTCNNTTTCCGTCCCNNTTNCNCTCNAG
TCGGCTNNNNNTTTTTTTTTTCNCNCGTTNNCTNCNCCNNTCTNCNCTANNNCNNCNCTT
NNANCNCTGNTTTNTNNCGCGCTNNNCACGTCCCAGTATCTNNTGCCAGCNCTCCCATCT
TNGNCCCCCCNGAACTTGCNNNNNGTNTTTCNNTCNCGTGCCTGTCNNANAATCNCGGTC
GCTCNNGATTTTCCATCNGNNCACANNTTNANTCTCNTCGGTNNNAGNNCAGTNNTTNCG
>A307
ATGNGCGCTACCATGCAGTCGAGCGCGTGGCAACACAAGCCGGCAGACGGGTGAGTAACG
CGCTGGGAATCTACCCATTTTCTACGGAATAACGCANGGAAACTTGNTGCTAATACCGTT
ATGAGCCCTTCNGGGGAAAGATTTATCGTGGAAAGGATGAGCCCGCGTTGGATTANCTAG
TTGGTGGGGTAAANGCCTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGC
CACATTGGGACTGAGACACGGCCCANACTCCTACGGGAGGCAGCANTGGGGAATATTGGA
CAATGGGCGCAAGCCTGATCCAGCCATGCCNCGTGAGTGATGAAGGCCCTAGGGTTGTAA
AGCTCTTTCACCGGTGAAGATAATGACGGTAACCGGACAAGAAGCCCCGGCTAACTTCGT
GCCAGCAGCCGCGGTAATACGAANGGGGCTAGCGTTGTTCGGAATTACTGGNCGTAAAGC
GCACNTAGGCGGACATTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGNAACTGCCT
TTGATACTGGNTGTCTAGANTCCGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAA
TTCGTANATATTCNNAGGAACACCAGTGGCGAAGGCGGCTCACTGNTNCGGTACTGACGC
TGAGGTGCGAAAGCNTGNGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAA
CGATNAATGTTAGCCGTCNGGCAGTTTACTGTTCGGTGGCGCATCTAACGCATTAAACAT
TCCGCCTGGGGAGTACGGTCGCANGATTAAAACTCAAANGAATTGACNGGGGGCCCGCAC
AANCGGTGGANCATGTGGTTTAATTCGAAGCAACGCGCAAAACCTTACCAGCCCTTGACA
TCCCNGTCGCGGATACGANANATCGTATCNTTCATTTTCGGCTGGACCNGAAACAGGTGC
TGCCTNGGCTGTCCTTCANCTCNTGTCGTGANATGTTNGGTTTAATTCCCGCAACNAACG
CANCCTTCNCCCTTANTTGCCNNCCATTTNNTTGGGCNCNTNNAAGGGNACTGCCNGTNN
ANAANCCNAAAGAAAGGTGGGGANNACNTNCANNTCCTNCNGGNCCTTCCGGNNGGCTTC
CCCCNTNCTNNCATGGNGGNNAANNGGNNNNNAAACCCCNNGNTCANNTNATCCCAAAAN
CNTNNTNTTTNNAATTCCTTNNNNNNCGGGNNNAANTGGNANCCNTNNTNNCNNACNCCC
N
>A308
GGGAAGGGGCGGNNAAAGGGNGGCAGNTCGNNGNCGGGGTACACAGGTATATNNTTATNG
TGNTACGGNAAGNCCAAANGCNCCCCCTTTNANANGGGNNCTGGTANNNCTNAGGNANTN
NGTNACTCGGGGGNANTCGNAGANCNTCNANNATNCNGNNGNNTGGTTTNTNTNGTCNCN
NNAATTAGNNTNNGCCGCCATNNNTTCNCNATNNNNNGCNCTANCNCNNTCCGNCCTNTT
TANTCANCCNTCAGNNCCCGNCNNGNAGGNGATNANNCNNNTNNGTNNTANTNTGTTNNA
ANGGANTCTGTGGGNGTAAAGGNCNGGAGCNNGACCANATNCCANTNNANCCGNNGCATC
AGNAGGACNNNGGCGNAACACNNAGCGAGCANNGATATGNCCGGNTCTATGTNAGTGNNC
TNNAATNNTNAGCNCACNACCNNGGNNCAATNNNNNGTNTGTNNANCCNCCNTATTNNCN
CCNTTNNNNTGNGCNAGGCGANAGAANTTGTATNNANNGANTTCGTNTCCCNANACCCGC
TNTTTNCTCCCCTTTTNTNTGTTTNTTCCNNNCCNNACNCNNNNNNANCNATNTTNTNTA
TNCGCNNANNANCGCTGTNTNNTNTNTTTNTTTNTTNNNNTCNTGNNNCNNNTCCTTNNC
CTGTTNNACGGTNAGNANNCGACAGTGGNNTTATATAGANANGNNAGNNCNCTNANTTNT
TATNNTGNNAGANNNNGGNTNGNNGAGTGNNNAGANTTAGNAGNGTNANNGAGNAGAAGT
NAGGTACGAGTNNTNANAAATGANAGTNNGCATANNANNCNNNNNTNGGNANTTGNATAN
TAAAGNNGTNTGTNGGCNGNNANNNNNNNATNGGGAANGAAACGCANTGNNGAANNGNAC
GATGGANGCTNNNNGGTCTNNAAGNACTNGCANNNANGNAGNNGGATTTNANNCNTNAAA
GTNTNANNNNNTGGGNNNACGAANNGANNNTANGAATANTGNTNNNTGGNGTGTNTNGNT
GCANAANGGNTACGNTNNCNNNGGNCTNCGTNTNTCTCNGGTNAGANANAANGGGNNTTG
TTGAGNGTAGGTTTCTGANCGANNNTNNGCCANCAACGGNNGCTTNATCNGGNGGNGGAC
CNNNTTNTTGTTTTATANCTTCNCNCCNCGNGANCTNTGNTTATNNNCCGNNGCGNGANT
TNNANCNNNGCANCGGNTNTGTTATTTNTNTNANNCNGNANNNNNNTATNCTCNGNCCTA
NNNANNTCGACNGGNAACNNGNNGTTGNTATTATTATTATTNNTNTNNCAGAANGACNNN
NGCGGGGNGAANNNGCGNGAGCGTGCTNNNNANNNNGTTTATNNGCNGGNNCCCATNGTA
TNGTNTTCNNCNCCTNNTATNNTANNTTACACCGCNCGGTCNNCNATTNANNTGGNNANA

NNNNNGATNTATTACNGNNNCNCGNNGNTCNNATANNAATNANTNTCCNCGGNNNNNACT
TNNTNNTNNNGTNANCGANNCGNAATNACCTGNGTCTTTTCNTCNTNNNNACCNGGTANG
NNANNTGNGAAATNTGNNACNGNNGGNNCCGTCNCGCGNAGNACGTNNANTNATNNAGNG
NNGTNNNCGNNNANAGCNANGNNCNAGACNTTGNTTACNNACGCGGGNGTNCNANNCGTN
GANTCAAAGNGGAANNGNANCTNNANAGACGTNNANANATNNANAGGNGGANAGGGCCAA
NNNANTNNCGGNCNNANGACGTAGNNTNNTAGTCGNGNNNATAGNNANCTCNGNTNNTAA
GCGTAGCGNTATNANTGANTNNACCACNTGNGAAGGTAGGGTNNTCCNTNCGNGGAGANT
ATNNNATAAGGCN
>F1
ANGCGCNGCTTACCATGCAAGTCGAACGCGTGTAGCAATACACGAGTGGCGCACGGGTGA
GTAACACGCTGGTATATCTGCCTTTTGGTTCGGAATAACCCTGGGAAACTAGGGCTAATA
CCGGATGGTTCCTACGGGATAAAGATTTATCGCCAAAAGATGAGTCCGCGTCCGATTAGC
TAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCGGTAGCTGGTCTGAGAGGACGATC
AGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATT
GGACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTG
TAAAGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAACTT
CGTGCCAGCAGCCGCGGTAAGACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAA
AGCGCGTGTAGGCGGTTGTTCAAGTCAGGTGTGAAAGCCTTGAGCTCAACTCAAGAAATG
CACTTGATACTGGATGACTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTGGTG
AAATACGTAGAGATTGGGAAGAACACCAGTGGCGAAGGCGGCTATCTGGACGGTTACTGA
CGCTAAGACGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGT
AAACGATGGGTGCTAGACGTTGGCGAGCTTGCTCGTCAGTGTCGCAGCTAACGCGTTAAG
CACCCCGCCTGGGGAAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCG
CACAAGCGGTGGAGCATGTGGTTCAATTCGACGCANCGCGAAGAACCTTACCAGCCCTTG
ACATGTGACTCGCCNGGCTCCAGANACGGANCCNTTCGGTTCGGCCGGAGTCAACACAGG
TGCCTGCATGGCTGTNCGTCAGCTCNTGTNCNTGANATGTTGGGNTTAANTCCCGCAACG
AGCGCCAACCCTCGTCTCCNGTTGCCATCAGGGTATGCTGGGCNCNTTTGGAAAAACTGC
CGNTGAACAACCCGANAAAGGTNGGGNATTAACTTCAATTCNNCNNGGCCCTTNCNGGGT
TGGGNTACNNCNTTNNAAANGGNCNTGAAAAATGGATTCCNATGNCCNCACCCCAAACCN
NANCNNAAAAACCTCTTCTANTTGNAANCNCCNNTCCCCCNNGGTTNAATTG
>F10
NGAGGGGGCGNTAAAAAATGNAANGCCNGAACGGCCCGNGNNNAGGGTNNANTTGGGNNN
GGANGNGAAAACGANCGNGCCNTANNAAGGNNGACNGGNNGGTAGCANANNNNNNCTGNG
GAGGNNNNTNNTNNNGNANACNNCATNANTTNCCCTNTTTTCGNTCGCNGAGGTTCNATA
TGCTNGNNANGNNCGTNCCCTTGTGCCCCNNNTACGAAAGGGGGNNNGAGGTTCNCGNTN
TAANAACTNGNNTTAAAGAANAAANGNNACCACCNNGNACNCNTTCNNTNCCGCGGGNTG
NGNNNNNNNANNNANANNNNNANGNNNNNTNTNNNCCCNNNCCCACTTNNNCNTGTAGNN
CCATGTNCGCNNNCNACCNCCTTNNNNNGAGGNTNTTTNNNATANTTCCCCTNTCNNNGG
TNNNANNTTTTGNCNCCCCCCTNCNAACTNNTGTNTNCACNNNCACGCNTGNNANNTNNA
TNATNTNNATNNCCTCNTCTNTNNNNGNACCGNNCNTNNNNACNNTNTNNNTNCNNGTGA
AANAAANTNCCGCNGNTTATCANANNGCNNNNNNGNCCNNNGTCCNTANCTNANNNNNTA
CGNTCNTAANAGCGCNTAGNNNGCAANCNATATTCTCTCANNGTNNGGCNAANAGNNNAN
CNCCAGNATGNTAAGNTANTAGANGANNATNNCACANATAGTCNCGNGCNGCNGAAGNTA
NNGGANNNCNCCCTGCANTANNNNCGGNTTATTNNCTTTNNANNANNNGNCNNAACNNNG
NGNCCCAGAANNNANNTTNGGTGGNAGANGTNTGNNTNNTATNNAGNNNGTCNGNAGNNC
CNNTTNCGAANGGNTTTTTATTNNTNGNNACGGNNNNGATNNANNGNCCTNNCNGCTNNA
ANCNAGNGTNNNNTATNTGAANNATNNNNANCATNGGNTGCCTNTANACTNNNTCNANNN
NTNGTACTNTTTNNNTNNTTNNTNGNNAGANNNNNACNNCCGGTNGNTNTNNAGCNTNNN
NNNNTNCAGCNTNNTATANNNNNACNTNTNTTNCGCGCCCCCNCNCNCANNNGACNANAN
AANNTNTACNANNNCNCNCNTNCNANTNANNGNGGCCGGCTTNCGGTTANGACNNCAACA
CGCNNNNNNGNTNNCTTNNNAANCATTACGNANNTACANNAAAGANNGACNGAACANTCA
AGAAGGANNAANTGNGANTTATGANCCCNNNCNGTGAANAGAGNNGTNNTAGGGCCNNNA
AACCANANNNNAGNNNNNNNTGGNACATNNNGATNAANGNAAGCNNGATACNNGCANAAN
NCGNGCNCCANNNNNNTNNNNNATGTANGAGCNCNANNANTATNAAACNACGCNAGNTNN

CATNNNNANNNNNAGGTGGAAGNCGTGCNGNGAGNAAANGANNGNGN
>F11
NAGGNGCGGCTACCATGCAGTCGAGCGGGCGTAGCAATACGTCAGCGGCAGACGGGTGAG
TAACGCGTGGGAACGTACCTTTTGGTTCGGAACAACACAGGGAAACTTGTGCTAATACCG
GATAAGCCCTTACGGGGAAAGATTTATCGCCGAAAGATCGGCCCGCGTCTGATTAGCTAG
TTGGTGAGGTAACGGCTCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGATCAGC
CACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGNAGTGGGGAATATTGGA
CAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAA
AGCTCTTTTGTGCGGGAAGATAATGACGGTACCGCAAGAATAAGCCCCGGCTAACTTCNT
GCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATCACTGGGCGTAAAGG
GTGCGTAGGCGGGTCTTTAAGTCAGGGGTGAAATCCTGGAGCTCAACTCCAGAACTGCCT
TTGATACTGAAGATCTTGAGTTCGGGAGAGGTGAGTGGAACTGCNAGTGTAGAGGTGAAA
TTCGTAGATATTCGCAAGAACACCAGCGGCGAAGGCGGTTCACTGGCCCGATACTGANNC
TGAGGCACGAAAGCGTGGGGAGCNNANANNGNNTAGATNCCCTGGTTGTCCCCNCCCGTA
AACNATNAATGNCAGTNNGTNNGCNGGGTTTACTCACTAGTGGCGCACCTTATTTTTTTN
TATTATTTCCNGCCNGGGGGAGTNCCNGTCTCNAGATTTAAATNTCCNNTNTNNTTTNAC
GNTGCTTCNTCCNCCCNGCNCCNNCNANCATGTGGNTTTTATTTCCGCACTCCANATTCT
TTNNNANCCTTNCCCNNTTCCTTTGTCNNTCTCCGTTCNCCGNNTTNCNCNNGACGNGNG
TTCNTTTNTCTTTTNCCCNNANCCNNCAACANGTTCCCTNTNTGNTTNATNTATATACCC
CNCGTCGCTGNNANCTNTCNGTTNANNCCCCCTTCAACCCNGCCGNNNCCCCNCCCCCTN
NGTTCNNTCCCTNTTTNTNTNNNCCTCCTTTTCNNNNTTANTNTCCTNCGTNNCNNTTCT
CCTATATNNNTTGNNTTNCNNNNNCGCCANCACTTATTNNNCCCTNNCNNTNTTCNGNTT
NTCCCCACTNCTANCNNGCNTNTTAAGTNANCGTNNTGCAGTAANTNNCCNNCC
>F12
AGGCCCCTNANNNNNGNGGGCAAAAGGGNNNGATGGCAGAGNGNNGGGGGGAGAACNTCC
CCNTTNAAAAAAAGGGNTNANNTNCCANTNTTACNNCAGCGCANGTAACNAGTTNCNNGG
NNGTACGTCNNCATGGNANCANGNNNGANNAGCNCGGGGGCTGAGTATCACNTTANTGTA
NCNGGNTGGANGCGNTNAACNAACNCNNCCNCCTNAAGCAAGCNAACTANTTTATCTTGN
GCCCNGNNNANNCNNCCANCNANNANCAACCNCTGCNTTANTNGNCTNNTANAATTNACN
GCCNNNCCANNTTNNGGGAGNTNACNGNGNGNATTNNNNCTNTCGGNCCCNCCNTGNCNG
TNGNGNNTNGGNATNATGGATNNACANTNTNCGCNGNGGNNANGANTTANCCNCNGNNCC
NNNNNANTNNTGNATNANCNACCANCNGGNCGNNAAGAAANNNAGAAGNAGNNNCTTTTT
NTNGNNACGCGCNGNNACGTCNCNGNTNANNTATANANTGAGTNNTGNGAGGNGNNTNAG
ANCANNNCNNNNAATNGNNNGNNNNTGCNNANCGANAGCCGNNGGCNGANGGAAAATNNN
NCNCNNNGGNNGANCTNGANNTTGTGNANACNNGANCCCGNCNNNANNCANTGNCNNNNG
NNANNGNTNGNNANNTANCNAGTAAAAANCCGNAGANAANTCGACCNNGGNNGNGGNTNT
NANNCTGTGCNNNCGCGGNGNGANGNNATNNNCNGGATNAANANNNNNATANCNNNGACG
CGNNNNNNNGNCTNNCGNNATNTNAATNGTTGTNNNTTTAGNACCCGNGTANGCNNNGAT
GCNGACNGGNATANANCNNNAGNTTGCNCACGTNCTATTTNNTGNNCCTNTNNGGATNCA
GTNTNTNGGCNANNACTCCTTNNNNCTGNGCCNTTNANANTANGTNNNNGCNGNGGGTGA
GAAGCNANNNNNAGGCGNCGNNNAGNTTNGTTTNNATNTNNNGAGNNGCNCGAGNNANTN
CCGGGGNGGNCNGGNNNNANNTATGNTGGCNNANTGTNTATAGCNCCNAGATNTTTTGNG
NNNGCGNNNCTNNNCNGGAANCNGGAGNANTNAACNCAANNCNTNNCTNAANNNATAAAG
CNCCCGNCNGCTTTCGNNNNNANCNGNNGCAAANNNCCNCNGTGAGGTCTGTCCAACCNC
TNNNAACTTANNANNNANCNNACNCGGANGACNGANNATNAGGAGNTNNTNNGANGNNNN
NANNGNGACNNNNGGNGCNANAGAGGGNCNAAGNGNGAGCAAGTNTAAGNACAAGGNNAC
NGNNGCAGNNAGNAACNNANGANAATANGNNNNCNAANNAGGCCANNNANNNNTNCNTCN
GCGNGACCNANNTCGTAGNNTAGNNGCNNNGCACNCNNGAGNCNNTNANNNGGNACANCA
CNANNAANNGNNGAACGNACGTANAGAANGAGNGNAAGAGTNANNGGCAGNANGAANGNG
CTNA
>F13
TCGCNCCTTNCATGCAAGTCGAACGGCAGCACGGGAGCAATCCTGGTGGCGAGTGGCGAA
CGGGTGAGTAATACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCGGAT
TAATACCGCATACGCTCTGAGGAGGAAAGCGGGGGATCTTTCGGGACCTCGCGCTCAAGG

GGCGGCCGATGGCAGATTAGGTAGTTGGTGGGGTAAAGGCCTACCAAGCCGACGATCTGT
AGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATTTTGGACAATGGGCGCAAGCCTGATCCAGCAATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGGAAGAAATCCCTGGTCCTAAT
ATGGCCGGGGGATGACGGTACTGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCG
GTGATGTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGCA
TCGCTCGAGTATGGCAGAGGGGGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGAT
GTGGAGGAATACCGATGGCGAAGGCAGCCCCCTGGGTCAATACTGACGCTCATGCACGAA
AGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAACT
GGTTGTCGGGCCTTCATTGGCTTGGTAACGTAGCTAACGCGTGAAGTTGACCGCCTGGGG
AGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATG
ATGTGGATTAATTCGATGCAACNCGAAAAACCTTACCTACCCTTGACATGTATGGAACCC
TGCTNANAGGTGGGGGTGCCCNAAAGGGNANCCATACCCAGGTGCTGCATGGCTNTCNTC
ANCTCTGTCNTGAAATTTTGGGTTANTNCCCCAACNAGNNCAACCCTTGTCCNNNNTNCT
TCCCAANANCCTCCAGGAAAACTGNCCGGTNACAANCCGANGAAGGGNGGGNATNACCTC
AAGTCCCCTGGNCCCTTTTNGGTAGGGCTTNCCCCCTCTTNATGGCGGA
>F14
NANGGGNNGCCTAACATGCAAGTCGAACGCTCGTANCAATACGGGAGTGGCAGACGGGGT
GAGTAACACGCTNGNTAACGTACCTTCAGGGTCTGGAATAACCCTGGGAAACTAGGGCTA
ATACCGGATATCCGAGAGATCGGAAAGGCTTGCTGCCTGAAGATCGGCCCGCGTCCGATT
AGCTTGTTGGTGGGGTAATGGCCTACCAAGGCTTCGATCGGTAGCTGGTCTGAGAGGATG
ACCAGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAAT
ATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGACGGCCTTAGGG
TTGTAAAGCTCTTTTACCTGGGAAGATCATGACGGTACCAGGAGAATAAGCCCCGGCTAA
CTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCG
TAAAGGGCGCGTAGGCGGACCTGTAAGTCAGGGGTGAAATCCCAAGGCTCAACCTTGGAA
CTGCCTTTGATACTGTGGGTCTTGAGTTCGAGAGAGGTAAGTGGAACTGCGAGTGTAGAG
GTGAAATTCGTAGATATTCGCAAGAACACCAGTGGCGAAGGCGGCTTACTGGCTCGAAAC
TGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTANATACCCTGGTAGTCCACGC
CTTAAACGATGGATGCTAGCCGTCGGGCAGCTTGCTGCTCTGTGGCGCCGTTAACACATT
AAGCATCCNGNCTGGGGAGTACGGTCGCAAGATTAAAACTCNAAGGAATTGACGGGGGCC
CNCNCAANCCGGTGGANCATGTGGTTTTAATTCGAAGCANCNCCGCAGAACCCTTANCCA
GCCTTTGACNTGNCANGGCTGCTGGAGAGATCCAGCTNTTNCCCTTCNGGGACCTGCACA
CAGGGTGCTGCCATGGCTGTNCNTTCACTTCGTGTTCGTGAAGANGTTGGGGTNAAGTNC
CCGCAACNGANCNCNANCCANCGTNCCTTNNTTGGCCATNCATTTGGGTTNGGCCNNTCT
ANGGGAAACTNCCNGNNNTAANCCCAAAAAGAAAGGGNNGGATAANCTCTAAGTTCTNNT
GGCCCTTNCGGNCTGGGNTNAACCCCNNNCTNCCAATGGCNGGTGAAAAANGGAACGCTA
ACCCCCGNGGGGGTCCANTCTCNTAAANCCCGNTNTNTTC
>F15
ATTCGGCNTCCTANCATTGCNAGNTCGNACGGCAGCTCGGGAGCCTNGTCNTGGCTGGCG
AAGTGGCGAACGGGTGAGTAATACATCGTGAACNNGTTCCTGNTATTGGGGGGATNNCCC
NNNCAAAGCNGNATTNATACCCCNTACNCATCTGANGACGAAAGCTGNGGGATCTTTCNN
NANCTCTCGCTCAANGNNNGGCCTATGGCACATTAGGTAGTTGGTGNGGTAANGGCCTAC
CAAGNCTACGATCTTTAGCTGATCTGAGAGGACGACCCCCCACACTGGGACTGAGACACG
GACCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGCGCAANCCTGATC
CAGCACATGCCCGCGTGTGTGAAGAACGCCTNNACGGTTGTANANACTTTTTTTCCNCGG
AAAAAATCCCNGGGTCCTATATATCCACCGNTGAGATGTACTGTACTGAATAAANTATCG
CACNNACTNCACTACCATGCNCACGCACGCCTNATANAATACGATCCNGTGCGTATATCN
TAGATCTTANATTACTTTAANGTNNNCCGTTGCGCCAGCGCTGGTTNATGTCAAGGACCT
GATAATGACAATCCCCTTAACTNATACGCTGGTGAACTGTNNATTGCTTGAACTGGCATT
CGAATCGTACTATGAGCGTAAGGCGGGAATAGTAATTCCACGTGTATGCAAATGAAATAG
CATATGATCATCATGAGAGACNATATNCGCATNAGGGCAACNNTACTCCNCCTNATGNTN
NTACTNTACAGNTCACTGAAACGCAATGGCGGTNTCNGANCNAGNATTAGATTACCCTGC

CTNTGTCNACTCCNCTACCNCTAATACAANTNCTCAACTGTTNTGTCATGTACTTCACNT
GCGCTTGAATCACAAANCCNTTANGCTGTTACATNTTNCCGCACTNNCGGCNNTNCNAGG
NTNAAAACTNAAAACTAATTGACAGTTGGCCCGCNACCCCCCANAGACCATGNAGTNTAA
TTTNTNNNATCCATGCNGCACCTTAANCCTTACTTACCCTTNACNTATNTGANACCTCNT
CACAGTCNTTCGTGACCCAAAANTAGCNCTACANGGGTGNNNNCNGCCNTNCCCAGNTNT
NTTCGNTNAATTTCNGNTNAATCCCANNCCGNGAANCTTTCNCCNATTNGNNGGGNATCT
TTCNANTGCCTNCCCNTTAAAACNAAGANGGGNGNANTAATCATNNCNCCTTCCTNGGGT
NNCCNCCCTCATNANGTCAAANAANTTCN
>F16
AGGGGCGGCTTANCATNCNAGTCGAGCGGGCGTAGCAATACGNTCAGCGGCAGACGGGTG
AGTAACGCGCGCTTAACGTACCTTTTGGTTCGGAACAACACAGGGAAACTTGTGCTAATA
CCGGATAAGCCCTTACGGGGAAAGATTTATCGCCGAAAGATCGGCCCGCGTCTGATTAGC
TAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGATC
AGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATT
GGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTG
TAAAGCTCTTTTGTGCGGGAAGATAATGACGGTACCGCANGAATAAGCCCCGGCTAACTT
CGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATCACTGGGCGTAA
AGGGTGCGTAGGCGGGTCTTTAAGTCAGGGGTGAAATCCTGGAGCTCAACTCCAGAACTG
CCTTTGATACTGAAGATCTTGAGTTCGGGAGAGGTGAGTGGAACTGCTNAGTGTAGAGGT
GAAATTCNTAGATATTCGCAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCGATACTG
ACGCTGAGGCACGAAAGCGTGGGGAGCAAACAGGATTTAGATACCCTGGTAGTCCACGCC
CGTAAACGATTNAATGCCAGNNTGTTATTGGGTTTTCTCNCTAGTGGGCGCAGCTAACGC
TTTTAAGNTATTTCCTCCTGGCGGAGNACNGGNCCGCNNAGATTTAAAAACTTTAAANGA
ACTTNANNGGNGGCCCTCCNCNANTNNGNTGGNCCCATGTGGNTTNATTTCGNTNTNANC
NCCNCATNAANCTTTACCAGNCCCCCTNGACNTTCCCGNTTCNCNGNACTCCCNTAANNT
GAANTTCCTTCCCCTTCCGCCTNGCACCGTCANACNAGNGGCTGNCNTNGTNTGTCGTCN
ANTTCTTNTTCNTGAGATNTNCGGGNTTAANTTCTCCNNACGANNNNAATCCCCGNTCCT
TTANGTGTNNCCACTTTAGNTCGANCANNNCNNAGGNNANCTGNCNNTGNANATTNCCCT
AGAAANGTGGGGANGAANTTCNNTTCNCNTNGCCCCNTANNNGCTTGGNCTCTCCCCTCT
NTTTTNANGAGGTNAACNATGTGNTATCNTATGGGNCANNCCTTTTAAAATCTTAAAANC
C
>F17
NGGGGNAACNANANGAATTNTAAGGTNATNCNGGACAAGGGAAGTGGGGGTGNGANGGGN
ANAGGGGNGCNNTNANAAAANNNGNNAGNGNTNNTANGAAAACGNNGNANNGCNCCNGNT
TTANNTGGNNCACCNCNTATANNTTNCCGGTNANANNCCANGCNNNTCCATAGNGGNCNG
NGCTTTCGACTNGTTCCANNCANACACGCAGCGGNGNNNACTNTNACNTTATGTANACNN
ATAAGNTANATNACTGANAANNTAGCATGACTCANTTANTTNGGGCGNGGNNANGGGNCT
CCCNNATCANNCNTGGNTTNNTNNNGNNCNTTTTCCCNCCCCNCCACNNCNGNATGNTTN
CCGNCCTGNNCNNCTTANCNGGNCCCNTGNCGNTTNNNNTNCAGCTANATNCATNANCNC
CTNTTCGCNTTTTTTNNTTTTNNCCCACNCGCACNCCGNNTTATNTATTANCANCNCAAN
CNGNNANNANNTANNNNNNTTGNCNNNCTANTNNCTANCNNNNNNNGNAGNNNNGNTGNN
NAATAGNANNNNNNNANNGGNGTNNNATAACCAANNGNNNNCCACCNNNGNCTNATATNA
NGNGNGANCCCNNCNTAANAANAGCGNNCGTANNCGNANNCNTNTNGNNGTANACCCNAN
CNCGGCGTNAAATTNACNCCNNTANTCANTTTNTNTNANANNNGNTCATNTNTGGANCNN
NNNGCNCNGCNCNATATGTNNANNNTGGGTNAGNCGANNAGNNAATCGNNTAAAACACNT
NGNGACTNNTNATNNTANANCANACNNTCCCGNANTNATATNNGCNATATTNGCGNGNNA
GGNANNNGTGNCNCAGNNGANCANANNNCTTNTNNCCAGTNNACNNTNNTNTNTTNNNTN
CCCCCNNNNATNTTTTNNNCNTTNNCGTANANNCCTNCNNNNTTTGTTNTTNTTGNCTNN
NNNNNCTNANTCNCNCNTNNANTCGCNACNNCANNNNTCCTTNNTTAANNATTTCTTNNC
CANANCCAAGNCANNNNNNNNNCNNTNACNCTTTTANAGCCCGANTTTGTNTTNGNCNNA
NCTTNNTTNTNNCCCNCCCNNCTCCANCNGCNNANNGACTNANCTNCGCGCTCNNCTANN
NATNATNAACNCGCCGCTNTNTNCCTNNTNAANTCNANCANCCCTGNATTGCTCTNTNGN
NCNGNNCNNTNTATNATGACTNCCNNTNCCGNCANCAANNGGCANAATGNATAANNNTAN
ANNCNGNNNGGTNGNGGGNNGNGACNNAGTCATTCNTNTTNAANACNNNNACGACTNACG

GCGGNAANGATANNNCGNCAAATAGTANAANGCANNNNATNTNNNGTTAANANNCGGGNC
GNAAAAAGCGGNTAACTNANTNANCNGNATGNNGNAAANNNGANNAAGAGCGANCNANNT
ANTNTNANNCACTNNNCANNNGAATNGTGAAGTNNAGNGNNNNGNAGAAAAGANTANAAA
TNN
>F18
NTGCGCTCCTTNCATGCAGTTCGAACGGCAGCGCGGGGGCAACCCTGGCGGCGAGTGGCG
AACGGGTGAGTAAGACATCGGAACGTGTCCTGGAGTGGGGGATAGCCCGGCGAAAGCCGG
ATTAATACCGCATACGCTCTGAGGAGGAAAGCGGGGGACCTTCGGGCCTCGCGCTCAAGG
GGCGGCCGATGGCAGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCTGT
AGCTGGTCTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCAATGCCGCGTGT
GTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTTGTCCGGAAAGAAAACGCGCGCTCTAAT
ACAGTGTGCGGATGACGGTACCGGAAGAATAAGCACCGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAAGCGTGCGCAGGCG
GTTCGCTAAGACCGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATTGGTGACTGGC
GGGCTAGAGTATGGCAGAGGGAGGTAGAATTCCACGTGTAGCAGTGAAATGCGTAGAGAT
GTGGAGGAATACCGATGGCGAAGGCAGCCTCCTGGGCCAATACTGACGCTCATGCACGAA
AGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGTCAACT
AGTTGTTGGGGATTCATTTCCTTAGTAACGTAGCTAACGCCGTGAAGTTGACCGCCTGGG
GAGTACGGTTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGACCCCCACAAGCGGTGG
ATGATGTTGGATTAATTTCNATGCAACGCCAAAAAACCTTTACCTACCCTTTACATGGTT
CNGAAACCCTGCTTAAAAGGTGGGGGGTNGCTCTAAANAAAACCTGNCNCANAGGTNCTT
CCNTGGNTNTTCNNNANCTTNNTTNTCTNGAAATTTTGGGTTNATTNCCCCCANCNANCG
CAACCTTTTTTCNTTTANTTNCTNCCNCANNAATCNNTTTTGGGAAACTTNCCNNTTAAC
AACCTAAANAANNNGGGATNANNTCNATNCTTTNNNCCCTTN
>F19
TNNNNNGATNNGGCGNCCTTTNNCATNCCAAGNTCGAAACGNGCAGCNNGCGNGGTANAC
AACCCTNGGCCGCGACGGAGNTGGACGTAACANCGNTGNAGATAATCANACTCGTTTNNC
GNTNCTCCGTGTAAGCTGCGTGGGGATAANCNACGGGCCGTAAATACCCGGAATTAATNC
CTNCGTNCGANTNATGAATTGNGTCAAANCAANGGNTACCGTTCCNGCGCCTCGATTATC
ATNGNTNGNCNGGATGACCAGCATTACCATAGGTTGNCTGANCGTAAAGGNCTACCNAGG
NGACGATCTGTNNCTNNGTCTNNTNAGACAACNNCCNANCNTGTNNCTGAGNCACCGCCC
AGACTNCTNCGANANNNAGNNCNNCNNNANTNNGGACATNGNGNGNNACNCTNATNCNNC
TTNGCCTTGTGNNTNNNTAATGCCTTCGANTNTTAAAACACTTTTNTCCGAANAGAANAC
GCGCTANTCTNATNCACATGTGCCGNTGACNGNCCCAAGAAGAATAAGCANTGGTCTAAC
TTNCTGCCAGCAANCGCGGTAATACGTAGTTGTGCCGAGNTTTAATTCTGAAATTACTTG
GAGCGTTAAACTCNTGCANNTGGNTTNTCGCATTNGANNCNANTATNAAATCCCCGGAGC
TNTAACCTGAATAACNGCATTGGTGACTGNANNNCTAGAGNNTGGCAAAAGCANCTANAA
TTCCACGTGNATATATTGAAATNCCTCNGAGANGTGGANNAATACCGCNTTGGTTNAAGC
NNACCCTCNTGGNTCANNNCTNTANTCTCCNTGCCNNTAAAANGNNTTGGGGNGCNTCAT
AACTTGATTNAAGNCNNGCCTCGGTTTAGTTCTCCGTCNCTTNAANCCCATTTTCAANCT
NCCCNTCTCTGGGGGAANGTTATTTTCCCTCANTTTTTTGATANCCTAACNTTGTTTANN
CCNTNNCCCTCCCCGCCNCAATTTCNTGNNCGCNNTNTTTTCTTTNTCTTTTTANCCNNN
NTNCCCTTGTGGCACCCCTCNCCAANCCCCTTNGATATANCNTATCCNCCNNTNTCCNTG
CNAATTNNGGANCNNATCCGNNCCCTNCCCCTNGNCNTNNNANTCGTANNTCCCCTTNCN
NNNAACNCTNNTCNNTTCCTNCNTNNATANNCCTGTNCTTNAATCCTCCCNCTCACCANA
CNTCNCTCNTACNTTTCCCNNTTTCCTCTTNNNNTTTTCNCTNCCCNCCTTNTAANNAAT
TCTCTNTNCNCCCNTNCCTTNTNNNNACNNNCNCATCNANCTTANCTNCCCTCNCGACAC
TTCNTNNNNNNGTTNNGNCNTNCNTCNCTTATNCCCTCC
>F2
GGGGNGNTNAAATGNAAGTCGAGCGGAAAGGNCCTTCGGGGTACTCGAGCGGCGAACGGG
TGANNNATNTGANGAGTAACCTGCCCCTGACTCTGGGATAAGCCTGGGAAACNGGGTCTA
ATACCGGATGTGACCTCCTCCGGCATCNGATGGTGGTGGAAAGTTTTTTCGGNGGGNNAN
GGGCCCNTGGNTAANNATNTNNNGGGGGGGGAANAGNCCCNCNAAGGNANCAACGGGAAT

CCGGCCNGANAGGGNNACCGGTNATNCGGGGATNNANACNCGTTTNANACCCCCCCNGNA
GNNNNNNCGGGNNAANNTTGCNCACGGGNGGAAAGCNTGATNTCATNAANNCCCTTNGNG
TTNATAANGNTNCNCCGGNTNNCAACNNNTTTTTCNNTNNGGNANAAAATATNATTTNTN
CTCCCCNTAANAAAGTNCCCNNGTTAACNCNNTTNCNACATTGCNNTNGTANNATTNNNT
NNNGGNNNCANTCNCCCNCCCGGCATTNCNTCNGGCTCCCNAACGNAGTTCCNTNAGGTN
GNCCNNGGNTNATTCNTNTNACCCTTNANCNANNNCCANNNCCGNTCANCTGTNGNTCNN
TCNNCGTCGATATTANCTGCTCANGGCCCANAANTTTTTNNTTANTGNCCCANNNGTNNA
NNATCTCNGNNNNANACNCANTANNAAATTTCNNCCGAAANCNCNNCNNGNNANNNTNNT
TGNNGNNCGAATAGNNGTCTTNGNTTGNCCCTNTNTNCNNNNACNCTTTCCCNCGCNCCA
NTTNTTTTTNNATTNCCCNCTNNCCNTTTATTTACCNANCNCCCNTNTTNNTCCNCNNTT
ATTTATTATTTTTNNNNCNCNCNTCCCTCGTCCTNNGNNTCNTNNNCNAACNCTTTCCTT
TTNTTTTTNTTCNNCNCCATNCNNNCCNCACCGCCTNCCNTNGGANACTNTANGNCNCNC
NNTTNTCTATANNNCANNANNTTTTNTNNCCNNNNCCCCCTCACNCNNNNATNANNANTT
GTTNCTNNCCCTTCTNATTTNNNTACCNCNCNAGCATTTCTGNCNATCATCTGNCAANCC
NNNTNTCCTAATCTNNGNGACGCNGAACTCTGATCTATTNACTTNAGNTGTCCAAANATT
AGNCTNANTNTNTANATANTTCAGANGGGTCGGNCNNAGNCNGNNATNATNCGNGTNANN
ATCACTCNATCAANTCNCNGNCNTANATCNTCCNNCNTTANCNNANNGCTNGNNNGANTA
NTGAATNCNCNNGNNNAAATNNAAANTCNNTNACNNCCNCAGNTNANCNACTACGGANAC
CCTTNNGNGTCTTAAGCGCGCAAAANTANAGCAATANNTNNNCGGCNAATGAGTCTNACN
>F20
CATGCGCNGCTANCATGCAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGAGTAAC
GCGTGGGAATCTACCCAATACTACGGAATAGCTCCGGGAAACTGGAATTAATACCGTATG
TGCCCTACGGGGGAAAGATTTATCGGTATTGGATGAGCCCGCGTTGGATTAGCTAGTTGG
TGAGGTAACGGCCCACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACA
CTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAAT
GGGCGAAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCT
CTTTCAGTGGGGAAGATAATGACGGTACCCACAGAAGAAGCCCCGGCTAACTTCGTGCCA
GCAGCCGCGGTAATACGAAGGGGGNTAGCGTTGTTCGGATTTACTGGGCNTAAAGCGNAC
GTAGGCGGATCGTTAAGTCGGGGGTGAAATCCTGGAGCTCAACTCCAGAACTGCCTTCGA
TACTGGCGATCTCGAGTCCGGAAGAGGTGAGTGGAACTCCTAGNGTANAGGTGGAATTCG
TANATATTNGGAAGANCACCTNCGGCGNAAGGNGGCTTNCTGGTCCGGGTNNTGNCGCTT
ACNTGCNNNAGCCNGCGNNNNTAACNNTGNTNNNANACCNNGGTNGNCCCACNTCCNTNN
ATTTATTANAANTTNANCCCTTTGNGNAGTTNCCNCCTNCACTGNTNTTTTTTTNATTTC
CTTTNCNCTCTCCCCCCCGGGGNTGTNCGGTTTTCNTCANTTNNTATTCACNCNAANTTT
NTCCNGNNCCCNTCCCNNCTCTTTTNGATNCTATCGNCCNTTNTATNTTNNTCNNNCGCN
CNCTNNTCCTCTTCCCCTNCCTCTTGACTTTGGTTTTCCTTCGNTTNCNCNAGAATTTCN
CATCCCTCCCCNTTCCNGGCGATCTTTGTAANCNATTTNTTTNTANTCCCNCGCNCTCCC
CCCTCNNTCTTNCTTNGATNGTTTTGTTCTNCTCTNCNCNCNCNCCCCCTACACCCCTGT
TTCNGTTTCCCCTCTANNTNCTNTNTNTNCCCTCTCCCNANTCNNNTTNTNCNTTNTTNT
CCCTCNCCCTCTCCTTNTTTTACCCCCTATCNNTTTTCTTCCCTCCTTTGNCTNNNNTNT
CTCNCNCTTNNTTNNCATGTTTCCCCC
>F21
TNAGGGGGNGCTAACATGCAGTCGAACGGTGAAGCCCTTCGGGGTGGATCAGTGGCGAAC
GGGTGAGCTAACACGTTGGGCAATCATGCCCTGCACTCTGGGACAAGCCTTGGAAACGAG
GTCTAATACCGGATATTACCCTTGACCGCATGGTCTTGGGTGGAAAGCTCCGGCGGTGCA
GGATGAGCCCGCGGCCTATCAGCTTGTTGGTGGGGTGATGGCCTACCAAGGCGACGACGG
GTAGCCGGCCTGANAGGGCGACCGGCCACACTGGGACTGAGACACGGCCCANACTCCTAC
NGGAGGCAGCAGTGGGGAATATTGCACAATGGGNGCAAGCCTGATGCAGCGACGCCNCGT
GAGGGATGAAGGCCNCCNGGTTGTAAACCTCCCTCANCACGGCAANAANCNCNNGTGACG
GCTANCNGNANAANAATCANCNCNNTAACTACATGACCAACNNCCGNGTTAATANNCANG
GCGCGNNNCTTTNTTTNNCCAATNNTTTNATAGCNATNAGNCTGNANTNGNTGCCNTNAT
CTTCCCCCGNNCTTANCNCNCTTCTTTNTTNNAACTNTCTNTTTGCNTTCGNCCCCCCCN
CTTTTTTNTTTCNCNCNCGTNNCCTNCATTCTCTNNNNATNNTTNTTATGNCCNNCNCCN
NANCCTCCCCNCNTTTTNTTTTNNTTTNCATGCCTCTNCTCGTTCACCCTATTTTTNCNC

NCNTTTNTCCCCCCTTTTTTTTCTNCTTTNCCNCTNCNCTTTTTTNACCTTCNCCCTNNN
TNCTNNCCNTTATTTTTTATTTTCCNCCNTCNTCTCTACNTNNANACTCNCNNCACNNCT
TCTTTTTATTTTTATACTTGCNTTNNCNCCANAGTCGTCCTCTTCTNNCCCTTTTCCNNC
CCCNTTNTTTATNCCACTNCTATTTTTGCCCCNNCCCCCTATNCCCCCNCTATTTTCTAT
TCNNCCTCCCTTCTTTCTATTTTNACCCCCCNCCCTNTTCCCCTTATNNCNTCCCAACNA
CCCTNNNTTCTCNTTCNANNCNNANNCNTTNCNCTTCTCCCNCTANCCNNCTCTTNCTTA
TTTTNTNTTNTTTTNCCCTCNCCNNGCCCNACTTNCTTGCTTTCNCNANTTNGCNCTNTT
NTNCCCCTNCCTTTNNTGCCNCCCTATTACTTNNNANCNCNATNCCTCTNTNACTNCCCT
CCCCNCTCNTNNTTCTTTNTNTTNCCCNCNNACCNCTAAANTTGTANCCTCNCNNTATNC
TCTTNCCCNTTCANNTTNATCACNGTNCCCCTCCCNNNNNNTCGATCCNCC
>F22
AANGNGCNGCTTTACCATGCAAGTCGAACGCGTGTAGCAATACACGAGTGGCGCACGGGT
GAGTAACACGTGGATATCTGCCTTTTGGTTCGGAATAACCCTGGGAAACTAGGGCTAATA
CCGAATGGTTCCTACGGGATAAAGATTTATCGCCAAAAGATGAGTCCGCGTCCGATTAGC
TAGTTGGTGAGGTAATGGCTCACCAAGGCAACGATCGGTAGCTGGTCTGAGAGGACGATC
AGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATT
GGACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTG
TAAAGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAACTT
CGTGCCAGCAGCCGCGGTAAGACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAA
AGCGAGTGTAGGCGGTTGTCCAAGTTGGATGTGAAAGCCTTGAGCTTAACTCAAGAAATG
CATTCAGGACTGGATGGCTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTGGTG
AAATACGTAGAGATTGGGAAGAACACCAGTGGCGAAGGCGGCTATCTGGACGGTTACTGA
CGCTAAGACTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGT
AAACGATGGGTGCTAGACGTTGGCGAGCTTGCTCGTCAGTGTCGCAGCTAACGCGTTAAG
CACCCCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGC
ACAAGCGGTGGAGCATGTGGTTCAATTCGACGCAACGCGAAGAACCTTACCAGCCCTTGA
CATGTGACTCGCCGGCTCCAGAGATGGAAGCCTTTCGGNTTCGGCCGGAGTCAACACAGG
TGCTGCATGGCTGTCGTCAGCTCGTGTCNTGAGATGTTGGGTTAATTCCCGCAACGAGCG
CAACCCTCGTCTCCGTTGCCATCANGTTATGCTGGGCNCTTTGNAAAAACTGCCGGTGAC
AANCCGNAGNAAAGGNGGGGATNACCTNCANTNNTCNNGNCCNTNACGGGCTGGGCTACC
CACCTNNCTACAATTGGGGGGGAAAATTGGGAATGCN
>F23
CAATGNGCNGCTTACCATGCAAGTCGAACGCGTGTAGCAATACACGAGTGGCGCACGGGT
GAGTAACGCGTGGATATCTGCCTTTTGGTTCGGAATAACCCCGGGAAACTGGGGCTAATA
CCGGATGGTTCCTTCGGGATAAAGATTTATCGCCAAAAGATGAGTCCGCGTACGATTAGC
TAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCGTTAGCTGGTCTGAGAGGACGATC
AGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATT
GGACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTG
TAAAGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAACTT
CGTGCCAGCAGCCGCGGTAAGACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAA
AGCGAGTGTAGGCGGTTGCCCAAGTCAGGTGTGAAAGCCTTGAGCTCAACTCAAGAAATG
CACTTGGTACTGGGTGGCTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTGGTG
AAATACGTAGAGATTGGGAAGAACACCAGTGGCGAAGGCGGCTATCTGGACGGTTACTGA
CGCTAAGACTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGT
AAACGATGGGTGCTAGACGTTGGCGAGCTTGCTCGTCAGTGTCGCAGCTAACGCGTTAAG
CACCCCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGC
ACAAGCGGTGGNAGCATGTGGTTCAATTCGACGCAACGCGCANAACCTTACCAGCCCNTT
GACATGGGNACTCGCCGGGGAGCAGAGACNCTTCCCNTTCGGTTCGGCCGGAGTCCGCAC
AGGTGCTGCATGGCTGTCGTCANCTCGTGTCCGTGAGATGTTGGGTTAAGTCCCNCCAAC
GAGNGCAACCCTCGTCTTCCNGTTNCCCCCAGNTTATGCTNGGGCACTTTGGAAAAACTG
CCCGNTAACAACCCGNANGAAAGNNGGGNATAACNTCAATTCCTCNGGGCCTTACCGGGN
TGGGNNTACCCCCTNCTTAAATGNCGGTAACATNGGNATNCATTGNCCCNANC
>F24
NANGNGCNGCTTACCATGCAAGTCGAACGCGTGTAGCAATACACGAGTGGCGCACGGGTG

AGTAACGCGTGGATATCTGCCTTTTGGTTCGGAATAACCCCGGGAAACTGGGGCTAATAC
CGGATGGTTCCTTCGGGATAAAGATTTATCGCCAAAAGATGAGTCCGCGTACGATTAGCT
AGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCGTTAGCTGGTCTGAGAGGACGATCA
GCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTG
GACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTGT
AAAGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAACTTC
GTGCCAGCAGCCGCGGTAAGACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAA
GCGAGTGTAGGCGGTTGCCCAAGTCAGGTGTGAAAGCCTTGAGCTCAACTCAAGAAATGC
ACTTGGTACTGGGTGGCTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTGGTGA
AATACGTAGAGATTGGGAAGAACACCAGTGGCGAAGGCGGCTATCTGGACGGTTACTGAC
GCTAAGACTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTA
AACGATGGGTGCTAGACGTTGGCGAGCTTGCTCGTCAGTGTCGCAGCTAACGCGTTAAGC
ACCCCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCA
CAAGCGGTGGAGCATGTGGTTCAATTCGACGCAACGCGCAGAACCNTTACCAGCCCCTTG
ACATGGGACTCGCCGGGAGCAGAGACNCTCCCCTTCGGTTCGGCCGGAGTCCGCACAGGT
GCTGCATGGCTGTCGTCAGCTCGTGTTCGTGAGATGTTGGGNTAANTNCCGCAACGAGCG
GCAACCCTNGTTCTCCAGTTGCCACCAGGTTATGCCTGGGCACTTTNGNANAANCTNNCC
GGTGACAANCCNGAAGNAANGGTGGGGAATAACCTCAAATCCNCATNGCCCTTACGGNCT
GGCTNNCNNCNTTCTNNAATGGNGGGGANAATGGGATTCAAN
>F25
GNGAGGGGNGNCTNAACATGCAAGTTCGAGCGGGGCGTTAGCAATACGATNAGCGGCAGA
CGGGTGAGTAACGCGCCGTTTNACGTACCTTTTGGTTCGGAACAACACAGGGCAAACTTG
TTGCTAATACCGGATAAGCCCTTACGGGGAAAGATTTATCGCCGAAAGATCGGCCCGCGT
CTGATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGTGACGATCAGTANCTGGTCTGAG
AGGATGATCAGNCACATTGGGACTGAGACACGGCCCAAACTCNTANGGNAGGNAGCAGTG
GGGAATATTGGACAATGGGCGNAAGCCTGATCCAGCCATGCCGNGTGAGTGATGAAGGCC
CTAGGGTTGTAAAGCTCTNTTGNGCNNGAAGATAATGACGGCACCGCAAGAATAAGCCCC
GGCTAACTTCGTGNCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATCAC
TGGGCGTNAAGGGTGCGTAGGCGGGTNTTTAAGTCAGGGGCGAAATCCTGGAGCTCANCT
CCAGAACTGCCTTTGATACTGAAGATCTTGAGNNCNGNAGAGGTNAGTGTAACTGCNAGN
GNNGAGGTGAAATTCGTATATATTNGCAAGAACNCCANCCGGCNNAAGGNGGNTCACTGT
NNCNTATANTGACCNTNNACGCACTAAAGCNTGCGGTAGTAANACTNCNNNTANNATATN
NCTGGTANNCCTAGCNCCGTCANANTTATTNANTNTCCACNCCGTTACTCGGCGTNTANC
NCNCTTNNCNGTCNTNTTNTATAANTCTTCNAANGAATTTCCCCCCTGGGGGAANNTNCC
CGTTTTTCCNNCNTNTCATTTTTTCNAAGTGAATTTTNNCGGGCGNCCCCCTACCCANCC
NTTTGTAATCTTNNTNCNCTTNTATATNTNNNTCACCGCCCCNNTNTTCCTNANCCNGTC
CNNNTGNNNNTNCNTTNNCTTCCNAACCTTCCNAANAACCNCNTNCCNCNCNNNTCCTNN
CNNNNNCCCGTNANNACTGGNTATTTTTNNNGTNTNCTCNCNNGCCTNCCGGCCNNNCGC
TTNNTTNNNNGCTNAANTCTCNCCTTNCNGCATCNACNNCCCCACNNNCNNTAANCTNCN
CNNCGCNTNTCTTTNTGCNGCGCCCCNCANAGGANNATNCNCNCGGNCTATNTAANCCTC
CCGACCNACGNTNTNGNTATANATCNCNATTTNTCTNCANTCCCACCCNCAAGNTNNTNG
TGCNCNCNCNNTNTNNNNTNNNAAGCNACNCGANNN
>F3
CANGNGCNGCTTACCATGCNAGTCGAACGCGTGTAGCAATACACGAGTGGCGCACGGGTG
AGTAACGCGTGGATATCTGCCTTTTGGTTCGGAATAACCCCGGGAAACTGGGGCTAATAC
CGAATGGTTCCTACGGGATAAAGATTTATCGCCAAAAGATGAGTCCGCGTCCGATTAGCT
AGTTGGTGGGGTAACGGCCTACCAAGGCGACGATCGGTAGCTGGTCTGAGAGGACGATCA
GCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTG
GACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTCGGGTTGT
AAAGCTCTTTTGGCGGGGACGATGATGACGGTACCCGCAGAATAAGCCCCGGCTAACTTC
GTGCCAGCAGCCGCGGTAAGACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAA
GCGCGTGTAGGCGGTTGTCCAAGTCGGGTGTGAAAGCCTTGAGCTCAACTCAAGAAATGC
ACTCGATACTGGGCGACTAGAGGACCGGAGAGGATAGTGGAATTCCCAGTGTAGTGGTGA
AATACGTAGAGATTGGGAAGAACACCAGTGGCGAAGGCGGCTATCTGGACGGTTTCTGAC

GCTAAGACGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTA
AACGATGGGTGCTAGACGTTGGCGAGCTTGCTCGTCAGTGTCGCAGCTAACGCGTTAAGC
ACCCCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCA
CAAGCGGTGGAGCATGTGGTTCAATTCGACGCAACGCGAAGAACCTTACCAGCCCTTGAC
ATGGGACTCGCCGGGACCAGAGACGGTCCCNTTCGGTTCGGCCGGAGTCCGCACAGGTGC
TGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTNGGGTTAAGTCCCGCANCGAGCGCA
ACCCTCGTCTCNGTTGCCATCANGTTAAGCTGGGCACTTTGGAAAAACTGCCGGTGACAA
CCCGANGAAAGTGGGGAATAACCTCAANTCTNNTGGNCCTTNCGGNTGGGCTACCACNTG
CTACANTGGCNGTNAANTGGGAATCCAAGGGCNACCCCGANCCATCCAAAANCCTCCCAT
TCGNATNNNCNTNCACTCNGGNNTNAA
>F4
GNGNGGNAAAANGACGGNCTNGNCGAAGGGGGGGAGCNACGCGAAANNACAATNATGACG
NNTGAANGGGGAGCAGANNNAAAAGNCCCCCTTTTNTAANGGGAGCNCCACGNGNGCANC
ANACGGTAATACNNTNAAGCGGCGNNANNNNCGNCCGNTTTTNNATAACGNCCAACCGTG
TTNNNTANGNNCNNAANTCNNNNNNTCCGCACNCGNNCCNANNTGCNNNCNCACNNCCNN
TCNCTNTTAANTCCANACGCAGNGACAAGAGACANCGGGATANNNCGGTCNGCGNNCNAN
TGNANANCCNANNNCGCANAGATANANTGANCAANNNNCACCGNACCNNNNNAACCCANC
TATNTGAGCGNCNCCNNNNTAGCNCGNCNGACGCANTATNGCCAGCNGAGGTAGTGANNA
GNNNTTNANNNCNNCCNNNANCGCACCNANGGNGGAANAANNNGCNCTNNTACGNCCTTN
ANAGGGCCNGGGNGAGATAGNGAAANGTAANNTNGANTNNNNGNCNGCGCTANNCCGNNG
GTNTTNNTTNNTCCCNCCCNCANNCCCCCACCTNNNTTGTNTCNCCANNCANCNNTAGNN
GNNNANNNNNTNTCCNNNNANGNCACNTCTNTNNTNGNGNGGTNNAGANGGTANGCGGCN
NNNAAAGACAGNNNNNTANATTNNNCGNNGAGAATNAANACCNNNANATNTNANGAACNG
NGCNNNCGCNNTNNTNGNNCNTANNGNGAGNAANAAANAGANGNCGAGGNNNNAGANCNN
ATACANNTGNAGTACGNGNNGCNNCCANAAANNNNNGCGCNGCACNAANAGACNNGNGGA
NCAACNANNGAANGANAANCNGAGANGNAGGGNGNGGGNCNNGCAAATNTGNNNGNNGNA
TNCCGGNTNGAACNANNAANNGNGNTNNNAANNGANANAAAAGTTNAGNCNGNCNNANCC
CCNAGGNNCCGAAAGAGANATGGNNANCNAAANACAGGAANGGGCGNANNNNNGCAGNCC
NNNANGNNAGNTTTATCNNNNNGGCNNNNNNTNANNNTNNTNANCNCCNNNNNNTTNGTN
AAGCCNNNACCNNNNTNNNNCNCNCGTCNNNTNNNTTANTNTTNCCANCCNNNGNNCGNT
CNNCNNAAACNTNGCNNGGNGNNNNGGTCTATTTNAATTTAANANCNCNGAANNNCNNAA
CNACANNCNNCTNCNGNGCNACTNGNNNNNGNCNCNCGCNNNTNGNNAACGGCNNNNNAN
TANGTGNCNNACCCAGNANAACNACGGACNANNNAAATGNTNACTNGNCCCGCTATAANA
NNNTNCNCCTGNNNAGGCNATANCGCGTAATGGCGNCACGNNGANGGTNNNTGGNCNNNC
GCNGNCNGCNNNNAANTATNTNTNNAGGNNNNNAGANANNGAGANNTAAACNNNNNAAAT
GATAAGAAATNTGAGNGNNAGCNGNNCNANNCCNGAGTTNNGAGGGNGGNNGCNAGNNAA
GNNGAGNGCGAGANACACNGGGGGNNAAGANGANCNNNNAGTTGNNNAATAANNTCGAGA
NGNNNNANNNNGNNCNNGNCNNNAAAANCAGGANGTNNCGNNNGCNCNGAAGACANAGNG
GANGANNNNCANNNNAGNNTGANNANCNNNAANNAGNAGTTNTAGNNTGNTAGNGGNNGA
NTNNGATNNGG
>F5
GCCCCCTTAANNGANGNGGGGAGCTGNANGAAATANANGNNNAANAANTAACCGGGCTNA
NGNGGTAACCNCTTGTGNAAANNNNANCCNTTNTNTTTTNCTCANNNNNTCNNCGCATCG
NCNNGTTCNCTNNTANNTCNCNCNCACTNTNNNCTTTTANCCNNCNNCTNNNNCNACGAG
GATNNANNNNNNAGNAGTTCANNNTNNCACGCNGACGGNCANAGNNGGAGGNGNAANCCN
GANNCATAAACNANCNANNTNGGNCCNCGNGNGCTACACNCCCCTNANTANCTTGAGTNA
TTTNGGNANTATGCCCNCAGNGCCGTTCCGGNNTNGCNCCNATTCCNCTTNNGTGCNNGA
GAGGAGTTNANTNTANTNNNTTGTANCGCTNTNCNTTTTNNTTNTTACCNNNGNGCNNNN
CTTTATTTNTNNGNACCNACGGNGNGNGNANNNNNGNNNTCNTGTCNNNNNTNCTTTNNA
CNNAGNTNCNNNGNNGTTTNTATAGNGNATNGNNNAGTGGCGGNCGNANCNAGCGCCNCA
NATTAAGCCNTNTNNNCGTNTNNTANNGGNGNTCNNGNCTANATNAGCNACGNGNNTNNN
NGCNANNAAANTCCNCNNCAGCANGTGNNGNAAANNCCNANANCCNCANAANNNNNNGGA
TGAATANNNANTGANATGGNGANNAAANNNTNAGNNNGNCGAGGTNTATNNCNNNNNNTA
AGGNNNTAAANAAAATNGGGNAAAAATGATNANNNNNGNANNGNNNNGANGNNACANCGN

NGCNNNTNAATTGNNNTNGNATANNATCNGGNNTNANGNNATATACNNNANNNNTNNCNC
NAANNNTNTCNNNAANNGNNNTTNNTGNATNCGTTANNNGCNNCNATTNTNTNGCNNNNN
NNACCTNNNANGGTCGTNNNNTNTTNTTNATTNNANNNGGNANCGNNGTAGNAGTNATAN
NNGNNGGCCANNAANAGTNGNGATTANNTTATANTACANNNNAAAAGANNGAAGAAGACG
CNGNNGNGANNNTNNNCGNNTTANTCGGNGCNCNGNATTNTTATGCNNNCCNANNNNTTA
NTNANNCNNACCNNGCTCCCANNANNNGNNNNTNNTAANANATNNACAACGCNNNCNNNA
TANAANATNNGNGCTGGNNNGANNTNANNNANATAGNGCNCCNGNAAAANNGNNCAANNN
NGTTCNTTNGTAAANGGNNGCAAGCTNANGATATNGTGAGCAGCANGNGGNNNNAAGAAG
GCGANAANNAATAANAAAGAGAGGGNGGAGCGGGANCGNCAGCNGTNNGNGGNGGANGAN
ANNNTTCAGCAGNANANNANNACNANGGAAGNNANNNNATAGANNTNNGNNNGGNNGTCN
ATNTNAGNNNGNGTGNNANAANNAAAAAATTATAGTANNNGTGGACNAAGGGGTNANNNN
GAGNCANNCCGTTNNATNGGNAGGACCANGNGANAAGNNAGAGNNGNTGGATGNAGGAGN
ATGNGAACNNNN
>F6
GCNTNAAAAGGAGGGGNTTACAACNACGGAGGAAAGACNGCAACANTNNGGATGTTTATT
TANANGGNTGGAGAGNGNNAGGACCCCCTTTNNANNGNGNCCAGTNGGNCGNNGANAAAT
ANNTACGNNCNTGANNNCNGNNNCNCNNNNGNTCCCNNCNANTNTNATANCNNTTTNCCN
NNCNCCTNNNTCCGTTACGNNCNNAACNTCCNCTTTNNANANCGNGTNTANNNNCATCNN
TTNATNNNGAATGAGAGGNCAGGGGACGAAAGGANAANACNAGGNCGACCAANCNNGANA
NCCCNNCTTTGCACCCGNGNGACCANGTNTAGAATCGTCGNNTNTTNNTATNGATNTATT
TANCCCCNCNNNNCAGCCGGNNNTGGGCCNATNNNCNAATATANGGGNGANNAAGNNTNG
ANTNNNGNNTACTTCCNTNGCNCCTNTTCTACNTTGATTNTTTTGNCNNNNNCNAAGNCT
NTTTTNNGTCANAGNCCACAGNTNNNTATGTTNTANNGCNGTNTNAGCCATNTNGANGNC
GCCNAGNCNNNANANNCGNTTGTTNAANCAGTNNAGNGAGTNAANNGCGNATNANGGAGA
GAANGNNTCCGCACGCGNGANANNANCAGNNGGNGTNAAAGGAAATNCNCTNGGATGCGC
CTAGNTCNNAACGGANAANGGNNAGNGGATNNAANNGTACCCAANNANNNNAGGANGAGA
NAATTGNGTATNTGCNNNCACNCGTNNAAGNNGCNGNGGATTAGGNGNTANNAANNGNNA
CCANNANGGCCNTAAAATTNNNGNAAATGNTGNNNGANGAAGAGAGNNNNNNANANNAAN
NNTTTNTNNNTCTNNNNCANTAANNTGNNGNTNGNANAANTANNNNNNCNCNACNTNTNC
AGCNNNNAATTTTNGTTTTNNANTTTANNNCCNCNATTNTTNNCNATTNNCNGNTGNNNC
CCCTCCCGNNTTNNTTNTATNTATCANNGCGNCNNTNACATNGNCNNNTCCNCGNNCCNN
NNNNNCNATCCTTTNTTATNTTNATNACCNCCGGNNAACANNNNCANNNCNGTNNGANAT
NNAACATTTNTNCGNCNCATTTNATGTTNTANCNCNAANTTNCTNACACCTNCCCNCGCN
NCCNCGCNACNTANNNATAGTTNANAGNNACCCNNCTNATNNNNTATNTNACNCCCCCCA
CNCAGTTANCGGNATANCTCNCGNCACCCCAANNTNNCNACTNNNCCGACNNNGANCNTT
CTNATNNCNACNCNCNANCGAACCANNNGCTTGNGNATTATANAGANGCACGGAGNGNNA
GGGCNNNNTNGNCNANCCNNACNNGNNAGNGNANNNAAANCCTGGGNANANNTCGCACAC
TNNANAACANCNNTCANGGCGNNANANAAGNCGANNGNANNTANAGAATNNNCGGNCCCN
NGNGGNNAANTNTNCNGGNGCCACTGCNNTANNNNNANCNCNNGTGGTANAAACNTNNAN
NNTANANTGGNNGNCNN
>F7
NGAGGAGGGGGNGGNTAAAAGGCNGAGGATTGTTAAACACGNNGNNNAGACAACAAAACN
GNGGTTGANNGTGTNNNGGAGNTNAANGGCGAGGGGGACCNCCTCTNTTNAAGNGGNCCC
TNCGTNGCANTCNANNGAAAACGNNNCNGAGNNGGACGNANCNCATACNTCNGTAGTGNA
CNCNCACTTCTNTGNGTATGNANCATNATTNGNNCGTCCNCCATACGTCNCNNNTANCTC
ACNGANNCNCTNNCNNCCNTNTNTNNTNTANNCGATTCGNGATNCAANGNGNTTTNNNNN
TANNCCCGNAGNGNGTGGCNCAGGGAGNGGGGNNGAGGACGGAGGNANNCNCGAACCANN
TAACGGNNAANCCNAAGNACACCGGNNAACACCNCNNCTCNNANTNGAGCCNGGTTANAN
TATTTTGCGNTCNNTCNNNTCCGCACCACCCGCNCGNNNTCNNANCTNTATGNNNCGCNN
CTATNTCGACNNTTNGANGGCGNGGAATNNAAAGNGNANNNNANNGGNNANNTTNGNTTN
CCCACTANTNCTNCCNNTGTTTNTTTTNTATTGGGNANAANNCGANCCCNNCCNTGTNTT
NNTCNAGNNNCNCANGGAGTATNNATTTNNNTCTNNNTNNCNTNTATACCCGCATCNTNT
ANTTTCNGCNCGCGGNTAANGCAGTCGTTGTCGCNNCGGGATNCNNATCAGAGNNNTNNN
TGNCAGNGGGNANTCANAANAGCNGGAGNACGCNTNNANGTNNAANGGNCGGNNAGACAT

AGGTNGNTNNCNCNNGGAGGTCGGAGNAGTTNTNNTGTGNNCGNNANGANNNNCANGNAG
TTNCNANGTGNNNCNACNTCCTTAANTGNNTATANNGACAGNGAANGATNTGGNNTNNNA
TGCTGNCGNNGGNNCNNGNNNNNTNNANNGCGNTNAAATNANACGNNTGAGCGNNAACNN
GNNTNTTNANTANGNTNNNNNGGGGTTATGNCNNGNNACCTNCGNATAANGNGAANNNNA
NTNAGTCNNNNNTAAANGNNNANNTTTNNNTAANGNTNGCNNCANTNATTANNGACACNN
CAGTTTTCNGNNCNGNNCCATTTTTTTNTTNTNNNTTNACNCCCNNCGANGTTTNTTGGC
NCNGGCNNNTNTNNNNCNNTNNNCTGTTNTTTTTTTGATATGTTANNCACGNNNNTNTTG
NCTCGNGCNTATTACTCAGNNCNNNANNCCTTNTTTTTTTNTNTNTNTGNCNCCNTANNA
ANNNTACNNCNANCNNGTGGNGNNNANGCANNCNTTNTGNNCACGNNCTTNNNNTNTTNC
NNCCNNANNNTTNTNNNANNNNGCNNNGCNTAACANCNCACCATTTANNNAGTTATTNGG
TGCNNCANTNCTTNTANTNATAANNNACCCNCNAANGCCGTNAANNNCGNCACNGNNANG
AANAGNCGTNANATNTNNNNTAGNAANCGGGTAGCNNTATANGNGTAGNAGNNNGCNNAN
NGCNGGNNAAGNGCGCTANNTTAATAANNTCTGAACNGNAGANACACCNTGNCNTTTNAG
CNCANGNGCATTNAGNNANNCATGATTAGNTNNANNANCGATANCGNCNACTNANNANNT
TACNNCANTGTNNGNNTGNANNGNNCGNCTNGNCNNNNAATNNATTAATNCGNNTNNCCG
CNCNNCCGANNTNTTAGGCNCANCANTNNATNACNANNNNGNGATNCATCGTNNNNANGN
ANNNGNANNTNAGTAAGGATAACCCNN
>F8
GGAGGGGGGGNNAAANNTGCCAGTTTNAANNCGGGGAACCAANACTTNNGTNGGGGAGTG
NTNGNAGCAANGCNCCNCTTNANNNNCCNTTCTGGGATCGGAACACNNAAGGGNAANTCN
GNCGCTNATANCGGACNANCCCTTGTACGGGNAAAGATGCNATCNGTCAAAAGACCGNCC
CGNCCTCTGATTAGCNNCNTGGTGAGGGANTNNTTCACCAAGGCGACNAGNAGNAGCTGG
TAGGAGAGGANGACCNACCCNCNTTGGNACCGCAGATACACCCCTCNNCTNNNACGGGNT
GTNNNCATNTGNGGAACACNNNGACNNNGGAGCGTTAANCCTGATCCANTNANGCCGCGT
GAGTGATNNANNATTTNNNGGTTNNNNNGCTCTTTTNTNTNNNAAGACAATGACNGTNTN
TNAAGAACNAGTTTCNGTTTTAATTTCTTGCCTCTANTNCGNTNNCAACACAAAGGGNGC
TTAGAGCCANCNCGNAATNNCTGGGCGTAAANGGCGCGNAGGTNTGNTCNNTTANGTNAG
GGNNNAAANCCTGGTNNCACAANTTTAGAACTGTTCTNCGATACTGAAGACCNTGAGNTC
GGNNATANNTTATCGNAACTGCGNGNGNNTNNTGGAAATTTCACAGANATNCGNANTAAN
ANTTNGTNGNTAANGCNGCCCNNTNGNCCCCCNNTNNTGATTTNTGANGTACTCAANGTN
NGCNGAANAANCANCTATNNCNTACCNTNTTAGTCCNCNCCCNTNNAANTNTTGANANNC
CNNTNNGTCNTGCNTTTTTGTTNTTTANNGGCCCCTTCTNANNTTCTNNANCANNTCCNC
TTGTNTTNNTNTTGTACNCCNNNATNCAAAACNCNNNNGGANTTCNANCCTTTNCCNCCA
CNTGTNTNNCNNCTNNCTTGNTTCNNNANCCCNCCAACNCCCNCTAANNTTNTCNGTCCC
TCCNNNTTNANTTTTCNCNGCTCNCCNTAGNTCNTTNTNTNTGTAGGNNNNCNGTTCCTT
NNNANCNGNCNCANATNGTTTTTCTCCCCCNTGNNCGNCTTNNAGTTNTTNTNANTNTTT
NTNTNCCCGCTGGATNCCCCCAGTNTTGTGNNGGNCNCATTAATGNCNACGNGTGTACGG
GGNCNACNTAGNAGCCNTCTNCANGANNNNTNGGCNTNCCCGTNNATCCCCGGNCCCCCT
AGATNGNGGCTACTNNNCGCCNNNGCGNCCNATNNTACCCNTCNGNTCTNNCTNNTNCAA
NNCANAAAGCNNCTNGGNTCCATTGGNNCNNNANTATGNNNACTNG
>F9
AGNCGAGCTTACCATGCNAGTCGAGCGGGCGTAGCAATACGGTCAGCGGCAGACGGGTGA
GTAACGCGCTGGNAACGTACCTTTTGGTTCGGAACAACACAGGGAAACTTGCTGCTAATA
CCGGATAAGCCCTTACGGGGAAAGATTTATCGCCGAAAGATCGGCCCGCGTCTGATTAGC
TAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCAGTAGCTGGTCTGAGAGGATGATC
AGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATT
GGACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCCTAGGGTTG
TAAAGCTCTTTTGTGCGGGAAGATAATGACGGTACCGCAAGAATAAGCCCCGGCTAACTT
CGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTAGCGTTGCTCGGAATCACTGGGCGTAA
AGGGTGCGTAGGCGGGTCTTTAAGTCAGGGGTGAAATCCTGGAGCTCAACTCCAGAACTG
CCTTTGATACTGAGGATCTTGAGTTCGGGAGAGGTGAGTGGAACTGCGAGTGTAGAGGTG
AAATTCGTAGATATTCGCAAGAACACCAGTGGCGAAGGCGGCTCACTGGCCCGATACTGA
CGCTGAGGCACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGT
ANACGATGAATGCCAGCCGTTAGTGGGTTTACTCACTAGTGGCGCAGCTAACGCTTTTAA

GCATTCCGTCTGGGGAGTACGGTCGCAAGATTAAAACTCANAGGAATTGACNGGGGCCCG
CACAAGNGGTGGAGCATGTGGNTTTNNTTCNACGCAAACNCNCAGAACCTTTAACCNTCC
TTTNACATGTNCCAGGANCCGGTTNCNAAANATGTTGACNCTCNTNTTCNGANCCCTGGN
ACCACAGGTGCTTGCATGGCTTGTTCTCCACCCCTGNTCGTGANATGTTGGGTTAAAGTC
CCCCCAACNANGCGNNACCCCCGTNCTTTANTTNNCTNCCATTTAATTNAACCNCTCTAA
GGAAANTTNCGGTNATAACCCNCNAAGNAAGGTNGGGATACTCCCAAGNCCNCNNGNCNC
NTNCNGGGCTGGNCTNCCNNCGTGNNTCATTGCNGGGANATNGGATGTCAAAACCCAATN
NTNGCAATTTCAAAACCCNTNCTTNCAANTGCCNC