4-22-2024

Exploration of Enterococcus faecalis Bacteriophages Diversity: Comprehensive Phylogenetic Analysis and Discovery Isolation of Enterococcus faecalis Phage from a Soil Sample in South Florida

Hailey N. Jauernick
Nova Southeastern University

Capstone of
Hailey N. Jauernick

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science
Biological Sciences

Nova Southeastern University
Halmos College of Arts and Sciences

April 2024

Approved:
Capstone Committee

Committee Chair: Katie Crump

Committee Member: Julie Torruellas Garcia

Committee Member: Navi Gill

NOVA SOUTHEASTERN UNIVERSITY

HALMOS COLLEGE OF ARTS AND SCIENCES

Exploration of *Enterococcus faecalis* Bacteriophages Diversity: Comprehensive Phylogenetic Analysis and Discovery Isolation of *Enterococcus faecalis* Phage from a Soil Sample in South Florida

By

Hailey N. Jauernick

Submitted to the Faculty of
Halmos College of Arts and Sciences
in fulfillment of the requirements for
the degree of Master of Science with a specialty in:
Biological Sciences

Nova Southeastern University
April 2024

Abstract
Enterococcus faecalis is a commensal bacterium found in the gastrointestinal tract, oral cavity, and vagina of humans, which can transition to pathogenic states, leading to diverse infections. Many *E. faecalis* infections are becoming progressively challenging to treat, stemming from the lack of effective antibiotic therapies against its virulence factors, including biofilm formation and antibiotic resistance. Phage therapy emerges as a promising alternative, targeting bacteria with minimal impact on host flora. However, despite phages' ubiquitous presence and potential applications in medicine and biotechnology, their diversity and ecological niches are vastly understudied. Here, we investigated the diversity of sequenced *E. faecalis* phages and isolated a phage from a new environmental source. All phages belonged to either the *Siphoviridae, Herelleviridae, Andrewesvirina, Rountreeviridae,* or *Autographiviridae* family, and their genomes ranged between 16,954 and 151,985 base pairs. The phylogenetic analysis focused on single 'marker' genes, specifically the portal protein and terminase large subunit. Applying single-gene analysis yielded distinct clusters, each corresponding to specific families and genera. Furthermore, we successfully isolated a lytic phage from soil capable of infecting *E. faecalis* OG1RF. From our results, we conclude that a single gene has the potential to serve as a molecular marker for classifying phage taxonomy, as evidenced by the distinct clustering of phages. Additionally, our findings indicate soil as an additional source for isolating *E. faecalis* phages. Understanding the diversity of *E. faecalis* phages will aid in identifying alternative therapeutic targets to treat *E. faecalis* infections, thereby addressing the challenge of antibiotic resistance and enhancing treatment options for patients.

Keywords: *Enterococcus faecalis* (*E. faecalis*), bacteriophage (phage), phage therapy, antibiotic resistance, nosocomial infections, ESKAPE, biofilm, phylogenetics, portal protein, large subunit terminase, enriched isolation

Acknowledgements

This study and all that has come along with it has been a gratifying and valuable experience. I want to thank my principal advisor, Dr. Katie Crump, for supporting me from my undergraduate years throughout my years as a graduate student. Your dedication to helping me grow my knowledge and skills in science has made this project successful. I express my gratitude to Dr. Emily Schmitt for her receptiveness and readiness to address all aspects of this study, contributing immensely to the research process and my success in the program. Thank you to Dr. Navi Gill for guiding me through the data collection and analysis related to this study. Thank you to Dr. Julie Torruellas Garcia for your support and feedback, which made this a seamless project. I'd also like to thank Dr. Breck Duerkop at the Department of Immunology & Microbiology University of Colorado Anschutz School of Medicine for providing me with the bacterial strain and positive control phage used in this study. All your assistance has been sincerely valued and deeply appreciated. Finally, I would like to thank my son Camden for being my motivation and my mother for her unconditional love and support. To my incredible family, friends, and mentors, thank you for being there to encourage and guide me throughout my academic journey.

Table of Contents

I. Introduction

The rise in antimicrobial resistance is becoming a critical worldwide health threat causing the death of 1.27 million people annually and is projected to kill more people than cancer by 2050 (Walsh, 2014). Bacteria acquire resistance to antibiotics naturally, but resistance has begun to accelerate due to the overuse and misuse of antibiotics by doctors and veterinarians (World Health Organization, 2023). Additionally, due to inadequate infection control and prevention in healthcare settings, bacteria are rapidly acquiring new mechanisms and developing resistance to antibiotics (World Health Organization, 2023). On the other hand, nosocomial infections, also known as hospital acquired infections, are either initially antibiotic resistant or have acquired resistances through a variety of mechanisms. For instance, over and improper use of antibiotics can lead to selective survival allowing the pre-existing resistant bacteria to share their resistant genes to non-resistant bacterium through horizontal gene transfer (HGT) making future use of antibiotics ineffective. The six main culprits of hospital acquired nosocomial infections (HAI) come from the acronym ESKAPE, which are infections caused by *Enterococcus, Staphylococcus, Klebsiella, Acinetobacter, Pseudomonas,* and *Enterobacter* spp (Mulani et al., 2019).

Among the six ESKAPE pathogens, all exhibit some degree of pathogenic characteristics including multiple-drug resistance and possess a variety of virulence factors (Mulani et al., 2019). In the United States, enterococci bacteria are ranked the second leading cause of nosocomial infections after staphylococci and is the third most common cause of community-acquired endocarditis (Said et al., 2024). Commonly, many medical advances rely on using antibiotics to combat infections that patients may acquire via direct contact or indirect contact, such as through insect bites or food contamination (Mayo Foundation for Medical Education and Research, 2022). However, in some cases, antimicrobial-resistant infections may not have any antibiotic treatment options or require the use of a second and third-line treatment. These treatments often cause serious harm to patients due to their severe adverse side effects such as organ failure and can require prolonged care and recovery from the infection (United States Food and Drug Administration, 2021).

The global decline in antibiotic effectiveness combined with the slowed discovery of new antibiotics has renewed research interest in the generation of alternative antimicrobial options to treat infectious diseases. Specifically, new treatment options against *Enterococcus faecalis* are imperative due to its remarkable ability to spread antibiotic resistance bacteria (ARB) via HGT to

other bacteria. In addition to HGT, *E. faecalis* forms biofilms which also renders antibiotics less effective. These mechanisms and pathogenetic factors of *E. faecalis* contribute to its widespread rise as the second leading cause of nosocomial infections (Bhardwaj, 2019).

Bacteriophages, also known as phages, are viruses that specifically target certain bacteria. Phages are found anywhere in the environment where there is life. Due to their ability to target, infect, and kill specific strains of bacteria, they are a promising candidate for antimicrobial therapy, and an alternative to antibiotics for treating and preventing bacterial infections. Unlike many antibiotics that destroy destructive bacteria while concurrently obliterating the microbiota, an individual phage can target a specific bacterial strain or species precisely.

Currently, there are an estimated 10^{31} bacteriophages that are found in any region of Earth where bacteria exist (White et al., 2019). However, despite their considerable abundance, most bacteriophages are undiscovered and uncharacterized. Interestingly, there are limited number of bacteriophages identified that infect *Enterococcus faecalis*. Therefore, our overall objective is to understand the diversity among *Enterococcus faecalis* bacteriophages. Our two-fold research project aims to perform a phylogenetic analysis on the portal proteins and large subunit terminases of *Enterococcus faecalis* phage sequences in the GenBank database and employ isolation techniques to extract an *E. faecalis* phage from a soil sample. Through these analyses, we can enhance our understanding of the diversity of *E. faecalis* phages and establish a novel and feasible source that is suitable in in enhancing the discovery and subsequent diversity of *Enterococcus faecalis* phages. Furthermore, by uncovering the diversity of these phages that target a well-documented, yet understudied area of research, our research will highlight knowledge necessary for future investigations into *Enterococcus* bacteriophages and drug development research.

I.1 The Microbiology of Enterococcus faecalis

Enterococcus faecalis is a Gram-positive, γ-hemolytic, non-spore-forming, commensal bacteria found in humans and animals' gastrointestinal tract, oral cavity, and vagina (Lee et al., 2019). Enterococci belong to the facultative anaerobic lactic acid bacteria species and are composed of short and medium chains. Due to *E. faecalis'* virulence factors such as aggregation substances and surface proteins, it can survive through an array of harsh conditions, including starvation, hot, salty, and acidic environments, hospital disinfection, and antimicrobial treatments (Gilmore et al., 2020; Hartke et al., 1998; Strateva et al., 2016; Watson, 2018). In addition, the

virulence and hardiness of *E. faecalis* make it one of the most preeminent multiple-drug resistance nosocomial infections (Bhardwaj, 2019).

I.2 Antimicrobial Resistant Mechanisms of Enterococcus faecalis

Although *E. faecalis* is a commensal bacterium, disruption to its environment triggered by various factors, such as changes in the host's immune system, antibiotic use, or alterations in the surrounding microenvironment, may influence the opportunity to transition into an opportunistic pathogen and activate a range of resistance mechanisms in response to the altered conditions. Consequently, activating resistance mechanisms in *E. faecalis* during these disturbances enhances its survival and proliferation. Thus, therapeutic interventions are substantially complicated, as the bacterium exhibits amplified resistance to conventional treatment approaches. Therefore, the inherent dilemma in effectively treating *Enterococcus faecalis* infections, require a thorough understanding of its responses to environmental changes for the development of more targeted and effective treatment strategies.

The fundamental mechanisms of *E. faecalis'* strong antibiotic resistance may occur from HGT of resistant determinants, its intrinsic genes, or gained by sporadic mutation to their intrinsic genes (Hollenbeck & Rice, 2012). Through multiple HGT or erratic mutations, *E. faecalis* acquired multiple-drug resistance to many antibiotics. For example, *E. faecalis* acquired *bla* genes via plasmid mediated HGT making it indistinguishable from those in *S. aureus*. The *bla* genes encode for β-lactamases, which lead to intrinsic resistance through the expression of low-affinity penicillin-binding proteins (PBP). These proteins bind weakly to β-lactam antibiotics, reducing efficacy to drugs such as penicillin and ampicillin, high-level resistance to semi-synthetic penicillin and a majority of cephalosporins (Hollenbeck & Rice, 2012; Kristich et al., 2014).

Additionally, it is also suggested that *E. faecalis* acquired the van gene causing it to become vancomycin-resistant enterococcus (VRE) through HGT from *P. popilliae* (Ahmed & Baptiste, 2018). Typically, vancomycin will bind to the D-ALA-D-ALA portion of the peptidoglycan, inhibiting cross-linking and preventing the cell wall from having any structural integrity, which allows for the effectiveness of vancomycin (Hollenbeck & Rice, 2012). However, the *van* gene modifies the pentapeptide cell wall precursor terminal end to D-Lactate or D-Serine. This modification by the *van* gene allows the binding of glycopeptides, blocking vancomycin from binding to the terminus of the pentapeptide precursor and inhibiting cell wall

synthesis. Thus, VRE can ensure their stability by continuing to cross-link their peptidoglycan and build a resistance to intense antibiotics such as vancomycin (Hollenbeck & Rice, 2012).

Moreover, the expression of the intrinsic *lsa* gene, which is responsible for the structure of ATP-binding cassette (ABC)-efflux pumps, suggests intrinsic resistance to drugs such as clindamycin, quinupristin and dalfopristin (Hollenbeck & Rice, 2012). In addition, when sub-cultured at sublethal levels, the intrinsic nature of *E. faecalis* may contribute to its development of tolerance to fatal levels of bile salts and detergents, which explains how it can survive cleaning protocols practiced in most hospitals as a measure to control the spread of infection (Fiore et al., 2019).

While not a gene associated with specific antibiotic resistance, the *esp* gene is an intrinsic gene expressed on the surface of *E. faecalis*. This particular gene is responsible for the formation of biofilms. Biofilms formed by *E. faecalis* create a symbiotic community on the surface of which they adhere to. Once fixated to the surface, a period of bacterial growth follows, and the bacterial community can achieve multiple layers and form a more 3-D structure. During this growth period, other microorganisms can also adhere and exchange nutrients and genes. Lastly, the bacteria then disperse causing horizontal and vertical cross-host transmission (Kaplan, 2010). The formation of biofilms aids in the survival of *E. faecalis* under stressful conditions such as changes in temperature, pH, and disruption by disinfectants and antibiotics (Yin et al., 2019).

Interestingly, the *esp* gene is responsible for the formation of biofilms to polystyrene, a material used for many medical indwelling devices (Al-Zubidi et al., 2019). Moreover, biofilm formation can also occur on tissues, cardiac valves, dental materials, and contact lenses (Dincer et al., 2020). Once the first of the three stages of biofilm formation begins, the risk of more severe infections increases, and treatment becomes increasingly difficult. Therefore, *Enterococcus* is a very tenacious bacteria and has a lot of both intrinsic and acquired resistant factors making it one of the most difficult to treat nosocomial infections explaining why it is the second leading cause of HAI.

I.3 Clinical Infections Caused by Enterococcus faecalis

As described previously, enterococci are highly resilient bacteria that can survive in extreme conditions such as a harsh variety of temperatures, salts, pH, antiseptics, and disinfectants which accounts for their ability to survive on surfaces for a prolonged time without nutrients, promoting them to spread quickly (Allarakha, 2021; Said et al., 2024). Of the estimated

58 species of enterococci, *Enterococcus faecalis* is the leading and most common opportunistic pathogen accounting for roughly 80% of enterococcal infections (Li et al., 2020). In addition, *Enterococcus* is responsible for approximately 110,000 urinary tract infections (UTIs), 40,000 wound infections, 25,000 cases of bacteremia, and 1,100 cases of endocarditis in the United States annually (WebMD, 2023) with a mortality rate that may surpass 50% in critically ill patients (Fraser, 2022). Furthermore, the prevalence of *E. faecalis* in the oral cavity demonstrates a direct link with periodontal disease, and due to the resistance mechanisms, as previously discussed, it remains persistent within the subgingival biofilm even after undergoing periodontal treatment (Bhardwaj et al., 2020).

Studies have shown that previous antibiotic treatment that disrupt the normal gut microbiome can decrease RegIIIγ, an antimicrobial peptide produced by Paneth cells, which in turn can increase susceptibility to VRE (Khosravi & Mazmanian, 2013). Additionally, overgrowth of *E. faecalis*, can result in gut translocation across the intestinal wall where the bacteria can avoid being killed by the immune system and can result in bloodstream infections (Chatterjee et al., 2021; Said et al., 2024.). Taken together, the intrinsic and acquired mechanisms for resistance combined with its ability to evade the host immune system, it has been determined that the risk of severe infections with *E. faecalis* increases, especially for patients with prolonged or repeated hospital and health care facility stays including those in, intensive care units, long-term care facility, patients receiving hemodialysis, diabetes, cancer, and transplant treatment (Said et al., 2024). Therefore, due to its virulence factors and its ability to evade the host immune system it is imperative to find alternative preventative and treatment measures that are more effective than current therapeutic options.

I.4 Current Strategies and Interventions to Prevent and Treat Enterococcus faecalis Infections

Enterococcus is a hardy bacterium that can survive on hands up to sixty minutes after inoculation and up to four months on inanimate surfaces (Levitus et al., 2024). Current research suggests that contact with an infected person or surface contaminated with *Enterococcus* is the most prominent way the infection can spread (Levitus et al., 2024). However, the recommendations for contact precautions around VRE remain controversial as there is a lack of quality data and conflicting studies (Levitus et al., 2024). Until further studies are done to establish an agreeable protocol and method to prevent the spread and treat *Enterococcus* infections, especially VRE, generic disinfectant and treatment measures remain the same.

In healthcare settings, frequent hand washing, sterilization of medical devices, and surfaces are the ideal standard precautionary methods to reduce the spread of infection (Allarakha, 2021). Medical devices are sterilized by an assortment of methods that include steam, dry heat, radiation, ethylene oxide gas, and vaporized hydrogen peroxide (United States Food and Drug Administration, 2021). In addition, medical personnel should use and change personal protective equipment (i.e., gloves, masks, gowns, and goggles) when encountering a patient (Allarakha, 2021). Furthermore, protocol suggests the replacement or removal of medical indwelling devices as soon as possible to prevent the risk of infection from biofilm formation (Wu et al., 2015). Lastly, proper use and disposal of antibiotics are vital to helping reduce the increased risk of resistance.

Even with the appropriate protocols and preventative measures, infections occur due to the increasing prevalence of intrinsic and acquired resistance of *E. faecalis* as mentioned previously. The treatment of antibiotic susceptible versus antibiotic resistant *E. faecalis* infections differs significantly. Currently, the primary monotherapy method of treating most *E. faecalis* infections that do not display high level resistance (HLR) to aminoglycosides is ampicillin or vancomycin for patients with a β-lactam allergy. However, due to some of the virulence factors previously discussed, such as the *bla* or *van* gene, *E. faecalis* can be resistant to ampicillin and vancomycin. Additionally, because β-lactam does not have a strong bactericidal effect the success rate of treatment for *E. faecalis* endocarditis is around 60% (Kristich et al., 2014). Therefore, requiring a combination treatment of antibiotics such as β-lactam along with an aminoglycoside is necessary to generate a synergistic bactericidal effect against *E. faecalis* infections (Kristich et al., 2014).

On the other hand, treatment of ampicillin and vancomycin-susceptible *E. faecalis* that exhibit HLR to aminoglycosides, such as in *E. faecalis* endovascular infections, requires a different approach. In this case, the synergistic bactericidal effect of the cell-wall aminoglycoside and agent combination is nullified by the HLR to gentamicin or to streptomycin (Kristich et al., 2014). Additionally, the toxic profile of aminoglycoside can be harmful to critically ill patients (Gonzalez et al., 2022). Therefore, studies show that it is possible to achieve a synergistic bactericidal effect for HLR *E. faecalis* that are susceptible to ampicillin by using a combination of cephalosporins with an amino-penicillin (Kristich et al., 2014). However, more clinical data is still needed before this treatment combination can be recommended.

Multiple studies conducted show that the multiple-drug resistance of various clinical/ HAI isolates of *E. faecalis* range from 74.6-100% (Esmail et al., 2019; Said & Abdelmegeed, 2019). Nonetheless, treatment of *E. faecalis* poses an enormous threat worldwide, due to its evolving mechanism and virulence factors in response to prevention and treatment protocols. As a result, novel therapeutic strategies concentrating on the prevention and treatment of these infections are critically needed.

I.5 Bacteriophage Structure and Life Cycle

Bacteriophages, also known as phages, are bacteria-specific viruses that infect and use the bacteria's machinery to reproduce. While the classification of phages depends on one of six phage morphologies, all phages consist of a nucleic acid genome composed of either circular or linear, single-stranded or double-stranded RNA or DNA. Their genomes are enclosed in a capsid surrounded by a lipid membrane (The Biology Notes Editors, 2022). Generally, a phage's structure consists of a head, which contains the genetic material attached to the collar. The collar connects to the sheath that contracts to deliver the genetic material. The base plate is made up of spikes that poke holes into the host membrane and the tail fibers used to attach to the host (The Biology Notes Editors, 2022). Classification of a phage into one of the six morphological groups is dependent on their nucleic acid, morphology and physicochemical properties of their virion particles (The Biology Notes Editors, 2022).

Like all viruses, phages must infect a host to ensure their survival and initiate replication. Once a phage infects a host its life cycle occurs in one of two ways: the lytic or the lysogenic cycle. Regardless of the phage's life cycle, the first step involves the tail fibers recognition of specific receptors on the surface of the bacterial host facilitating attachment (Leprince & Mahillon, 2023). Then, the phage injects its genetic material via a hollow tail. A phage with a lytic life cycle will hijack the host's machinery to create more progeny until the host cell lysis, which results in cell death and the released phages continue their life cycle by infecting additional bacteria (Clokie et al., 2011). In contrast, a lysogenic phage will not immediately kill its host. However, once the phage genetic material is inside the bacterial host, the phage will recombine its DNA with the host's chromosome creating a prophage (Howard-Varona et al., 2017). Thus, the bacterial cell can continue to divide, creating many more bacterial cells containing the prophage (Howard-Varona et al., 2017). Subsequently, once triggered by a stimulus, the prophage becomes active and completes the remaining steps of the lytic cycle

(Libretexts, 2023). Therefore, determining a phage's structure and life cycle can aid in understanding the diversity of phages.

I.6 Diversity of Bacteriophages: Taxonomy and Classification

Phages are the most common and diverse entity, with an estimated 10^{31} phages in our biosphere (White et al., 2019). They can be isolated from all places in the environment where bacteria exist, such as soil, food, plants, sewage, feces, the Earth's crusts, plants, animals, freshwater, and salt water (White et al., 2019). The classification of a bacteriophage is dictated by the specific genera, species, and strains of bacteria it can infect (Gonzalez & Spencer, 1998). However, the diversity of phages is apparent in their morphology, structure, and genome composition and in the dynamic evolutionary relationships depicted through phage phylogenetics.

According to National Center for Biotechnology Information (NCBI) data, Enterococcus phages are classified into twelve families and one unclassified group (Dion et al., 2020). Exploring the diversity of phages involves analyzing their genomic characteristics and structural variations. As mentioned earlier, phages present an array of genomic compositions, with their genetic material packaged as either RNA or DNA within the capsid, and the capsid itself can be single-stranded or double-stranded (The Biology Notes Editors, 2022). Additionally, understanding the diversity spans to the morphology of the capsid, including polyhedral shapes found in *Microviridae*, *Corticoviridae, Tectiviridae*, *Leviviridae,* and *Cystoviridae*, filamentous structures characteristic of *Inoviridae,* pleomorphic forms observed in *Plasmaviridae,* and those associated with tails, as seen in the order *Caudovirales* (Dion et al., 2020). Furthermore, identifying the tail morphology, phage life cycle, and phage serotype contributes to identifying the diversity of phages. In addition, the range of environments where phages reside contributes to their assorted population dynamics and ecological roles. Nonetheless, the exceptional diversity is underscored by the number of phages currently discovered.

Additionally, phylogenetics is pivotal in understanding phage diversity, providing essential insights into the evolutionary relationships among these viral entities. By investigating the evolutionary dynamics within the currently sequenced *E. faecalis* phages, we can observe their genetic variations and illuminate the potential implications of host interactions.
Some critical components of phages, such as portal proteins and terminase large subunits, significantly contribute to comprehending phage diversity. Portal proteins are a multifunctional

structural segment found in the capsid of bacteriophages, which operates as a central channel for DNA packaging (Dedeo et al., 2019). Additionally, portal proteins play an integral role in sensing and transmitting signals during the viral life cycle. These proteins comprise distinctive structural domains, including the clip, stem, wing, crown, and barrel domains (Dedeo et al., 2019). Despite low sequence similarity, portal proteins display unique conservation in their fold across diverse phages. During the DNA packaging, the portal protein undergoes conformational changes in response to rising pressure from packaged DNA, acting as a sensor to denote when the capsid is complete (Dedeo et al., 2019). Understanding the phylogeny of portal proteins allows us to gain insight into the evolutionary relationships among phages, which illustrates both the conserved elements and structural variations (Dedeo et al., 2019). This knowledge assists in decoding the diversity of phages, highlighting their evolutionary history and the adaptive mechanisms that have contoured portal protein functions across various viral families.

On the other hand, the terminase large subunit is responsible for ATPase activity to supply energy and nuclease activity for packaging initiation and termination (Sun et al., 2012). Typically, the genome exists as a branched concatemer without unrestricted free ends, necessitating cleavage to create a free end for packaging initiation (Sun et al., 2012). Upon packaging fulfillment, the genomic DNA undergoes cleavage again, and the residual DNA goes to another vacant procapsid (Sun et al., 2012). Moreover, by understanding the phylogeny of the terminase large subunit, we can understand variations in the DNA packaging mechanisms, host specificity, and other aspects of phage biology.

As previously discussed, there are many phages that remain undiscovered and unclassified in comparison to the number of phages that exist. Many of these undiscovered phages can potentially be used as an alternative therapeutic for prevention and treatment of nosocomial infections that are highly resistant to multiple drugs. Therefore, phage discovery and understanding the diversity among phages is an essential area of research for the potential future development of alternative therapeutics for multiple-drug resistant bacteria.

I.7 Phage Application in Medicine and Biocontrol

Bacteriophages were first discovered by William Twort in 1915 (UC Health - UC San Diego, n.d.). In 1917, a microbiologist, Felix d'Herelle from the Institute Pasteur in Paris, coined the term “bacteriophage” in a paper he published explaining the lysis of bacteria by an invisible microbe (UC Health - UC San Diego, n.d.). In 1919, d'Herelle first tested the use of phage

therapy in a 12-year-old boy with severe dysentery by administering a phage cocktail (UC Health - UC San Diego, n.d.). With a single dose, the child fully recovered (UC Health - UC San Diego, n.d.). However, phage therapy research began to decline a few decades later when antibiotics were discovered and easily administered (UC Health - UC San Diego, n.d.). Today, we experience the immense repercussions of multiple-drug resistant infections caused from the decades of misuse and overuse of antibiotics. Therefore, the science community aims to find an alternative course of action through a more exemplary source, bacteriophages.

Phage abundance, limited host-specificity, ability to self-amplify, ability to degrade biofilms, and their low toxicity to humans and the environment make them a more desirable alternative as opposed to conventional antibiotic treatment. Ultimately, because phages can produce a high number of themselves during the bacterial killing process, and are self-limiting, a single dose of phage may be enough to kill off specific bacterial infection and destroy biofilms, a mechanism known as "auto-dosing" (Khalifa et al., 2015; Loc-Carrillo & Abedon, 2011). Additionally, phages show very little, if at all, evidence of disrupting the normal microbiome, in turn leaving little to no side effects (Khalifa et al., 2015). Moreover, the difference in phage-killing mechanism is due to phages' inherent specificity for their target host, which differentiates them from antibiotics. Unlike antibiotics, phages continue to exert their action until they have eliminated their target host, averting the consequence of antibiotic cross-resistance (Loc-Carrillo & Abedon, 2011). Finally, phages exhibit a low environmental impact in contrast to antibiotics (Loc-Carrillo & Abedon, 2011). Despite their natural origin, phages are more targeted and specific in their action, decreasing the likelihood of broad-spectrum effects on non-target organisms and ecosystems (Loc-Carrillo & Abedon, 2011).

Recently, due to the rise in antibiotic-resistances, phage therapy has shown an effective and promising treatment against highly resistant antibiotic infections (Nasr Azadani et al., 2020; Tkachev et al., 2022; Wright et al., 2009). As described in various mainstream news outlets as well as published studies a fifteen-year-old cystic fibrosis patient received phage therapy to treat a *Mycobacterium* infection (Howard Hughes Medical Institute, 2019). After six weeks of treatments, scans on her liver revealed that the infection had substantially cleared.

Additionally, due to the phage's ability to penetrate and disrupt biofilms they have potential use as part of a disinfectant protocol in hospitals for medical equipment and patient indwelling devices. However, this concept is not new as phages are already used in the food

industry to preserve the quality of food and reduce the spread of bacterial disease. For example, phages in the food industry are used in bio-sanitization to reduce and prevent the formation of biofilms on production equipment (Połaska & Sokołowska, 2019). Therefore, researchers should aim at continuing to contribute to phage identification to transfer the mechanism used in the food industry as an additional disinfectant protocol in hospitals.

I.8 Constraints in Enterococcus faecalis Bacteriophages Discovery

Enterococcus faecalis infection rate continues to increase and prove challenging to treat and prevent due to its multiple-drug resistance. Therefore, the search for bacteriophages of *E. faecalis* is a promising target for the use in phage therapy and biocontrol. Recent studies have isolated phages from environmental sources that are capable of infecting *E. faecalis* (Askora et al., 2020; Del Rio et al., 2021; Olsen et al., 2021; Tkachev et al., 2022). These sources include sewage water, freshwater streams, fecal samples, and cheeses (Askora et al., 2020; Del Rio et al., 2021; Olsen et al., 2021; Tkachev et al., 2022). However, when compared to other types of bacterial hosts such as *Mycobacterium* with 13,290 phages identified, only 100 phages for *E. faecalis* have a complete and partial genome sequenced and deposited in NCBI as of February 2024, indicating they are grossly understudied in relation to their ability to solve a global health burden (Song et al., 2021). The broad reliance on antibiotics as the prime treatment for various HAIs may contribute to the substantial oversight in studying phages that infect *E. faecalis*. Additionally, implications of why *E. faecalis* phages are so vastly understudied could be attributed to the lack of published data, difficulty isolating phage, and challenges extracting phage DNA.

Moreover, by increasing the number of identified phages, researchers can investigate the combined potential of phage "cocktails" to target multiple bacteria host ranges at once (Abedon et al., 2021). Therefore, the scientific community should aim investigations into the efficacy of synthetic bacteriophages, cocktails, and the synergistic use of bacteriophages with antibiotics to generate powerful therapeutics against ESKAPE infections. However, it is essential to understand the genetic diversity among these phages and mechanisms in which to isolate them. Thus, our primary aim is to contribute meaningful information to enhance our knowledge on the genetic diversity of the currently sequenced *Enterococcus faecalis* phages. In addition, through isolating *E. faecalis* phage from a new environmental source, we aim to demonstrate the diverse

environments in which these phages can be isolated. Thus, unveiling additional regions for *E. faecalis* phage discovery that has the potential for further research in therapeutic applications.

I.9 In Vitro and In Vivo Applications of Using Phage Therapy to Treat Enterococcus faecalis

Due to *E. faecalis'* pathogenic abilities, strong resistance mechanisms, and ease of transmission through hospitals, various studies have implemented laboratory techniques to test the capabilities of identified phages. To test the infecting capabilities of theses phages against *E. faecalis,* an array of research was carried out throughout these studies to determine the antimicrobial therapy against invitro infections of mice, biofilm inhibition or degradation, and applications in therapeutic treatments for periodontal disease. Based upon examination of these studies, scientist can acquire strategies to develop alternative therapies to fight against pathogenic *E. faecalis*.

Among these studies, one demonstrates that isolated phages of *E. faecalis* has potential use for antimicrobial therapy, as indicated in an infected mouse study (Tkachev et al., 2022). Researchers treated septic mice with phages from the *Herelleviridae* and *Siphoviridae* family, which resulted in their survival until the end of the experiment, whereas majority of the untreated mice died (Tkachev et al., 2022). Thus, this study indicates the studied phages are suitable for further examination as a treatment of septic *E. faecalis* infections.

In addition, among the limited studies available, most demonstrate applications of phages' ability to combat biofilms caused by *E. faecalis.* A recent study showcases the lytic activity of a *Siphoviridae* family phage known as V583. This phage's specificity and potent lytic activity are effective against all thirteen strains of vancomycin-resistant *E. faecalis* (VREfs) (Goodarzi et al., 2022). Ultimately, the most notable factor in quantifying the inhibitory capabilities of phage V583 against biofilms was done through performing a crystal violet stain and taking absorbance measurements over a seven-day period (Goodarzi et al., 2022). Thus, validating this phage's ability to successfully inhibit biofilm formation and disrupt mature biofilms, which further highlights its potential as a biocontrol agent. Additionally, this phage may be an ideal candidate as an alternative or supplementary approach to conventional antibiotic treatments for biofilm-associated infections caused by a diverse array of antibiotic-resistant *E. faecalis* strains. Subsequently, further research should aim to study the potential of this phage in invitro or clinical studies as an alternative therapeutic treatment to highly ARB *E. faecalis* infections.

Moreover, a study demonstrates a significant promise in using phages as a therapeutic agent against *E. faecalis* in periapical periodontitis (PP). Specifically, experiments tested the treatment potential of phage PEf771, a calcium hydroxide preparation, and 2% chlorhexidine gel in rats with PP (Bhardwaj et al., 2020). Throughout a consecutive seven-week period, bone destruction and inflammation were evaluated by micro-computed tomography and hematoxylin-eosin staining. Results indicated that the regions of bone destruction in rats treated with these agents were significantly smaller than those in the untreated group. Subsequently, the researchers performed further monitoring with imaging and histopathology assessments over three-consecutive weeks to observe the healing process of the rats. The results indicate that treatment modalities demonstrate the potential to promote the healing of apical lesions in infected *E. faecalis* PP rat model. Strikingly, the therapeutic effects of phage PEf771 on periapical inflammation increased over time, presenting a promising new approach for treating refractory PP using bacteriophages.

Most importantly, the results from the following studies yield a promising foundation for the future of alternative antimicrobial treatments. The studies discussed highlight the potent lytic ability of phages to fight against *E. faecalis* and its potential use for an alternative antimicrobial treatment and use in biocontrol. Furthermore, subsequent studies should aim at conducting experiments to determine the abilities of *E. faecalis* phage among hospital associated infections and sanitation protocols. Thus, giving rise to a promising alternative to treat such a prevalent and highly resistant bacterium.

II. Statement of Purpose

The traditional method of treating bacterial infections is mainly through antibiotics. However, bacteria are ever evolving through mutations and natural selection leading to a rise in antibiotic resistance. Antibiotics are a broad-spectrum way of treatment; they not only target the bacteria causing the infection, but also commensal bacteria. Additionally, antibiotic research is severely underfunded, leading to the problem rising faster than the development of new antibiotics. Phage therapy research aims to combat these issues through bacteria-specific treatment, each unique to the infection. In our efforts to advance the knowledge of *Enterococcus faecalis* phages, we have designed a two-fold objective. Our primary goal is understanding the diversity within the sequenced *E. faecalis* phages. We aim to decipher the complex relationships among these phages via a phylogenetic analysis, specifically focusing on portal proteins and large subunit terminases. This primary objective aims to contribute valuable insights into the genomic variations of *Enterococcus faecalis* phages deposited in GenBank.

Moreover, our secondary aim is to explore that *Enterococcus faecalis* phages can be isolated from a soil sample. By isolating phages from a new environment and applying purification techniques, we aim to increase the number of identified phages. While genetic characterization is pending, this effort is vital as only a limited number of *E. faecalis* phages have been documented. Expanding their diversity is crucial for a comprehensive understanding of phage ecology and implies resource efficiency. **Therefore, given that enterococci are an ESKAPE pathogen severely understudied and have the remarkable ability to persist in various environments and conditions, our dualistic research provides insights toward expanding our knowledge on the diversity and discovery of *Enterococcus faecalis* phages.**

II.1 Objective 1: To analyze phylogenetic relationships on the portal proteins and terminase large subunits of complete *E. faecalis* phage genome sequences deposited in GenBank.

Currently, the NCBI GenBank database contains a total of 100 complete and partial genome sequences of *E. faecalis* phages, which is a significantly low opposed to *Mycobacterium* with 2,390 phages sequenced (https://phagesdb.org/hosts/genera/1/) and given the pressing need to discover alternative therapeutics for a leading cause of infections. However, the most recent phylogenetic analysis of complete *E. faecalis* phage genomes were conducted in 2020 using 33 *Enterococcus* phage genomes (Nasr Azadani et al., 2020). Subsequently, the number of sequenced *E. faecalis* genomes has significantly increased within the past four years, leading to the need for a more up-to-data analysis. Thus, we aim to determine the relatedness of the current 100 *E. faecalis* genomes deposited in the GenBank database in hopes of uncovering the genetic diversity among this population of phages.

II.2 Objective 2: To discover and isolate an *Enterococcus faecalis* infecting bacteriophage from a soil sample.

Our preliminary literature review demonstrates the successful isolation of phages specific to *Enterococcus faecalis* from environmental sources such as diverse bodies of water, sewage, and a variety of dairy products (Askora et al., 2020; Del Rio et al., 2021; Olsen et al., 2021; Tkachev et al., 2022; White et al., 2019). However, our goal is to isolate *Enterococcus faecalis* phage from a soil sample, an achievement that has not yet been accomplished in the literature. Thus, we can further contribute to understanding the diverse ecological distribution of these phages by exploration within a soil environment.

III. Materials and Methods

III.1 Phylogenetics

The relatedness of 100 available *Enterococcus faecalis* phages with complete genome sequences was assessed by performing a phylogenetic analysis based on the portal proteins and terminase large subunits. First, "*Enterococcus faecalis* phages" was entered into the NCBI GenBank database. Then, to narrow the search, filters were applied to include only viruses with complete genomic DNA/RNA sequences from the INSDC (GenBank) source. The selection process focused on nucleotide sequences to accurately display *Enterococcus faecalis* phages with complete genome information. The compiled portal protein and terminase large subunit sequence data was recovered from GenBank files and created into two distinct plain text files—one for portal proteins and another for large subunit terminases. Each file contained the phage name followed by the corresponding sequence, which prioritized the longer sequence in cases of multiple variants. The analysis of these text files was performed on the Phylogeny.fr platform and comprised of the following steps as shown in Figure 1 for portal proteins and Figure 2 for terminase large subunit (Dereeper et al., 2008). Subsequently, the sequences were aligned with MUSCLE (v3.7) and configured for the highest accuracy by maintaining the MUSCLE default settings. Following the alignment, ambiguous regions (i.e., containing gaps and/or poorly aligned) were withdrawn with Gblocks (v0.91b) for the portal proteins by adhering to the following criteria:

- The minimum length of a block after gap cleaning equals 10.
- The final alignment did not permit any gap positions.
- Every segment with contiguous, nonconserved positions exceeding eight was excluded.
- The minimum number of sequences for a flank position equals 85%.

After alignment of the terminase large subunits, positions with gaps were removed from the alignment using built-in-curer. The phylogenetic tree was reconstructed utilizing the maximum likelihood (ML) method established in the PhyML program (*v3.1/3.0 aLRT*). The WAG substitution model was selected, presuming an estimated proportion of invariant sites *(of 0.000)* and four gamma-distributed rate categories to factor in the rate heterogeneity across sites. Subsequently, the data allowed for directly estimating the gamma shape parameter (gamma=3.990 for portal proteins and gamma=6.922 for terminase large subunits). The aLRT test (SH-Like) was employed to evaluate the reliability of the internal branch. Then, TreeDyn

(v198.3) was used to create and edit the graphical representation of the phylogenetic tree. Lastly, the newly constructed trees will be compared to the portal protein and terminase large subunit trees created in a previous study (Nasr Azadani et al., 2020).

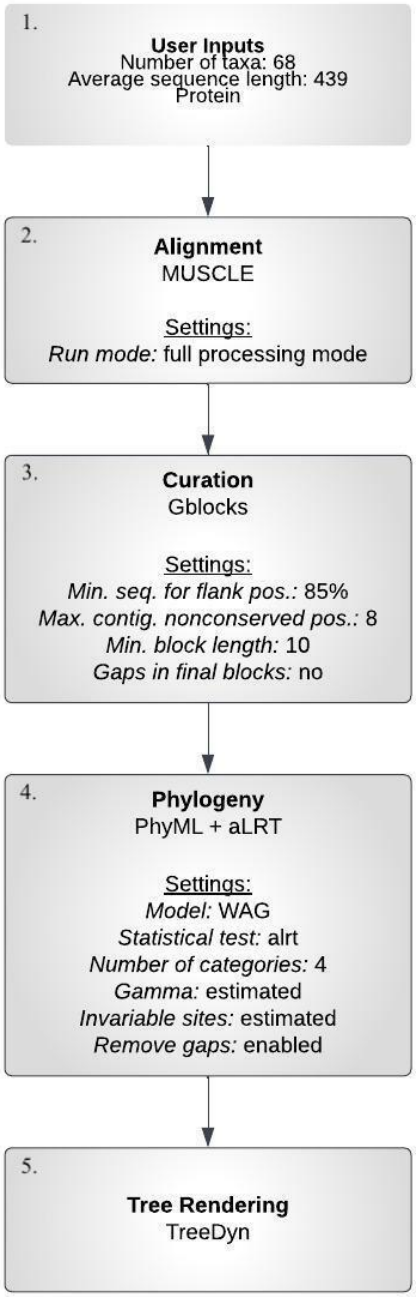


Figure 1: Flowchart of the steps for curating a maximum-likelihood phylogenetic tree on *Enterococcus faecalis* phage portal proteins. Step 1: Input 68 portal protein sequences. Step 2: Align sequences with MUSCLE. Step 3: Remove gaps or poorly aligned sequences with Gblocks. Step 4: Construct phylogenetic tree using PhyML (v3.1/3.0 aLRT) with WAG substitution model, four gamma-distributed rate categories (gamma=3.990), and aLRT test for internal branch reliability. Step 5: Generate phylogenetic tree using TreeDyn (v198.3)

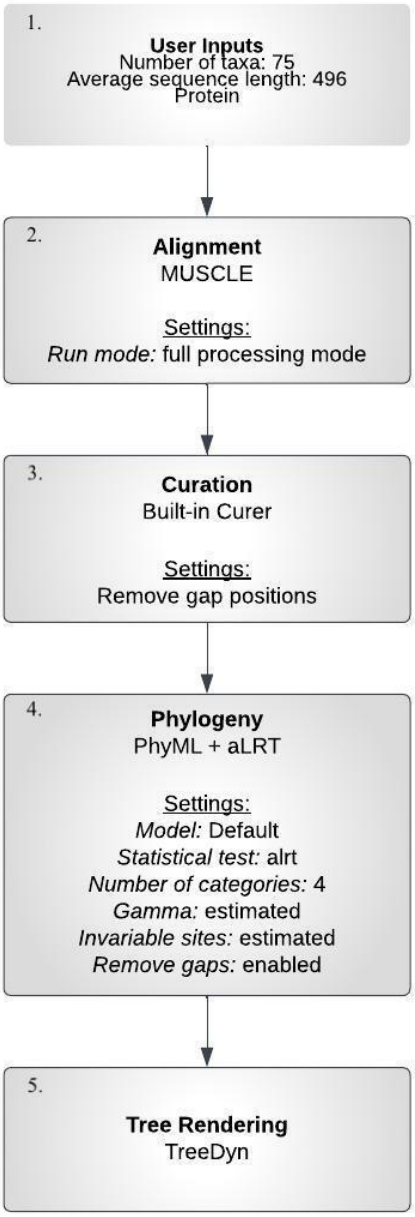


***Figure 2*: Flowchart of the steps for curating a maximum-likelihood phylogenetic tree on *Enterococcus faecalis* phage terminase large subunit.** Step 1: Input 75 terminase large subunit sequences. Step 2: Align sequences with MUSCLE. Step 3: Remove gap positions using built-in curer. Step 4: Construct phylogenetic tree using PhyML (v3.1/3.0 aLRT) with WAG substitution model, four gamma-distributed rate categories (gamma=6.922), and aLRT test for internal branch reliability. Step 5: Generate phylogenetic tree using TreeDyn (v198.3)

III.2 Bacterial Strain and Conditions

E. faecalis OG1RF was used as the phage host due to its ability to form biofilms (Núñez-Sánchez et al., 2020). This particular strain of bacteria was generously provided by Breck A. Duerkop, Ph.D. from the Department of Immunology and Microbiology at the University of Colorado Anschutz School of Medicine. The frozen cell stock was streaked onto an agar plate composed of Brain Heart Infusion supplemented with 1M of $MgSO_4$ (BHI-$MgSO_4$) and incubated at 37°C overnight to allow for the growth of single colonies. Then, a single isolated colony was inoculated and cultured in 5mL of BHI broth overnight at 37°C.

III.3 Phage Isolation from Soil

A soil sample was collected and filled to the 15mL mark of a 50mL conical tube. TMG (Tris-HCl, magnesium sulfate, gelatin pH 7.4) phage buffer was added to the 35mL mark in the conical tube containing the soil sample and inverted several times to ensure an even mixture. Subsequently, the sample was incubated at 37℃ with shaking at 220 rpm for 1-2 hours. Following incubation, the sample was left untouched for 30 minutes to allow settlement of debris and filtered (0.22 μm filter) to remove any soil particles, resulting in ~20-25 mL of filtrate containing the phage supernatant. Once the supernatant was obtained, a spot test was performed (III.4 described below) to test for the presence of phage. If no presence of phage was detected, the sample was enriched with 0.5mL of *E. faecalis* OG1RF and incubated at 220 rpm in a sterile aerobic environment for two days at 37℃. Following the two days of incubation the supernatant was aseptically filtered (0.22 μm filter) into a sterile 50mL conical tube and stored at 4℃.

III.4 Spot Test from Direct and Enriched Soil Sample

A spot test is utilized to quickly identify the presence/absence of phage. First, a bacterial lawn using *E. faecalis* OG1RF was prepared using aseptic technique by-transferring 3 mL of molten BHI-$MgSO_4$ top agar to a culture tube containing 500μL of *E. faecalis* OG1RF. The top agar-bacteria mixture was flooded onto a BHI-$MgSO_4$ agar plate and allowed to solidify completely. On the prepared bacterial lawn, 10μL of the enriched sample was added in one "spot" in a particular quadrant of the plate, along with phage VPE25, isolated and supplied by Breck Duerkop Ph.D., as a positive control to another quadrant. The plate was incubated at 37 °C, aerobically overnight. The presence of phage was determined by the presence of spots, which are zones of clearing on the bacteria lawn.

III.5 Preparation for Plaque Assay and Serial Dilution

A single colony of *E. faecalis* OG1RF was inoculated in 5mL of BHI broth overnight at 37°C. The following day the broth was divided into three microcentrifuge tubes and centrifuged for 10 minutes at 4,000 RPMs. The supernatant was discarded, and the remaining pellet was vortexed to break it up. Afterwards an equal amount of TMG phage buffer was added to each tube and briefly vortexed. The resulting 5mL mixture was then added to a clean 15mL conical. Then 500µL of this mixture was transferred to 4.5mL of TMG phage buffer and the tube was carefully inverted. Finally, 130µL of TMG bacteria mixture was transferred into ten sterile microcentrifuge tubes to begin serial dilution from 10^0 to 10^{-8}. Phage VPE25 was used as a positive control and the TMG bacterial mixture as a negative control.

To prepare the discovered phage from the enrichment culture, 90µL of TMG buffer was added to each tube containing the TMG bacteria mixture. Next, 10µL of the enrichment culture was added to the microcentrifuge tube labeled 10^0, then continued with a 10-fold serial dilution until 10^{-8}.

III.6 Plaque Assay for Phage Purification

Following the confirmation of phage from the spot test, the supernatant collected from the enriched isolation was used to perform a plaque assay to assess the presence of multiple phages. After preparing and serially diluting the enriched phage sample (III.5 described above) each phage and TMG bacterial mixture was transferred to 3 mL of molten BHI-$MgSO_4$ top agar and flooded onto a BHI-$MgSO_4$ agar plate and allowed to solidify completely. Once the plates were solidified, they were incubated overnight at 37°C.

The next day, multiple phages were detected, which initiated the purification of the phage by picking one plaque about 2mm in diameter and at least 1.5 cm away from other plaques. Using a micropipette with a 200µL sterile tip, the center of plaque was carefully pierced to excise the agar that contains the phage. The end of the tip with the phage and agar was transferred into the microcentrifuge tube containing the bacterial and TMG mixture (described in III.5 above). Subsequently, the above steps were repeated for another three rounds to serially dilute the phage and perform plaque assays on the dilutions to generate a clonal phage population.

III.7 Collecting Plate Lysates from Webbed Plates for Phage Amplification

Once a single clonal phage population was obtained, it was necessary to generate a highly concentrated liquid phage sample to use for future DNA extraction, electron microscopy, and

long-term storage. Briefly, we identified a webbed plate from the last round of amplification (described in III.6 above). Webbed plates exhibit densely packed confluent plaques with a "web" of bacteria left between them. The webbed plate was used for high titer lysate collection. To the webbed plate, 8mL of sterile phage buffer was added. The plates at 4°C overnight. After incubation, a syringe was used to draw up the phage lysate. The phage lysate was filtered and sterilized using a 0.22 um syringe filter. The lysate was then collected and placed in 4 °C for long term storage to be used for future experimentation (Poxleitner et al., n.d.).

III.8 Spot Titer to Predict Dilutions for Lysate Titer

Following lysate collection, it is necessary to determine the total number of phage particles (known as titer) in the lysate. Because the lysate titer is unknown because it was pooled from multiple webbed plates, it is hard to judge which dilutions should be made for titer assessment. Spot titer utilizes one plate to approximate the titer from various dilutions, reducing the number of plates needed initially. A plate with a lawn of bacteria was made by pipetting 3 mL of molten BHI-$MgSO_4$ top agar to a culture tube containing 500µL of host bacteria and pipetting the solution onto a BHI-$MgSO_4$ agar plate. The lysate generated in III.7 will be serially diluted from 10^0 to 10^{-8} (previously described in III.5). The serially diluted samples (3µL each) were "spotted" on the prepared bacterial lawn. The liquid spots were allowed to absorb the bacterial lawn undisturbed for 30 minutes and then incubated overnight at 37 °C. After incubation, the highest dilution factor was determined by the highest factor that contained observable plaques. Subsequently, a full plate titer was used to determine an accurate lysate titer concentration.

III.9 Full Plate Titer

The spot titer gives an approximate titer for the collected and pooled lysate. However, it is not accurate as 3µL were plated. Therefore, it is necessary to confirm the final titer using serial dilutions followed by a plaque assay. Based on the titer generated from the spot titer, only serial dilutions 10^{-6} to 10^{-8} were plated because 10^{-7} was the last dilution that there was an observable plaque. The full plate titer is calculated from the formula:

$\left(\frac{pfu}{mL}\right)\left(\frac{\#plaques}{10\mu L}\right) x \left(10^3 \frac{\mu L}{mL}\right) x\ serial\ dilution\ factor$ (Poxleitner et al., n.d.)

IV. Results

IV.1 Phylogenetic Analysis Demonstrating Diversity of E. faecalis Phage Portal Proteins and Terminase Large Subunits by Clustering Genomes by Size, Family, and Genus

A previous study, completed four years ago, shows the diverse relationship of 33 *Enterococcus* phage portal proteins and terminases through a comparative phylogenetic analysis (Nasr Azadani et al., 2020). A significant increase in the number of sequenced *Enterococcus faecalis* warrants a new comparative phylogenetic analysis of their portal proteins and terminase large subunits. Through this analysis, we can expect to see a better clustering of phage groups as more intermediate phages or phage groups become available. To address this hypothesis, we re-created and explored the phylogenetic relationship between 100 sequenced phages using the portal protein and terminase large subunit as mentioned above in Figure 1 and Figure 2. Publicly available *Enterococcus faecalis* phages with a complete genome were used for this analysis. Supplement Table 1 and 2 show the aligned sequences used to construct the phylogenetic portal protein tree and phylogenetic terminase large subunit tree. Subsequently, our new trees revealed more variations and new relationships in comparison with the earlier study (Nasr Azadani et al., 2020). In contrast to the earlier findings of 33 portal proteins and terminase proteins, this analysis indicates that only 24 portal proteins and 18 terminase remain conserved. Thus, there are 40 new phages with portal proteins and 57 phages with terminase. Additionally, branches in this new analysis show closer relationship among the phages than shown in the previous study (Nasr Azadani et al., 2020).

In the portal protein phylogenetic tree (Figure 3) all phages belong to the order of *Caudoviricetes.* However, the majority of phages appear from the *Siphoviridae* family and are grouped within three genera (*Saphexavirus, Phifelvirus,* and *Efaquatrovirus*). The average genome size of the phages from this tree that belong to the *Siphoviridae* family is 43,661 bp and are as follows for those grouped within each genus; *Saphexavirus* 57,092 bp, *Phifelvirus* 38,632 bp, and *Efquatrovirus* 40,082 bp. Additionally, the second most common family of phages appear grouped together from the *Herelleviridae* family and belong to two genera (*Kochikodavirus* and *Schiekvirus*). The average genome size of phages in this tree belonging to the *Herelleviridae* family is 131,407 bp and are as follows for the two grouped genera *Kochikodavirus* 142,366 bp and *Schiekvirus* 150,0252 bp. Furthermore, phages vB_EfaS_IME197 and EFC-1 are grouped together and have an average genome size of 40,692

bp, but their family and genus are undetermined. In addition, a branch containing only phage phiEf11with a genome size of 42,822 bp has an undetermined family and genus.

Similar to the portal protein phylogenetic tree, all phages in the terminase large subunit phylogenetic tree (Figure 4) also belong to the order of *Caudoviricetes.* Additionally, most phages are from the *Siphoviridae* family and are grouped within the same three genera (*Saphexavirus*, *Phifelvirus*, and *Efaquatrovirus*) as observed in the portal protein phylogenetic tree. The average genome size of the phages in the terminase tree that belong to the *Siphoviridae* family is 44,170 bp and are as follows for each genus; *Saphexavirus* 57,232 bp, *Phifelvirus* 38,358 bp, and *Efquatrovirus* 39,188 bp. Additionally, the second most common family of phages are from the *Herelleviridae* family and are grouped by two genera (*Kochikodavirus* and *Schiekvirus*). The average genome size for the *Herelleviridae* family is 145,486 bp and is as follows for each genus, *Kochikodavirus* 149,009 bp and *Schiekvirus* 140,554 bp. Furthermore, three additional branches show phages that belong to the *Andrewesvirina* family and *Vipetofemvirus* genus with a genome size of 85,520 bp, the *Rountreeviridae* family and *Copernicusvirus* genus with a genome size of 18,466 bp, and the *Autogrphiviridae* family has a genome size of 40,338 bp but does not have an identified genus.

Moreover, comparison of both trees indicates that phages Nonaheksakonada, vB_EfaS-DELF1, and EFAP-1 do not have a determined genus but are grouped within branches of phages that belong to the *Efquatrovirus* genus. Lastly, phages from both tree's (LY0322, FX417 vB_EfaS_Ef785CC, vB_EfaS_Ef785CS, EfaCPT1) and from the portal protein tree (LY0323, SANTOR1, phiSHEF2, phiSHEF4, and phiSHEF5) are all grouped with phages belonging to the *Efquatrovirus* genus, but their family is undetermined. In summary, the data reveal taxonomic diversity among *Caudoviricetes* order of *E. faecalis* phages, with distinct groups based on family and genus, as well as variation in genome size within and between these groups.

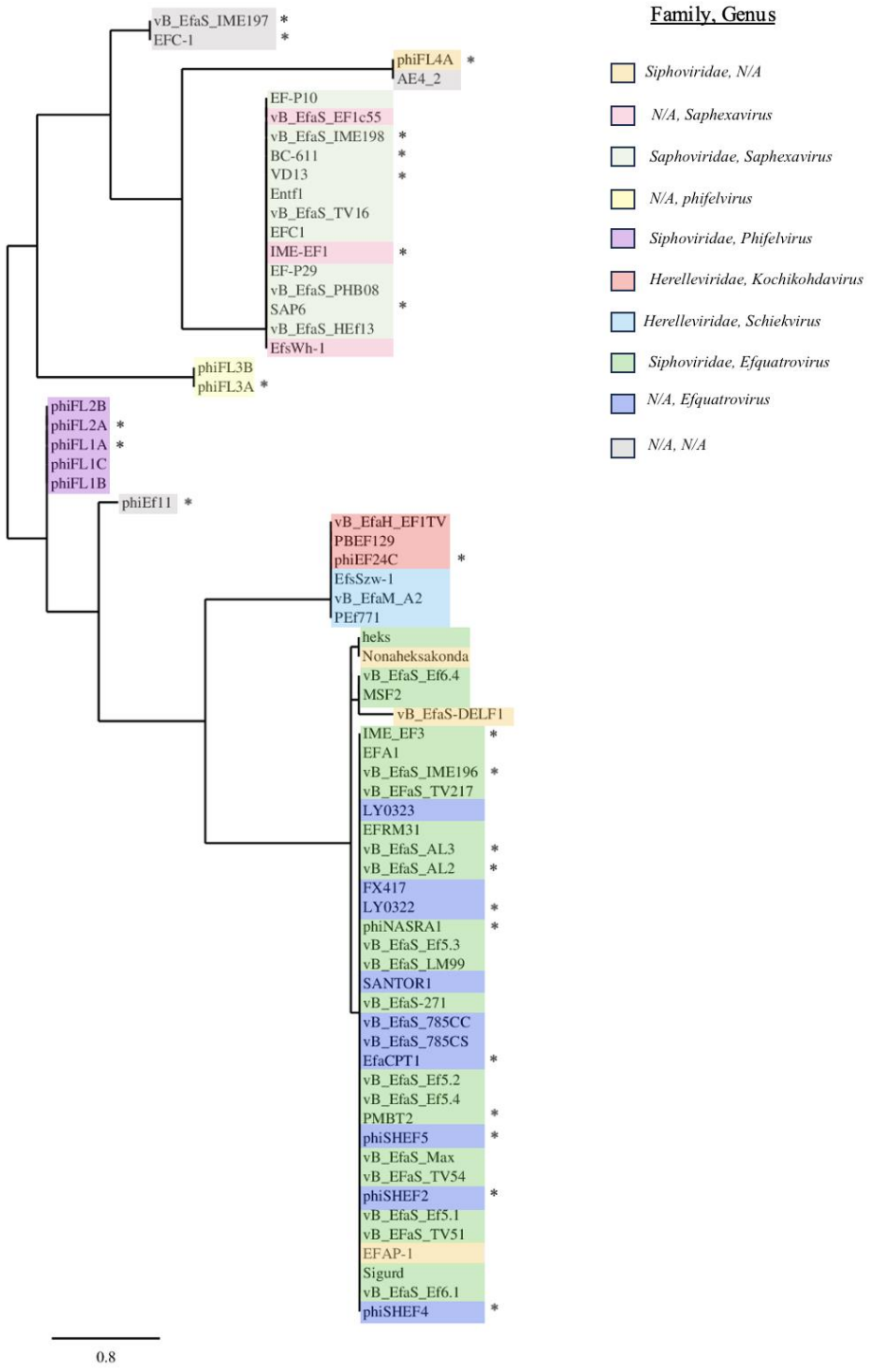


Figure 3: The portal protein phylogenetic tree of E. faecalis phages exhibits clustering based on evolutionary relationships, organized by groups of family and genus. Portal protein sequences of *E. faecalis* phage's complete genome was extracted, then aligned and trimmed on Phylogeny.fr platform to produce a phylogenetic tree. As shown '*' represent phages analyzed from a previous 2020 study (Nasr Azadani et al., 2020). Color coded bar represents the family and genus of the bacteriophage (key is at top right).

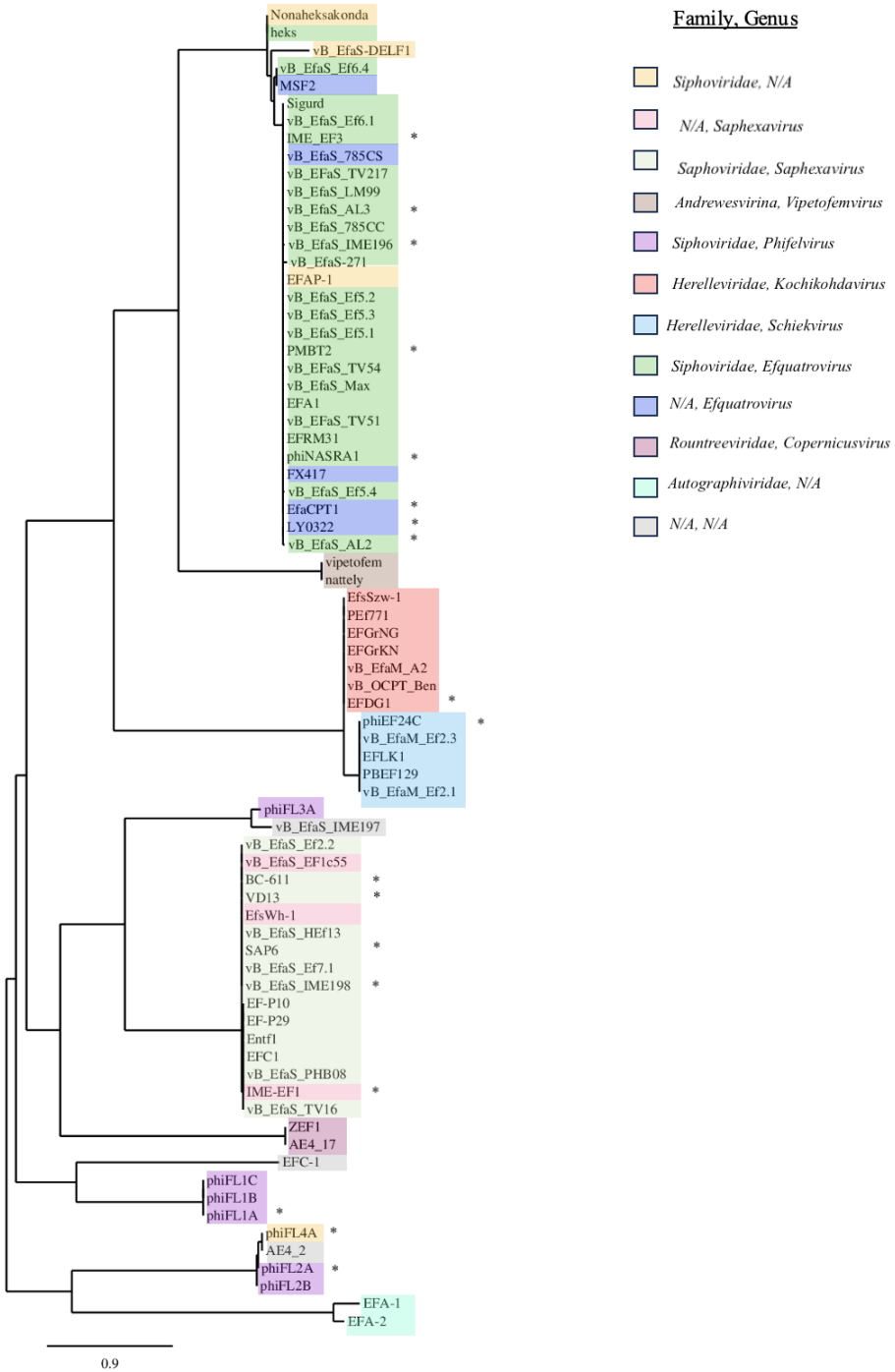


***Figure 4*: The terminase large subunit phylogenetic tree of *E. faecalis* phages exhibits clustering based on evolutionary relationships, organized by groups of family and genus.** Terminase large subunit sequences of *E. faecalis* phage's complete genome was extracted, then aligned and trimmed on Phylogeny.fr platform to produce a phylogenetic tree. As shown '*' represent phages analyzed from a previous 2020 study (Nasr Azadani et al., 2020). Color coded bar represents the family and genus of the bacteriophage (key is at top right).

IV.2 E. faecalis Phage CLH23 Isolated from Soil Exhibits a Lytic Life Cycle

Previous research has successfully isolated *Enterococcus* phage from diverse environments such as sewage water, freshwater streams, fecal samples, and cheeses (Askora et al., 2020; Del Rio et al., 2021; Olsen et al., 2021; Tkachev et al., 2022). However, our exhaustive review of the primary literature does not indicate the isolation of *Enterococcus faecalis* phage from soil. Consequently, we hypothesize that because phages exhibit a diverse range of genetic and ecological characteristics, some phages may also be found in soil environments because *Enterococcus* strains should also be present in soil where there is fecal matter. Therefore, we employed a direct and enriched isolation method to achieve successful isolation of a lytic and lysogenic *Enterococcus faecalis* phage under conditions described in Table 2 from soil samples described in Table 3. Soil samples that tested negative for the presence of *E. faecalis* phage are detailed in Table 4. The isolated phage was first assessed for activity against *E. faecalis* OG1RF using spot tests. The phage's lytic properties were evident by the formation of clear zones on the bacterial lawn (Figure 5, top), opposing the possibility of being a lysogenic phage (HNJ19), which would display turbid or "fuzzy" edges as phage shown in (Figure 5 bottom). This lytic phage was named ‘CLH23’ and underwent three rounds of purification, which produced small but clear plaques of approximately 1.5 ± 0.29 mm diameter of similar morphology on top agar (Figure 6). By repeated plaque purification method, a stock of $2.4x10^{10}$ PFU/ mL was prepared. Therefore, this result indicates that *E. faecalis* phage has a greater ecological diversity as it can successfully be isolated from an alternative source, such as soil and manure.

Table 1: Growth Conditions for *E. faecalis* Phage.

Growth Conditions for Isolation of *Enterococcus faecalis* Phage									
Phage Identifier	**Isolation source**	**Host Strain**	**Growth Medium**	**Culture Temperature**	**Incubation Period**	**Isolation Method**	**Detection Method**	**Isolation Success**	**Life Cycle**
CLH23	Soil	OG1RF	BHI	37°C	18 hours	Enriched	Spot Test	Yes	Lytic
HNJ19	Manure	OG1RF	BHI	37°C	18 hours	Enriched	Spot Test	Yes	Lysogenic

Table 2: Soil Sample Conditions on Date of Collection.

Soil Sample Conditions						
Phage Identifier	**GPS Coordinates**	**Collection Date**	**Soil Type**	**Moisture Content**	**Outside Temperature**	**Organic Matter**
CLH23	26° 3' 53.1" N80° 17' 2.832" W	5/23/2023	Loamy	Medium	31°C	Avian-enriched
HNJ19	26° 3' 53.1" N80° 17' 2.832" W	5/23/2023	Manure	High	31°C	Fecal-enriched

Table 3: Information for Soil Samples That Tested Negative for The Presences of *E. faecalis* Phage.

Sample Conditions					
GPS Coordinats	**Collection Date**	**Soil Type**	**Moisture Content**	**Outside Temperature**	**Organic Matter**
26.19765°N, 80.14861°W	4/11/2023	Silt	Light	24°C	Plant Litter
28.99702°N, 82.24268°W	4/17/2023	Silt	Medium	26.1°C	Animal Waste
28.99702°N, 82.24268°W	4/17/2023	Silt	Medium	26.1°C	Animal Waste
26.0832°N, 80.24871°W	4/25/2023	Loamy	Light	26.7°C	Rocks and Plant Litter
26.12734°N, 80.23832°W	4/25/2023	Loamy	Light	26.7°C	Plant Litter
26.0832°N, 80.24871°W	4/25/2023	Loamy	Light	26.7°C	Plant Litter
26.10189°N, 80.23431°W	5/1/2023	Silt	Medium	26.1°C	Animal Waste
26.12712°N, 80.23852°W	5/1/2023	Loamy	Dry	26.1°C	Plant Litter
26.12709°N, 80.23834°W	5/1/2023	Loamy	Dry	26.1°C	Plant Litter
26.12824°N, 80.23812°W	5/1/2023	Loamy	Dry	26.1°C	Plant Litter
26.12824°N, 80.23812°W	5/6/2023	Silt	Dry	25.6°C	Plant Litter
26.15403°N, 80.14507°W	5/6/2023	Peaty	Medium	25.6°C	Coir and Plant Litter
26.05435°N, 80.11192°W	5/6/2023	Sand	High	25.6°C	Detritus and Shell
26.05626°N, 80.11404°W	5/6/2023	Sandy Clay Loam	High	25.6°C	Detritus, Shell, and Plant Litter
26.10237°N, 80.23402°W	5/8/2023	Silt	Dry	25°C	Animal Waste
26.05248°N, 80.24709°W	5/8/2023	Sandy	Dry	25°C	Rocks
26.05485°N, 80.24673°W	5/8/2023	Sandy	Medium	25°C	Plant Litter
26.02504°N, 80.25035°W	5/15/2023	Loamy	Medium	25.7°C	Animal Waste
26.12640°N, 80.23851°W	5/23/2023	Clay	High	31°C	Composte

***Figure 5*: Discovery of lytic and lysogenic *E. faecalis* phage from soil.** Spot test on double layer BHI-$MgSO_4$ agar of enriched samples with *E. faecalis* OG1RF show phage CLH23 (**top**) is lytic and phage HNJ19 (**bottom**) is lysogenic.

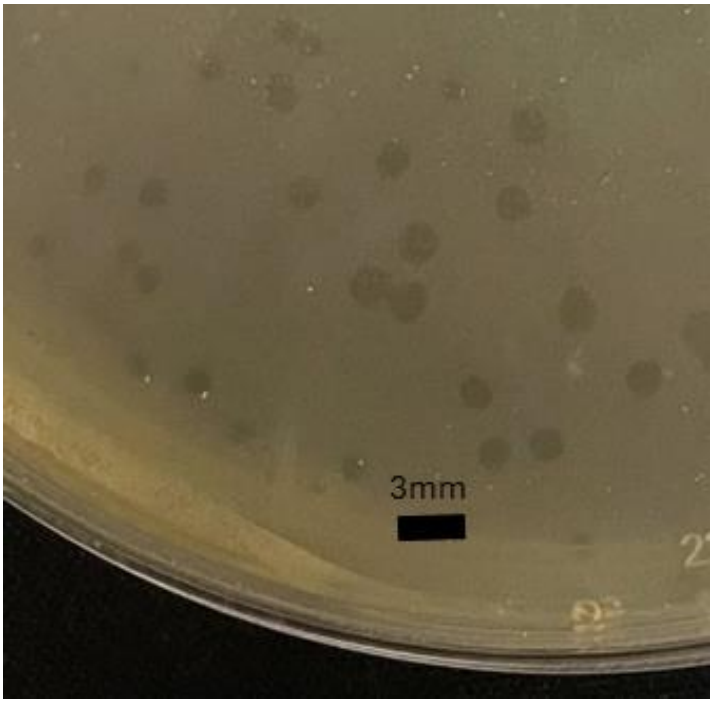


***Figure 6*: CHL23 is a lytic phage.** Three rounds of purification by plaque assays yielded lytic plaques of approximately 1.5 ± 0.29 mm in diameter of phage CLH23 on double layer BHI-$MgSO_4$ agar.

V. Discussion

In this study, we examined *Enterococcus faecalis* phage diversity through phylogenetic analysis using the portal protein and terminase large subunit, in addition to the isolation of a phage in a new environmental source. Here, we report three distinct observations that broaden our knowledge of *E. faecalis* phage diversity. First, the distinct clustering of phages in the phylogenetic portal protein tree and phylogenetic terminase large subunit tree is indicative of their genetic relatedness. Second, portal proteins and terminase large subunits serve as molecular markers, providing insights into the evolutionary dynamics and genetic diversity of *Enterococcus faecalis* phages. Third, isolating *E. faecalis* phage from a novel source highlights their prevalence in diverse ecological niches, suggesting potential implications for phage ecology and biotechnology. Together, these results highlight the significance of phylogenetic analysis for understanding and expanding *E. faecalis* phage diversity and the potential of exploring diverse environmental sources, emphasizes the importance of continued research to address emerging challenges in treating *E. faecalis* in healthcare.

What are the implications of these findings? *Enterococcus* is one of the leading ESKAPE pathogens, posing a severe threat to global health and is becoming increasingly resistant to current treatments. Therefore, current research efforts focus on expanding the understanding of phage diversity and identifying novel phages to aid in the future development of alternative therapeutics. Previous studies demonstrate *Enterococcus* phage diversity through phylogenetic analysis and isolation of novel phages (Nasr Azadani et al., 2020). Since nucleotide and gene content tend to show greater diversity in phages as compared to the protein structures, we used the protein sequences presumed to be highly conserved across phage families to perform phylogenetic analysis. Our research expands these studies by comprehensively analyzing the genetic diversity of currently sequenced *E. faecalis* phages. Specifically, we observed variations among different phages, indicating potential differences in their virulence, host range, and mechanisms of action. The first correlation addresses the clustering of *E. faecalis* phages within the phylogenetic portal protein tree and phylogenetic terminase large subunit tree. A previous study documents the phylogenetic trees of 33 *Enterococcus* phage portal protein and terminase large subunits (Nasr Azadani et al., 2020). Our phylogenetic tree analysis of 68 portal proteins and 75 terminase large subunits exhibits that the newly added *E. faecalis* phages are evolutionarily similar to the previous study. Therefore, inferring a shared evolutionary history

among these phages is possible, indicating conserved genetic traits and potential functional similarities. Additionally, we identified unique genetic sequences in specific phages belonging to a family that has not been previously documented, suggesting the presence of novel genetic elements or functional proteins. However, our results also indicate that not all phages from the previous study remained present in our phylogenetic trees, which could be because we used a stringent criterion of only including the full-length protein sequences and eliminating any truncated sequences. Using a larger diversity of phage sequences tends to resolve the protein structures to provide additional information on new relationships between the phages under study. This might result in the identification of intermediate ancestors that enables better clustering of phylogenies. Therefore, by uncovering these genetic diversities, our study provides valuable insights into the complex interplay between phages and their bacterial hosts, paving the way for targeted therapeutic interventions and developing more effective treatment strategies against antibiotic-resistant *E. faecalis* infections.

Second, the role of portal proteins and terminase large subunits serves as a promising molecular marker for *E. faecalis* phage. Portal proteins and terminase large subunits are essential for stages of the viral life cycle, such as DNA packaging and genome entry/exit (Lin et al., 1999). In addition, portal proteins have a tryptophan residue ring, which is highly conserved across diverse phages and even present in herpesviruses like HSV1, indicating a fundamental role in viral replication (Dedeo et al., 2019). This conservation makes portal proteins excellent phylogenetic markers, providing stable data for constructing evolutionary trees and revealing genetic relationships among viruses. In contrast, we noticed the terminase large subunit has a highly conserved function but a less conserved sequence. Therefore, understanding the role of these proteins in phage taxonomy will aid researchers in facilitating the selection of phages in studying their biotechnological and therapeutic applications, such as novel antiviral therapies and phage engineering.

Nevertheless, the study of phage taxonomy has presented difficulties due to the absence of a universally conserved marker gene spanning all phage families. A previous study utilized a variety of single genes to determine their viability as a molecular marker for identifying phage taxonomy (Sørensen et al., 2020). By targeting phages that infect *E. faecalis*, we can provide insights directly applicable to understanding and combating *Enterococcus faecalis* infections. Furthermore, while the previous study has laid the groundwork for using these molecular

markers, our research further advances the field by applying them to a relevant and understudied area, ultimately contributing to a more thorough understanding of phage diversity and evolution. Therefore, our research used the portal protein and terminase large subunit as molecular markers to help determine *E. faecalis* phages family and genus based on clustering organization.

Moreover, the portal protein and terminase large subunit analysis results indicate a clear distinction among genera of the *Siphoviridae* family (*Saphexavirus, Phifelvirus,* and *Efquatrovirus*) and *Herelleviridae* family (*Kochikohdavirus* and *Schiekvirus*). Additionally, the terminase large subunit indicates three clear distinctions of clustering between novel *E. faecalis* phage families (*Andrewesvirina, Rountreeviridae,* and *Autographiviridae*) that are not compared in other studies (Nasr Azadani et al., 2020). In comparison, our study indicates a greater grouping amongst these phages within their family and genera, whereas the previous study looked at less phages and showed less grouping between the phages (Nasr Azadani et al., 2020). However, despite specific phages lacking an identified family or genus, it is reasonable to infer their potential affiliation with a particular family or genus based on the clustering patterns observed in our phylogenetic analysis. Therefore, our research further expands on the possibility of a single-gene-based analysis as a valuable preliminary indication of a phage's taxonomic cluster. Nonetheless, relying solely on a single gene may exclude phages in which the gene is not detected in their genome. Thus, a whole-genome-based analysis should accompany a single-gene analysis for a more comprehensive understanding of phage taxonomy.

In addition, we further accentuate these studies by isolating phage CLH23 and HNJ19, which is capable of infecting *E. faecalis* OG1RF, a biofilm-producing bacterium, from a previously unexplored environmental source. Several studies note the successful isolation of *E. faecalis* phage from a variety of environmental samples, such as sewage water, freshwater streams, fecal samples, and cheeses (Askora et al., 2020; Del Rio et al., 2021; Olsen et al., 2021; Tkachev et al., 2022; White et al., 2019). Therefore, our discovery of *E. faecalis* phage CLH23 from soil further reinforces these phages' remarkable diversity within the environment. Thus, our finding highlights the potential of discovering *E. faecalis* phages in a broader range of environmental samples. Additionally, identifying *Enterococcus faecalis* phages in soil supports a more extensive diversity of phylogenetic trees of *E. faecalis* portal proteins and terminase large subunits. Including *E. faecalis* phages isolated from the soil in a phylogenetic analysis would enhance the representation of their diversity and evolutionary relationships.

What are the caveats of trying to isolate *E. faecalis* phage from soil? While discovering *Enterococcus faecalis* phage in soil suggests an exciting opportunity for further study, isolating them from any soil sample may pose challenges. The optimal soil environments for *E. faecalis* phage isolation may include fecal matter, compost bins, or soil mixed with manure, as these environments are more likely to harbor *Enterococcus faecalis* bacteria. Therefore, searching for specific soil areas with these characteristics is necessary to increase the chance of encountering a sample with *Enterococcus faecalis* bacteria. Nevertheless, despite these challenges, isolating *E. faecalis* phage from soil opens new avenues for researchers to explore previously uncharted territory and gain valuable insights into their diversity, distribution, and ecological roles. Thus, our findings contribute to an advancement in phage research and augments a new area researchers can explore to expand the number of identified *E. faecalis* phages.

Furthermore, *E. faecalis* phage VPE25 (positive control used in this study) shows a much smaller plaque size of 0.44± 0.27 mm in diameter (Supplemental Figure 1) compared to our phage CLH23 that has a plaque size of 1.5 ± 0.29 mm in diameter (Figure 6). What could this difference in plaque sizes between these two phages mean? Previous studies show differences in plaque sizes may indicate variations in phage morphology and lytic cycle efficiency. A study suggests that plaque size may correlate with phage adsorption, influenced by factors such as the presence of side-tail fibers (Stf) and tail fiber proteins (J alleles) (Gallet et al., 2011). Specifically, higher adsorption rates tend to result in larger plaques, while lower adsorption rates are associated with smaller plaques (Gallet et al., 2011). This correlation implies that the ability of phages to attach to host cells efficiently can impacts their capability of infecting and lysing the cell, thereby affecting plaque size. In addition, another study demonstrates that larger plaques indicate a faster lytic cycle, resulting in the rapid destruction of host cells and consequent expansion of clear zones on the bacterial lawn (Ameh et al., 2020). Therefore, we can infer that the significant difference in plaque sizes between phages VPE25 and CLH23 suggests varying adsorption rates, morphology, and lytic cycle efficiencies. The smaller plaque size of VPE25 implies a potentially slower lytic cycle and lower adsorption rate, while the larger plaque size of CLH23 indicates a faster lytic cycle and higher adsorption rate. These differences may impact their ability to attach to host cells, infect, and lyse cells efficiently, ultimately influencing their plaque sizes on bacterial lawns. However, further studies should aim to measure and compare

adsorption rates, conduct morphological analysis, and study the lytic cycle dynamics between phages VPE25 and CLH23 to determine this difference in plaque size.

What are the prospective research objectives for phage CLH23 and HNJ23 that we isolated? Due to phages CLH23 and HNJ19 ability to infect *E. faecalis* OG1RF, future work should aim to characterize it through electron microscopy and whole genome sequencing. Electron microscopy will aid in characterizing this phage at the ultrastructural level, offering us insight into its morphology and structure. Additionally, performing whole-genome sequencing will enable us to conduct a comparative analysis of the genome of this phage. This analysis can further enhance our current phylogenetic investigation, elucidating the taxonomic cluster to which our phage belongs. Moreover, the analysis of its genome enables the identification of virulence factors, such as toxins, adhesion proteins, and immune evasion mechanisms, which contribute to phage pathogenicity. As a result, studies can aim to understand what the implications of phage CLH23 are in the disruption and inhibition of biofilms.

VI. Conclusion

The present study significantly enhances our knowledge of *Enterococcus faecalis* phage diversity. By unraveling the genetic relationships within the broader context of phage evolution, our phylogenetic analysis of portal proteins and terminase large subunits sheds light on the diverse nature of these phages. Furthermore, our findings emphasize the potential value of portal proteins and terminase large subunits as phylogenetic markers, contributing to our knowledge of *E. faecalis* phage's evolutionary relationships and taxonomic classification. Future research should aim to characterize phage CLH23 through electron microscopy and whole genome sequencing, which promises to deepen our understanding of its morphology, genetic composition, and potential applications. As phage research continues to evolve, the findings presented here pave the way for further studies and foster a deeper understanding of the intricate diversity between *E. faecalis* phages.

VI. References

Abedon, S. T., Danis-Wlodarczyk, K. M., & Wozniak, D. J. (2021). Phage cocktail development for bacteriophage therapy: Toward improving spectrum of activity breadth and depth. *Pharmaceuticals, 14*(10), 1019. https://doi.org/10.3390/ph14101019

Ahmed, M. O., & Baptiste, K. E. (2018). Vancomycin-resistant enterococci: A review of antimicrobial resistance mechanisms and perspectives of human and Animal Health. *Microbial Drug Resistance, 24*(5), 590–606. https://doi.org/10.1089/mdr.2017.0147

Al-Zubidi, M., Widziolek, M., Court, E. K., Gains, A. F., Smith, R. E., Ansbro, K., Alrafaie, A., Evans, C., Murdoch, C., Mesnage, S., Douglas, C. W. I., Rawlinson, A., & Stafford, G. P. (2019). Identification of novel bacteriophages with therapeutic potential that target *Enterococcus faecalis*. *Infection and Immunity*, *87*(11), e00512-19.

Allarakha, S. (2021, October 13). *How do you get* Enterococcus faecalis *infection?* MedicineNet. https://www.medicinenet.com/how_do_you_get_enterococcus_faecalis_infection/article.htm

Alrafaie, A. M., & Stafford, G. P. (2023). Enterococcal bacteriophage: A survey of the tail associated lysin landscape. *Virus Research, 327*, 199073. https://doi.org/10.1016/j.virusres.2023.199073

Ameh, E. M., Tyrrel, S., Harris, J. A., Pawlett, M., Orlova, E. V., Ignatiou, A., & Nocker, A. (2020) Lysis performance of bacteriophages with different plaque sizes and comparison of lysis kinetics after simultaneous and sequential phage addition. *PHAGE, 1*(3), 149–157. https://doi.org/10.1089/phage.2020.0005

Askora, A., El-Telbany, M., El-Didamony, G., Ariny, E., & Askoura, M. (2020). Characterization of φef-VB1 prophage infecting oral *Enterococcus faecalis* and enhancing bacterial biofilm formation. *Journal of Medical Microbiology, 69*(9), 1151–1168. https://doi.org/10.1099/jmm.0.001246

Baum, D. (2008) Reading a phylogenetic tree: The meaning of monophyletic groups. *Nature Education, 1*(1):190

Bhardwaj, S. B. (2019, December 16). *Enterococci: An important nosocomial pathogen*. IntechOpen. https://www.intechopen.com/chapters/70494

Bhardwaj, S. B., Mehta, M., & Sood, S. (2020). Enterococci in the oral cavity of periodontitis patients from different urban socioeconomic groups. *Dental Research Journal*, *17*(2), 147–151.

The Biology Notes Editors. (2022, May 2). *Bacteriophage- structure, classification, application*. The Biology Notes. https://thebiologynotes.com/bacteriophage-structure-classification-application/

Centers for Disease Control and Prevention. (2021). *2019 antibiotic resistance threats report*. Centers for Disease Control and Prevention. https://www.cdc.gov/antimicrobial-resistance/dataresearch/threats/?CDC_AAref_Val=https://www.cdc.gov/drugresistance/biggest-threats.html

Chatterjee, A., Willett, J. L. E., Dunny, G. M., & Duerkop, B. A. (2021). Phage infection and sub-lethal antibiotic exposure mediate *Enterococcus faecalis* type VII secretion system dependent inhibition of bystander bacteria. *PLoS genetics*, *17*(1), e1009204. https://doi.org/10.1371/journal.pgen.1009204

Christie, G. E. (1999). Propagation of viruses | Bacteria. In A. Granoff & R. G. Webster (Eds.), *Encyclopedia of virology* (2nd ed.) (pp. 1413-1418). Elsevier. https://doi.org/10.1006/rwvi.1999.0237

Clokie, M. R., Millard, A. D., Letarov, A. V., & Heaphy, S. (2011). Phages in nature. *Bacteriophage*, *1*(1), 31–45. https://doi.org/10.4161/bact.1.1.14942

Dedeo, C. L., Cingolani, G., & Teschke, C. M. (2019). Portal protein: The orchestrator of capsid assembly for the dsDNA tailed bacteriophages and herpesviruses. *Annual Review of Virology*, *6*(1), 141–160. https://doi.org/10.1146/annurev-virology-092818-015819

Del Rio, B., Sánchez-Llana, E., Martínez, N., Fernández, M., Ladero, V., & Alvarez, M. A. (2021). Isolation and characterization of *Enterococcus faecalis*-infecting bacteriophages from different cheese types. *Frontiers in Microbiology*, *11*, 592172. https://doi.org/10.3389/fmicb.2020.592172

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., & Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, *36*, W465-W469. https://doi.org/10.1093/nar/gkn180

Dincer, S., Musume, F., Delik, U., & Delik, A. (2020). *Antibiotic resistance in biofilm*. IntechOpen. https://www.intechopen.com/chapters/72109

Dion, M. B., Oechslin, F., & Moineau, S. (2020). Phage diversity, genomics and Phylogeny. *Nature Reviews Microbiology*, *18*(3), 125–138. https://doi.org/10.1038/s41579-019-0311-5

Esmail, M. A., Abdulghany, H. M., & Khairy, R. M. (2019). Prevalence of multidrug-resistant *Enterococcus faecalis* in hospital-acquired surgical wound infections and bacteremia: Concomitant analysis of antimicrobial resistance genes. *Infectious Diseases: Research and Treatment*, *12*, 117863371988292. https://doi.org/10.1177/1178633719882929

Fiore, E., Van Tyne, D., & Gilmore, M. S. (2019). Pathogenicity of enterococci. *Microbiology Spectrum*, *7*(4), 10.1128/microbiolspec.GPP3-0053-2018.

Fraser, S.L, (2022). *Enterococcal infections*. Medscape. https://emedicine.medscape.com/article/216993-overview?reg=1#a6

Gallet, R., Kannoly, S., & Wang, I. N. (2011). Effects of bacteriophage traits on plaque formation. *BMC Microbiology*, *11*, 181. https://doi.org/10.1186/1471-2180-11-181

Gilmore, M. S., Salamzade, R., Selleck, E., Bryan, N., Mello, S. S., Manson, A. L., & Earl, A. M. (2020). Genes contributing to the unique biology and intrinsic antibiotic resistance of *Enterococcus faecalis*. *mBio*, *11*(6), e02962-20. https://doi.org/10.1128/mBio.02962-20

Gonzalez, L. S., III., & Spencer, J. P. (1998). Aminoglycosides: A practical review. *American Family Physician*, *58*(8), 1811–1820.

Goodarzi, F., Hallajzadeh, M., Sholeh, M., Talebi, M., Pirhajati Mahabadi, V., & Amirmozafari, N. (2022). Anti-biofilm activity of a lytic phage against vancomycin-resistant *Enterococcus faecalis*. *Iranian Journal of Pathology*, *17*(3), 285–293. https://doi.org/10.30699/IJP.2022.541855.2760

Hartke, A., Giard, J. C., Laplace, J. M., & Auffray, Y. (1998). Survival of *Enterococcus faecalis* in an oligotrophic microcosm: Changes in morphology, development of general stress resistance, and analysis of protein synthesis. *Applied and Environmental Microbiology*, *64*(11), 4238–4245. https://doi.org/10.1128/AEM.64.11.4238-4245.1998

Hill, C. (2019). Evaluation of phage therapy in the context of *Enterococcus faecalis* and its associated diseases. *Viruses*, *11*(4), 366. https://doi.org/10.3390/v11040366

Hollenbeck, B. L., & Rice, L. B. (2012). Intrinsic and acquired resistance mechanisms in enterococcus. *Virulence*, *3*(5), 421–433. https://doi.org/10.4161/viru.21282

Howard Hughes Medical Institute. (2019). Phage therapy treats patient with drug-resistant bacterial infection. HHMI. https://www.hhmi.org/news/phage-therapy-treats-patient-with-drug-resistant-bacterial-infection

Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., & Sullivan, M. B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *The ISME Journal*, *11*(7), 1511–1520. https://doi.org/10.1038/ismej.2017.16

Kaplan J. B. (2010). Biofilm dispersal: Mechanisms, clinical implications, and potential therapeutic uses. *Journal of Dental Research*, *89*(3), 205–218. https://doi.org/10.1177/0022034509359403

Khalifa, L., Brosh, Y., Gelman, D., Coppenhagen-Glazer, S., Beyth, S., Poradosu-Cohen, R., Que, Y.-A., Beyth, N., & Hazan, R. (2015). Targeting *Enterococcus faecalis* biofilms with

phage therapy. *Applied and Environmental Microbiology*, *81*(8), 2696–2705. https://doi.org/10.1128/aem.00096-15

Khosravi, A., & Mazmanian, S. K. (2013). Disruption of the gut microbiome as a risk factor for microbial infections. *Current Opinion in Microbiology*, *16*(2), 221–227. https://doi.org/10.1016/j.mib.2013.03.009

Kristich, C. J., Li, Y. H., Cvitkovitch, D. G., & Dunny, G. M. (2004). Esp-independent biofilm formation by *Enterococcus faecalis*. *Journal of Bacteriology*, *186*(1), 154–163. https://doi.org/10.1128/JB.186.1.154-163.2004

Kristich, C. J., Rice, L. B., & Arias, C. A. (2014). Enterococcal infection—Treatment and antibiotic resistance. In M. S. Gilmore, D.B. Clewell, Y. Ike, & N. Shankar (Eds.), *Enterococci: From commensals to leading causes of drug resistant infection*. Massachusetts Eye and Ear Infirmary.

Kutter, E. (2009). Phage host range and efficiency of plating. *Methods in Molecular Biology*, 141–149. https://doi.org/10.1007/978-1-60327-164-6_14

Lee, D., Im, J., Na, H., Ryu, S., Yun, C.-H., & Han, S. H. (2019, November 28). The novel enterococcus phage vb_efas_hef13 has broad lytic activity against clinical isolates of *Enterococcus faecalis*. *Frontiers in Microbiology, 10*. https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2019.02877/full#B17

Levitus, M., Rewane, A., & Perera, T. B. (2024). Vancomycin-resistant enterococci. In *StatPearls* [Internet]. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK513233/

Li, M., Yang, F., Lu, Y., & Huang, W. (2020). Identification of *Enterococcus faecalis* in a patient with urinary-tract infection based on metagenomic next-generation sequencing: A case report. *BMC Infectious Diseases*, *20*(1). https://doi.org/10.1186/s12879-020-05179-0

Libretexts. (2023, October 31). *21.2b: The lytic and lysogenic cycles of bacteriophages*. Biology LibreTexts. https://bio.libretexts.org/Bookshelves/Introductory_and_General_Biology/General_Biology_(Boundless)/21%3A_Viruses/21.02%3A_Virus_Infections_and_Hosts/21.2B%3A_The_Lytic_and_Lysogenic_Cycles_of_Bacteriophages

Lin, H., Rao, V. B., & Black, L. W. (1999). Analysis of capsid portal protein and terminase functional domains: Interaction sites required for DNA packaging in bacteriophage T4. *Journal of Molecular Biology, 289*(2), 249- 260. https://doi.org/10.1006/jmbi.1999.2781

Loc-Carrillo, C., & Abedon, S. T. (2011). Pros and cons of phage therapy. *Bacteriophage*, *1*(2), 111–114. https://doi.org/10.4161/bact.1.2.14590

Lossouarn, J., Briet, A., Moncaut, E., Furlan, S., Bouteau, A., Son, O., Leroy, M., DuBow, M., Lecointe, F., Serror, P., & Petit, M.-A. (2019). *Enterococcus faecalis* countermeasures defeat a virulent Picovirinae bacteriophage. *Viruses, 11*(1), 48. https://doi.org/10.3390/v11010048

Mayo Foundation for Medical Education and Research. (2022, February 18). *Infectious diseases*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/infectious-diseases/symptoms-causes/syc-20351173

Mulani, M. S., Kamble, E. E., Kumkar, S. N., Tawre, M. S., & Pardesi, K. R. (2019). Emerging strategies to combat ESKAPE pathogens in the era of antimicrobial resistance: A review. *Frontiers in Microbiology, 10*, 539. https://doi.org/10.3389/fmicb.2019.00539

Nasr Azadani, D., Zhang, D., Hatherill, J. R., Silva, D., & Turner, J. W. (2020). Isolation, characterization, and comparative genomic analysis of a phage infecting high-level aminoglycoside-resistant (HLAR) *Enterococcus faecalis*. *PeerJ, 8*. https://doi.org/10.7717/peerj.9171

Núñez-Sánchez, M. A., Colom, J., Walsh, L., Buttimer, C., Bolocan, A. S., Pang, R., Gahan, C. G. M., & Hill, C. (2020). Characterizing phage-host interactions in a simplified human intestinal barrier model. *Microorganisms*, *8*(9), 1374. https://doi.org/10.3390/microorganisms8091374

Olsen, N. S., Nielsen, K. J., Plöger, M., Kot, W., & Hansen, L. H. (2021). Enterococcus phage Nonaheksakonda infecting clinical isolates of *Enterococcus faecalis* represents a new lineage in the family Siphoviridae. *Archives of Virology*, *166*(2), 593–599. https://doi.org/10.1007/s00705-020-04905-y

Phage 101. (n.d.). UC San Diego School of Medicine, Division of Infectious Diseases & Global Public Health, Center for Innovative Phage Applications and Therapeutics. https://sites.medschool.ucsd.edu/som/medicine/divisions/idgph/research/center-innovative-phage-applications-and-therapeutics/research/Pages/Phage%20101.aspx

Połaska, M., & Sokołowska, B. (2019). Bacteriophages-a new hope or a huge problem in the food industry. *AIMS Microbiology*, *5*(4), 324–346. https://doi.org/10.3934/microbiol.2019.4.324

Poxleitner, M., Pope, W., Jacobs-Sera, D., Sivanathan, V., & Hatfull, G. (n.d.). *Sea Phages Discovery Guide*. Howard Hughes Medical Institute. https://seaphagesphagediscoveryguide.helpdocsonline.com/home

Rogers, C. G., & Sarles, W. B. (1963). Characterization of an *Enterococcus faecalis* bacteriophages from the small intestine of the rat. *Journal of Bacteriology*, *85*(6), 1378–1385. https://doi.org/10.1128/jb.85.6.1378-1385.1963

Said, H. S., & Abdelmegeed, E. S. (2019). Emergence of multidrug resistance and extensive drug resistance among enterococcal clinical isolates in Egypt. *Infection and Drug Resistance, 12*, 1113–1125. https://doi.org/10.2147/idr.s189341

Said, M., Tirthani, E., & Leash, E. (2024). Enterococcus infections. In *StatPearls* [Internet]. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK567759/

Song, M., Wu, D., Hu, Y., Luo, H., & Li, G. (2021). Characterization of an *Enterococcus faecalis* bacteriophage vB_EfaM_LG1 and its synergistic effect with antibiotic. *Frontiers in Cellular and Infection Microbiology*, *11*, 698807. https://doi.org/10.3389/fcimb.2021.698807

Sørensen, P. E., Van Den Broeck, W., Kiil, K., Jasinskyte, D., Moodley, A., Garmyn, A., Ingmer, H., & Butaye, P. (2020). New insights into the biodiversity of coliphages in the intestine of poultry. *Scientific Reports*, *10*(1), 15220. https://doi.org/10.1038/s41598-020-72177-2

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Strateva, T., Atanasova, D., Savov, E., Petrova, G., & Mitov, I. (2016). Incidence of virulence determinants in clinical *Enterococcus faecalis* and *Enterococcus faecium* isolates collected in Bulgaria. *The Brazilian Journal of Infectious Diseases*, *20*(2), 127–133. https://doi.org/10.1016/j.bjid.2015.11.011

Sun, S., Gao, S., Kondabagil, K., Xiang, Y., Rossmann, M. G., & Rao, V. B. (2012). Structure and function of the small terminase component of the DNA packaging machine in T4-like bacteriophages. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(3), 817–822. https://doi.org/10.1073/pnas.1110224109

Tkachev, P. V., Pchelin, I. M., Azarov, D. V., Gorshkov, A. N., Shamova, O. V., Dmitriev, A. V., & Goncharov, A. E. (2022). Two novel lytic bacteriophages infecting Enterococcus spp. are promising candidates for targeted antibacterial therapy. *Viruses, 14*(4), 831. https://doi.org/10.3390/v14040831

United States Food and Drug Administration. (2023). *Sterilization for medical devices*. U.S. Food and Drug Administration. https://www.fda.gov/medical-devices/general-hospital-devices-and-supplies/sterilization-medical-devices

Walsh, F. (2014, December 11). *Superbugs to kill 'more than cancer' by 2050*. BBC News. https://www.bbc.com/news/health-30416844

Watson, S. (2018, September 29). Enterococcus faecalis*: Causes, symptoms, and treatments*. Healthline. https://www.healthline.com/health/enterococcus-faecalis

WebMD (2023). *Enterococcal infections: What they are, how they spread, and how they're treated*. WebMD. https://www.webmd.com/a-to-z-guides/what-are-enterococcal-infections

Welsh, B., Abdalla, N. M., Aldana, E., Alvarado Fernandez, V. M., Arenales Salgado de Oliveira, B., Fakhre, D., Haymond, A. J., Helton, K. M., Kanchibhatta, A., Knight, J., Marshall, S., Martinez, M. L. N., Merkher, A., Morrow, S. E., Nguyen, K. P., Patel, J. J., Patel, S. R., Rayala, P., Ruiz-Houston, K. M., ... Pollenz, R. S. (2022). Genome sequence of CaiB, a DR cluster actinobacteriophage that infects *Gordonia rubripertincta*. *Microbiology Resource Announcements, 11*(7), e0037622. https://doi.org/10.1128/mra.00376-22

White, H. E., & Orlova, E. V. (2019). Bacteriophages: Their structural organization and function. In R. Savva (Ed.), *Bacteriophages - perspectives and future*. DOI: 10.5772/intechopen.85484.

World Health Organization. (2023). *Antibiotic resistance*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance

Wright, A., Hawkins, C. H., Anggård, E. E., & Harper, D. R. (2009). A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant Pseudomonas aeruginosa; a preliminary report of efficacy. *Clinical Otolaryngology, 34*(4), 349–357. https://doi.org/10.1111/j.1749-4486.2009.01973.x

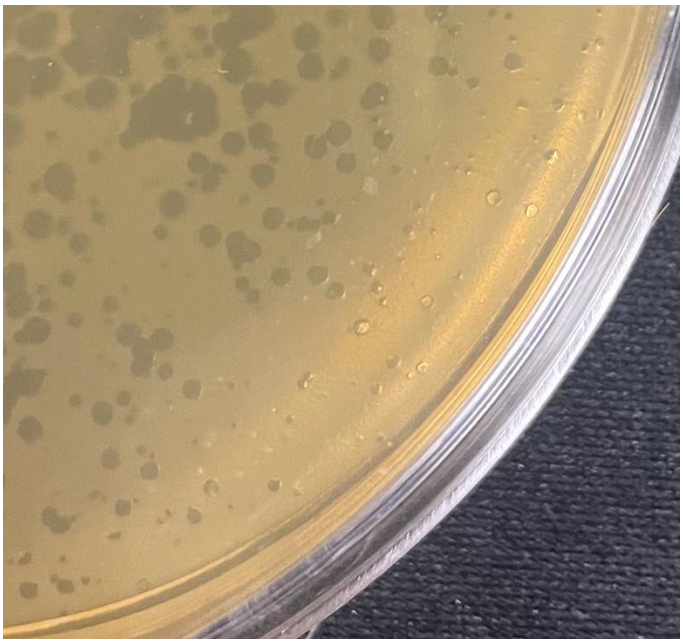
Wu, H., Moser, C., Wang, H.-Z., Høiby, N., & Song, Z.-J. (2014). Strategies for combating bacterial biofilm infections. *International Journal of Oral Science, 7*(1), 1–7. https://doi.org/10.1038/ijos.2014.65

Yin, W., Wang, Y., Liu, L., & He, J. (2019). Biofilms: The microbial "protective clothing" in extreme environments. *International Journal of Molecular Sciences, 20*(14), 3423. https://doi.org/10.3390/ijms20143423

Zhang, W., Mi, Z., Yin, X., Fan, H., An, X., Zhang, Z., Chen, J., & Tong, Y. (2013). Characterization of *Enterococcus faecalis* phage IME-EF1 and its endolysin. *PLoS One, 8*(11), e80435. https://doi.org/10.1371/journal.pone.0080435

Zou, J., & Shankar, N. (2016). Surface protein Esp enhances pro-inflammatory cytokine expression through NF-κB activation during enterococcal infection. *Innate Immunity, 22*(1), 31-39. doi:10.1177/1753425915611237

VII. Appendix



Supplemental Figure 1: Lytic phage VPE25 used as positive control. Phage VPE25 has lytic plaques of approximately 0.44± 0.27 mm in diameter double layer BHI-$MgSO_4$ agar.

Supplemental Table 1: Gblocks aligned sequence used for creation of phylogenetic portal protein tree. Selected alignment blocks (underlined in blue)

Processed file: **input.fasta**
Number of sequences: **68**
Alignment assumed to be: **Protein**
New number of positions: **16** (selected positions are underlined in blue)

```
                    10        20        30        40        50        60
                ========+=========+=========+=========+=========+=========+
vB_EfaH_EF1TV   ----------------------------MESIEKG-------------MNGKTTAYMQ
PBEF129         MPKWLDKALGIEKSSIEETRNMENYKMHLREIDTNVVNNEPYSMESIEKGMNGKTTAYMQ
phiEF24C        MPKWLDKALGIEKSSIEETRNMENYKMHLREIDTNVVNNEPYSMESIEKGMNGKTTAYMQ
EfsSzw-1        -MPKWLDKVLGVEEDADIVKSEEVYQTLLKEIQAQP------LTETVEKGMKGKPVAYMQ
vB_EfaM_A2      -MPKWLDKVLGVEEDTDIVKSEEVYQTLLKEIQAQP------LTETVEKGMNGKPVAYMQ
PEf771          -MPKWLDKVLGVEEDTDIVKSEEVYQTLLKEIQAQP------LTETVEKGMNGKPVAYMQ
EF-P10          ----------------------------M-----------------------------
vB_EfaS_EF1c55  ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNYDIR
vB_EfaS_IME198  ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNYDIR
BC-611          ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNFDLR
VD13            ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNYDIR
Entf1           ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNYDIR
vB_EfaS_TV16    ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNYDIR
EFC1            ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNYDIR
IME-EF1         ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNYDIR
EF-P29          ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNYDIR
vB_EfaS_PHB08   ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNYDIR
SAP6            ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNYDIR
vB_EfaS_HEf13   ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNYDIR
EfsWh-1         ----------------------------MAIPNGQINAGDIITTNIRRKHFIRRNYDIR
phiFL3B         ----------------------------MSIVVNRQIAGDLNKPSAELLNFC-INKHVE
phiFL3A         ----------------------------MSIVVNRQIAGDLNKPSAELLNFC-INKHVE
vB_EfaS_IME197  METEEILKVNEFEHG-------------SDISYSSDMNENFVSFGVESNIHY-RYSSAE
EFC-1           MENEEVLKVNEFEHG-------------SDISYSSDVNENFVSFGVESNIHY-RYSSAD
phiEf11         MNNKLLNGSRFDKEANLV-----------YKVPVSKLPTRIMQYSNGEKEEV--VDFEHQ
phiFL2B         ---------------------------MAQEDFES-------------------LEFDSE
phiFL2A         ----------------------MNKKNEVTFLRNHRYHKNANAVFRMAQEDFESLEFDSE
phiFL1A         ----------------------MNKKNEVTFLRNHRYHKDANAVFRMSQEDFDDIDFEKK
phiFL1C         ------------------------------------------------------MLNCKK
phiFL1B         ------------------------------------------------------MLNCKK
phiFL4A         ----------------------------MEALLSE-----------------DVKIIA
AE4_2           ----------------------------MEALLSE-----------------DVKIIA
vB_EfaS-DELF1   ----------------------------MN-----------------------------
EFA1            ----------------------------MGVFVAP-----------------RSTTKQ
vB_EfaS_IME196  ----------------------------MGVFVAP-----------------RSTTKQ
vB_EFaS_TV217   ----------------------------MGVFVAP-----------------RSTTKQ
LY0323          ----------------------------MGVFVAP-----------------RSTTKQ
EFRM31          ----------------------------MGVFVAP-----------------RSTTKQ
vB_EfaS_AL3     ----------------------------MGVFVAP-----------------RSTTKQ
```

```
vB_EfaS_AL2      ----------------------------MGVFVAP-----------------RSTTKQ
FX417            ----------------------------MGVFVAP-----------------RSTTKQ
LY0322           ----------------------------MGVFVAP-----------------RSTTKQ
phiNASRA1        ----------------------------MGVFVAP-----------------RSTTKQ
vB_EfaS_Ef5.3    ----------------------------MGVFVAP-----------------ISTTKQ
vB_EfaS_LM99     ----------------------------MGVFVAP-----------------RSTTKQ
SANTOR1          ----------------------------MGVFVAP-----------------RSTTKQ
vB_EfaS-271      ----------------------------MGVFVAP-----------------RSTTKQ
vB_EfaS_785CC    ----------------------------MGVFVAP-----------------RSTTKQ
vB_EfaS_785CS    ----------------------------MGVFVAP-----------------RSTTKQ
EfaCPT1          ----------------------------MGVFVAP-----------------RSTTKQ
vB_EfaS_Ef5.1    ----------------------------MGVFVAP-----------------RSTTKQ
vB_EfaS_Ef5.4    ----------------------------MGVFVAP-----------------RSTTKQ
PMBT2            ----------------------------MGVFVAP-----------------RSTTKQ
phiSHEF5         ----------------------------MGVFVAP-----------------RSTTKQ
vB_EfaS_Max      ----------------------------MGVFVAP-----------------RSTTKQ
vB_EFaS_TV54     ----------------------------MGVFVAP-----------------RSTTKQ
phiSHEF2         ----------------------------MGVFVAP-----------------RSTTKQ
vB_EfaS_Ef5.2    ----------------------------MGVFVAP-----------------RSTTKQ
vB_EFaS_TV51     ----------------------------MGVFVAP-----------------RSTTKQ
EFAP-1           ----------------------------MGVFVAP-----------------RSTTKQ
Sigurd           ----------------------------MGVFVAP-----------------RSTTKQ
vB_EfaS_Ef6.1    ----------------------------MGVFVAP-----------------RSTTKQ
phiSHEF4         ----------------------------MGVFVAP-----------------RSTTKQ
IME_EF3          ----------------------------MGVFVAP-----------------RSTTKQ
vB_EfaS_Ef6.4    ----------------------------MGVFIAP-----------------QTSAEK
MSF2             ----------------------------MGVFIAP-----------------QTKAEK
Nonaheksakonda   ----------------------------MGVFIAP-----------------KSMTSD
heks             ----------------------------MGVFIAP-----------------KSMTSD



                          70        80        90       100       110       120
                 =========+=========+=========+=========+=========+=========+
vB_EfaH_EF1TV    PII---GEMSVNPGYKTKPSIRNSQDLHKTLKKFGNNIILNAIINTRSNQVSMYCKPARN
PBEF129          PII---GEMSVNPGYKTKPSIRNSQDLHKTLKKFGNNIILNAIINTRSNQVSMYCKPARN
phiEF24C         PII---GEMSVNPGYKTKPSIRNSQDLHKTLKKFGNNIILNAIINTRSNQVSMYCKPARN
EfsSzw-1         PII---GDMSVNPGFKTKPSIRNSQDLHKMLKKFGNNIILNSIINTRSNQVSMYCKPARN
vB_EfaM_A2       PII---GDMSVNPGFKTKPSIRNSQDLHKMLKKFGNNIILNSIINTRSNQVSMYCKPARN
PEf771           PII---GDMSVNPGFKTKPSIRNSQDLHKMLKKFGNNIILNSIINTRSNQVSMYCKPARN
EF-P10           ------------------------------------------------------------
vB_EfaS_EF1c55   ELV---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRTFDD---------ANKPN
vB_EfaS_IME198   ELI---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRTFDD---------TNKPN
BC-611           ELI---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRTFDD---------TNKPN
VD13             ELI---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRTFDD---------TNKPN
Entf1            ELI---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRTFDD---------TNKPN
vB_EfaS_TV16     ELI---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRTFDD---------TNKPN
EFC1             ELI---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRAFDD---------TNKPN
IME-EF1          ELI---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRTFDD---------TNKPN
EF-P29           ELI---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRTFDD---------TNKPN
vB_EfaS_PHB08    ELI---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRTFDD---------TNKPN
SAP6             ELI---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRTFDD---------TNKPN
vB_EfaS_HEf13    ELI---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRTFDD---------TNKPN
EfsWh-1          ELI---TLAEMHSRSSS------AYGVLYDYYKGNHIAIQSRTFDD---------TNKPN
phiFL3B          ELL---RLQKLSDYYDG-----------------KHDILKRTKEN---------EGAPN
phiFL3A          ELL---RLQKLSDYYDG-----------------KHDILKRTKEN---------EGAPN
vB_EfaS_IME197   ELLGNTDILAKMIAHHNEHQVPRLQILDDYYKAKNTNIIKNRRRKE---------KEKAD
```

```
EFC-1              ELLENVDVLAKMIAHHNEHQVPRLQVLDDYYKAKNTNIIKNRRRKE---------KEKAD
phiEf11            DVF---NMIVKFVRHHKEKQVPRLKELKRYSLAQNNI----KFTEDKS-------ENRAD
phiFL2B            TWI---TQLKRFVNRHLAEQVPRLKELKRYYLADNNIKYRPPKTDE----------YAAD
phiFL2A            TWI---TQLKRFVNRHLAEQVPRLKELKRYYLADNNIKYRPPKTDE----------YAAD
phiFL1A            SWI---TRLEKFVNRHIVEQVPRLRELKRYYLADNNIKYRPPKTDE----------YAAD
phiFL1C            KID---EFNKQNN-------------------------------------------EAYHN
phiFL1B            KID---EFNKQNN-------------------------------------------EAYHN
phiFL4A            NAL---KAAIDKDRKSSSKREAETGIRYYNHENDIMNNRIFYVDDE---------GILRE
AE4_2              NAL---KAAIDKDRKSSSKREAETGIRYYNHENDIMNNRIFYVDDE---------GILRE
vB_EfaS-DELF1      ------------------------------------------------------------
EFA1               NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_IME196     NFI---DFIESSDIYST-----------------------DITSE---------KALRN
vB_EFaS_TV217      NFI---DFIESSDIYST-----------------------EITGE---------KALRN
LY0323             NFI---DFIESSDIYST-----------------------EITGE---------KALRN
EFRM31             NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_AL3        NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_AL2        NFI---DFIESSDIYST-----------------------EITGE---------KALRN
FX417              NFI---DFIESSDIYST-----------------------EITGE---------KALRN
LY0322             NFI---DFIESSDIYST-----------------------EITGE---------KALRN
phiNASRA1          NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_Ef5.3      NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_LM99       NFI---DFIESSDIYST-----------------------EITGE---------KALRN
SANTOR1            NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS-271        NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_785CC      NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_785CS      NFI---DFIESSDIYST-----------------------EITGE---------KALRN
EfaCPT1            NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_Ef5.1      NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_Ef5.4      NFI---DFIESSDIYST-----------------------EITGE---------KALRN
PMBT2              NFI---DFIESSDIYST-----------------------EITGE---------KALRN
phiSHEF5           NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_Max        NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EFaS_TV54       NFI---DFIESSDIYST-----------------------EITGE---------KALRN
phiSHEF2           NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_Ef5.2      NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EFaS_TV51       NFI---DFIESSDIYST-----------------------EITGE---------KALRN
EFAP-1             NFI---DFIESSDIYST-----------------------EITGE---------KALRN
Sigurd             NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_Ef6.1      NFI---DFIESSDIYST-----------------------EITGE---------KALRN
phiSHEF4           NFI---DFIESSDIYST-----------------------EITGE---------KALRN
IME_EF3            NFI---DFIESSDIYST-----------------------EITGE---------KALRN
vB_EfaS_Ef6.4      NFI---DYIESGMDNLG-----------------------YLTGE---------RALQN
MSF2               NFI---DYIESGMDNLG-----------------------YLTGE---------RALQN
Nonaheksakonda     RFI---DYINNGMMDIP-----------------------PLTGL---------DALKN
heks               RFI---DYINNGMMDIP-----------------------PLTGL---------DALKN



                          130       140       150       160       170       180
                   =========+=========+=========+=========+=========+=========+
vB_EfaH_EF1TV      SETGVGYEIRLKD----------------IEAEPTSHDIANIKRIESFLENTAQFR----
PBEF129            SETGVGYEIRLKD----------------IEAEPTSHDIANIKRIESFLENTAQFR----
phiEF24C           SETGVGYEIRLKD----------------IEAEPTSHDIANIKRIESFLENTAQFR----
EfsSzw-1           SETGVGYEVRLKD----------------IEQEPSTHDIANIKRIENFLENTAVFK----
vB_EfaM_A2         SETGVGYEVRLKD----------------IEQEPSTHDIANIKRIESFLENTAVFK----
PEf771             SETGVGYEVRLKD----------------IEQEPSTHDIANIKRIENFLENTAVFK----
EF-P10             ------------------------------------------------------------
```

```
vB_EfaS_EF1c55   SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
vB_EfaS_IME198   SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
BC-611           SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
VD13             SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
Entf1            SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
vB_EfaS_TV16     SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
EFC1             SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
IME-EF1          SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
EF-P29           SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
vB_EfaS_PHB08    SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
SAP6             SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
vB_EfaS_HEf13    SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
EfsWh-1          SKIVHNFPKLLVD------TSTAY-----LAGEPI-----TESGDEKTIKAMQ-------
phiFL3B          NKVLINHAKYVVD------MNVGF-----MVGNPISYVAESDKNIGPILEAYD-------
phiFL3A          NKVLINHAKYVVD----------MNVGFMVGNPISYVAESDKNIGPILEAYD-------
vB_EfaS_IME197   HRAAHNFGKVLAT--------------FDVGYNTGNPLKVQINNKASQEAID-------
EFC-1            HRAAHNFGKVLAT----------F-----DVGYNTGNPIKVQINNKASQEAID-------
phiEf11          NKIANDWARFIVN----------FKKGVLLGNPL-----KYNGDKTIADKIN-------
phiFL2B          NRIASDFARYITI----------FEQGYMLGQPVQYKNENDELQVKIDDFNK-------
phiFL2A          NRIASDFARYITI------FEQGY-----MLGQPVQYKNENDELQVKIDDFNK-------
phiFL1A          NRIASDFARYITI----------FEQGYMLGQPVQYKNENVELQKKIDEFNK-------
phiFL1C          VLIKTDLSIYG----------------------RAYDALQEEQAFVKLVK-------
phiFL1B          VLIKTDLSI--------------------YGRAYELETIALQEEQAFVKLVK-------
phiFL4A          DKYASNVRIPHGFFPEIVDQKTQY-----LLSNPVEYETENEELKEYLAEYYNSEFQVVL
AE4_2            DKYASNVRIPHGFFPEIVDQKTQY-----LLSNPVEYETENEELKEYLAEYYNSEFQVVL
vB_EfaS-DELF1    ------------------------------------NSAVDESLLTLLNKE-----
EFA1             SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
vB_EfaS_IME196   SDIFTGINIISGD----------------LGQSKFKPTKNSEVDEQFLKMINKR-----
vB_EFaS_TV217    SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
LY0323           SDIFTGINIISGD----------------LGQSKFRPTKNSEVDEQFLKMINKR-----
EFRM31           SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
vB_EfaS_AL3      SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
vB_EfaS_AL2      SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
FX417            SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
LY0322           SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
phiNASRA1        SDIFTGINIISGD----------------LGQSKFRPTKNSEVDEQFLKMINKR-----
vB_EfaS_Ef5.3    SDIFTGINIISGD----------------LGQSSFRPTKNSEVDEQFLKMINKR-----
vB_EfaS_LM99     SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
SANTOR1          SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
vB_EfaS-271      SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
vB_EfaS_785CC    SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
vB_EfaS_785CS    SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
EfaCPT1          SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
vB_EfaS_Ef5.1    SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
vB_EfaS_Ef5.4    SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
PMBT2            SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
phiSHEF5         SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
vB_EfaS_Max      SDIFTGINIISGD----------------LGQSKFRPTKNSEADEQFLKMINKR-----
vB_EFaS_TV54     SDIFTGINIISGD----------------LGQSKFRPTKNSEVDEQFLKMINKR-----
phiSHEF2         SDIFTGINIISGD----------------LGQSKFRPTKNSEVDEQFLKMINKR-----
vB_EfaS_Ef5.2    SDIFTGINIISGD----------------LGQSKFRPTKNSEVDEQFLKMINKR-----
vB_EFaS_TV51     SDIFTGINIISGD----------------LGQSKFRPTKNSEVDEQFLKMINKR-----
EFAP-1           SDIFTGINIISGD----------------LGQSSFRPVKDIPVDEEFLKMINKR-----
Sigurd           SDIFTGINIISGD----------------LGQSSFRAVKDIPVDEEFLKMINKR-----
vB_EfaS_Ef6.1    SDIFTGINIISGD----------------LGQSSFRAVKDIPVDEEFLKMINKR-----
phiSHEF4         SDIFTGINIISGD----------------LGQSSFRAVKDIPVDEEFLKMINKR-----
IME_EF3          SDIFTGINIISGD----------------LGQSSFRAVKDIPVDEEFLKMINKR-----
```

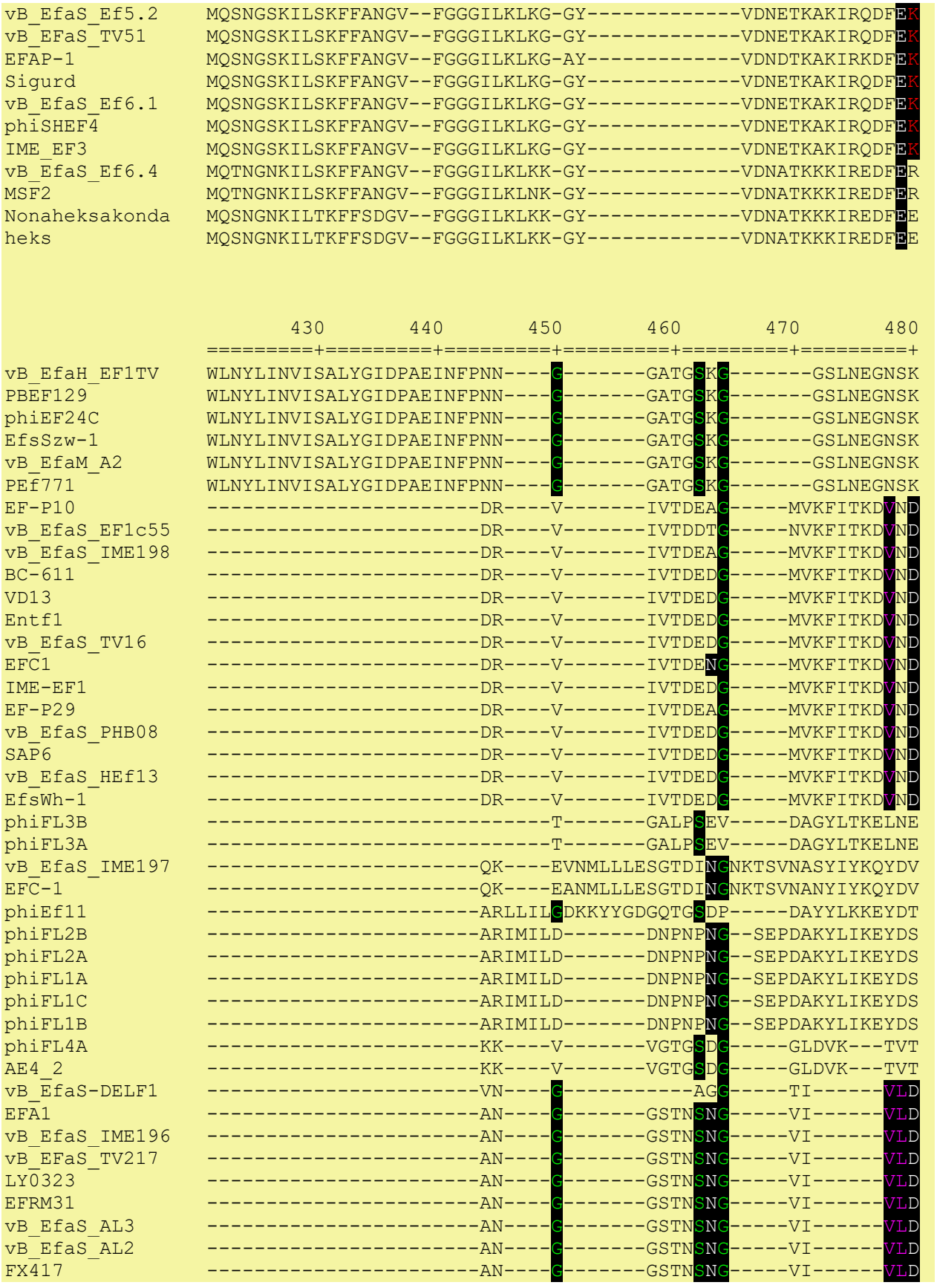
```
vB_EfaS_Ef6.4    SDVFTGVNIIGGD----------------IGKSLFRKVPDEYADKDFLQLINKR-----
MSF2             SDVFTGVNIIGGD----------------IGKSLFRKVPDEHADQSFLQLINKR-----
Nonaheksakonda   SDVYTGVNIIAGD----------------IGKTMFRNVPDEPSDKNLITLLNKR-----
heks             SDVYTGVNIIAGD----------------IGKTMFRNVPDEPSDKDLIKLLNKR-----



                        190       200       210       220       230       240
                 =========+=========+=========+=========+=========+=========+
vB_EfaH_EF1TV    ----DPNRDNFTTFCKK-----------LVRATYMYDQVNFEKV-----FDKDGN----F
PBEF129          ----DPNRDNFTTFCKK-----------LVRATYMYDQVNFEKV-----FDKDGN----F
phiEF24C         ----DPNRDNFTTFCKK-----------LVRATYMYDQVNFEKV-----FDKDGN----F
EfsSzw-1         ----DQNRDNFTAFCKK-----------MVRATYMYDQVNFEKV-----FDKDGN----F
vB_EfaM_A2       ----DQNRDNFTAFCKK-----------MVRATYMYDQVNFEKV-----FDKDGN----F
PEf771           ----DQNRDNFTAFCKK-----------MVRATYMYDQVNFEKV-----FDKDGN----F
EF-P10           ------------------------------------------------------------
vB_EfaS_EF1c55   ----PVMKENYVTDVNS----------EEIKLSGIFG--HCFEI-HWIDRNNKHR-----
vB_EfaS_IME198   ----PVFKENYVTDVNS----------EEVKLSGIFG--HCFEI-HWIDRNKKHR-----
BC-611           ----PVFKENYVTDVNS----------EEVKLSGIFG--HCFEI-HWIDRNKKHR-----
VD13             ----PVFKENYVTDVNS----------EEVKLSGIFG--HCFEI-HWIDRNKKHR-----
Entf1            ----PVFKENYVTDVNS----------EEVKLSGIFG--HCFEI-HWIDRNKKHR-----
vB_EfaS_TV16     ----PVFKENYVTDVNS----------EEVKLSGIFG--HCFEI-HWIDRNKKHR-----
EFC1             ----PVFKENYVTDVNS----------EEVKLSGIFG--HCFEI-HWIDRNKKHR-----
IME-EF1          ----PVFKENYVTDVNS----------EEVKLSGIFG--HCFEI-HWIDRNKKHR-----
EF-P29           ----PVFKENYVTDVNS----------EEVKLSGIFG--HCFEI-HWIDRNKKHR-----
vB_EfaS_PHB08    ----PVFKENYVTDVNS----------EEVKLSGIFG--HCFEI-HWIDRNKKHR-----
SAP6             ----PVFKENYVTDVNS----------EEVKLSGIFG--HCFEI-HWIDRNKKHR-----
vB_EfaS_HEf13    ----PVFKENYVTDVNS----------EEVKLSGIFG--HCFEI-HWIDRNKKHR-----
EfsWh-1          ----PVFKENYVTDVNS----------EEVKLSGIFG--HCFEI-HWIDRNKKHR-----
phiFL3B          ----RVDIVSHDTELEK----------DLSTFG--VGYEL---VYLKKTKDNDATQ
phiFL3A          ----RVDIVSHDTELEK----------DLSTFG--VGYEL---VYLKKTKDNDATQ
vB_EfaS_IME197   ----NFNQDNDIDGLNG----------ELWLDMDKYG--RAYEI---IYRDEDDV-----
EFC-1            ----NFNQDNDIDGLNG----------ELWLDMDKYG--RAYEI---IYRDEDDV-----
phiEf11          ----DFSSKSNEDYHNQ----------LMLDDLLVYG--RAFEY---IGRDEYGK-----
phiFL2B          ----QNNEAYHNVLIKT----------DLSIYG--RAYELETIALQDEQAF-----
phiFL2A          ----QNNEAYHNVLIKT----------DLSIYG--RAYELETIALQDEQAF----V
phiFL1A          ----QNNEAYHNVLIKT----------DLSIYG--RAYEL-ETIALQEEQA----F
phiFL1C          ----LNPEQTFIVYDDT----------TDSNSLFG--VYYYS---------------I
phiFL1B          ----LNPEQTFIVYDDT----------TDSNSLFG-VYYYSI-----DYGDGV-----
phiFL4A          QELVEGSSQKGFEYVYARTNAEDRLCFQVADSLNVFGVYNEYNE---LQRICRHY----I
AE4_2            QELVEGSSQKGFEYVYARTNAEDRLCFQVADSLNVFGVYNEYNE---LQRICRHY----I
vB_EfaS-DELF1    ----PKTNQSHYTMMYA----------VVSNLILTG--NAYIL---IHRNDDNS----V
EFA1             ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDRMGT----I
vB_EfaS_IME196   ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EFaS_TV217    ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDRMGG----I
LY0323           ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
EFRM31           ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EfaS_AL3      ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EfaS_AL2      ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
FX417            ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
LY0322           ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
phiNASRA1        ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EfaS_Ef5.3    ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EfaS_LM99     ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
SANTOR1          ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EfaS-271      ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EfaS_785CC    ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
```

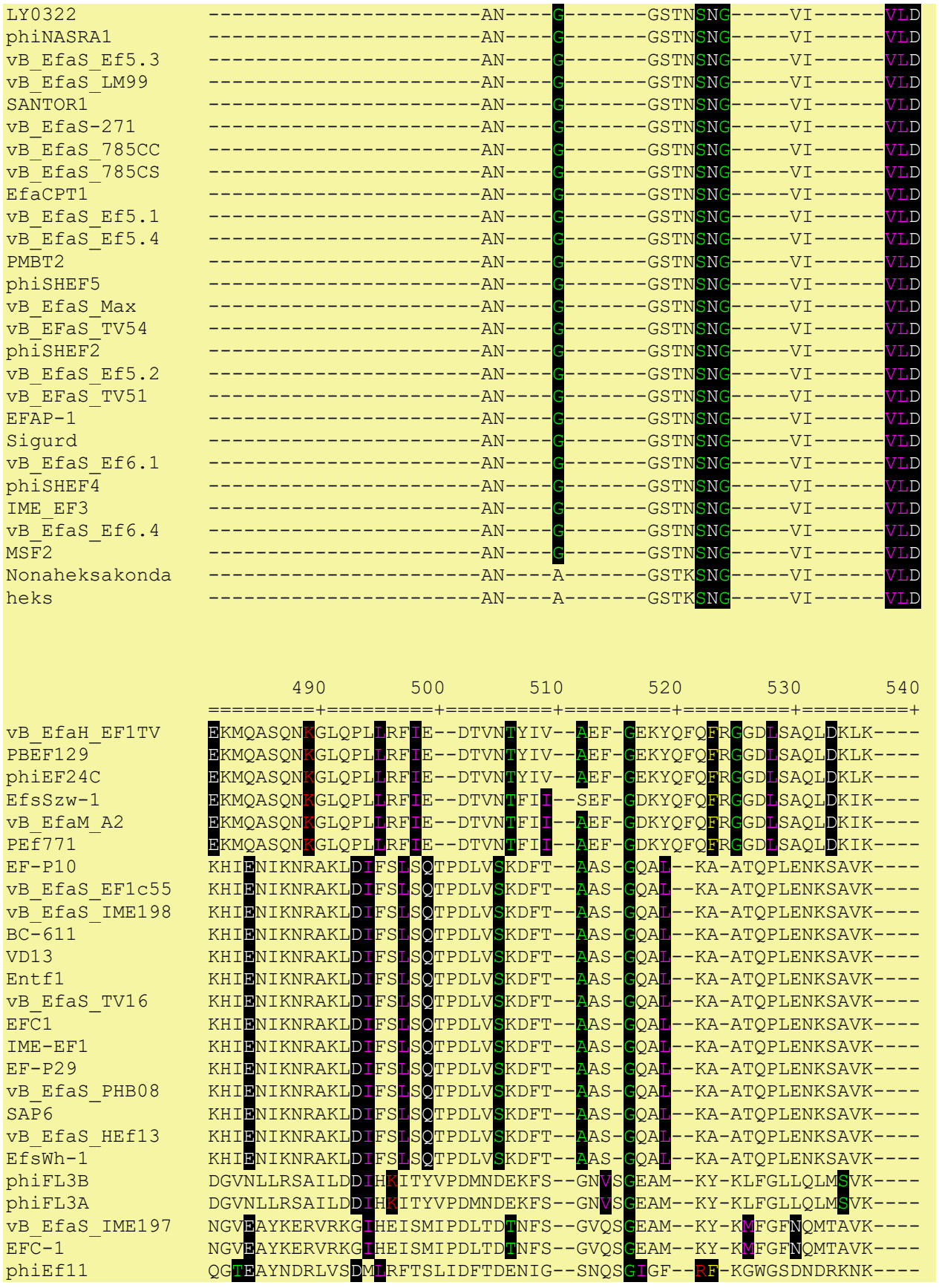
```
vB_EfaS_785CS    ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
EfaCPT1          ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EfaS_Ef5.1    ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EfaS_Ef5.4    ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
PMBT2            ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
phiSHEF5         ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EfaS_Max      ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EFaS_TV54     ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
phiSHEF2         ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EfaS_Ef5.2    ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
vB_EFaS_TV51     ----PNNKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDKLGV----V
EFAP-1           ----PNDKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDRMGG----I
Sigurd           ----PNEKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDRMGS----I
vB_EfaS_Ef6.1    ----PNEKQSHYMFMYA----------TVANLILSG--NSYAL---IHRDRMGS----I
phiSHEF4         ----PNDKQSHYMFMYA----------TVANLILSG--NSYAL---IHRNRMGD----I
IME_EF3          ----PNDKQSHYMFMYA----------TVANLILSG--NSYAL---IHRNRMGD----I
vB_EfaS_Ef6.4    ----PHEKQSHYTFMYA----------TVAQLILYG--NSYAI---IHREKENDYNSPI
MSF2             ----PHEKQSHYTFMYA----------TVAQLILYG--NSYAI---IHREKENDYNSPI
Nonaheksakonda   ----PHKKQSHYTFLYA----------TVAQLILYG--NAYAV---IHRETRNDYTSPI
heks             ----PHKKQSHYTFLYA----------TVAQLILYG--NAYAV---IHRETRNDYTSPI



                        250       260       270       280       290       300
                 =========+=========+=========+=========+=========+=========+
vB_EfaH_EF1TV    IKFDTVDPTTIFLATNGE---------------GKLIKNGERFVQVIDNRIVAKFNERE-
PBEF129          IKFDTVDPTTIFLATNGD---------------GKLIKNGERFVQVIDNRIVAKFNERE-
phiEF24C         IKFDTVDPTTIFLATNGE---------------GKLIKNGERFVQVIDNRIVAKFNERE-
EfsSzw-1         IKFDTVDPTTIFLATNGK---------------GKIIEKGERFVQVLDNRIVAKFNERE-
vB_EfaM_A2       IKFDTVDPTTIFLATNGK---------------GKIIEKGERFVQVLDNRIVAKFNERE-
PEf771           IKFDTVDPTTIFLATNGK---------------GKIIEKGERFVQVLDNRIVAKFNERE-
EF-P10           ------------------------------------------------------------
vB_EfaS_EF1c55   --FKAVSPMNCLIAYSADLDEEPLAAIYYNTVISDITGHVIRTYEIYTEDKILKFSTDDE
vB_EfaS_IME198   --FKAVSPMNCLIAYSADLDEEPIAAIYYNTVISDITGHQIRTYEIYTEDLIYKFSTDDE
BC-611           --FKAVSPMNCLIAYSADLDEEPVAAIYYNTVISDITGHQIRTYEVYTEDLIYKFSTDDE
VD13             --FKAVSPMNCLIAYSADLDEEPIAAIYYNTVISDITGHQIRTYEVYTEDLIYKFSTDDE
Entf1            --FKAVSPMNCLIAYSADLDEEPIAAIYYNTVISDITGHQIRTYEVYTEDLIYKFSTDDE
vB_EfaS_TV16     --FKAVSPMNCLIAYSADLDEEPIAAIYYNTVISDITGHQIRTYEIYTEDLIYKFSTDDE
EFC1             --FKAVSPMNCLIAYSADLDEEPIAAIYYNTVISDITGHQIRTYEVYTEDLIYKFSTDDE
IME-EF1          --FKAVSPMNCLIAYSADLDEEPIAAIYYNTVISDITGHQIRTYEVYTEDLIYKFSTDDE
EF-P29           --FKAVSPMNCLIAYSADLDEEPIAAIYYNTVISDITGHQIRTYEVYTEDLIYKFSTDDE
vB_EfaS_PHB08    --FKAVSPMNCLIAYSADLDEEPIAAIYYNTVISDITGHQIRTYEVYTEDLIYKFSTDDE
SAP6             --FKAVSPMNCLIAYSADLDEEPIAAIYYNTVISDITGHQIRTYEVYTEDLIYKFSTDDE
vB_EfaS_HEf13    --FKAVSPMNCLIAYSADLDEEPIAAIYYNTVISDITGHQIRTYEVYTEDLIYKFSTDDE
EfsWh-1          --FKAVSPMNCLIAYSADLDEEPIAAIYYNTVISDITGHQIRTYEVYTEDLIYKFSTDDE
phiFL3B          LEIKCIDPRGIFLVTDDTVDKNPLFAVHYQPVLTLQGG--------IDHYVIKYYNDNRV
phiFL3A          LEIKCIDPRGIFLVTDDTVDKNPLFAVHYQPVLTLQGG--------IDHYVIKYYNDNRV
vB_EfaS_IME197   DYVDLCNVFETFVVYDTTVKRRPILAVRYPKTKFTIDAD-------KQYIQPIIYTKDKI
EFC-1            DYVDLCNVFETFVVYDTTVKRRPILAVRYPKTKFTIDAD-------KQYIQPIIYTKDKI
phiEf11          EMLAKFSAEETFVIYDTT------------------TNK-NSVCAIHCYDLEFNDETFSY
phiFL2B          VKLVKLNPEQTFIVYDDTTDSNSLFGVYYYSIDYG-DEVSKDFINVYTSDMLYIYVN---
phiFL2A          KLVKL-NPEQTFIVYDDTTDSNSLFGVYYYSIDYG-DEVSKDFINVYTSDMLYIYVN---
phiFL1A          VKLVKLNPEQTFIVYDDTTDSNSLFGVYYYSIDYG-DGVRKDFINVYTSDMLYIYVN---
phiFL1C          DYGDGVRKDFINV-------------------------------YTSDMLYIYVN---
phiFL1B          ------RKDFINV-------------------------------YTSDMLYIYVN---
phiFL4A          TEIEK-DGETVDIHHAEV--------------------------WTDQNVYFFVAEDN
AE4_2            TEIEK-DGETVDIHHAEV--------------------------WTDQNVYFFVAEDN
```

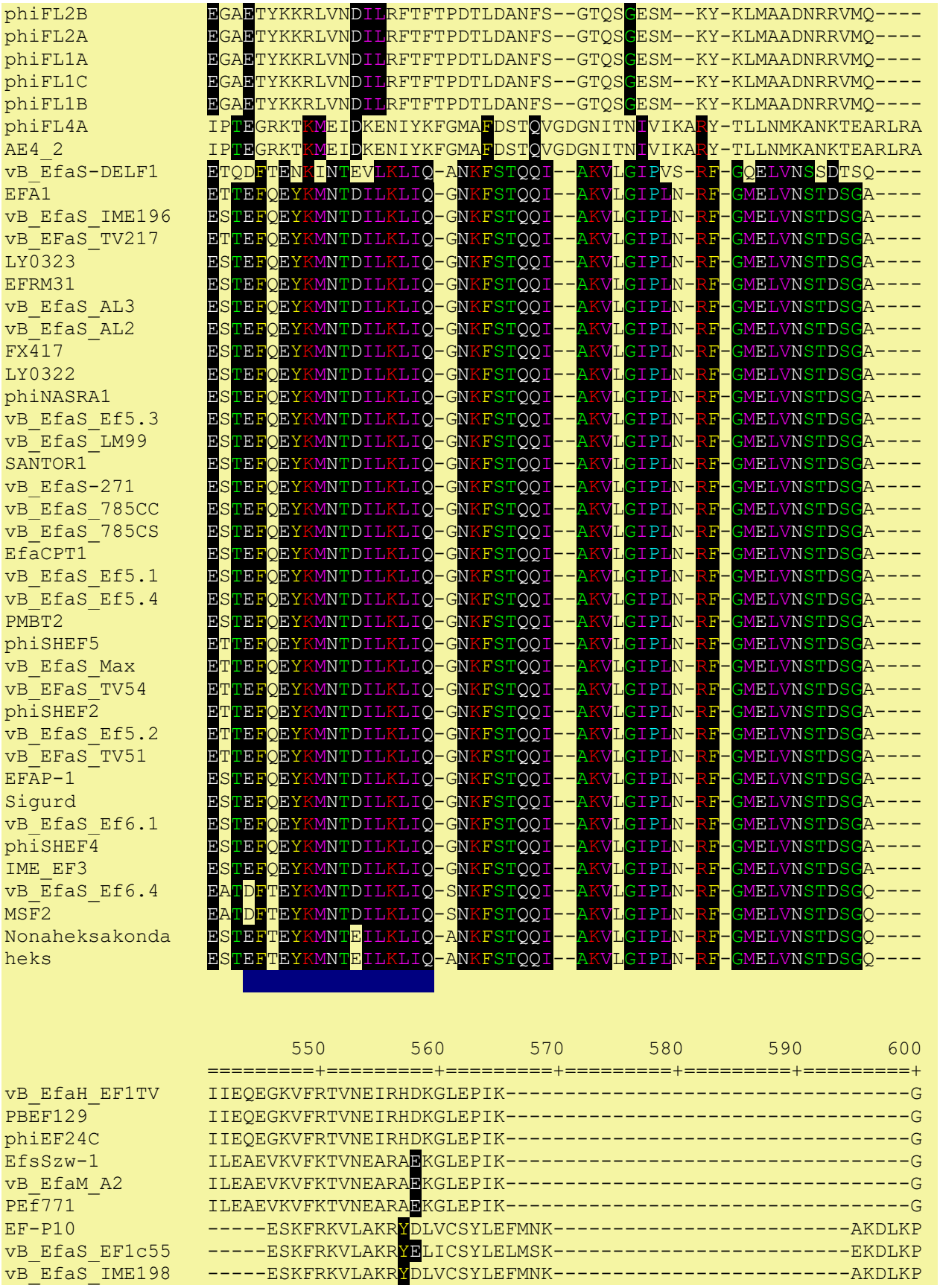
```
vB_EfaS-DELF1    KELEFVETQQVNVIRDLV-----------------TGE-------YRYEVIMPYGN---
EFA1             ESLEFVRPEQVNVIQNVT-----------------TGE-------WSYDVTMDYGS---
vB_EfaS_IME196   ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMDYGS---
vB_EFaS_TV217    ESLEFVKPEQVNVIQNVT-----------------TGE-------WTYDVTMDYGS---
LY0323           ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMEYGS---
EFRM31           ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMEYGS---
vB_EfaS_AL3      ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMEYGS---
vB_EfaS_AL2      ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMEYGS---
FX417            ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMEYGS---
LY0322           ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMEYGS---
phiNASRA1        ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMDYGS---
vB_EfaS_Ef5.3    ESLEFVKPEQVNVIQNVT-----------------TGE-------WTYDVTMDYGS---
vB_EfaS_LM99     ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMDYGS---
SANTOR1          ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMDYGS---
vB_EfaS-271      ESLEFVKPEQVNVIQNVT-----------------TGE-------WTYDVTMDYGS---
vB_EfaS_785CC    ESLEFVKPEQVNVIQNVT-----------------TGE-------WTYDVTMDYGS---
vB_EfaS_785CS    ESLEFVKPEQVNVIQNVT-----------------TGE-------WTYDVTMDYGS---
EfaCPT1          ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMDYGS---
vB_EfaS_Ef5.1    ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMDYGS---
vB_EfaS_Ef5.4    ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMDYGS---
PMBT2            ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMDYGS---
phiSHEF5         ESLEFVKPEQVNVIQNVT-----------------TGE-------WSYDVTMDYGS---
vB_EfaS_Max      ESLEFVKPEQVNVIQNVT-----------------TGE-------WTYDVTMDYGS---
vB_EFaS_TV54     ESLEFVKPEQVNVIQNVT-----------------TGE-------WNYDVTMDYGS---
phiSHEF2         ESLEFVKPEQVNVIQNVT-----------------TGE-------WTYDVTMDYGS---
vB_EfaS_Ef5.2    ESLEFVKPEQVNVIQNVT-----------------TGE-------WTYDVTMDYGS---
vB_EFaS_TV51     ESLEFVKPEQVNVIQNVT-----------------TGE-------WTYDVTMDYGS---
EFAP-1           ESLEFVRPEQVNVIQNVT-----------------TGE-------WSYDVTMDYGS---
Sigurd           ESLEFVRPEQVNVIQNVN-----------------TGE-------WSYDVTMDYGS---
vB_EfaS_Ef6.1    ESLEFVRPEQVNVIQNVN-----------------TGE-------WSYDVTMDYGS---
phiSHEF4         ESLEFVRPEQVNVIQNVN-----------------TGE-------WSYDVTMDYGS---
IME_EF3          DSLEFVRPEQVNVIQNVN-----------------TGE-------WSYDVTMDYGS---
vB_EfaS_Ef6.4    KSLEFVTADQVNLIQDMT-----------------TGE-------WRYDVTLDYGN---
MSF2             KSLEFVTADQVNLIQDMT-----------------TGE-------WRYDVTLDYGN---
Nonaheksakonda   KSLEFVRADQVNVIQDMT-----------------TGE-------WRYDVTLDYGS---
heks             TSLEFVRADQVNVIQDMT-----------------TGE-------WRYDVTLDYGS---



                        310       320       330       340       350       360
                 =========+=========+=========+=========+=========+=========+
vB_EfaH_EF1TV    ------LAFAVRNPRADIEVGQYG-YPELEIALKQFIAHENTEVFNDRFFSHGGTTRGIL
PBEF129          ------LAFAVRNPRADIEVGQYG-YPELEIALKQFIAHENTEVFNDRFFSHGGTTRGIL
phiEF24C         ------LAFAVRNPRADIEVGQYG-YPELEIALKQFIAHENTEVFNDRFFSHGGTTRGIL
EfsSzw-1         ------MAFAVRNPRADIEVGQYG-YPELEIALKQFIAHENTETFNDRFFSHGGTTRGIL
vB_EfaM_A2       ------MAFAVRNPRADIEVGQYG-YPELEIALKQFIAHENTETFNDRFFSHGGTTRGIL
PEf771           ------MAFAVRNPRADIEVGQYG-YPELEIALKQFIAHENTETFNDRFFSHGGTTRGIL
EF-P10           ------------------------------------------------------------
vB_EfaS_EF1c55   R-----DVYKEIPE----VLDIRG-YEEHPNLLKKF----PVLEIIANEERLGDFEA-QL
vB_EfaS_IME198   R-----EVYKEIPE----ELEIKD-YEVHPNLLQKF----PVLEIIANEERLGDFEA-QL
BC-611           R-----EVYREIPE----ELEIKD-YEVHPNLLQKF----PVLEIIANEERLGDFEA-QL
VD13             R-----EVYREIPE----ELEIKD-YEVHPNLLQKF----PVLEIIANEERLGDFEA-QL
Entf1            R-----EVYREIPE----ELEIKD-YEVHPNLLQKF----PVLEIIANEERLGDFEA-QL
vB_EfaS_TV16     R-----EVYKEIPE----ELEIKD-YEVHPNLLQKF----PVLEIIANEERLGDFEA-QL
EFC1             R-----EVYKEIPE----ELEIKD-YEVHPNLLQKF----PVLEIIANEERLGDFEA-QL
IME-EF1          R-----EVYKEIPE----ELEIKD-YEVHPNLLQKF----PVLEIIANEERLGDFEA-QL
EF-P29           R-----EVYKEIPE----ELEIKD-YEVHPNLLQKF----PVLEIIANEERLGDFEA-QL
```

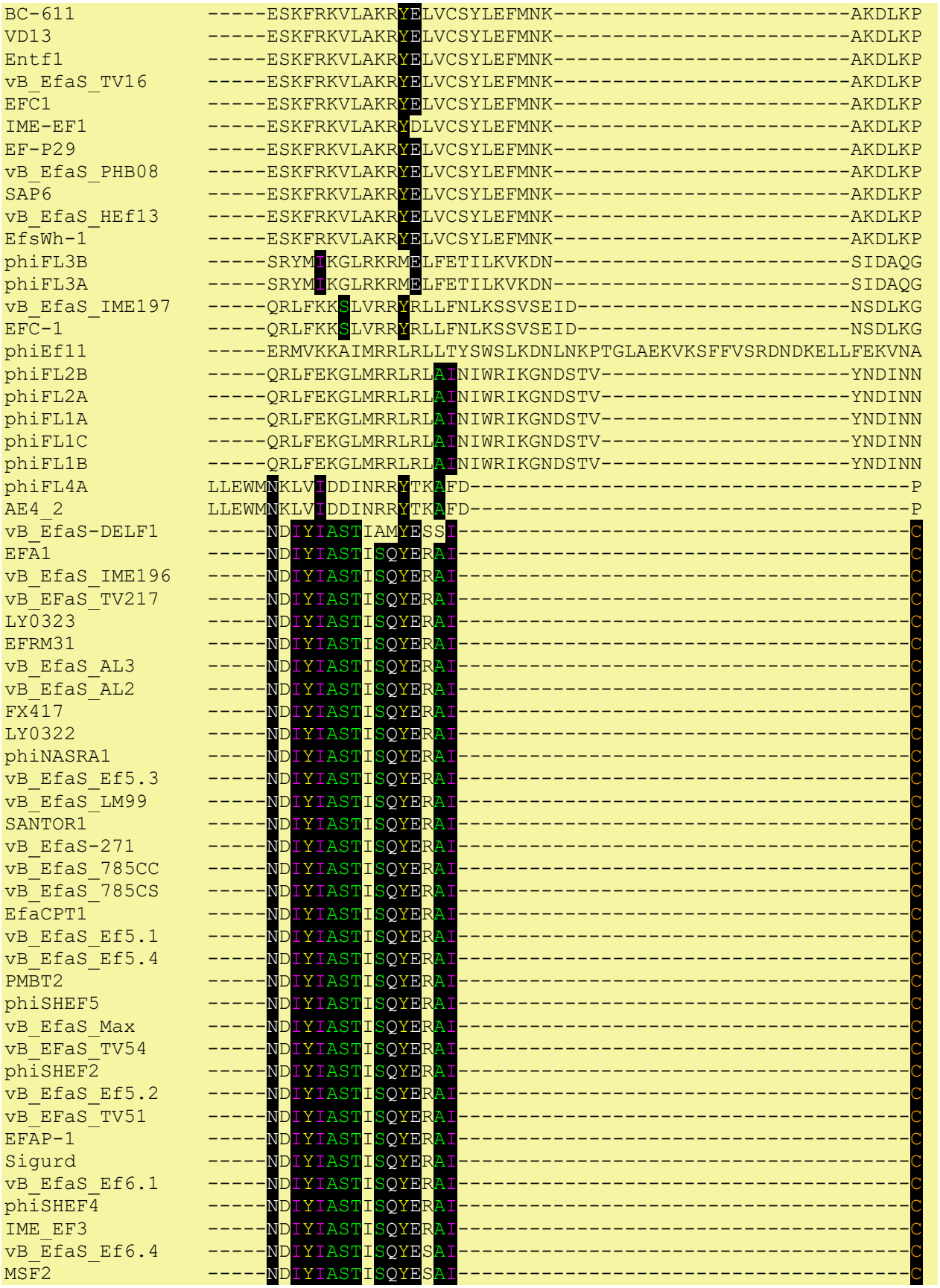
```
vB_EfaS_PHB08    R-----EVYKEIPE----ELEIKD-YEVHPNLLQKF----PVLEIIANEERLGDFEA-QL
SAP6             R-----EVYKEIPE----ELEIKD-YEVHPNLLQKF----PVLEIIANEERLGDFEA-QL
vB_EfaS_HEf13    R-----EVYKEIPE----ELEIKD-YEVHPNLLQKF----PVLEIIANEERLGDFEA-QL
EfsWh-1          R-----EVYKEIPE----ELEIKD-YEVHPNLLQKF----PVLEIIANEERLGDFEA-QL
phiFL3B          ------ITYHAKSRG---SGEYLF-INAKPHYFKAV----PVIEYRNNEEKQGDFEQ-AI
phiFL3A          ------ITYHAKSRG---SGEYLF-INAKPHYFKAV----PVIEYRNNEEKQGDFEQ-AI
vB_EfaS_IME197   ------ITYAETTL----ATIKLS-DPKEDSHDYKEV---QITEFSPNRFRMGLYED-IL
EFC-1            ------ITYAETTL----ATIKLS-DPKEDSHDYKEV---QITEFSPNRFRMGLYED-IL
phiEf11          IDIYANDGYFYQHESKNQDYEQSKLIDKYQTFFDSI----QVNEWINNEERLGDFET-VL
phiFL2B          ------DNQNKKGL----TLLDFE-DYAFNGV--------PINEFANNEDRTGAYEP-VL
phiFL2A          ------DNQNKKGL----TLLDFE-DYAFNGV--------PINEFANNEDRTGAYEP-VL
phiFL1A          ------DNQNEKGL----TLLDFE-DYAFDGV--------PINEFANNEDRTGAYEP-VL
phiFL1C          ------DNQNEKGL----TLLDFE-DYAFDGV--------PINEFANNEDRTGAYEP-VL
phiFL1B          ------DNQNEKGL----TLLDFE-DYAFDGV--------PINEFANNEDRTGAYEP-VL
phiFL4A          KDYELDEAEPINPR----PHVLAV-DSENESLLQRSYGQIPFYRLSNNKQETTDLKP-IK
AE4_2            KDYELDEAEPINPR----PHVLAV-DSENESLLQRSYGQIPFYRLSNNKQETTDLKP-IK
vB_EfaS-DELF1    ------IMYKCTPK----DILHFK-LTTTDGWLGRS----PLLSLNDE----------IS
EFA1             ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_IME196   ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EFaS_TV217    ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
LY0323           ------IMYKCTPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
EFRM31           ------IMYKCTPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_AL3      ------IMYKCTPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_AL2      ------IMYKCTPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
FX417            ------IMYKCTPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
LY0322           ------IMYKCTPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
phiNASRA1        ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_Ef5.3    ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_LM99     ------IMYKCKPD----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
SANTOR1          ------IMYKCKPD----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS-271      ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_785CC    ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_785CS    ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
EfaCPT1          ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_Ef5.1    ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_Ef5.4    ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
PMBT2            ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
phiSHEF5         ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_Max      ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EFaS_TV54     ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
phiSHEF2         ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_Ef5.2    ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EFaS_TV51     ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
EFAP-1           ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
Sigurd           ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_Ef6.1    ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
phiSHEF4         ------IMYKCKPE----DILHFR-ITTIDGFLGRS----PLLSLKDE----------VA
IME_EF3          ------IMYKCKPE----DILHFR-ITTVDGFLGRS----PLLSLKDE----------VA
vB_EfaS_Ef6.4    ------HMLRCEPR----DILHFR-ISVVDGFIGRS----PLLSLRDE----------IG
MSF2             ------HMLKCEPR----DILHFR-ISVIDGFIGRS----PLLSLRDE----------IG
Nonaheksakonda   ------KMYHCNEA----DILHFR-ISSVNGFLGRS----PLLSLNDE----------IA
heks             ------KMYHCNES----DILHFR-ISSVDGFLGRS----PLLSLNDE----------IA



                        370       380       390       400       410       420
                 =========+=========+=========+=========+=========+=========+
```

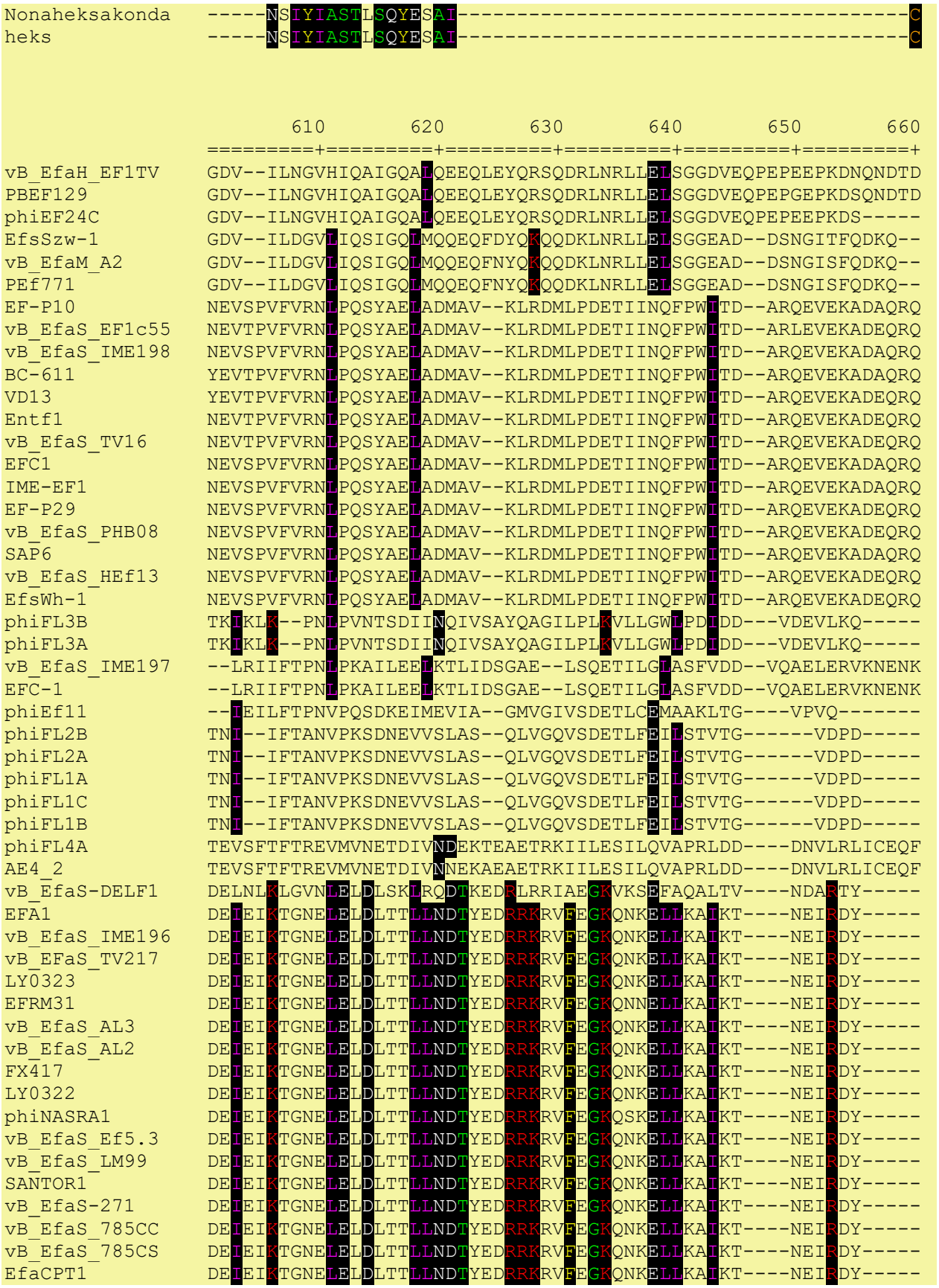
```
vB_EfaH_EF1TV    HVKTGQQQSQQALDIFRREWRSSLAGING-SWQIPVVSAEDVK--FVNMTPSANDMQFEK
PBEF129          HVKTGQQQSQQALDIFRREWRSSLAGING-SWQIPVVSAEDVK--FVNMTPSANDMQFEK
phiEF24C         HVKTGQQQSQQALDIFRREWRSSLAGING-SWQIPVVSAEDVK--FVNMTPSANDMQFEK
EfsSzw-1         HVKAGQQQSQQALDIFRREWRSSLSGING-SWQIPVVSAEDVK--FVNMTPSANDMQFEK
vB_EfaM_A2       HVKAGQQQSQQALDIFRREWRSSLSGING-SWQIPVVSAEDVK--FVNMTPSANDMQFEK
PEf771           HVKAGQQQSQQALDIFRREWRSSLSGING-SWQIPVVSAEDVK--FVNMTPSANDMQFEK
EF-P10           ------------------------------------------------------------KN
vB_EfaS_EF1c55   SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADSDSISNMKN
vB_EfaS_IME198   SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADSNSISNMKN
BC-611           SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADSDSISNMKN
VD13             SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADSDSISNMKN
Entf1            SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADSDSISNMKN
vB_EfaS_TV16     SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADSDSISNMKN
EFC1             SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADSDSISNMKN
IME-EF1          SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADSDSISNMKN
EF-P29           SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADSKSISNMKN
vB_EfaS_PHB08    SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADGDSISNMKN
SAP6             SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADSDSISNMKN
vB_EfaS_HEf13    SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADSDSISNMKN
EfsWh-1          SLIDAYNLAVSDSVNDIAYWNDAYLWLQG---------------FDLSADSDSISNMKN
phiFL3B          SLIDAYNLLQSDRLNDKEAFVDAILFIRGFML------------EDGDGEKLAKEKMLQ
phiFL3A          SLIDAYNLLQSDRLNDKEAFVDAILFIRGFML------------EDGDGEKLAKEKMLQ
vB_EfaS_IME197   SQIDLYDAGQSDTANYMTDLNDALLVISG----------------DIEAAGLSTEDAIK
EFC-1            SQIDLYDAGQSDTANYMTDLNDALLVISG----------------DIEAAGLSTEDAIK
phiEf11          DNIDAYDLSQSSMANFQQDSSEAYLVIKGNPETAIGD--------EEGNSAVDVLNDMIK
phiFL2B          DSIDAYDLSQSELANYQQDTMDAILLIKGNPYTGTAQNDLDEDGNIVPNSRLAVSLAFKK
phiFL2A          DSIDAYDLSQSELANYQQDTMDAILLIKGNPYTGTAQNDLDEDGNIVPNSRLAVSLAFKK
phiFL1A          DSIDAYDLSQSELANYQQDTMDAILLIKGNPYTGTAQNDLDEDGNIVPNSRLAVSLAFKK
phiFL1C          DSIDAYDLSQSELANYQQDTMDAILLIKGNPYTGTAQNDLDEDGNIVPNSRLAVSLAFKK
phiFL1B          DSIDAYDLSQSELANYQQDTMDAILLIKGNPYTGTAQNDLDEDGNIVPNSRLAVSLAFKK
phiFL4A          ALIDDYDLMNCFLSNNLQDFAEAIYVVSG----------------FQGDDLSKLRQNVKS
AE4_2            ALIDDYDLMNCFLSNNLQDFAEAIYVVSG----------------FQGDDLSKLRQNVKS
vB_EfaS-DELF1    LQTNGLKVLNNFFSKGV--FSGGILKLLN-GT------------VNNKTKAKIREDFEK
EFA1             MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
vB_EfaS_IME196   MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
vB_EFaS_TV217    MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
LY0323           MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
EFRM31           MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
vB_EfaS_AL3      MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
vB_EfaS_AL2      MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
FX417            MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
LY0322           MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
phiNASRA1        MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
vB_EfaS_Ef5.3    MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
vB_EfaS_LM99     MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
SANTOR1          MQSNGSKILSKFFANGV--FGGGILKLKG-GL------------VDNETKAKIRQDFEK
vB_EfaS-271      MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
vB_EfaS_785CC    MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
vB_EfaS_785CS    MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
EfaCPT1          MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
vB_EfaS_Ef5.1    MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
vB_EfaS_Ef5.4    MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
PMBT2            MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
phiSHEF5         MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
vB_EfaS_Max      MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
vB_EFaS_TV54     MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
phiSHEF2         MQSNGSKILSKFFANGV--FGGGILKLKG-GY------------VDNETKAKIRQDFEK
```

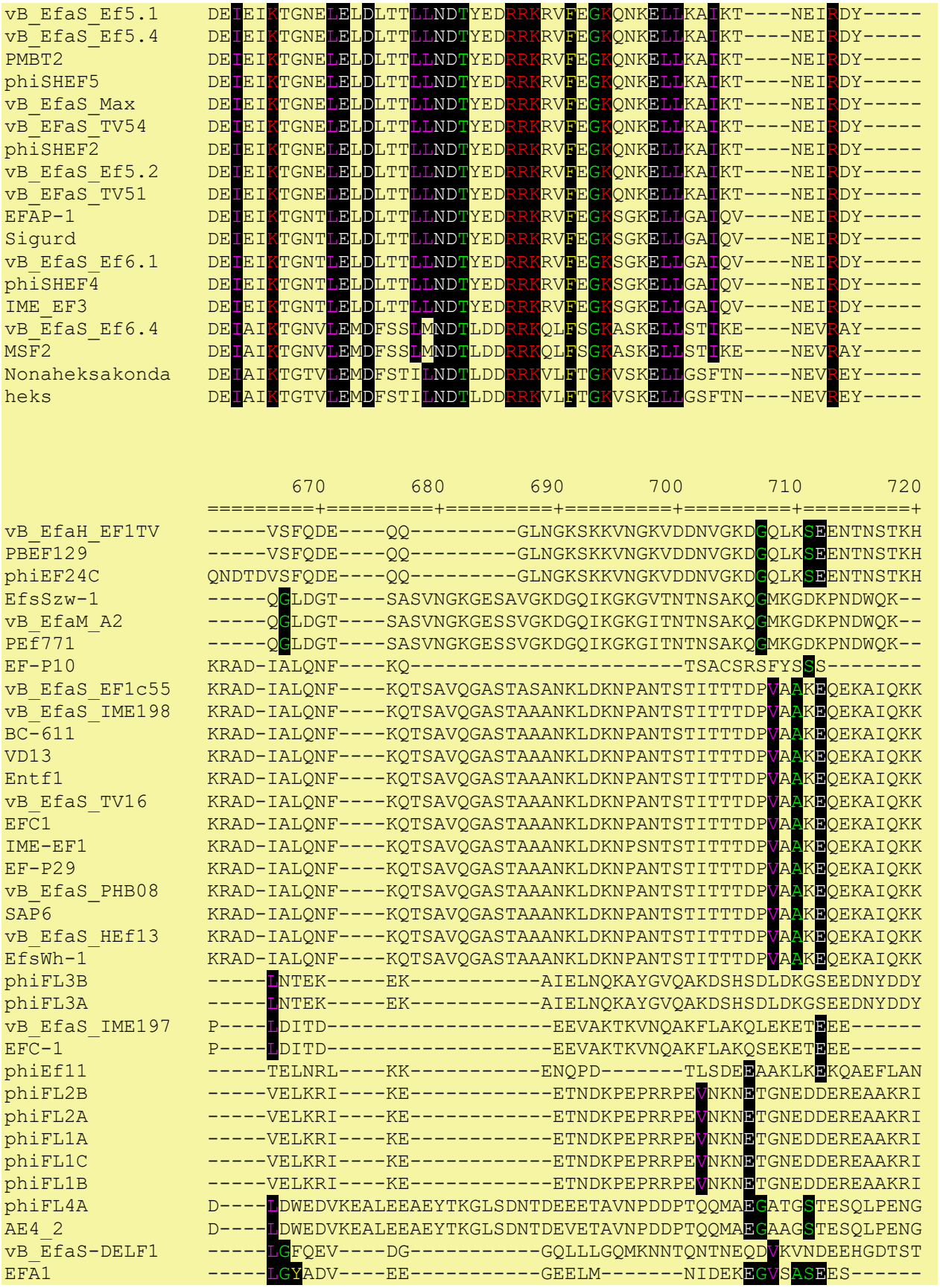

vB_EfaS_Ef5.2 MQSNGSKILSKFFANGV--FGGGILKLKG-GY----------VDNETKAKIRQDFEK
vB_EFaS_TV51 MQSNGSKILSKFFANGV--FGGGILKLKG-GY----------VDNETKAKIRQDFEK
EFAP-1 MQSNGSKILSKFFANGV--FGGGILKLKG-AY----------VDNDTKAKIRKDFEK
Sigurd MQSNGSKILSKFFANGV--FGGGILKLKG-GY----------VDNETKAKIRQDFEK
vB_EfaS_Ef6.1 MQSNGSKILSKFFANGV--FGGGILKLKG-GY----------VDNETKAKIRQDFEK
phiSHEF4 MQSNGSKILSKFFANGV--FGGGILKLKG-GY----------VDNETKAKIRQDFEK
IME_EF3 MQSNGSKILSKFFANGV--FGGGILKLKG-GY----------VDNETKAKIRQDFEK
vB_EfaS_Ef6.4 MQTNGNKILSKFFANGV--FGGGILKLKK-GY----------VDNATKKKIREDFER
MSF2 MQTNGNKILSKFFANGV--FGGGILKLNK-GY----------VDNATKKKIREDFER
Nonaheksakonda MQSNGNKILTKFFSDGV--FGGGILKLKK-GY----------VDNATKKKIREDFEE
heks MQSNGNKILTKFFSDGV--FGGGILKLKK-GY----------VDNATKKKIREDFEE
430 440 450 460 470 480
vB_EfaH_EF1TV WLNYLINVISALYGIDPAEINFPNN----G-------GATGSKG-------GSLNEGNSK
PBEF129 WLNYLINVISALYGIDPAEINFPNN----G-------GATGSKG-------GSLNEGNSK
phiEF24C WLNYLINVISALYGIDPAEINFPNN----G-------GATGSKG-------GSLNEGNSK
EfsSzw-1 WLNYLINVISALYGIDPAEINFPNN----G-------GATGSKG-------GSLNEGNSK
vB_EfaM_A2 WLNYLINVISALYGIDPAEINFPNN----G-------GATGSKG-------GSLNEGNSK
PEf771 WLNYLINVISALYGIDPAEINFPNN----G-------GATGSKG-------GSLNEGNSK
EF-P10 ----------DR----V-------IVTDEAG-----MVKFITKDVND
vB_EfaS_EF1c55 ----------DR----V-------IVTDDTG-----NVKFITKDVND
vB_EfaS_IME198 ----------DR----V-------IVTDEAG-----MVKFITKDVND
BC-611 ----------DR----V-------IVTDEDG-----MVKFITKDVND
VD13 ----------DR----V-------IVTDEDG-----MVKFITKDVND
Entf1 ----------DR----V-------IVTDEDG-----MVKFITKDVND
vB_EfaS_TV16 ----------DR----V-------IVTDEDG-----MVKFITKDVND
EFC1 ----------DR----V-------IVTDENG-----MVKFITKDVND
IME-EF1 ----------DR----V-------IVTDEDG-----MVKFITKDVND
EF-P29 ----------DR----V-------IVTDEAG-----MVKFITKDVND
vB_EfaS_PHB08 ----------DR----V-------IVTDEDG-----MVKFITKDVND
SAP6 ----------DR----V-------IVTDEDG-----MVKFITKDVND
vB_EfaS_HEf13 ----------DR----V-------IVTDEDG-----MVKFITKDVND
EfsWh-1 ----------DR----V-------IVTDEDG-----MVKFITKDVND
phiFL3B ----------T-------GALPSEV-----DAGYLTKELNE
phiFL3A ----------T-------GALPSEV-----DAGYLTKELNE
vB_EfaS_IME197 ----------QK----EVNMLLLESGTDINGNKTSVNASYIYKQYDV
EFC-1 ----------QK----EANMLLLESGTDINGNKTSVNANYIYKQYDV
phiEf11 ----------ARLLILGDKKYYGDGQTGSDP-----DAYYLKKEYDT
phiFL2B ----------ARIMILD-------DNPNPNG--SEPDAKYLIKEYDS
phiFL2A ----------ARIMILD-------DNPNPNG--SEPDAKYLIKEYDS
phiFL1A ----------ARIMILD-------DNPNPNG--SEPDAKYLIKEYDS
phiFL1C ----------ARIMILD-------DNPNPNG--SEPDAKYLIKEYDS
phiFL1B ----------ARIMILD-------DNPNPNG--SEPDAKYLIKEYDS
phiFL4A ----------KK----V-------VGTGSDG-----GLDVK---TVT
AE4_2 ----------KK----V-------VGTGSDG-----GLDVK---TVT
vB_EfaS-DELF1 ----------VN----G----------AGG-----TI------VLD
EFA1 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_IME196 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EFaS_TV217 ----------AN----G-------GSTNSNG-----VI------VLD
LY0323 ----------AN----G-------GSTNSNG-----VI------VLD
EFRM31 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_AL3 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_AL2 ----------AN----G-------GSTNSNG-----VI------VLD
FX417 ----------AN----G-------GSTNSNG-----VI------VLD

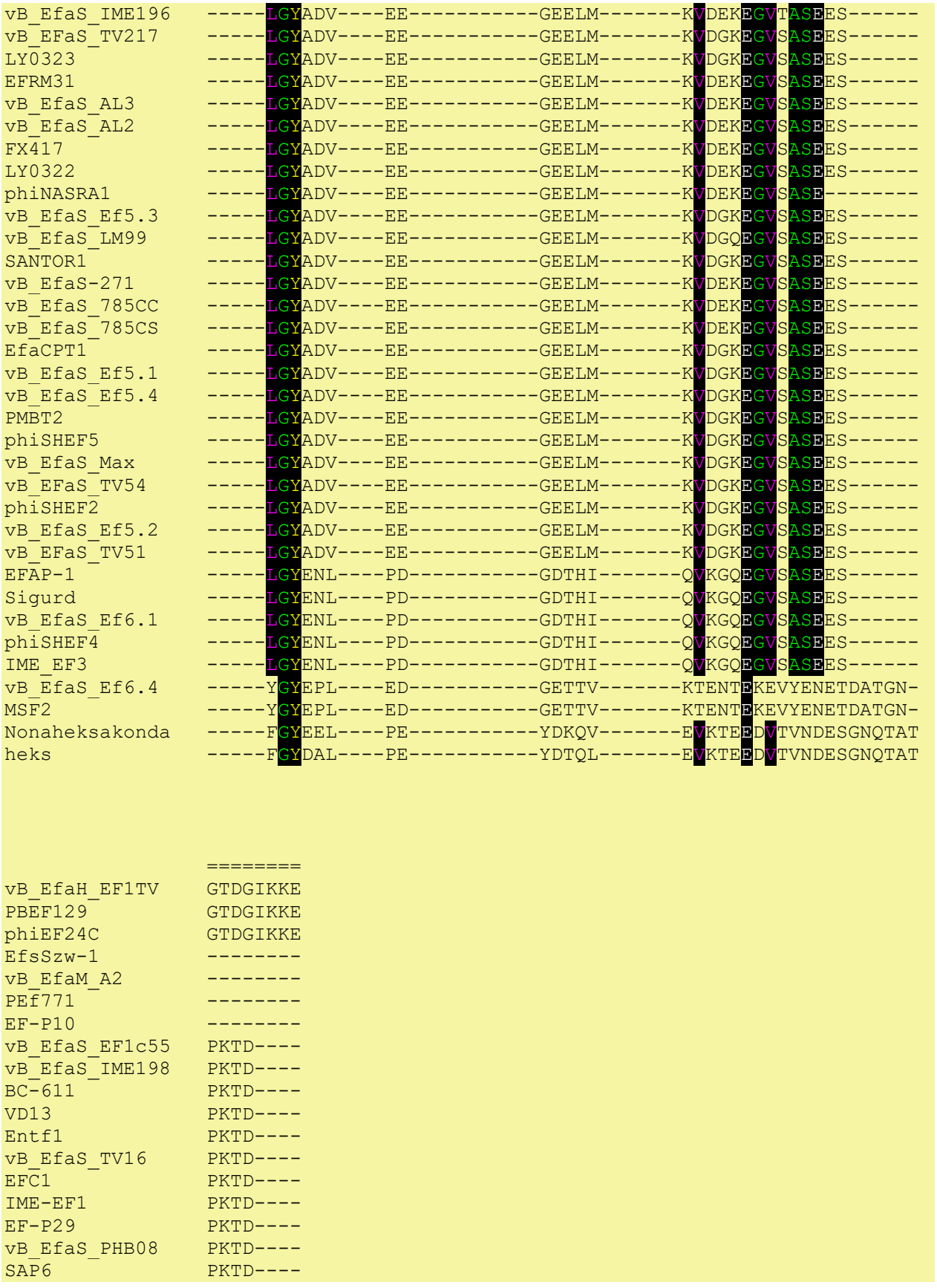

LY0322 ----------AN----G-------GSTNSNG-----VI------VLD
phiNASRA1 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_Ef5.3 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_LM99 ----------AN----G-------GSTNSNG-----VI------VLD
SANTOR1 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS-271 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_785CC ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_785CS ----------AN----G-------GSTNSNG-----VI------VLD
EfaCPT1 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_Ef5.1 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_Ef5.4 ----------AN----G-------GSTNSNG-----VI------VLD
PMBT2 ----------AN----G-------GSTNSNG-----VI------VLD
phiSHEF5 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_Max ----------AN----G-------GSTNSNG-----VI------VLD
vB_EFaS_TV54 ----------AN----G-------GSTNSNG-----VI------VLD
phiSHEF2 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_Ef5.2 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EFaS_TV51 ----------AN----G-------GSTNSNG-----VI------VLD
EFAP-1 ----------AN----G-------GSTNSNG-----VI------VLD
Sigurd ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_Ef6.1 ----------AN----G-------GSTNSNG-----VI------VLD
phiSHEF4 ----------AN----G-------GSTNSNG-----VI------VLD
IME_EF3 ----------AN----G-------GSTNSNG-----VI------VLD
vB_EfaS_Ef6.4 ----------AN----G-------GSTNSNG-----VI------VLD
MSF2 ----------AN----G-------GSTNSNG-----VI------VLD
Nonaheksakonda ----------AN----A-------GSTKSNG-----VI------VLD
heks ----------AN----A-------GSTKSNG-----VI------VLD
490
500
510
520
530
540
vB_EfaH_EF1TV EKMQASQNKGLQPLLRFIE--DTVNTYIV--AEF-GEKYQFQFRGGDLSAQLDKLK----
PBEF129 EKMQASQNKGLQPLLRFIE--DTVNTYIV--AEF-GEKYQFQFRGGDLSAQLDKLK----
phiEF24C EKMQASQNKGLQPLLRFIE--DTVNTYIV--AEF-GEKYQFQFRGGDLSAQLDKLK----
EfsSzw-1 EKMQASQNKGLQPLLRFIE--DTVNTFII--SEF-GDKYQFQFRGGDLSAQLDKIK----
vB_EfaM_A2 EKMQASQNKGLQPLLRFIE--DTVNTFII--AEF-GDKYQFQFRGGDLSAQLDKIK----
PEf771 EKMQASQNKGLQPLLRFIE--DTVNTFII--AEF-GDKYQFQFRGGDLSAQLDKIK----
EF-P10 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
vB_EfaS_EF1c55 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
vB_EfaS_IME198 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
BC-611 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
VD13 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
Entf1 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
vB_EfaS_TV16 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
EFC1 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
IME-EF1 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
EF-P29 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
vB_EfaS_PHB08 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
SAP6 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
vB_EfaS_HEf13 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
EfsWh-1 KHIENIKNRAKLDIFSLSQTPDLVSKDFT--AAS-GQAL--KA-ATQPLENKSAVK----
phiFL3B DGVNLLRSAILDDIHKITYVPDMNDEKFS--GNVSGEAM--KY-KLFGLLQLMSVK----
phiFL3A DGVNLLRSAILDDIHKITYVPDMNDEKFS--GNVSGEAM--KY-KLFGLLQLMSVK----
vB_EfaS_IME197 NGVEAYKERVRKGIHEISMIPDLTDTNFS--GVQSGEAM--KY-KMFGFNQMTAVK----
EFC-1 NGVEAYKERVRKGIHEISMIPDLTDTNFS--GVQSGEAM--KY-KMFGFNQMTAVK----
phiEf11 QGTEAYNDRLVSDMLRFTSLIDFTDENIG--SNQSGIGF--RF-KGWGSDNDRKNK----


phiFL2B EGAETYKKRLVNDILRFTFTPDTLDANFS--GTQSGESM--KY-KLMAADNRRVMQ----
phiFL2A EGAETYKKRLVNDILRFTFTPDTLDANFS--GTQSGESM--KY-KLMAADNRRVMQ----
phiFL1A EGAETYKKRLVNDILRFTFTPDTLDANFS--GTQSGESM--KY-KLMAADNRRVMQ----
phiFL1C EGAETYKKRLVNDILRFTFTPDTLDANFS--GTQSGESM--KY-KLMAADNRRVMQ----
phiFL1B EGAETYKKRLVNDILRFTFTPDTLDANFS--GTQSGESM--KY-KLMAADNRRVMQ----
phiFL4A IPTEGRKTKMEIDKENIYKFGMAFDSTQVGDGNITNIVIKARY-TLLNMKANKTEARLRA
AE4_2 IPTEGRKTKMEIDKENIYKFGMAFDSTQVGDGNITNIVIKARY-TLLNMKANKTEARLRA
vB_EfaS-DELF1 ETQDFTENKINTEVLKLIQ-ANKFSTQQI--AKVLGIPVS-RF-GQELVNSSDTSQ----
EFA1 ETTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_IME196 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EFaS_TV217 ETTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
LY0323 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
EFRM31 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_AL3 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_AL2 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
FX417 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
LY0322 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
phiNASRA1 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_Ef5.3 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_LM99 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
SANTOR1 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS-271 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_785CC ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_785CS ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
EfaCPT1 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_Ef5.1 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_Ef5.4 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
PMBT2 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
phiSHEF5 ETTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_Max ETTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EFaS_TV54 ETTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
phiSHEF2 ETTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_Ef5.2 ETTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EFaS_TV51 ETTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
EFAP-1 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
Sigurd ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_Ef6.1 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
phiSHEF4 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
IME_EF3 ESTEFQEYKMNTDILKLIQ-GNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGA----
vB_EfaS_Ef6.4 EATDFTEYKMNTDILKLIQ-SNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGQ----
MSF2 EATDFTEYKMNTDILKLIQ-SNKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGQ----
Nonaheksakonda ESTEFTEYKMNTEILKLIQ-ANKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGQ----
heks ESTEFTEYKMNTEILKLIQ-ANKFSTQQI--AKVLGIPLN-RF-GMELVNSTDSGQ----
550 560 570 580 590 600
vB_EfaH_EF1TV IIEQEGKVFRTVNEIRHDKGLEPIK----------G
PBEF129 IIEQEGKVFRTVNEIRHDKGLEPIK----------G
phiEF24C IIEQEGKVFRTVNEIRHDKGLEPIK----------G
EfsSzw-1 ILEAEVKVFKTVNEARAEKGLEPIK----------G
vB_EfaM_A2 ILEAEVKVFKTVNEARAEKGLEPIK----------G
PEf771 ILEAEVKVFKTVNEARAEKGLEPIK----------G
EF-P10 -----ESKFRKVLAKRYDLVCSYLEFMNK----------AKDLKP
vB_EfaS_EF1c55 -----ESKFRKVLAKRYELICSYLELMSK----------EKDLKP
vB_EfaS_IME198 -----ESKFRKVLAKRYDLVCSYLEFMNK----------AKDLKP


BC-611 -----ESKFRKVLAKRYELVCSYLEFMNK----------AKDLKP
VD13 -----ESKFRKVLAKRYELVCSYLEFMNK----------AKDLKP
Entf1 -----ESKFRKVLAKRYELVCSYLEFMNK----------AKDLKP
vB_EfaS_TV16 -----ESKFRKVLAKRYELVCSYLEFMNK----------AKDLKP
EFC1 -----ESKFRKVLAKRYELVCSYLEFMNK----------AKDLKP
IME-EF1 -----ESKFRKVLAKRYDLVCSYLEFMNK----------AKDLKP
EF-P29 -----ESKFRKVLAKRYELVCSYLEFMNK----------AKDLKP
vB_EfaS_PHB08 -----ESKFRKVLAKRYELVCSYLEFMNK----------AKDLKP
SAP6 -----ESKFRKVLAKRYELVCSYLEFMNK----------AKDLKP
vB_EfaS_HEf13 -----ESKFRKVLAKRYELVCSYLEFMNK----------AKDLKP
EfsWh-1 -----ESKFRKVLAKRYELVCSYLEFMNK----------AKDLKP
phiFL3B -----SRYMIKGLRKRMELFETILKVKDN----------SIDAQG
phiFL3A -----SRYMIKGLRKRMELFETILKVKDN----------SIDAQG
vB_EfaS_IME197 -----QRLFKKSLVRRYRLLFNLKSSVSEID----------NSDLKG
EFC-1 -----QRLFKKSLVRRYRLLFNLKSSVSEID----------NSDLKG
phiEf11 -----ERMVKKAIMRRLRLLTYSWSLKDNLNKPTGLAEKVKSFFVSRDNDKELLFEKVNA
phiFL2B -----QRLFEKGLMRRLRLAINIWRIKGNDSTV----------YNDINN
phiFL2A -----QRLFEKGLMRRLRLAINIWRIKGNDSTV----------YNDINN
phiFL1A -----QRLFEKGLMRRLRLAINIWRIKGNDSTV----------YNDINN
phiFL1C -----QRLFEKGLMRRLRLAINIWRIKGNDSTV----------YNDINN
phiFL1B -----QRLFEKGLMRRLRLAINIWRIKGNDSTV----------YNDINN
phiFL4A LLEWMNKLVIDDINRRYTKAFD----------P
AE4_2 LLEWMNKLVIDDINRRYTKAFD----------P
vB_EfaS-DELF1 -----NDIYIASTIAMYESSI----------C
EFA1 -----NDIYIASTISQYERAI----------C
vB_EfaS_IME196 -----NDIYIASTISQYERAI----------C
vB_EFaS_TV217 -----NDIYIASTISQYERAI----------C
LY0323 -----NDIYIASTISQYERAI----------C
EFRM31 -----NDIYIASTISQYERAI----------C
vB_EfaS_AL3 -----NDIYIASTISQYERAI----------C
vB_EfaS_AL2 -----NDIYIASTISQYERAI----------C
FX417 -----NDIYIASTISQYERAI----------C
LY0322 -----NDIYIASTISQYERAI----------C
phiNASRA1 -----NDIYIASTISQYERAI----------C
vB_EfaS_Ef5.3 -----NDIYIASTISQYERAI----------C
vB_EfaS_LM99 -----NDIYIASTISQYERAI----------C
SANTOR1 -----NDIYIASTISQYERAI----------C
vB_EfaS-271 -----NDIYIASTISQYERAI----------C
vB_EfaS_785CC -----NDIYIASTISQYERAI----------C
vB_EfaS_785CS -----NDIYIASTISQYERAI----------C
EfaCPT1 -----NDIYIASTISQYERAI----------C
vB_EfaS_Ef5.1 -----NDIYIASTISQYERAI----------C
vB_EfaS_Ef5.4 -----NDIYIASTISQYERAI----------C
PMBT2 -----NDIYIASTISQYERAI----------C
phiSHEF5 -----NDIYIASTISQYERAI----------C
vB_EfaS_Max -----NDIYIASTISQYERAI----------C
vB_EFaS_TV54 -----NDIYIASTISQYERAI----------C
phiSHEF2 -----NDIYIASTISQYERAI----------C
vB_EfaS_Ef5.2 -----NDIYIASTISQYERAI----------C
vB_EFaS_TV51 -----NDIYIASTISQYERAI----------C
EFAP-1 -----NDIYIASTISQYERAI----------C
Sigurd -----NDIYIASTISQYERAI----------C
vB_EfaS_Ef6.1 -----NDIYIASTISQYERAI----------C
phiSHEF4 -----NDIYIASTISQYERAI----------C
IME_EF3 -----NDIYIASTISQYERAI----------C
vB_EfaS_Ef6.4 -----NDIYIASTISQYESAI----------C
MSF2 -----NDIYIASTISQYESAI----------C


Nonaheksakonda -----NSIYIASTLSQYESAI-----C
heks -----NSIYIASTLSQYESAI-----C
610 620 630 640 650 660
vB_EfaH_EF1TV GDV--ILNGVHIQAIGQALQEEQLEYQRSQDRLNRLLELSGGDVEQPEPEEPKDNQNDTD
PBEF129 GDV--ILNGVHIQAIGQALQEEQLEYQRSQDRLNRLLELSGGDVEQPEPGEPKDSQNDTD
phiEF24C GDV--ILNGVHIQAIGQALQEEQLEYQRSQDRLNRLLELSGGDVEQPEPEEPKDS-----
EfsSzw-1 GDV--ILDGVLIQSIGQLMQQEQFDYQKQQDKLNRLLELSGGEAD--DSNGITFQDKQ--
vB_EfaM_A2 GDV--ILDGVLIQSIGQLMQQEQFNYQKQQDKLNRLLELSGGEAD--DSNGISFQDKQ--
PEf771 GDV--ILDGVLIQSIGQLMQQEQFNYQKQQDKLNRLLELSGGEAD--DSNGISFQDKQ--
EF-P10 NEVSPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADAQRQ
vB_EfaS_EF1c55 NEVTPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARLEVEKADEQRQ
vB_EfaS_IME198 NEVSPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADAQRQ
BC-611 YEVTPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADAQRQ
VD13 YEVTPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADEQRQ
Entf1 NEVTPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADEQRQ
vB_EfaS_TV16 NEVTPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADEQRQ
EFC1 NEVSPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADAQRQ
IME-EF1 NEVSPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADAQRQ
EF-P29 NEVSPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADEQRQ
vB_EfaS_PHB08 NEVSPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADEQRQ
SAP6 NEVSPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADAQRQ
vB_EfaS_HEf13 NEVSPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADEQRQ
EfsWh-1 NEVSPVFVRNLPQSYAELADMAV--KLRDMLPDETIINQFPWITD--ARQEVEKADEQRQ
phiFL3B TKIKLK--PNLPVNTSDIINQIVSAYQAGILPLKVLLGWLPDIDD---VDEVLKQ-----
phiFL3A TKIKLK--PNLPVNTSDIINQIVSAYQAGILPLKVLLGWLPDIDD---VDEVLKQ-----
vB_EfaS_IME197 --LRIIFTPNLPKAILEELKTLIDSGAE--LSQETILGLASFVDD--VQAELERVKNENK
EFC-1 --LRIIFTPNLPKAILEELKTLIDSGAE--LSQETILGLASFVDD--VQAELERVKNENK
phiEf11 --IEILFTPNVPQSDKEIMEVIA--GMVGIVSDETLCEMAAKLTG----VPVQ-------
phiFL2B TNI--IFTANVPKSDNEVVSLAS--QLVGQVSDETLFEILSTVTG------VDPD-----
phiFL2A TNI--IFTANVPKSDNEVVSLAS--QLVGQVSDETLFEILSTVTG------VDPD-----
phiFL1A TNI--IFTANVPKSDNEVVSLAS--QLVGQVSDETLFEILSTVTG------VDPD-----
phiFL1C TNI--IFTANVPKSDNEVVSLAS--QLVGQVSDETLFEILSTVTG------VDPD-----
phiFL1B TNI--IFTANVPKSDNEVVSLAS--QLVGQVSDETLFEILSTVTG------VDPD-----
phiFL4A TEVSFTFTREVMVNETDIVNDEKTEAETRKIILESILQVAPRLDD----DNVLRLICEQF
AE4_2 TEVSFTFTREVMVNETDIVNNEKAEAETRKIILESILQVAPRLDD----DNVLRLICEQF
vB_EfaS-DELF1 DELNLKLGVNLELDLSKLRQDTKEDRLRRIAEGKVKSEFAQALTV----NDARTY-----
EFA1 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
vB_EfaS_IME196 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
vB_EFaS_TV217 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
LY0323 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
EFRM31 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNNELLKAIKT----NEIRDY-----
vB_EfaS_AL3 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
vB_EfaS_AL2 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
FX417 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
LY0322 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
phiNASRA1 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQSKELLKAIKT----NEIRDY-----
vB_EfaS_Ef5.3 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
vB_EfaS_LM99 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
SANTOR1 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
vB_EfaS-271 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
vB_EfaS_785CC DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
vB_EfaS_785CS DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
EfaCPT1 DEIEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----


vB_EfaS_Ef5.1 DETEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
vB_EfaS_Ef5.4 DETEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
PMBT2 DETEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
phiSHEF5 DETEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
vB_EfaS_Max DETEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
vB_EFaS_TV54 DETEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
phiSHEF2 DETEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
vB_EfaS_Ef5.2 DETEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
vB_EFaS_TV51 DETEIKTGNELELDLTTLLNDTYEDRRKRVFEGKQNKELLKAIKT----NEIRDY-----
EFAP-1 DETEIKTGNTLELDLTTLLNDTYEDRRKRVFEGKSGKELLGAIQV----NEIRDY-----
Sigurd DETEIKTGNTLELDLTTLLNDTYEDRRKRVFEGKSGKELLGAIQV----NEIRDY-----
vB_EfaS_Ef6.1 DETEIKTGNTLELDLTTLLNDTYEDRRKRVFEGKSGKELLGAIQV----NEIRDY-----
phiSHEF4 DETEIKTGNTLELDLTTLLNDTYEDRRKRVFEGKSGKELLGAIQV----NEIRDY-----
IME_EF3 DETEIKTGNTLELDLTTLLNDTYEDRRKRVFEGKSGKELLGAIQV----NEIRDY-----
vB_EfaS_Ef6.4 DEIAIKTGNVLEMDFSSLMNDTLLDDRRKQLFSGKASKELLSTIKE----NEVRAY-----
MSF2 DEIAIKTGNVLEMDFSSLMNDTLLDDRRKQLFSGKASKELLSTIKE----NEVRAY-----
Nonaheksakonda DEIAIKTGTVLEMDFSTILNDTLLDDRRKVLFTGKVSKELLGSFTN----NEVREY-----
heks DEIAIKTGTVLEMDFSTILNDTLLDDRRKVLFTGKVSKELLGSFTN----NEVREY-----
670 680 690 700 710 720
vB_EfaH_EF1TV -----VSFQDE----QQ---------GLNGKSKKVNGKVDDNVGKDGQLKSEENTNSTKH
PBEF129 -----VSFQDE----QQ---------GLNGKSKKVNGKVDDNVGKDGQLKSEENTNSTKH
phiEF24C QNDTDVSFQDE----QQ---------GLNGKSKKVNGKVDDNVGKDGQLKSEENTNSTKH
EfsSzw-1 -----QGLDGT----SASVNGKGESAVGKDGQIKGKGVTNTNSAKQGMKGDKPNDWQK--
vB_EfaM_A2 -----QGLDGT----SASVNGKGESSVGKDGQIKGKGITNTNSAKQGMKGDKPNDWQK--
PEf771 -----QGLDGT----SASVNGKGESSVGKDGQIKGKGITNTNSAKQGMKGDKPNDWQK--
EF-P10 KRAD-IALQNF----KQ----------TSACSRSFYSSS--------
vB_EfaS_EF1c55 KRAD-IALQNF----KQTSAVQGASTASANKLDKNPANTSTITTTDPVAAKEQEKAIQKK
vB_EfaS_IME198 KRAD-IALQNF----KQTSAVQGASTAAANKLDKNPANTSTITTTDPVAAKEQEKAIQKK
BC-611 KRAD-IALQNF----KQTSAVQGASTAAANKLDKNPANTSTITTTDPVAAKEQEKAIQKK
VD13 KRAD-IALQNF----KQTSAVQGASTAAANKLDKNPANTSTITTTDPVAAKEQEKAIQKK
Entf1 KRAD-IALQNF----KQTSAVQGASTAAANKLDKNPANTSTITTTDPVAAKEQEKAIQKK
vB_EfaS_TV16 KRAD-IALQNF----KQTSAVQGASTAAANKLDKNPANTSTITTTDPVAAKEQEKAIQKK
EFC1 KRAD-IALQNF----KQTSAVQGASTAAANKLDKNPANTSTITTTDPVAAKEQEKAIQKK
IME-EF1 KRAD-IALQNF----KQTSAVQGASTAAANKLDKNPSNTSTITTTDPVAAKEQEKAIQKK
EF-P29 KRAD-IALQNF----KQTSAVQGASTAAANKLDKNPANTSTITTTDPVAAKEQEKAIQKK
vB_EfaS_PHB08 KRAD-IALQNF----KQTSAVQGASTAAANKLDKNPANTSTITTTDPVAAKEQEKAIQKK
SAP6 KRAD-IALQNF----KQTSAVQGASTAAANKLDKNPANTSTITTTDPVAAKEQEKAIQKK
vB_EfaS_HEf13 KRAD-IALQNF----KQTSAVQGASTAAANKLDKNPANTSTITTTDPVAAKEQEKAIQKK
EfsWh-1 KRAD-IALQNF----KQTSAVQGASTAAANKLDKNPANTSTITTTDPVAAKEQEKAIQKK
phiFL3B -----LNTEK-----EK----------AIELNQKAYGVQAKDSHSDLDKGSEEDNYDDY
phiFL3A -----LNTEK-----EK----------AIELNQKAYGVQAKDSHSDLDKGSEEDNYDDY
vB_EfaS_IME197 P----LDITD----------EEVAKTKVNQAKFLAKQLEKETEEE------
EFC-1 P----LDITD----------EEVAKTKVNQAKFLAKQSEKETEEE------
phiEf11 -----TELNRL----KK----------ENQPD-------TLSDEEAAKLKEKQAEFLAN
phiFL2B -----VELKRI----KE----------ETNDKPEPRRPEVNKNETGNEDDEREAAKRI
phiFL2A -----VELKRI----KE----------ETNDKPEPRRPEVNKNETGNEDDEREAAKRI
phiFL1A -----VELKRI----KE----------ETNDKPEPRRPEVNKNETGNEDDEREAAKRI
phiFL1C -----VELKRI----KE----------ETNDKPEPRRPEVNKNETGNEDDEREAAKRI
phiFL1B -----VELKRI----KE----------ETNDKPEPRRPEVNKNETGNEDDEREAAKRI
phiFL4A D----LDWEDVKEALEEAEYTKGLSDNTDEEETAVNPDDPTQQMAEGATGSTESQLPENG
AE4_2 D----LDWEDVKEALEEAEYTKGLSDNTDEVETAVNPDDPTQQMAEGAAGSTESQLPENG
vB_EfaS-DELF1 -----LGFQEV----DG----------GQLLLGQMKNNTQNTNEQDVKVNDEEHGDTST
EFA1 -----LGYADV----EE----------GEELM-------NIDEKEGVSASEES------


vB_EfaS_IME196 -----LGYADV----EE----------GEELM-------KVDEKEGVTASEES------
vB_EFaS_TV217 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
LY0323 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
EFRM31 -----LGYADV----EE----------GEELM-------KVDEKEGVSASEES------
vB_EfaS_AL3 -----LGYADV----EE----------GEELM-------KVDEKEGVSASEES------
vB_EfaS_AL2 -----LGYADV----EE----------GEELM-------KVDEKEGVSASEES------
FX417 -----LGYADV----EE----------GEELM-------KVDEKEGVSASEES------
LY0322 -----LGYADV----EE----------GEELM-------KVDEKEGVSASEES------
phiNASRA1 -----LGYADV----EE----------GEELM-------KVDEKEGVSASE--------
vB_EfaS_Ef5.3 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
vB_EfaS_LM99 -----LGYADV----EE----------GEELM-------KVDGQEGVSASEES------
SANTOR1 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
vB_EfaS-271 -----LGYADV----EE----------GEELM-------KVDEKEGVSASEES------
vB_EfaS_785CC -----LGYADV----EE----------GEELM-------KVDEKEGVSASEES------
vB_EfaS_785CS -----LGYADV----EE----------GEELM-------KVDEKEGVSASEES------
EfaCPT1 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
vB_EfaS_Ef5.1 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
vB_EfaS_Ef5.4 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
PMBT2 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
phiSHEF5 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
vB_EfaS_Max -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
vB_EFaS_TV54 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
phiSHEF2 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
vB_EfaS_Ef5.2 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
vB_EFaS_TV51 -----LGYADV----EE----------GEELM-------KVDGKEGVSASEES------
EFAP-1 -----LGYENL----PD----------GDTHI-------QVKGQEGVSASEES------
Sigurd -----LGYENL----PD----------GDTHI-------QVKGQEGVSASEES------
vB_EfaS_Ef6.1 -----LGYENL----PD----------GDTHI-------QVKGQEGVSASEES------
phiSHEF4 -----LGYENL----PD----------GDTHI-------QVKGQEGVSASEES------
IME_EF3 -----LGYENL----PD----------GDTHI-------QVKGQEGVSASEES------
vB_EfaS_Ef6.4 -----YGYEPL----ED----------GETTV-------KTENTEKEVYENETDATGN-
MSF2 -----YGYEPL----ED----------GETTV-------KTENTEKEVYENETDATGN-
Nonaheksakonda -----FGYEEL----PE----------YDKQV-------EVKTEEDVTVNDESGNQTAT
heks -----FGYDAL----PE----------YDTQL-------EVKTEEDVTVNDESGNQTAT
========
vB_EfaH_EF1TV GTDGIKKE
PBEF129 GTDGIKKE
phiEF24C GTDGIKKE
EfsSzw-1 --------
vB_EfaM_A2 --------
PEf771 --------
EF-P10 --------
vB_EfaS_EF1c55 PKTD----
vB_EfaS_IME198 PKTD----
BC-611 PKTD----
VD13 PKTD----
Entf1 PKTD----
vB_EfaS_TV16 PKTD----
EFC1 PKTD----
IME-EF1 PKTD----
EF-P29 PKTD----
vB_EfaS_PHB08 PKTD----
SAP6 PKTD----

```
vB_EfaS_HEf13    PKTD----
EfsWh-1          PKTD----
phiFL3B          SNI-----
phiFL3A          SNI-----
vB_EfaS_IME197   --------
EFC-1            --------
phiEf11          QSETEED-
phiFL2B          NEDE----
phiFL2A          NEDE----
phiFL1A          NEDE----
phiFL1C          NEDE----
phiFL1B          NEDE----
phiFL4A          --------
AE4_2            --------
vB_EfaS-DELF1    TDTGEE--
EFA1             --------
vB_EfaS_IME196   --------
vB_EFaS_TV217    --------
LY0323           --------
EFRM31           --------
vB_EfaS_AL3      --------
vB_EfaS_AL2      --------
FX417            --------
LY0322           --------
phiNASRA1        --------
vB_EfaS_Ef5.3    --------
vB_EfaS_LM99     --------
SANTOR1          --------
vB_EfaS-271      --------
vB_EfaS_785CC    --------
vB_EfaS_785CS    --------
EfaCPT1          --------
vB_EfaS_Ef5.1    --------
vB_EfaS_Ef5.4    --------
PMBT2            --------
phiSHEF5         --------
vB_EfaS_Max      --------
vB_EFaS_TV54     --------
phiSHEF2         --------
vB_EfaS_Ef5.2    --------
vB_EFaS_TV51     --------
EFAP-1           --------
Sigurd           --------
vB_EfaS_Ef6.1    --------
phiSHEF4         --------
IME_EF3          --------
vB_EfaS_Ef6.4    --------
MSF2             --------
Nonaheksakonda   EYTE----
heks             EYTE----
```

Parameters used
Minimum Number Of Sequences For A Conserved Position: 35
Minimum Number Of Sequences For A Flanking Position: 58
Maximum Number Of Contiguous Nonconserved Positions: 8
Minimum Length Of A Block: 10
Allowed Gap Positions: None
Use Similarity Matrices: Yes

Flank positions of the 1 selected block(s)
Flanks: [484 499]

New number of positions in input.fasta-gb: **16** (2% of the original 728 positions)

Supplemental Table 2: Built-in curer aligned sequences used for creation phylogenetic terminase large subunit tree.

EFC-1 EKNIIQSELG IIGDLVTYHS LLKYKTNGED NKARWKKVLG GYIDEINADM
EfsSzw-1 MTKFVQTRLD PVSTIIDPED SLKAKKRNSK GAVEGVDIDY LSMDEYDVVT
PEf771 MTKFVQTRLD PVSTIIDPED SLKAKKRNSK GAVEGVDIDY LSMDEYDVVT
EFGrNG MTKFVQTRLD PVSTIIDPED SLKAKKRNSK GAVEGVDIDY LSMDEYDVVT
EFGrKN MTKFVQTRLD PVSTIIDPED SLKAKKRNSK GAVEGVDIDY LSMDEYDVVT
v.EfaM MTKFVQTRLD PVSTIIDPED SLKAKKRNSK GAVEGVDIDY LSMDEYDVVT
v.O.Ben MTKFVQTRLD PVSTIIDPED SLKAKKRNSK GAVEGVDIDY LSMDEYDVVT
EFDG1 MTKFVQTRLD PVSTIIDPED SLKAKKRNSK GAVEGVDIDY LSMDEYDVVT
phiEF24C MTKFVQTRLD PVSTIVDQEN SLKAKKRNSK GAVEGVDIDY LSMDEYDVVN
vBEfME1 MTKFVQTRLD PVSTIVDQEN SLKAKKRNSK GAVEGVDIDY LSMDEYDVVN
EFLK1 MTKFVQTRLD PVSTIVDQEN SLKAKKRNSK GAVEGVDIDY LSMDEYDVVN
PBEF129 MTKFVQTRLD PVSTIVDQEN SLKAKKRNSK GAVEGVDIDY LSMDEYDVVN
v.EfM.Ef2. MTKFVQTRLD PVSTIVDQEN SLKAKKRNSK GAVEGVDIDY LSMDEYDVVN
EFA-1 RADANSIFIK RILQELKPKI SFDVGPKPDH PSVKSVGITG QLTGSRAIFD
EFA-2 RADANSIFIK NILSELKPRI SFDVGPKPDH PSVKSVGITG QLTGSRAIFA
phiFL2B DTDVVEIILQ VAFKTLVFAE LLKETTEVDG SQLLGMGIYA SLTGKHAIVS
phiFL2A DTDVVEIILQ VAFKTLVFAV LLKETTEVDG SQLLGMGIYA SLTGKHAIVS
phiFL4A DTDVVEIILQ VAFKTLVFAV LLKETTEIDG SQLLGMGIYA SLTGKHAIVS
AE4_2 DTDVVEIILQ VAFKTLVFAV LLKETTEIDG SQLLGMGIYA SLTGKHAIVS
phiFL1C QKNILQELYN TYDKHGNFKK VVQAYTNIGG GSIRGMTAYG AYINEASAFA
phiFL1B QKNILQELYN TYDKHGNFKK VVQAYTNIGG GSIRGMTAYG AYINEASAFA
phiFL1A QKNILQELYN TYDKHGNFKK VVQAYTNIGG GSIRGMTAYG AYINEASAFA
vipetofem LSSIVKKEME QQKSPLINKV RGEVRCLTKS NRMDGRKANV FVADEVGLYE
nattely LSSIVKKEME QQKSPLINKV RGEVRCLTKS NRMDGRKANV FVADEVGLYE
v.EfS.DELF QAGIVFRMLK SQVSKEVKKR RYDIEASDED DTLDGLDVLC GILDEYGATK
vBEfS_I8 QAGIVFNMLK SQVSDGTRKN KKDIEHGDED DSLDGLDVLC GVLDEYGATK
Sigurd QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
v.EfS.Ef6. QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
I.EF3 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
vBEfS_7C QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
v.EfaS QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
FX417 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
LY0322 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
v.EfS.AL3 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
phiNASRA1 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
vBEfS_Ef54 QASIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
v.EfS.LM99 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
v.EF.TV217 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
EfaCPT1 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
EFAP-1 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
v.EfS.Ef5. QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
vBEfS_Ef52 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
vBEfS_Ef51 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
PMBT2 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK

v.EF.TV54 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
v.EfS.Max QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
EFA1 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
v.EF.TV51 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
EFRM31 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
v.EfS.AL2 QAGIVFNMLK SQVSDGTRKN KKDIEHEDED DSLDGLDVLC GVLDEYGATK
vBEfS_7S QAGIVFNMLK SQVSDGTRKN KKDIEHVDED DSLDGLDVLC GVLDEYGATK
vBEfS_Ef64 QAGIVFRMLK SQSSDSVRKN KYDIEHEDED DSLDGLDVLC GVLDEYGATK
MSF2 QAGIVFRMLK SQSSDSVRKN KYDIEHEDED DSLDGLDVLC GVLDEYGATK
Nonaheksak QAGIVFRMLK SQVSESVRKN KYDIEQEDED DSLDGLDVLC GVLDEYGATK
heks QAGIVFRMLK SQVSESVRKN KYDIEQEDED DSLDGLDVLC GVLDEYGATK
ZEF1 RQSVTELLWE VFSKELDFER GDSYTNEYQD NKSTDLKNYS NFLKKFPIID
AE4_17 RQSVTELLWE VFSKELDFER GDSYTNEYQD NKSTDLKNYS NFLKKFPIID
v.EfS.Ef2. DNTIRDSIFQ EFWNLLDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
v.EfS.EF1c DNTIRDSIFQ EFWNLLDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
E.P10 DNTIRDSIFQ EFWNILDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
E.P29 DNTIRDSIFQ EFWNILDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
Entf1 DNTIRDSIFQ EFWNILDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
BC-611 DNTIRDSIFQ EFWNLLDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
VD13 DNTIRDSIFQ EFWNLLDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
EfsWh-1 DNTIRDSIFQ EFWNLLDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
v.EfS.HEf1 DNTIRDSIFQ EFWNLLDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
SAP6 DNTIRDSIFQ EFWNLLDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
v.EfS.Ef7. DNTIRDSIFQ EFWNLLDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
v.EfS.IME1 DNTIRDSIFQ EFWNLLDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
EFC1 DNTIRDSIFQ EFWNILDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
v.EfS.PHB0 DNTIRDSIFQ EFWNILDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
I.EF1 DNTIRDSIFQ EFWNILDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
v.EfS.TV16 DNTIRDSIFQ EFWNILDFCS YMTIKLNKTD ERIKSIQGLT DIIMEEAEQT
phiFL3A DNTIELSIFE QIEEQGVSKN KMKITKRGNN ERIKSLKDSR FWIEELAFNS
vBEfS_I7 DNTLEQSIYE QMEQQGVTHN KLRITIRGNN ERIKSLKDSR FWIEELGFNS

EYVREIFRAD YVMATLNPDD PELPIYHEYI DQLNQSWFFG FSHNEKKKQI
RWSTPSADVG IHALFYNQMN YEDYVPEAPV VDGSFQEPLD RWYNPKYPDR
RWSTPSADVG IHALFYNQMN YEDYVPEAPV VDGSFQEPLD RWYNPKYPDR
RWSTPSADVG IHALFYNQMN YEDYVPEAPV VDGSFQEPLD RWYNPKYPDR
RWSTPSADVG IHALFYNQMN YEDYVPEAPV VDGSFQEPLD RWYNPKYPDR
RWSTPSADVG IHALFYNQMN YEDYVPEAPV VDGSFQEPLD RWYNPKYPDR
RWSTPSADVG IHALFYNQMN YEDYVPEAPV VDGSFQEPLD RWYNPKYPDR
RWSTPSADVG IHALFYNQMN YEDYVPEAPV VDGSFQEPLD RWYNPKYPDR
RWSTPSADMG IHGLFYNEMS YDAYTPEAPV VDGSFQEPLD RWYNPKYPDR
RWSTPSADMG IHGLFYNEMS YDAYTPEAPV VDGSFQEPLD RWYNPKYPDR
RWSTPSADMG IHGLFYNEMS YDAYTPEAPV VDGSFQEPLD RWYNPKYPDR
RWSTPSADMG IHGLFYNEMS YDAYTPEAPV VDGSFQEPLD RWYNPKYPDR

RWSTPSADMG IHGLFYNEMS YDAYTPEAPV VDGSFQEPLD RWYNPKYPDR
AILKPGGTII YLGTPSYGDR LAPMLQAELE ERELSYFALQ FMLNPNLSDE
ALLKPLPTSR VIYLGYYSQR LAPMLRAEYE ERELEYFTLQ FMLNPNLSDE
RAERERTKLQ YQELQRFINT GTPWHKEDAI SKMPNVCYET GLIDQALQQM
RAERERTKLQ YQELQRFINT GTPWHKEDAI SKMPNVCYET GLIDQALQQM
RAEREKTKLQ YQELQRFINT GTPWHKEDAI SKMPNVCYET GLIDKALQQM
RAEREKTKLQ YQELQRFINT GTPWHKEDAI SKMPNVCYET GLIDKALQQM
EIVSRCSATG ARILATNPDN PEHWLKKEYI SKSIQSLDDN TFLSRTNIKS
EIVSRCSATG ARILATNPDN PEHWLKKEYI SKSIQSLDDN TFLSRTNIKS
EIVSRCSATG ARILATNPDN PEHWLKKEYI SKSIQSLDDN TFLSRTNIKS
SLNNPMTEVD YATKVYRPDD SKKWLKANPL AQKNFKKHMN IFVDEVYIRE
SLNNPMTEVD YATKVYRPDD SKKWLKANPL AQKNFKKHMN IFVDEVYIRE
NLNGPMFTID YAFVESEVDD EANWIKANPL GKMDLSKEEN FWVQDSFVTE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMYSIE YPFITSEVDD EANWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHAIE YPFITSEVDD ESNWIKSNPL AKGDMSKEMN FWVQDSFIDE
NLNGPMHSIE YPFITSEVDD QANWIKSNPL AKGDMSKEMN FWVQDSFMDE
NLNGPMHSIE YPFITSEVDD QANWIKSNPL AKGDMSKEMN FWVQDSFMDE
NLNGPMHSIE YPFITSEVDN QENWIKANPL AKGDMSKEMN FWVQDSFMDE
NLNGPMHSIE YPFITSEVDN QENWIKANPL AKGDMSKEMN FWVQDSFMDE
RWDNPIPYIG TPKVFGNPEN FASPLMALLD PMNTGACYDI VYLEENQNAR

RWDNPIPYIG TPKVFGNPEN FASPLMALLD PMNTGACYDI VYLEENQNAR
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
RLRHPTARHQ QVFVMYNPAS KDNWVYQYFH PKGSKVYKDN RFLPDHLQDK
LLRGELAGLF YKFFYNPPKR RQSWVNKKYE PENTFVYKDN PFIAEEVNAR
LLRGELGGLF YKFFFNPPKR RQSWVNKKYE PDNTFVYKDN PFISKEVEAR

ISVP
TKGL
TKGL
TKGL
TKGL
TKGL
TKGL
TKGL
TKGL
TKGL
TKGL
TKGL
TKGL
KYPL
KYPL
TPLF
TPLF
TPLF
TPLF
TPGM
TPGM
TPGM
DLKN
DLKN

EWAI
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAV
SWAI
SWAI
SWAV
SWAV
NSAL
NSAL
NTPV
NTPV
NTPV
NTPV
NTPV
NTPV
NTPV
NTPV
NTPV
NTPV
NTPV

NTPV
NTPV
NTPV
NTPV
NTPV
AKPK
DRPL