Universidade do Minho
Escola de Engenharia

Carlos André Ramalho Simões

Construction of an innocuous *Salmonella* phage

Outubro de 2015

Universidade do Minho
Escola de Engenharia

Carlos André Ramalho Simões

Construction of an innocuous *Salmonella* phage

Dissertação de Mestrado

Mestrado em Bioengenharia

Trabalho efetuado sob a orientação da

Doutora **Sanna Maria Sillankorva**

Outubro de 2015

DECLARAÇÃO

Nome: Carlos André Ramalho Simões

Endereço eletrónico: carlosandrersimoes@gmail.com Telefone:926814219

Bilhete de Identidade/Cartão do Cidadão: 14161592

Título da dissertação: Construction of an innocuous *Salmonella* phage

Orientadora

Professora Doutora Sanna Maria Silankorva

Ano de conclusão: 2015

Mestrado em Bioengenharia

Universidade do Minho, _____/_____/_____

Assinatura:

ACKNOWLEDGEMENTS

Desde já expresso o meu profundo agradecimento a todas as pessoas que me ajudaram neste último ano e que me conseguiram ajudar a alcançar este patamar de satisfação.

Em primeiro lugar, à Doutora Sanna Sillankorva pelo apoio dado, pelas sugestões e críticas, pelo à vontade com que me pôs sempre, mas acima de tudo por me ter dado a oportunidade de trabalhar com ela.

À Catarina por estar sempre presente para qualquer dúvida que precisasse no laboratório, pelo à vontade que me pôs desde o primeiro dia em laboratório e pela grande ajuda que me deu na escrita da tese. Ao Luís e à Graça por serem duas pessoas sempre dispostas a ajudar e a explicar qualquer dúvida (ao Luís principalmente pela ajuda que me deu no estudo da Bioinformática e claro aos grandes debates que tivemos). Ao Sílvio e à Ana pela grande e boa disposição que a presença deles trazia ao trabalhar no laboratório.

Aos meus amigos Joana e Macedo, pois este ano de trabalho não seria a mesma coisa sem a boa companhia e boa disposição deles. "Aturar" a Joana sem a ajuda do Macedo seria difícil.

Aos meus amigos Andreia, Rita, Sara e Pedro pelo apoio dado e pela "dor" partilhada ao longo do último ano (bem ditos foram os intervalos que fazíamos nas máquinas) foi trabalhoso, mas todos conseguimos no final.

Aos meus amigos Diana, Marco, Humberto e Ezequiel por todo o apoio dado e claro pela paciência por me aturarem, uma irmandade que nunca se vai separar. Agradeço profundamente pela grande ajuda que me deram e claro que espero retribuir já no próximo ano.

À minha família, principalmente, à minha irmã, mãe e pai por permitir que tudo isto fosse possível, por me aguentarem nas alturas de maior pressão, mas acima de tudo por todo o apoio dado durante esta fase e pelo exemplo que são para mim.

RESUMO

A *Salmonela* é um dos principais agentes patogénicos alimentares responsáveis por doenças e hospitalizações, sendo responsável por 38% das 3000 mortes reportadas anualmente, devido a intoxicações alimentares nos EUA (Centro de Controlo de Doenças dos EUA). A *Salmonela*, presente no trato gastrointestinal das aves domésticas, pode ser passada de ave para ave através de comida, água contaminada, fontes/vetores ambientais. Os Bacteriófagos (fagos) são vírus bacterianos e por isso são considerados os predadores naturais das bactérias. Os fagos são inofensivos para os humanos e animais e têm vindo a ser utilizados extensivamente para a prevenção e controlo de patogéneos bacterianos. Experiências *in vitro* têm vindo a demonstrar que os fagos quando estão sozinhos ou misturados com outros fagos (*cocktail*), conseguem diminuir significativamente os níveis de *Salmonela* em alimentos contaminados. No entanto, o futuro dos fagos na segurança alimentar encontra-se mais dependente das agências reguladoras, que ainda mostram algum desconforto aquando da utilização dos fagos. Isto deve-se principalmente ao facto de haver uma escassez de fortes evidências científicas, a uma emergência de fenótipos bacterianos resistentes a fagos e devido à alta percentagem de genes fágicos com funções desconhecidas, que podem ser fatores de virulência ou genes de toxinas.

O principal objetivo deste trabalho foi o de abordar o problema de genes com função desconhecida no genoma de um fago. Para se conseguir isso a sequência do genoma deste fago foi analisada através do uso de ferramentas bioinformáticas, todos os genes desconhecidos foram identificados e mais tarde todo o seu genoma será introduzido num cromossoma artificial bacteriano de maneira a ser geneticamente manipulado.

O fago vB_SenS_Φ38, que ataca *Salmonella enterica* serovar Enteritidis, foi geneticamente caracterizado e apresentou um genoma com um tamanho de 42439 pb, DNA de cadeia dupla, possuindo 60 *open reading frames* (ORFs). O estudo foi mais focado nas 28 proteínas hipotéticas que existem no genoma do fago, as quais todas foram clonadas e tiveram a sua atividade antimicrobiana avaliada. Estas 28 proteínas hipotéticas não demonstraram funções antimicrobianas, quer quando expressas em meio líquido ou em meio sólido.

Palavras-Chave: *Salmonella* Enteritidis, Bacteriófagos e Bioinformática.

ABSTRACT

Salmonella is one of the leading foodborne pathogens responsible for illnesses and hospitalizations, and responsible for 38% of the 3000 deaths reported annually (US Center for Disease and Control). *Salmonella*, present in the gastrointestinal tract of poultry, is passed from bird to bird for instance through contaminated feed and water, environmental sources/vectors, etc. Bacteriophages are bacterial viruses and therefore natural predators of bacteria. Bacteriophages (phages) are harmless to humans and animals, and have been extensively used for prevention and control of bacterial pathogens. *In vitro* experiments show that phages, alone and in cocktail, can decrease significantly *Salmonella* levels in contaminated foods; however, the future of phages in food safety is further dependent on the regulatory agencies that still display uneasiness when it comes to the use of phages. This is mostly due to a scarcity of strong scientific evidence, emergence of phage resistant bacterial phenotypes, and the high percentage of genes in phages with unknown function, which can be virulence factors or toxin genes.

The main objective of this work was to address the problem of genes of unknown function in a phage genome. To achieve this, the genome sequence of a phage was analysed using bioinformatics tools, all unknown genes were identified and later the whole genome of the phage will be introduced in a bacterial artificial chromosome so that it could be genetically engineered.

Phage vB_SenS_Φ38, targeting *Salmonella enterica* serovar Enteritidis, was genetically characterized and was shown to be 42439 bp in size, dsDNA and to contain 60 open reading frames (ORFs). The study further focused in the 28 hypothetical proteins that exist in the phage genome which were effectively cloned and their antimicrobial activity evaluated. These 28 hypothetical proteins have not shown any antibacterial function, either when expressed in liquid or in solid medium.

KEYWORDS: *Salmonella* Enteritidis, Bacteriophage and Bioinformatics.

OUTLINE OF THESIS

This thesis is structured in six chapters.

The thesis starts with a thorough introduction to *Salmonella* and the associated problems, phages and their applications (Chapter I). The following chapter enumerates and describes all materials and methods used during the work performed. Chapter III focuses not only in the bioinformatics tools used and the consequent results, but also in results obtained in the expression of phage proteins. Chapter IV describes the main conclusions and also reports some possible future experiences that can be done.

Finally, includes a list of the references cited in this thesis and presents some important attachments.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

HACCP - Hazard Analysis Critical Control Point

EU – European Union

EFSA - European Food Safety Authority

ECDC - European Centre for Disease Prevention and Control

EU/EEA - European Union/ *European* Economic Area

Spp. – species

DNA – Desoxyribonucleic Acid

RNA – Ribonucleic Acid

DEB - Department of Biological Engineering

LB - Lysogeny Broth

MCS - Multiple Cloning Site

GFP - Green Fluorescent Protein

IPTG - Isopropyl β-D-1-thiogalactopyranoside

ORF – Open Reading Frame

tRNA - transfer RNA

RNase – Ribonuclease

EDTA - Ethylenediaminetetraacetic acid

SDS - Sodium dodecyl sulfate

RT – Room Temperature

PCR – Polymerase Chain Reaction

T_m°C – Melting temperature

TAE - Tris-Acetate-EDTA

Amp – Ampicillin

CHAPTER I

INTRODUCTION

1. FOOD CONTAMINATIONS

Every day the quality and the food safety are put at risk and despite all the improvements that have been made, such as the application of new technologies, good manufacturing practices, guarantees of quality and hygiene. Food safety is continually put into question due changes in lifestyle and requirements customers and also because the number of international exchanges between different countries nowadays are bigger than ever (Rocourt, Moy, Vierk, & Schlundt, 2003).

World population is continually rising. It is therefore expected that food consumption, particularly of ready-to-eat foods, will also rise. From such increment arises the need of creating security measures that control food quality (Westrell, Ciampa, Boelaert, Helwigh, Korsgaard, Chríel, Ammon, 2009).

Contaminations are still an issue and according to the current legislation no kind of food should contain microorganisms or toxins that present a danger to public health (Westrell, Ciampa, Boelaert, Helwigh, Korsgaard, Chríel, Ammon, 2009). It is important to note that food products contamination may occur at any moment between production and consumption of food items (Guntupalli et al., 2007) (see Food Production Chain, Figure 1). In order to avoid microbial growth in a determined environment, good hygiene, biocides, and disinfectants should be used (Maukonen et al., 2003). However, even applying these security measures bacteria continue to be found in food (Holah, Taylor, Dawson, & Hall, 2002). Contamination problems may be further aggravated due to cross contamination, i.e., when there is transference of pathogens from raw items to the surface of others (this kind of contamination is responsible for about 40% of food related disease outbreaks) (Oliveira, Oliveira, Teixeira, Azeredo, 2007; Teixeira, Silva, Araújo, Azeredo, & Oliveira, 2007). Despite all the efforts made to fight these microorganisms, they possess defence mechanisms which, in the end, allow them to surpass the measures put in place to prevent their growth (Maukonen et al., 2003) in such a way it is still possible to find them in food and on food processing surfaces (Holah et al., 2002).

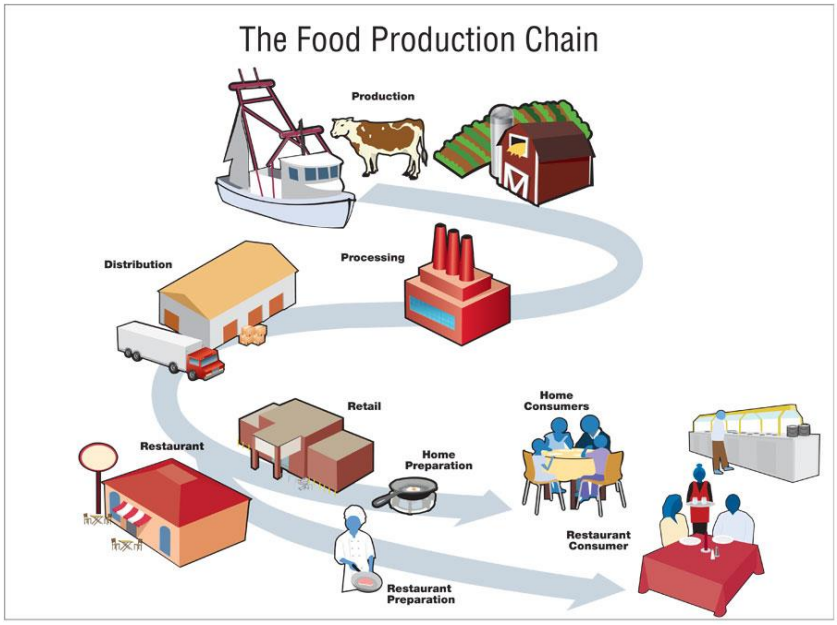


Figure 1 | Food Production Chain shows that food products contamination may happen at any time since the moment they are produced until they are finally consumed (The Food Production Chain, 2013)

Because of this daily increase of food contamination various security measures have been adopted in order to ensure the necessary food security. These measures passed by the creation of entities responsible for implementing effective programs of quality control, implementation of HACCP programs and above all the use of safer methods at all stages of the Food Production Chain. Even after all the implementation of these security measures, it is also the essential education of food handlers and consumers (Havelaar et al., 2010; Seaman, 2010).

1.1 Foodborne Pathogens

Demand for food is increasing at a global level however the huge diversity of food has led to an increase of the number of illnesses that are connected to the enormous variety of foodborne pathogens that exist today.

According to the data obtained in recent years there are several foodborne pathogens responsible for illnesses, hospitalizations and deaths. The best known pathogens are *Salmonella* spp., *Listeria monocytogenes*, *Campylobacter* spp., *Bacillus cereus*, *Staphylococcus aureus* and *Escherichia coli*. This type of contamination can occur by different ways such as contaminated food and water, infected peole, air, insects, pets and kitchen utensils (Teixeira et al., 2007).

1.1.1 Foodborne outbreaks in EU

In line with the European Union (EU) summary report on trends and sources of zoonoses, zoonotic agents and foodborne outbreaks in 2013, a total number of 5,196 foodborne outbreaks (including waterborne outbreaks) were reported in the EU. 43,183 human cases, 5,946 hospitalisations and 11 deaths were data obtained. 839 of outbreaks were connected to food vehicles such as eggs and egg products followed by mixed food, and fish and fish products (The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013, 2015).

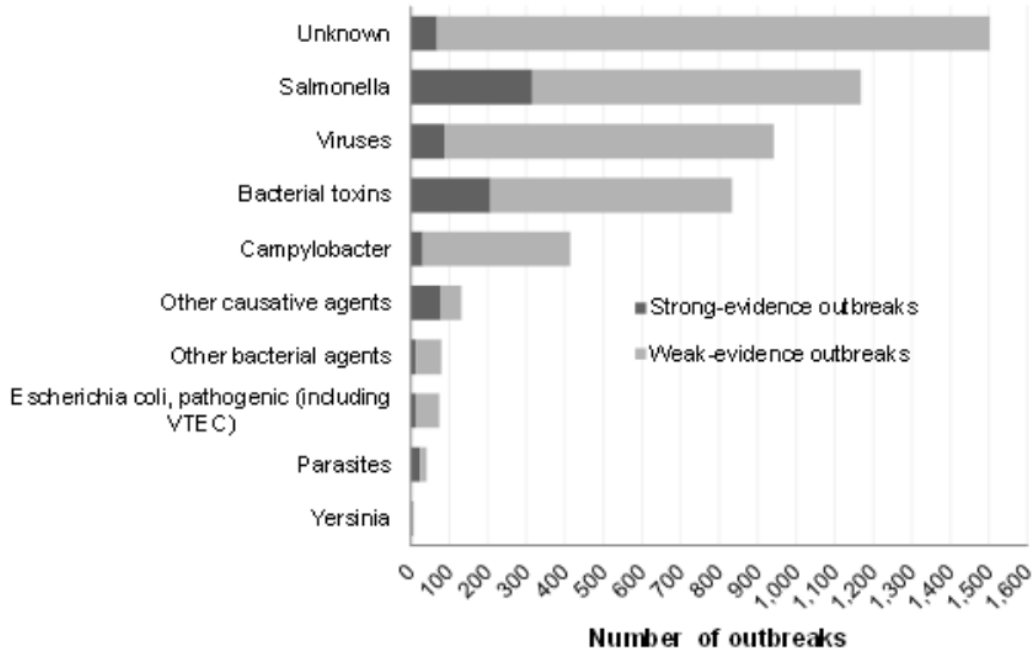


Figure 2 | Distribution of all foodborne outbreaks per causative agent in the EU, 2013. Adapted from "The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013".

Figure 2 shows that *Salmonella* caused the largest number of reported foodborne outbreaks (22.5% of all outbreaks), followed by viruses (18.1%), bacterial toxins (16.1%), and *Campylobacter* (8.0%). The culpable agent was not known for 28.9 % of the outbreaks (Food et al. 2015).

1.2 *Salmonella* Problems

Salmonella is one of the main foodborne pathogens worldwide, being responsible for disease outbreaks, hospitalizations and deaths (Centers for Disease Control and Prevention, 2013). It is one of the main virulent bacteria in domestic context, being this the reason why outbreaks of *Salmonella* scare so much the general public and, above all, the consumers. Of the 3000 annual deaths due to

pathogens transmitted by food in USA, according to US Center for Disease and Control, *Salmonella* is responsible for 38% of these deaths (Barbara et al., 2000). Symptoms of an infection by this microorganism are mainly fever, diarrhea and abdominal cramps (Westrell et al., 2009).

Salmonella is primarily found in poultry meat, pork and eggs. *Salmonella* is easily disseminated to hands, clothes and surfaces that are in contact with foods during preparation of meals (Cogan et al., 1999). These pathogens are capable of surviving on food contact surfaces during long periods of time (De Cesare et al., 2003; Redmond et al., 2004).

In 2013 it was reported a total number of 85,268 salmonellosis cases in EU. It shows a decrease of 7.9% when compared with values of 2012. The lowest rates were reported by Portugal and Greece and the highest notification rates in 2013 were reported by the Czech Republic and Slovakia, on the other hand the highest proportion of travel-related cases were linked to the Nordic countries, like Finland, Sweden and Norway (The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013, 2015).

Table 1 | Reported cases and notification rates per 100,000 of human salmonellosis in the EU/EEA, 2009-2013. Adapted from "The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013".

Country	2013					2012		2011		2010		2009	
	National Coverage	Data format	Total cases	Confirmed Cases & Rates		Confirmed Cases & Rates		Confirmed Cases & Rates		Confirmed Cases & Rates		Confirmed Cases & Rates	
				Cases	Rate	Cases	Rate	Cases	Rate	Cases	Rate	Cases	Rate
Portugal	Y	C	171	167	1.6	185	1.8	174	1.7	205	2.0	220	2.1
EU Total	-	-	85268	82694	20.4	90883	22.1	96682	20.9	101589	22.1	110179	24.0

*Y – yes. C – case-based data

1.2.1 *Salmonella* spp.

This organism is characterised to be Gram-negative, non-spore forming rod-shaped and flagellated facultative anaerobic bacterium, that can be classified by its O, H, and Vi antigens. It is a member of the family *Enterobacteriaceae* (Hocking, 2003) and the genus is divided in two species: *S. enterica* and *S. bongori* (Cianflone, 2008). The first is divided into six subspecies: *enterica*, *salamae*, *arizonae*, *diarizonae*, *houtenae* and *indica* (Pitt & Simpson, 2006).

In terms of growth and survival characteristics, the growth and survival of *Salmonella* spp. is affected by some factors like temperature, pH, water activity and preservatives (Podolak et al., 2010). *Salmonella* spp. presents a temperature range for growth of 5.1 – 46.2°C and an optimal temperature between 35 – 43°C. Heat resistance of these organisms in food is dependent on the composition, pH and water activity of the food. For pH they present a broad range where they can

grow and survive (3.8 – 9.5) nevertheless the optimum pH range for growth is 7 – 7.5. Water activity (a_w) also influences *Salmonella* spp. growth. The optimum value for a_w is 0.99 and the lower limit is 0.93. Presevatives have some impact in the growth of *Salmonella* spp., for example benzoic acid, sorbic acid or propionic acid can inhibit its growth (Hocking, 2003; Podolak et al., 2010).

2. BACTERIOPHAGES

Bacteriophages (phages) are viruses that specifically infect bacteria and are harmless to humans, animals and plants (Brown, 2010; Park & Nakai, 2003). They have been used for the prevention and control of bacterial pathogens. They are among the simplest and most abundant organisms on earth (it is estimated that there are 10 phages for each bacterial cell) (Park & Nakai, 2003).

Phage structure is very simple having mainly a DNA molecule (sometimes can also be an RNA molecule) that transport a certain number of genes that are responsible for phage replication and are protected by a protein capsid.

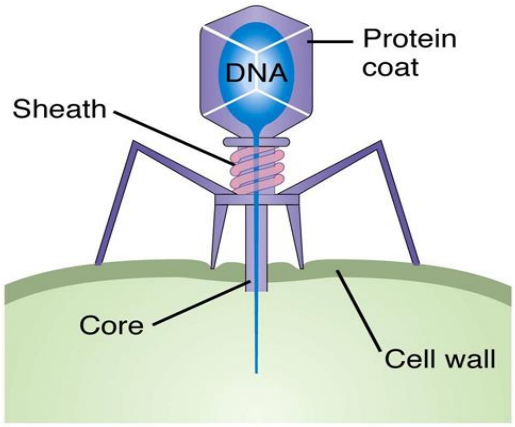


Figure 3 | Image of a typical bacteriophage. It can be seen a DNA molecule surrounded by a protein coat (Microbial Food Webs, 2010)

2.1 Bacteriophages Discovery

Bacteriophage discovery is not exactly recent and all information that is currently known is the result of several worldwide researches.

From Ernest Hanking observations in 1896 who noted the existence of high antibacterial activity against *Vibrio cholerae* and reported the presence of a substance responsible for the

decrease in cholera epidemics (Leverentz et al., 2001) to Frederick W. Twort's experiments with the varicella virus which led to a publication of an article in 1915 about the discovery of a virus capable of lysing bacterial cells (Levine, 1991). The 'real' discovery of phages however commonly is associated with d'Hérelle's work during an outbreak of hemorrhagic dysentery in 1915. D'Hérelle was the first to call this group of viruses bacteriophages (Leverentz et al., 2001). Then continuing his research he was able to isolate other phages for bacteria which cause diseases such as cholera, diphtheria, bubonic plague among others, and he was also responsible for developing the major virus quantitation method that allows the analysis of the phage replication cycle (Levine, 1991).

Associated with this theme, in 1969, Max Delbrück, Alfred Hershey and Salvador Luria won the Nobel Prize in Physiology and Medicine for their studies about phages, more specifically the discovery of the replication of viruses and their genetic structure (The Nobel Prize in Physiology or Medicine 1969, 2014).

2.2 Bacteriophages Classification

There are several types of phages, with some displaying greater specificity and others displaying a wider host range. They are obligatory parasites and can only replicate in host cells by controlling their cellular machinery. They are composed mainly of DNA - which contains genes that encode proteins involved in the replication of the phage - surrounded by a protein capsid (Brown, 2010). Phages can consist of double-stranded or single-stranded DNA or RNA (Ackermann, 2011).

Morphologically, phages can be tailed, polyhedral, filamentous and pleomorphic and there are 17 recognized families. Tailed phages belong to the order of *Caudovirales* and can be divided into three families: *Myoviridae*, *Siphoviridae*, and *Podoviridae* that vary in the characteristics of their tails: contractile, long and noncontractile, or short tails respectively. Tailed phages represent over 96% of all isolated phages (Ackermann, 2011).

2.3 Bacteriophages Infection Cycles

The overall pattern of infection cycle of a phage is the same for all tailed phage and consists in three steps: adsorption and penetration, viral eclipse and maturation phase, and extrusion (Figure 4). In the first step, phage particles connect to receptors located in the host cell surface and the

DNA is injected into the cell. Then, in the second phase, replication of phage DNA molecules occurs. Finally, new phage particles are accumulated and freed from the host through the action of enzymes holin and lysin (Brown, 2010).

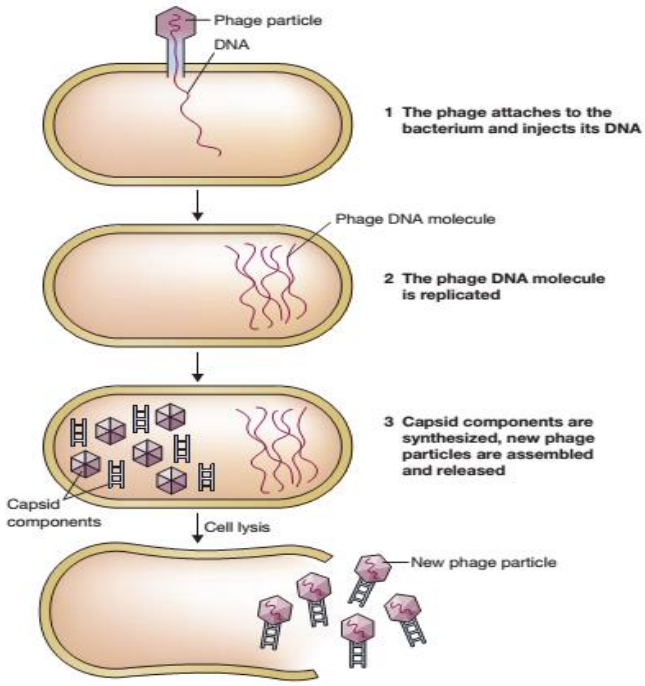


Figure 4 | The general pattern of infection of a bacterial cell by a bacteriophage (Brown, 2010)

There are two types of cycles that phages can adopt: the lytic cycle and the lysogenic cycle. The first is characterized by the insertion of the virus genetic material into the cell followed by replication and simultaneous synthesis of the proteins that constitute the capsid. After this, the viral particles are organized and cause cellular lysis, by which they are released (Brown, 2010). On the other hand, the lysogenic cycle is based on the incorporation of viral genome into the cell genetic material without cell death, thus being able of using intrinsic cell mitosis for its replication and ensuring its propagation to future generations. When the phage is in this state it is called prophage. By behaving in this manner, when phage proteins finally start to be translated (i.e. when the phage switches from the lysogenic cycle to the lytic cycle), their effect on bacterial population will be cell lysis (Brown, 2010).

2.4 Bacteriophages and their applications

The study of phage genomes has brought many advantages in the utilization of these viruses for a large number of applications. Today, we live in a time of scientific revolution where increasingly

faster and more efficient molecular methods arise. Therefore it is not surprising that phage genome sequencing has seen an increase in the last years (>1500 phage genomes have been already sequenced) (Genomes Pages – Phage, 2014). New sequencing methods are being developed in order to allow the control of pathogens and also study their impact on a global scale (WHO: Antimicrobial resistance, 2013).

The scientific community has been aware of phages' ability to kill antibiotic resistant bacteria for a long time and this explains the main importance of studying these entities nowadays (WHO: Antimicrobial resistance, 2013). However, in spite of these advantages, the use of phages at the level of food safety is dependent on various regulatory agencies since there are only a few phage-based products approved due to lack of scientific support through controlled clinical trials supervised by the competent authorities (The use and mode of action of bacteriophages in food production, 2009). The existence of several genes in the genome of phages with unknown functions worsens the situation since these genes may be virulence factors or toxin genes (Bru et al., 2004; Sillankorva et al., 2011; S. Sillankorva et al., 2012). Furthermore, bacteria that survive a phage infection can create new phenotypes with different receptors that phages do not recognize, and in this way evade the phage action and eventually survive (Pires et al., 2011). Despite the large genomic information available, routine genetic manipulation of phages falls only in the engineering of vectors used for phage display (Lu & Collins, 2009; Pande et al., 2010; Weber & Fussenegger, 2012).

Food products contamination may happen at any time since the moment they are produced until they finally are consumed (Lewis, 2008). Treatment of foods with phages has seen an increase, since they can help to prevent food deterioration as well as the spread of bacterial diseases. It is a sound strategy in both animal and vegetal food maintenance. Phages are therefore good tools that can be used for various purposes in the food industry such as treatment of infected animals, as prophylactic agents to prevent animal illness, food biocontrol agents, and biosanitizers for food contact surfaces.

2.4.1 Bacteriophage as biocontrol agents in food

It has been experimentally shown that phages are very effective in acting against growing bacteria but they lose effectiveness in nongrowing bacteria (Snyder & Champness, 2007). In the latter situation, control could be achieved by applying high concentrations of phages to fight pathogens by "lysis form without" (it is possible to control infection by administration of a high

phage dose that will overwhelm the pathogen and lyse the bacterial cells) (Berchieri et al., 1991; Goode et al., 2003).

The utilization of phages as biocontrol agents brings a number of advantages, such as high specificity to a certain host determined by bacteria cell wall receptors which allows fighting a specific microorganism while leaving other microbiota unharmed; self-replication and self-limiting, meaning low dosages are the minimum necessary to affect big bacteria populations and they will continue as long as there is a host present to allow virus reproduction; phages' capacity of adapting in response to defence mechanisms developed by bacteria; since they consist mostly in nucleic acids and proteins the risk to have contamination is too low (low inherent toxicity); cheap and easy to isolate and propagate; resistant to food processing and environmental stresses and long shelf life (Sillankorva et al., 2012). Along the food chain it is possible to use phages to promote food protection at three different phases that are as therapy, as biosanitizers of food contact surfaces and as biopreservation agents (Sillankorva et al., 2012).

2.4.2 Bacteriophage as therapy agents

The fact that bacterial pathogens have the capability to evolve resistance to antibiotics has led to their increasing ineffectiveness. Given this state of events, some small pharmaceutical companies and several laboratories worldwide are starting to focus on the use of phages as a viable alternative (Pirnay et al., 2011).

Nowadays, phage therapy is already used to promote food safety, but also used in animals against the most common foodborne pathogens to prevent and treat experimentally induced infections (Sillankorva et al., 2012). The capacity of isolating phages and finding phages for antibiotic resistant bacteria makes them the perfect candidates to be used in therapy (Pirnay et al., 2015) along with their specificity characteristics and lack of animal and plant toxicity. Furthermore, their production is simple, fast and inexpensive (Azeredo, 2008).

The use of phages as therapy agents presents some advantages such as: phages can be bactericidal, during the treatment phage can grow and raise their number, influence only a few bacteria from the normal flora, effective against antibiotic-sensitive and antibiotic-resistant bacteria, low inherent toxicities and are easily noticed. As disadvantages it is important to note that not all phages are good for therapeutics, some phages possess a small host range, and phages with lysogenic cycle are not accepted for therapeutic purposes (Loc-Carrillo & Abedon, 2011). Moreover,

one of the main weaknesses of phage therapy which receives apprehension of health authorities is the potential emergence of phage-resistant mutant bacteria and since they are viruses, the public opinion recognizes them as invaders by the immune system and create a response to be eliminated from systemic circulation (Teng-hern & Kok-gan, 2014).

It is therefore important to develop tests to guarantee public safety. So, to modern phage therapy evolve the main goal will be not to forget that before applying a product, validation and licence of competent authorities is needed to ensure quality and safety in the process (Pirnay et al., 2015).

2.4.3 Bacteriophage as biosanitizers of food contact surfaces and biopreservation agents

Phages can be used on food surfaces to reduce bacterial colonization of foods such as meats, seafood, milk, or processed foods (Atterbury et al., 2003; Hsu et al., 2002; Kennedy et al., 1986; Suárez et. al, 2002). The importance of phage use in equipment surfaces is further inflated due to pathogen´s capacity of adhering to inert surfaces and forming biofilms that allow them to persist on foods and food contact surfaces (Lewis, 2008). Biofilms on equipment are a common problem in food industry, especially in sites that are not easy to clean or sanitize. Phage use on biosanitation is promising as they have been shown to significantly reduce viable microbial cells in biofilms under ideal conditions, although it has to be noted that their specificity may limit their usefulness in this kind of environments where there is great bacterial diversity (S. M. Sillankorva et al., 2012).

Phages have been reported to lyse hosts at temperatures as low as 1°C (Greer, 1982, 1988), which makes them excellent as food biopreservation agents since they can limit bacterial growth on refrigerated foods (which is especially important when facing psychotropic bacteria) (Bigwood et al.,2008).

2.5 Bacteriophage infection of *Salmonella*

The use of bacteriophages when dealing with foodborne pathogens can be performed through two main strategies: postharvest and preharvest. Postharvest strategy is defined to be the usage of phages when animals are already dead in order to prevent the proliferation of pathogens. Preharvest strategy on the other hand relies on the administration of phages to live animals as a way to both

avoid illness and reduce pathogen presence in the gastrointestinal tract, thus diminishing the amount of pathogens that can later be present in food. Optimization of both strategies has been the main focus of investigation regarding the use of phages against foodborne pathogens (Sillankorva et al., 2012).

The majority of *Salmonella* outbreaks can be traced back to the *Salmonella enterica* serovars Enteritidis and Typhimurium. Eggs and poultry are the main source of the first, while pork and beef are the main source of the latter (EFSA, 2011). Under artificial laboratory conditions, both poultry and swine animals experimentally infected with *Salmonella* phages saw a significant decrease of the bacterium in major tissues, such as ileum and cecal tonsils. Despite this results, only studies on poultry farms can truly determine the success of this postharvest strategy in decreasing the risk of cross-contamination (Sillankorva et al., 2012).

The vast majority of recent *in vivo* experiments were carried out using cocktails of two to six phages (Table 2). The gastrointestinal tract of poultry is characterized by different pH values in distinct segments (2.5 in proventriculus/gizzard up to 8.0 in the colon). Particularly acidic conditions have a significant negative impact on phage viability (Sillankorva et al., 2012). In order to overcome this, Ma and colleagues (2008) have encapsulated phage Felix01 in chitosan-coated Ca-alginate spheres and discovered that in *in vitro* studies this approach is capable of increasing phage viability when exposed to lower pH. It should be noted, however, that *in vivo* experiments are lacking in regards to this encapsulated phages' resilience in such conditions (Ma et al., 2008).

Table 2 | Preharvest *Salmonella* phages application (Sillankorva et al., 2012)

Year	Animal/product	Phage(s)	Strategy	Main outcome
			Preharvest application	
2001	Poultry (chicken)	Phage cocktail	Oral delivery (direct and via feed)	Reduction of CFU in cecal counts between 0.3 and 1.3 log compared to controls birds
2001	Swine (pig)	Felix01	Oral delivery and i.m.	Reduction of CFU in the tonsils and cecum
2005	Poultry (broiler chickens)	CNPSA1, CNPSA3, CNPSA4	Oral delivery	Reduction of CFU by 3.5 orders of magnitude after five days
2005	Poultry (chickens)	Phage cocktail (Sa2, S9, S11)	Oral delivery phage/competitive exclusion	Reduction of CFU in cecum and ileum after phage cocktail and/or competitive exclusion treatment
2007	Poultry (broiler chickens)	Φ151, Φ25, Φ10	Oral delivery (antacid suspension)	Reduction of 4.2 log and 2.19 log with phages Φ151 and Φ25 within 24 h compared with control
2007	Poultry (broiler chickens)	Phage cocktail (CB4ϕ, WT45ϕ)	Oral delivery	Reduction of CFU in cecal tonsils after 24 h. No significant differences at 48 h compered to controls
2008	Poultry (chickens)	Phage cocktail	Oral delivery (coarse spray/drinking water)	Reduction of intestinal colonization of ten-day-old experimentally contaminated birds
2010	Swine (pig)	Phage cocktail		Reduction of colonization by 99.0 to 99.9% in the tonsils, ileum, and cecum
2011	Swine (weaned pigs)	Phage cocktail	Oral delivery	Significant reduction of CFU in the rectum
2011	Poultry (chickens)		Oral delivery (via feed)	Phage prevented horizontal transmission on six-week-old infected chickens

Regarding experiments involving postharvest strategies (Table 3), the use of a single phage as oppose to a cocktail has been the preferred choice. As of 2012, all *Salmonella* phages had been able to decrease the number of live pathogens present on processed and ready-to-eat foods, raw meats, and fresh produce (Guenther et al., 2012). Moreover, the combined use of phage and the *Salmonella* antagonistic bacteria *Enterobacter asburiae* is a remarkably auspicious approach as a way of decreasing the presence of *Salmonella* on mung bean sprouts, tomatoes, and alfalfa seeds without the need of chemicals (Guenther et al., 2012). In some cases, however, phages were found to become immobilized by the food matrix which rendered them unable to diffuse and consequently infect cells (Guenther et al., 2012).

Various studies performed shortly after the discovery of phages presented negative results due to improper use of phages, namely the treatment of viral and unknown agent diseases. With the increasing use of phages it becomes important not to repeat the same errors (Sillankorva et al., 2012). It is essential to take note of several parameters (e.g. concentrations and timings of application) upon administration of phages, since their effectiveness depends on the phage-host systems. When studying phage effectiveness, it is also important to faithfully simulate the conditions in which they are going to be used when in the industrial setting (Sillankorva et al., 2012).

The increase of phage-resistant phenotypes should also be always taken into account although it is possible to overcome this obstacle by using other phages which target these resistant phenotypes. Nevertheless, the full understanding of the resistance mechanisms obtained by the

hosts to the phages used and the percentage of disposal of the phages by the animal body are still scarcely studied (Sillankorva et al., 2012).

Table 3 | Postharvest *Salmonella* phage applications (Sillankorva et al., 2012)

			Postharvest application	
2001	Processed food (ripened cheese)	SJ2	Added to milk	No survival during 89 days in pasteurized cheeses containing phages (MOI 10^4)
2001	Fresh produce (fresh-cut melon and apple)	Phage cocktail	Added to foods	Significant CFU reduction on melon but not on apple
2003	Meat (chicken skin)	P22, 29C	Applied on top	MOI 1 caused less than 1 log reduction in CFU; MOI 100-1,000 caused 2 log reductions in CFU and eradicated resistant strains
2003	Meat (chicken frankfurters)	Felix O1		Approx. 2 log reduction with a MOI of 1.9×10^4
2004	Fresh produce (sprouting seeds)	A, B	Applied by immersion	Phage-A reduced CFU by 1.37 logs on mustard seeds. Cocktail resulted in a 1.5-log reduction in CFU in the soaking water of broccoli seeds
2005	Meat (broiler, turkey)	PHL 4	Sprayed	Phage treatments reduced frequency of *Salmonella* recovery as compared with controls
2008	Meat (raw/cooked beef)	P7	Applied on top	Reduction in CFU of 2-3 log at 5°C and approx. 6 log at 24°C
2009	Fresh produce (tomatoes)	Phage cocktail	Phage + *E. asburiae* JX1 added to food	Prevalence reduction of internalized *S. Javiana*, although the major suppressing effect was via antagonistic activity of *E. asburiae* JX1
2010	Fresh produce (mung bean sprouts and alfalfa seeds)	Phage cocktail	Phage + *E. asburiae* JX1 added to foods	Combined biocontrol with *E. asburiae* and phage suppressed pathogen growth on mung beans and alfalfa seeds
2011	Meat (pig skin)	Phage cocktail (PC1)	Applied on top	Above 99% reduction in CFU for MOI of 10 or above at 4°C for 96 h
2012	Ready-to-eat foods and chocolate milk	FO1-E2	Added to foods and mixed in milk	At 8°C no viable cells. At 15°C reduction of CFU by 5 logs on turkey deli meats and in chocolate milk and by 3 logs on hot dogs

3. OBJECTIVES

The main objective of this work was to analyse the genome sequence of phage vB_SenS_Φ38 to identify all unknown and non-essential genes. In order to do that, common bioinformatics tools currently used in genome annotation and analysis were used.

This task was divided in four parts: the first one was dedicated to the annotation of the genome through myRAST (automatic annotation software) and the identification of all genes. The second part focuses on the use of other bioinformatic tools for annotating transmembrane domains, promoters, terminators, tRNAs and also to perform comparative analysis of the phage vB_SenS_Φ38 genome with those of similar phages existing in the genome database. The third part will be focused on the expression of vB_SenS_Φ38 proteins in its host. Finally, in order to promote a more thorough evaluation and discussion of the discovered features, all identified hypothetical proteins will be analysed using HHpred; their effect on *Salmonella* growth will be determined; and proteins that are

shown to have no effect will be deleted from the genome through an *in silico* study performed with Geneious.

CHAPTER II

MATERIALS AND METHODS

1. BACTERIAL STRAINS AND PLASMIDS

In this study *E. coli* TOP 10 was used to keep the newly done plasmidic constructions and *Salmonella enterica* Serovar Enteritidis 821 was used to perform all the other assays. These two strains, which belong to the Department of Biological Engineering (DEB) of University of Minho, were grown on solid LB medium (400 ml of H_20, 10 g of LB and 6 g Agar) supplemented with 100 µg/ml ampicillin, or in liquid LB medium. They were cryopreserved at -80°C in glycerol 15% (v/v).

Table 2 | Microorganisms and vectors used in work development

Microorganisms	**Vectors**
E.coli TOP 10	pZE11G
Salmonella enterica Serovar Enteritidis 821	pZE11G

The studies of expression of proteins were carried out with pZE11G, a vector expressing the green fluorescent protein (GFP) (Figure 1).

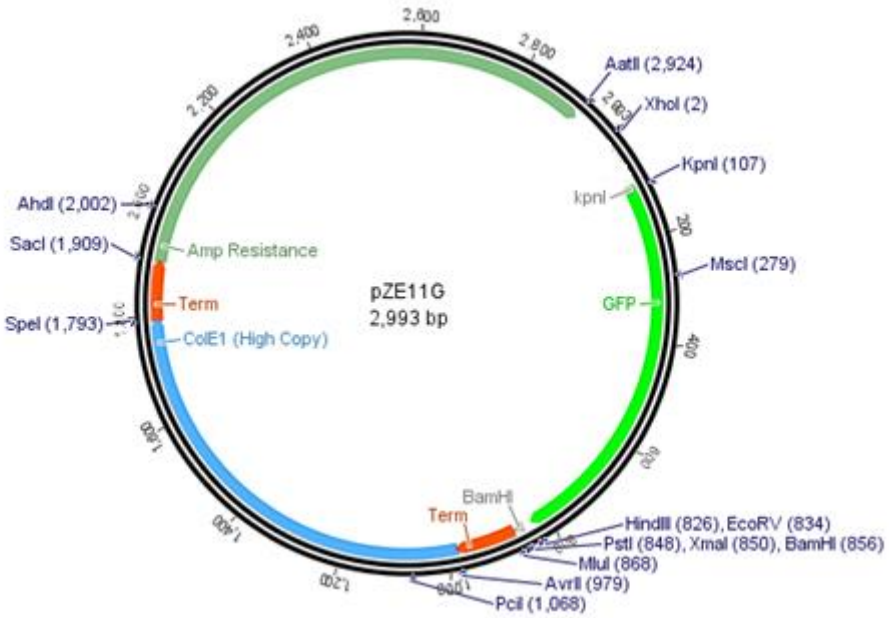


Figure 2 | Representation scheme of pZE11G cloning vector used to clone each of unknown gene sequences of *Salmonella* phage vB_SenS_Φ38

2. DNA ISOLATION AND GENOME SEQUENCING

To isolate genomic DNA from phage vB_SenS_Φ38, it was added 2.5 volumes of absolute ethanol and sodium actetate 3M to 500 µl of purified phage, being the DNA precipitated, air-dried and resuspended in MilliQ water (Sambrook, 2001) .Genome sequencing was performed on a 454 sequencing platform (pyrosequencing) (Plate-forme d'Analyses Genomiques at Laval University, Quebec, Canada) to 50-fold coverage. Sequence data was assembled using SeqMan NGen4 software (DNASTAR, Madison, WI, USA). Protein pI and molecular mass were predicted using Compute pI/Mw (Gasteiger et al., 2005).

3. PRIMERS CONSTRUCTION FOR CLONING

Primers presented in Table 2 were designed for amplification of each gene fragment with unknown function. Through OligoAnalizer 3.1 (OligoAnalyzer 3.1, 2014) it was possible to predict T_m°C, GC content and the presence of secondary structures like primer dimers, hairpins, self-dimers, hetero-dimers. Restriction enzymes are also described in Table 2 for each gene. *HindIII/BamHI* were used to most of the fragments, except for vB_SenS_CEB2_0052 and vB_SenS_CEB2_0054 fragments, in which *PstI/BamHI* were used.

Table 2 | Sequences of forward and reverse primers with restriction enzymes used. Where red nucleotides are *HindIII* restriction sequence and purple nucleotides are *BamHI* restriction sequence

Fragments	**Restriction Enzymes**	**Primer Fw**	**Primer Rv**
vB_SenS_CEB2_0007	*HindIII/BamHI*	5' GATC **AAGCTT** GTG AAG CGG GCG TCA CAA - 3' (59°C)	5' - GC **GGATCC** TTA AAC TCG TAT ATC TGT AAA CCT CAT ACG GCG - 3' (59°C)
vB_SenS_CEB2_0009	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG GCA TTG CCA AGT AGT GAT G - 3' (56°C)	5' - GC **GGATCC** CTA TTC CAG AAT ATT CCT TAT GTA CTC CTG - 3' (55°C)
vB_SenS_CEB2_0013	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG GCA GAT AAT TAT GTA GTA CGG GA - 3' (55°C)	5' - GC **GGATCC** TTA TTC CTC TAC ACT TTT ACG GCG - 3' (55°C)
vB_SenS_CEB2_0016	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG AGC ACC GCT TTT AGT AAA CG - 3' (56°C)	5' - GC **GGATCC** TCA TTT TCT CAC CTG AAT AAA GTA TGC C - 3' (55°C)

vB_SenS_CEB2_0028	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG GTT TCA TCC CCT CTT GTT G - 3' (55°C)	5' - GC **GGATCC** TTA CGG AAA TAC CGT AGT AAT ATT GAT GAG - 3' (55°C)
vB_SenS_CEB2_0029	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG ACT AAC AAA TAC AAT CGC ACA ATG - 3' (55°C)	5' - GC **GGATCC** TCA CTC ATC AAT ATT ACT ACG GTA TTT CC - 3' (55°C)
vB_SenS_CEB2_0030	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG GTC ACT CGC AAA ATA ACA GAT - 3' (55°C)	5' - GC **GGATCC** TCA TCC CAG CAT CTC CG - 3' (54°C)
vB_SenS_CEB2_0031	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG AGC GTT TTT ATC GGT ATT ATC GC - 3' (56°C)	5' - GC **GGATCC** TTA TTT ACC ATT AAC ATT AGA GTT ACA GC - 3' (55°C)
vB_SenS_CEB2_0032	*HindIII/BamHI*	5' - GATC **AAGCTT** GTG GTT CCC GCC TGG CG - 3' (63°C)	5' - GC **GGATCC** CTA CTT TTC CTG CAA TTG ACG CAA CTG TTC GTG - 3' (63°C)
vB_SenS_CEB2_0033	*HindIII/BamHI*	5' - GATC **AAGCTT** TTG TTT CAG TTT GCC CGT TG - 3' (54°C)	5' - GC **GGATCC** TCA GCC ATC GTT TCT ACC C - 3' (54°C)
vB_SenS_CEB2_0035	*HindIII/BamHI*	5' - GATC **AAGCTT** GTG TGG CGA GTA GAC CTG - 3' (55°C)	5' - GC **GGATCC** TCA TTG CAT TGG GTC CCA AAT T - 3' (55°C)
vB_SenS_CEB2_0037	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG AAT ATT AAT GAC TAC ACC GGT CTG - 3' (55°C)	5' - GC **GGATCC** CTA GCG CCA AAA TTC AAT CTC C - 3' (54°C)
vB_SenS_CEB2_0040	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG AGT AGC ATC GAA AAA GCT ATA GAA G - 3' (55°C)	5' - GC **GGATCC** TCA TTT AGC GCC CCT CTC - 3' (54°C)
vB_SenS_CEB2_0042	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG AAA CAC GAA TAT GAC CGC AAG - 3' (56°C)	5'- GC **GGATCC** TCA TTT CGC CAC CAG AAC C - 3' (56°C)
vB_SenS_CEB2_0045	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG GGA ATC AAA CTT AAT CTT CGT AAA G - 3' (54°C)	5' - GC **GGATCC** TTA ACG GTT ACG ACG GCG - 3' (55°C)
vB_SenS_CEB2_0046	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG ACT TTA AAA GCT AAA GAT AGG AGC G - 3' (55°C)	5' - GC **GGATCC** TCA AGC CTC TAA ATC GTC TTC AG - 3' (54°C)
vB_SenS_CEB2_0047	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG ATT GAC CAG GTG GGG GGG - 3' (62°C)	5' - GC **GGATCC** TCA GGC GAG CGC CGC TTC - 3' (63°C)
vB_SenS_CEB2_0048	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG ATG TTA GAA CAA TTT ATT AAA TTA TTT G - 3' (50°C)	5' - GC **GGATCC** TTA TTC TTC TTC GAA ATA CTT GTT C - 3' (50°C)

vB_SenS_CEB2_0052	*PstI/BamHI*	5' - GATC **CTGCAG** ATG ACG CAG AAT GAA GTA GCT - 3' (54°C)	5' - GC **GGATCC** TCA CTT CAC ATT CTC CCT AAT CC - 3' (54°C)
vB_SenS_CEB2_0053	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG AAA CCA AAT GAC CTC GTA ACC - 3' (56°C)	5' - GC **GGATCC** CTA TTT TAC GAT TAA TTT ATC CTG GCA CAT CA - 3' (56°C)
vB_SenS_CEB2_0054	*PstI/BamHI*	5' - GATC **CTGCAG** ATG ACT AGC GTA CTA TTC ATC TGG G - 3' (56°C)	5' - GC **GGATCC** TCA CTT CTC CTG TTT ACC ACC AA - 3' (56°C)
vB_SenS_CEB2_0055	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG AGT CTC GCA ACC GAT ATA C - 3' (54°)	5' - GC **GGATCC** TTA CCT GAA CGT ATA AAG TCG AAT AGA TG - 3' (55°C)
vB_SenS_CEB2_0056	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG ACT ACT ATA GCT TTT GAC GGT G - 3' (55°C)	5' - GC **GGATCC** TTA TCC TAG TCG ATG TGC TAT CAC T - 3' (55°C)
vB_SenS_CEB2_0059	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG AAC TTA CAA AGC GAT AAA GTT TTT TAC C - 3' (55°C)	5' - GC **GGATCC** TCA TTT CTT CTC TCC TGC ATA AGC - 3' (55°C)
vB_SenS_CEB2_0060	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG AAA GTA TAT ATC ATA TCT GGA TGG TAT TAC G - 3' (55°C)	5' - GC **GGATCC** TCA TTC CTG CAC CTC CC - 3' (54°C)
vB_SenS_CEB2_0061	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG AGC GAC AAC GGG CA - 3' (58°C)	5' - GC **GGATCC** TCA TTT CTT CTT GCC CCA CAT TCG - 3' (58°C)
vB_SenS_CEB2_0063	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG TCA CTA GCG TAT CGC C - 3' (54°C)	5' - GC **GGATCC** TTA GCA ATG CTC GTC TTT ATA AAC C - 3' (54°C)
vB_SenS_CEB2_0064	*HindIII/BamHI*	5' - GATC **AAGCTT** ATG GGC ACT AAA TTT GAA GTA ATA G - 3' (52°C)	5' - GC **GGATCC** TTA CAG TTC TAC AGA TGA TAA ATG GT - 3' (52°C)

4. BIOINFORMATICS

4.1 Genomic Properties

Phage vB_SenS_Φ38 genome was annotated using MyRAST (Aziz et al., 2008) and Glimmer (Salzberg, Deicher, Kasif, & White, 1998). MEME (Bailey et al., 2009) helped to identify putative promoter regions searching 150 bp upstream of each identified ORF followed by manual verification, and was complemented with PHIRE (Lavigne et al., 2004) and the promoter sequence logos were built with WebLogo (Crooks et al., 2004). ARAGORN (Laslett & Canback, 2004) and tRNAscan (Lowe & Eddy, 1997) were used to predict tRNAs. ARNold (Gautheret & Lambert, 2001; Macke et al.,

2001) was used to find rho-independent terminators and the folding structure of putative terminators and their secondary structures were calculated using Mfold (Zuker, 2003).

4.2 Compartive Genomics

BLASTN and BLASTX (Altschul et al., 1990; Programs available for the BLAST search, n.d.) were used to compare phage vB_SenS_Φ38 sequence with similar phages sequences deposited in database. Phage genome sequences resembling vB_SenS_Φ38 were *Salmonella* phage SETP7 (accession no. KF562865.1), *Salmonella* phage vB_SenS-Ent1 (accession no. HE775250.1), *Salmonella* phage SETP3 (accession no. EF177456.2) and *Salmonella* phage vB_SenS-Ent3 (accession no. HG934470.1). Using progressive MAUVE (Darling et al., 2004) and EMBOSS Stretcher (Rice et al., 2015) it was possible to compare DNA homology between vB_SenS_Φ38 genome and other phage genomes. With CG View (Grant & Stothard, 2008) a graphical map was built to compare all similar sequences against vB_SenS_Φ38 genome and ClustalW2 (Larkin et al., 2007) was used to perform progressive alignments based on phylogenetic tree for phage DNA or proteins.

4.3 Protein Analysis

Transmembrane domains were projected using TMHMM (TMHMM Server v. 2.0 – Prediction of transmembrane helices in proteins, n.d.) and Phobius (Käll et al., 2007). vB_SenS_Φ38 protein sequences were queried against non-redundant protein databases using HHpred (Söding et al., 2005). The sequences of predicted ORFs and potential alternative start codons were checked by BLASTP (Altschul et al., 1990; Programs available for the BLAST search, n.d.).

4.4 Deletion of Hypothetical Proteins

Geneious is a bioinformatic tool that allows manipulation of genomic sequences. This tool reads genomic maps, allows manipulation of repeated regions and is able to perform comparative analysis of the phage genomes. In this study, it was used to deletion of each hypothetical protein without antimicrobial properties and to comparison of the initial phage genome with the phage genome obtained after each deletion (Geneious, 2015).

5. TRANSFORMATION

5.1 Electrocomptent cells

To produce electrocompetent cells an inoculum of one colony from a fresh plate of the strain *E.coli* TOP 10 was done in LB broth supplemented with 100 μg/ml of ampicillin and incubated at 37°C, 120 rpm agitation (Environmental Shaker incubator ES- 20/60), overnight. Then, a dilution of 1:100 from the overnight culture was made in fresh LB medium and incubated at the same conditions of temperature and rotation until the culture reached OD_{600nm}≈0.5. The culture was transferred to 2 falcons of 50ml and kept on ice for 20 min. Then, a centrifugation step was made (7000 x g, 10min, 4°C). This step was repeated three times however the pellet was ressuspend in different solutions and quantities and the supernatant was always discarded. The first pellet was ressupended in 50 ml of cold sterile H_2O_d, the second pellet was resuspended in 8 ml of cold sterile glycerol 10% solution and the last on in 1 ml of sterile glycerol 10% (v/v). Aliquots of 80 μl were stored at -80°C. Along this process, ice was always used to prevent thermal shock and to maintain the efficiency of competent cells.

The procedure used to make *S.* Enteritidis 821 electrocompetent cells was the same as the one adopted for *E. coli* TOP 10.

5.2 Electroporation

To *E. coli* TOP 10 cells, 2 μl of ligation (section 7.c below) was used. This step was performed using Gene Pulser Xcell (Bio-Rad) with 1 mm cuvettes at 1800 V, 25 μF and 200 Ω. The ligation was added to 80 μl of chemicompetent cells, mixed, and added to a cuvette. After an electric pulse the cuvette was removed and 800 μl of LB was added. The final solution was passed to an eppendorf and incubated at 37°C with 120 rpm agitation for 1 hour for recovery and cellular growth. Finally, the cells were plated in LB plates supplemented with ampicillin.

To *Salmonella* cells 1 μl of plasmid construction previously done was used. The rest of the protocol is similar to *E. coli* TOP 10 cells.

6. DNA PHAGE EXTRACTION

Before cloning, it was necessary to digest the vector and also the amplified fragments that code for proteins. DNA phage Extraction Protocol was used. Phage DNA extraction was carried out with 500 µl of phage ($1.2x10^{12}$ pfu/ml) according to Sambrook (Sambrook, 2013). To analyse the integrity of DNA, an agarose gel electrophoresis was performed (1% (w/v)) and concentration measured with NanoDrop 1000™ (Thermo Scientific). Phage DNA stock was stored at -18 ˚C.

7. CLONING

Phage DNA was used to amplify each gene fragment. Specific primers were used for each fragment. All reagents used in the PCR are described in Table 3 and the conditions presented in Table 4.

Table 3 | PCR components for amplification of each fragment. The volumes were calculated for 50 µl of final reaction volume

Components	**50 µl**
Phusion DNA Polymerase	0.5 µl
5x Phusion Buffer	10 µl
10 mM dNTPs	1 µl
H_2O	36.5 µl
10 mM Fw	0.5 µl
10 mM Rv	0.5 µl
DNA	1 µl

Table 4 | PCR Program conditions for amplification of each fragment using Phusion DNA Polymerase

Step	**Temperature**	**Time**
Initial denaturation	98˚C	30 sec
Denaturation, primer annealing, extension (34 cycles)	98˚C	10 sec
	***T_mºC + 3ºC**	15 sec
	72˚C	30 sec
Final Extension	72˚C	5 min
	4˚C	∞

*Annealing temperature was 3°C higher than the melting temperature of the primers, since this is the recommended by the manufacturer (New England Biolabs).

An agarose gel was run, the DNA fragment was excised and each fragment was cleaned through Zymoclean Gel DNA Recovery Kit (ZymoPURE™).

The vector pZE11G and the gene fragments were double digested with either *HindIII/BamHI* or *PstI/BamHI*. 0.5 µl of each enzyme and 2 µl of Cut Smart buffer (10x) were used in a final volume of 20 µl.

T4 DNA Ligase (New England Biolabs) was used to catalyse the ligation between the digested pZE11G and each fragment that code for proteins of unknowns function described in Table 3.

Colony PCR was used to confirm correct insertion of the fragments in the vector. From a transformation plate, colonies were selected, picked into 50 µl of sterile H_20. From this, 1 µl was pipeted into a PCR tube and then PCR master mix components described in Table 5 were added. Amplification was performed as described in Table 6.

Table 5 | Colony PCR Components used

Components	**25 µl**
Kapa Taq DNA Polymerase	0.1 µl
Buffer 10x	2.5 µl
10 mM dNTPs	0.5 µl
H_2O	18.9 µl
10 mM Fw	1 µl
10 mM Rv	1 µl
DNA	1 µl

Table 6 | PCR Program Steps

Step	**Temperature**	**Time**
Denaturation	95˚C	3 min
Denaturation, primer annealing, extension (34 cycles)	95˚C	30 sec
	***T_mºC – 5ºC**	30 sec
	72˚C	30 sec
Final Extension	72˚C	5 min
	4˚C	∞

***** Annealing temperature was 5°C lower than than the melting temperature of the primers, since this is the recommended by the manufacturer (KapaBiosystems).

The correct sizes of the amplified products was checked by electrophoresis in 1% or 2% (w/v) agarose gels depending on the size of the PCR product. SYBR Safe (Invitrogen) was used as DNA stain for visualization of DNA and the agarose gel visualized in ChemicDoc XRS (BioRad).

After Colony PCR, each positive clone was grown overnight in 20 ml of LB supplemented with ampicillin (100 μg/ml) at 37°C, 120 rpm. Plasmid DNA was isolated with the Plasmid Miniprep Kit (ZymoPURE™). In the end, to ensure integrity of DNA an electrophoresis was performed and the concentration of plasmid measured in NanoDrop 1000™ (Thermo Scientific). 700 μl of the overnight culture was added to 300 μl of glycerol, to a final concentration of 15% (v/v) of glycerol and the sample stored at -80°C.

S. Enteritidis 821 transformed cells were prepared according to section 5.2 and after overnight growth on 20 ml LB supplemented with ampicillin (100 μg/ml), cultures were cryopreserved at -80°C (700 μl of culture and 300 μl of glycerol, to a final concentration of 15% (v/v) of glycerol).

8. EXPRESSION OF PHAGE PROTEINS

To evaluate the effect of each hypothetical protein on the growth of *S.* Enteritidis 821, two methodologies were used (solid and liquid media) and two controls were used: a negative control (*S.* Enteritidis 821 cells with pZE11G) and a positive control (*S.* Enteritidis 821 cells with a pZE11G expressing an endolysin).

The induction of protein expression was accomplished by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG). Expression of phage proteins was performed both in solid and liquid media.

8.1 Solid Medium

Overnight cultures of the transformed *S.* Enteritidis 821 cells carrying each hypothetical protein were spotted (5 μl) on petri dishes with: LB, LB + Amp (100 μg/ml) and LB + Amp (100 μg/ml) + IPTG (2mM). In petri dishes with IPTG were used different dilutions of the strain (overnight

culture, dilution 1:100 and dilution 1:10000 (Figure 2)). After the spots were completely dried, they were incubated at 37°C overnight.

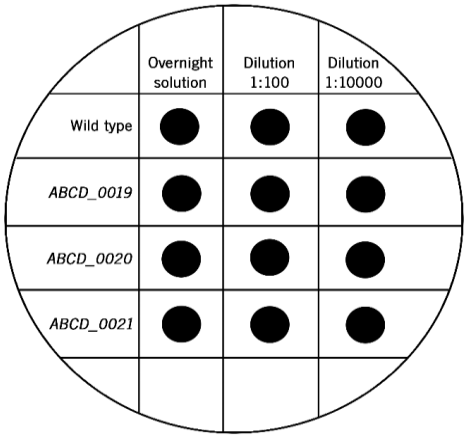


Figure 2 | Petri dish scheme used to analyse expression of phage proteins (Simões, 2015)

8.2 Liquid Medium

A dilution of an overnight culture of the transformed *S.* Enteritidis 821 cells carrying each hypothetical protein were introduced in three different medium: LB, LB + Amp (100 μg/ml) and LB + Amp (100 μg/ml) + IPTG (2mM). From each overnight inoculum 100 μl were pipetted and then added to the different media already done (5 ml). The cultures were incubated at 37°C (120 rpm) and to determine if there was a decrease in cell mass, a sample (100 μl) was hourly taken for eight hours and the optical density measured (OD_{620nm}).

CHAPTER III

RESULTS AND DISCUSSION

1. GENOMIC PROPERTIES

vB_SenS_Φ38 has a linear dsDNA genome of 42 439 bp, with a GC content of 49.9%, which is lower than 52% that was described for serovars of *S.enterica* (McClelland et al., 2001; Thomson et al., 2008). The phage vB_SenS_Φ38 genome was run in myRAST and Glimmer was used to annotate automatically the genes. Glimmer was used as an additional tool to search for potential genes which myRAST could have missed. Crossing the results of both programs (Attachment I and Attachment II), 60 ORFs were found. The putative proteins were blasted against InterPro database and 32 were found to correspond to proteins with known function and other 28 to hypothetical proteins. Using Compute pI/Mw, the theoretical isoelectric point (pI) and molecular weight (Mw) of the phage (Gasteiger et al., 2005) were determined and resulted in a pI of 4.22 and a Mw of 3522895.31.

Promoter presence was searched with three different programs: PHIRE, MEME and WebLogo. The first one was used to find their possible localization. In order to confirm PHIRE results, MEME was used to complete this information, which was further manually verified. Candidate promoter sequences were identified by searches of 150 bp sequences upstream of ORFs and 5 possible promoters were identified. Then WebLogo was used to build the promoter consensus sequence (Figure 1). Each logo consists in a stack of symbols, one letter for each position in the sequence and the higher the height of the letter, higher will be its relative frequency (Crooks et al., 2004). The promoter consensus sequence obtained was **AxAxATAATxxxxxxCxxTTx**.

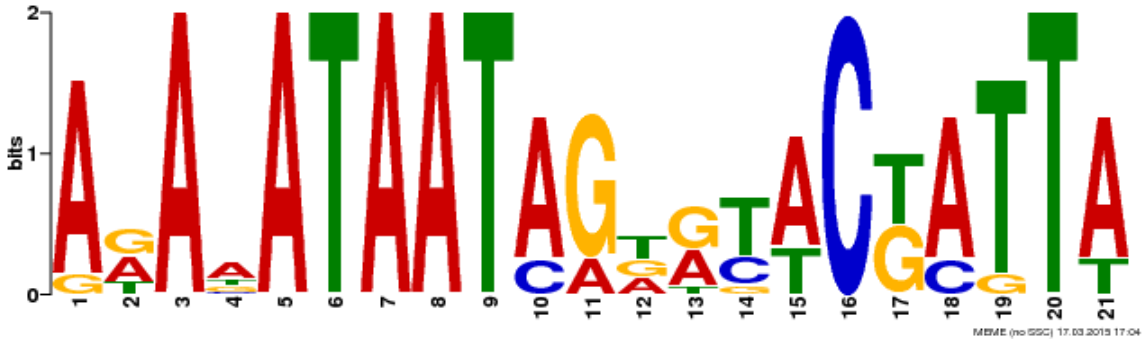


Figure 3 | Weblogo multiple sequence alignment of consensus sequence

The search for tRNAs with ARAGORN and tRNA-scan showed that no tRNAs were found in the genome sequence of phage vB_SenS_Φ38.

With the help of TMHMM and Phobius the presence of transmembrane domains was verified and resulted in no transmembrane domain found for phage vB_SenS_Φ38.

At last, ARNold was used to find rho-independent terminators in nucleic acid sequences and Mfold to complete the analysis of terminators presence. ARNold predicted that there were 28 transcription terminators and after this all the sequences with a score of free energy value less than -9 were excluded. The remaining sequences were analysed using Mfold to assure that only terminators with a loop were chosen. A cross-check of terminators with the genome sequence was carried out to understand the terminators position in the sequence and only the better positioned terminators were chosen. At the end of this analysis, only 6 terminators were found in the genome.

2. COMPARATIVE GENOMICS

The genome alignment was based on the BLASTN search results and the remaining conclusions were confirmed by a progressive Mauve alignment. With BLASTN it was possible to find the most similar phage genome sequences to the phage in study. Four genome sequences with homologies ≥ 90% are apresented in Table 1.

Table 2 | NCBI sequences of phages with more homology with *Salmonella* phage vB_SenS_Φ38

Sequences	Identity	Query Cover
Salmonella phage vB_SenS-Ent1	94%	91%
Salmonella phage vB_SenS-Ent3	93%	90%
Salmonella phage SETP3	93%	88%
Salmonella phage SETP7	90%	90%

Using progressive Mauve, all genomes were compared, and even though the genomes of the selected phages are all similar, there are regions which lack homology (Figure 2, white areas). Not only a higher homology between different phage genome sequences is verified but it can also be seen that fragments of different phages are aligned. These results were expected since BLASTN

showed a great proximity between phage sequences and *Salmonella* phage vB_SenS_Φ38 genome sequence (90-94% identical).

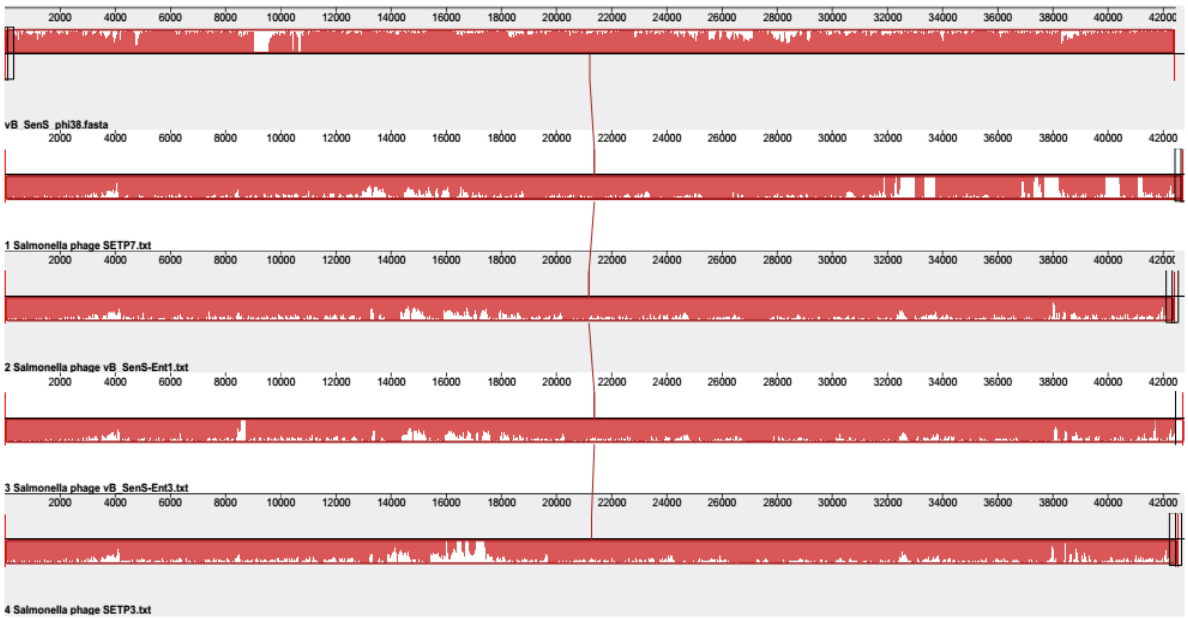


Figure 4 | Mauve results showing genome sequence comparison

Emboss Stretcher was also used to align the sequences and the results show that the genome sequence of *Salmonella* phage SETP7 is 47.5% identical at nucleotide level to *Salmonella* phage vB_SenS_Φ38, followed by *Salmonella* phage SEPT3 with 46.3% homology, *Salmonella* phage vB_SenS-Ent1 with 46.1% and *Salmonella* phage vB_SenS-Ent3 with 45.9% (Table 2).

Table 3 | Alignment results of Emboss Stretcher with sequences with higher homology to *Salmonella* phage vB_SenS_Φ38

	***Salmonella* phage SETP7**	***Salmonella* phage vB_SenS-Ent1**	***Salmonella* phage SETP3**	***Salmonella* phage vB_SenS-Ent3**
***Salmonella* phage vB_SenS_Φ38**	Identity: 47.5% Similarity: 47.5% Gaps: 13.6%	Identity: 46.1% Similarity: 46.1% Gaps: 12.7%	Identity: 46.3% Similarity: 46.3% Gaps: 12.9%	Identity: 45.9% Similarity: 45.9% Gaps: 12.7%

A filogenetic tree was built with ClustalW2. Through analysis of Attachment III, it can be verified that *Salmonella* phage SETP7 is the phage that presents a higher homolgy with *Salmonella* phage vB_SenS_Φ38. This result was expected since they share a common ancestor. These two phages are paralogs since they derive from a unique species that suffered a duplication event, originating two new species. Aditionally, the remaining phages come from different ancestors, so they present a

bigger distance relative to *Salmonella* phage vB_SenS_Φ38. *Salmonella* phage SETP3 is a special case because it is an ortholog, since it is from a different evolutionary descent.

ClustalW2 also allowed to obtain a table that compares not only *Salmonella* phage vB_SenS_Φ38 with the other phage genome sequences but also acomparison between them (**Table 3**).

Table 3 | Score values comparison between different phage sequences. 90-100% dark blue, 70-90% blue and 50-70% light blue

	***Salmonella* phage vB_SenS_Φ38**	***Salmonella* phage SETP7**	***Salmonella* phage vB_SenS-Ent1**	***Salmonella* phage SETP3**	***Salmonella* phage vB_SenS-Ent3**
***Salmonella* phage vB_SenS_Φ38**	100%	51.52%	50.68%	51.05%	51.01%
***Salmonella* phage SETP7**	51.52%	100%	84.15%	83.07%	82.87%
***Salmonella* phage vB_SenS-Ent1**	50.68%	84.15%	100%	88.09%	98.2%
***Salmonella* phage SETP3**	51.05%	83.07%	88.09%	100%	87.48%
***Salmonella* phage vB_SenS-Ent3**	51.01%	82.87%	98.2%	87.48%	100%

Analyzing the results it can be verified that score values obtained for phage vB_SenS_Φ38 are among 50.68-51.52 % of identity and similarity, yet, when other phage genomes are compared to each other, score values vary between 82.87-98.2 %. So, it can be said that these phage genomes have more similarity between them than between *Salmonella* phage vB_SenS_Φ38.

Analysis of phage vB_SenS_Φ38 genome was also done using CG View, even though the genome of the phage is not circular but rather linear. Through the graphical map obtained (Attachment IV) it can be seen that there are genes that are similar between all phage genomes. Once again, some sequences of the phage genome have a lot of similarity and others no similarity at all. This is something that was expected since previous tools already warned for the existence of similarities and differences.

In summary, the results obtained by comparative genomics are not conclusive. BLASTN suggested that *Salmonella* phage vB_SenS-Ent1 genome was the one with greater similarity to phage vB_SenS_Φ38 genome; however, Emboss Stretcher and ClustalW2 determined *Salmonella* phage SETP7 to be the most identical genome to vB_SenS_Φ38 genome. Based on these outcomes the determination of each phage protein function is very important in order to ensure the results obtained are more conclusive, and that comparative genomics studies of unknown genomes is easier to perform. Recently sequenced genomes have had a constant rise in the number of hypothetical genes present. This state of affairs has made it more difficult to describe all hypothetical proteins as the most sizable chunk of these do not emerge in the databases (Hatfull, 2012).

BLASTX was used as a complementary accessory tool for the analysis of the results provided by BLASTP. Both programs were used to search each ORF, and also to predict functions (*e value* threshold of $<1 \times 10^{-5}$). Of this 60 ORFs, about 32 encode for proteins with known function and 28 for proteins with unknown function. A more thorough analysis of these proteins was completed using HHPred. With Pfam, for protein motif search, an *e value* threshold of <1x10-5 was always used. The main obtained results are illustrated in Attachment V. As mentioned previously, some proteins had known functions, however 28 proteins had unknown function. In certain cases, when no significant resemblances were found, conserved domains or protein families were searched for. With the *e value* obtained, some homologues of the 28 proteins were obtained. Most of them were hypothetical proteins from other *Salmonella* phages, although some of the results obtained were putative proteins or not possible to find the best homologue. Then, Pfam database enabled to go farther and obtain a possible function to some proteins, but still some of proteins presented unknown function. From these results stands out a phage lysozyme (CDS 7 with a *e value* of 1.3e-33), a proteasome subunit (CDS 10 with a *e value* of 1.8e-20), some proteins responsible for tail components (CDS 32 with a *e value* of 4.9e-10 and CDS 43 with a *e value* of 9.4e-15), gp6 and gp20 (CDS 57 with a *e value* of 4.1e-29 and CDS 51 with a *e value* of 2.3e-07, respectively). In the end, only a few protein functions were predicted, and with BLASTP and HHpred no virulence genes were found, which is important for choosing a good candidate for later studies in phage therapy (Loc-Carrillo & Abedon, 2011).

After this in-depth analysis, 32 proteins had their functions predicted and the others are examples of proteins with unknown functions (no match in databases). It was not possible to detect proteins with antimicrobial activity, although it is probable that phage vB_SenS_Φ38 expresses

proteins with potential antimicrobial action. However, with the current knowledge of phage genomes, the prediction that can be done is dependent on finding the function of each protein. It is, therefore, imperative to complete protein databases, to avoid problems like this in the future.

3. DELETION OF HYPOTHETICAL PROTEINS

Through software Geneious it was possible to visualize the complete genome sequence of phage vB_SenS_Φ38 with all ORFs that constitute its genome (Figure 3).

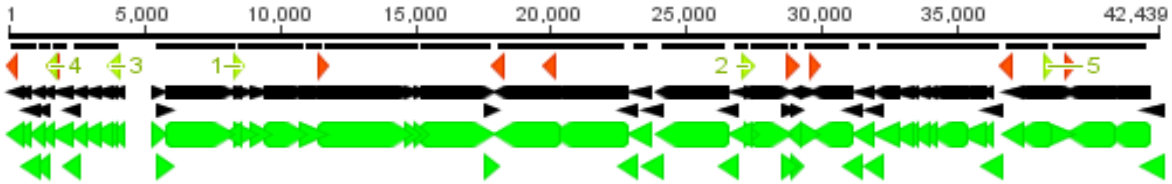


Figure 3 | Phage genome sequence with all ORFs highlight. The arrows represent the ORFs and point in the direction of transcription

For this task Geneious was used to delete all genes that were not essential for the phage infectivity. Based on the results previously obtained with other bioinformatic tools, it was kwown that none of the hypothetical proteins that were studied presented antimicrobial activity. Hence, they were all *in silico* deleted from the genome sequence. Deletion of each gene was performed while being mindful of some their caractherisctis like the position of the gene (if it was in the middle of two genes and if its deletion would halt their functionality), his orientation, and the presence of terminators or promoters dependent on their presence to operate. After a detailed analysis and taking into account the bioinformatic results previously obtained, the deletion of the 28 genes that encode for unknown proteins was carried out. The obtained genome has a size of 30 337 bp (Figure 4) which is a small size, allowing a faster replication, an advantage in the infection process.

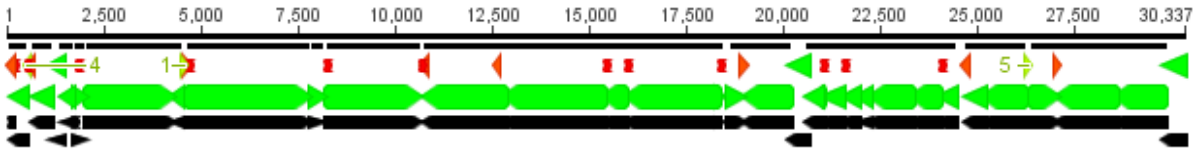


Figure 4 | Phage genome sequence with ORFs and without hypothetical proteins. The arrows represent the ORFs and point in the direction of transcription

4. EXPRESSION OF PHAGE PROTEINS

Phages usually start the establishment of suitable conditions for their survival right from the point the infection process begins (Labrie et al., 2010). This is achieved by anticipating the bacterial defence mechanisms in certain ways, such as producing proteins capable of hindering these same mechanisms or modifying bacterial metabolic pathways to the phage's own betterment (Häuser et al., 2012) in this way generating optimal surroundings for phage proliferation (Wagemans et al., 2014).

In order to determine the effect of each hypothetical protein in the growth of *S.* Enteritidis*,* the genes coding for the 28 hypothetical proteins were cloned in the pZE11G vector in *E. coli* TOP 10. After having all the plasmid constructions, these were transformed into *S.* Enteritidis 821. At this point two controls were done: a positive control (*S.* Enteritidis 821 expressing, from pZE11G, an endolysine, which was expected to effect *S.* Enteritidis growth) and negative control (*S.* Enteritidis 821 cells containing an empty pZE11G). Induction of protein expression was accomplished by using IPTG inductor. This was done not only to verify the importance of each hypothetical protein but also to study their effect on bacterial growth. In the end, the effect of the hypothetical proteins on bacterial growth was compared to the control strain.

The growth of the 30 strains was done in solid and liquid media. In both cases, 3 different media were used: LB, LB + Amp (100 μg/ml) and LB + Amp (100 μg/ml) + IPTG (2mM).

The results obtained in both tests were the same. All used strains showed a normal growth when compared to control strain. No clear difference could be observed in all media, even in those that were supplemented with IPTG. So, it can be suggested that the growth of *Salmonella* was not influenced by expression of the hypothetical phage proteins, given that in the presence or absence of IPTG the results were the same.

A solution for this problem would be changing vector used. Vector pET is a system developed for cloning and expression of proteins in *E. coli* and it is very well studied and developed, unlike the pZE11G vector (Studier et al., 1990). Other factor is the bacterial host. In this study, at first *E. coli* was used and only then *Salmonella*. The vector used in this study consists in a vector which was developed especifically to *E. coli* which can explain why the results obtained were not the expected. Also, some proteins may require the presence of other phage proteins in order to be expressed or to

display antibacterian activity. Since, in this study, all proteins were studied separately, that might be the reason why no growth inhibition was seen in the end.

During the process, some mutations can occur and have effects on the protein sequence, leading to protein inactivation or even the loss of its expression.

In summary, simple changes in protocol used can have a big effect in results obtained. Unfortunately, it was impossible to find a protein that showed antimicrobial activity for later used in studies, like the construction of a 'new' phage that can effectively combat *Salmonella*.

CHAPTER IV

CONCLUSIONS AND FUTURE APPROACHES

This work had three main objectives. In first place, the genome sequence of *Salmonella* phage vB_SenS_Φ38 was analysed using bioinformatic tools, so all unknown gene sequences could be identified. The second objective of this work was to find proteins that showed antimicrobial properties. Finally, to engineer the genome sequence of phage vB_SenS_Φ38 in order to build a compact, safe and infective phage was also one of the aims of this thesis.

The study of the genomic properties, *in silico*, of phage vB_SenS_Φ38 revealed the presence of 60 ORFs, where 32 were found to correspond to proteins with known function and 28 to hypothetical proteins. Then, 5 possible promoter sequences were identified and the promoter consensus sequence achieved was AxAxATAATxxxxxxCxxTTx. 6 terminators were found and no tRNAs or transmembrane domains could be detected. Unfortunately, comparative genomic studies were not conclusive. In one case, doing a BLASTN *Salmonella* phage vB_SenS-Ent1 showed to be the phage more similar to phage vB_SenS_Φ38. On the other hand, using the Emboss Stretcher and ClustalW2, *Salmonella* phage SETP7 genome showed greater similarity to phage vB_SenS_Φ38. HHpred allowed to obtain a more detailed analysis of the studied proteins. However, of the 28 hypothetical proteins, none seemed to have antimicrobial activity. Using Geneious it was possible to preview how the genome of phage vB_SenS_Φ38 would be after deleting all the genes coding for the 28 hypothetical proteins. In the end, this 'new' phage vB_SenS_Φ38 genome presented a size of 30 337 bp.

The expression of these 28 hypothetical phage proteins was tested using solid and liquid media. Although it was possible to express almost all proteins, the obtained results were inconclusive. For none of the *S.* Enteritidis strains tested, each one expressing a different hypothetical protein, was observed a difference in growth when compared to the strain containing only the empty vector. This could mean that none of these proteins have an antimicrobial activity against *S.* Enteritidis. However, protein expression was not confirmed, for example, by running a SDS-PAGE gel, so it is not possible to make this conclusion. Hence, it becomes crucial to perform further studies in order to confirm that these hypothetical proteins are being expressed and have no antimicrobial activity against *S.* Enteritidis. Future approaches will be focused in the deletion of the phage vB_SenS_Φ38 hypothetical proteins which have no effect on the growth of *Salmonella* and on the infectivity capacity of the phage. To do that, a genome modification of phage vB_SenS_Φ38 will be done. Basically, the modification consists in a deletion of the selected genes that code for the

hypothetical proteins from the genome which has been successfully captured in a Bacterial Artificial Chromossome (BAC) (unpublished data). The deletions will be done sequentially, since after each deletion, using recombineering techniques, it is important to check the viability of the engineered genome by electroporation of the BAC into the bacterial host. By the end of this task, it is expected to have deleted the majority of the hypothetical proteins present, having a very robust and compact end phage genome. The non-essential genes will be also deleted. In each step, phage viability will be checked to make sure it remains active. In the end, it will be always done an effectiveness test of the "new" phage against the original phage, to know if the presence of only one antibacterial protein is better for the phage. Other interesting future approaches may involve the investigation of the nature of toxicity of each protein using time-lapse microscopy to monitor cell growth in the presence of IPTG and the impact on bacterial expression of each protein.

Unfortunately, since it was not possible to conclude if any of the 28 hypothetical phage proteins have any antimicrobial activity against *S.* Enteritidis, the objective of deleting the non-essential genes in the genome of phage vB_SenS_Φ38 was not accomplished.

In the end of this work, it was expected to obtain a phage with a robust genome, free of non-essential genes while maintaining its viability, so it can be used in applications against *Salmonella*.

REFERENCES

Ackermann, H.-W. (2011). Bacteriophage taxonomy. *Viruses of Prokaryotes. Tom I*, 90–94.

Altschul, S., Gish, W., & Miller, W. (1990). Basic Local Alignment Search Tool. *J Mol Biol.*, *215*(3), 403–410.

Atterbury, R. J., Connerton, P. L., Christine, E. R., Rees, C. E. D., Connerton, I. F., & Dodd, C. E. R. (2003). Application of Host-Specific Bacteriophages to the Surface of Chicken Skin Leads to a Reduction in Recovery of *Campylobacter jejuni*. *Applied and Environmental Microbiology*, *69*(10), 6302–6306.

Aziz, R. K., Bartels, D., Best, A. a, DeJongh, M., Disz, T., Edwards, R. a, ... Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, *9*(1), 75.

Bailey, T. L., Boden, M., Buske, F. a, Frith, M., Grant, C. E., Clementi, L., ... Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, *37*(Web Server issue), 1–7.

Barbara, G., Stanghellini, V., Berti-Ceroni, C., De Giorgio, R., Salvioli, B., Corradi, F., ... Corinaldesi, R. (2000). Role of antibiotic therapy on long-term germ excretion in faeces and digestive symptoms after *Salmonella* infection. *Alimentary Pharmacology & Therapeutics*, *14*(9), 1127–1131.

Berchieri, A., Lovell, M. A., & Barrow, P. A. (1991). The activity in the chicken alimentary tract of bacteriophages lytic for *Salmonella* Typhimurium. *Research in Microbiology*, *142*(5), 541–549.

Bigwood, T., Hudson, J. a, Billington, C., Carey-Smith, G. V, & Heinemann, J. a. (2008). Phage inactivation of foodborne pathogens on cooked and raw meat. *Food Microbiology*, *25*(2), 400–406.

Brook, S. (2013). Molecular cloning: A laboratoy manual (4th ed., Vol. 33, pp. 75–78).

Brown, B. (2010). Gene Cloning and DNA Analysis: An Introduction (6th editio, pp. 17–19).

Bru, H., Canchaya, C., & Hardt, W. (2004). Phages and the Evolution of Bacterial Pathogens : from Genomic Rearrangements to Lysogenic Conversion, *68*(3), 560–602.

Centers for Disease Control and Prevention. (2013). Tracking and Reporting Foodborne Disease Outbreaks. Retrieved from http://www.cdc.gov/features/dsfoodborneoutbreaks/

Cianflone, N. F. C. (2008). Salmonellosis and the GI Tract: More than Just Peanut Butter. *Current Gastroenterology Reports*, *10*(4), 424–431.

Cogan, T. A., Bloomfield, S. F., & Humphrey, T. J. (1999). The effectiveness of hygiene procedures for prevention of cross-contamination from chicken carcases in the domestic kitchen. *Letters in Applied Microbiology*, *29*(5), 354–358.

Crooks, G., Hon, G., Chandonia, J., & Brenner, S. (2004). WebLogo: a sequence logo generator. *Genome Res*, *14*, 1188–1190.

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research* , *14* (7), 1394–1403.

De Cesare A, Sheldon B, Smith K, J. L. (2003). Survival and Persistence of *Campylobacter* and *Salmonella* Species under Various Organic Loads on Food Contact Surfaces. *Journal of Food Protection*, *66*, 1587–1594.

EFSA. (2011). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2009. *The European Food Safety Authority Journal*, *9*(3), 2090.

G. G. Greer. (1982). Psychrotrophic bacteriophages for beef spoilage *Pseudomonas*. *Journal of Food Protection*, *45*, 1318–1325.

G. G. Greer. (1988). Effect of phage concentration, bacterial density, and temperature on phage control of beef spoilage. *Journal of Food Science*, *53*, 1226–1227.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M., Appel, R., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In J. Walker (Ed.), *The Proteomics Protocols Handbook SE - 52* (pp. 571–607). Humana Press.

Gautheret, D., & Lambert, A. (2001). Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *Journal of Molecular Biology*, *313*(5), 1003–1011.

Geneious. (2015). Retrieved July 15, 2015, from http://www.geneious.com/

Genomes Pages – Phage. (2014). Retrieved from http://www.ebi.ac.uk/genomes/phage.html

Goode, D., Allen, V. M., & Barrow, P. A. (2003). Reduction of Experimental *Salmonella* and *Campylobacter* Contamination of Chicken Skin by Application of Lytic Bacteriophages. *Applied and Environmental Microbiology*, *69*(8), 5032–5036.

Grant, J. R., & Stothard, P. (2008). The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Research*, *36*(Web Server issue), W181–W184.

Green, M. R., & Sambrook, J. (n.d.). *Molecular cloning : a laboratory manual*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.

Guenther, S., Herzig, O., Fieseler, L., Klumpp, J., & Loessner, M. J. (2012). Biocontrol of *Salmonella* Typhimurium in RTE foods with the virulent bacteriophage FO1-E2. *International Journal of Food Microbiology*, *154*(1-2), 66–72.

Guntupalli, R., Hu, J., Lakshmanan, R. S., Huang, T. S., Barbaree, J. M., & Chin, B. A. (2007). A magnetoelastic resonance biosensor immobilized with polyclonal antibody for the detection of *Salmonella* Typhimurium. *Biosensors and Bioelectronics*, *22*(7), 1474–1479.

Hatfull, G. F. (2012). Chapter 7 - The Secret Lives of Mycobacteriophages. In M. Ł. and W. T. S. B. T.-A. in V. Research (Ed.), *Bacteriophages, Part A* (Vol. Volume 82, pp. 179–288). Academic Press.

Häuser, R., Blasche, S., Dokland, T., Haggård-Ljungquist, E., von Brunn, A., Salas, M., … Uetz, P. (2012). *Bacteriophage protein-protein interactions. Advances in virus research* (Vol. 83).

Havelaar, A. H., Brul, S., de Jong, A., de Jonge, R., Zwietering, M. H., & ter Kuile, B. H. (2010). Future challenges to microbial food safety. *International Journal of Food Microbiology*, *139, Suppl*, S79–S94.

Hocking, A. D. (2003). *Foodborne microorganisms of public health significance / Editor in Chief Ailsa D. Hocking*. (6th ed.).

Holah, J. T., Taylor, J. H., Dawson, D. J., & Hall, K. E. (2002). Biocide use in the food industry and the disinfectant resistance of persistent strains of Listeria monocytogenes and *Escherichia coli*. *Journal of Applied Microbiology*. Campden & Chorleywood Food Research Association, Chipping Campden, UK.

Hsu, Fc, Shieh, Ysc, Sobsey, M. (2002). Enteric bacteriophages as potential fecal indicators in ground beef and poultry meat. *Journal of Food Protection*, *65*(1), 93–99.

J. E. Kennedy, C. I. Wei, and J. L. O. (1986). Distribution of coliphages in various foods. *Journal of Food Protection*, *49*(12), 944–951.

J.Azeredo, I. W. S. (2008). The Use of Phages for the Removal of Infectious Biofilms. *Current Pharmaceutical Biotechnology*, *9*, 261–266.

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Research*, *35*(Web Server issue), W429–W432.

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*, *8*(5), 317–327.

Larkin, M. a., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. a., McWilliam, H., ... Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947–2948.

Laslett, D., & Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, *32*(1), 11–16.

Lavigne, R., Sun, W. D., & Volckaert, G. (2004). PHIRE, a deterministic approach to reveal regulatory elements in bacteriophage genomes. *Bioinformatics*, *20*(5), 629–635.

Leverentz, B., Conway, W. S., Alavidze, Z., Janisiewicz, W. J., Fuchs, Y., Camp, M. J., ... Sulakvelidze, a. (2001). Examination of bacteriophage as a biocontrol method for salmonella on fresh-cut fruit: a model study. *Journal of Food Protection*, *64*(8), 1116–1121.

Levine, A. J. (1991). *Viruses: A Scientific American Library Book* (1st ed.). New York: Scientific American Library: Distributed by W.H. Freeman and Co.

Lewis, K. (2008). Multidrug tolerance of biofilms and persister cells. *Current Topics in Microbiology and Immunology*, *322*, 107–131.

Loc-Carrillo, C., & Abedon, S. T. (2011). Pros and cons of phage therapy. *Bacteriophage*, *1*(2), 111–114.

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*(5), 955–964.

Lu, T. K., & Collins, J. J. (2009). Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(12), 4629–4634.

Ma, Y., Pacan, J. C., Wang, Q., Xu, Y., Huang, X., Korenevsky, A., & Sabour, P. M. (2008). Microencapsulation of bacteriophage Felix O1 into chitosan-alginate microspheres for oral delivery. *Applied and Environmental Microbiology*, *74*(15), 4799–4805.

Macke, T. J., Ecker, D. J., Gutell, R. R., Gautheret, D., Case, D. A., & Sampath, R. (2001). RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Research* , *29* (22), 4724–4735.

Maukonen, J., Mättö, J., Wirtanen, G., Raaska, L., Mattila-Sandholm, T., & Saarela, M. (2003). Methodologies for the characterization of microbes in industrial environments: a review. *Journal of Industrial Microbiology & Biotechnology*, *30*(6), 327–356.

McClelland, M., Sanderson, K. E., Spieth, J., Clifton, S. W., Latreille, P., Courtney, L., ... Wilson, R. K. (2001). Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature*, *413*(6858), 852–856.

Microbial Food Webs. (2010). Retrieved August 30, 2015, from http://oceanworld.tamu.edu/resources/oceanography-book/microbialweb.htm

OligoAnalyzer 3.1. (2014). Retrieved September 15, 2014, from http://eu.idtdna.com/calc/analyzer

Oliveira K, Oliveira T, Teixeira P, Azeredo J, O. R. (2007). Factors involved in attachment of *Salmonella* Enteritidis, *Pseudomonas aeruginosa* and *Serratia* marcescens to stainless steel. *Brazilian Journal of Microbiology*, *38*, 318–323.

Pande, J., Szewczyk, M. M., & Grover, A. K. (2010). Phage display: concept, innovations, applications and future. *Biotechnology Advances*, *28*(6), 849–58.

Park, S. C., & Nakai, T. (2003). Bacteriophage control of *Pseudomonas plecoglossicida* infection in ayu Plecoglossus altivelis. *Dis Aquat Organ*, *53*, 33–39.

Pires, D., Sillankorva, S., Faustino, A., & Azeredo, J. (2011). Use of newly isolated phages for control of *Pseudomonas aeruginosa* PAO1 and ATCC 10145 biofilms. *Research in Microbiology*, *162*(8), 798–806.

Pirnay, J. P., De Vos, D., Verbeken, G., Merabishvili, M., Chanishvili, N., Vaneechoutte, M., ... Adamia, R. (2011). The phage therapy paradigm: Prêt-à-porter or sur-mesure? *Pharmaceutical Research*, *28*, 934–937.

Pirnay, J.-P., Blasdel, B. G., Bretaudeau, L., Buckling, A., Chanishvili, N., Clark, J. R., ... Van den Eede, G. (2015). Quality and Safety Requirements for Sustainable Phage Therapy Products. *Pharmaceutical Research*, 2173–2179.

Pitt, T. L., & Simpson, A. J. H. (2006). *Principles and Practice of Clinical Bacteriology, Second Edition*.

Podolak, Richard and Enache, Elena and Stone, Warren and Black, Darryl G. and Elliott, P. H. (2010). Sources and Risk Factors for Contamination, Survival, Persistence, and Heat Resistance of *Salmonella* in Low-Moisture Foods. *Journal of Food Protection*, *73*(10), 1919–1936.

Programs available for the BLAST search. (n.d.). Retrieved June 11, 2015, from http://www.ncbi.nlm.nih.gov/BLAST/blast_program.shtml

Redmond, E. C., Griffith, C. J., Slader, J., & Humphrey, T. J. (2004). Microbiological and observational analysis of cross contamination risks during domestic food preparation. *British Food Journal*, *106*(8), 581–597.

Rice, P., Longden, I., & Bleasby, A. (2015). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, *16*(6), 276–277.

Rocourt, J., Moy, G., Vierk, K., & Schlundt, J. (2003). The present state of foodborne disease in OECD countries. *Public Health*.

Salzberg, S. L., Deicher, A. L., Kasif, S., & White, O. (1998). Microbial gene identification using interpolated Markov models. *Nucleic Acids Research*, *26*(2), 544–548.

Seaman, P. (2010). Food hygiene training: Introducing the Food Hygiene Training Model. *Food Control*, *21*(4), 381–387.

Sillankorva, S., Kluskens, L. D., Lingohr, E. J., Kropinski, A. M., Neubauer, P., & Azeredo, J. (2011). Complete genome sequence of the lytic *Pseudomonas fluorescens* phage ϕIBB-PF7A. *Virology Journal*, *8*(1), 142.

Sillankorva, S., Kropinski, A. M., & Azeredo, J. (2012). Genome sequence of the broad-host-range *Pseudomonas* phage Φ-S1. *Journal of Virology*, *86*(18).

Sillankorva, S. M., Oliveira, H., & Azeredo, J. (2012). Bacteriophages and their role in food safety. *International Journal of Microbiology*, *2012*.

Simões, C. (2015). *Construction of an innocuous Salmonella phage*. University of Minhio.

Snyder, L., & Champness, W. (2007). *Lytic bacteriophages: genetic analysis and use in transduction*. *Molecular genetics of bacteria* (Ed.3 ed.). Washington, D.C: ASM Press.

Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, *33*(Web Server issue), W244–W248.

Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J. W. (1990). Methods Enzymol (pp. 60–89).

Teixeira, P., Silva, S. C., Araújo, F., Azeredo, J., & Oliveira, R. (2007). *Bacterial adhesion to food contacting surfaces*.

Teng-hern, T. L., & Kok-gan, C. (2014). Application of Bacteriophage in Biocontrol of Major Foodborne Bacterial Pathogens, *1*(1), 1–9.

The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013. (2015), *13*(1).

The Food Production Chain. (2013). Retrieved May 7, 2014, from http://www.cdc.gov/foodsafety/outbreaks/investigating-outbreaks/figure_food_production.html

"The Nobel Prize in Physiology or Medicine 1969." (2014). Retrieved September 1, 2015, from http://www.nobelprize.org/nobel_prizes/medicine/laureates/1969/index.html

The use and mode of action of bacteriophages in food production. (2009). Retrieved June 15, 2014, from http://www.efsa.europa.eu/en/efsajournal/pub/1076

Thomson, N. R., Clayton, D. J., Windhorst, D., Vernikos, G., Davidson, S., Churcher, C., ... Parkhill, J. (2008). Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Research*, *18*(10), 1624–1637.

TMHMM Server v. 2.0 – Prediction of transmembrane helices in proteins. (n.d.). Retrieved from http://www.cbs.dtu.dk/services/TMHMM-2.0/

V. B. Suárez, A. Quiberoni, A. G. Binetti, and J. A. R. (2002). Thermophilic lactic acid bacteria phages isolated from Argentinian dairy industries. *Journal of Food Protection*, *65*(10), 1597–1604.

Wagemans, J., Blasdel, B. G., Van den Bossche, A., Uytterhoeven, B., De Smet, J., Paeshuyse, J., ... Lavigne, R. (2014). Functional elucidation of antibacterial phage ORFans targeting *Pseudomonas aeruginosa*. *Cellular Microbiology*, *16*(12), 1822–1835.

Weber, W., & Fussenegger, M. (2012). Emerging biomedical applications of synthetic biology. *Nat Rev Genet*, *13*(1), 21–35.

Westrell T, Ciampa N, Boelaert F, Helwigh B, Korsgaard H, Chríel M, Ammon A, M. P. (2009). Zoonotic infections in Europe in 2007: a summary of the EFSA-ECDC annual report, *14*(3).

WHO: Antimicrobial resistance. (2013). Retrieved from http://www.who.int/mediacentre/factsheets/fs194/en/

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, *31*(13), 3406–3415.

ATTACHMENTS

ATTACHMENT I – "FRED" FILE RESULT OBTAINED WITH MYRAST

LOCUS vB_SenS_phi38 42439 bp DNA linear BCT 04-MAR-2015
DEFINITION Salmonella phage vB_SenS_phi38
ACCESSION Unknown Unknown
VERSION Unknown
KEYWORDS WGS.
SOURCE Salmonella phage vB_SenS_phi38
ORGANISM Salmonella phage vB_SenS_phi38
Salmonella phage vB_SenS_phi38
REFERENCE 1 (bases 1 to 42439)
AUTHORS [Insert Names here]
TITLE Direct Submission
JOURNAL [Insert paper submission information here]
COMMENT [Insert project information here]
FEATURES Location/Qualifiers
source 1..42439
/organism="Salmonella phage vB_SenS_phi38"
/mol_type="genomic DNA"~
/strain="vB_SenS_phi38"
/db_xref="taxon:6666667"
gene complement(95..325)
/gene="orf"
CDS complement(95..325)
/locus_tag="ABCD_0001"
/product="hypothetical protein"
/translation="MGTKFEVIDENTVTVTEEGSHYILIDKVNGGWSAKAYFNGGAIV
VSKGVHMSFTSAYDSLKDAVKSAANHLSSVEL"
gene complement(342..641) /locus_tag="ABCD_0002"
CDS complement(342..641) /locus_tag="ABCD_0002"
/product="hypothetical protein"
/translation="MNDFLILMGIGFTTLLLAVVIYLMWILFLWPFVEAVSLTRMSLA
YRRLNAHKIGAATAVKLFVMWYGETVFGRRFCAIHSRGWRWEAVGKWWVYKDEHC"
gene complement(638..994) /locus_tag="ABCD_0003"
CDS complement(638..994) /locus_tag="ABCD_0003"
/product="hypothetical protein"
/translation="MTNNEYEKMMVEAANGRFKIEAVDAQIEAHQAVPDYLNEHRREL
ITSIPKLIETHGTLADTCRETGLNEMTLSKYRHDTKCEQHVIYNNRLMTHTKTSPVIY
TKRGVSRNDRMKLEGL"
gene complement(991..1146) /locus_tag="ABCD_0004"
CDS complement(991..1146) /locus_tag="ABCD_0004"
/product="hypothetical protein"
/translation="MSDNGQVTVTLKVGGKVGNTPFPTREELLKRNSFPGPDKNKYLN
RMWGKKK"
gene complement(1143..1301) /locus_tag="ABCD_0005"
CDS complement(1143..1301) /locus_tag="ABCD_0005"
/product="hypothetical protein"
/translation="MKVYIISGWYYYDGADILGVYLSEEKAEAAAELERAKGDYDGVE
VREWEVQE"
gene complement(1298..1483) /locus_tag="ABCD_0006"
CDS complement(1298..1483) /locus_tag="ABCD_0006"

/product="hypothetical protein"
/translation="MNLQSDKVFYHCKPVLDTEAFEDAQLMAHIAVNNLSSNIPADAF
WAAAMQALKAAYAGEKK"
gene complement(1667..2155) /locus_tag="ABCD_0007"
CDS complement(1667..2155) /locus_tag="ABCD_0007"
/product="hypothetical protein"
/translation="MSNRNISNNGIKFTAAFEGFRGTAYRATKNEKYLTIGYGHYGAD
VKEGQKITEGQGLLLLNRDMVKAVAAVDAVAHPSLNQSQFDAMCDLVYNAGAGVIAAS
TGTGHALRKGDVATLRNKLTQFHYQNGKSLLGLRRRAAGRVALFDGMLWQQAEAIGRG
AK"
gene complement(2136..2423) /locus_tag="ABCD_0008"
CDS complement(2136..2423) /locus_tag="ABCD_0008"
/product="hypothetical protein"
/translation="MLDSPIGVICLAAIIVTSLINIYAHWIEDGLFGRLLYMASVVTA
AGGLAHLFTGGIPPFIITTLVAIFALKSVRHICVKGARYYKYRRMYVKPKH"
gene complement(2425..2706) /locus_tag="ABCD_0009"
CDS complement(2425..2706) /locus_tag="ABCD_0009"
/product="hypothetical protein"
/translation="MDDKYLWLSVAGLAGGAVSQIKKREAISPWLRLCHLVASACCAV
YASPIIISYYELSQSEGQYLVPFGVGMFWLKLFEAADSSLSNFKLPWGK"
gene complement(2785..3219) /locus_tag="ABCD_0010"
CDS complement(2785..3219) /locus_tag="ABCD_0010"
/product="hypothetical protein"
/translation="MTTIAFDGETMACDTCVTGNFKYYTDTKIYENDHFVMGVSGDAG
VGRLLVVDAEILTPKYYDFDFSALVFVKEDNRIFRVEFFKSWDSPLSSVIPIAGNAAA
VGSGAPYALTAMFMGATATAAV0+EIAKRFDPNTDGGVIAHRLG"
gene complement(3225..3602) /locus_tag="ABCD_0011"
CDS complement(3225..3602) /locus_tag="ABCD_0011"
/product="hypothetical protein"
/translation="MSLATDILKHAGINLAPPSSAVTKKVTRVKEKVKVKVKRKPKPR
AKPVNEMPDVYPRIPGVHQPKYCVGKGLWRAHSYDGKKVVNLGEFSSQARAHMAVKLY
KLWRKRGYSDIPHKPSIRLYTFR"
gene complement(3605..3808) /locus_tag="ABCD_0012"
CDS complement(3605..3808) /locus_tag="ABCD_0012"
/product="hypothetical protein"
/translation="MTSVLFIWVLTAGQMQLAASETFYTLEACQTAARAVENAHFLFQ
GDKPIDSEVRAICSPKRLGKQEK"
gene complement(3871..4035) /locus_tag="ABCD_0013"
CDS complement(3871..4035) /locus_tag="ABCD_0013"
/product="hypothetical protein"
/translation="MKPNDLVTWTGRNGGTRHGKVTSLHGIYARVERWRAHAKKPRYF
LMCQDKLIVK"
gene 5476..5646 /locus_tag="ABCD_0014"
CDS 5476..5646 /locus_tag="ABCD_0014"
/product="hypothetical protein"
/translation="MTQNEVAKLIGVTRRTLNNWLRDGKFPDCCVRIMGRRMPGTFDR
EKVEAWIRENVK"
gene 5643..5876 /locus_tag="ABCD_0015"
CDS 5643..5876 /locus_tag="ABCD_0015"
/product="hypothetical protein"
/translation="MTDFQKRVFRAVAALTFKKGSCSVLDLRRTYFQYYSSSLIENTL
KVLVKDGAVKSKAGKYSAVAEVKCSQATPEDLE"
gene 5933..8119 /locus_tag="ABCD_0016"
CDS 5933..8119 /locus_tag="ABCD_0016"

/product="hypothetical protein"
/translation="MSTGNLAYSWSRSNRREERADINVYTTTSGTWDEFVELMQPLKR
SRRNPKTDPGYITAACTATVSSTGKEAAEGMFYRCNASVTSSSLAYADVDSATPEEFA
TDCEMVRESRFAMMLYTTASHTEEAPRYRVVMPVRTPVTGGDIIRIRYGLLTHFLKGR
DVDSAGFTLSQPMYRPPVGSQVIVSESSRMITASKLMEGVPEINVTGASDYRVPEGEQ
SELTDLFEEFAFEFGGRMTDRGLQMPATPEHAAQYTTGEPKQDDFLFCWPRDGFERPN
VTLYHDTDLVATGGMKPGGRDMWAYACAATGLPFDRVEAALGWAEGVTCDEEDLADEE
PPAPQADFIVEGYMPTNCIWDIVGESGTYKSFYTLGMMYISAAGYRFAGADTRKAHHF
YIDGEGGEFTHTRIAALAAKYGDEGMRYMHVLDAGEFADTKKLVRKIRQIAGSEPVGM
VAFDTLNQTFGNWIDKFNENSAGQDGMGRVVAILKEVRDGTKGAVGVVHHTPKGGSKA
RGSGALYAGVDVELTLVRATEKQINVAHTKNKNGMQQKTIGMVLEPVQFREAPPPKEF
QAVEFVGGEGYGEIVNLDLPEPHKALVLMPWGFQPFETDEEKERNEGLDSKGKDSVKD
TVKRSKDASARESVMSALEDLQQADDTGRGFTQRQIVARAGDHSITNLVLEKMLREGE
LMLGCDENGEVVTNTYRLPTGIDDRKRPKNRYEPNDNIKTTEGDLE"
gene complement(8134..8352) /locus_tag="ABCD_0017"
CDS complement(8134..8352) /locus_tag="ABCD_0017"
/product="hypothetical protein"
/translation="MQQSELGARVERRRKEIGMGQTELAFKAGVSQSLITHLATGRVS
KVDCFKIFHIADALGVDPRWLSFGDTGA"
gene 8483..9010 /locus_tag="ABCD_0018"
CDS 8483..9010 /locus_tag="ABCD_0018"
/product="hypothetical protein"
/translation="MMLEQFIKLFERFVVAHELIAANSAKQTVYQLKSDITAVNKELI
EKSVETAAPDEYRASSAVCKASEKELPVEGEDVVDTKPAEEEKPKRKTRKAKVEEPEE
EKEEIDYQSLRDQIQAIDDAINEGPSDAACDDSDELLEEFTGKKMKIAAIKDEDLAEY
LERLTAIKNKYFEEE"
gene 9054..9536 /locus_tag="ABCD_0019"
CDS 9054..9536 /locus_tag="ABCD_0019"
/product="hypothetical protein"
/translation="MIDQVGGGPGMNWHELFNYDPDTGKLFWKISPSQAVKAGDEAGT
LCGGGRSCDKKYLRVCYRRKFKQVHHVVWEMFNGPLKDGEEIDHINHIRTDNRIENLR
KVYRGDNAKNRSMPINNTSGVHGVSYFSRIDMWVARIGVNGKSIFLGRYNNLEEAALA
"
gene 9556..10839 /locus_tag="ABCD_0020"
CDS 9556..10839 /locus_tag="ABCD_0020"
/product="hypothetical protein"
/translation="MGFIKITELRRMTLKAKDRSGSNDVHALLSPSGAKKWLTCPASL
ICEKDIPNTSGKAAVLGTAMHSLSEYHLNAYIRGTALPLERDVGAYVLEEGKGAVKAL
IKPMKGAVLVTDDMVEQVRKYTDYCKAIIDVSTYTKLEMRVNLTEVLHPGYEGVETFG
TADFVAVQELANTDKHMLIIGDLKTGRHRVEAKENKQLMLYALGVYRRLKRRYNITTV
RLVIFQPYAGGASEWDISVEGLELFAKFAQKRAVAALDAYSRGKKNLKASDFKPSVEG
CQWCRFSEQCAARTKTVNAVLAQELEDDFALELTPEQLVAEYEKLPLLRQHIDKVEKA
MFAALHSGKKVPGYKLVEGRPGSRTWKDAKKVASKYGIALMKQVLLTPTEAIKVVPEE
ELKDFITRKPGAPCVTTVEDKRPEWKNVTEDDLEA"
gene 10921..11547 /locus_tag="ABCD_0021"
CDS 10921..11547 /locus_tag="ABCD_0021"
/product="hypothetical protein"
/translation="MGIKLNLRKVQTAWLNVFERAKDRENNDGSVTKGTYNGTFILTP
EHPQIEELRDTVFAVVSEALGEAAAEKWMKQNYGEGKHMDKCAVRDIAERDNPFEDFP
EGFYFQAKNKQQPLILTSVKGEKQVEPDFNIDGEQIEGEQVYSGCVANISIEIWFSEQ
YKVLGAKLNGIKFAGEGKAFGGSAVSASVDDLEDDEDETPRRERRRNR"
gene 11605..14703 /locus_tag="ABCD_0022"
CDS 11605..14703 /locus_tag="ABCD_0022"
/product="DNA polymerase I , phage-associated"

/EC_number="2.7.7.7"
/translation="MNLLYLDTETFSEADLKKVGSYAYAEHPTTEIVICTYAFDEGPV
QVWDATDGSDMPRDLRRAMLKLQKPDSNLKLVGQNFLMFDRPVIKHCWGFELLVENII
DTMIVAFRHALPGSLAALCEVLNIDASMAKDKRGKALIQRFSKPTPKNYKIRRYTADT
HPKEWAEFIAYAKSDITSMREVYKKMPKWGNSEFEDRVLHLDQVINDRGFKVDVALAE
AAIEAVTRHKEELQEEAQRKYGGSLTGKDFLPILQELAPAHRIHNAQKSTLNDLLADE
DLPDDARTIIEMRLGAASTASTKYAPLLLGRSSDDRRRGCLQYGGAKRTLRWAGKGFQ
PQNLARGYYHDDELDRGIAALLKGRAHRRFDVAKLTASTVRSCIIPEAGHKFVVADYS
NVEGRGLAWLAGEETALDTFRAGLDIYCVTAGKMFGMDPDDIKKERKDLRQIGKACLH
RHTQVLTDGGFKDIMAVTSTDKVWSGEKWVNTKGAHLMGWKPVINVDGVLMTEDHKIL
THSWKQAKQLVSNKYMMDRALEIGMDAWLSCASYQNDKAKDNYSSNVIVERCLGGSIM
TMSGRVKPLNATAAPLKQQKNIVNSISATKTQCRMMRTERDCSTGCPRQSCGQQAPET
RVTKTMGPGVSRYLTSGGPIKGRFLDMFKLWKAGTTQTSKWIESTQTVTTNLETFGLS
AVLKTAATDGRYPSFNESMMQPLPALLNLSGKLTYCEPVYDLIDVEDGNRFLIASDSG
FLVAHNCELGLGYEGGVGAFVTFAKNLGLDLIEMAKTMDGTFPDHIWAATARGYEWAR
IQEAKRPPHPGEKDDRPSYILDKKVWRTCDAIKRMWRESHPETVAFWRDLKDGILAAV
RNPGREFWAGAHLRRNGERAIRIWRTVEFDSSGRKVPGWWLCMGLPSGRILSYPGIGV
SVTKETDEDGRINTNVRIKYQGENQLTRQWTTLYTHGGKACENIVQALCRDLLAYAML
NVEAGGYPIVLSVHDELVCETPDTPEYSVAELEKLMCALPEWAEGFPLVAEGAELKRY
AK"
gene 14790..15077 /locus_tag="ABCD_0023"
CDS 14790..15077 /locus_tag="ABCD_0023"
/product="hypothetical protein"
/translation="MSTPEGRVQKYAKERFEALGGLVRKLSYEGRSGAPDLLVILPGG
IIWFVEVKKDENTKPDPHQLREHERMCKRGANVFVVGSKKQVDKLIEHYYI"
gene 15109..15300 /locus_tag="ABCD_0024"
CDS 15109..15300 /locus_tag="ABCD_0024"
/product="hypothetical protein"
/translation="MKHEYDRKPARDIVPGDMIFNVKTRQPVAVDTVYVESNGKLVIE
DVTGNVTAFGRKELVLVAK"
gene 15381..17762 /locus_tag="ABCD_0025"
CDS 15381..17762 /locus_tag="ABCD_0025"
/product="DNA helicase, phage-associated"
/translation="MGSGKCLKRGTEVIMFDGTTKKVEDVIVGDVLMGPDSTPRNVLS
LGRGREMMYEVKPRKGESYTVNESHILSLRTTSGIAKGSWPDNTVFDISVRDWLKLPK
YVTGPNGYLKGWRVPVDFPRKEQDEALLPPYLMGLWLGDGTSSSGAITSGENEKEIRA
YLESYAARNGMQIRKEGLTWSISHGNTGHKKHGFTHALKSAGVLNNKHIPHNYKCGDR
RQRLELLAGLLDSDGYCDLSKAGFDWISVSERLADDFCYLCRSLGFAAYKKKTRKRCA
NTDVWGDYFRVSVSGDFSDVPFVRGRHQNLPKRNINKNVLNVGIESITPVGVDDYYGF
TIDGDHRFLLGDFTVTHNTSSTMWVLNRLFRNGQLNDDDRVLILAPLRVASGTWPAEQ
ARWNFPCLSVIDATGSEKRRIAALESDANVVCTNYEVIEWLIDYYGKDDWPFTVIVAD
ESTKLKSFRSRSGGSKRARALSKVAFGKVKRFINLTGTPSPNGLKDLWGQNWFIDAGE
RLGSSYTAFTDRWFNSVQKGKSAMAREYHARPGADNEIHQKMKDISLTIDAAEWFGCE
APVIVPVEIDLPKKARQAYIDMEEKLFAELESGEVEAANAAAKTSKCLQIASGAVYVS
GPDGEATKDWEKVHDAKLDALESIVEELQGAPLLVAYQFKHELERILKRFPQAQAFAK
GTKGNKQMEAWNRGEIEILCVHPASAGHGLNLQDGGHHLAFISQGWNLEHYLQVVERI
GPVRQKQAGHERPVFLYHIVAKDTLDEVVAARTDEKKSVQEELLNYMKRRGKK"
gene 17759..17929 /locus_tag="ABCD_0026"
CDS 17759..17929 /locus_tag="ABCD_0026"
/product="hypothetical protein"
/translation="MSSIEKAIEEARLPLLARITELEVEVLRLRQQRDAANAQLDWLL
EQQEVDKERGAK"
gene complement(18048..20102) /locus_tag="ABCD_0027"
CDS complement(18048..20102) /locus_tag="ABCD_0027"

/product="Phage tail fibers"
/translation="MSSGCGDVLSLNDLQIAKKHQIFEAEVITGKQGGVAGGADIDYA
TNQVTGQTQKTLPAVLRDAGFSPASFNFTTGGTLGVNDADKAVLWPKEDGGDGNYYAW
RGPLPKVIPAASTPLTTGGISDSAWVAFGDITFRAEADKKFKYSVKLSDFTTLQQLAD
AAVDSVLIDRDYNFSNNETVNFGGKTLTIDCKAKFIGDGNLVFTQLGRGSVVVGAYME
SVTTPWVIKPWTDDNQWITDPAAIVATLKQSKTDGYQPTVNDYAKFPGIESLLPPEAK
DQNISSVLEIRECTGVEVHRASGLMACFLFRGCHFCKMVDADNPSGGKDGVITFENLS
GDWGKGNYVIGGRTSYGSVSSAQFLRNNGGFARDGGVIGFTSYRAGESGVKTWQGTVG
STTSRNYNLQFRDSAVLYPVWDGFDLGADTDMNPEDDRPGDFPISQYPVHMLPLNHLI
DNLFVRGSLGVGFGMDGQGLYVSNITVEDCAGSGAYILAHETVFTNIAIIDTNTKNFP
ANQIYISGACRVNGLRLVGIRSTSEQGLTIDAPNSTVSGITGFVDPSRINVANLMEEG
LGNSRINSFNNGSAALRFRIHKLSKTLDSGSVYSHLNGGPGSGSAWTEITAIAGSLPD
AVSLKINRGDYRAVEIPVAMSVLPDNAVRDNGSISLYLEGDSLKALVKRADGSYTRLT
LA"
gene complement(20115..22673) /locus_tag="ABCD_0028"
CDS complement(20115..22673) /locus_tag="ABCD_0028"
/product="Phage tail fiber protein"
/translation="MALVIHYTRNEDGTFDVKRYRDNPMNFVVNHVPDGVPVRVFIDE
IGEDNDVTEDFEALKENATFHIVESAGGGAIKGVMKIFSVILKPLAKLLSPSVKGASS
NLANSQADSPNNSLTDRNNKARPYERSYDICGTVQTIPNNLMSTYKVFNAAGKIVEYG
YYDAGRGYLDIHPEGITDGDTRVSDITGTSVAVYAPYTSPNNTSTPQVMVGDPIEQGL
YITVESNEVDGVVLKAPNGLGISFSYMSGYPSLSGNIGTIYDPTGGSDFSGVLVPNDT
FSLVSAWTNTDVDLSGGGYQVVSVSEGTVTFIVPGGLIGRWQEIRPGSFFRGDGEASL
QPDNAYEKTLTDWVSINRTEVERIVANIAAANGMYKDNGKSKTLASVTAEIQYQLLDE
NSTPYGPIYTAQGTVSGRTPDYNGVTIYADLPVVSRVRVRARRVTDLDFNFEGSVVDE
ITYVNLYGQTRDNTPHYGNRTTVHSMRKQTPRAAEVKQPQLRMIATEMVYKYLGNGVF
EDTMTPNTQAVQSLIRLARDPDVGGLNLTVRNMDKLLAVQNEVEAYFGDKQAGEFCYT
FDDYKTTMQDIVSTIADAIFCTPYRRGADILLDFERPRMGPEMVFTHRSKAGTSEKWT
RTFNDSQVFDSLKFSYIDPKTNVKETITIPETGGLKTETYDSKGIRNYKQAFWAANRR
HQKNILKKISVSFTATEEGIFALPNRAVSVVKGSRMSTYDGYVTAVNGLTVELSQPVK
FTSGDDHYLVLKLRDGGVQSVRVVPGAHDRQVIMTSVPQEAIYTGNSALKTEFSFGNE
ARHNAQMILVSTVDPGDDRTVKITGFNYDKDFYKFDNVPPFGRAFSSGFDNGFN"
gene complement(22664..23029) /locus_tag="ABCD_0029"
CDS complement(22664..23029) /locus_tag="ABCD_0029"
/product="hypothetical protein"
/translation="MNINDYTGLPYDFRRRNCWHHVRNVRADAGLSTPMFDVTSPTAI
DAAFDDGHSDPKGLRRVVTPQNFDAVLLGVKHRGRIVWHAGVYYEGMVSHCELASRQV
RLDSLEDLKDTYSEIEFWR"
gene complement(23026..23541) /locus_tag="ABCD_0030"
CDS complement(23026..23541) /locus_tag="ABCD_0030"
/product="hypothetical protein"
/translation="MSQESVEAAYRRKLASNPDGEMDFITLEISHPLLSKRWLLVRGV
NDLTATLETGEVVTFEGTPMEAKNAANNNDMDQTASFSLPDVLNILDEEMDRIPYDNK
ELPKFIFRRYVSTDLSYPCDGPVVYELQTLTQEKGVFTAETGTPMLNQRATGILMTPE
EIPLLRGILTS"
gene complement(23538..23981) /locus_tag="ABCD_0031"
CDS complement(23538..23981) /locus_tag="ABCD_0031"
/product="hypothetical protein"
/translation="MTRGNNVWRVDLAGGGVRQGRDTYFDVFPISVTLVVSPLGRQAF
LSFMEKVDGGASSFWMKHDLGQGIEDYQVTLTSTWNESTDDGKNWVITFTATAEKSPF
QEVGSACLNQNLPDLYGCYGDCLGEFLKTYGVYQTTFPRIWDPMQ"
gene complement(24040..26373) /locus_tag="ABCD_0032"
CDS complement(24040..26373) /locus_tag="ABCD_0032"
/product="hypothetical protein"

/translation="MADVASLVVKVTEQGAKATSDRLDNLSKSAKVAGAAVTGLAAVV
AATAYKAAQELVESQRQLDKMSASLKTLTGSTQGAKQALSILQDFARDTPYGLEQAVE
GFRKLVALGLTPSEEALRSYGNTASAMGKDLNQMIEAVADASTFEFERLKEFGIKAKQ
NQSDIEFTFQGTTTVVKKNAADIEQYLLNIGNVNFAGAMADQANTLNGAIASAEDSWS
QLKMTLATSLDVGSLAEPLRYIDDLIQEINAQVASGEFVAEMRMWGDMASDVGGAIEA
SFDAAFGMVADALNALNSAWTYTSESITGSGEETASTIAESAADALDFIAQEFTAMER
FFEDMVKGAQDAGRLVKAALTPGESVAEAKNLNFQLALAMDTQRAVTDLTRKSFREQV
EAQEDLIALKRAAYDIDKEAAKAEGLGKFKVSGKDNGSTGDSADKAAKKSVDAFERQK
KAAEDFYYQSIHLNDDVFQKIQANQEEQLTKLQEFFSNRLLSDQQYETAKTQIMLEAD
TARQAELDKREKERLEKQFSADAYVAQMQALAEGEFAELDRQYEVKLQKLNDFHAQGL
IAEETYQQTLNAMNDTYALDRAKATGTAFGNMASNIGAALGEASTAYKAFAIAQATIA
TYTSAVEAYKSTAAIPVVGPYLAPVAAAAAVAAGLANVGKIRSAREQGGNLAAGQIST
IAERGKPEVIMPASASRVRTAEQMRQIMGENDAKSGGDNVTIVNNTTGRIDSAATERD
DEGRLRIIISETVSSALLDSNSAISKSRRATRGQPGY"
gene complement(26366..26725) /locus_tag="ABCD_0033"
CDS complement(26366..26725) /locus_tag="ABCD_0033"
/product="hypothetical protein"
/translation="MFQFARWNFITRHERRKFDSIADGHKAALIAMGVIKDAGETTQD
AGPECPPELLTTFEKYRDAKFTRRVDDDGVKLYPREQLSWSDLVAYSTISGQNIGMFE
SEIIMGLDAIFEGRNDG"
gene complement(26731..27234) /locus_tag="ABCD_0034"
CDS complement(26731..27234) /locus_tag="ABCD_0034"
/product="hypothetical protein"
/translation="MASTVAPEKRVVPAWRNYQPVTDQPKGITMKLSDFYYEAEAEKG
ARMPIPLKDGTDSGEWLNVVSPEADVAVKAMRAFTLAYRAAVGKLKPLRDKCEEQKDF
SEYNLKMEDAAGDLNRQLALELVNGWSLDDEFTKENLKTLLAQYKRLAEHVVVFHHEQ
LRQLQEK"
gene 27317..27496 /locus_tag="ABCD_0035"
CDS 27317..27496 /locus_tag="ABCD_0035"
/product="hypothetical protein"
/translation="MSVFIGIIAFIYFIPFLVALLRNHKAKLGIFVCNLLLGWILLPW
VGALIWACNSNVNGK"
gene 27559..28689 /locus_tag="ABCD_0036"
CDS 27559..28689 /locus_tag="ABCD_0036"
/product="hypothetical protein"
/translation="MVTRKITDEQLQQELNAGLGPTEIAKKYNMSRRNVQLRSARLAK
KGVGHGRDVSHLVPDGYKIKGTSSLVDEFGNTKLQWVKTDADAERQVELMRAVIDGMK
SDITPVSSVPRPKKRLNEKLLNLYTVSDFHLGMLAWADESGDDWDMKIAEDLFSRWFD
AAFQKAPDAGVGVINLLGDFAHFDSLDAVTPASGHVLDADTRYQKLVRYMIRMVRRVV
NMALVKHKNVHLLIVQGNHDESGMIWLAEMFNTLYDNEPRVFVDTSADVYKMVQHGKT
TLFFHHGHKARFDAIEPVMIAKFRKAFGESVYSYAHVGHLHHQKIVESRNMIVEQHRT
LAAKDAYASRGGWMSGRSANVITYSAEYGEVARLTISPEMLG"
gene 28686..28916 /locus_tag="ABCD_0037"
CDS 28686..28916 /locus_tag="ABCD_0037"
/product="hypothetical protein"
/translation="MTNKYNRTMTNTDGDSITCDVYDVLRAFDIRDPALQHALKKLLC
MGLRGHKDTGTDLAEAIESLEKLRKYRSNIDE"
gene complement(28884..29030) /locus_tag="ABCD_0038"
CDS complement(28884..29030) /locus_tag="ABCD_0038"
/product="hypothetical protein"
/translation="MVSSPLVGVDIFNFEHSYRSNQILYRDDKKKAPKGAFFSLINIT
TVFP"
gene 29029..29169 /locus_tag="ABCD_0039"
CDS 29029..29169 /locus_tag="ABCD_0039"

/product="hypothetical protein"
/translation="MNALISVSEAQTMSSREIAEKKKNGERTRQRTTRYIKNGTGTFP
QF"
gene 29370..29714 /locus_tag="ABCD_0040"
CDS 29370..29714 /locus_tag="ABCD_0040"
/product="hypothetical protein"
/translation="MAADLWEEKERLALENKEMAPKADVYDRIIDRNNLYNATQVAQK
FGQSAVWMNKQLEQFGVYNRSVKRGRVFQQWFVDKGYGIMRETETGHSQAMFFAEGEM
WIIGKLTAEGLI"
gene complement(29743..30912) /locus_tag="ABCD_0041"
CDS complement(29743..30912) /locus_tag="ABCD_0041"
/product="hypothetical protein"
/translation="MALQPYKGAMTAQFYVLETTPGVTPDNPVWQPLRNTGGIPAVTR
DALISNELDGSRETSSIRTGNRQVTGEYAIELSATSQDELLAGAMTSSWVAGSTKSGI
SVTVDPAVKTFTRATGSFVTDGVEVGDLVQFDGLSGNNDKAFLVTAVTATVVTGAGIQ
HTLTAESDAQADLRIADKLETGNLCKTYSILTWLKGKCGNPDSYIITRGVEFTGFTIE
QAVNAMVTGSFPFIGLNQEILQTPPSGSDFTTNFSARPFASVDVSAYDGAAPLKLIDT
FTITNDNSASAQFELGNNSVAFVERGRAANTFSLAGKLYDMTLLNKFLNETQMEVSSV
LNGPDGAMSFTLKRASLTSATPEIGGPESVTLSLEGQATGNQFQSSIVIQRIKYT"
gene complement(30912..31331) /locus_tag="ABCD_0042"
CDS complement(30912..31331) /locus_tag="ABCD_0042"
/product="hypothetical protein"
/translation="MATYFEDLTKAFDTALAAFGTDNDIMVALENIDAPTSTDVPYLA
SYMLLSDTEQADLFWTEQRAGIYQVDINVGSTLGSAPLNRLADKLNATFAAGNCFSRN
EICAEVQSVSLGPLIVENGWAKRPLSINFIAFTARIR"
gene complement(31331..31726) /locus_tag="ABCD_0043"
CDS complement(31331..31726) /locus_tag="ABCD_0043"
/product="hypothetical protein"
/translation="MSFALDVSKFVEKAKKNPEKVMRQVSIKLFSAIIKASPVDTGRF
RMNWMASGSTPADGTTDATDKSGNTATGNATSFVLNAADWHTFTLTNNLPYAQRLEYG
WSQQAPQGMVRTNISRFQQLLNEEASKVK"
gene complement(31723..32082) /locus_tag="ABCD_0044"
CDS complement(31723..32082) /locus_tag="ABCD_0044"
/product="hypothetical protein"
/translation="MSTAFSKRMQGVGTRLLSKYGSTVNLVRKGQKTWDPVLGEYVWG
PDVVLPLKAVPVPVNAGLVNGTTIQAGDMMVKADYSVVPKMDDKVQFSGEQWSVVAIE
KKMVNDDVVAYFIQVRK"
gene complement(32082..32687) /locus_tag="ABCD_0045"
CDS complement(32082..32687) /locus_tag="ABCD_0045"
/product="hypothetical protein"
/translation="MAKTKSEIFALIGASFPDNQSGLITPEKLREVTTQMADSMLYRA
KEVEVLRASSTDIQAPSTTGTALTVAFGGAQKTSADPVMINASGVVTFNAAGNYAIRV
KLQAGRTGASGTSILLSRVLLGGAQFGSTAATKLASAESTIPIESRVVVNAAAGQTFT
VEIMRDAAGSNFGGLYPQAATVTSWGVAPSALLVVSRLEGV"
gene complement(32690..33199) /locus_tag="ABCD_0046"
CDS complement(32690..33199) /locus_tag="ABCD_0046"
/product="hypothetical protein"
/translation="MPLIVETGQGLPNADSYVSLEDGRALAAKYGLELPEDDIAAEAS
LRNGAVYVGLFESQMCGRRVSANQSLAFPRTGVTLHGFPQPSNVIPSLVIQAQVMAAV
EYGAGTDVRGTTDGREVQMERVEGAVTVSYFKNGYSGGTVSITAADDALRPILCGNNN
AYSFNVFRG"
gene complement(33203..33391) /locus_tag="ABCD_0047"
CDS complement(33203..33391) /locus_tag="ABCD_0047"
/product="hypothetical protein"

/translation="MADNYVVREQYKGVVEVDGQLMPTREEVNPEALIETQPVAEEPH
YNGGREPKQRRRRKSVEE"
gene complement(33428..33778) /locus_tag="ABCD_0048"
CDS complement(33428..33778) /locus_tag="ABCD_0048"
/product="Phage neck whiskers"
/translation="MVDVIKRRIVGVSDDSPQDGQVEIDMENVTPLRFSTGLDDTTAV
TAGQAITLTVALADGMGPKTVQWYKDNNAISGATGLTYTKANSAAADSGTYKVVAHDG
YGNIISDSTVVTVS"
gene complement(33792..34076) /locus_tag="ABCD_0049"
CDS complement(33792..34076) /locus_tag="ABCD_0049"
/product="Phage capsid and scaffold"
/translation="MSKTGKGLPRSLVNAELDIPAATTTAIGGVKKSATVAAPPAITA
GSGAAAAAAPTKEEFDALVTNYNKLRTDVTSLRTAVTNLLTALKNAGTVS"
gene complement(34137..35186) /locus_tag="ABCD_0050"
CDS complement(34137..35186) /locus_tag="ABCD_0050"
/product="Phage capsid protein"
/translation="MAITTIGDIVTGNIPVLASYMTEDPVEKTAFFNSGILTPTPYAA
EIANGPSNIANLPFWKAIDTSIEPNYSNDVYQDIATPRAIQTGEMMARVAYLNEGFGQ
ADLTVELTSQNPLQSVASRLDNFWQRQAQRRLIATALGLYNDNVSATDAYHEQNDMVV
DVSATLGFDAGAFIDATQTMGDALMGNGGEVLGAIAMHSFVYAQARKAQLIDFIRDAE
NDTMFATYQGYRVIVDDSMTVVGQGAQRKFISIIFGQGAIGYGEGNPEMPLEYEREAS
RANGGGVETLWTRKTWLLHPFGYSFTSAVITGNGTETIARSASWQDLANATNWNRVVD
RKHVPIAFLVTGVGA"
gene complement(35190..35891) /locus_tag="ABCD_0051"
CDS complement(35190..35891) /locus_tag="ABCD_0051"
/product="hypothetical protein"
/translation="MNRFLRYPLHEEAGVEDKPGAGDAPKMYTAEEVQALIEKEVAGL
KANQEALLSEKKEAARRAKEAEEERQRAHQEALKAAGKMDEFEKTIRSQYEPVLKEKE
ERYASLAARILDSERKAVLGSFAGDFIAPEAVEILAPFVKTEFDGEDVVTKFMGADGN
VVTTDPEQFRKYLREHKAFSHLIKANAASGGGASGNKGGGAAPAFKDMSESERLALYK
SNPAEFERQLKALRK"
gene complement(35976..36101) /locus_tag="ABCD_0052"
CDS complement(35976..36101) /locus_tag="ABCD_0052"
/product="hypothetical protein"
/translation="MALPSSDADEDLAIDVQNAECTRQLRLKVFRLQEYIRNILE"
gene complement(36085..36471) /locus_tag="ABCD_0053"
CDS complement(36085..36471) /locus_tag="ABCD_0053"
/product="hypothetical protein"
/translation="MNWKVTAIATAAGILSLWLYGQYNYRSGWVEGRANLVSQQQQKA
QAELAKKTQRQRQNDTRAAAAETEGKENAEAITREVIKYVTRPGRVVCVFPPERVSIK
RRAAENANSIPGYDTNAAAVQNGIAK"
gene complement(36790..37248) /locus_tag="ABCD_0054"
CDS complement(36790..37248) /locus_tag="ABCD_0054"
/product="Phage neck whiskers"
/translation="MGFFKVKDVPSRRVVQYSRVSGAGEGVVYIKDESVLGESVDEMP
FADKTGLAAIPDGILYEVPYLDGAGDVYFDTQPADVELKDGSAKLTVVVKGGKAPYDL
QWFKNGKEVINVPYVEGELTVKDPGEYFVRAVDADGISVVSKAAKVSEPK"
gene complement(37251..38294) /locus_tag="ABCD_0055"
CDS complement(37251..38294) /locus_tag="ABCD_0055"
/product="hypothetical protein"
/translation="MSLLESIISHQIWLQRNASSEVKDLAPFISQMRDEVKRQVLLFG
DDSRTAARLTIMLRELEQALNGITSGWYEKLLADARELSDYEVNWNVKTLSTNVNANF
VTPAAEQVWAAATFAPLELSEKPVDFVSLMNGWRQTEVNRLVMGVKSGFVQGMTTRQI
VKNVVGPGGLADISERNAATVIRTALAHVSNEARQQVYAQNGDIITKYEWVSTLDSRT

SAVCRSRDSMQYEIGKGPLPPAHPNCRSSTAPVISSEFDFLDKGAKRAARGADGGQQV
SADTTYYEFLKQQPAWFQDEALGPVRGKIFRNSGITPEEFRVISVDGFGRPLTLKEMA
ELDKRVADYLKEE"
gene 38471..39121 /locus_tag="ABCD_0056"
CDS 38471..39121 /locus_tag="ABCD_0056"
/product="hypothetical protein"
/translation="MNVQTLEFPYGEVEFTRDEYIDASTLTIQMNRWRFDNGLQPRNL
AQILKTDDCKNFMRVCEEETGIVPLKTSRGRNGKTWLCLHLAVYIAEQYSAYFHFLVI
DRFITQRQVELRNIGAVSFVELNAAVSRMLERTEGRIGHSGHFIQVAKAIKESIDIRE
VTGFDTWDSQDAKTNQLRSEIQKSMVTLLDMEAVNSWDELKETIPRVVRKCAANIC"
gene complement(39151..40620) /locus_tag="ABCD_0057"
CDS complement(39151..40620) /locus_tag="ABCD_0057"
/product="hypothetical protein"
/translation="MLTANVQGSGVKTKHREWLHYAPKWQKVRHALAGELVGYLRNVG
LNEPDKAYGEARQAEYEAGGIVYNFTRRTLSGMVGSVMRKEPEINIPKELEYLLKNAD
GSGVGLIQHAQDTLMEIDSVGRGGLLVDAPETGAATAAEQNAGLLNPTIAFYTTENIV
NWRLTRVGSVNRVTMVVLRETWEYNEPGNEFETKYGEQYRVLDIDTDDNYRQRLFRFD
AEGGAQEDVVEIYPDLGESLRGVIPFTFIGATNNDATIDDAPLLPLAELNIGHYRNSA
DNEESSFVVGQPTLFIYPGENLTPQSFKEANPNGIKFGSRCGHNLGYGGSAQLIQAGE
NNLARQNMLDKEQQAIQIGAQLITPTQQITAESARIQRGADTSVMATIARNVSQAYTD
ALRWVAIMLGKPEDTEVEFRLNMDFFLQPMTAQDRAAWMADINAGLLPATAYYAALRK
AGVTDWTDADIKDAVADQPLPVATEVQGEIPQSAQQQEK"
gene complement(40633..41904) /locus_tag="ABCD_0058"
CDS complement(40633..41904) /locus_tag="ABCD_0058"
/product="Putative phage terminase"
/translation="MNVDITATEPQGAFLNLHCKFPAFVAGFGTGKSEVMCNSALLDS
MEGGSDSLIAMYEPTYDLVRLILAPRMEEKLSDWGIRYKYNKSDNIIYTSSGQFGDFV
LRTLDNPARIVGYESFRAKIDELDTLNKDHAEHAWNKVIARNRQLPRTYRPITPKPAN
TVSVFTTPEGFRFVHDRWAVKKKPGYEMIQASTTSNPFLPEDYVQSLRDTYPGQLIDA
YIDGEFVNLTSGSVYYAYDRRKNSSRETIQPGETLYIGQDFNVGHMASTVYVQREYVW
HAVAELVDMFDTPDVVREITERWKRHGHHIVVYPDASGKNRKSTDASTSDIAQLQNAG
FEIRAKSVNPAVKDRVASVNKALESGKLMVNEQTCPVTARCLEQQAYDKNGIPDKTSG
NDHQNDATGYPIAYEMPLVKPVSHIPVTFAL"
gene complement(41894..42400) /locus_tag="ABCD_0059"
CDS complement(41894..42400) /locus_tag="ABCD_0059"
/product="hypothetical protein"
/translation="MKLKLKQPDAEVVAAAHEEAVTANRRRKRPRGKQSLYQSSRHSA
ELWDPDYCDELIRFFDRASWELVPTSKGDERPLIQDKPPSLARFALHIGVTIPIIKLW
LREVPAFAEAYETAQALEEAYFTETGAAGISATFAAAKLGLNKTVVEETRDEPISEVT
IKVVSGER"

ORIGIN
0001 cagacacccc ctgagcgtct attaaggatt gtacagggga atgaaggcct gtgtacagaa
0061 aagcccgcag aagcgggctt ggtggtaagc tattttacag ttctacagat gataaatggt
0121 tggccgcgga cttaaccgca tctttaaggc tatcgtacgc ggaagtaaaa gacatatgca
0181 ctcctttcga gacaactata gcgccgccgt taaagtacgc ttttgcactc cagccaccat
0241 ttaccttatc aataaggatg tagtgactac cttcttccgt taccgtaacc gtgttttcgt
0301 ctattacttc aaatttagtg cccatgttcg ccctctcctc tttagcaatg ctcgtcttta
0361 taaacccacc atttacctac ggcttcccat cgccacccgc gactatggat ggcacagaat
0421 ctacggccaa atactgtttc cccgtaccac atgacaaata gcttcaccgc ggtagcagca
0481 ccaattttat gagcattcaa acggcgatac gctagtgaca tcctcgtaag acttaccgcc
0541 tcaacaaatg gccaaagaaa cagaatccac atcaggtata taacaacggc aagtagcaag
0601 gttgtaaatc cgatacccat gagaatcaga aaatcattca taacccctcc agcttcatgc
0661 ggtcgttgcg agataccccg cgtttcgtgt agattaccgg tgatgttttc gtatgggtca
0721 tcagtctgtt gttatagata acatgttgct cgcattttgt atcatggcgg tacttagata
0781 acgtcatctc gttaagccca gtctcgcggc acgtatctgc taatgtgccg tgggtctcga

0841 ttagcttggg gatactggta atcagttccc gccgatgctc attaagatag tccggcacgg
0901 cttggtgcgc ctcgatttgc gcgtctacag cttcaatctt aaacctgccg ttggcggctt
0961 ctaccatcat cttttcatat tcgttgttag tcatttcttc ttgccccaca ttcgattgag
1021 atacttattc ttgtcgggcc cgggaaaact gttgcgtttt aacagttcct cgcgtgtcgg
1081 gaaaggcgta tttccaacct taccaccaac cttaagcgtc accgttactt gcccgttgtc
1141 gctcattcct gcacctccca ctcgcgtact tctacgccat cataatcgcc tttagctcgc
1201 tccagctcgg cggcggcttc ggccttttct tcgctgagat aaacaccgag aatgtccgca
1261 ccatcgtagt aataccatcc agatatgata tatactttca tttcttctct cctgcataag
1321 ctgctttcag tgcctgcatc gcagcggccc agaacgcgtc agccgggata ttgctgctca
1381 ggttattgac ggcgatgtgc gccatcagtt gcgcgtcttc gaatgcttcg gtgtccagca
1441 ccggtttgca atggtaaaaa actttatcgc tttgtaagtt catcttctct actccgttct
1501 cgctgtcgat gagctaaata taataaggtt aatgctacta gtcaactatt atttttggac
1561 acagataaaa aaatcccggt aagggtcagt taccgggata aaggagctag agggatgaat
1621 aagcaggagc aggtttagta tcttaggaat actcctatgt gtcaacctac tttgcaccac
1681 ggccgatagc ttcggcctgt tgccacagca taccgtcgaa tagcgcaaca cgaccggcag
1741 cacgacgacg taatccgagg agtgatttgc cgttctgata atggaactga gttaacttat
1801 tccgcagtgt ggcaacgtcg cccttacgca gggcatgccc tgttccggta gaagccgcaa
1861 tcacaccggc accagcgtta tacaccaggt cgcacatcgc atcgaactgt gactgattga
1921 gcgacggatg cgcaacggcg tctaccgcgg ctacagcctt aaccatgtcc ctgttcagta
1981 gcaaaagacc ctggccttcg gtaatcttct ggccctcttt cacatctgcg ccgtagtggc
2041 cgtagccaat tgtcaggtac ttctcgttct tcgttgccct gtacgcggtt ccccggaacc
2101 cctcgaatgc cgcggtgaat ttgatgccgt tgttgctaat gtttcggttt gacatacatc
2161 ctccggtatt tgtagtagcg ggcgcctttc acgcaaatat gacgcaccga tttaagcgcg
2221 aatatagcga cgagagtcgt aattatgaat ggcgggatgc caccggtaaa gagatgagct
2281 agcccaccgg cggcggtaac aacggacgcc atatacaaca gacgaccgaa aagaccgtct
2341 tcaatccagt gagcgtagat attaattagc gaggtgacga tgattgcggc caggcaaatg
2401 acaccaattg gtgagtcgag catgttattt cccccagggt aacttaaagt tgctgaggga
2461 cgagtcggcg gcctcgaata atttaagcca gaacatacct acgccaaaag gaaccaggta
2521 ctgaccctca gactgtgata gttcatagta acttataatg atgggggatg cgtacaccgc
2581 acaacaggca gatgctacga gatggcataa tcgcaaccac ggcgaaatag cctcacgctt
2641 cttaatctgg gatacggcac ccccggcgag gcccgccaca ctaagccaaa ggtatttatc
2701 gtccacaata aaaaatctcc gaaagttggt ttacctttgg agattaacat ttatacgaca
2761 attttagtaa cagataattc ggatttatcc tagtcgatgt gctatcactc ccccatcagt
2821 attagggtcg aagcgcttcg ctatctcgac agcggcagta gcggtggcac ccataaacat
2881 cgccgtaagt gcatacggag caccggagcc aacggccgct gcgttacccg caattggtat
2941 aactgagctt aacggcgagt cccatgactt gaagaattct accctgaaaa tacggttatc
3001 ctctttaaca aacaccagcg ctgaaaagtc gaagtcgtag tacttaggcg tcaatatctc
3061 ggcatcgaca accaacagcc tacccacccc ggcgtccccg gatacaccca taacgaagtg
3121 gtcgttctcg taaatcttgg tatccgtata gtatttaaaa ttccccgtga cacatgtatc
3181 gcaggccatc gtttcaccgt caaaagctat agtagtcata taagttacct gaacgtataa
3241 agtcgaatag atggcttgtg agggatatcg gaatacccac gcttgcgcca cagcttatac
3301 agtttaaccg ccatatgggc cctggcctgg ctactgaact caccgaggtt taccactttc
3361 ttgccgtcgt aagagtgcgc acgccacaga cctttgccta cacaatactt aggctggtgc
3421 accccaggga tgcgcggata cacgtccggc atctcattaa ctggctttgc gcgcggcttc
3481 ggcttacgct taaccttaac cttaaccttt tctttaacgc gagtaacctt ctttgttacc
3541 gccgatgaag gaggtgccag gttgatgccg gcgtgtttca gtatatcggt tgcgagactc
3601 ataatcactt ctcctgttta ccaagtcgct taggcgagca gatagcgcgt acctctgaat
3661 caattggctt gtcgccctga aacaggaagt gagcgttctc aacagcccgt gccgctgtct
3721 gacaagcttc taatgtgtag aacgtctccg atgccgcgag ttgcatttga cccgcggtca
3781 atacccagat gaatagtacg ctagtcattt atatctctcc cctcgttccg atgaagtaat
3841 agtgccctat tatgtttggt tacgtcaata ctattttacg attaatttat cctggcacat
3901 cagaaagtaa cgaggcttct tggcatgagc gcgccaccgc tcgacacggg catagatgcc
3961 gtgtagcgat gttactttgc cgtgtcttgt tcccccgtta cgcccggtcc aggttacgag
4021 gtcatttggt ttcattgttg catctcctgc aatagtcgct tcgattgttg cgtctagtga
4081 aatagaaaaa tattaacaaa aatggtacgc tgctactcag cagcgtaagt cactcagtaa
4141 aatcctaaat agttgttaca atatagcatt gcacttccag cggttccggc aacggcatcc
4201 agtgcgacgg ttcccacgat gcgccgggat tatccaccca tcattcgcca gtcaccttta
4261 cagccaagct cagtagctac gagcacgcct gtcaatttag gcggcatacg ctcactacat
4321 ttaatccact tattcatagt tcctcctctc ttctcaaacg cccacgcagg acgcttctga
4381 acgttagaag gcttctctca taacttatcc cccacgtaac gcctgagtcg cgacctggat

4441 gatagctgat aagcactgcg cgcaaggctc ttgctcgaat ccgttttcta cagcacacag
4501 ggcgtgttgc gcgtcctgga acagccattc cccgtacgat ggagcagccc cacaccacgt
4561 ttctatttta ttcgtcggca atctaacccc atccagatgg tctaccttga ttatgtgttt
4621 cccccatacc aacccccctca atgcatgcac cggcaattaa aataaacggc aatctccttt
4681 aagattcaca gcccgaagac tcaaactctt cagaatcttt tgccgtgtag ttatttccga
4741 ttaacgttac ataccgcgcg tccataaacc ctgcacaccc caacttacac tcctgacagc
4801 taacatcttt cttttctggt caagaactat gtaatattcg tgagagatag cacaccgcac
4861 taactgaccc ttcttaagca tgactcatcc tcttttctcg ctaacctgca ataatagtac
4921 tctactatat taggttctgc aagtgccttt ctattcttgt ctcccgtccc tcgtcgggta
4981 agcgcttgcc gcaaggcaag tgcgttaatc cctgagtagt agactaagaa tctttcggca
5041 tttcagttac gttgccggtg tggttcttca accagggttg tgcctttatt cttattacct
5101 tgtaactaaa gggcgggcct taagcctgcc ctctaatagt tacattttac agttgcgaat
5161 ctacggcaca tcaaggaaat cagatactta agagatatgc tgccgatgag aatatagctc
5221 gcccgaggca gttcagatac taaggaagga agttacttag ttttaagttt gcggggtttc
5281 agtaagcctt aacgcgtttc cgcttagggc ggaaccgcta agctctcaca aggtgaaaca
5341 gaaacttaag taacttgaag ttacaagaga caacttacac gaaaaaaata agaatgtcaa
5401 taggtgcgtt tatagttgat ttcccattat gccgtgggta tacttgcttt actaactgag
5461 atggagggtt tgataatgac gcagaatgaa gtagctaagc ttattggggt gacccgccgt
5521 acactgaata actggctaag agacggcaag ttcccggatt gttgtgtccg gattatgggt
5581 cgccggatgc caggaacatt cgacagggag aaagtggaag cgtggattag ggagaatgtg
5641 aagtgactga tttccagaaa agagtattta gagctgtagc tgctcttaca tttaaaaaag
5701 ggtcgtgtag tgttcttgac ttgcgccgga cttattttca gtactatagt tcctctctca
5761 tcgagaacac gttgaaggtt ctggttaaag acggagcagt taaaagtaag gccggtaaat
5821 attcggcagt agctgaagtt aaatgctctc aggcgacgcc cgaagacctt gaataaaaga
5881 aaagccccaa cgcgcgaacg tcagggcttt aaacactaca aaggaaaagc acatgtcaac
5941 aggtaatcta gcatactcat ggtcgcgtag caatcgccgt gaggaacgcg ctgacattaa
6001 cgtttacacc acgacttccg gcacctggga tgagttcgtt gaacttatgc agccgcttaa
6061 acgttcgcgc cggaacccca agacagaccc cggatatatc accgccgcgt gtaccgccac
6121 ggtaagctct actggtaaag aagccgccga aggtatgttc tatcgctgca atgcgtctgt
6181 tacgtcttca tccctggcct atgccgacgt ggatagtgcg actccggaag agttcgcaac
6241 tgactgtgag atggtgcgcg agtcgcgttt cgcaatgatg ctctacacca cggcatccca
6301 caccgaagaa gcgccgcgct atcgcgtcgt tatgcctgtg cgcactccgg taaccggcgg
6361 cgacatcatc cgcatacggt acggcctgtt gacgcatttc cttaaagggc gtgacgtgga
6421 tagcgccggg ttcaccctgt ctcagccgat gtaccgcccg cctgtcggga gtcaggtcat
6481 cgtgtccgaa agtagccgca tgattacggc aagcaagctt atggaggggg tgccggaaat
6541 taacgttacc ggtgcttccg attatagggt gcctgagggc gagcaatccg aattaaccga
6601 cctgttcgaa gagttcgctt ttgaattcgg cggacgtatg accgaccgcg gcctgcaaat
6661 gcccgccacg ccggagcacg ccgctcaata cactaccggc gaacctaagc aggatgactt
6721 tttgttctgt tggccgcgcg acggcttcga gcgccccaac gttaccctgt accacgatac
6781 cgacctggta gctacgggcg ggatgaaacc cggcggacgg gatatgtggg cttacgcttg
6841 tgcagctacc ggcttaccgt tcgaccgtgt agaggcggca cttgggtggg ccgagggggt
6901 cacttgcgac gaggaagacc tggccgacga agaaccacca gcaccccagg ctgacttcat
6961 cgtcgaaggg tacatgccga ccaattgtat ctgggatatc gttggtgaat cagggacgta
7021 taagtcgttc tacacactcg gcatgatgta tatcagtgct gccgggtatc ggttcgcggg
7081 ggcggacacc cgtaaagccc atcatttcta catagatggt gagggtgggg agtttacgca
7141 cacccgtatc gctgcgttgg ccgctaaata cggcgatgaa ggtatgcggt atatgcatgt
7201 gcttgatgct ggtgaattcg cggatacgaa gaaactggtg cgtaaaatac gccagatagc
7261 cggtagtgaa cctgtcggga tggttgcatt cgacaccctt aaccagacct tcggtaactg
7321 gatagacaag tttaatgaaa actccgccgg gcaggacggt atgggtcgtg tcgtggcgat
7381 actcaaagaa gtgcgcgatg gtactaaagg cgctgtgggt gtcgttcacc atactcctaa
7441 aggcgggagt aaggctcgcg gaagtggtgc gctgtatgcc ggtgtcgatg tggaactaac
7501 gcttgtccgc gctaccgaga aacaaataaa cgttgcgcac accaagaaca aaaacggtat
7561 gcagcagaaa acaataggca tggtgcttga accagtacag tttcgcgagg ccccgccgcc
7621 gaaagagttc caggctgtag aattcgttgg gggtgaggga tacggggaaa tcgttaacct
7681 agacctgccg gaaccgcata aagctcttgt actcatgccg tggggcttcc aaccgttcga
7741 aacggacgag gaaaaagagc gcaacgaagg actggacagt aaaggtaaag attcagtaaa
7801 ggatactgtt aaacgctcta aagatgcttc ggcgcgcgaa tcggttatgt ctgctcttga
7861 ggatttacag caggcagatg ataccgggcg aggcttcaca caacggcaaa tagtcgcgag
7921 agctggagac cacagcatca caaacctggt acttgagaag atgctacgtg agggtgaatt
7981 gatgctcggt tgcgatgaaa acggtgaagt tgtaacgaat acttacaggt tgccgacggg

8041 tatagatgac cgtaagcggc caaagaacag gtatgaacct aacgacaaca taaagacaac
8101 agaaggggat ttagagtaaa aagaaagggg ccatcaggcc cctgtatccc caaagcttaa
8161 ccatcggggg tcaacaccga gggcgtcggc aatgtggaag attttaaaac agtctacctt
8221 agatacccga ccggttgcca ggtgggttat aaggctttgg gatacaccag ctttgaaggc
8281 gagttcagtc tggcccatgc cgatttcttt acggcggcgc tctacgcggg cgcctaattc
8341 ggattgctgc atatgtttac tccttagtta gatatgtagt aatagtacag taataaaata
8401 cacggtgcaa tttacgaaaa agtgcttgca taaataatag agtacgatta cagttattat
8461 caccagctaa acaggagaac aaatgatgtt agaacaattt attaaattat ttgaacgttt
8521 cgtagttgct cacgaactga ttgctgcgaa cagcgcgaaa cagactgtgt accaacttaa
8581 aagtgatatc accgcagtaa acaaagagtt gatagaaaaa tcggtagaaa cggcggcgcc
8641 ggacgagtat cgggctagtt ctgcagtatg caaggcgtca gaaaaagaac tccctgtcga
8701 aggtgaagac gtcgtcgaca ctaaacccgc agaagaagag aagccgaaac gtaagacgcg
8761 taaagctaaa gtagaggagc cggaagaaga gaaagaagaa atcgattacc agtctcttag
8821 agaccagatt caggctatcg acgatgcgat taacgaaggt ccgagcgatg ccgcgtgcga
8881 cgattcagat gaactgctgg aagagttcac cggcaagaag atgaagattg ccgcgattaa
8941 agacgaagac ctggctgaat acctggagcg cctgacagca atcaagaaca agtatttcga
9001 agaagaataa ttacccagcg gccttcaggc cgctttaact gagggtcgga attatgattg
9061 accaggtggg gggggggcca ggtatgaatt ggcatgaact gttcaattat gaccccgata
9121 caggtaagtt gttttggaaa ataagccctt ctcaagcggt aaaggctggc gatgaggccg
9181 gtactctttg cggcggagga cgtagttgtg acaagaaata tttaagggtg tgctatagga
9241 gaaagtttaa gcaggtacac cacgtagtat gggaaatgtt caacgggcct ttaaaagatg
9301 gtgaggagat tgaccacata aatcatataa gaacggataa taggatagag aatttaagga
9361 aggtgtaccg cggcgataac gccaaaaata gaagtatgcc aataaacaat acttccgggg
9421 tgcatggtgt ttcgtacttc tctcgtattg atatgtgggt tgcaagaata ggggtaaatg
9481 gtaagtctat atttttagga agatacaata atttagaaga agcggcgctc gcctgacggc
9541 gggcagaagt aacatatggg tttcataaaa atcacggagc taagaagaat gactttaaaa
9601 gctaaagata ggagcgggag taacgacgta cacgccttat tgtcgccatc cggcgctaag
9661 aaatggttaa catgccctgc gtcgctgata tgcgaaaaag atatccctaa cacttccggc
9721 aaggccgcag tgttaggaac ggcaatgcat agcctttctg aatatcattt aaacgcgtat
9781 atacgcggca ctgcgttgcc gttagagcga gatgtagggg cttatgtgtt ggaagaaggt
9841 aaaggcgcgg ttaaagcact aattaagccg atgaaaggtg cggtgctggt tacagatgac
9901 atggtagagc aggtgcgaaa gtacactgac tactgcaaag cgattataga cgtatctact
9961 tacaccaaac tggaaatgcg cgtaaatctt actgaggtgt tgcacccggg ttatgaaggc
10021 gtggagacgt tcggaaccgc cgactttgtg gccgttcagg aactggcgaa caccgacaag
10081 cacatgctca tcatcggcga cctgaaaaca gggcggcatc gtgtcgaagc gaaagaaaat
10141 aaacagctta tgctctacgc tctcggtgtt tatcgcagac taaagagacg ctataacata
10201 acaacagttc gtctggtcat tttccaacca tacgctggtg gtgcgtcaga gtgggacatc
10261 tcagttgaag gccttgaact gttcgctaag ttcgcgcaga aacgcgcggt agcagcgctt
10321 gacgcgtatt cccgcggcaa gaagaacctg aaagcatcag acttcaagcc gtcagtagaa
10381 ggatgtcagt ggtgccggtt ctccgagcag tgcgccgcgc gtacaaaaac cgttaacgcg
10441 gtactggcgc aagaactgga agacgacttt gcgctagaac ttacgccgga gcaacttgta
10501 gctgagtatg agaagttgcc gttgctgcgt cagcacatcg acaaggttga gaaagctatg
10561 ttcgccgctt tgcattcagg taagaaagtg cctggataca aactggttga agggcgcccc
10621 ggcagtcgca cgtggaaaga cgcgaagaaa gtggctagca aatatggcat agcactaatg
10681 aagcaggtac tccttacgcc gaccgaagcg attaaagtag taccagaaga ggaactgaaa
10741 gacttcatta cccgcaagcc aggcgcgcca tgtgtgacaa cggtcgaaga caaacggcca
10801 gaatggaaaa atgttactga agacgattta gaggcttgac atcttaatcg tactctatta
10861 tattcttaat cactggccgg gcagcttacc ggataaactc taaaacgcga gaaaactaaa
10921 atgggaatca aacttaatct tcgtaaagta caaactgcat ggttgaacgt attcgaacgc
10981 gctaaagacc gtgaaaataa tgatggttcc gtgactaaag gtacttataa cggtacttt
11041 atcctaactc cggagcatcc gcaaattgaa gaacttcgcg atacagtatt cgcagtagtg
11101 tcggaagcat tgggagaggc cgctgccgag aaatggatga agcagaacta cggcgaaggt
11161 aagcacatgg ataaatgcgc cgtgcgcgac atcgcggaac gtgacaaccc gttcgaagac
11221 ttcccggaag gtttttattt ccaggcaaag aacaagcaac aaccattaat actgacttca
11281 gttaagggtg aaaagcaggt agaaccagac ttcaacattg atggggagca gattgaaggt
11341 gagcaggtat atagtggttg tgtggctaat atctcaatcg aaatctggtt ttccgagcaa
11401 tataaagttt taggcgcaaa gctaaacggc attaaatttg ccggtgaagg gaaggcattc
11461 ggtggttccg cggtttctgc cagcgtcgac gacctggaag atgatgaaga cgaaacaccg
11521 cgtcgcgaac gccgtcgtaa ccgttaatat attttctttt caattaaggc ggcttcggtc
11581 gccttttcta taagggtcaa ataaatgaat ctgctttatc tcgatactga aacattttca

11641 gaagccgatt taaaaaaagt cggttcctat gcttacgccg aacatccgac taccgaaatt
11701 gttatctgca cctacgcttt cgatgaaggc cctgtgcaag tatgggacgc caccgacggc
11761 agcgatatgc cgcgtgattt gcgtcgggcg atgctaaagc tgcaaaaacc agacagcaat
11821 ctcaaactgg taggccaaaa cttccttatg ttcgaccgac cagttattaa gcattgctgg
11881 ggattcgaac tcctggtaga aaacattata gacactatga tagtcgcgtt ccgacatgcc
11941 ctcccgggtt cactggccgc gctgtgtgag gttttaaaca ttgacgcaag catggctaag
12001 gataaacgcg gtaaggcgct gatacagcga ttcagtaagc ctacgcccaa gaactataag
12061 attcgacgtt atactgccga tacccaccca aaagagtggg cagaatttat tgcatacgca
12121 aaaagcgaca ttacgtccat gcgtgaagtg tataagaaaa tgcctaagtg ggggaattct
12181 gagttcgaag accgcgtgct gcacttagac caggtgatta atgaccgagg atttaaggtt
12241 gatgtggcat tggcggaagc agcgattgaa gcagtcacgc gccataaaga agagttacag
12301 gaagaagccc aacgtaaata tggcggttca ctaaccggca aagacttctt gcctatttta
12361 caggaactag caccggcgca ccgcatccac aacgcacaga agtcgacgct taacgacttg
12421 ctggcagatg aggatttacc ggacgatgcg cgcactatta tcgagatgcg tctcggggct
12481 gcttccaccg catcgacgaa atacgcgccg ttgctgttag gtcgttcttc cgatgaccgc
12541 cgtcgcggtt gtctgcaata cggcggagcg aagcgaacac tgcggtgggc ggggaaaggt
12601 tttcagccgc aaaacctggc gcgcgggtat tatcacgacg atgaactgga taggggtatc
12661 gccgcgttac ttaaaggtcg cgcacatcgc cgcttcgatg tggctaaact aacggcatcg
12721 acggtgcgta gctgcatcat cccggaagca ggacataagt ttgtagttgc cgattactct
12781 aacgtagaag gccgtggttt agcgtggctg gcgggcgaag aaaccgcgct tgatacgttc
12841 cgcgccgggc tggatattta ctgcgtaacg gcaggaaaga tgttcggcat ggacccagat
12901 gatattaaga aagaacgtaa agatttacgc cagattggta aagcctgcct tcaccgccat
12961 acccaagttt tgaccgatgg cggatttaag gatattatgg cggtcacatc gacagataaa
13021 gtatggagtg gtgaaaaatg ggtgaataca aaaggcgcgc accttatggg gtggaaaccg
13081 gtgataaatg tggacggggt gctaatgacc gaagaccaca aaatcttgac gcattcctgg
13141 aagcaggcaa agcaactcgt ttcaaacaaa tatatgatgg accgcgccct ggagataggt
13201 atggacgcct ggttatcctg cgcgagttac caaaacgaca aggccaaaga caactactcg
13261 tccaatgtga ttgtggagag atgcctgggc gggtctatta tgacaatgtc agggcgggta
13321 aaaccactca atgcaaccgc tgcgcccttg aagcaacaaa aaaatatcgt aaacagtatt
13381 tctgctacga agacgcaatg ccggatgatg cgcacagaac gcgattgctc aaccggttgt
13441 ccgcggcaat cgtgcggaca acaagccccg gaaacaagag ttacaaaaac tatggggcca
13501 ggggtatcac ggtatttgac cagtggaggg ccgataaaag gtcgtttctt agatatgttc
13561 aaactttgga aggctgggac gacccaaacc tcgaaatgga tagaatcgac acagacggta
13621 actacgaacc tggaaacatt cggtttgtca gccgttctga aaactgccgc aacagacggc
13681 agatacccga gcttcaacga aagtatgatg cagccgttgc ccgcattgct gaacttgagc
13741 gggaaattaa cctactgcga gccagtttac gacctgattg atgtggaaga cggcaaccgt
13801 tttcttatag cctcagactc gggctttctt gtggcgcata actgcgaact gggcctcggc
13861 tacgaggggg gtgtcggagc gttcgttacg tttgctaaaa acctgggtct tgaccttatt
13921 gagatggcaa aaacaatgga cgggactttc cccgaccaca tctgggctgc tactgcacgt
13981 ggatatgagt gggcgcgtat tcaggaagcc aagagaccac cgcatcccgg tgaaaaggat
14041 gatcgaccct cgtatattct tgacaagaaa gtatggcgca catgtgacgc gatcaagcgt
14101 atgtggcgtg agtcacaccc tgaaacagta gcgttctggc gcgaccttaa agacggaatt
14161 ttagccgctg ttcgtaatcc tggtcgtgaa ttttgggctg gggcacactt gcgcaggaat
14221 ggggaaaggg ctattcgcat atggcgtact gtagaattcg attcgtcggg taggaaggtt
14281 cctggctggt ggttgtgtat ggggttgccg tcaggccgta tcctatcgta tccgggaatc
14341 ggcgttagtg tgacaaagga aacagacgaa gacggacgga taaacaccaa tgtaagaatt
14401 aagtaccagg gtgagaacca gttaacacgg caatggacca ccctgtacac acacggcgga
14461 aaggcttgtg aaaacattgt tcaggcgttg tgccgtgact tattggccta tgcgatgctt
14521 aatgtagaag ccggtgggta tccaatagtt ctttctgttc acgatgaact ggtatgcgag
14581 actccggata ctccggaata cagtgtggct gaactggaaa aattaatgtg tgcattgcca
14641 gaatgggctg agggttttcc tcttgtagcg gaaggtgcgg agttaaaacg gtatgctaag
14701 taaactgatt atttattatt tccgggaaga agactgtcga atctgccctc gctgcggtat
14761 tgaacatacg aaacgcgagg ggtgcatcga tgagtacgcc tgaagggcgc gtgcaaaaat
14821 acgcaaagga gcgatttgag gccctgggtg gcctcgttcg gaaactgtct tatgaagggc
14881 gctccggcgc tcctgacctg ctggtaattc tccccggcgg catcatctgg ttcgtcgagg
14941 tgaagaaaga cgaaaacacg aagcctgacc cgcaccagtt acgggagcat gagcggatgt
15001 gcaaacgcgg ggcaaatgtt tttgttgtcg ggtctaagaa acaggttgat aaattaatag
15061 aacactatta tatttagttc acaccaaaca gaaataagga attgagaaat gaaacacgaa
15121 tatgaccgca agcctgcacg tgacatcgta ccgggtgaca tgattttcaa tgttaagacc
15181 cgccagcctg ttgccgttga tacggtgtac gtcgagtcta atggtaagct ggttattgaa

15241 gacgtaactg gtaatgttac ggcgttcggg cggaaggaat tggttctggt ggcgaaatga
15301 gtaaattcac tagaaggcct tatcaaaagc tcatgacctc gttcatgctg cgacacccgc
15361 gctgtaatat ctgggcaagt atgggcagcg gcaagtgcct gaaacgcggg actgaagtta
15421 ttatgtttga cggtacgacc aagaaagtag aagacgtcat cgtcggtgac gtgcttatgg
15481 gaccagattc aacgcctaga aatgtactgt ctctcggtag aggccgggag atgatgtacg
15541 aggttaaacc gcgaaaaggc gagtcatata cggtaaacga aagccatata ttatcactgc
15601 gaactacttc aggtatcgca aaaggttcat ggccagataa caccgttttc gacataagcg
15661 tccgcgactg gcttaagttg ccgaaatatg taaccgggcc taatgggtat cttaagggtt
15721 ggcgggtacc tgtcgatttt ccaaggaaag agcaggacga ggctttattg cctccgtacc
15781 ttatggggct atggctgggc gacggtactt cgtcgtctgg ggctataaca tcaggagaga
15841 acgaaaagga aataagggcg tatcttgagt cgtacgcagc acgtaatggg atgcaaatac
15901 gaaaagaagg acttacatgg agtatatctc acggcaacac aggtcataag aagcacgggt
15961 ttacgcacgc gctgaaatcg gcgggcgtcc tgaacaataa gcatataccg cataattaca
16021 aatgcggcga ccggagacag cgtttagagt tgctggcagg gcttctggat agtgatggtt
16081 actgcgattt gtctaaagca ggttttgact ggatttcggt aagcgagcga ctcgctgatg
16141 atttttgcta tctatgccgt tctcttggtt tcgccgctta caagaagaaa actcgtaagc
16201 gttgtgcaaa cactgatgtt tgggggggatt acttccgcgt atcggtatca ggcgattttt
16261 ctgatgtgcc tttcgtgcgt ggccgccacc agaatttacc gaagcgtaat attaataaaa
16321 acgtacttaa cgttggaata gaatctataa ccccggtcgg cgtagatgat tattacggct
16381 tcactatcga cggtgaccac cggtttctat taggcgactt caccgtaacg cacaacacaa
16441 gctcgacaat gtgggtgctt aaccggctat tccgtaatgg tcaacttaat gacgacgacc
16501 gtgtgttaat ccttgccccg ttacgtgttg cgtccggtac gtggcctgcg gaacaggctc
16561 gttggaattt cccatgtctg agtgtcatcg atgcaaccgg ttcggagaag cgccgcatag
16621 cggcgctgga gtcagacgct aacgtggtgt gcactaacta tgaagttatc gagtggttga
16681 ttgactacta cggcaaagac gactggcctt ttaccgttat cgttgccgac gagagcacga
16741 aactgaaatc gttccgcagc cgttctggcg ggagcaagcg ggcgagggcg cttagtaagg
16801 tggcgttcgg taaagttaag cggttcatca acctgactgg tacaccgtca ccaaacggcc
16861 taaaagactt atggggccag aactggttta tcgatgcagg tgagcgcctc ggttcttcct
16921 acacggcatt caccgaccgt tggtttaact cggtgcagaa agggaagtcc gctatggcac
16981 gggagtacca cgcccgcccc ggcgcggata acgagattca ccagaaaatg aaggatatca
17041 gtctgaccat tgatgccgcc gagtggttcg gttgtgaagc gccggttatc gtaccggttg
17101 aaattgacct gccgaagaaa gcacgccagg cctacattga tatggaggag aaattattcg
17161 ccgaactgga aagcggggag gttgaagccg ctaacgccgc ggctaagacg tcgaaatgtt
17221 tgcagattgc ttccggtgcc gtgtatgtgt cagggccgga tggtgaagca acgaaagact
17281 gggagaaagt gcacgacgcg aaactcgatg cgttagagtc cattgtcgag gagttgcagg
17341 gtgcgccgct gctggtggcc tatcagttta agcacgaact tgagcgcata cttaagcgat
17401 tccccccaggc gcaggcgttc gcgaaaggta ccaagggtaa taagcagatg gaagcgtgga
17461 accgcgggga aatcgagatt ttgtgcgtgc accctgcatc ggcgggccat ggtttgaatt
17521 tacaggacgg agggcatcat ctggcgttta tttcgcaagg ctggaacctg gagcactatt
17581 tgcaggttgt cgagcgtata ggtcctgtac gccagaaaca ggctggccac gagcgtccag
17641 tgttcctgta tcacatagtc gctaaagaca cgctggatga ggtcgttgcc gcgcgtacgg
17701 acgagaaaaa atctgtccag gaggagttgc ttaattatat gaagagacga ggtaagaaat
17761 gagtagcatc gaaaaagcta tagaagaagc gcggctgccg ttgttagcgc gtatcacgga
17821 actagaagtg gaagttctac gactacggca gcaaagagac gctgctaatg cgcaacttga
17881 ctggctacta gaacaacagg aagttgacaa agagaggggc gctaaatgaa atactcaatt
17941 ggcgcgaaat gctttttaat aggctacgga ttggctgcgg tcgctttact tatcgctgcc
18001 gtggttctgt tatcataaat taaggcccct tacggggcct ttactattta tgccaaagtt
18061 aatcttgtat agcttccgtc ggcccgctta actaacgctt taaggctatc gccctccaga
18121 tacagtgata tagacccgtt atccctgaca gcgttgtctg gtaggacgga catcgctacc
18181 ggtatctcaa cagcacgata atcgcccctg tttatttta atgacacggc atcaggcaag
18241 gaccccgcaa tagcggtaat ttcggtccat gctgaaccag aacccggccc tccgttaagg
18301 tgggagtaca cggacccact atcaagggtc tttgacagtt tatgaatacg aaaccgaagc
18361 gccgcagaac cattattgaa actgtttatg cgcgagttac caagaccttc ctccatcaaa
18421 ttggctacgt taatccttga tgggtccacg aagcccgtta ttccgcttac agtagagtta
18481 ggtgcgtcta tcgtcaggcc ctgttcgctg gtagaccgta ttccgaccaa acgaaggccg
18541 tttacacgac aggccccga gatatatatc tggttcgcgg ggaagttttt agtattggtg
18601 tcgattattg cgatattagt gaatactgtt tcgtgggcaa gaatataagc cccagaacca
18661 gcgcaatcct cgacggttat gttagagaca tacagacctt gcccgtccat accgaaacct
18721 acccccagcg aacctctaac aaatagattg tctatcaaat ggtttaatgg gagcatatgt
18781 accgggtact gagaaatggg gaaatcccct gggcggtcat cttcggggtt catgtcagta

18841 tctgcgccta aatcgaagcc gtcccataca gggtacagta ccgccgaatc ccggaattgc
18901 aggttgtagt tacgagaggt cgtagaacct accgtacctt gccacgtctt aacaccactt
18961 tccccccgcac gatacgaggt aaacccgatg accccgccat cgcgcgcgaa accgccattg
19021 ttgcgtaaga attgcgcgct acttaccgaa ccgtaacttg tgcgcccacc gataacatag
19081 ttacctttgc cccaatctcc gcttaagttt tcgaaggtga ttacgccgtc ttttccgccg
19141 ctcgggttgt cagcgtctac catcttacag aaatggcatc cgcggaacag gaaacacgcc
19201 ataagaccac tcgcccggtg aacctcgacg cctgtacatt cccgtatctc aagaacagac
19261 gaaatgttct ggtctttagc ttccggaggg agaagggatt ctatacccgg gaacttggcg
19321 taatcgttta ccgtaggctg gtatccatct gttttagact gtttaagtgt ggccacgatt
19381 gctgcggggt cggttatcca ctgattatcg tcagtccacg gtttaatcac ccacggcgtt
19441 gtgacactct ccatgtaagc accaactaca acggaccctc taccaagttg tgtgaaaaca
19501 aggttaccgt cgccaataaa cttcgcttta cagtcgatgg ttagggtctt accgccgaag
19561 ttaacggttt cgttattgct gaaattgtag tcacggtcga taagaacgct gtcgacggca
19621 gcatcagcta attgctgtaa agtggtaaaa tccgacaact taacggagta tttaaatttc
19681 ttatccgctt ctgcgcgaaa ggtgatatcc ccgaacgcga cccaagcgga atctgaaatg
19741 ccgcccgtcg taagaggtgt cgacgccgca gggataactt tcggcagggg gccacgccat
19801 gcgtaatagt taccgtcccc gccatcttct ttcggccaaa gaaccgcttt atccgcgtcg
19861 tttacgccga gagtaccccc ggtcgtaaag ttaaaagatg ccggggagaa accagcatca
19921 cgtaagactg cgggaagcgt cttctgcgtc tgcccggtta cctggttagt ggcgtagtcg
19981 atatccgcac caccagctac accgccttgt ttaccggtga tcacctcggc ttcgaaaatc
20041 tggtgttttt tagctatttg taaatcatta agtgacaata catcaccgca accgctggac
20101 atatggctat cctcttagtt aaaaccgtta tcgaatccgc tggagaacgc acgaccgaaa
20161 ggaggcacgt tgtcaaactt atagaaatcc ttgtcatagt taaacccggt tattttgact
20221 gttctgtcat cgccagggtc taccgtagaa acaagaatca tctgagcatt atgccttgct
20281 tcgttgccga atgaaaattc agttttcaaa gcgctattac cagtgtaaat ggcttcttgc
20341 ggcacagacg tcataattac ttgtcggtca tgtgcgccag ggacaacacg aacactttgg
20401 actccgccat cacgtaactt cagaaccaaa taatggtcat ctccggatgt gaacttaacc
20461 ggctgggata gctctacggt aagaccgtta accgcggtta cgtagccgtc gtaagtagac
20521 atacgcgaac ccttaaccac actaacggca cgattcggaa gggcaaaaat accctcttca
20581 gtggcggtaa acgacaccga aattttcttt aaaatgttct tctggtggcg acggtttgcc
20641 gcccagaaag cctgcttata gttgcggatt ccttttgagt cgtaagtctc cgttttaagg
20701 cccccggttt cgggtatggt tatagtctct ttgacgttcg tcttagggtc tatgtacgag
20761 aatttaaggc tgtcaaaaac ctgagaatcg ttaaatgttc tggtccactt ttcagaagta
20821 ccggccttgc ttcggtgggt gaacaccatc tcggggccca tgcgagggcg ctcaaaatcg
20881 agaaggatat ccgccccacg ccgatatggg gtgcagaata tggcgtctgc tatagtacta
20941 actatgtcct gcatggtggt tttatagtca tcaaacgtgt aacaaaattc tccagcctgt
21001 ttgtcgccaa aatacgcttc gacctcgttc tgcacagcaa gtaacttatc catgttgcgt
21061 accgtcaggt ttaaaccccc cacatccgga tcacgcgcca ggcggataag agattgcacg
21121 gcttgcgtat tgggggtcat cgtgtcttca aaaacgccat taccgaggta tttgtacacc
21181 atttcagtag caatcatacg caactgcggt tgctttactt ccgcagcacg gggggtctgc
21241 ttgcgcatcg agtgtactgt tgttctgttg ccgtagtgtg gggtgttgtc acgtgtttgc
21301 ccgtataagt taacgtacgt tatctcatcg actacagacc cctcgaaatt aaagtccagg
21361 tctgttaccc ttctggcgcg cacccgcacc cgtgacacaa ccggcaggtc ggcgtaaata
21421 gtgacgccgt tgtagtccgg tgtacgcccg gacacagttc cttgcgcagt atatatcggt
21481 ccgtaaggag tgctgttttc atcgagcaat tgatactgta tttcggcggt tacagacgcc
21541 agtgttttcg atttgccgtt gtctttatac atgccgttcg cggcggcgat attggcgact
21601 atacgctcaa cctcggtacg gtttattgaa acccaatcgg ttaatgtttt ctcatacgcg
21661 ttgtctggtt gcagcgaggc ctctccatca ccgcggaaaa atgaaccggg tcttatttct
21721 tgccaccgcc cgataaggcc gccaggcacg ataaaggtaa cggtcccttc ggacacgctg
21781 actacctgat atccgccgcc ggagaggtca acgtctgtat ttgtccacgc cgacaccagc
21841 gaaaatgtat cattaggcac cagtacccca gaaaaatccg aaccacctgt tgggtcatat
21901 atcgtgccaa tgtttccgga caaagacgga tacccggaca tgtaagagaa agaaatgccc
21961 agaccgttcg gtgctttaag aaccacgccg tcaacttcgt tagattctac ggtaatgtat
22021 aggccttgct ctatggggtc cccgaccatg acctggggtg tagatgtgtt attcggcgat
22081 gtatatggtg cgtacacggc aaccgacgtg cctgttatgt ctgatacgcg ggtgtccccg
22141 tcagttatac cttcagggtg tatgtcgagg tatccacgcc cggcgtcgta atagccgtac
22201 tctacaattt taccggcggc gttaaacacc ttgtaagtag acataaggtt attggggatg
22261 gtttgcaccg ttccacagat gtcgtaactg cgttcgtacg ggcgcgcctt gttgttacgg
22321 tcggtgagac tgttgttcgg ggaatccgcc tgcgagttcg caaggttaga tgacgcccct
22381 ttcacggatg gtgataatag tttcgctagc ggtttaagga taacgctaaa aatcttcatg

22441 acgcctttaa tagcgccacc accggcagat tccacaatgt ggaaagtcgc gttttctttc
22501 agtgcttcga agtcttctgt tacgtcgtta tcttccccga tttcgtcgat gaaaacgcga
22561 accggcaccc catcgggaac gtggttcacg acgaagttca tcggattatc gcgataacgt
22621 ttaacgtcaa atgtgccgtc ttcgttacgg gtatagtgga ttactagcgc caaaattcaa
22681 tctccgaata agtatcttta aggtcttcca gactatccag tctaacctgt ctggacgcca
22741 gctcacagtg gctaaccatt ccttcgtaat ataccccagc gtgccacact attcgccctc
22801 gatgtttaac accgagtaga acagcgtcaa aattctgcgg agtaaccact cgcctaagac
22861 ctttagggtc tgaatgtccg tcatcgaaag ctgcatctat cgccgttggg cttgtaacat
22921 cgaacatcgg ggttgataac cccgcgtccg cgcgcacatt gcgtacatga tgccagcaat
22981 tacgacggcg gaagtcatac ggcagaccgg tgtagtcatt aatattcatg acgtcagtat
23041 cccgcgaagt aatggaatct cttccggggt catcagaatg cctgttgccc gttggttaag
23101 catcggtgtg ccagtttccg ccgtaaatac acctttttct tgtgtgagtg tttgcaactc
23161 ataaaccacc gggccgtcgc atggatatga caggtcggtg ctcacataac gccggaagat
23221 gaatttaggt aattccttat tgtcgtaagg gattctgtcc atttcttcat cgagaatgtt
23281 gagcacatcc ggcaaagaga aagaggcggt ctggtccata tcgttattgt tagccgcgtt
23341 cttggcctcc atcggcgtac cctcgaacgt cactacttcc ccggtttcaa gagttgcggt
23401 taagtcgtta accccgcgca caagcaacca gcgcttcgaa agaagagggt gcgatatctc
23461 cagggtaata aaatccatct cgccgtcggg gtttgatgct agcttgcgac ggtatgccgc
23521 ttctactgat tcctgactca ttgcattggg tcccaaattc gagggaatgt agtttggtac
23581 actccgtaag tttttagaaa ttcgccaaga caatcgccgt agcatccata taagtcgggc
23641 agattctggt taagacaggc gctgccgact tcctggaatg gtgatttctc ggctgtggcc
23701 gtgaaagtta ttacccagtt cttcccgtcg tcggtggact cgttccacgt agatgttaga
23761 gtaacctggt aatcctcgat accctgaccc aagtcgtgtt tcatccagaa actggaagcc
23821 cctccgtcta ccttctccat gaaactgagg aatgcttgtc tacccagtgg cgacacgacc
23881 agcgtaacgc taatcgggaa cacatcaaag tatgtatcac gcccctggcg aaccccgcca
23941 ccggccaggt ctactcgcca cacgttgtta cctcgcgtca tcgagtagcc tttcgatact
24001 atcggcctca aagaggccgg gaaatggtaa tcgctcatat caatatcctg gttggccgcg
24061 tgtagcgcga cgtgatttag aaatggcgct gttactatcc agcaacgctg aacttacggt
24121 ttcactgatt ataatacgca accggccttc atcgtcgcgt tctgttgcag cagaatcaat
24181 tcttccagtg gtattgttca caatagtaac attatccccg ccggatttag cgtcgttctc
24241 gcccataatc tgtcgcatct gctccgccgt ccggacacgg gaggcgctag cgggcatgat
24301 tacttctggt ttaccacgtt ctgcaatagt ggatatctgc cctgcggcca ggttaccacc
24361 ctgttctcgg gcggagcgaa tcttacctac gtttgctaaa ccagccgcca ccgcagcagc
24421 cgcagcaaca ggcgccagat acgggccgac taccggtatg gctgctgttg atttatacgc
24481 ttctactgct gacgtgtacg tagcgattgt cgcctgtgcg atggcaaatg ctttgtatgc
24541 cgtggaagcc tctcctagcg cggccccgat gttactagcc atattaccaa atgcggtgcc
24601 tgtcgccttc gctcggtcta atgcgtaagt atcattcatt gcgtttaggg tctgctggta
24661 tgtttcttcc gcgattaaac cctgcgcatg aaagtcgtta agtttttgta gcttgacctc
24721 gtactgacgg tctaactcgg cgaattcccc ttcggcaagg gcctgcatct gagcgacgta
24781 ggcatctgcg gagaattgtt tttccaggcg ctctttctcg cgtttatcta actcagcctg
24841 gcgggccgta tccgcctcga gcataatctg cgtcttagcg gtttcgtatt gctggtcact
24901 aaggagacgg ttgctgaaga actcctgtag cttagtaagt tgctcttctt ggttagcttg
24961 tatcttctgg aatacgtcat cgttaaggtg gattgactgg taatagaaat cctcagcggc
25021 tttcttctgg cgttcgaatg cgtcgacgga tttcttagcg gccttgtctg cggaatctcc
25081 tgtagaaccg ttatctttac ccgacacttt aaacttacca agaccctcgg ccttggccgc
25141 ttctttatcg atgtcgtaag ccgcacgctt caatgcgata aggtcttcct gagcttctac
25201 ctgttcacgg aaacttttac gtgtcaggtc agtcacagct ctttgagtat ccatagcaag
25261 cgctaactgg aagttaaggt tcttggcctc agctaccgac tctcccggcg tcaatgcggc
25321 ctttacaaga cgtcctgcat cctgagcacc tttaaccata tcttcaaaaa accgctccat
25381 tgccgtaaat tcctgggcaa tgaagtccag cgcatccgcc gctgactcag ctatcgtaga
25441 tgcggtctcc tctccgcttc cggtaatgct ttcactggtg taagtccagg cggagttcaa
25501 agcgtttaac gcgtctgcga ccatcccaaa cgccgcgtcg aatgaagcct ctatcgcacc
25561 cccaacatcg gatgccatgt caccccacat ccgcatctcg gctacgaact caccggatgc
25621 gacctgagcg tttatctcct gtatcaggtc gtcaatataa cgtaaaggtt ccgccagcga
25681 ccctacatcc agactagtag cgagagtcat cttcaactgg gaccatgagt cctccgcact
25741 agcgatagct ccgttcaggg tattagcctg gtcagccatc gccccggcga agttaacgtt
25801 gcctatgttg aggagatatt gctcaatatc ggcagcgttt ttcttaacta cagtagtcgt
25861 cccctggaag gtaaattcga tatcactttg gttctgcttg gctttaatgc cgaattcttt
25921 cagacgttcg aattcaaagg tgctcgcatc tgcgacggcc tcaatcatct ggttaaggtc
25981 tttacccatc gctgatgcag tattgccgta agagcgtagc gcttcttctg atggggtgag

26041 gcctagagct accagtttac ggaacccttc cactgcttgt tcaaggccgt aaggtgtgtc
26101 acgggcgaag tcctgtaaaa tactcagggc ctgttttgcg ccttgagtac tgcctgtgag
26161 tgttttcagg ctagccgaca tcttatccag ttgccgttgt gactcgacta gttcctgcgc
26221 cgccttatag gccgtagcgg ctacaacggc agccagtccg gtgaccgcgg cccccgcaac
26281 ttttgcagat ttagagaggt tatcaagacg gtctgatgtg gccttcgcgc cttgttcggt
26341 taccttttact actaagctag ctacatcagc catcgtttct accctcaaaa atggcgtcta
26401 agcccatgat aatttcagat tcaaacatcc ctatattctg acctgaaata gtgctatacg
26461 ccactaaatc tgaccaacta agttgctctc ttgggtatag ctttacgccg tcgtcatcga
26521 cgcggcgggt aaatttagca tcacgatact tctcaaaagt ggttagtagt tcaggggggc
26581 attcaggccc ggcgtcctgc gttgtttctc ccgcgtcttt tattaccccc atagcgataa
26641 gcgcggcttt gtgcccatcg gcgatactgt caaacttgcg cctttcgtgg cgggttatga
26701 agttccaacg ggcaaactga aacaacgcgt ctacttttcc tgcaattgac gcaactgttc
26761 gtggtggaat acgactacat gttccgccag gcgcttatat tgggcgagaa gagtcttaag
26821 attttcctta gtgaactcat catcgagact ccagccattc accaattcga gagccaattg
26881 tcggtttagg tctcctgccg cgtcttccat ttttaaattg tattctgaga agtctttttg
26941 ctcttcgcat ttatctcgaa gcggttttaa tttacctacc gccgctcggt acgccagggt
27001 gaaggcacgc atagctttaa ctgcgacatc ggcctccggg gagacaacgt tcaaccattc
27061 tcctgaatct gtaccatctt ttaaagggat cggcatgcgc gcgcctttct cggcttcggc
27121 ttcgtagtaa aaatcgctaa gtttcatagt aatgcccttt ggttggtcgg ttactggttg
27181 gtagttgcgc caggcgggaa ccacccgctt ttccggtgcg accgtagacg ccaaataaga
27241 ataacaaaat agcttgcata accgttaata gtgtcctatt atctctacgt caactaaaca
27301 atatgaggtt aagaacatga gcgtttttat cggtattatc gcgtttatct atttcattcc
27361 gttcctggta gccttgctac gtaatcataa agcgaaatta ggaatattcg tatgtaatct
27421 tttactgggg tggattcttc tgccgtgggt tggggctctc atctgggcct gtaactctaa
27481 tgttaatggt aaataatctg gctaaattaa acttacatta ttagaatcag atttgtctta
27541 aacaatcagg aaataagcat ggtcactcgc aaaataacag atgaacagtt gcagcaagag
27601 ctaaacgctg gtctcgggcc gaccgaaatc gcaaagaagt acaacatgtc acgccgtaat
27661 gttcagctaa ggtcagctcg tctggccaag aaaggtgttg gtcacggtcg tgacgtaagc
27721 catctggtgc cggatgggta taagattaag ggtacgtcgt cactggtgga tgagttcggc
27781 aatacgaaac ttcagtgggt taagactgac gctgatgccg agcgccaggt cgagttgatg
27841 cgagccgtaa tagatggaat gaagtccgat attacgccag tttcttcggt ccctcggcct
27901 aaaaagcgac taaacgaaaa gttgctaaat ctttacacgg tttctgactt ccatttaggt
27961 atgttggcct gggcggatga gagtggcgac gactgggata tgaagattgc agaagacctg
28021 ttctcgagat ggttcgacgc ggcgtttcag aaggcgcctg atgccggtgt gggtgttatt
28081 aacctattag gggactttgc gcattttgac agccttgatg ccgttacgcc tgccagtggc
28141 catgtactgg acgcggacac gcgttaccag aaactggtgc gctacatgat tcgtatggtc
28201 aggcgcgtgg ttaatatggc gcttgttaag cataaaaatg ttcatctcct aattgtgcaa
28261 ggaaaccatg acgagtcagg catgatttgg ttggctgaaa tgtttaacac tctttacgat
28321 aatgaacctc gtgtttttgt agatacatcg gctgatgtct acaagatggt ccaacacggc
28381 aaaacgaccc ttttctttca ccatgggcat aaggcgagat tcgatgctat cgaaccagtt
28441 atgatcgcca agttccgtaa ggcgttcgga gagagcgttt acagttacgc ccatgtgggg
28501 caccttcacc atcagaagat tgtggaaagc cgtaacatga ttgttgaaca acaccgcact
28561 ctcgcggcga aagatgctta cgcatcacgt ggtggatgga tgtcaggccg cagcgcgaat
28621 gttattactt acagtgccga atacggcgaa gtcgcgcgtt taactatttc accggagatg
28681 ctgggatgac taacaaatac aatcgcacaa tgacaaatac tgacggagat agcattacct
28741 gtgatgtgta cgacgttctg agggcatttg atatccgcga cccggcacta cagcatgcgt
28801 tgaagaaact gttgtgtatg ggcttgcggg ggcacaagga cacaggaacc gacttagcag
28861 aagcaattga aagtctggag aagttacgga aataccgtag taatattgat gagtgagaaa
28921 aaggcccctt tcggggcctt ttttttatcg tcacgataaa gaatctgatt tgaacgataa
28981 gaatgttcga agttaaatat atctacacca acaagagggg atgaaaccat gaacgcattg
29041 attagtgtta gtgaagcaca aaccatgtcc agccgtgaga ttgcggaaaa aaaaaaaaac
29101 ggggaaagaa cacgacaacg tacgacgcga tatattaaaa atggcacagg aactttccct
29161 caattttgag gaaaaggttt tgccttcaaa cggtggcaga ccaagtaagg tatttctctt
29221 agacaaagaa aatacattaa tcctgatttc agggtacagc atcaaaatgc gcgctgccat
29281 catccgccgt tggcaggaac tggagtcaca agcgagcaaa ccatccttac cggtgcccaa
29341 gacaataaaa aaaaaaggag gctttaaggt tggccgcgga cttgtgggag gagaaagaac
29401 gccttgcgct tgagaacaag gaaatggcgc caaaagctga tgtctacgac cgcatcatcg
29461 acagaaataa cttgtacaac gcgacgcagg tggctcagaa gtttggtcaa tctgccgtgt
29521 ggatgaataa acaacttgaa caatttggcg tatacaatcg ttccgttaag cgaggacgtg
29581 tctttcagca atggtttgtg gataaaggct acggcatcat gagggaaaca gaaactgggc

29641 attcacaggc catgttcttt gcagaaggtg aaatgtggat tatcggtaag ttaacggcag
29701 aaggcttgat ttaaaactaa ggcccctttc ggggcctttt tcttaggtgt acttaatgcg
29761 ctggataaca attgaagact ggaactggtt accggttgcc tgtccctcaa gagaaagggt
29821 gacagattcg ggaccaccga tttccggagt tgctgatgtc aacgaagcgc gttttaaggt
29881 gaaactcatg gcaccgtctg ggccgttcag aacagaagat acctccattt gcgtttcgtt
29941 caggaattta ttcaataacg tcatgtcgta cagcttaccc gccaacgaga aggtgttagc
30001 cgcgcggcca cgttcgacaa atgccacgct gttatttcct aactcgaact gtgcggacgc
30061 gctgttgtcg ttagtaatgg tgaacgtgtc gataagttta agcggcgcgg ctccgtcata
30121 agcggataca tcaaccgatg caaacgggcg ggcgctaaaa ttggtcgtga aatctgaacc
30181 gctcggcggt gtttgtagga tttcctgatt caagccaatg aacgggaaag agccggtcac
30241 catcgcgtta actgcctgtt cgatagtgaa cccggtaaac tcgactccgc gggttatgat
30301 gtatgaatcc gggtttccgc atttaccttt caaccatgtc aagattgaat aggttttaca
30361 caagttacca gtttcaagtt tatctgcgat acgcaaatcg gcttgggcgt ctgattcggc
30421 ggtaagggta tgctggatac cggcaccggt tacaaccgta gcagttactg cggtaacgag
30481 gaaagctttg tcattattgc cagacaggcc gtcgaattgc accaggtcgc ctacttcaac
30541 gccatcggtt acaaagctac cagtagcgcg tgtgaaagtt ttcactgctg ggtctacggt
30601 aacgctgatt ccagatttag tggaacctgc tacccaggaa ctggtcattg cgcccgcaag
30661 caactcatcc tggcttgtcg cgcttagttc gatagcgtat tcaccagtta cctgacggtt
30721 accggtgcgg atagatgatg tttcgcggct gccgtccagc tcgttagaga tgagggcgtc
30781 gcgcgttacg gctggaatgc cgccggtgtt acggagcggc tgccataccg gattatcggg
30841 cgtcacaccc ggcgtcgtct caagaacgta aaattgcgcg gtcatcgcgc ccttatatgg
30901 ttgtaacgcc attatctaat cctcgctgta aaggctatga aattaattga gagaggcctc
30961 ttcgcccatc cattctcaac aataagagga ccgaggctta ctgattgtac ctcagcacag
31021 atttcgttac gactaaaaca gttaccggcg gcgaaggtgg cgtttagctt atcagctaac
31081 cggttgagag gggcgctacc caaggtagac ccaacgttaa tgtccacctg ataaatacct
31141 gcccgttgct cggtccagaa caggtcagct tgttctgtat cagacaacag catataactc
31201 gccaaatatg gtacatccgt agacgtcggt gcgtctatgt tctctaacgc aaccatgatg
31261 tcgttgtcgg tgccaaatgc cgccagtgcc gtgtcgaatg ctttcgttaa gtcctcaaaa
31321 tatgttgcca ttatttcacc ttagaggctt cttcatttaa tagttgctgg aaacgcgata
31381 tgttagtcct taccatgccc tgcggggcct gttgtgacca accgtactcc agacgttgcg
31441 catacggcaa attgttagta agagtgaagg tatgccagtc cgcggcgttt agcacaaaac
31501 tagtagcatt accggttgcc gtgttgccgg atttatcggt ggcgtccgta gttccatcgg
31561 cgggggtgct tcctgacgcc atccagttca tacgaaatcg cccggtatct acgggactcg
31621 ctttgataat cgcagaaaac aactttatag aaacctgacg cattaccttc tcgggattct
31681 tcttagcctt ttccacgaac ttggacacat caagcgcaaa actcattttc tcacctgaat
31741 aaagtatgcc acaacatcat cgttaaccat cttcttctca atagcaacta cagaccattg
31801 ttcaccgcta aactgcacct tgtcatccat cttcggaact acgctgtaat ctgctttaac
31861 catcatgtca cccgcctgta tagtagttcc gtttaccagt ccagcgttaa ccggtacagg
31921 aaccgcttta aggggaagta cgacgtcggg cccccataca tactccccta aaacagggtc
31981 ccatgttttc tgccctttgc gcaccaaatt cactgtgcta ccgtatttcg atagtagacg
32041 agtacctacg ccttgcatcc gtttactaaa agcggtgctc atcacacacc ctccagtctt
32101 gaaacgacca gtagtgcgga tggagctact ccccaagaag taaccgttgc cgcttgcggg
32161 tacagtcctc cgaagttaga accggcggcg tcgcgcataa tctctaccgt aaaagtctgc
32221 ccggctgcgg cattcacaac cacacgagat tcaatgggga ttgtggattc cgcactcgcc
32281 agtttagttg cggctgttga tccgaattgc gcacctccga gaagaacacg cgacagaagg
32341 atggatgttc cgctcgcccc ggtgcgccca gcctgcaatt taacacggat agcgtagtta
32401 ccagcggcgt tgaacgtaac caccccggag gcattaatca ttaccgggtc ggcgcttgtt
32461 ttctgagcgc caccaaaagc tactgttaac gctgtaccag tggtagatgg cgcctggata
32521 tctgtagacg acgcgcgaag tacctctact tctttagctc tgtacaacat agagtccgcc
32581 atctgagtgg taacttctcg caacttttca ggcgtaatta ggccggactg gttatcaggg
32641 aagcttgccc caatcaaagc gaatatttcc gatttagtct tagccatgtt tagcccctaa
32701 aaacattaaa ggagtaggca ttattgttcc cacataagat agggcgtagc gcatcatcgg
32761 ctgccgtgat gcttacggtg ccacctgagt agccattctt aaagtaggac acggtcaccg
32821 cgccttctac ccgctccatc tgcacctcgc gcccgtctgt ggtccctcta acgtccgtac
32881 ccgcgccgta ctcaaccgcg gccattacct gagcttgaat aactaacgat ggaattacgt
32941 tggatggctg agggaacccg tgcagggtaa cgcctgttct cgggaaagcc agtgactggt
33001 ttgctgaaac gcgacgcccg cacatctgag actcaaaaag acccacgtat acagcgccat
33061 tacggagaga ggcctccgct gcgatgtcat cttctggtag ttcgagaccg tacttagcgg
33121 ccagcgcccg cccgtcttcc aggctaacat aagagtcggc atttggcagg ccttgaccgg
33181 tttccacgat aagcggcata aattattcct ctacactttt acggcgacgt cgttgcttag

33241 gctccctgcc accgttgtaa tgtggttctt ctgcgaccgg ttgggtctcg attaatgctt
33301 ctgggtttac ctcttcgcgc gtcggcatta actgcccatc aacctcaacc acgcctttgt
33361 attgttcccg tactacataa ttatctgcca tgtctaatcc ttaaagcggc ccggaggccg
33421 cgcatattta acttacggtt actactgtgc tatctgagat aatgttaccg tacccgtcat
33481 gtgctacgac tttataggtg ccagaatccg ctgcggcgga gttggccttc gtataagtca
33541 aaccggttgc gccggagatg gcgttattat ccttatacca ttgaaccgtt ttagggccca
33601 ttccgtcggc aagcgctaca gtcagggtga tagcctgccc ggctgttacc gcggtggtat
33661 catcgagacc tgtagagaaa cgcaacggcg tcacgttttc catatcaatc tcaacctgcc
33721 catcctgcgg gctgtcatca gataccccaa caatacgacg tttaattaca tcaaccattt
33781 ttttatcctc ctcaggatac agtaccagcg tttttaaggg cggttaacag attagtaacg
33841 gcggtgcgca atgaagtcac atccgtacgc aacttattgt agttggtgac cagtgcgtca
33901 aactcttctt tcgtcggtgc ggccgcggct gctgcaccgc tccccgcggt aatcgctgga
33961 ggtgccgcta ccgttgcgga ctttttaacc cctccgattg ccgttgtggt tgcagcggga
34021 atatctagtt ccgcgttaac caggctgcgg ggtagccctt tacctgtctt tgacattata
34081 tagcctccgt ataaatgaaa agagggaccg aagtccctca agagtatacc ctacgattaa
34141 gcaccgacgc cagttaccag gaacgcaatc ggcacatgct tacggtctac tacacggttc
34201 cagttggtgg cgttagccag gtcctgccag ctagcggaac gtgcgatagt ctcggtaccg
34261 ttgccggtga ttacggcgct ggtgaaactg tatccgaacg ggtgcagcaa ccaggtctta
34321 cgggtccaca gggtttcaac accaccacca ttggcacgag atgcttcacg ctcgtattcc
34381 aacggcatct ctgggtttcc ttcgccgtag ccgatagcgc cctggccgaa gatgatggag
34441 atgaacttgc gctgcgcacc ctggcctact acagtcatgc tgtcgtcgac gatgacgcga
34501 tagccttggt aggtggcgaa catggtgtcg ttctcagcat cacggatgaa gtcgataagc
34561 tgggctttac gggcctgtgc gtatacgaag ctgtgcattg cgatagcacc gagaacctca
34621 ccgccgttac ccatcaatgc gtcacccata gtctgagtag catcaatgaa agcgccagcg
34681 tcgaagccaa gggtagcgga aacgtcaaca accatgtcgt tctgctcgtg atacgcatca
34741 gtagcggata cgttgtcgtt gtacaggccg agtgcagttg caataaggcg acgctgcgcc
34801 tgacgctgcc agaagttatc cagacgggat gcgacggact gcaatggatt ctgactggtc
34861 aattctacgg tcaggtcagc ctggccaaaa ccttcgttca ggtaggcgac gcgggccatc
34921 atttcaccgg tctggatagc gcgcggcgta gcgatatcct ggtacacatc gttcgaatag
34981 ttaggttcga tagaggtgtc gatagctttc cagaacggca ggttagcgat gttggatgga
35041 ccattagcaa tctcggcggc atacggagtt ggagtcaaaa tgccggagtt gaaaaacgct
35101 gttttctcta ccgggtcttc cgtcatgtag gacgccagta cagggatgtt gccagttacg
35161 atgtcaccga tagtagtaat tgccattatt tatttcctca gggctttaag ttgccgttca
35221 aattcggcag ggtttgattt atagagagcc aatcgctcac tttcactcat atctttaaac
35281 gctggtgcgg ccccgccgcc tttattcccg gaagccccgc caccggaagc tgcatttgct
35341 ttaatcaaat gtgaaaaagc tttatgttcg cgcagatatt tgcggaactg ctcagggtcg
35401 gtagtaacaa cattaccatc cgcacccatg aatttagtca ctacgtcttc gccgtcgaac
35461 tcggtcttaa cgaaaggtgc caggatttct accgcttctg gggcgataaa atcaccagcg
35521 aaagaaccta aaacagcctt acgttcactg tcaagaatac gcgcagctaa agatgcgtaa
35581 cgctcttctt tctctttcag tacaggttca tactgactac gaatcgtctt ttcaaattcg
35641 tccattttac cggcagcctt cagcgcttcc tgatgtgcac gctgccgttc ttcttcggct
35701 tctttagcac ggcgtgcagc ttccttttc tcgctaagga gcgcttcctg gtttgcctta
35761 agccccgcca cttctttctc aatcagggcc tgcacttctt cggcggtgta cattttggt
35821 gcgtcacctg cgcccggttt atcttccacc ccggcttctt catgaagcgg gtaacgtaaa
35881 aaacgattca tagtcaatat gtcccctgga ctttaggaca ccgggccacc cggtgtttct
35941 gtaacaagaa taaattattc cacttactaa ggcaactatt ccagaatatt ccttatgtac
36001 tcctgcaacc ggaatacttt caagcgaagt tgccgtgtgc actcagcgtt ctgtacgtcg
36061 atggccaggt cttcgtcggc atcactactt ggcaatgcca ttttgcacgg cggctgcatt
36121 agtgtcatat ccggggatgg aattagcatt ctcggcggcg cgccgtttga ttgacacccg
36181 ctcaggaggg aacacgcaaa caacacgacc aggacgggta acgtatttga taacttcacg
36241 ggtgatagcc tccgcgtttt ctttaccttc ggtttcggcg gcggcggccc tggtatcgtt
36301 ctgccgttgc cgttgtgttt tcttcgctag ctcggcctgt gctttctgct gttgctgtga
36361 aacaagattg gcgcggcctt ctacccaccc gcttctgtag ttgtactggc cgtatagcca
36421 caaggagagg ataccggctg ccgtggctat cgctgttact ttccaattca tagtatcccc
36481 tttctttata cgctaagccc ggcgtaagcc gggcgttaaa ctcgtatatc tgtaaacctc
36541 atacggcgaa tcttagctaa atactgaaag cagtatagct cggcccttcc gaatgcgttc
36601 gcggttacgc tagggcgttg tgacgcccgc ttcacgcgac ccgcaccgcg gtacgctgaa
36661 agcgtaggta gcgtaacata aaaaataata aaagtctatt tttggaataa agtattcctg
36721 aatttacatt ggtaaggaat aactggaata atggtgtgta aaattaaggc cccttgacgg
36781 ggcctttcgt tacttaggtt ctgagacctt agccgcttta cttaccaccg atatcccgtc

36841 agcatctacc gccctgacga aatactcccc cgggtcttta acagttagtt ccccttcaac
36901 ataagggacg ttgattacct ctttaccatt tttaaaccat tgcaaatcgt agggcgcttt
36961 gccacctttc acgactacgg ttagtttagc ggaaccgtct ttcaattcta catctgctgg
37021 ctgtgtgtcg aagtacacat cacctgctcc gtccagatac ggcacttcgt acaggatacc
37081 gtcagggatt gctgccagtc cggtcttatc agcgaacggc atttcatcta ccgactcacc
37141 aagaacagat tcatctttaa tgtacacaac accttcgccc gcgccggaca cacgggagta
37201 ctgaactaca cgtcgtgaag gcacatcttt aactttgaaa aagcccatct ctattcctct
37261 ttcagataat cggcaacacg cttatcgagt tccgccatct ctttaagagt taacggacgc
37321 ccgaatccat ctaccgatat tacacgaaat tcttccgggg ttatcccact attacgaaaa
37381 atcttacccc tgacaggacc gagtgcttca tcctgaaacc atgccggttg ttgtttaagg
37441 aactcgtagt aagtggtgtc tgcgcttacc tgctgacctc catctgcgcc ccgggccgcc
37501 cgtttcgcgc ccttatcaag gaagtcaaat tcggaactaa ttaccggtgc tgtgcttgac
37561 cgacagttag ggtgcgcggg cggtagcgga cctttaccta tttcgtattg catagaatcc
37621 cgcgaccgac aaacggctga ggttctggag tcaagggtag atacccactc gtatttcgtg
37681 atgatgtcgc cgttctgagc gtaaacctgt tgccgtgctt cgttggatac gtgcgccagc
37741 gctgtccgga taacggttgc agcgttacgc tcggagatat cggccagacc accgggccca
37801 acaacgttct taactatctg ccgcgtggtc ataccctgta cgaacccaga cttaaccccc
37861 atcaccaggc gattcacctc ggtttgccgc cagccattca ttagcgaaac aaaatcaacc
37921 ggtttttcgc ttagttcaag tggcgcgaaa gtcgcagcgg cccacacctg ctccgcagcc
37981 ggtgttacaa agttagcatt aacgttggtc gacaatgttt ttacattcca gttaacctca
38041 tagtcggaaa gctcacgggc gtcggccagt agcttctcat accacccgga cgtaattccg
38101 ttcagtgctt gctcaagttc ccgcagcatg atggtgagtc tcgctgccgt tctgctgtca
38161 tcgccaaaca gcagcacctg tcgcttaacc tcatcacgca tttgcgagat aaatggcgcc
38221 aggtctttta cttcgctact ggcgttgcgt tgcagccata tttgatggct gataatcgat
38281 tctaaaagac tcatatgccc ctcctgcctt tggtgtggtg attatacact caactcatca
38341 gggtgcataa tcaaggccat ttagtctgcg agaaataatc atattcgctt tcattttggt
38401 agttaaaggt tgcataacca caaaaattag gaatataatt taaacgatac aattaagagg
38461 agaaaaacag atgaatgtgc agactttaga atttccgtat ggcgaggtgg aatttacacg
38521 tgatgagtat atagatgcct caacattaac catccagatg aacagatgga ggttcgataa
38581 cggactccaa ccaagaaatc ttgcgcagat actaaaaacc gatgactgca aaaactttat
38641 gcgtgtgtgt gaggaggaga caggtattgt tccgctaaaa acatcaagag ggaggaacgg
38701 gaagacgtgg ttgtgcctgc acctggccgt gtacattgca gaacagtata gtgcgtattt
38761 tcacttcctg gttatagacc gatttattac gcaacggcag gtagagctgc gtaacattgg
38821 tgcggtaagc ttcgtggagc ttaatgccgc cgttagccgg atgctcgagc gcacagaagg
38881 gcggataggc catagtggac atttcataca ggtagcgaaa gcaatcaagg agtctataga
38941 tattcgtgaa gtcacaggtt ttgatacatg ggactcccag gacgcaaaaa ctaatcaatt
39001 gcggagcgag attcaaaagt ctatggtgac gttgttggat atggaagcag tcaactcgtg
39061 ggatgagctt aaagagacaa taccgcgtgt tgttcgcaaa tgcgcggcaa atatatgtta
39121 aaactaaggc cccgtcatgg ggccttttgt tcacttctct tgttgttgcg ccgattgagg
39181 gatttccccc tgaacctcag tggcgaccgg taatggctgg tcagccacag cgtccttaat
39241 atcagcgtct gtccagtcgg tgacaccggc cttacgcaac gcggcatagt aggccgtggc
39301 aggcagcagc ccggcgttaa tgtccgccat ccatgcagcc cggtcttgcg cggtcatagg
39361 ttgcaggaag aaatccatgt taaggcggaa ctcgacttcc gtatcttcag gtttaccgag
39421 cattatcgcc acccaacgta aagcatctgt atacgcctgg cttacgttac gtgcgattgt
39481 tgccatgacg gacgtatccg cgccgcgttg gatgcgtgcc gattctgcgg taatttgctg
39541 agtcggggtg ataagctgcg caccaatctg tatcgcctgc tgctctttgt ccagcatatt
39601 ctggcgggcc aggttattct cacccgcctg aataagctga gcgctaccgc catatcccag
39661 gttatgccca caccgactac caaatttgat accgttaggg ttcgcttcct tgaatgattg
39721 aggtgttaag ttctcgcctg ggtagataaa cagggtcggc tgcccaacaa cgaagctgga
39781 ttcttcgttg tccgcgctgt tcctgtagtg cccgatgtta agttccgcga gtggtagtag
39841 cggagcgtcg tcaatggtgg cgtcgttatt ggtcgcccca ataaaggtaa acggaattac
39901 ccctcgtaat gattccccta agtccgggta aatctccacg acgtcttcct gcgcaccacc
39961 ttcggcatcg aaacgaaaaa gacgttgacg gtagttgtcg tcggtatcga tgtccaggac
40021 ccggtactgt tccccatatt ttgtttcgaa ctcgtttccc ggttcgttgt actcccatgt
40081 ctcgcgcagc acaaccatag ttacccggtt tacagaacct acgcgcgtga gtcgccagtt
40141 aacgatattc tcggtagtgt agaaagcgat tgtagggtta agcaagcccg cgttctgttc
40201 ggcggccgtg gctgcgccgg tttccggggc gtcaacaaga agaccaccgc gacctactga
40261 gtcaatctcc atgagcgtgt cctgcgcatg ctgtattagg cctacaccag acccatctgc
40321 atttttaagc aagtattcca gctccttcgg gatattgatt tccggttctt tacgcatgac
40381 gctaccaacc atcccggaca gcgtgcgacg cgtgaagttg tatacgatgc cgccagcctc

```
40441 gtattcggcc tggcgcgctt cgccataggc tttgtccggt tcgttcagcc ctacgttgcg
40501 tagatagccc accagctcac ccgcgagcgc gtggcgcacc ttctgccatt tcggtgcgta
40561 gtgcagccat tcacgatgct tggtttttac gccagaaccc tggacgtttg ctgttaacat
40621 ttaataatcc tcttaaagtg cgaaagtaac cgggatatga gatacgggtt taaccaacgg
40681 catttcgtag gcgatagggt atcctgttgc atcgttctgg tggtcgttgc cgcttgtctt
40741 atccggtatg ccgtttttat cgtaagcctg ttgctccagg caacgtgcag taacagggca
40801 agtctgttcg ttgaccatta atttaccaga ctccagcgct ttattcaccg atgctacacg
40861 gtctttaacc gcagggttaa ctgatttcgc acgtatctcg aaacccgcgt tctgtaattg
40921 ggcaatatct gacgtactgg catcagtcga tttacggttc ttgccgctgg cgtccgggta
40981 cacgacgatg tggtggccgt gtcgtttcca ccgctcggta atctccctga ccacatccgg
41041 ggtgtcgaac atatccacca gctcggctac cgcgtgccag acatactcac gctgaacgta
41101 tacggtgcta gccatatgcc cgacgttgaa gtcctgaccg atatacagcg tttccccagg
41161 ttgtatagtc tctcgactgc tgttcttgcg tcggtcgtaa gcataataca cgctgccgga
41221 tgtcaggttt acaaattcac catcgatgta ggcatcaatt aactggcctg gatacgtatc
41281 ccgcaacgac tgcacataat cttccggtag aaagggattg gatgttgtcg aggcctgaat
41341 catttcatag cctggtttct tttttacagc ccatcggtcg tgcacaaaac ggaagccttc
41401 tggcgtcgta aataccgaaa ctgtattagc gggcttcggg gtaatcggac gatatgtacg
41461 cggcaactga cggttacggg cgataacttt gttccaggcg tgctcggcgt ggtctttatt
41521 taacgtgtcc aattcgtcta tttttgcgcg aaaagattcg taaccgacga ttcgcgccgg
41581 gttatccaat gtacgcagga caaaatcccc gaattgcccg gatgatgtat agattatgtt
41641 gtctgattta ttgtacttgt aacgaatacc ccaatcagac aacttctctt ccatacgcgg
41701 ggcgaggata aggcgcacca ggtcgtatgt cggttcatac atggcgataa gtgaatcact
41761 accgccctcc atgctgtcga gtagggcgga gttacacatg acctccgatt tacctgtgcc
41821 gaagcccgca acgaaggccg ggaatttaca atgtagatta aggaacgcgc cttgcggttc
41881 cgtagccgtt atatcaacgt tcaccggaca ccaccttaat agttacttcg ctaattggtt
41941 cgtcgcgcgt ttcttccaca acagttttat taagcccaag ttttgctgcg gcgaacgtag
42001 cagatatccc ggcggcaccg gtctcagtga aatacgcctc ttccagggct tgtgccgtct
42061 catatgcttc ggcgaatgca ggaacctctc ggagccatag cttgataatc gggatggtta
42121 caccgatgtg taatgcgaag cgggccagcg acggcggttt gtcctggata agcgggcgtt
42181 cgtcgccttt agacgtgggt acgagttccc atgacgcgcg gtcgaagaac cggattaatt
42241 cgtcgcaata gtccgggtcc catagttcgg cggaatgtct ggatgactga tacaggcttt
42301 gtttaccacg cggacgttta cggcgacggt ttgcggtaac cgcttcttcg tgcgcggcag
42361 ccactacttc ggcgtcaggc tgttttaatt tgagtttcat gcaatcctct cttaaatgcg
42421 tatgaccatc atagcgcat
//
```

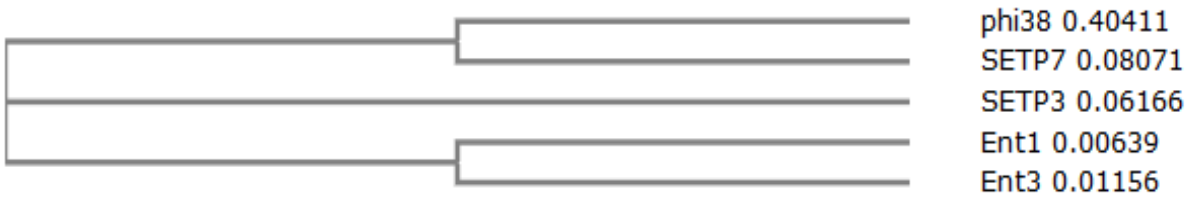
ATTACHMENT II – RESULTS OBTAINED WITH GLIMMER

- GLIMMER (ver. 3.02; iterated) predictions:
- orfID start end frame score
- ——— —— —— – ——
- >Sanna vB_SenS_Ent1 type phage assembled with 2 Ns
- orf00002 325 95 -2 9.46
- orf00003 521 342 -3 0.97
- orf00004 994 638 -2 4.45
- orf00005 1146 991 -1 5.88
- orf00006 1301 1143 -3 12.70
- orf00007 1483 1298 -2 9.86
- orf00008 2155 1667 -2 10.72
- orf00010 2423 2136 -3 1.88
- orf00012 3219 2785 -1 7.84
- orf00013 3602 3225 -3 6.13
- orf00015 3808 3605 -2 2.99
- orf00016 4035 3871 -1 4.18
- orf00017 4581 4405 -1 0.65
- orf00019 5476 5646 +1 3.96
- orf00020 5721 5876 +3 1.92
- orf00021 5933 8119 +2 5.96
- orf00022 8352 8134 -1 5.82
- orf00023 8483 9010 +2 14.42
- orf00024 9054 9536 +3 5.38
- orf00025 9589 10839 +1 10.16
- orf00026 10921 11547 +1 11.84
- orf00029 11605 14703 +1 3.13
- orf00030 14790 15077 +3 5.24
- orf00031 15109 15300 +1 11.92
- orf00032 15381 17762 +3 8.33
- orf00033 17759 17929 +2 4.88
- orf00034 20102 18048 -3 8.49
- orf00036 22673 20115 -3 9.21
- orf00037 23029 22664 -2 2.24
- orf00038 23541 23026 -1 4.48
- orf00039 23981 23538 -3 2.40
- orf00041 26373 24040 -1 6.32
- orf00042 26725 26366 -2 4.27
- orf00043 27234 26731 -1 7.24
- orf00044 27317 27496 +2 5.11
- orf00045 27559 28689 +1 8.03
- orf00046 28686 28916 +3 2.35
- orf00047 29030 28884 -3 2.89
- orf00048 29029 29169 +1 4.42
- orf00049 29370 29714 +3 6.75
- orf00050 30912 29743 -1 10.34
- orf00051 31331 30912 -3 6.63
- orf00052 31726 31331 -2 9.81
- orf00053 32082 31723 -1 6.28
- orf00054 32687 32082 -3 5.67
- orf00055 33199 32690 -2 8.24
- orf00057 33391 33203 -2 8.65

-	orf00058	33778	33428	-2	9.48
-	orf00059	34076	33792	-3	6.45
-	orf00060	35186	34137	-3	6.80
-	orf00062	35891	35190	-3	8.52
-	orf00063	36119	35976	-3	4.37
-	orf00064	36471	36085	-1	10.86
-	orf00065	36635	36516	-3	0.00
-	orf00066	37248	36790	-1	12.26
-	orf00067	38294	37251	-3	9.16
-	orf00068	38471	39121	+2	1.71
-	orf00070	40620	39151	-1	10.00
-	orf00071	41904	40633	-1	8.77
	orf00072	42400	41894	-2	10.80

ATTACHMENT III – PHYLOGENETIC TREE OBTAINED BY CLUSTALW2

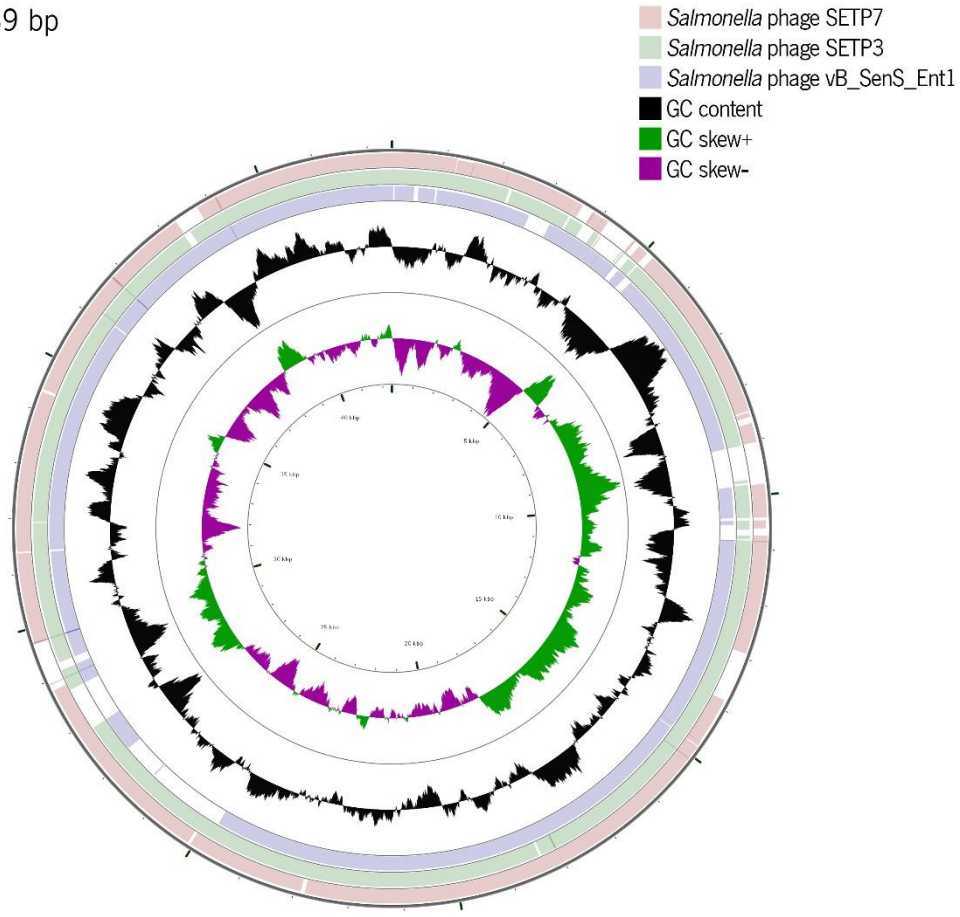
ALREADY WITH DISTANCE CORRECTIONS



ATTACHMENT IV – CIRCULAR MAP OF THE *SALMONELLA* PHAGE vB_SenS_Φ38 GENOME PREPARED USING CGVIEW

Three opaque inner rings correspond to BLASTN from different phages: red – Salmonela phage SETP7, grenn – Salmonella phage SETP3 and blue – Salmonella phage vB_SenS_Ent1. The inner rings show GC content.

Length: 42,439 bp



ATTACHMENT V – FEATURES OF THE PREDICTED ORFS OF BACTERIOPHAGE VB_SENS_Φ38

For each predicted ORF the start and stop position, the transcription strand, the length, Mass, pI and the number of the corresponding gene product in amino acids, the putative protein function, best homologue, genbank accession no., motif and e-value are shown.

	Nucleotide			Protein									
CDS	Coordinates	Strand	Length	Mass	pI	AA residues	Product	Best Homologue	Predicted function	E-value	Gen Bank Accession no.	Motif	E-value
1	95..325	-	230	8124.10	5.42	76	hypothetical protein	hypothetical protein [Salmonella phage SETP3]	-	1,00E-45	YP_001110828	-	-
2	342..641	-	299	11690.98	9.62	99	hypothetical protein	-	-	9,00E-62	-	-	-
3	638..994	-	356	13691.56	7.02	118	hypothetical protein	hypothetical protein [Salmonella phage SETP13]	Bacterial regulatory protein	8,00E-82	YP_008767066	HTH_8 (PF02954)	3.1E-13
4	991..1146	-	155	5763.68	10.29	51	hypothetical protein	hypothetical protein [Salmonella phage SETP7]	-	7,00E-28	YP_008767235.1	DUF2737 (PF10930)	4.9E-31

5	1143..1301	-	158	6008.62	4.15	52	hypothetical protein	hypothetical protein [Salmonella Phage vB_SenS-Ent1]	-	6,00E-12	YP_007010507.1	-	-
6	1298..1483	-	185	6699.65	5.43	61	hypothetical protein	hypothetical protein [Salmonella phage vB_SenS-Ent1]	-	4,00E-32	YP_007010506.1	-	-
7	1667..2155	-	488	17433.83	9.89	162	hypothetical protein	endolysin 68 [Salmonella phage phi68]	Phage Lysozyme	4,00E-109	AHY18890.1	Phage_lysozyme (PF00959)	1.3e-33
8	2136..2423	-	287	10507.69	9.80	95	hypothetical protein	hypothetical protein [Salmonella phage ST4]	-	5,00E-59	AFO70789.1	-	-
9	2425..2706	-	281	10293.02	7.69	93	hypothetical protein	hypothetical protein [Salmonella phage SETP3]	-	4,00E-60	YP_001110821.1	-	-
10	2785..3219	-	434	15530.61	4.61	144	hypothetical protein	hypothetical protein	Proteasome subunit	3,00E-99	YP_005098116.1	Proteasome (PF00227)	1.8e-20

								[Salmonella phage SE2]					
11	3225..3602	-	377	14257.93	10.57	125	hypothetical protein	hypothetical protein [Salmonella phage SETP3]	-	5,00E-83	YP_001110820.2	-	-
12	3605..3808	-	203	7447.59	6.54	67	hypothetical protein	hypothetical protein [Salmonella phage SETP3]	-	4,00E-41	YP_001110819.1	-	-
13	3871..4035	-	165	6322.43	10.72	54	hypothetical protein	hypothetical protein [Salmonella phage SETP3]	-	2,00E-29	YP_001110818.1	-	-
14	5476..5646	+	170	6636.76	10.04	56	hypothetical protein	hypothetical protein [Salmonella phage SETP3]	helix-turn-helix domain	2,00E-32	YP_008859654.1	HTH_17 (PF12728)	1.4e-16
15	5643..5876	+	233	8587.91	9.39	77	hypothetical protein	putative uvsX-like protein [Salmonella phage vB_SenS-Ent1]	-	1,00E-45	YP_007010497.1	-	-

16	5933..8119	+	2186	80427.05	4.99	728	hypothetical protein	hypothetical protein [Salmonella phage ST4]	-	0,00E+00	AFO70781.1	KaiC (PF06745)	9.4E-21
17	8134..8352	-	218	7890.05	9.10	72	hypothetical protein	hypothetical protein [Salmonella phage SETP7]	helix-turn-helix domain	2,00E-44	YP_008767214.1	HTH_19 (PF12844)	1.1e-19
18	8483..9010	+	527	19907.27	4.51	175	hypothetical protein	hypothetical protein [Salmonella phage SETP3]	-	5,00E-101	YP_001110813.1	-	-
19	9054..9536	+	482	18171.61	9.17	160	hypothetical protein	gp59 [Escherichia phage phiEB49]	-	1,00E-34	YP_009018673.1	-	4.5e-18
20	9556..10839	+	1283	47641.17	8.88	427	hypothetical protein	hypothetical protein SP101_00190 [Salmonella phage FSL SP-101]	-	0,00E+00	AGF87741.1	DUF2800 (PF10926)	1.3E-87
21	10921..1154	+	626	23485.14	4.89	208	hypothetic	hypothetical	-	5,00E-	YP_008767039.	DUF2815	1.2e-

	7						al protein	protein [Salmonella phage SETP13]		151	1	(PF10991)	63
22	11605..14703	+	3098	115643.25	8.27	1032	hypothetical protein	hypothetical protein [Salmonella phage SETP13]	-	0,00E+00	YP_007010490.1	-	-
23	14790..15077	+	287	10882.61	9.3	95	hypothetical protein	hypothetical protein [Salmonella phage vB_SenS_AG11]	VRR-NUC domain	5,00E-61	AFO12435.1	VRR_NUC (PF08774)	1.2e-24
24	15109..15300	+	191	7059.17	8.06	63	hypothetical protein	hypothetical protein [Escherichia phage K1-dep(4)]	-	2,00E-36	ADA82278.1	-	-
25	15381..17762	+	2381	88808.74	8.49	793	hypothetical protein	putative intein containing helicase	-	0,00E+00	YP_009009958.1	-	-

								precursor [Salmonella phage vB_SenS-Ent2]					
26	17759..17929	+	170	6501.42	4.85	56	hypothetical protein	hypothetical protein [Salmonella phage vB_SenS-Ent1]	-	4,00E-28	YP_007010486.1	-	-
27	18048..20102	-	2054	72970.63	5.06	684	hypothetical protein	tailspike protein [Salmonella phage L13]	-	0,00E+00	YP_008058250.1	-	-
28	20115..22673	-	2558	93905.02	5.27	852	hypothetical protein	hypothetical protein [Salmonella phage SE2]	-	0,00E+00	YP_005098156.1	-	-
29	22664..23029	-	365	13951.65	6.28	121	hypothetical protein	hypothetical protein [Salmonella phage SETP3]	-	9,00E-85	YP_008859651.1	-	-
30	23026..23541	-	515	19176.73	4.46	171	hypothetical protein	hypothetical protein	-	5,00E-121	AFO12377.1	DUF1833 (PF08875)	6.4e-46

								[Salmonella phage wksl3]					
31	23538..23981	-	443	16459.44	4.59	147	hypothetical protein	hypothetical protein [Salmonella phage wksl3]	-	4,00E-103	AFO12378.1	-	-
32	24040..26373	-	2333	83141.84	4.77	777	hypothetical protein	tape measure protein [Salmonella phage wksl3]	Prophage tail length tape measure protein	0,00E+00	AFO12379.1	TMP_2 (PF06791)	4.9e-10
33	26366..26725	-	359	13524.21	4.99	119	hypothetical protein	hypothetical protein [Salmonella phage SS3e]	-	4,00E-82	YP_308657.1	-	-
34	26731..27234	-	503	19069.74	5.70	167	hypothetical protein	hypothetical protein [Salmonella phage SE2]	Phage tail assembly chaperone	5,00E-95	YP_005098150.1	DUF1789 (PF08748)	2,00E-19
35	27317..27496	+	179	6603.13	9.70	59	hypothetical protein	hypothetical protein [Salmonella phage SETP3]	Superinfection immunity protein	1,00E-31	YP_001110850.1	Imm_superinfect (PF14373)	2.1e-25
36	27559..2868	+	1130	42443.52	8.76	376	hypothetic	hypothetical	Calcineurin-like	0,00E+0	YP_008058258.	Metallophos_2	1.3e-

	9						al protein	protein [Salmonella phage L13]	phosphoesterase superfamily domain	0	1	(PF12850)	15
37	28686..28916	+	230	8690.86	6.05	76	hypothetical protein	hypothetical protein [Salmonella phage SETP3]	-	6,00E-48	YP_001110848.1	-	-
38	28884..29030	-	146	5492.34	9.30	48	hypothetical protein	hypothetical protein [Salmonella phage L13]	-	6,00E-25	YP_008058260.1	-	-
39	29029..29169	+	140	5262.95	10.52	46	hypothetical protein	-	-	0,00E+00	-	-	-
40	29370..29714	+	344	13270.12	5.86	114	hypothetical protein	putative DNA-binding protein [Salmonella phage vB_SenS-Ent1]	Phage antirepressor protein KilAC domain	2,00E-77	YP_007010473.1	ANT (PF03374)	9,00E-35
41	29743..30912	-	1169	41177.06	4.62	389	hypothetical protein	putative major tail protein [Salmonella phage vB_SenS-Ent2]	-	0,00E+00	YP_009009943.1	-	-

42	30912..31331	-	419	15164.11	4.28	139	hypothetical protein	hypothetical protein [Salmonella phage SETP7]	-	3,00E-95	YP_008767190.1	-	-
43	31331..31726	-	395	14494.39	9.69	131	hypothetical protein	hypothetical protein [Salmonella phage SETP7]	Bacteriophage HK97-gp10, putative tail-component	2,00E-88	YP_008767189.1	HK97-gp10_like (PF04883)	9.4e-15
44	31723..32082	-	359	13106.29	9.4	119	hypothetical protein	hypothetical protein [Salmonella phage SETP7]	-	2,00E-77	YP_008767188.1	-	-
45	32082..32687	-	605	20587.46	9.05	201	hypothetical protein	hypothetical protein [Salmonella phage SETP3]	-	8,00E-135	YP_001110842.1	-	-
46	32690..33199	-	509	17867.02	4.71	169	hypothetical protein	hypothetical protein [Salmonella phage SS3e]	-	3,00E-114	YP_224025.1	-	-
47	33203..33391	-	188	7239.04	5.12	62	hypothetical protein	hypothetical protein SP101_00055	-	2,00E-33	AGF87714.1	-	-

								[Salmonella phage FSL SP-101]					
48	33428..33778	-	349	12083.41	4.38	116	hypothetical protein	putative major tail subunit [Salmonella phage SETP3]	-	2,00E-74	YP_001110839.1	-	-
49	33792..34076	-	284	9353.70	9.77	94	hypothetical protein	putative head protein [Salmonella phage SETP13]	-	1,00E-55	YP_008767011.1	-	-
50	34137..35186	-	1049	37897.43	4.58	349	hypothetical protein	putative coat protein [Salmonella phage wksl3]	-	0,00E+00	AFO12396.1	-	-
51	35190..35891	-	701	25749.02	5.78	233	hypothetical protein	hypothetical protein [Salmonella phage SETP13]	Phage minor structural protein gp20	4,00E-162	YP_008767009.1	Phage_GP20 (PF06810)	2.3e-07
52	35976..36101	-	125	4778.43	4.51	41	hypothetical protein	p11.5 [Salmonella	-	3,00E-20	CAA60564.1	-	-

								phage MB78]					
53	36085..36471	-	386	14148.96	9.93	128	hypothetical protein	hypothetical protein [Salmonella phage vB_SenS_AG11]	-	3,00E-83	AFO12408.1	-	-
53a	36516...36635	-	119	-	-	39	hypothetical protein	Salmonella phage SETP3, complete genome	-	3,00E-53	EF177456.2	-	-
54	36790..37248	-	638	16396.58	4.66	212	hypothetical protein	putative tail protein [Salmonella phage SETP3]	-	1,00E-98	YP_001110834.1	-	-
55	37251..38294	-	1043	38640.69	5.64	462	hypothetical protein	head morphogenesis protein [Salmonella phage SETP3]	Phage Mu protein F like protein	0,00E+00	YP_001110833.1	Phage_Mu_F (PF04233)	7.8e-26
56	38471..39121	+	650	24957.49	5.87	216	hypothetical protein	hypothetical protein [Salmonella	KilA-N domain	3,00E-159	YP_001110832.1	KilA-N (PF04383)	5.5e-05

								phage SETP3]					
57	39151..40620	-	1469	53908.31	4.75	489	hypothetical protein	putative portal protein [Salmonella phage vB_SenS-Ent2]	Phage portal protein, SPP1 gp6-like	0,00E+00	YP_009009927.1	Phage_prot_Gp6 (PF05133)	4.1e-29
58	40633..41904	-	1271	47603.58	6.03	423	hypothetical protein	hypothetical protein [Salmonella phage ST4]	-	0,00E+00	AFO70799.1	-	-
59	41894..42400	-	506	18821.40	5.99	168	hypothetical protein	hypothetical protein [Salmonella phage SS3e]	helix-turn-helix domain of resolvase	3,00E-116	YP_224037.2	HTH_7 (PF02796)	9.7e-09