The Type II-A CRISPR-Cas system of Streptococcus mutans: Characterisation of Bacteriophage-Insensitive Mutan(t)s

Thèse

Cas Mosterd

Doctorat en microbiologie
Philosophiæ doctor (Ph. D.)

Québec, Canada

Résumé

Les bactéries sont continuellement exposées à un danger, la prédation par des bactériophages. Pour se défendre, elles ont développé une grande variété de mécanismes. Parmi ceux-ci, on retrouve CRISPR-Cas (« clustered regularly interspaced palindromic repeats »), un système adaptatif que possèdent environ 45% des bactéries. Une caractéristique unique du système CRISPR-Cas est qu'il constitue en quelque sorte la mémoire de l'hôte. Par exemple, le système peut emmagasiner des petits fragments d'un génome viral, appelés espaceurs, et les introduire dans son CRISPR. Cette mémoire lui permet de se défendre contre une réinfection par le même virus ou un virus hautement apparenté. Par contre, malgré que l'acquisition de nouveaux espaceurs semble fréquente dans la nature, ce phénomène n'est que très rarement observé en conditions de laboratoire. Néanmoins, quelques bactéries font exception à la règle et l'une d'entre elles est *Streptococcus mutans*. Dans le cadre de cette étude, l'interaction entre la souche *S. mutans* P42S et le bactériophage virulent M102AD a été analysée en détail. De plus, certaines applications potentielles du système CRISPR-Cas ont également été approfondies.

Le premier objectif de cette thèse était de caractériser le système CRISPR-Cas de *S. mutans* P42S au niveau moléculaire et de déterminer son rôle dans les interactions phage-bactérie. Le deuxième objectif était d'établir le potentiel de la protéine Cas9 de *S. mutans* P42S (SmutCas9) comme nouvel outil d'édition génomique. *S. mutans* P42S possède un système CRISPR-Cas de type II-A. Bien que ce type de système soit probablement le plus étudié, celui de *S. mutans* P42S présente plusieurs caractéristiques uniques lui permettant de se démarquer. En effet, ce dernier reconnaît un PAM différent de ce qui était auparavant connu pour cette espèce bactérienne, l'acquisition simultanée de multiples espaceurs semble fréquente, ce qui est probablement dû au phénomène de « priming ». Malgré le rôle de CRISPR-Cas dans la défense antivirale, *S. mutans* P42S dispose d'autres mécanismes de défense contre les phages. Des cellules mutantes sont résistantes aux phages en empêchant l'adsorption de particules virales à la cellule ont notamment été observées. D'autres mécanismes sont assurément impliqués dans la défense antivirale de *S. mutans*. Finalement, SmutCas9 s'est montrée efficace dans l'édition de génomes viraux et elle apparaît comme une candidate à explorer pour cette application.

Abstract

Bacteria are exposed to the constant threat of viral predation. To defend themselves, bacteria have developed a wide variety of different mechanisms. One of these mechanisms is CRISPR-Cas (clustered regularly interspaced palindromic repeats), an adaptive immune mechanism found in approximately 45% of bacteria. A unique feature of CRISPR-Cas systems compared to other antiviral defence mechanisms is that it has a memory. The system is capable of remembering previous viral encounters and protects the bacterial host from re-infection by the same or highly-related viruses. This memory is due to the acquisition of virus-derived genome fragments called spacers. Despite common acquisition of novel spacers in nature, and thereby the emergence of new immunity, acquisition of new spacers under laboratory conditions has been rarely observed. One of the few exceptions is *Streptococcus mutans*. In this study, the interactions between *S. mutans* strain P42S and its virulent bacteriophage M102AD are investigated in detail. In addition, possible applications of the CRISPR-Cas system are analysed.

The first objective of this thesis was to characterise the CRISPR-Cas system of *S. mutans* P42S on the molecular level and to determine its role in antiviral defence. The second objective was to determine the potential of the Cas9 protein of *S. mutans* P42S (SmutCas9) in genome editing. *S. mutans* P42S possesses a type II-A CRISPR-Cas system. Although this is arguably the best studied system, the one found in the strain *S. mutans* P42S has several features that makes it stand out. It recognises a PAM different from what was known for this species, multiple spacer acquisitions are frequent, and this appears to be partially due to priming. Although CRISPR-Cas plays a role in antiviral defence, there are additional antiviral defence mechanisms that protect *S. mutans* against phages. Adsorption resistance is one of them, although additional unidentified antiviral defence mechanisms are likely involved. Finally, SmutCas9 has been shown functional in editing of viral genomes and appears to be a candidate for human genome editing.

Table of contents

List of figures

List of tables

List of abbreviations

ABI	abortive infection
aca	anti-CRISPR associated
acr	anti-CRISPR
BHI	Brain Heart Infusion
BIM	bacteriophage-insensitive mutant
bp	base pair
BREX	bacteriophage exclusion
$CaCl_2$	calcium chloride
cas	CRISPR-associated
Cascade	CRISPR-associated complex for antiviral defence
CBASS	cyclic oligonucleotide-based anti-phage signalling system
CDM	chemically defined medium
cDNA	complementary DNA
CEM	CRISPR escaping mutant
Chi	Crossover hotspot instigator
Cmr	CRISPR RAMP module
CO_2	carbon dioxide
cpf	CRISPR-Cas substype as in *Prevotella* and *Francisella*
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR-RNA
csf	CRISPR-Cas subtype as in *Acidithiobacillus ferrooxidans*
CSP	competence stimulating peptide
DISARM	defence islands systems associated with restriction-modification
DNA	deoxyribonucleic acid
DSBs	double-stranded DNA breaks
dsDNA	double-stranded DNA
DTT	DL-Dithiothreitol
EOP	efficiency of plaquing
GBP	glucan-binding protein
GM17	M17 supplemented with glucose

iap	isozyme alkaline phosphatase
kb	kilo base
LB	Luria Broth
LC-MS	liquid chromatography–mass spectrometry
LDH	lactate dehydrogenase
LM17	M17 supplemented with lactose
MOI	multiplicity of infection
ng	nanogram
nt	nucleotide
OD_{600}	optical density at 600 nanometres
ORF	open reading frame
pAgo	prokaryotic Argonaute
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PFU	plaque forming unit
pH	potential of hydrogen
RGP	rhamnose-glucose polysaccharide
R-M	restriction-modification
RNA	ribonucleic acid
rpm	rotations per minute
rRNA	ribosomal RNA
SIE	superinfection exclusion
SmaCas9	*Streptococcus macacae* Cas9
SmutCas9	*Streptococcus mutans* Cas9
SpyCas9	*Streptococcus pyogenes* Cas9
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
TCTS	two-component signal transduction system
tracrRNA	trans-activating crRNA
TSYE	Tryptic Soy Broth supplemented with yeast extract and K_2HPO_4
UV	ultra-violet

WT	wild-type

Deze is voor jou, mama, dankjewel voor alles.

Acknowledgments

When you stay at the same place for a long time, you will end up thanking a lot of people. I would like to start with my supervisor Sylvain Moineau. If not for him, I would not have even started this amazing adventure of moving far away from home to not only learn so much about science, but also about life. With his enthusiasm and positive attitude he convinced me to cross the ocean and the positive attitude is something that helped me throughout the PhD. In addition, I would want to thank the research staff of the lab that keeps the lab running on a daily basis and has been a great all-round help. Therefore, many thanks to Denise Tremblay and Stéphanie Loignon, and especially to Geneviève Rousseau. She often thought along with me in the lab on moments that I needed it the most and also made a lot of effort to make me feel at home. Also the rest of the colleagues who helped me deserve a great thank you. In addition, thank you to my advisory committee consisting of Alex Culley and Jacques Corbeil as well, for taking the time to give advice throughout the years. And thank you to external examinator Peter Fineran for his great suggestions. NSERC of Canada is also acknowledged for funding the project.

Over the years, many people have been trained in the Moineau lab ("the bacteriophage bunker") in the dentistry building of the Université Laval. I got the chance to meet some of the most wonderful and brilliant people that I've met in my life. Some stayed for a long time, some for a too short a period (particularly international visitors), but many people have helped me during my project and not only that. Many helped me to develop as a person and to know more about life. They showed me many new insights and different views that will stay with me for the rest of my life. Upon arrival, I got helped so much by Lynn el Haddad and Siham Ouennane, who were pretty much like two big sisters. They helped me with all the paper work and gave me tips and tricks of how everything works in Université Laval as a foreign student. Alex Hynes and Simon Labrie were also always available to help out whenever I had a question in the lab. It was a great opportunity to work with them. Giovanni Eraclio was a great companion during evenings and weekends in the lab. I will never forget the laughs and great conversations I've had with Alexia Lacelle-Côté. She was always there at moments when things were tough. Caroline Renaud and

Ariane Renaud always improved the atmosphere when they were there. Jessy Bélanger was such a great addition to the lab with his positive attitude. The ultimate good guy. Alessandra Gonçalves de Melo has always been there to give the advice a big sister would give to her little brother. Also, I'm sure neither of us will forget how we were both there during the great flood of the lab. Sana Hamdi was a great companion during weekends and holidays in the lab and we will always remember how we are survivors of that hell of a snowstorm during Christmas. Sébastien Lévesque definitely made the basement a better place with his sense of humour and always showed us that you could be more unlucky in the lab. Hanne Hendrix was a great colleague and housemate, although for a very short period. Françoise Leblanc-Bourque and Rachel Morin-Pelchat are among the funniest girls that I know and they always made me smile. Honghui Liu was such an inspiration to have around. She was truly a model colleague with good work ethics and with an amazing attitude towards other people and towards life. Nicolas 'Timazing' Lemire has been great to have as a colleague, but also as a drinking and sports buddy. It's just difficult not to like Tim. Speaking of difficult not to like, of course there was Xiaojun 'Xiaomazing' Zhu. He is not only a very kind guy, but also the master of crystals and hotpots. Gabriel Byatt has among the most contagious laughs I've heard. It's difficult not to be cheerful when Gab is around. Alice Perrault-Jolicoeur always has a fun anecdote to tell. Although short, during his stay, Torben Sølbeck Rasmussen was great to have around. It was nice to have a fellow northerner in the basement. Yuyu Shao is among the craziest, genius, creative and dynamic people I've ever met. With his unique sense of humour, his energy and brilliance he was a great addition to the team both inside and outside the lab. Jeffrey Cornuault always gave new insights when discussing a scientific problem, as well as when discussing football over a drink. The stay of Ville Hoikkala was too short, but he was good to have around and contributed to a positive vibe in the lab. Victoria Bureau-Lagarde was lots of fun in the lab and could switch easily between being the sweetest and the meanest person of the lab in a matter of seconds. Her sense of humour made the lab a good place to go to. The same can be said for Neil Chennoufi. Having Neil in the lab was like having a little brother. We shared so many great nights out and I will always keep those precious memories. Clément Fage was another great addition to the gang. His cheeky sense of humour was well appreciated in the basement or during nights out with the guys.

Outside the lab I've also had the chance to meet some wonderful people. First of all there were former housemates Edmond Haxaire, Félix Vallet and Toni Stoeber who welcomed me into the French community of Québec and with whom I will always share countless memories of our great nights in Québec. Through them I've also met my next former housemates, Manuel Pinard and Marc-André Rioux. I have had so many laughs and great nights with them and I feel lucky to have had them as my housemates. They also introduced me to other great people of the agronomie gang, such as Gabrielle Dumas, Renelle Giroux, Philippe Laliberté and Maxime Côté. And of course, there is my current housemate Alireza Yaseri. We went through the quarantine together and motivated each other while working from home. I'm happy to have spent this period in such good company. Apart from them, there is Martyne Audet, who accepted me in her family during Christmas and when I locked myself out. She was really someone I could count on. There was my aunt Tamam Ahmed Jama, who provided me with so much familial warmth far away from home and is one of the best and most impressive people I know. Besides them there were Deo Florence Onda and Adeel with whom I had many interesting discussions. The same can be said for my gym buddies Amar Laidani and Oliver Ernst. And of course, my tennis, badminton and drinking buddies Minja Velimirović and Billie-Tanisha Moar.

Back home in the Netherlands, there have been also people who have been regularly sending me long-distance support. In particular, I want to mention Dieke van Rees, Jeroen Koomen, Mirella Feenstra, Brenda van der Lee, Anne van Leeuwen, Daisy Hofman, Arja Grolleman, Lisette Erens, Tirsa Spruitenburg, Niels Floor and Wilbert Walraven. Your kind words and support have helped me through difficult moments.

I already mentioned a long list of colleagues who I owe a thank you. Some of them deserve some special attention. First of all, there were Maxime Bélanger and Bruno Martel, who warmly welcomed me into the lab. Both of them have always been available to think along with me during my project, but they have also been great friends and are just exceptionally thoughtful and caring people. They introduced me to the wonderful Canadian winter sports and have been great company during nights out, nights at home or late

evenings in the office. In the case of Maxime, I unfortunately only got to work with him for a couple of months. Bruno spent two periods in the lab so we spent a bit more time in the workspace. It was an absolute joy to work with them and get to know them. Also, there was Hany Geagea. He was there to welcome me into the lab and he has been there for me since the start. As my neighbour in the lab he was usually the first to know about my latest lab results and the first person I would talk to about it. But most of all, he has been an incredible friend who has supported me every single moment of my PhD. Together we got to know this new environment we lived in, shared our experiences and helped each other out with every struggle. His friendship has meant an awful lot and I could never have done this without Hany. Another great person I met in the lab was André da Silva Xavier. Very few people on this planet are as kindhearted as André. He has provided tremendous emotional support and friendship and his incredible work ethic was very motivational. There was also Witold Kot who unfortunately did not stay with us in the lab very long. But during the time he was there, he was an incredibly supportive friend, both inside and outside the lab. He is also among the people that had the biggest positive influence on the general atmosphere in the lab. Of course, I cannot forget Jéssica Fernandes Carvalhais. With her I had an amazing friend to keep me company in the lab during late evenings, weekends and holidays. But not only that, with her I had someone with whom I could talk about anything and with whom I have countless memories that I will cherish forever. I would like to mention Frank Oechslin, one of the most brilliant people I've met. His ideas and view on things have not only been a great help in the lab, but have also resulted in countless unforgettable nights out. I am happy to have such a friend that I can count on, both in and out of the lab. And finally there was the latest addition to our lab and twin sister Martina Scarrone, who has been an incredible help during the period I was writing my thesis and preparing for the defence, thinking along at so many important moments. In addition, her sense of humour, energy and friendship make life a more fun. And thanks to her being my training partner in confinement home workouts she helped me not to get fat during these last months of pizza-fueled thesis writing.

Then there have been some dear friends I've met outside the lab that also deserve a special mentioning. Deborah Tomaz has been a wonderful person to talk to who has

provided a lot of moral support. The same can be said about Erfan Dabaghi Zadeh. The many talks and good company in the gym, in the bar and wherever have meant a lot to me. Myriam Labbé, Krisztián Ratkovszki, Chenour Hasani and Samira Ebrahim have also been great friends throughout the PhD. The many nights out and good talks have been a great source of positive energy. I feel lucky to have met such wonderful friends. And last but definitely not least, Julien Fleuret. I have rarely met someone as warm, generous and loyal as Julien. We have spent a lot of hours together and I will always cherish all those good talks, the drinks, gym sessions that would take way too long and all the well meant advice. He has always been there for me and really made Québec city a better place.

There is a very special group that means the world to me. Donna dela Cruz is one of the people on this planet who understands me the most and with whom I got to share everything that was on my mind throughout this crazy journey called a PhD. No matter how far away we are from each other, I feel her support. Alice Lévesque has helped me in so many ways that I do not even know where to begin. She may actually sometimes know me better than I know myself. She also is the person who saw me at my worst, but still has my back. Alice and her mums really gave me the feeling of having a family of my own, far away from home. Words can never express my gratitude. Roberta Dagher has really improved my life ever since I know her. I consider myself very lucky to have her in my life. And whereas for many people 2020 will not be remembered as a highlight in their lives, for me it will always be the year when I met Miruna Anohim. Miruna has been amazing company in everyday life and gives me all the love and positive energy I need to make the most out of every day. I cannot thank her enough for her support.

But finally, the most important has been my mum. She has always done everything in her power and she has provided a solid base to make me succeed. If not for her, I would have never even started a PhD in the first place. She has always believed in me and had my back for the full 100%. All the motivational talks and all the love she has provided me with for all my life, including those years I have been far away, have made a huge difference. She has given me so much in this life and I will never be able to repay her, but I hope this

PhD will make her proud. I’m lucky to have her not only as my mum, but also as my best friend.

Foreword

This thesis consists of an introduction, which includes a review article; a section with the objectives of the study; and three chapters organised as research articles. These chapters are followed by a conclusion section containing, among others, perspectives.

Introduction

In the introduction section, I discuss bacteriophages, *Streptococcus mutans* and its phages.

First research article (Chapter 1)

"Characterisation of a Type II-A CRISPR-Cas System in *Streptococcus mutans*". This article was published online in mSphere on the 24th of June 2020. The paper version of the article was published in the May/June edition of 2020 (Volume 5, Issue 3, e00235-20). I am the first author of this article which was prepared in collaboration with Sylvain Moineau. All authors are affiliated with the Département de biochimie, de microbiologie et de bio-informatique of Université Laval and with the Groupe de recherche en écologie buccale. Sylvain Moineau is also affiliated to the Félix d'Hérelle Reference Center for Bacterial Viruses.

Second research article (Chapter 2)

"CRISPR-Cas and adsorption resistance provide combined phage protection in *Streptococcus mutans*".

This article is in preparation for submission. I am the first author of this article which was prepared in collaboration with Sylvain Moineau. The affiliations of the authors are as in the previous chapters.

Third research article (Chapter 3)

"Application of Cas9 from *Streptococcus mutans* P42S in viral genome editing".

This article is in preparation for submission. I am the first author of this article which was prepared in collaboration with Sylvain Moineau. Again, the affiliations of the authors are as previously mentioned.

Fourth research article (Annex A)

"A mutation in the *methionine aminopeptidase* gene provides phage resistance in *Streptococcus thermophilus*".

This article was published online in Nature Scientific Reports on the 25th of September 2019. The paper version of the article was published in Volume 9, Issue 13816. I am the second author of this paper, which was prepared in collaboration with Simon Labrie, Stéphanie Loignon, Marie-Ève Dupuis, Philippe Desjardins, Geneviève Rousseau, Denise Tremblay, Dennis Romero, Philippe Horvath, Christophe Fremaux and Sylvain Moineau. I performed the experiments related to *Streptococcus mutans*. Simon Labrie headed the project. Simon Labrie and Sylvain Moineau wrote the article and designed the experiments. Simon Labrie also performed the adsorption assays and growth curves. Marie-Ève Dupuis isolated the natural non-CRISPR BIMs. Geneviève Rousseau performed the mutation stability test and the growth curves. Denise Tremblay sequenced the genomes. Philippe Desjardins assisted Simon Labrie during Map experiments. Dennis Romero, Philippe Horvath and Christophe Fremaux helped with the design of the project. At the time of the project, Simon Labrie, Stéphanie Loignon, Marie-Ève Dupuis, Philippe Desjardins, Geneviève Rousseau, Denise Tremblay, Sylvain Moineau and myself were all affiliated with the Département de biochimie, de microbiologie et de bio-informatique of Université Laval and with the Groupe de recherche en écologie buccale. In addition, Simon Labrie is affiliated with SynthBio Lab Inc. and Sylvain Moineau with the Félix d'Hérelle Reference Center for Bacterial Viruses. Dennis Romero is affiliated with DuPont Nutrition and Biosciences (USA) and Philippe Horvath and Christophe Fremaux are affiliated with DuPont Nutrition and Biosciences (France).

Review (Annex B)

"A short overview of the CRISPR-Cas adaptation stage"

This review article was published online in the Canadian Journal of Microbiology on the 19th of June 2020. I am the first author of this review which was prepared in collaboration with Geneviève Rousseau and Sylvain Moineau. I performed the literature research and wrote most of the review. All authors are affiliated with the Département de biochimie, de microbiologie et de bio-informatique of Université Laval and with the Groupe de recherche en écologie buccale. In addition, Sylvain Moineau is affiliated with the Félix d'Hérelle Reference Center for Bacterial Viruses.

Introduction

Streptococcus mutans

Streptococcus mutans is a facultative anaerobic gram-positive bacterial species, primarily associated with dental caries (Loesche, 1986). It was first isolated from a human tooth cavity sample in 1924. The isolated bacterium appeared to consist of ovaloid cocci rather than round ones, hence the name *S. mutans* (Clarke, 1924). There are a variety of virulence factors produced by *S. mutans* that lead to dental caries. These factors include, among others, adhesion factors and acid tolerance (Banas, 2004).

S. mutans encodes a wide variety of adhesion factors enabling the cell to adhere to tooth surfaces. Glucose-independent adhesion occurs primarily by the means of the antigen I/II family. Proteins from this family enable the cell to adhere to various salivary components (Petersen et al., 2002). Furthermore, *S. mutans* makes use of glucose-dependent adhesion by relying on its glucosyl transferases (Munro et al., 1995; Ooshima et al., 2001). This group of enzymes splits sucrose into separate glucose and fructose moieties and catalyses a reaction to form glucans (glucose polymers) from glucose (Monchois et al., 1999). The glucans favour adherence of both the bacteria and the salivary pellicle, which is a protein-film covering the tooth surface (Kawabata and Hamada, 1999). *S. mutans* also encode several glucan-binding proteins (GBPs), of which several have been associated with virulence (Russell, 1979; Sato et al., 1997; Shah and Russell, 2004; Smith et al., 1994). In addition, the glucan-binding protein A (GbpA) is also involved in binding proteins and exopolysaccharides for biofilm formation (Banas et al., 2007). Other GBPs are equally essential for biofilm formation (Lynch et al., 2007). The biofilm structure allows *S. mutans* to withstand the constant fluctuation in nutrient availability and pH (Lemos and Burne, 2008; Yoshida and Kuramitsu, 2002).

Another factor of interest is the production of several bacteriocins. The bacteriocins encoded by *S. mutans* have been named mutacins (Hamada and Ooshima, 1975). These antimicrobial peptides can eliminate competitors for nutrients in the oral environment, perhaps functioning as virulence factors. Various stress-related genes, including several mutacin encoding genes are regulated by the competence ComDE system, a two-component signal

transduction system (TCSTS) (Li et al., 2002b, 2001; Perry et al., 2009; Van Der Ploeg, 2005). Mutacin-producing strains have been shown to encode bacteriocin-immunity proteins, also regulated by the ComCDE system (Matsumoto-Nakano, 2018). These observations have resulted in the hypothesis that *S. mutans* uses its mutacins to lyse competing bacteria and to take up free DNA (Kreth et al., 2005; Van Der Ploeg, 2005). This would lead to genetic diversity or for the nucleic acids to serve as nutrients (Spoering and Gilmore, 2006).

S. mutans is a highly acidogenic bacterium encoding a wide variety of fermentation pathways with end products as lactate, formate and acetate (Ajdić et al., 2002). Depending on the conditions encountered, the ratios of these fermentation products may vary. Lactate is the primary end product of fermentation in high glucose conditions (Dashper and Reynolds, 1996). Although acidogenicity differs from strain to strain, acid production of *S. mutans* along with *Streptococcus sobrinus* is significantly higher than that of other streptococci within the environmental pH range of 5 to 7 (De Soet et al., 2000). The acid production of *S. mutans* alters the composition of the dental plaque flora, as growth of more acid-resistant bacteria is favoured. In addition, the low pH leads to demineralisation of tooth enamel which in the long term results in dental caries (Banas, 2004). A key factor for this outcome is lactate, as *S. mutans* strains deficient in lactate dehydrogenase (LDH) are significantly less cariogenic (Fitzgerald et al., 1989; Johnson et al., 1980). The importance of LDH is further underlined since its absence may even be lethal for *S. mutans* (Hillman et al., 1996).

Not only is *S. mutans* capable of producing a variety of acids, it is also highly resistant to them. In acidic conditions, *S. mutans* utilises its acid-tolerance response (Svensäter et al., 1997), which consists of altering the expression of genes involved in protecting the cell from the low pH and in particular an F_1-F_0-ATPase proton pump. Indeed, expression of the F_1-F_0-ATPase proton pump is increased as the pH decreases (Kuhnert et al., 2004). By pumping out protons the cell is capable of maintaining the internal pH at desired levels (Bender et al., 1986; Dashper and Reynolds, 1992; Kobayashi, 1985; Kobayashi et al., 1986). Furthermore, membrane fatty acid profiles are altered upon lower pH levels, thereby reducing the permeability to protons (Quivey et al., 2000). Acid-induced DNA repair enzymes have also been reported in *S. mutans* (Hahn et al., 1999). In addition, malolactic fermentation is employed by *S. mutans* to transform malate

into the less acidic lactic acid and CO_2, thereby increasing the internal pH (Sheng and Marquis, 2007).

S. mutans is almost exclusively found in mixed-population biofilm conditions on the tooth surface. Environmental factors determine the structure and composition of these biofilms. In order to respond to environmental changes within biofilms, TCSTSs are used to sense the variations and regulate gene expression accordingly. At least 14 TCTSs have been found in *S. mutans* (Ajdić et al., 2002; Biswas et al., 2008). Mechanisms that are regulated by TCSTS include virulence gene expression (Chen et al., 2008; Lévesque et al., 2007; Senadheera et al., 2005; Van Der Ploeg, 2005; Zeng et al., 2006), biofilm production (Li et al., 2002a; Qi et al., 2004), stress tolerance (Biswas et al., 2008; Chen et al., 2008; Deng et al., 2007; Li et al., 2001, 2002a; Qi et al., 2004) and competence (Li et al., 2002b; Qi et al., 2004; Senadheera et al., 2005; Van Der Ploeg, 2005).

Natural competence is a phenomenon found in multiple species of the *Streptococcus* genus (Håvarstein et al., 1997). It is driven by the Competence Stimulating Peptide (CSP), a signal peptide that is excreted by the cell to regulate competence-related genes in neighbouring cells. The unprocessed version of CSP is encoded by *comC* and together with *comDE* it forms the competence cassette *comCDE*. *comD* encodes a histidine kinase receptor protein for CSP, anchored in the membrane and *comE* encodes the intracellular response regulator which induces other competence related genes downstream in the process, ultimately sigma factor *comX* (Li et al., 2001), the master regulator of competence. Whereas in the orthologous system in *Streptococcus pneumoniae*, *comE* directly activates *comX*, in *S. mutans* the intermediate regulating system *comSR* is required downstream of the ComCDE system. ComE induces a number of pathways of which one undefined pathway results in expression of peptide precursor ComS, which is processed into a peptide regulating the expression of signal peptide ComR, ultimately activating *comX* (Mashburn-Warren et al., 2010).

Competence was correlated to proper biofilm formation as *comCDE* mutants generated biofilms with less integrity and biomass (Li et al., 2002b). Transformation efficiencies are highest under biofilm conditions since competence-related genes are upregulated in this setting (Li et al., 2001; Rathsam et al., 2005). In addition, at high concentrations of CSP, cell lysis is

induced (Qi et al., 2005), which suggests that the competence pathway functions in regulating cell density within biofilms (Lemos and Burne, 2008). The natural competence of *S. mutans* has been exploited in laboratory settings (Dufour et al., 2011). The exact sequence of CSP, the processed form of ComC, was identified (Hossain and Biswas, 2012) and its addition to an exponentially growing culture resulted in efficient transformation (Dufour et al., 2011).

Bacteriophages (phages)

Viruses are capable of infecting virtually all life forms, including bacteria. Viruses that infect only bacteria are called bacteriophages, or phages for short. With an estimated 10^{31} phage particles on the planet, they are considered the most numerous biological entities (Suttle, 2005) as well as the most diversified (Breitbart and Rowher, 2005). Since phages are responsible for a significant fraction of bacterial deaths in every ecosystem (Fuhrman, 1999), they have a role in the regulation of bacterial populations, particularly in marine environments (Suttle, 2005). Phages can also regulate the bacterial composition of our microbiota, such as in the gut (Kim and Bae, 2018) or the oral cavity (Bachrach et al., 2003).

The first description of filtrates capable of killing bacteria was in 1915 by Frederick Twort. Several theories were suggested. One of them was the responsibility of a bacterial virus (Twort, 1915). Independently, Félix d'Hérelle in 1917 stated the presence of bacterial viruses, invented the word "bacteriophage" and was able to isolate some of them (d'Herelle, 1917). Phage research has made remarkable advances over the past decades as they were the first models studied in virology. Nevertheless, phage research is currently still limited by the fact that the genomes of the majority of phages have not been sequenced. Many intact (or defective) phages are also found as prophages (dormant integrated phages) in bacterial genomes (Hendrix, 2003). In order to isolate or identify phages, first the bacterial host needs to be cultured, which is a significant bottleneck as the vast majority of the bacteria found in metagenomic data have yet to be cultured (Breitbart and Rohwer, 2005; Grimes et al., 1986; Torsvik et al., 1990; Torsvik and Øvreås, 2002). Development and improvement of culture-independent methods can help overcome this technical problem. For this, phage detection and identification methods have been developed involving techniques such as PCR (del Rio et al., 2007; Del Rio et al., 2008; Labrie and Moineau, 2000; Ly-Chatain et al., 2011; Moisan and Moineau, 2012). Whereas for bacteria,

the 16S rRNA harbours slow evolving sequences that allow a targeting site for bacteria-specific sequencing and identification (Woese et al., 1990), no such gene exists in phages (Breitbart and Rohwer, 2005). Therefore, PCR-based detection and identification methods have primarily been useful in cases where there is prior knowledge regarding the viral presence within the tested sample. A common viral signature gene used to identify viral sequences is *g23*, present in almost all viral genomes sequenced thus far (Adriaenssens and Cowan, 2014). However, this gene is identical in many related viruses and therefore does not allow the distinction between viruses as the 16S rRNA PCR does for bacteria (Nakayama et al., 2010).

The most common method to identify phages of which prior knowledge is lacking is metagenomics. Metagenomic approaches have high potential to reveal the true diversity of phages present on our planet (Dutilh, 2014; Koonin and Yutin, 2020). The taxonomy of phages is currently undergoing significant changes due to the avalanche of new and diverse viral genomes that have been identified through metagenomic studies and bioinformatic approaches (Walker et al, 2019; Koonin and Yutin, 2020).

Depending on the context, the capability of phages to infect and kill bacteria is considered disadvantageous or advantageous. Industrial food fermentations often make use of beneficial bacteria to change the organoleptic properties of a food product. The food fermentation process often occurs under non-sterile conditions in large vessels and phage infection is a significant undesired risk which can lead to fermentation failure or low-quality fermented products (Samson and Moineau, 2013). The most well-known advantageous application is phage therapy. Widespread antibiotic resistance among pathogenic bacteria is a major concern and due to the sharp decline of newly discovered antibiotics, alternatives such as phages may provide a positive avenue for control. Since phages are non-toxic to human cells, self-dosing and highly specific, they have been considered promising agents to combat bacterial infections. However, their high strain specificity is simultaneously considered a downside, since it requires identification of the exact bacterial pathogen (Domingo-Calap and Delgado-Martínez, 2018).

As phage infection is a constant threat to bacteria, they have evolved a variety of defences to protect themselves from predation. In turn, phages have developed mechanisms to circumvent

these defences which have led to the extension of the antiviral defence arsenal of bacteria. Many bacteria encode several antiviral defence mechanisms. In fact, they are estimated to make up 10% of bacterial genomes (Koonin et al., 2017). They are regularly gained due to horizontal gene transfer, but can as easily be lost if they create a fitness disadvantage in the absence of phage infection (van Houte et al., 2016). Because a single strain could never harbour all of these defence mechanisms, the pan-immune system has been proposed. This theory proposes that when strains with different defence mechanisms coexist within a population, this creates a reservoir of systems (the pan-genome), increasing the chance of survival of some of the members, and therefore the population, upon phage infection (Bernheim and Sorek, 2020).

The infection or replication cycle of phages consists of multiple steps, of which every single step can be a target of these antiviral defence mechanisms. The first step of phage infection involves adsorption to a receptor (or multiple receptors) on the surface of its host. The bacterium can defend itself by masking or modifying the receptor (Labrie et al., 2019). If the infection continues, the second step would involve the introduction of its genome, commonly by injection. A mechanism that interferes with the injection of DNA is the superinfection exclusion (SIE) system (Sun et al., 2006). Once the phage genome is introduced, the next line of defence interferes with the replication of this genome. For this purpose, bacteria can encode a wide variety of defence mechanisms, such as restriction-modification (R-M) systems (Oliveira et al., 2014), defence islands systems associated with R-M (DISARM) (Ofir et al., 2018), bacteriophage exclusion (BREX) (Barrangou and Oost, 2015; Goldfarb et al., 2015), prokaryotic Argonaute (pAgo) proteins (Swarts et al., 2014), and CRISPR-Cas (Barrangou et al., 2007). If the phage manages to bypass all these lines of defence, the phage DNA gets replicated, transcribed and translated and the phage proteins self-assemble to form new phages, ready to be released and infect their next prey. However, bacteria may encode a last resort in the form of abortive infection (ABI) systems (Samson et al., 2013), or cyclic oligonucleotide-based anti-phage signalling systems (CBASS) (Cohen et al., 2019). These systems kill the infected cell but protect the bacterial population from further infection. Many more antiviral systems exist or are under investigation (Doron et al., 2018). As the arm race between phages and bacteria continues, new antiviral mechanisms are still awaiting discovery.

Streptococcus mutans and its phages

Despite the significance of *S. mutans* in oral health, very few *S. mutans* phages have been isolated and described in the literature. The isolation of phages specific to *S. mutans* from saliva appears to be an inefficient process (Bachrach et al., 2003; Van Der Ploeg, 2007). Isolating *S. mutans* specific phages from dental plaques has been proposed as a potentially more efficient alternative (Mohamedhussein and Foley, 2020). There are currently only three *S. mutans* phage genomes available in public databases and described in the literature, namely phage M102 (Van Der Ploeg, 2007), M102AD (Delisle et al., 2012), and ϕAPCM01 (Dalmasso et al., 2015). These three phages are highly similar and have a narrow host range. This narrow host range has been proposed as another bottleneck in isolating novel *S. mutans* phages (Bachrach et al., 2003; Mohamedhussein and Foley, 2020). The genome of phage smHBZ8 has recently been made public, but no additional information is available. The first genome sequence of a *S. mutans* strain became available in 2001, when strain UA159 was sequenced (Ajdić et al., 2002). This strain is still often used as reference strain, including for CRISPR-Cas studies, which led to the identification of two CRISPR-Cas systems in this strain (CRISPR1 and CRISPR2). In a large-scale study of spacers acquired by different *S. mutans* strains, 172 of the 305 spacers acquired in CRISPR1 and 55 of the 155 spacers acquired in CRISPR2 showed significant sequence identity to the genome of phage M102. This indicates frequent encounters between *S. mutans* and M102-related phages (van der Ploeg, 2009) and even perhaps that *S. mutans* phages are not diverse. A screening of 171 *S. mutans* genomes revealed prophage-like elements in 24 of them. A total of 35 genomic elements that resembled (partial) prophages were identified, but only three appeared to be complete (Fu et al., 2017).

Streptococcus mutans P42S and phage M102AD

S. mutans P42S, also referred to as P42S-M (Delisle and Rostkowski, 1993), is a spontaneous streptomycin-resistant mutant. It serves as the host strain for the virulent phage M102AD. Phage M102 was first isolated in 1988 in France and the sequencing of its dsDNA genome in Switzerland revealed a size of 31,147 bp (Van Der Ploeg, 2007). A stock of M102 was sent from France to the United States and re-sequenced. Initially thought to be M102, it appeared to be a different but closely related phage, and it was named M102AD. With 30,664 bp,

the M102AD genome is slightly shorter than the M102 genome. The genomes share 90.8% identity at the nucleotide level and out of the 40 M102AD *orfs*, 32 of them share at least 90% identity to their equivalent in M102 (Delisle et al., 2012). Electron microscopy revealed M102AD to be a member of the *Siphoviridae* family with a capsid size of approximately 67 nm and a non-contractile tail of approximately 283 nm in length and 8 nm in width, comparable to M102 (Delisle et al., 2012; Van Der Ploeg, 2007). Nine out of 25 tested *S. mutans* strains were sensitive to phage M102AD and all these strains belonged to serotype c. Among these, there was the strain *S. mutans* P42S (Delisle et al., 2012; Delisle and Rostkowski, 1993). The serotype-dependant host range was also observed for M102 (serotype c), e10 (serotype e) and f1 (serotype f) (Delisle and Rostkowski, 1993). Phage M102 was able to adsorb to the surface of all strains belonging to serotype c, even though it was not able to infect all these strains. The serotypes are differentiated from one another by the different linkage of the glucoside side chains of the rhamnose-glucose polysaccharides (RGPs) found in their cell walls. Transforming strains belonging to serotype e with the serotype c-specific locus resulted in drastically increased phage adsorption rates. This indicates that the receptor for M102 is found in the glucoside side chains of RGPs (Shibata et al., 2009). Considering the serotype-dependant host range of M102AD, the receptor may well be found in these glucoside side chains of RGPs. Phage M102 was not able to infect M102AD host strain P42S (unpublished), which suggests that these two phages do not use the same receptor.

CRISPR-Cas is discussed in detail in Annex B.

Problematic, hypothesis and objectives of the study

Phage-host interactions are complex and despite much progress regarding this topic, still a lot is unknown for many bacterial species. It is even magnified considering that the large majority of host bacteria and phages have yet to be isolated in the laboratory. The evolution rate of bacteria and especially phages are the fastest on the planet. A better understanding of how bacteria and phages evolve, will allow a better understanding of evolution in general.

CRISPR-Cas is a fascinating natural mechanism by which bacteria have learned to defend themselves against viral predators. One aspect that makes it stand out among all other mechanisms is its capacity to acquire new spacers, ie new immunity. This microbial adaptability makes one think of the immune system as found in humans and animals. Yet, the adaptation phase of the CRISPR-Cas systems is the least understood. And with the current development of CRISPR-Cas9 genome editing technology and the set of derivative tools, which mostly rely on the interference step, attention has somewhat shifted away from the adaptation stage of this natural defence system.

Despite the diversity of CRISPR-Cas systems in microbes, very few systems are known to acquire new immunity under laboratory conditions. *Streptococcus mutans* is one of the few species known to harbour active CRISPR-Cas systems. To increase our knowledge regarding the adaptation phase of CRISPR-Cas, the strain *S. mutans* P42S was investigated. We hypothesise that *S. mutans* P42S harbours an active CRISPR-Cas system that enables it to protect itself from phage M102AD infection but using a different specifity compared to other *S. mutans* strains. This hypothesis is based on the fact that this strain carried a seemingly unique Cas9 protein (at least at the beginning of the project). The first objective of this study was therefore to characterise the CRISPR-Cas system of *S. mutans* P42S on the molecular level and to determine its role in antiviral defence (chapters 1 and 2).

As mentioned, CRISPR-Cas has received much attention in recent years due to the remarkable application in genome editing. One bottleneck is the lack of variety in PAMs targeted by Cas9 proteins. The discovery of new Cas9 proteins that recognise different

PAMs would allow a wider scope of sequences to be targeted by genome editing. We hypothesise that the Cas9 of *S. mutans* P42S can be applied in genome editing. The second objective of this study was to determine the potential of SmutCas9 in genome editing (chapter 3).

Chapter 1 – Article 1

Characterisation of a type II-A CRISPR-Cas system in *Streptococcus mutans*

Cas Mosterd[1,2] **and Sylvain Moineau**[1,2,3]*

[1] Département de biochimie, de microbiologie, et de bio-informatique, Faculté des sciences et de génie, Université Laval, Québec City, QC, G1V 0A6, Canada

[2] Groupe de recherche en écologie buccale, Faculté de médecine dentaire, Université Laval, Québec City, QC, G1V 0A6, Canada

[3] Félix d'Hérelle Reference Center for Bacterial Viruses, Faculté de médecine dentaire, Université Laval, Québec City, QC, G1V 0A6, Canada

* Corresponding author. Tel: +1 418 656 3712; Email: Sylvain.Moineau@bcm.ulaval.ca

mSphere 2020. 5, e00235-20. doi: 10.1128/mSphere.00235-20

Keywords

CRISPR, CRISPR-Cas, Cas9, Streptococcus, mutans, bacteriophages, phage, resistance, plasmids, spacers

Résumé

Streptococcus mutans et les phages virulents pouvant l'infecter sont des membres du microbiote oral humain. D'ailleurs, *S. mutans* est le principal agent responsable des caries dentaires. Pour survivre dans cette niche écologique, *S. mutans* code pour différents mécanismes de défense contre les phages, dont le système CRISPR-Cas. Dans ce chapitre, le système CRISPR-Cas de type II-A de *S. mutans* souche P42S est caractérisé en détail. Ce dernier démontre une activité naturelle d'adaptation et d'interférence en réponse à l'entrée d'ADN exogène, soit suite à une infection virale ou à la transformation d'un plasmide. Les espaceurs nouvellement acquis sont ajoutés à l'extrémité 5' du locus CRISPR, mais aussi de manière ectopique dans quelques cas. La comparaison des gènes *cas* de la souche P42S à ceux d'autres souches de *S. mutans* a permis de constater que les gènes *cas1*, *cas2* et *csn2* sont fortement conservés chez cette espèce. Cependant, la diversité est plus élevée du côté de *cas9*. Bien que les domaines nucléases demeurent similaires, l'extrémité C-terminale de la protéine qui inclue le domaine de reconnaissance du PAM est moins conservée. En support à ces résultats, nos expériences ont démontré que les PAMs associés avec SmutCas9 de la souche P42S sont NAA et NGAA. Ces PAMs sont différents de ceux publiés précédemment pour le système CRISPR-Cas de la souche modèle *S. mutans* UA159. Cette étude illustre la diversité des systèmes CRISPR-Cas de type II-A que l'on retrouve à l'intérieur d'une même espèce bactérienne.

Abstract

Streptococcus mutans and its virulent phages are important members of the human oral microbiota. *S. mutans* is also the primary causal agent of dental caries. To survive in this ecological niche, *S. mutans* must encode phage defence mechanisms, which include CRISPR-Cas systems. Here, we describe the CRISPR-Cas type II-A system of *S. mutans* strain P42S, which was found to display natural adaptation and interference activity in response to phage infection and plasmid transformation. Newly acquired spacers were integrated both at the 5' end of the CRISPR locus and ectopically. In comparisons of the *cas* genes of P42S to those of other *S. mutans* strains, *cas1*, *cas2* and *csn2* appear to be highly conserved within the species. However, more diversity was observed with *cas9*.

While the nuclease domains of SmutCas9 are conserved, its C-terminus, including the PAM recognition domain, is less conserved. In support of these findings, we experimentally demonstrated that the PAMs associated with SmutCas9 of strain P42S are NAA and NGAA. These PAMs are different from those previously reported for the CRISPR-Cas system of the model strain *S. mutans* UA159. This study illustrates the diversity of CRISPR-Cas type II-A systems that can be found within the same bacterial species.

Importance

CRISPR-Cas is one of the mechanisms used by bacteria to defend against viral predation. Increasing our knowledge of the biology and diversity of CRISPR-Cas systems will also improve our understanding of viral-bacterial interactions. As CRISPR-Cas systems acquiring novel immunity under laboratory conditions are rare, *S. mutans* P42S provides an alternative model to study the adaptation step, which is still the least understood step in CRISPR-Cas biology. Furthermore, the availability of a natural Cas9 protein recognising an AT-rich PAM opens up new avenues for genome editing purposes.

Abbreviations

BHI	Brain Heart Infusion
BIM	bacteriophage-insensitive mutant
bp	base pair
cas	CRISPR-associated
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR-RNA
CSP	competence stimulating peptide
OD_{600}	optical density at 600 nanometres
PAM	protospacer adjacent motif
PFU	plaque forming unit
tracrRNA	trans-activating crRNA
UV	ultra-violet

WT	wild-type
SmaCas9	*Streptococcus macacae* Cas9
SmutCas9	*Streptococcus mutans* Cas9
SpyCas9	*Streptococcus pyogenes* Cas9

Introduction

More than 500 different bacterial species can be found in the human oral cavity, although very few of them can cause diseases (Paster et al., 2001). *Streptococcus mutans* is a Gram-positive bacterial species associated with dental caries, which is the most common oral disease. *S. mutans* metabolises carbohydrates transiently passing through the mouth into various acids, including lactic acid (Loesche, 1986). The resulting pH reduction demineralises the hard tissue of the teeth, and the net loss of minerals over time leads to the formation of dental caries (Fejerskov, 1997). *S. mutans* is also resistant to many environmental conditions (Lemos and Burne, 2008), and its capacity to favour dental caries is likely due to a combination of its adhesion abilities, production of acids and relative resistance to low pH (Banas, 2004).

Viruses are the most abundant biological entities in characterised Earth ecosystems and globally they can infect all hosts, including bacteria (Breitbart and Rohwer, 2005). Bacterial viruses (phages) play a role in the regulation of bacterial populations including in the oral microbiota (Bachrach et al., 2003). However, very few lytic *S. mutans* phages have been isolated and described in the literature (Delisle and Rostkowski, 1993; Van Der Ploeg, 2007). For example, only the genomic sequences of phage M102 (Van Der Ploeg, 2007), M102AD (Delisle et al., 2012) and ϕAPCM01 (Dalmasso et al., 2015) are currently available in public databases.

To survive in phage-containing environments, bacteria have developed an impressive arsenal of anti-phage mechanisms (Doron et al., 2018; Labrie et al., 2010). One of these numerous mechanisms is the CRISPR-Cas system. CRISPR (clustered regularly interspaced, short palindromic repeats) refers to a series of short palindromic nucleotide repeats interspaced with similarly sized ‘spacers’ and these arrays are found in less than half of bacteria (Ishino et al., 2018), including in *S. mutans* (van der Ploeg, 2009). Along with a set of associated genes (*cas)*, this system acts as a microbial adaptive immune system (Barrangou et al., 2007). To date, six different types of CRISPR-Cas systems have been identified and divided into several subtypes (Makarova et al., 2019, 2018). Although

there are significant differences at the molecular level between the various types, they mostly function using a similar process.

First, short DNA proto-spacers from infecting phages (often from defective) (Hynes et al., 2014) or plasmid sequences (Garneau et al., 2010) are integrated into the CRISPR array as spacers in a process known as adaptation. An AT-rich sequence called the leader sequence, is often found directly upstream of the CRISPR array and usually contains a promoter that allows transcription of the array into pre-crRNA (Jansen et al., 2002; Pougach et al., 2010; Pul et al., 2010). The pre-crRNA is then matured into small RNA molecules (Brouns, 2008; Deltcheva et al., 2011). In type II systems, the small RNAs, also known as crRNA, are associated with Cas9 inside bacterial cells to recognise and cleave subsequent invading nucleic acids with sequences identical to the spacer (Garneau et al., 2010). The DNA cutting activity observed with type II systems also requires the presence of a short nucleotide motif called the proto-spacer adjacent motif (PAM), next to the target DNA (Deveau et al., 2008; Horvath et al., 2008). This ability to target and to specifically cleave DNA has led to many applications, including in genome editing (Gasiunas et al., 2012; Jinek et al., 2012). Another unique feature of type II systems is the requirement of tracrRNA (trans-activating RNA). These are small RNA molecules, which possess nucleotides of complementarity with the repeat regions of crRNAs. The complementarity will allow the formation of an RNA duplex which in turn facilitates crRNA maturation (Deltcheva et al., 2011).

In a previous study (van der Ploeg, 2009), it was noted that 19 out of 27 (70%) examined *S. mutans* strains possessed a type II-A CRISPR-Cas system, which consists of the four genes *cas9*, *cas1*, *cas2* and *csn2*. Moreover, 9 of the same 27 strains (33%) possessed a type I-C CRISPR-Cas system, consisting of seven genes, which are *cas3*, *cas5*, *cas8c*, *cas7*, *cas4*, *cas1* and *cas2*, while 15% of them (4 out of 27) possessed both types. Interestingly, 56% of the spacers (172 out of the 305 spacers) in the various CRISPR arrays of these *S. mutans* strains had homology to the genome of the virulent siphophage M102. Bioinformatic analyses also suggested that the type I-C system in the model *S. mutans* strain UA159 was inactive due to truncated *cas1* and *cas8c*. On the other hand, spacer acquisition was experimentally demonstrated for the type II-A system. Indeed, a 5'-end

expansion of the CRISPR array was observed in bacteriophage-insensitive mutants (BIMs) isolated following exposure of the wild-type *S. mutans* strain UA159 to the phage M102 (van der Ploeg, 2009). Surprisingly, disruption of the type II CRISPR-Cas system did not restore phage sensitivity, suggesting the presence of additional antiviral systems (Serbanescu et al., 2015; van der Ploeg, 2009). Nonetheless, the PAM sequence recognised by the CRISPR-Cas system of strain UA159 was proposed to be 5'-NGG-3' (van der Ploeg, 2009).

During the characterisation of the virulent siphophage M102AD (Delisle et al., 2012) it was demonstrated that it shares 90.8% identity at the nucleotide level with phage M102. Phage M102AD replicates on the host strain *S. mutans* P42S but not on *S. mutans* strain UA159. Here, we investigated the interactions between phage M102AD and its host *S. mutans* P42S. We showed the presence of an active type II-A CRISPR-Cas system in *S. mutans* P42S following the characterisation of bacteriophage-insensitive mutants (BIMs) obtained after a challenge with phage M102AD. However, bioinformatic analyses and functional studies indicated that this system recognises a different PAM.

Results

Analysis of the CRISPR-Cas systems of *S. mutans* P42S

Whole genome sequencing of *S. mutans* P42S revealed one CRISPR locus consisting of five spacers of 29 to 31 bp in length (Figure 1) separated by the five identical 36-bp repeat sequences (5'-GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC-3') and a terminal repeat that possessed a mutation at the final bp (5'-GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAA**T**-3'). The same repeat sequence was observed in CRISPR arrays of other *S. mutans* genomes, including in UA159 (van der Ploeg, 2009). Of note, the third spacer in the CRISPR array of P42S had a stretch of 19 out of 20 bp identical to a segment of a gene of unknown function in the genome of phage M102AD. The other four spacers did not share any significant sequence identity with sequences in public databases, including spacers found in other *S. mutans* strains in the CRISPR database. Upstream of the CRISPR array, four *cas* genes associated to a type II-A

system were found, namely *cas9*, *cas1*, *cas2*, and *csn2* (Figure 1). A *tracrRNA* was also found upstream of the *cas9* gene. No type I-C system was detected in this strain.

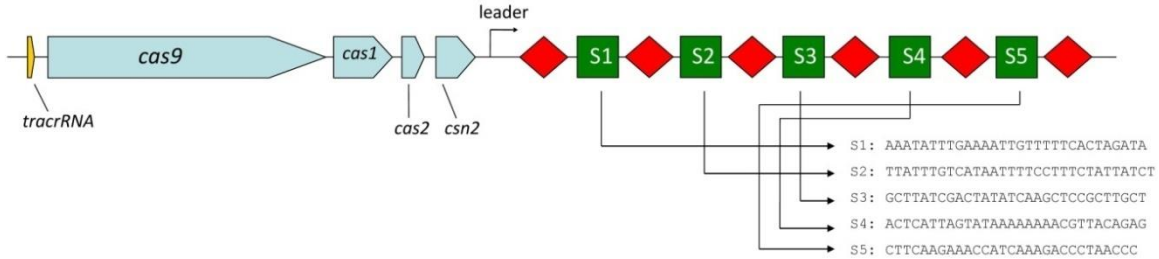


Figure 1: Type II-A CRISPR-Cas system of *S. mutans* P42S. From 5' to 3'.The *tracrRNA* is in yellow, *cas* genes are in blue, leader is a black arrow, repeats are represented as red diamonds and spacers as green squares.

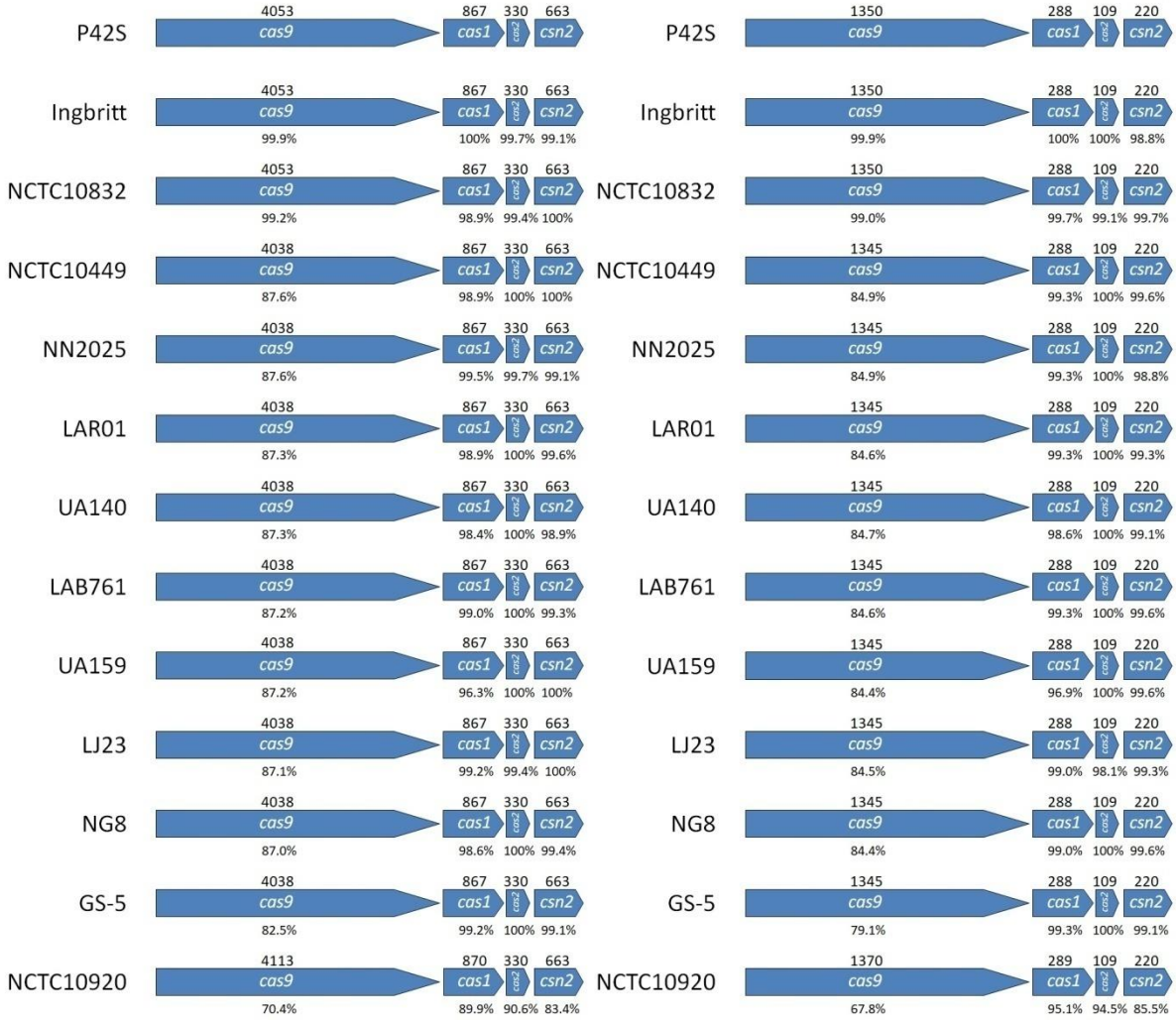
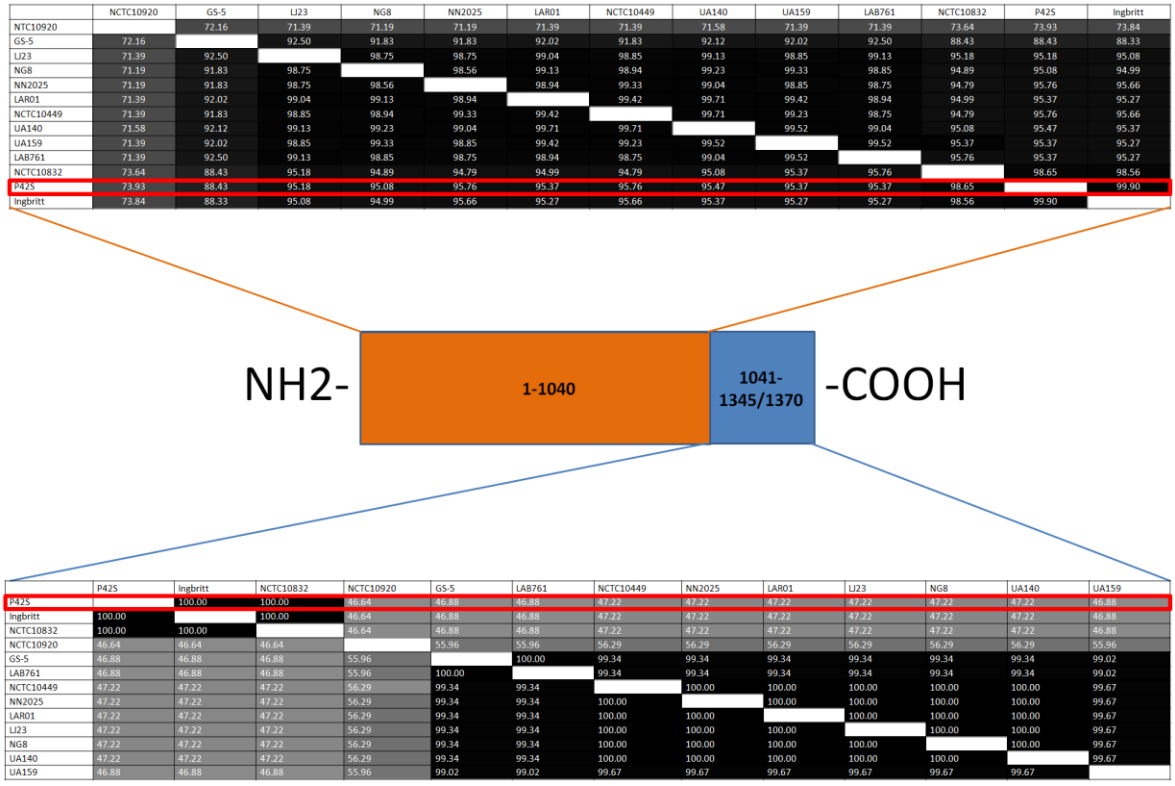


Figure 2: Comparison of *cas* genes (left) and Cas proteins (right) of nine *S. mutans* strains. Percent identities to the sequences found in *S. mutans* P42S are indicated.

The whole genome sequences are publicly available for 12 strains of *S. mutans* which possess a CRISPR array with the same repeat sequence as found in *S. mutans* P42S. The CRISPR arrays in these strains contained between 3 and 70 spacers, with an average of 21 spacers. Remarkably, out of the 245 spacers detected in these strains, 89 of them partially matched phage M102AD genome. Nine of these spacers matched 100% part of the phage M102AD genome. Only one (strain NCTC10920) did not possess any spacer with a level of identity to the genome of phage M102AD. The nucleotide sequences of the *cas* genes and deduced proteins of *S. mutans* P42S were compared to those found in the 12 strains mentioned above. The results are illustrated in Figure 2.

There was a high degree of identity with 11 of the 12 *S. mutans* strains in the *cas1* (96.3% to 100%), *cas2* (99.4% to 100%), and *csn2* (99.1% to 100%) sequences. Interestingly, more diversity was observed between the *cas9* sequences of these strains (82.5% to 99.9%), with the *cas9* sequence of P42S being only similar to that of strains Ingbritt (99.9%) and NCTC10832 (99.2%). Strain NCTC10920 was somewhat of an outlier sharing the lowest identity to *cas9* (70.4%), *cas1* (89.9%), *cas2* (90.6%) and *csn2* (83.4%) of strain P42S.

Similar observations were made with the deduced amino acid sequences of Cas1 (96.9% to 100% identity), Cas2 (98.1% to 100%), Csn2 (98.8% to 99.7%), and Cas9 (79.1% to 99.9%). The lowest identity of P42S Cas proteins again was with those of NCTC10920 (Cas9/67.8%, Cas1/95.1%, Cas2/94.5%, Csn2/85.5%). Further investigations revealed that the difference between the Cas9 protein sequences mostly lies at the C terminus. In comparing the first 1040 amino acids (out of the total of 1345 to 1370) of P42S Cas9 to the ones found in other *S. mutans* strains, the identity to 10 of the 12 other Cas9 proteins ranged between 95.1% and 99.9%. The sequence identity dropped to 88.4% for the Cas9 of *S. mutans* strain GS-5 and 73.9% for NCTC10920. For the remaining 305 to 330 amino acids at the C terminus, the percentage of sequence identity between P42S Cas9 and the other *S. mutans* Cas9 proteins drops drastically. Sequence identity of P42S Cas9 C-terminus was 100% with strain Ingbritt and NCTC10832 but declined to between 46.6% and 47.2% with the other 10 strains. This analysis is summarised in Figure 3.



	NCTC10920	GS-5	LJ23	NG8	NN2025	LAR01	NCTC10449	UA140	UA159	LAB761	NCTC10832	P42S	Ingbritt
NCTC10920		72.16	71.39	71.19	71.19	71.39	71.39	71.58	71.39	71.39	73.64	73.93	73.84
GS-5	72.16		92.50	91.83	91.83	92.02	91.83	92.12	92.02	92.50	88.43	88.43	88.33
LJ23	71.39	92.50		98.75	98.75	99.04	98.85	99.13	98.85	99.13	95.18	95.18	95.08
NG8	71.19	91.83	98.75		98.56	99.13	98.94	99.23	99.33	98.85	94.89	95.08	94.99
NN2025	71.19	91.83	98.75	98.56		98.94	99.33	99.04	98.85	98.75	94.79	95.76	95.66
LAR01	71.39	92.02	99.04	99.13	98.94		99.42	99.71	99.42	98.94	94.99	95.37	95.27
NCTC10449	71.39	91.83	98.85	98.94	99.33	99.42		99.71	99.23	98.75	94.79	95.76	95.66
UA140	71.58	92.12	99.13	99.23	99.04	99.71	99.71		99.52	99.04	95.08	95.47	95.37
UA159	71.39	92.02	98.85	99.33	98.85	99.42	99.23	99.52		99.52	95.37	95.37	95.27
LAB761	71.39	92.50	99.13	98.85	98.75	98.94	98.75	99.04	99.52		95.76	95.37	95.27
NCTC10832	73.64	88.43	95.18	94.89	94.79	94.99	94.79	95.08	95.37	95.76		98.65	98.56
P42S	73.93	88.43	95.18	95.08	95.76	95.37	95.76	95.47	95.37	95.37	98.65		99.90
Ingbritt	73.84	88.33	95.08	94.99	95.66	95.27	95.66	95.37	95.27	95.27	98.56	99.90	

	P42S	Ingbritt	NCTC10832	NCTC10920	GS-5	LAB761	NCTC10449	NN2025	LAR01	LJ23	NG8	UA140	UA159
P42S		100.00	100.00	46.64	46.88	46.88	47.22	47.22	47.22	47.22	47.22	47.22	46.88
Ingbritt	100.00		100.00	46.64	46.88	46.88	47.22	47.22	47.22	47.22	47.22	47.22	46.88
NCTC10832	100.00	100.00		46.64	46.88	46.88	47.22	47.22	47.22	47.22	47.22	47.22	46.88
NCTC10920	46.64	46.64	46.64		55.96	55.96	56.29	56.29	56.29	56.29	56.29	56.29	55.96
GS-5	46.88	46.88	46.88	55.96		100.00	99.34	99.34	99.34	99.34	99.34	99.34	99.02
LAB761	46.88	46.88	46.88	55.96	100.00		99.34	99.34	99.34	99.34	99.34	99.34	99.02
NCTC10449	47.22	47.22	47.22	56.29	99.34	99.34		100.00	100.00	100.00	100.00	100.00	99.67
NN2025	47.22	47.22	47.22	56.29	99.34	99.34	100.00		100.00	100.00	100.00	100.00	99.67
LAR01	47.22	47.22	47.22	56.29	99.34	99.34	100.00	100.00		100.00	100.00	100.00	99.67
LJ23	47.22	47.22	47.22	56.29	99.34	99.34	100.00	100.00	100.00		100.00	100.00	99.67
NG8	47.22	47.22	47.22	56.29	99.34	99.34	100.00	100.00	100.00	100.00		100.00	99.67
UA140	47.22	47.22	47.22	56.29	99.34	99.34	100.00	100.00	100.00	100.00	100.00		99.67
UA159	46.88	46.88	46.88	55.96	99.02	99.02	99.67	99.67	99.67	99.67	99.67	99.67	

Figure 3: Percentage identity between Cas9 N-terminus and C-terminus found in several *S. mutans* strains.

A *tracrRNA* sequence was found upstream of the *cas9* gene of *S. mutans* P42S. The tracrRNA of *S. mutans* strain UA159 was previously estimated to be 107 nucleotides long (Deltcheva et al., 2011). The tracrRNA of *S. mutans* P42S was estimated to be 93 nucleotides due to a deletion. However, within the anti-repeat region of the tracrRNA there was almost full identity to the tracrRNA of *S. mutans* UA159, with the exception of two positions, where there were inversions. Still, there was a stretch of 25 nucleotides where 24 nucleotides of the tracrRNA matched the repeat sequence (Figure 4). RNA duplex formation is necessary for the maturation of the pre-crRNA (Deltcheva et al., 2011).

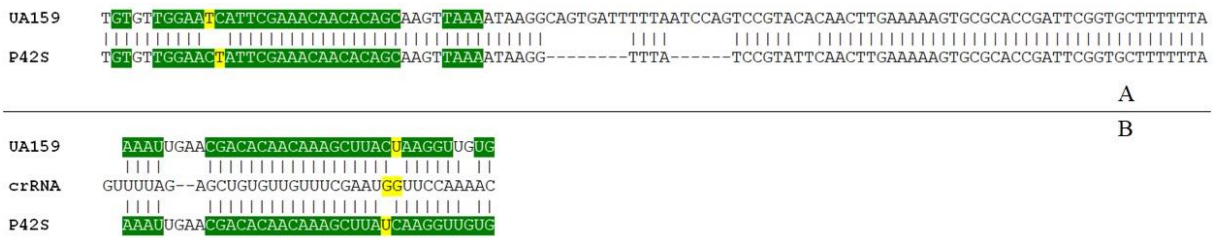


Figure 4: tracrRNA in *S. mutans*. (A): Comparison of predicted *tracrRNA* in *S. mutans* UA159 and P42S. Complementarity to crRNA is highlighted in green. (B): The anti-repeat region within the tracrRNA of *S. mutans* UA159 and P42S compared to crRNA.

BIM assays

To determine if the type II-A CRISPR-Cas system of *S. mutans* P42S was functional, we performed BIM (bacteriophage-insensitive mutant) assays as described previously (Barrangou et al., 2007). Based on the random screening of 100 colonies, 20% of them acquired at least one new spacer after exposure to the virulent phage M102AD. In separate assays, when the phage lysate was exposed to UV light prior to the phage challenge assay (Hynes et al., 2014), the fraction of BIMs that had acquired a new spacer increased to 68% (19 out of 28 colonies tested). Overall, most BIMs acquired one or two new spacers, but the acquisition of up to seven new spacers was even observed in one BIM.

Newly acquired spacers were between 28 and 32 bp long, with 30 bp being the most frequent. Out of the 168 unique acquired spacers, 114 were 30 bp (68%), 46 were 31 bp (27%), five were 32 bp (3%), two were 29 bp (1%) and one was 28 bp (<1%) long. The details of all the 168 acquired spacers can be found in Table S1 (Supplementary Data). Of these 168 acquired spacers, 148 of them (88%) were identical (100%) to a section of the phage M102AD genome. In addition, 10% (16 out of 168) of them had one or two nucleotides mismatches compared to the genome of M102AD. The remaining 2% (4 out of 168) had low (1 spacer) or no identity (3 spacers) to the genome of M102AD. Surprisingly, among these there was one spacer (spacer 101) that perfectly matched the genome of *S. mutans* P42S.

In related type II-A systems, new spacers are typically integrated at the 5'-end of the CRISPR locus. The sequence upstream of the first repeat in the CRISPR array was shown to have an important role in specifying the integration site, since mutations in this leader sequence result in the integration of spacers within the CRISPR array (McGinn and Marraffini, 2016). This ectopic spacer acquisition within the array following phage infection was observed in another *Streptococcus* species, namely *S. thermophilus* (Achigar et al., 2017). In *S. mutans* P42S, novel spacers were integrated both at the 5'-end or within the CRISPR locus. Ectopic spacers in *S. mutans* P42S were acquired between spacers 4 and 5, except one, which was integrated between spacers 1 and 2. Ectopic spacer acquisition was broadly observed in BIMs which acquired multiple spacers, with at least one at the 5'-

end of the CRISPR locus. Indeed, only two BIMs acquired a single spacer within the array. In both of these BIMs, the acquired spacer perfectly matched the genome of phage M102AD.

Identification of the proto-spacer adjacent motif (PAM)

Of the 168 newly acquired spacers, 165 could be mapped as proto-spacers on the phage M102AD genome (Table S1, Supplementary Data). The 10 bp upstream and downstream of each of the corresponding proto-spacers in the phage genome were analysed for the presence of the proto-spacer adjacent motif (PAM). No particular motif was detected upstream of the proto-spacers. However, there was a clear preference for a 5'-NAAA-3' motif downstream of the proto-spacers (Figure 5).

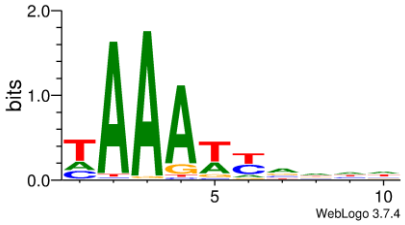


Figure 5: PAM downstream of protospacers.

Indeed, 79% of the proto-spacers (131 out of 165) were flanked by the NAAA sequence. The adenine on position 4 was the least conserved. When considering only the NAA motif, the percentage of proto-spacers flanked by this sequence increased to 93% (153/165). The most commonly observed PAM was TAAAT, which was found next to 21% (35/165) of the proto-spacers. When considering only the spacers acquired at the 5'-end of CRISPR loci, 97% (85/88) of the proto-spacers were flanked by the trinucleotide NAA and 27% (24/88) were flanked by TAAAT. When analysing just the spacers integrated at an ectopic position, 82% (19/23) of the proto-spacers were flanked by NAA and 13% (3/23) were flanked by TAAAT. Of interest, in BIMs that have acquired a single spacer, 100% (35/35) of the proto-spacers were flanked by NAA (35/35), while 35% (12/35) of them were flanked by TAAAT. An overview of the PAMs and their relative abundance is found in Table 1.

Table 1: Relative frequencies of acquired PAMs.

PAM (5'-3')	Frequency	Percentage
TAAAT	35	21.2
TAAAA	23	13.9
CAAAT	19	11.5
AAAAT	18	10.9
CAAAA	11	6.7
TAAGT	10	6.1
AAAAA	7	4.2
TAAAG	7	4.2
AAAAG	3	1.8
AAAGT	3	1.8
TTAAA	3	1.8
TAAAC	2	1.2
GAAAT	2	1.2
AAAAC	1	0.6
AAATT	1	0.6
AAATC	1	0.6
AAACC	1	0.6
AAAGG	1	0.6
AAATG	1	0.6
AAGTG	1	0.6
AAGCT	1	0.6
ATAAA	1	0.6
TAATA	1	0.6
TAACA	1	0.6
TAAGA	1	0.6
TAAGC	1	0.6
TACAG	1	0.6
TCAAA	1	0.6
TCGCC	1	0.6
TGAAA	1	0.6
CAAAC	1	0.6
CAAAG	1	0.6
CAAGT	1	0.6
CAAGG	1	0.6
GAAAC	1	0.6

Phage resistance assays

As indicated above, between one and seven new spacers were acquired per BIM. It was previously shown that the acquisition of multiple spacers increases the overall phage resistance (Deveau et al., 2008). Thus, we evaluated whether the number of acquired spacers influences the level of resistance towards phage M102AD. This was accomplished by comparing the phage titres on the various BIMs and the phage-sensitive WT host. No effect on the number of newly acquired spacers was observed as all tested BIMs were fully phage-resistant.

Table 2: Phage adsorption assay.

Strain	No. of spacers	Spacers acquired (5'-3')	Targeted sequence on M102AD genome	Percentage of phage M102AD adsorption (n = 2)
WT	N/A	N/A	N/A	84 ± 1
BIM4	7	sp3: AACGCTCTGATTTCGTGTTTGTGTTATCGCC sp4: GTCAAGCATACGTATATATGCTTGTGCACT sp5: TGATGAAGTAAACCTCTTTTGTGAAAGGATT sp6: TTAGCGCGAGTGATGATGGGTTGGTAATTGCC sp7: CTCTAGCTTTATCTATTTTGATAAAGACAC sp8: GTACACTCTGCAACTAACCCATCGGCACCA sp9: CACGTCGAGTAAAATTGTACTAGCGCCTAA	14311-14284 25547-25518 24368-24338 16891-16922 14862-14891 7587-7616 X	54 ± 4
BIM5	7	sp10: AACGCTCTGATTTCGTGTTTGTGTTATCGCC sp4: GTCAAGCATACGTATATATGCTTGTGCACT sp5: TGATGAAGTAAACCTCTTTTGTGAAAGGATT sp6: TTAGCGCGAGTGATGATGGGTTGGTAATTGCC sp7: CTCTAGCTTTATCTATTTTGATAAAGACAC sp8: GTACACTCTGCAACTAACCCATCGGCACCA sp9: CACGTCGAGTAAAATTGTACTAGCGCCTAA	14314-14284 25547-25518 24368-24338 16891-16922 14862-14891 7587-7616 X	35 ± 5
BIM6	2	sp11: AAATTTTATAGCATATGCGAATATTGTTGT sp12: GTTAACCGCAAGCGTAAAGTTTGCATATGC	27729-27700 27896-27867	56 ± 5
BIM7	4	sp13: AGAATTTTTCCATTCTTGCTCTTGGTTGGT sp14: AGATGATAGTGACTTGTTTGCGGTAATTAA sp15: AACTCTAACACTGGCTATTACTGATAAGAC sp16: GAATTTTTCCATTCTTGCTCTTGGTTGGTT	15737-15708 3074-3103 15938-15967 15736-15707	38 ± 3
BIM9	3	sp18: TGGTTTGCACATTTTTTTTCCTTCCTTTTT sp19: TAAGATTACATTTTGCAAGTAATCTTTCTT sp21: GAATTGGGTTTTCCACAGTAGTAGCAAAGA	26698-26669 22856-22827 294-265	20 ± 7
BIM24	5	sp38: GCTAGTGACGTTAAAGATTTTGATGATAAT sp39: TAAACACAAAGAAGCTTTGCAAGCCGTCGG sp40: GCAGACAAAAGCTAAACAAGCTCTTGACTAT sp41: GCTACTCGTATGTTGGATGTTATCGACGCC sp1: CTCTTTTAGCAATTGTGAAAGGACGTAATT	25233-25262 10138-10167 8584-8614 9878-9907 24306-24335	57 ± 12
BIM29	3	sp11: AAATTTTATAGCATATGCGAATATTGTTGT sp51: ACGATGAAGTAAACCTCTTTTGTGAAAGGATT sp52: CTAAAACCGCAAGACACAGAGCCACAGGCT	27729-27700 24367-24338 4463-4492	5 ± 3
BIM31	3	sp54: TTCTTTACTAGTAAAACTTCTGTACATTTA sp51: ACGATGAAGTAAACCTCTTTTGTGAAAGGATT sp52: CTAAAACCGCAAGACACAGAGCCACAGGCT	196-167 24367-24338 4463-4492	17 ± 17
BIM33	2	sp56: GCCTTTAGACGAATGTATCCAAAATGTATCC sp55: CATTAAAAGCTATGCGCAATAAGGACTATGT	22183-22213 26770-26799	9 ± 2
BIM34	6	sp57: GTTTTGGCTGTAACGTCTTTGACAACGCCG sp58: TTTTGTAACTGCGTATCATCAGCGCTCGAG sp51: ACGATGAAGTAAACCTCTTTTGTGAAAGGATT sp59: CTTGCGATGTGGACAAATTGGGGCACGGTCA sp51: ACGATGAAGTAAACCTCTTTTGTGAAAGGATT sp52: CTAAAACCGCAAGACACAGAGCCACAGGCT	5500-5471 16567-16538 24367-24338 6897-6927 24367-24338 4463-4492	9 ± 9
BIM35	4	sp60: TAAAGCGTTTTGGATTAACTGCGCTTTAGC sp61: CGCATAGAGTTTTGAGAGGTGAAGAATGTTT sp51: ACGATGAAGTAAACCTCTTTTGTGAAAGGATT sp52: CTAAAACCGCAAGACACAGAGCCACAGGCT	4783-4754 23905-23935 24367-24338 4463-4492	10 ± 10

One striking observation was that one of the BIMs acquired a single spacer that did not target phage M102AD, yet the strain was still fully resistant to this phage. In this BIM, the phage resistance was non-CRISPR related and another resistance mechanism is likely at play. Another frequent defence system leading to BIMs is the mutation of a phage receptor at the cell surface (Labrie et al., 2010). To investigate this possibility, phage adsorption assays were performed using the WT strain and eleven BIMs that have acquired various spacers. While adsorption of phage M102AD to the WT strain was at 84% after 15 minutes of co-incubation, adsorption to the BIMs was reduced between 5 to 57% (Table 2). The genome of one adsorption-resistant BIM was fully sequenced (BIM5). Two mutations were found in BIM5 compared to the WT sequence, outside the CRISPR locus. However, PCR amplifications of these regions in other BIMs did not show the same mutations.

Plasmid interference assays

To confirm the CRISPR interference activity, we designed a plasmid-based assay. - In the first experiment, the construct pNZ123-sp1 was transformed into *S. mutans* P42S WT. This construct harbours a proto-spacer sequence that is targeted by one of the native spacers already present in the CRISPR locus of the *S. mutans* P42S WT strain, along with the most common PAM sequence (TAAAT). Therefore, if the interference activity of the type II-A CRISPR-Cas system of *S. mutans* P42S is functional, the transformation of this plasmid should be prevented or greatly reduced. Indeed, while pNZ123 was transformable in this strain, pNZ123-sp1 was not transformable, validating the CRISPR interference activity as well as the PAM (Table 3).

A similar experiment was performed by transforming pNZ123 and the construct pNZ123-sp2 into BIM1. This *S. mutans* BIM derivative acquired a single spacer at the 5'-end of the CRISPR array from the phage M102AD genome and the matching proto-spacer (and its PAM) was cloned into pNZ123-sp2. Transformation of pNZ123 was successful in BIM1 but transformation of pNZ123-sp2 was prevented, confirming the CRISPR-Cas interference activity and the PAM (Table 3).

To confirm that ectopically acquired spacers also conferred protection, we used pNZ123-sp3 and BIM2. This derivative strain acquired a single spacer between the native spacers 4 and 5 of the P42S CRISPR array. pNZ123-sp3 was constructed by cloning the matching proto-spacer and its PAM into pNZ123. Transformation of pNZ123 was successful in BIM2, but transformation of pNZ123-sp3 was inhibited, confirming that the ectopic spacer within the CRISPR array of *S. mutans* P42S provides protection against invading nucleic acids (Table 3).

Table 3: Plasmid interference assays.

Strain	Construct name	Insert sequence (5'-3')	Clones per 10 µg plasmid DNA*
WT	pNZ123	N/A	110 ± 28
	pNZ123-sp1	AAATATTTGAAAATTGTTTTTCACTAGATA-*TAAAT*	0
BIM1	pNZ123	N/A	71 ± 30
	pNZ123-sp2	CTCTTTTAGCAATTGTGAAAGGACGTAATT-*TAAAT*	0
BIM2	pNZ123	N/A	33.5 ± 7.5
	pNZ123-sp3	TTTTGGTCTAAAATTCTCAGGAATTTCACC-*TAAAT*	0

No CSP was used during this experiment. *n=2

Plasmid-based determination of PAM sequence

Because the abovementioned plasmid interference assay was effective, we further investigated the 3'-end PAM associated with the type II-A CRISPR-Cas system of *S. mutans* P42S. Although there was evidence that the PAM is either NAA or NAAA, a plasmid interference assay was designed in which plasmid constructs with various PAMs were transformed. Each pNZ123-derived construct harboured a proto-spacer sequence targeted by the native spacer at the 5-end of the CRISPR locus of *S. mutans* P42S. The proto-spacer was flanked by a 5-bp PAM, which was modified by at least one bp. A total of 18 constructs with alternative PAMs were transformed into P42S. Plasmids pNZ123 and pNZ123-sp1 were used as controls. An overview of the transformation efficiencies can be found in Table 4.

When the PAM was changed for any other bp on position 1, 4 or 5, the interference activity was not affected, and the transformation efficiencies of these plasmids were similar to pNZ123-sp1. This implies that these nucleotides do not play a significant role in CRISPR recognition. However, when the nucleotide was changed at positions 2 and 3,

there was a drastic change in transformability. If any nucleotide other than the adenine was found at position 3, the CRISPR-Cas system did not interfere with transformation. If the adenine located at position 2 was exchanged for either a cytosine or a thymine, the same phenomenon occurred. However, if the adenine at position 2 was exchanged for a guanine, transformation interference was noted.

Table 4: Plasmid interference assays to determine the PAM. Modified nucleotide of PAM is underlined. The darker the boxes, the stronger the interference.

PAM	Clones per µg DNA	PAM	Clones per µg DNA	PAM	Clones per µg DNA
AAAAT	$3.0 \pm 2.0 \times 10^{1}$	CAAAT	$2.5 \pm 2.5 \times 10^{0}$	GAAAT	$7.5 \pm 2.5 \times 10^{0}$
TCAAT	$9.0 \pm 2.0 \times 10^{4}$	TGAAT	$9.8 \pm 1.8 \times 10^{1}$	TTAAT	$1.9 \pm 1.3 \times 10^{4}$
TACAT	$4.3 \pm 0.8 \times 10^{4}$	TAGAT	$2.5 \pm 2 \times 10^{5}$	TATAT	$1.9 \pm 1.6 \times 10^{5}$
TAACT	$2.5 \pm 1.0 \times 10^{1}$	TAAGT	$2.8 \pm 0.3 \times 10^{1}$	TAATT	$4.0 \pm 1.5 \times 10^{1}$
TAAAA	$4.0 \pm 2.5 \times 10^{1}$	TAAAC	$3.8 \pm 2.8 \times 10^{1}$	TAAAG	$5.5 \pm 0 \times 10^{1}$
TGATT	$1.5 \pm 0 \times 10^{4}$	TGACT	$3.3 \pm 1.3 \times 10^{4}$	TGAGT	$1.8 \pm 1 \times 10^{5}$
TAAAT	$4.5 \pm 1.5 \times 10^{1}$				
pNZ123	$1.1 \pm 0.4 \times 10^{5}$				

CSP was used during this experiment. *n=2

To investigate the importance of the flanking nucleotides when a guanine is at position 2 of the PAM, an additional series of plasmid constructs were made in which the adenine at position 4 was also changed. When the transformation efficiencies of the new constructs were compared with the previously described plasmids, it became apparent that the guanine at position 2 was accepted only if the nucleotide as position 4 was an adenine.

Discussion

A functional type II-A CRISPR-Cas system was uncovered in the genome of *S. mutans* strain P42S. The *tracrRNA* as well as the *cas1*, *cas2* and *csn2* genes shared a high percentage of identity with the similar loci found in other *S. mutans* genomes. Less conservation is found between the *cas9* genes. Functional and structural studies have previously shown that nuclease domains of Cas9 are found in the first part of the protein, which appears to be highly conserved in *S. mutans*. The C-terminal of Cas9 is where the PAM-interacting domain is usually found (Anders et al., 2014; Nishimasu et al., 2014). Previously, the PAM recognised by the type II-A system of *S. mutans* strain UA159 was

determined to be NGG based on DNA cleavage assays with purified Cas9 (Fonfara et al., 2014). While the C-terminal sequence of UA159 Cas9 has 99% or more identity with the same Cas9 region found in eight other *S. mutans* strains, the identity significantly drops when compared to the Cas9 C-terminal from strains P42S, Ingbritt and NCTC10832 (46.9%) and from NCTC10920 (56.0%) (Figure 3). These data suggest that these latter Cas9 proteins are structurally different in terms of their PAM recognition domain and, therefore, likely recognise different PAMs, which was experimentally confirmed here with Cas9 of strain P42S.

The adaptive nature of the CRISPR-Cas system of *S. mutans* P42S was also confirmed when spacer acquisition was observed after exposure to the virulent siphophage M102AD. Ectopic spacer acquisition was also noted, a phenomenon previously described in *S. thermophilus*. In a particular *S. thermophilus* strain in which BIM-derivatives had acquired new ectopic spacers, the leader sequence had a deletion of the last bp before the beginning of the first repeat compared to other strains (Achigar et al., 2017). In comparisons of this last bp of the leader sequence of BIMs of *S. mutans* P42S that have acquired ectopic spacers, no such deletion was observed. In *S. thermophilus,* the dinucleotide AG was at the 3'-end of the spacers found upstream of the newly acquired ectopic spacers (Achigar et al., 2017). The same AG dinucleotide was found in spacer 4 of *S. mutans* P42S which preceded most of the ectopic spacers that were acquired in the strain, but not in spacer 1.

Phage adsorption assays revealed that in parallel to spacer acquisition, surface receptor mutations likely also play a role in phage resistance in *S. mutans*. This finding is in sharp contrast with *S. thermophilus* where surface receptor mutations are less frequent (Hynes et al., 2017; Labrie et al., 2019). The receptor for phage M102AD is currently not known and no universal mutation responsible for adsorption resistance could be found. Of interest, in *Streptococcus thermophilus* it was already shown that multiple point mutations in the genome of various BIMs can result in phage resistance. One of these mutations was a specific point mutation in the methionine aminopeptidase gene, which also resulted in phage resistance in *S. mutans* P42S (Labrie et al., 2019). Taken altogether, a variety of mutations is likely responsible for the phenotype.

The interference activity of the CRISPR-Cas system of *S. mutans* P42S was also confirmed using several plasmids constructs that harboured proto-spacers targeted by the native spacers. The data obtained from the spacer acquisition and interference assays led to the identification of a PAM at the 3-end of the proto-spacer. First, we showed that 93% of the proto-spacers acquired from the genome of phage M102AD were flanked by 5'-NAA-3', while 79% were flanked by 5'-NAAA-3'. Then, plasmid interference assays revealed that the type II-A CRISPR-Cas system of *S. mutans* P42S interferes only when the targeted proto-spacer sequences are flanked at the 3'-end by NAA or NGAA. While the PAM sequence of NAA was identified in both spacer acquisition and interference activities, the NGAA sequence was only observed in one spacer acquisition event (spacer 145 in BIM93). Interestingly, the dinucleotide AA is found 6828 times in the AT-rich genome of phage M102AD, while GAA is found 1150 times. While there is a significant number of proto-spacers in the 30,664 bp genome of this phage, the PAM difference associated with the frequency of spacer acquisition is intriguing.

It has been previously suggested for type I systems that PAM recognition during the interference stage occurs through a different mechanism than of that during the acquisition stage, since the PAM requirements during the interference stage are less stringent (Shah et al., 2013; Swarts et al., 2012). However, in these systems, the expression of Cas1 and Cas2 is sufficient for spacer acquisition (Datsenko et al., 2012; Díez-Villaseñor et al., 2013; Nuñez et al., 2014; Yosef et al., 2012). In type II-A systems, Cas1 is not able to recognise PAM sequences to guide adaptation while Cas1, Cas2, Csn2 and Cas9 have been reported to be all essential for spacer acquisition. The PAM recognition domain of Cas9 was shown to be involved during spacer acquisition but not its nuclease activity (Heler et al., 2015; Wei et al., 2015). Whereas differences in PAM requirements in adaptation and interference may be explained by the different proteins involved in the two stages in type I systems, in type II-A systems PAM recognition occurs through Cas9 in both stages.

The finding of a type II-A CRISPR-Cas system using a distinct PAM may also have biotechnological applications. For example, one of the limitations of the current CRISPR-Cas9 genome editing technology, which mostly uses Cas9 of *Streptococcus pyogenes*, is

the reliance on its NGG PAM, which limits the sequences that can be targeted (Gleditzsch et al., 2019). Cas9 of *S. pyogenes* (SpyCas9) has been engineered at the so-called PAM interacting motif to recognise other sequences, such as NAAG (Anders et al., 2016). Overall, amino acid sequence identity is very low between *S. mutans* P42S Cas9 (SmutCas9) and *S. pyogenes* Cas9 (SpyCas9). When focusing on the PAM interacting motif of SpyCas9 and its NAAG-recognising variant, no motif with any significant identity could be found with SmutCas9. To date, type II-A Cas9 proteins naturally recognising AT-rich PAMs such as NAAAA have been found in *Lactobacillus buchneri* (Briner and Barrangou, 2014) and *Treponema denticola* (Esvelt et al., 2013). The Cas9 protein of *Streptococcus macacae* (SmaCas9) has also been shown to recognise the shorter NAA PAM (Chatterjee et al., 2020). The Cas9 of *S. mutans* P42S is 76.5 % identical to SmaCas9 (75.3% N-terminal, 80.9% C-terminal).

An efficient Cas9 protein that is able to recognise the NAA or NGAA sequence would allow a wider range of targets for genome editing. For example, the human genome is known to be AT-rich (Romiguier et al., 2010) and the Cas9 protein of *S. mutans* P42S may offer additional biotechnological benefits.

Materials and Methods

Strain, phage, and culture conditions

S. mutans strain P42S and the lytic phage M102AD were obtained from the Félix d'Hérelle Reference Center for Bacterial Viruses (www.phage.ulaval.ca). The bacterial strain was grown in Brain Heart Infusion (BHI) medium at 37°C with 5% CO_2. For growth on plates, 1.25% agar was added to BHI medium. Phage M102AD was amplified using an exponentially growing culture of P42S. Phage-infected cultures were incubated at 37°C until lysis. The resulting lysate was then filtered (0.45 μm) and stored at 4˚C until use. Phage titration was performed using the double-layer plaque assay, and the top agar consisted of BHI supplemented with 0.75% agar.

Identification and analysis of the CRISPR-Cas system in *S. mutans* P42S

The *cas* genes and the CRISPR array sequences were obtained from the whole-genome sequence analysis of *S. mutans* P42S. The genomic DNA was first extracted as described elsewhere (Leblond et al., 1996) with the following modifications. Briefly, the proteinase K and SDS steps were separated into a 15-minute proteinase K (0.4 mg/ml) step followed by a two-hour SDS (1%) step. After the potassium acetate step and subsequent centrifugation, the supernatant was treated with RNAse A (2 µg/ml) for one hour at 37°C. Then, the protocol was resumed with the isopropanol step as described (Leblond et al., 1996). The genomic DNA of *S. mutans* P42S was prepared for sequencing using the Nextera XT DNA library preparation kit according to the manufacturer's instructions. The library was sequenced on a MiSeq apparatus using a MiSeq Reagent Kit v2 (Illumina). Sequences were assembled into 18 contigs using Ray Assembler 2.3.0 (Boisvert et al., 2010) and fused using Mauve Assembly Metrics (Darling et al., 2011). CRISPR loci were identified by searching for repeat sequences as listed in the CRISPR database (www.crispr.i2bc.paris-saclay.fr/). The *cas* gene sequences from other *S. mutans* strains were obtained from NCBI and compared to the genome of *S. mutans* P42S. Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) was used to determine the percentage of identity between the DNA sequences and between the translated amino acid sequences.

BIM assay

An overnight culture of *S. mutans* P42S was transferred (1%) to fresh BHI medium and grown to an OD_{600nm} of 0.3-0.5. Mixtures of 100 µL of the *S. mutans* P42S culture and 100 µL of phage M102AD lysate (titre between 10^7 to 10^9 PFUs/mL) were mixed in BHI top agar and poured directly onto solid medium. Plates were incubated at 37 °C for 48-72 hours and surviving cells were analysed for spacer acquisition by amplifying the CRISPR locus as described elsewhere (Barrangou et al., 2007). The primers CR-F (5'-AATGTCGTGACGAAAATTGG-3') and CR-R (5'-GAAGTCATCGGAACGGTCAT-3') were used to amplify the CRISPR locus found in *S. mutans* P42S. PCR products were sequenced with an ABI 3730xl analyser at the Plateforme de Séquençage et de Génotypage

des Génomes at the CHU of Québec city. BIM assays were also performed with UV-damaged phage lysates as described previously (Hynes et al., 2014).

Phage adsorption assay

An overnight culture of *S. mutans* P42S was transferred (1%) to fresh BHI medium and grown until an OD_{600nm} of 0.7. Phage M102AD (10^3 PFUs) was added to 900 µL of this culture and allowed to adsorb for 15 minutes at 37°C. Cultures then were centrifuged for one minute at maximum speed in a microcentrifuge, and the titre of the supernatant was determined to estimate the phage fraction that did not adsorb to the host cells.

Plasmid interference assay

An approach similar to the one described by Serbanescu et al. was performed (Serbanescu et al., 2015), except that we used plasmid pNZ123 (De Vos, 1987). pNZ123 contains 2497 bp, provides chloramphenicol resistance to the host cells and is readily transformable in *S. mutans*. The 24-bp between the XhoI and EcoRI restriction sites (positions 149 to 173) were removed, and the resulting linearised plasmid (2473 bp) was purified from an agarose gel using the QIAquick Gel Purification Kit as described by the manufacturer. Various DNA inserts were ligated between the XhoI and EcoRI sites of the gel-purified plasmid (see below).

A 30-bp proto-spacer sequence, targeted by one of the spacers already present in the CRISPR locus of the wild-type (WT) strain *S. mutans* P42S and flanked by the nucleotide sequence TAAAT (see below) at the ‘3 end, was first cloned between the XhoI and EcoRI sites to generate pNZ123-sp1. The recombinant plasmid was confirmed by sequencing. Then, pNZ123 and pNZ123-sp1 were independently transformed (see below) into *S. mutans* P42S.

Another 30-bp proto-spacer targeted in strain BIM1 was cloned between the XhoI and EcoRI sites of pNZ123. The cloned proto-spacer was flanked by five bp found downstream in the phage genome (positions 24306-24335) to generate pNZ123-sp2. pNZ123 and pNZ123-sp2 were independently transformed into *S. mutans* P42S BIM1.

Finally, the activity of ectopically acquired spacers (see below) was assayed by a similar experiment. The 30-bp proto-spacer (positions 18743-18714) targeted in BIM2 and flanked by the five bp downstream was cloned between the XhoI and EcoRI sites of pNZ123 to generate pNZ123-sp3. Then, pNZ123 and pNZ123-sp3 were independently transformed into *S. mutans* P42S BIM2.

Transformation of *S. mutans*

The plasmid constructs were transformed into *S. mutans* using natural competence (Håvarstein et al., 1996) through the addition of the Competence Stimulating Peptide (CSP). The active form of this peptide has the following sequence: NH_2-SGSLSTFFRLFNRSFTQA-COOH (Hossain and Biswas, 2012; Mashburn-Warren et al., 2010). The first peptide batch was kindly provided by Prof. Céline Lévesque from the University of Toronto. All subsequent batches were ordered from Biomatik (www.biomatik.com). An overnight culture of *S. mutans* P42S was transferred to fresh BHI medium and grown at 37 ˚C until the OD_{600nm} reached 0.1. Aliquots of 500 µL were then collected and 1 µg of plasmid DNA was added to them. Along with the plasmid construct, the CSP was added at a concentration of 1 µM to the growing culture (Dufour et al., 2011). If no CSP was added, the quantity of plasmid DNA was increased to 10 µg. The cultures were incubated at 37 °C and 5% CO_2 for 2.5 hours, spun down, and the cell pellets were re-suspended in 100µL of BHI. Samples were plated onto BHI agar plates supplemented with 10 µg/mL chloramphenicol. Plates were incubated at 37 °C for 72 hours.

Determination of PAM sequence

Based on the CRISPR analysis of various *S. mutans* BIMs obtained after the challenge with the virulent phage M102AD, we identified several newly acquired spacers. The analysis of the sequences flanking the proto-spacers (Deveau et al., 2008) in the genome of phage M102AD led to the identification of a preferred 5-bp PAM motif at the 3' end of the proto-spacer. To determine the importance of each of these five bps, we designed a plasmid-based interference experiment as described above. Between the XhoI and EcoRI sites of pNZ123, one of the spacers already present in the CRISPR locus of the *S. mutans*

P42S was flanked by a 5-bp motif and several derivatives. The pNZ123-sp1 plasmid was used as a control, as it contains the proto-spacer flanked by the most commonly observed PAM (TAAAT). Other versions of the plasmid included one or two mismatches in the motif, as a nucleotide was replaced by one of the three other alternatives. All plasmids were transformed in duplicate into *S. mutans* P42S WT and derivatives and their transformability was compared. The constructs and insert sequences are listed in Table S2 of the Supplementary Data.

Data availability

The complete genome sequence of phage M102AD has been previously deposited in GenBank under accession number DQ386162(Delisle et al., 2012). The sequences of the *cas* genes of *S. mutans* P42S are available in GenBank under accession numbers MT008463 (*cas9*), MT008464 (*cas1*), MT008465 (*cas2*), MT008466 (*csn2*), and MT008467 (*tracrRNA*).

Acknowledgments

We thank Barbara-Ann Conway for editorial assistance. We are also grateful to Céline Lévesque and Delphine Dufour from the University of Toronto for their practical advice regarding CSP-mediated transformation of and genomic DNA extraction from *S. mutans*. We thank Geneviève Rousseau and Alex Hynes for discussion, Denise Tremblay for preparing the genome library for sequencing as well as Pier-Luc Plante and Simon Labrie for genome assembly. This work was funded by the Natural Sciences and Engineering Research Council of Canada (Discovery program). S.M. holds a T1 Canada Research Chair in Bacteriophages.

References

Achigar, R., Magadán, A.H., Tremblay, D.M., Julia Pianzzola, M., Moineau, S., 2017. Phage-host interactions in *Streptococcus thermophilus*: Genome analysis of phages isolated in Uruguay and ectopic spacer acquisition in CRISPR array. Sci. Rep. 7, 43438.

Anders, C., Bargsten, K., Jinek, M., 2016. Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease Cas9. Mol. Cell 61, 895-902.

Anders, C., Niewoehner, O., Duerst, A., Jinek, M., 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513, 569-573.

Bachrach, G., Leizerovici-Zigmond, M., Zlotkin, A., Naor, R., Steinberg, D., 2003. Bacteriophage isolation from human saliva. Lett. Appl. Microbiol. 36, 50-53.

Banas, J.A., 2004. Virulence properties of *Streptococcus mutans*. Front. Biosci. 9, 1267.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709-1712.

Boisvert, S., Laviolette, F., Corbeil, J., 2010. Ray: Simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J. Comput. Biol. 17, 1519-1533.

Breitbart, M., Rohwer, F., 2005. Here a virus, there a virus, everywhere the same virus? Trends Microbiol. 13, 278-284.

Briner, A.E., Barrangou, R., 2014. *Lactobacillus buchneri* genotyping on the basis of clustered regularly interspaced short palindromic repeat (CRISPR) locus diversity. Appl. Environ. Microbiol. 80, 994-1001.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J.,Makarova, K.S., Koonin, E.V., van der Oost J., 2008. Small CRISPR RNAs guide antiviraldefense in prokaryotes. Science 321, 960-964.

Chatterjee P., Lee J., Nip L., Koseki S.R.T., Tysinger E., Sontheimer E.J., Jacobson J.M., Jakimo N.2020. A Cas9 with complete PAM recognition for adenine dinucleotides. Nat. Comm., 11, 2474.

Dalmasso, M., De Haas, E., Neve, H., Strain, R., Cousin, F.J., Stockdale, S.R., Ross, R.P., Hill, C., 2015. Isolation of a novel phage with activity against *Streptococcus mutans* biofilms. PLoS One 10, e0138651.

Darling, A.E., Tritt, A., Eisen, J.A., Facciotti, M.T., 2011. Mauve assembly metrics. Bioinformatics 27, 2756-2757.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., Semenova, E., 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. 3, 945.

De Vos, W.M., 1987. Gene cloning and expression in lactic streptococci. FEMS Microbiol. Lett. 46, 281-295.

Delisle, A.L., Guo, M., Chalmers, N.I., Barcak, G.J., Rousseau, G.M., Moineau, S., 2012. Biology and genome sequence of *Streptococcus mutans* phage M102AD. Appl. Environ. Microbiol. 78, 2264-2271.

Delisle, A.L., Rostkowski, C.A., 1993. Lytic bacteriophages of *Streptococcus mutans*. Curr. Microbiol. 27, 163-167.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., Charpentier, E., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-607.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., Moineau, S., 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J. Bacteriol. 190, 1390-1400.

Díez-Villaseñor, C., Guzmán, N.M., Almendros, C., García-Martínez, J., Mojica, F.J.M., 2013. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. RNA Biol. 10, 792-802.

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., Sorek, R., 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359, eaar4120.

Dufour, D., Cordova, M., Cvitkovitch, D.G., Lévesque, C.M., 2011. Regulation of the competence pathway as a novel role associated with a streptococcal bacteriocin. J. Bacteriol. 193, 6552-6559.

Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S., Church, G.M., 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat. Methods 10, 1116-11121.

Fejerskov, O., 1997. Concepts of dental caries and their consequences for understanding the disease. Community Dent. Oral Epidemiol. 25, 5-12.

Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lécrivain, A.L., Bzdrenga, J., Koonin, E. V., Charpentier, E., 2014. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res. 42, 2577-2590.

Garneau, J.E., Dupuis, M.È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., Moineau, S., 2010. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67-71.

Gasiunas, G., Barrangou, R., Horvath, P., Siksnys, V., 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. 109, E2579-E2586.

Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., Randau, L., 2019. PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. RNA Biol. 16, 504-517.

Håvarstein, L.S., Gaustad, P., Nes, I.F., Morrison, D.A., 1996. Identification of the streptococcal competence-pheromone receptor. Mol. Microbiol. 21, 863-869.

Heler, R., Samai, P., Modell, J.W, Weiner, C., Goldberg, G., Bikard, D., Marraffini, L.A., 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature 519, 199-202.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., Barrangou, R., 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J. Bacteriol. 190, 1401-1412.

Hossain, M.S., Biswas, I., 2012. An extracelluar protease, SepM, generates functional competence-stimulating peptide in *Streptococcus mutans* UA159. J. Bacteriol. 194, 5886-5896.

Hynes, A.P., Rousseau, G.M., Lemay, M.L., Horvath, P., Romero, D.A., Fremaux, C., Moineau, S., 2017. An anti-CRISPR from a virulent streptococcal phage inhibits *Streptococcus pyogenes* Cas9. Nat. Microbiol. 2, 1374-1380.

Hynes, A.P., Villion, M., Moineau, S., 2014. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. Nat. Commun. 5, 4399.

Ishino, Y., Krupovic, M., Forterre, P., 2018. History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. J. Bacteriol. 200, e00580-17.

Jansen, R., Embden, J.D.A. van, Gaastra, W., Schouls, L.M., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. 43, 1565-1575.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821.

Labrie, S.J., Mosterd, C., Loignon, S., Dupuis, M.È., Desjardins, P., Rousseau, G.M., Tremblay, D.M., Romero, D.A., Horvath, P., Fremaux, C., Moineau, S., 2019. A mutation in the methionine aminopeptidase gene provides phage resistance in *Streptococcus thermophilus*. Sci. Rep. 9,13816.

Labrie, S.J., Samson, J.E., Moineau, S., 2010. Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8, 317-327.

Leblond, P., Fischer, G., Francou, F.X., Berger, F., Guérineau, M., Decaris, B., 1996. The unstable region of *Streptomyces ambofaciens* includes 210kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA. Mol. Microbiol. 19, 261-271.

Lemos, J.A., Burne, R.A., 2008. A model of efficiency: Stress tolerance by *Streptococcus mutans*. Microbiology 154, 3247-3255.

Loesche, W.J., 1986. Role of *Streptococcus mutans* in human dental decay. Microbiol. Rev. 50, 353-380.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Scott, D., Shah, S.A., Siksnys, V., Terns, M.P., Venclovas, Č., White, M.F., Yakunin, A.F., Yan, W., Zhang, F., Garrett, R.A., Backofen, R., van der Oost, J., Barrangou, R., Koonin, E. V, 2019. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. 18, 67-83.

Makarova, K.S., Wolf, Y.I., Koonin, E. V., 2018. Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? Crispr J. 1, 325-336.

Mashburn-Warren, L., Morrison, D.A., Federle, M.J., 2010. A novel double-tryptophan

peptide pheromone controls competence in Streptococcus spp. via an Rgg regulator. Mol. Microbiol. 78, 589-606.

McGinn, J., Marraffini, L.A., 2016. CRISPR-Cas systems optimize their immune response by specifying the site of spacer integration. Mol. Cell 64, 616-623.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O., 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 156, 935-949.

Nuñez, J.K., Kranzusch, P.J., Noeske, J., Wright, A. V., Davies, C.W., Doudna, J.A., 2014. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. Nat. Struct. Mol. Biol. 21, 528-534.

Paster, B., Boches, S., Galvin, J., Ericson, R., Lau, C., Levanos, V., Sahasrabudhe, A., Dewhirst, F., 2001. Bacterial diversity in human subgingival plaque. J. Bacteriol. 183, 3770-3783.

Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L., Severinov, K., 2010. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. Mol. Microbiol. 77, 1367-1379.

Pul, Ü., Wurm, R., Arslan, Z., Geißen, R., Hofmann, N., Wagner, R., 2010. Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. Mol. Microbiol. 75, 1495-1512.

Romiguier, J., Ranwez, V., Douzery, E.J.P., Galtier, N., 2010. Contrasting GC-content dynamics across 33 mammalian genomes: Relationship with life-history traits and chromosome sizes. Genome Res. 20, 1001-1009.

Serbanescu, M.A., Cordova, M., Krastel, K., Flick, R., Beloglazova, N., Latos, A., Yakunin, A.F., Senadheera, D.B., Cvitkovitch, D.G., 2015. Role of the *Streptococcus mutans* CRISPR-Cas systems in immunity and cell physiology. J. Bacteriol. 197, 749-761.

Shah, S., Erdmann, S., Mojica, F., Garrett, R., 2013. Protospacer recognition motifs. RNA Biol. 10, 891-899.

Swarts, D.C., Mosterd, C., van Passel, M.W.J., Brouns, S.J.J., 2012. CRISPR interference directs strand specific spacer acquisition. PLoS One 7, e35888.

van der Ploeg, J.R., 2009. Analysis of CRISPR in *Streptocccus mutans* suggests frequent

occurrence of acquired immunity against infection by M102-like bacteriophages. Microbiology 155, 1966-1976.

Van Der Ploeg, J.R., 2007. Genome sequence of *Streptococcus mutans* bacteriophage M102. FEMS Microbiol. Lett. 275, 130-138.

Wei, Y., Terns, R.M., Terns, M.P., 2015. Cas9 function and host genome sampling in type II-A CRISPR–Cas adaptation. Genes Dev. 29, 356-361.

Yosef, I., Goren, M.G., Qimron, U., 2012. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Res. 40, 5569-5576.

Chapter 2 – Article 2

CRISPR-Cas and adsorption resistance provide combined phage protection in *Streptococcus mutans*

Cas Mosterd[1,2] **and Sylvain Moineau**[1,2,3]*

[1] Département de biochimie, de microbiologie, et de bio-informatique, Faculté des sciences et de génie, Université Laval, Québec City, QC, G1V 0A6, Canada

[2] Groupe de recherche en écologie buccale, Faculté de médecine dentaire, Université Laval, Québec City, QC, G1V 0A6, Canada

[3] Félix d'Hérelle Reference Center for Bacterial Viruses, Faculté de médecine dentaire, Université Laval, Québec City, QC, G1V 0A6, Canada

* Corresponding author. Tel: +1 418 656 3712; Email: Sylvain.Moineau@bcm.ulaval.ca

Keywords

CRISPR, Cas, Streptococcus, mutans, primed, adaptation, adsorption, resistance

Résumé

Streptococcus mutans P42S est connu pour posséder un système CRISPR-Cas de type II-A qui protège contre les phages et les plasmides. En effet, en exposant la souche *S. mutans* P42S au phage M102AD, l'acquisition et la perte d'espaceurs ont été observés. Plusieurs espaceurs ont souvent été acquis de manière simultanée lors d'une ronde d'exposition, ce qui est dû à l'adaptation avec ou sans *priming*. L'analyse de 293 mutants naturels résitants aux phages a mené à l'identification de 399 espaceurs uniques. Dans 88% des cas, des nouveaux espaceurs sont identiques à une région du génome du phage M102AD. Les autres 12% ont des mésappariements, et ce principalement à l'extrémité 5' de l'espaceur. Malgré ces mésappariements, la plupart de ces espaceurs conservent une identité élevée avec le phage dans la région 3', suggérant le maintien d'une activité d'interférence. L'acquisition simultanée d'espaceurs et l'émergence de souches résistantes à l'adsorption du phage M102AD ont auparavant été observées avec la souche *S. mutans* P42S. Dans cette étude, nous avons observé que des mutants ayant une modification au niveau de l'adsorption des phages émergent surtout lorsque des grandes quantités de phages sont présentes. Il semble également que d'autres mécanismes anti-viraux sont impliqués dans les interactions entre le phage M102Ad et la souche *S. mutans* P42S.

Abstract

Streptococcus mutans P42S possesses a type II-A CRISPR-Cas system that protects against phage infection and plasmid transformation. The analysis of 293 Bacteriophage-Insensitive Mutants (BIMs) obtained upon exposure to phage M102AD revealed the acquisition of 399 unique spacers, including several (30%) ectopic spacer acquisitions and a few cases of native spacer deletions. The acquisition of multiple spacers was also observed and appears to be due to both primed and non-primed adaptation. In 88% of the cases, the acquired spacers are identical to a region of the phage genome. The remaining 12% of the spacers had mismatches with the phage genome, but the mistmatches were primarily at the 5' end of the spacer. Mismatches were far less frequent at the 3'-end of the spacer. Experimental evidence indicated that despite multiple mismatches, spacers displayed interference activity as the seed sequence was likely intact. We also observed the emergence of reduced phage adsorption when a high multiplicity of infection (MOI) was used in the phage infection

assays to generate CRISPR BIMs. The simultaneous acquisition of spacers and the emergence of adsorption resistance was less frequent at low MOI. The high level of phage resistance acquired by some of the BIMs suggests the presence of additional antiviral factors in *S. mutans* P42S.

Abbreviations

BHI	Brain Heart Infusion
BIM	bacteriophage-insensitive mutant
bp	base pair
cas	CRISPR-associated
CRISPR	clustered regularly interspaced short palindromic repeats
kb	kilo base
MOI	multiplicity of infection
nt	nucleotide
OD_{600}	optical density at 600 nanometres
PAM	protospacer adjacent motif
rpm	rotations per minute
TSYE	Tryptic Soy Broth supplemented with yeast extract and K_2HPO_4
WT	wild-type

Introduction

Similar to all life forms, bacteria are under the constant threat of viral infection. To prevent infection by bacterial viruses called bacteriophages or phages, bacteria encode multiple antiviral defence mechanisms, of which CRISPR-Cas is one example (Doron et al., 2018; Labrie et al., 2010). CRISPR is an acronym for clustered regularly interspaced short palindromic repeats and along with CRISPR associated (Cas) proteins, they protect bacteria from invasion by nucleic acids such as phage genomes (Barrangou et al., 2007). These CRISPR-Cas systems are highly diverse as they are currently classified in two classes, six types and several subtypes (Makarova et al., 2019). Overall, they have been detected in approximately 40-45% of all bacterial species (Makarova et al., 2019), including *Streptococcus mutans* (van der Ploeg, 2009). *S. mutans* is an important member of the oral microflora as it is a coloniser of the tooth surface and also the main cause of dental caries. Although not life-threatening, these infections are considered a major financial burden (Loesche, 1986).

CRISPR-Cas systems naturally function through a multistep process. The first step, called the adaptation phase, is by far the least understood. Two genes, called *cas1* and *cas2*, are known to be essential for adaptation and are found in almost all types of CRISPR-Cas systems (Makarova et al., 2019, 2018). In the subtype II-A systems, such as those found in *S. mutans* strains, two additional genes are needed for adaptation, namely *csn2* and *cas9* (Heler et al., 2015). CRISPR loci typically contain spacers that match mobile genetic elements such as phage genomes, indicating past frequent phage-bacteria interactions and novel spacer acquisition or immunity. However, such new spacer acquisition events following phage infection is seldomly observed under laboratory conditions for most phage-host pairs. There are nevertheless some examples where CRISPR-Cas systems do acquire novel spacers in laboratory settings. Spacer acquisition following phage challenges has been observed for type II-A systems (Barrangou et al., 2007; Mosterd and Moineau, 2020; van der Ploeg, 2009), and from plasmids in the type II-C system (He et al., 2018). Under particular conditions, spacer acquisition has also been observed in type I-B (Li et al., 2014), I-E (Datsenko et al., 2012; Swarts et al., 2012) and I-F systems (Richter et al., 2014).

One of the conditions that increases the frequency of spacer acquisition is called priming. This primed adaptation refers to the acquisition of new spacers from a nucleic acid that was previously the subject of a spacer acquisition event in the same cell. Therefore, spacer acquisition can be either naïve (from a phage genome that has never been encountered before) or primed. The latter significantly increases the spacer acquisition events and the acquisition of multiple spacers can increase the overall phage resistance (Barrangou et al., 2007, Deveau et al. 2008). Although primarily observed in type I systems (Datsenko et al., 2012; Li et al., 2014; Richter et al., 2014; Swarts et al., 2012), priming has been recently associated with type II-A systems by bioinfomatic analyses (Nicholson et al., 2019) and as well as experimentally to a certain degree with the *Streptococcus pyogenes* (Nussenzweig et al., 2019) and *Streptococcus thermophilus* (Pyenson and Marraffini, 2020) systems.

The spacer acquisition activity of type II-A CRISPR-Cas systems in *S. mutans* has already been demonstrated from phage genomes (Mosterd and Moineau, 2020; van der Ploeg, 2009) and plasmids (Mosterd and Moineau, 2020; Serbanescu et al., 2015), including the acquisition of multiple spacers in some cases. Phage adsorption was also significantly impaired in many bacteriophage-insensitive mutants (BIMs) of *S. mutans* P42S that naturally emerged after phage infection (Mosterd and Moineau, 2020). Here, we further investigated spacer acquisition in the type II-A CRISPR-Cas systems of *S. mutans*, including if priming is occurring as well as whether this influences the overall antiviral response of the cells.

Results

BIM assays

In this study, we randomly generated 174 BIMs that have acquired spacers using various growth conditions of *S. mutans* P42S and its virulent phage M102AD (see Materials and Methods). When adding these to the 119 BIMs obtained in a previous study (Mosterd and Moineau, 2020), this resulted in a set of 293 BIMs. Altogether, these BIMs acquired 399 unique spacers. Spacer size ranged from 28 to 32 base pairs (bp), with the

majority of spacers being 30 bp in length. Of the 399 spacers, 279 were 30 bp (70%), 105 were 31 bp (26%), 11 were 32 bp (3%), three were 29 bp (1%) and one was 28 bp (<1%) in length. This is visualised in Figure 6.

Figure 6: Distribution of spacer lengths.

Of the 399 acquired spacers, 352 (88%) were identical to genomic regions of phage M102AD and 38 (9%) had one to three mismatches, with 31 of them having only one mismatch. One spacer had 9 mismatches with a region of the phage genome. Two spacers (1%) perfectly matched the genome of the *S. mutans* P42S. The remaining six spacers (2%) did not match any known sequence. Overall, we could retrieve the origin of 393/399 (99%) of the acquired spacers.

We could identify a PAM sequence flanking the 393 matching protospacers. Among these protospacers, 95% (375/393) were flanked by the 5'-NAA-3' motif while 78% (306/393) were flanked by 5'-NAAA-3'. However, while the most commonly observed PAM was previously identified as 5'-TAAAT-3' (Mosterd and Moineau, 2020), here the most commonly observed PAM was 5'-TAAAA-3' (17%, 66/393). The motif 5'-TAAAT-3' was associated with 13% of the protospacers (52/393). Consistent with previous observations, when considering only the BIMs that have acquired a single spacer, 100% (68/68) of the protospacers were flanked by 5'-NAA-3'. The PAMs 5'-TAAAA-3' and 5'TAAAT-3' were both observed in 32% (22/68) of the cases. A comparison of the the PAM downstream of the protospacers considering all acquired spacers versus those from single acquisition events can be observed in Figure 7. An overview of the PAMs is found in Table 5.

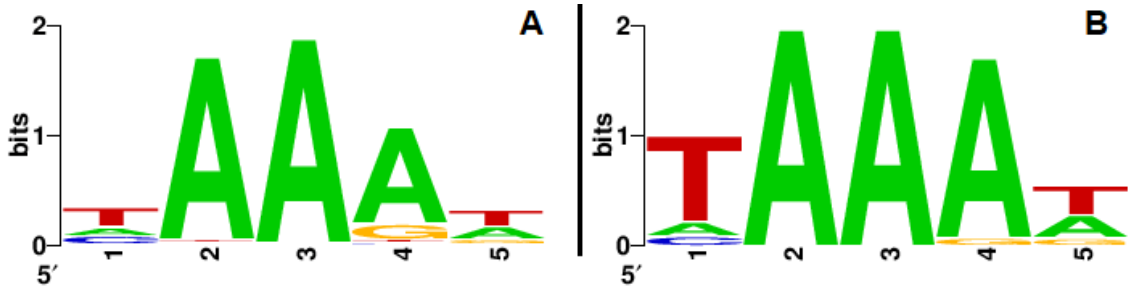


Figure 7: PAM of all protospacers (A) versus PAM during single spacer acquisition events (B).

Table 5: Total and relative frequencies of each acquired PAM. All the PAM sequences observed are listed in the first column. Theire frequencies and relative frequencies can be found in the second and third columns, respectively. The number of times the PAM can be found in the M102AD genome is listed in the fourth column. The fifth column demonstrates the ratio between the frequency of acquisition versus the prevalence on the phage genome of the respective PAM.

PAM (5'-3')	Number of spacers	Ratio number of proto-spacers / unique proto-spacers (393)	PAM prevalence in the phage M102AD genome	Ratio frequency of acquisition / genome prevalence
TAAAA	66	16.8	212	0.31
TAAAT	52	13.2	81	0.65
AAAAT	40	10.2	263	0.15
CAAAT	33	8.4	123	0.27
TAAAG	29	7.4	136	0.21
CAAAA	26	6.6	249	0.10
AAAAA	20	5.1	336	0.06
TAAGT	20	5.1	68	0.29
CAAAG	9	2.3	157	0.06
AAAGT	7	1.8	141	0.05
CAAGT	7	1.8	77	0.09
AAAAG	6	1.5	202	0.03
GAAAA	6	1.5	184	0.03
GAAAT	6	1.5	125	0.05
AAAAC	4	1.0	180	0.02
TAAGG	4	1.0	25	0.16
TTAAA	4	1.0	216	0.02
CAAAC	4	1.0	107	0.04
TAAAC	3	0.8	92	0.03
TAACT	3	0.8	87	0.03
TGAAA	3	0.8	128	0.02
AAATT	2	0.5	183	0.01
AAATC	2	0.5	175	0.01
AAAGA	2	0.5	196	0.01
AAAGG	2	0.5	73	0.03
TAATT	2	0.5	136	0.02
TAACA	2	0.5	88	0.02
TAAGA	2	0.5	78	0.03
TAAGC	2	0.5	103	0.02
CAAGA	2	0.5	112	0.02
CAAGG	2	0.5	45	0.04
GAAAC	2	0.5	86	0.02

AAATA	1	0.3	121	0.01
AAATG	1	0.3	112	0.01
AAACC	1	0.3	64	0.02
AAGTG	1	0.3	64	0.02
AAGCT	1	0.3	176	0.01
ATAAA	1	0.3	111	0.01
TAATA	1	0.3	96	0.01
TAATG	1	0.3	71	0.01
TATAT	1	0.3	101	0.01
TACAG	1	0.3	74	0.01
TCAAA	1	0.3	212	<0.01
TCGCC	1	0.3	33	0.03
TGAAG	1	0.3	86	0.01
CAATT	1	0.3	106	0.01
CAACG	1	0.3	35	0.03
CTAAA	1	0.3	122	0.01
CTATT	1	0.3	68	0.02
CCAAG	1	0.3	44	0.02
GAAGT	1	0.3	70	0.01

```
sp24      1          CGATAAAATTTTAATGCGACAAACAGAAATC    31
                      |||||||||||||||||||||||||||||||
M102AD    23554      GGATAAAATTTTAATGCGACAAACAGAAATC    23584


sp51      1          ACGATGAAGTAAACCTCTTTTGTGAAAGGATT   32
                       ||||||||||||||||||||||||||||||
M102AD    23469      TTGATGAAGTAAACCTCTTTTGTGAAAGGATT   24338


sp157     1          TCCACTAATTTTGTCATCACTAAAATCAAC     30
                     |||||||||||| ||||||||||||||||||
M102AD    19006      TCCACTAATTTCGTCATCACTAAAATCAAC     18977


sp314     1          TGATACAATCAACAAACTAGCTGACAAACC     30
                      ||||||| ||| ||||||||||||||||||
M102AD    21056      CGATACAAACAATAAACTAGCTGACAAACC     21028
```

Figure 8: Non-perfectly matching spacers and their targets on M102AD.

Among those spacers with one to three mismatches, 34 out of the 38 (92%) had a mismatch at the first bp of the spacer. Interestingly, 27/31 spacers that had a 1-bp mismatch, were in fact 31-bp long (and 1/31 was 32-bp long) and contained a stretch of 30 bp perfectly matching a region of the phage genome. Similarly, all of the five spacers with two mismatches were 32 bp long and included a 30-bp stretch identical to the phage genome. Therefore, 384/399 (96%) of the acquired spacers had a perfect match of at least 30 bp to the phage genome. This is visualised in Figure 8.

And interesting observation was that all 27 spacers of 31 bp in length that possessed a mismatch at the first bp, had a C at the first position. In the 4 spacers of 32 bp in length with two mismatches at the first two bp, these first two bp were always AC.

Ectopic spacer acquisition

The CRISPR array of the wild-type phage-sensitive *S. mutans* strain P42S contains 5 native spacers and 6 repeats. Usually, new spacers are added at the 5'-end of the CRISPR array. However, the acquisition of spacers within the array (ectopic) has been observed with the type II-A systems of *S. pyogenes* (McGinn and Marraffini, 2016), *S. thermophilus* (Aguichar et al., 2017; Wei et al., 2015) and *S. mutans* (Mosterd and Moineau, 2020). Here, 30% (89/293) of the BIMs have acquired ectopic spacers. Typically, only one ectopic spacer was acquired in each of the BIMs. Still, multiple ectopic spacers were acquired in 12% (11/89) of them. In the vast majority (91%, 81/89) of the BIMs that have acquired ectopic spacers, at least one additional spacer was also integrated at the 5'-end of the array. All ectopic spacers, except one, were acquired between the native spacers 4 and 5. The other was acquired between spacers 1 and 2.

A total of 78 different spacers were acquired ectopically. Of these, 25 have also been acquired at the 5'-end of the CRISPR locus in other BIMs. Of the remaining 53 spacers acquired ectopically, 52 have an identifiable PAM associated with the corresponding proto-spacer. PAM preferences did not differ significantly compared to the others as 91% (47/52) were flanked by 5'-NAA-3' and 67% (35/52) by 5'-NAAA-3'. Among the BIMs isolated during this study, one BIM acquired a total of 10 spacers during a single infection assay with phage M102AD. Nine new spacers were acquired at the 5'-end of the CRISPR array while one was acquired ectopically. The CRISPR locus of this BIM and the targets of the acquired spacers are highlighted in Figure 9.

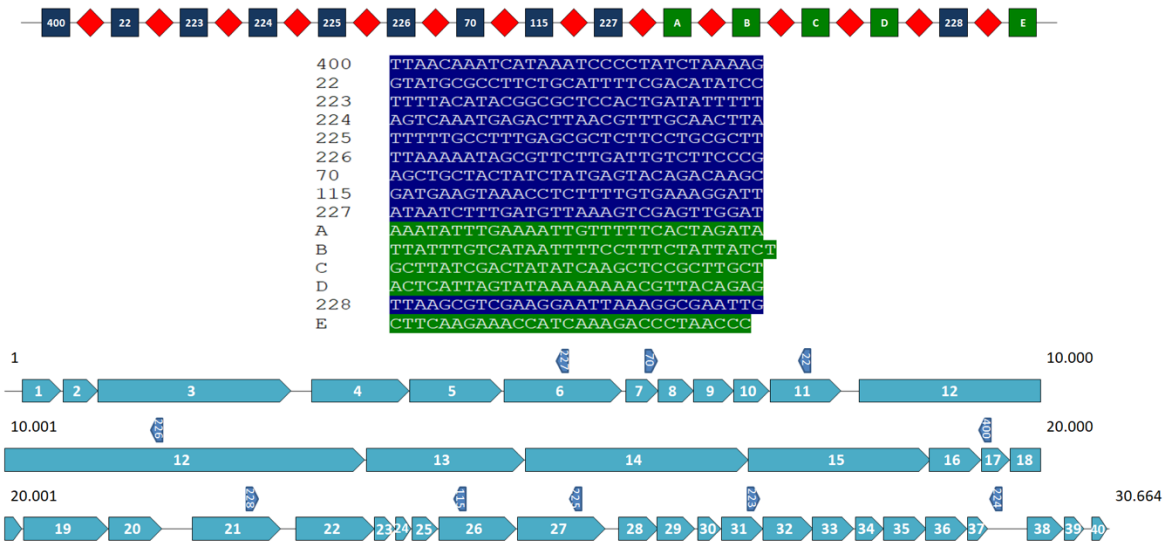


Figure 9: CRISPR locus of BIM 2.2-43CA and the protospacers on the genome of phage M102AD. In the top of the figure, repeats are illustrated as red diamonds and the native spacers in the wild-type CRISPR locus are green squares. Newly acquired spacers are displayed as blue squares. The sequences of the spacers are listed below the CRISPR locus. The colour code corresponds to the colour code used for the CRISPR locus. The genome of phage M102AD is found at the bottom of the figure. The protospacers are highlighted by arrows and are identified by the corresponding spacer numbers. The 10 acquired spacers targeted phage genes coding for hypothetical proteins (ORF17, ORF27, and ORF31), a major tail protein (ORF11), a tape measure protein (ORF12), a DNA packaging protein (ORF7), a single-strand annealing protein (ORF26), a capsid protein (ORF6), a replisome organiser (ORF21) as well as a non-coding region (between *orf37* and *orf38*).

Finally, we also observed the deletion of four of the five native spacers in two BIMs (BIM 4.6-9B and BIM 9.2-8C). Both BIMs acquired one new spacer targeting the phage M102AD genome while maintaining only the WT spacer at the 3' end of the locus.

Acquisition of non-perfecly matching spacers

Of the 293 characterised BIMs, 58 BIMs acquired at least one spacer that did not perfectly match a region of the phage genome. The vast majority of these BIMs (51/58) also acquired at least one other spacer perfectly matching a region of the phage M102AD genome. Among those 51 BIMs, 19 (37%) acquired the non-perfectly matching spacers at the 5' end of the CRISPR locus, while 12 of them acquired multiple spacers at the 5'-end and the non-perfectly matching spacers were located furthest away from the 5' end of the locus. Additionally, 11 BIMs acquired non-perfectly matching spacers ectopically.

Only 7 BIMs acquired a single, non-perfectly matching spacer. Among this fraction, 71% (5/7) acquired a spacer with one, two or three mismatches, while 29% (2/7) acquired a

spacer with low sequence identity compared to M102AD. The interference activity of two non-perfectly matching spacers was examined using a plasmid transformation assay. The mismatches between the spacers (sp157 and sp314) are highlighted in Figure 8. In these assays, plasmid constructs with a proto-spacer derived from phage M102AD were transformed into BIMs with a targeting (non-perfectly matching) spacer. Next, their transformability in these BIMs was compared to the transformability of pNZ123, which does not contain any targeted proto-spacers. pNZ123, pNZsp157 (one mismatch) and pNZsp314 (three mismatches) were all transformable into the wild-type strain *S. mutans* P42S. However, transformation of pNZsp157 into BIM 107, which had previously acquired spacer 157, was negligible. Transformation of pNZsp314 into BIM 18.3-8 was equally close to zero. pNZ123 could be transformed efficiently into both BIMs (Figure 10). Altogether, these data indicate that these non-perfectly matching spacers were still providing interference activity.

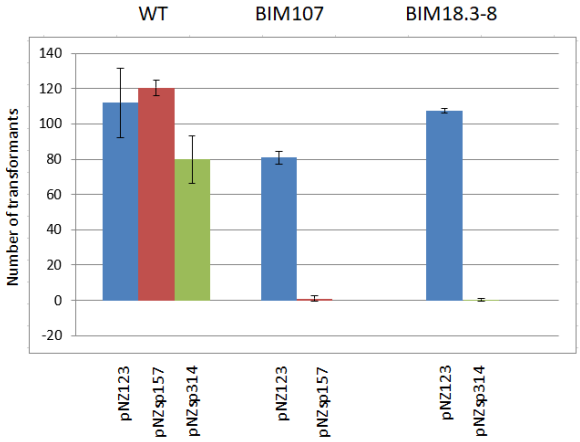


Figure 10: Interference activity of non-perfectly matching spacers. Plasmids containing the targeted protospacers were transformed into the wild-type S. mutans P42S or BIMs containing the non-matching spacer. The number of transformants are per µg of plasmid DNA used in the transformation assay. pNZ123 was used for comparison as this plasmid harbours no proto-spacer.

Priming

Candidate BIMs were selected to examine whether priming could have taken place in the type II-A system of *S. mutans* P42S. Of the 293 BIMs, 193 of them acquired multiple spacers. In the case of ectopic spacers, it could not be determined which spacer was acquired first. As such, we subtracted 81 BIMs that also acquired ectopic spacers as well as one BIM in which the second spacer was not targeting M102AD and could therefore not be

used for the priming analysis. Of those 111 remaining BIMs, 59 BIMs have acquired two spacers targeting M102AD genome and are located at the 5'-end of the CRISPR locusand 52 BIMs that acquired more than two spacers. We assumed that the spacer closest to the native spacers was acquired first.

First, we focused on the 59 BIMs that have acquired two spacers. Specifically, we looked at the distance on the phage M102AD genome between the protospacers targeted by the first and second acquired spacers (Figure 11A). In 36% of the cases (21/59), the second spacer (the one at the 5'-end of the CRISPR array) was acquired from within 2.5 kb from the original spacer. In 12% of the cases (7/59), the second spacer was even acquired from within 500 bp from the first spacer. Of the 59 spacers, 37 (63%) of them were acquired from the same strand as the original spacer (target strand) and 22 (37%) spacers originated from the other strand (non-target strand). Of the 37 spacers acquired from the same strand, 10 (27%) were acquired from within 2.5 kb from the original spacer, while 11 (50%) of the 22 spacers from the non-targeted strand were acquired from within 2.5 kb.

Then, we performed the same analyses on the 52 BIMs that had acquired more than two spacers. In this case, only the second spacer next the last native spacer was considered in the analysis, since this one was presumably acquired after the first new spacer. In addition to the 59 spacers above, these 52 additional spacers gave us a set of 111 spacers to analyse for possible priming (Figure 11B). The percentages remained essentially the same as during the analysis of the 59 spacers. In 32% of the cases (35/111), the spacer was acquired within 2.5 kb from the original spacer and in 10% of the cases (11/111) within 500 bp. Of the 111 spacers, 60 (54%) of them were acquired from the same strand as the original spacer (target strand) and 51 (46%) spacers originated from the other strand (non-target strand). These percentages are closer to each other as compared to the 59 set of spacers. In the case of those spacers acquired from the target strand, 30% of them (18/60) were acquired from within 2.5 kb and 7% (4/60) within 500bp. As for the non-targeted strand, 33% of the spacers (17/51) were acquired from within 2.5 kb and 14% (7/51) within 500 bp.

Finally, when considering all the spacers acquired by these 111 BIMs, 32% of the spacers (67/211) were found within 2.5 kb from the original spacer and 11% (24/211) was found within 500 bp. In 28% of the spacers acquired from the target strand (30/106) they were found within 2.5 kb from the original spacer and 8% (8/106) was found within 500 bp. Of the spacers acquired from the non-target strand, 35% (37/105) was acquired within 2.5 kb from the original spacer and 15% (16/105) was acquired within 500 bp from the original spacer (Figure 11C).

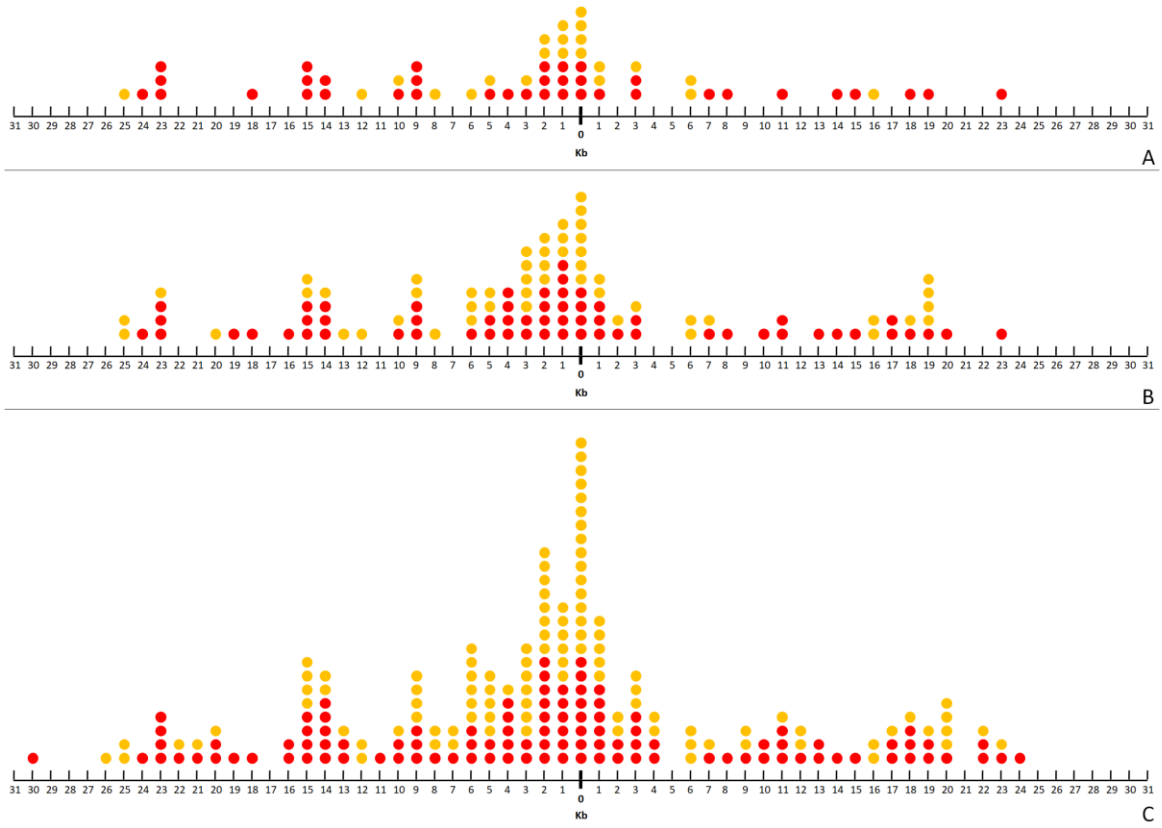


Figure 11: Priming in *Streptococcus mutans* P42S. Spacers acquired from the target strand are illustrated as red dots and those acquired from the non-target strand as yellow dots. The distance of the spacers to the original spacer is illustrated in the figure by their position on the x axis in respect to the value 0. On the right of value 0 the spacers are shown that have been acquired downstream of the original spacer and on the left of value 0 those spacers that have been acquired upstream of the original spacer. The spacers shown at 0 are those spacers that are found <500 bp upstream or downstream from the original spacer. The spacers shown at 1 are those that are found ≥ 500 bp and <1500 downstream of the original spacer. The spacers shown at -1 are those that are found ≥ 500 bp and <1500 upstream of the original spacer. (A) The 59 second spacers (the one at the 5'-end of the CRISPR array) acquired by BIMs with two new spacers at the 5' end of the array. (B) Only the second spacers of all BIMs that acquired multiple spacers (111 spacers). (C) All spacers that were acquired by BIMs that acquired multiple spacers on the 5' end of their loci (211 spacers).

We also noticed that one native spacer of *S. mutans* P42S partially matched a region of the phage M102AD genome (position 25,002-24,973 bp). The spacer has mismatches

with phage M102AD at position 2, 8, 23 and 27. This raised the question of whether this partial identity could play a role in priming. To assay this, the distance of the 111 spacers found closest to the native spacers in the CRISPR locus of the abovementioned BIMs, compared to the M102AD genome fragment that partially matches spacer 3 (25,002-24,973 bp on the phage genome) was determined (Figure 12). This revealed that 55% (61/111) of these spacers were found within 2.5 kb from the M102AD genome fragment partially matching spacer 3.

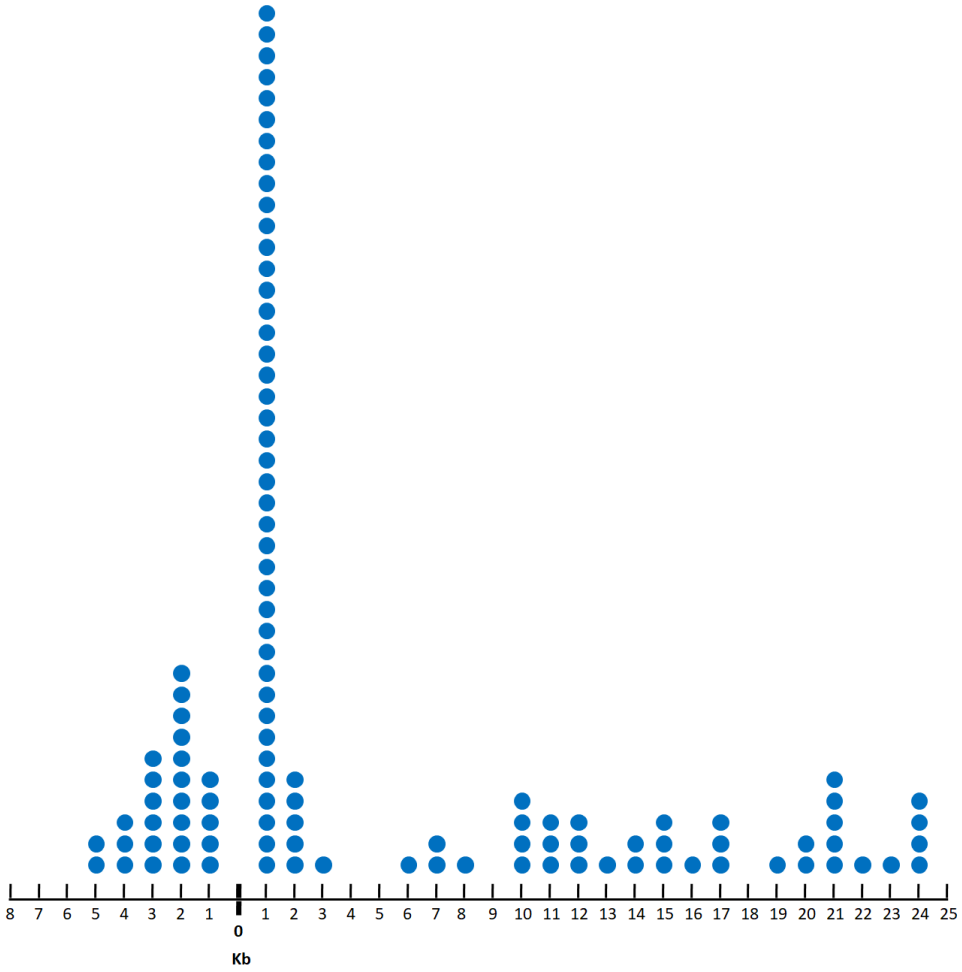


Figure 12: Priming in *S. mutans* P42S in relation to the native spacer 3. The distance of the first acquired spacers to the region of phage M102AD genome that partially matches spacer 3 is illustrated by their position on the x axis in respect to the value 0. On the right of value 0 the spacers are shown that are found downstream of this sequence and on the left of value 0 those spacers that are found upstream of the original spacer**.**

Impact of the multiplicity of phage infection on spacer acquisition

In a previous study of the *S. mutans* P42S CRISPR-Cas system, spacer acquisition was found in 20% of the screened BIMs (Mosterd and Moineau, 2020). Here we investigated in greater details whether the multiplicity of infection (MOI) had an impact on spacer acquisition. In this study, the average rate of spacer acquisition in the tested BIMs at all MOIs was 25% but BIMs acquired more spacers when the MOI was between 1 and 10 (Table 6). Multiple spacer acquisition and ectopic spacer acquisition occurred at all MOIs, with no particular preference (data not shown).

Table 6: Spacer acquisition at different MOIs.

MOI	Number of BIMs screened	Number of BIMs that acquired a spacer	Percentage of the screened BIMs that acquired a spacer	Number of BIMs that acquired an ectopic spacer
< 1	189	24	13	6
1 – 10	266	97	37	27
> 10	225	46	20	14

Phage adsorption assays

We previously observed that spacer acquisition was often accompanied by reduced phage adsorption in *S. mutans* (Mosterd and Moineau, 2020). We determined the phage adsorption rates for 132 BIMs that were obtained at different MOIs. When MOIs >1 were used in phage challenge assays, over 50% of the BIMs that had acquired at least one new spacer also displayed reduced phage adsorption (Table 7). On the other hand, when low MOIs were used, reduced phage adsorption was rarely observed in BIMs that had acquired at least one spacer.

Table 7: Phage adsorption on BIMs obtained at various MOIs.

MOI	Number of BIMs tested	Percentage of BIMs with reduced phage adsorption
< 1	16	12.5
1 – 10	72	64
> 10	45	58

Discussion

The adaptation and interference activities of the type II-A CRISPR-Cas system of *S. mutans* P42S were previously demonstrated (Mosterd and Moineau, 2020). Spacer size and PAM preferences of this CRISPR-Cas system were confirmed in this study. One key observation here was that the PAM 5'-TAAAA-3' (17%) was apparently more often used than 5'-TAAAT-3' (13%). However, it should also be noted that the 5'-TAAAA-3' motif is found 212 times in the genome of phage M102AD, compared to 81 times for 5'-TAAAT-3'. In fact, when considering the relative frequency of acquisition compared to the frequency that this PAM is found in the M102AD genome, TAAAT scores higher (0.65) than TAAAA (0.31). Nonetheless, these two PAMs were most commonly found to flank protospacers, both in absolute frequencies and in relative frequencies (Table 5).

Of the acquired spacers, 88% were identical to a region of the phage M102AD genome, with most of the remaining spacers having one to three mismatches. These mismatches were usually found in the first bp of the spacer, furthest away from the PAM, and they still had a 30-bp stretch that was identical to phage M102AD, likely indicating that they are still providing interference activity (Martel and Moineau, 2014). Non-perfectly matching spacers were found in several BIMs, but in most cases, multiple spacers were acquired in these BIMs during exposure to phage M102AD. The few others that acquired only a non-matching spacer were likely resistant to phages due to other antiviral resistance mechanisms.

An interesting observation was that all 31-bp spacers with a mismatch with phage M102AD at the first bp had a C at this position, and all 32-bp spacers with a mismatch at the first two bp had AC at these positions. A terminal AC-motif is found in the flanking repeat sequences and therefore these additional bases may be the result of an error during the replication of the repeat sequence during spacer integration.

The interference activity of some of the non-perfectly matching spacers were tested using a plasmid transformation assay. Spacer 157 had a mismatch at position 12, leaving an 18-bp stretch of sequence at the 3' end of the spacer that was identical to the targeted dsDNA. Spacer 314 had mismatches at positions 1, 9 and 13, still resulting in a 17-bp

stretch of sequence identity at the 3' end of the spacer. Both plasmids carrying a mismatched protospacer could not be transformed into the BIMs carrying the phage-derived spacer, indicating that the short stretches of sequence identity are sufficient to provide interference. It has been previously shown that within a spacer at the PAM-proximal end, there is a short sequence called the seed sequence, which is essential for spacer target specificity. This phenomenon is observed both in type I (Wiedenheft et al., 2011) and type II systems (Semenova et al., 2011). The length of this seed sequence in type II systems has been defined as being 12 to 13 bp (Cong et al., 2013; Jiang et al., 2013a; Jinek et al., 2012; Martel and Moineau 2014; Sternberg et al., 2014).

Still, 12% of the acquired spacers were not identical to phage M102AD, perhaps suggesting an error rate in the acquisition machinery of the CRISPR-Cas system of *S. mutans* P42S. We cannot exclude that these spacers originated from phage mutants that may have been present at low frequency in the lysate of phage M102AD (Hynes et al. 2014). Two spacers that had mismatches compared to phage M102AD were identical to the closely related phage M102, which was not used in this study and is unable to infect *S. mutans* P42S (Delisle and Rostkowski, 1993). Two spacers that did not match the phage M102AD genome were identical to the genome of *S. mutans* P42S. The sequences on the bacterial genome were not flanked by an appropriate PAM, explaining the viability of these BIMs, but not why they were acquired in the first place.

The acquisition of multiple spacers into the type II-A associated CRISPR array has been observed previously in *S. mutans* P42S duirng infection by phage M102AD (Mosterd and Moineau, 2020). Here, we even isolated a BIM that acquired ten spacers, tripling the size of the native array from 5 to 15 spacers. One explanation for the acquisition of multiple spacers is that when phage exposure occurs on solid media, the ongoing interaction between BIMs and CRISPR-escaping phages likely results in multiple rounds of infection, and thus the acquisition of multiple spacers (Pyenson and Marraffini, 2020). It has been shown by others that the acquisition of multiple spacers leads to increased phage resistance (Barrangou et al., 2007; Deveau et al., 2008).

We also noticed cases of spacer deletions in BIMs of *S. mutans* P42S. Spacer deletion is likely why the CRISPR locus of this strain remains relatively small, even though it can actively acquire multiple spacers. Although deletion of spacers has been previously reported in literature, it is not extensively studied. It is considered unlikely that the incease in the number of spacers has an impact on the fitness of the majority of the BIMs, as it has been reported that acquisition of up to four novel spacers does not affect the fitness of a strain (Vale et al., 2015), eliminating it as an explanation for the deletion events. *S. mutans* is a naturally competent species and the CRISPR-Cas system could interfere with the uptake of beneficial foreign DNA. It has been proposed that the deletion of spacers is a mechanism by which bacteria circumvent this evolutionary downside of CRISPR-Cas, without deleting the entire system (Jiang et al., 2013b; Sanozky-Dawes et al., 2015; Stout et al., 2018). It has been reported in *S. thermophilus* that most deletion events take place at the 3' end of the CRISPR locus, possibly because these ancient spacers protect against foreign nucleic acid sources that no longer provide a threat (Horvath et al., 2008). However, in *Lactobacillus gasseri*, deletion events appear to be more common at the 5' end of the CRISPR locus (Stout et al., 2018). In the case of the two S. mutans P42S BIMs, the four spacers on the 5' end of the CRISPR locus were deleted, leaving only the last spacer on the 3' end. Deletion events within the CRISPR locus have been proposed to be the result of recombination between repeat sequences (Tyson and Banfield, 2008; Jiang et al., 2013b). The last repeat of the CRISPR locus carries a mutation at the final bp, perhaps reducing the possibility of recombination between the final repeat and the previous repeat within the CRISPR locus.

Even though priming has been primarly associated with type I CRISPR-Cas systems, it has also been recently observed in some type II-A systems (Nicholson et al., 2019; Nussenzweig et al., 2019), although it appears to be less pronounced than in type I systems. We also investigated the priming phenomenon in the type II-A system of *S. mutans* P42S. Almost a third of the acquired spacers were found within 2.5-kb of the first acquired (proto)spacer, and over 10% were within 500 bp. Considering that the genome of phage M102AD consists of 30,664 bp, our data suggest a preference of acquiring new spacers within the vicinity of the first acquired spacer. This trend was even more pronounced when considering only the spacers that were acquired from the non-target

strand as the spacers that were acquired within 2.5 kb of the first (proto)spacer in the phage genome increases to 35% when considering all spacers and 50% when considering just the second spacers. In addition, when comparing the distance of this first acquired spacer in relation to a M102AD genomic fragment partially matching native spacer 3 of *S. mutans* P42S, more than half (55%) of the first acquired spacers were found within 2.5 kb from this genomic fragment. Taken altogether, it is likely that priming plays a role in multiple spacer acquisition in *S. mutans* P42S.

The mechanism by which priming occurs in type I systems is relatively well understood. Priming in type I systems requires the activity of Cas3 to generate substrates for subsequent acquisition events (Datsenko et al., 2012; Swarts et al., 2012; Fineran et al., 2014). In the type II-A system of *S. mutans* P42S Cas9 target cleavage can result in the generation of substrates for the type II-A acquisition mechanism (Nussenzweig et al., 2019). Cas9 has been reported to remain bound to its cleavage products (Sternberg et al., 2014), from where it can guide the acquisition complex to a new acquisition target (Nussenzweig et al., 2019).

Originally, primed acquisition in type I systems was thought to only occur from the same strand as the original spacer (Datsenko et al., 2012; Swarts et al., 2012; Fineran et al., 2014). However, priming in type I systems has also been observed without this strand bias (Li et al., 2014; Richter et al, 2014). In type II-A systems, this strand bias was not detected during previous studies (Nicholson et al., 2019; Nussenzweig et al., 2019), which is confirmed in this study. Furthermore, whereas in type I systems priming typically occurs when the original spacer has mismatches with the targeting protospacer or a suboptimal PAM (Fineran et al., 2014; Richter et al., 2014), priming in the type II-A system was determined optimal with fully functional spacers (Nussenzweig et al., 2019). In the case of the system found in *S. mutans* P42S, the spacers identified as the first acquired spacer would typically 100% match a sequence on the phage genome flanked by a functional PAM. However, within the WT CRISPR locus, there is the case of spacer 3, which shows partial identity with the M102AD genome. The targeted region is flanked by the functional PAM TAAAC. Nevertheless, the spacer contains mutations in the seed region, explaining

why the spacer does not target the M102AD genome. Despite the lack of interference activity, the spacer appears to play a role in priming.

In addition to CRISPR-Cas systems, reduced phage adsorption also plays a role in phage resistance of *S. mutans* (Mosterd and Moineau, 2020). We show that when we used a low MOI under laboratory conditions, the CRISPR BIMs did not concomitantly develop an adsorption resistance. However, at higher MOIs, a secondary phage defence mechanism emerged in the BIMs, likely affecting the unknown phage receptors of *S. mutans* P42S. Presumably, this second level of resistance is to cope with the higher number of phages in the infected cultures. Still, all the examined BIMs were fully resistant to phages using plaque assays (data not shown), with or without reduced phage adsorption. This could imply that in these BIMs, CRISPR-Cas resistance was enough to achieve full resistance as observed with other streptococci (Barrangou et al., 2007; Deveau et al., 2008) or that unknown mechanisms are at play.

Materials and Methods

Strain, phage and culture conditions

The bacterial strain *S. mutans* P42S and the lytic siphophage M102AD were obtained from the Félix d'Hérelle Reference Center for Bacterial Viruses (www.phage.ulaval.ca). *S. mutans* strain P42S is a spontaneous streptomycin-resistant mutant of strain P42SM (Delisle et al., 2012; Delisle and Rostkowski, 1993). The conditions for growing the bacterial strain and for phage amplification were as described previously (Mosterd and Moineau, 2020).

BIM assays

An overnight culture of *S. mutans* P42S was transferred (1%) to fresh BHI and grown until an OD_{600nm} of 0.4 was reached. The culture was then mixed in BHI top agar with an appropriate titre of phage M102AD to obtain the desired MOI. The mixture was poured on solid medium and plates were incubated at 37°C for 72 hours. Surviving cells were analysed for spacer acquisition through the amplification of the CRISPR locus. The

CR-F (5'-AATGTCGTGACGAAAATTGG-3') and CR-R (5'-GAAGTCATCGGAACGGTCAT-3') primers were used to amplify the CRISPR locus of *S. mutans* P42S. PCR products were sequenced at the Plateforme de Séquençage et de Génotypage des Génomes at the CHUL centre.

Plasmid interference assays

Plasmid interference assays were performed as described previously (Mosterd and Moineau, 2020). Plasmid constructs were prepared using the vector pNZ123 (De Vos, 1987), which contains a chloramphenicol resistance gene and is transformable in *S. mutans*. The 24-bp fragment between the XhoI and EcoRI restriction sites was removed and the linearised plasmid was purified from an agarose gel using the QIAquick Gel Purification Kit as described by the manufacturer. DNA inserts were ligated between these sites. These inserts consist of sequences from the M102AD genome that have been targeted by various spacers.

The first insert was the target of spacer 157. This 30-bp spacer matches a region of the phage M102AD genome (position 19006 to 18977) in 29 out of 30 bp and had a mismatch at the position 12. The protospacer in the M102AD genome is flanked by the PAM 5'-TAAAG-3'. This 30-bp target and the flanking PAM were cloned between the XhoI and EcoRI restriction sites to generate pNZsp157. A second insert was the target of spacer 314, a 30-bp long spacer with mismatches at positions 1, 9 and 13 compared to the genome of M102AD. The spacer targets a 30-bp stretch on the M102AD genome (21055-21028) that is flanked by 5'-TAAAA-3'. This targeted sequence and PAM were cloned into pNZ123 to generate pNZsp314. The sequences of the inserts of the various constructs can be found in Table 8.

Constructs were transformed into *S. mutans* P42S using natural competence (Dufour et al., 2011). An overnight culture of *S. mutans* P42S was grown in sterile-filtered Tryptic Soy Broth supplemented with 0.5% yeast extract and 0.5% K_2HPO_4 (TSYE) and was transferred to fresh medium and grown at 37˚C until the OD_{600nm} reached 0.1. Then, aliquots of 500 µL were collected and 10 µg of plasmid DNA was added. The cultures were incubated at 37˚C for 4 hours, spun down, and the cell pellets were re-suspended in 100 µL

of TSYE medium. Samples were plated onto TSYE agar plates supplemented with 5 μg/mL chloramphenicol. Plates were incubated at 37 °C for 120 to 168 hours.

Table 8: Fragments cloned into pNZ123 for plasmid interference assays. The PAM sequence is in *italics.* Mismatches between the protospacer and the spacer sequences are underlined.

Construct	Insert/protospacersequence (5'-3')	Targeting spacer sequence (5'-3')
pNZsp157	TCCACTAATTTCGTCATCACTAAAATCAAC*TAAAG*	TCCACTAATTTTGTCATCACTAAAATCAAC
pNZsp314	GATACAAACAATAAACTAGCTGACAAACC*TAAAA*	TGATACAATCAACAAACTAGCTGACAAACC

Phage adsorption assay

An overnight culture of *S. mutans* P42S was transferred (2%) to fresh BHI medium and grown until an OD_{600nm} of 0.7 was reached. Phage M102AD (10^3 PFUs) was added to 900 μL of this culture and allowed to adsorb for 15 minutes at 37°C. Cultures were then centrifuged for one minute at 13,200 rpm and the supernatant was titered to determine the fraction of the phages that did not adsorb to the host cells. A BIM was considered to have a reduced phage adsorption phenotype if the percentage of phage adsorption was below 80%.

Acknowledgments

We thank Amanda Toperoff and Michi Waygood for editorial assistance. This work was funded by the Natural Sciences and Engineering Research Council of Canada (Discovery program). S.M. holds a T1 Canada Research Chair in Bacteriophages.

References

Achigar, R., Magadán, A.H., Tremblay, D.M., Julia Pianzzola, M., and Moineau, S. 2017. Phage-host interactions in *Streptococcus thermophilus*: genome analysis of phages isolated in Uruguay and ectopic spacer acquisition in CRISPR array. Sci. Rep. 7, 43438.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science315, 1709-1712.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., Semenova, E., 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. 3, 945.

De Vos, W.M., 1987. Gene cloning and expression in lactic streptococci. FEMS Microbiol. Lett. 46, 281-295.

Delisle, A.L., Guo, M., Chalmers, N.I., Barcak, G.J., Rousseau, G.M., Moineau, S., 2012. Biology and genome sequence of *Streptococcus mutans* phage M102AD. Appl. Environ. Microbiol. 78, 2264-2271.

Delisle, A.L., Rostkowski, C.A., 1993. Lytic bacteriophages of *Streptococcus mutans*. Curr. Microbiol. 27, 163-167.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., Moineau, S., 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J. Bacteriol. 190, 1390-1400.

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., Sorek, R., 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359, eaar4120.

Dufour, D., Cordova, M., Cvitkovitch, D.G., Lévesque, C.M., 2011. Regulation of the competence pathway as a novel role associated with a streptococcal bacteriocin. J. Bacteriol. 193, 6552-6559.

Fineran, P.C., Gerritzen, M.J.H., Suárez-Diez, M., Kuñne, T., Boekhorst, J., Van Hijum,

S.A.F.T., Staals, R.H.J., Brouns, S.J.J., 2014. Degenerate target sites mediate rapid primed CRISPR adaptation. Proc Natl Acad Sci U S A. 111, E1629-E1638.

He, Y., Wang, M., Liu, M., Huang, L., Liu, C., Zhang, X., Yi, H., Cheng, A., Zhu, D., Yang, Q., Wu, Y., Zhao, X., Chen, S., Jia, R., Zhang, S., Liu, Y., Yu, Y., Zhang, L., 2018. Cas1 and Cas2 from the type II-C CRISPR-Cas system of *Riemerella anatipestifer* are required for spacer acquisition. Front. Cell. Infect. Microbiol. 8, 195.

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., Marraffini, L.A., 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature 519, 199-202.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., Barrangou, R., 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J. Bacteriol. 190, 1401-1412.

Hynes, A.P., Villion, M., Moineau, S., 2014. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. Nat. Commun. 5, 4399.

Jiang, W., Bikard, D., Cox, D., Zhang, F., Marraffini, L.A., 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31, 233-239.

Jiang W, Maniv I, Arain F, Wang Y, Levin BR, Marraffini LA. 2013. Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. PLoS Genet. 9: e1003844.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821.

Labrie, S.J., Mosterd, C., Loignon, S., Dupuis, M.È., Desjardins, P., Rousseau, G.M., Tremblay, D.M., Romero, D.A., Horvath, P., Fremaux, C., Moineau, S., 2019. A mutation in the methionine aminopeptidase gene provides phage resistance in *Streptococcus thermophilus*. Sci. Rep. 9, 13816.

Labrie, S.J., Samson, J.E., Moineau, S., 2010. Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8, 317-327.

Li, M., Wang, R., Zhao, D., Xiang, H., 2014. Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. Nucleic Acids Res. 42, 2483-2492.

Loesche, W.J., 1986. Role of *Streptococcus mutans* in human dental decay. Microbiol. Rev. 50, 353-380.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Scott, D., Shah, S.A., Siksnys, V., Terns, M.P., Venclovas, Č., White, M.F., Yakunin, A.F., Yan, W., Zhang, F., Garrett, R.A., Backofen, R., van der Oost, J., Barrangou, R., Koonin, E. V, 2019. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. 18, 67-83.

Makarova, K.S., Wolf, Y.I., Koonin, E. V., 2018. Classification and nomenclature of CRISPR-Cas systems: where from here? Crispr J. 1, 325-336.

Martel, B., Moineau, S., 2014. CRISPR-Cas: An efficient tool for genome engineering of virulent bacteriophages. Nucleic Acids Res. 42, 9504-9513.

McDonnell, B., Mahony, J., Hanemaaijer, L., Kouwen, T.R.H.M., van Sinderen, D., 2018. Generation of bacteriophage-insensitive mutants of *Streptococcus thermophilus* via an antisense RNA CRISPR-Cas silencing approach. Appl. Environ. Microbiol. 84.e01733-17.

McGinn, J., and Marraffini, L.A. 2016. CRISPR-Cas systems optimize their immune response by specifying the site of spacer integration. Mol. Cell 64, 616-623.

Mosterd, C., Moineau, S., 2020. Characterization of a Type II-A CRISPR-Cas system in *Streptococcus mutans*. mSphere 5, e00235-20.

Nicholson, T.J., Jackson, S.A., Croft, B.I., Staals, R.H.J., Fineran, P.C., Brown, C.M., 2019. Bioinformatic evidence of widespread priming in type I and II CRISPR-Cas systems. RNA Biol. 16, 566-576.

Nussenzweig, P.M., McGinn, J., Marraffini, L.A., 2019. Cas9 cleavage of viral genomes primes the acquisition of new immunological memories. Cell Host Microbe 26, 515-526.e6..

Pyenson, N.C., Marraffini, L.A., 2020. Co-evolution within structured bacterial communities results in multiple expansion of CRISPR loci and enhanced immunity. Elife 9, e53078.

Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N.J., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H.J., Fineran, P.C., 2014. Priming in the type I-F CRISPR-Cas system

triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Res. 42, 8516-8526.

Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., Van Der Oost, J., Brouns, S.J.J., Severinov, K., 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc. Natl. Acad. Sci. U. S. A 108, 10098-10103.

Serbanescu, M.A., Cordova, M., Krastel, K., Flick, R., Beloglazova, N., Latos, A., Yakunin, A.F., Senadheera, D.B., Cvitkovitch, D.G., 2015. Role of the *Streptococcus mutans* CRISPR-Cas systems in immunity and cell physiology. J. Bacteriol. 197, 749-761.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., Doudna, J.A., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62-67.

Swarts, D.C., Mosterd, C., van Passel, M.W.J., Brouns, S.J.J., 2012. CRISPR interference directs strand specific spacer acquisition. PLoS One 7, e35888.

Sanozky-Dawes, R., Selle K., O'Flaherty S., Klaenhammer T., Barrangou R., 2015. Occurrence and activity of a type II CRISPR-Cas system in *Lactobacillus gasseri*. Microbiol. 161, 1752-1761.

Stout E.A., Sanozky-Dawes R., Goh Y.J., Crawley A.B., Klaenhammer T.R., Barrangou R., 2018. Deletion-based escape of CRISPR-Cas9 targeting in *Lactobacillus gasseri*. Microbiol. 164, 1098-1111.

Tyson G.W., Banfield J.F., 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol. 10, 200-207.

Vale, P.F., Lafforgue, G., Gatchitch, F., Gardan, R., Moineau, S., Gandon, S. 2015. Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. Proc. Biol. Sci. 282, 20151270.

van der Ploeg, J.R., 2009. Analysis of CRISPR in *Streptocccus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. Microbiology 155, 1966-1976.

Wei, Y., Terns, R.M., Terns, M.P., 2015b. Cas9 function and host genome sampling in type II-A CRISPR–Cas adaptation. Genes Dev. 29, 356-361.

Wiedenheft, B., Van Duijn, E., Bultema, J., Waghmare, S., Zhou, K., Barendregt, A.,

Westphal, W., Heck, A., Boekema, E., Dickman, M., Doudna, J.A., 2011. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proc. Natl. Acad. Sci. U. S. A 108, 10092-10097.

Chapter 3 – Article 3

Application of Cas9 from *Streptococcus mutans* P42S in viral genome editing

Cas Mosterd[1,2] and Sylvain Moineau[1,2,3]*

[1] Département de biochimie, de microbiologie, et de bio-informatique, Faculté des sciences et de génie, Université Laval, Québec City, QC, G1V 0A6, Canada

[2] Groupe de recherche en écologie buccale, Faculté de médecine dentaire, Université Laval, Québec City, QC, G1V 0A6, Canada

[3] Félix d'Hérelle Reference Center for Bacterial Viruses, Faculté de médecine dentaire, Université Laval, Québec City, QC, G1V 0A6, Canada

* Corresponding author. Tel: +1 418 656 3712; Email: Sylvain.Moineau@bcm.ulaval.ca

Keywords

CRISPR, Cas9, *Streptococcus mutans*, genome, editing, *Lactococcus lactis*, phage, p2

Footnote: The data in this chapter will be eventually merged with the work of another graduate student from a different Faculty at the Université Laval into a larger manuscript. The data from the other research group are not included in this chapter.

Résumé

CRISPR-Cas est un système immunitaire adaptatif retrouvé dans les bactéries et les archées. Ce système offre une protection contre les éléments génétiques mobiles. Des éléments du système naturel ont été utilisés pour construire un puissant outil d'édition du génome, nommé CRISPR-Cas9. Plus précisément, la technologie CRISPR-Cas9 utilise la capacité de la nucléase Cas9 (généralement provenant de *Streptococcus pyogenes*) à couper l'ADN à une position présence via un ARN guide. L'activité de coupure de Cas9 nécessite également la présence d'un court motif nucléotidique (5'-NGG-3 'pour le Cas9 de *S. pyogenes*) appelé PAM et qui flanque la région génomique ciblée. Une limitation de cette technologie est le recours à ce motif PAM strict. Dans cette étude, nous avons évalué le potentiel de Cas9 de *Streptococcus mutans* P42S (SmutCas9) dans l'édition du génome car il utilise des PAM différents (NAA et NGAA). Pour tester son efficacité, nous avons édité le génome du phage virulent p2, qui infecte la bactérie Gram-positive *Lactococcus lactis*dont le génome est à faible GC. D'abord, nous avons déterminé la séquence du tracrRNA du système type II-A CRISPR-Cas chez *S. mutans* P42S. Ensuite, nous avons cloné le gène codant pour SmutCas9, tracrRNA et une unité de répétition-espaceur-répétition, don't l'espaceur cible un gène non essentiel du phage p2, dans le vecteur à faible copie pTRKL2. En parallèle, un vecteur de réparation contenant une version tronquée du gène ciblé de p2 a été préparé. Les deux plasmides ont été transformés dans *L. lactis* MG1363 et le transformant a été infecté avec le phage p2. Des phages mutés dans le gène d'intérêt ont été obtenus, démontrant le potentiel de SmutCas9 pour l'édition du génome, en particulier pour les génomes riches en AT.

Abstract

CRISPR-Cas is an adaptive immune system found in bacteria and archaea and provides resistance against invading mobile genetic elements. Elements of the natural system have use to contruct a powerful genome editing tool, named CRISPR-Cas9. Specifically, the CRISPR-Cas9 technology makes use of the ability of the nuclease Cas9 (usually from *Streptococcus pyogenes*) to cut DNA at a position that is guided by a guide RNA. The cutting activity of Cas9 also requires the presence of a short nucleotide motif (5'-NGG-3' for the Cas9 of *S. pyogenes*) called PAM flanking the targeted genomic region.

A limitation of this technology is the reliance on this strict PAM motif. In this study, we assessed the potential of Cas9 from *Streptococcus mutans* P42S (SmutCas9) to be applied in genome editing as it requires different PAMs (NAA and NGAA). To test its efficacy, we edited the genome of the virulent phage p2, which infects the low-GC Gram-positive bacterium *Lactococcus lactis*. First, we determined the sequence of the tracrRNA of the type II-A CRISPR-Cas system in *S. mutans* P42S. Then, we cloned *SmutCas9*, tracrRNA and a repeat-spacer-repeat unit, with a spacer targeting a non-essentiel gene of phage p2, into the low copy vector pTRKL2. In parallel, a repair template vector harbouring a truncated version of the targeted gene of p2 was prepared. Both plasmids were transformed into *L. lactis* MG1363 and the transformant was infected with phage p2. The resulting escaping phages were mutated in the expected gene, demonstrating the potential of SmutCas9 for genome editing purposes, particulary for AT-rich genomes.

Abbreviations

BHI	Brain Heart Infusion
bp	base pair
$CaCl_2$	calcium chloride
cas	CRISPR-associated
cDNA	complementary DNA
cpf	CRISPR-Cas substype as in *Prevotella* and *Francisella*
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR-RNA
EOP	efficiency of plaquing
GM17	M17 supplemented with glucose
ng	nanogram
ORF	open reading frame
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
SmutCas9	*Streptococcus mutans* Cas9
SpyCas9	*Streptococcus pyogenes* Cas9
tracrRNA	trans-activating crRNA

WT	wild type

Introduction

CRISPR-Cas is an antiviral defence mechanism naturally found in roughly 45% of all bacteria (Makarova et al., 2019). CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) consists of an array of multiple palindromic nucleotide repeats that are separated from one another by variable sequences so-called spacers. These spacers often share homology with viral and plasmid sequences (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). CRISPR arrays are often flanked by *cas* (CRISPR-associated) genes (Jansen et al., 2002; Makarova et al., 2006) which cooperate with the CRISPRs to provide resistance against mobile genetic elements such as phages genomes (Barrangou et al., 2007) by precisely cutting invading nucleic acids (Garneau et al. 2010). The CRISPR arrays are transcribed and processed into short crRNAs that will bind and guide Cas nucleases to their target through base pairing. Thanks to the substantial advances made in the understanding of the biology and diversity of CRISPR-Cas systems over the past decade (Mosterd et al., 2020), a significant segment of CRISPR-Cas research has shifted towards its potential in biotechnology, particularly in genome editing applications. Genome editing by CRISPR-Cas exploits the nuclease activity of Cas effector proteins combined with the specificity that mature (because of tracrRNA activity) CRISPR-RNA (crRNA) provides. CRISPR-Cas systems are divided into Class 1 and Class 2, where Class 1 systems contain multiple Cas proteins that form an effecter complex while Class 2 systems only possess a single Cas effector protein (Makarova et al., 2019). Despite the occasional applications of Class 1 CRISPR-Cas systems in genome editing (Cameron et al., 2019; Fricke et al., 2020; Hidalgo-Cantabrana et al., 2019), Class 2 systems have been more commonly applied due to their less complex gene architecture. The most studied Class 2 systems for their genome editing potential are type II, with Cas9 as the effector protein, and type V, with Cas12a (previously known as Cpf1) as the effector protein. Cas9 and Cas12a can be programmed to cut any target DNA sequence if a crRNA is provided that directs the nucleases towards a target sequence. However, there are limitations in the sequences that can be targeted due to the requirement of so-called PAMs (proto-spacer adjacent motifs) (Anders et al., 2014; Zetsche et al., 2015). The absence of such a PAM within a target renders the site inaccessible to editing. To overcome this limitation, Cas9 and Cas12a variants have been subject to engineering in order to alter or reduce their PAM requirements (Hu et al., 2018; Kleinstiver et al., 2015; Nishimasu et al., 2018). Reducing

the PAM requirements reduces the specificity of the system and therefore increases the chances of off target activity. This is already considered a major concern in genome editing technology (Zhang et al., 2015). Alternatively, one could search for other CRISPR-Cas nucleases that recognise different PAMs, enlarging the scope of possible targets.

One such promising nuclease is Cas9 from *Streptococcus mutans* P42S (SmutCas9), which recognises the PAMs 5'-NAA-3' and 5'-NGAA-3'(Mosterd and Moineau, 2020). The type II-A CRISPR-Cas system from where its derived was previously characterised and plays a role in antiviral defence in *S. mutans* P42S. In addition, SmutCas9 has already been shown to interfere with plasmid transformation (Mosterd and Moineau, 2020). The CRISPR-Cas system of *S. mutans* UA159 has been previously applied in genome editing of its own genome by the means of self targeting CRISPR loci (Gong et al., 2018), however this Cas9 recognises the same PAM as the one from widely used Cas9 from *S. pyogenes*(SpyCas9), namely 5'-NGG-3'.

We previously used SpyCas9 and the CRISPR-Cas9 technology in a heterologous host to knock-out non-essential genes in the genome of the virulent lactococcal phage p2 (Lemay et al., 2017). This phage belongs to the genus of Sk1virus (Deveau et al., 2006; Mahony et al., 2012), and it infects *Lactococcus lactis* MG1363, a model strain for basic research on low GC Gram-postive bacteria (Kok et al., 2017; Wegmann et al., 2007). This model was used in this study to evaluate the potential of SmutCas9 in genome editing.

Results

The CRISPR-Cas9 technology is based on the knowledge of three natural components of the type II-A CRISPR-Cas system: Cas9, crRNA and a tracrRNA. The activity and PAM preferences for SmutCas9 were established previously but information was lacking for crRNA and the tracrRNA (Mosterd and Moineau, 2020).

RNA sequencing

Total RNA was extracted from *S. mutans* P42S and sequenced, which resulted in 9,374,170 identified reads. The native CRISPR array of *S. mutans* P42S contains 6 repeats

and 5 spacers. Of note, the last repeat (3'-end) has one nucleotide difference at its 3'-end (C --> T). Reads corresponding to four of the five expected crRNAs were observed. From the 5' end of the CRISPR locus, crRNAs were identified for spacer 1, 2, 3 and 5. It is unclear why no reads could be observed for spacer 4. The crRNAs of spacer 1 (crRNA1), 2 (crRNA2) and 3 (crRNA3) were 43 nucleotides long, consisting of the last 19 nucleotides of the spacer followed by the first 24 nucleotides of the repeat. The crRNA of spacer 5 (crRNA5) was 39 nucleotides long, consisting of the last 19 nucleotides of the spacer but followed by only the first 20 nucleotides of the repeat.

The RNA sequencing revealed the first 63 nucleotides of the tracrRNA. It was previously determined that the tracrRNA found in *S. mutans* UA159 (Deltcheva et al. 2011, Chylinski et al. 2013) was 107 nt long and the tracrRNA in *S. mutans* P42S was previously estimated to be 93 nt in length (Mosterd and Moineau, 2020). Therefore, we hypothesised that the 63-nucleotide tracrRNA may have been processed or be incomplete. We performed poly-A tailing of RNA, cDNA library construction and sequencing, which led to the identification of an 88- or 89-nt long tracrRNA in *S. mutans* P42S. The sequences of the crRNAs and tracrRNA are presented in Table 9.

Table 9: crRNA and tracrRNA sequences. CRISPR repeats in crRNAs are in *italics*.

RNA ID	**Sequence 5'-3'**
crRNA1	AAUUGUUUUUCACUAGAUA*GUUUUAGAGCUGUGUUGUUUCGAA*
crRNA2	AUUUUCCUUUCUAUUAUCU*GUUUUAGAGCUGUGUUGUUUCGAA*
crRNA3	AUAUCAAGCUCCGCUUGCU*GUUUUAGAGCUGUGUUGUUUCGAA*
crRNA5	CCAUCAAAGACCCUAACCC*GUUUUAGAGCUGUGUUGUUU*
tracrRNA	UUGGAACUAUUCGAAACAACACAGCAAGUUAAAAUAAGGUUUAUCCGUAU UCAACUUGAAAAAGUGCGCACCGAUUCGGUGCUUUUUUA

Construction of pTRKL-SmutCas9

The 11,110-bp plasmid pTRKL2-SmutCas9 was prepared as described in the Materials and Methods section. The presence of the correct sequence of *tracrRNA*, *cas9* and repeat-spacer-unit (R-S-R) and the pTRKL2 backbone were confirmed by Illumina sequencing. The complete sequence of the plasmid can be found in Annex C. After pTRKL2-SmutCas9 was obtained, we generated two constructs in which the non-targeting original spacer present within the R-S-R was removed and exchanged for a spacer that

targets a protospacer (including the appropriate PAM) in *orf49* of phage p2. The R-S-R of the resulting pTRKL2-SmutCas9-49 was sequenced to confirm the construct.

Genome editing of *orf49*

pTRKL2-SmutCas9 (containing a non-targeting spacer) and pTRKL2-SmutCas9-49 (containing a spacer targeting *orf49*) were transformed into *L. lactis* MG1363-pKO49, a strain containing a repair template with a truncated version of *orf49* designed for homologous recombination with the infecting and targeted wild-type phage p2 genome. In addition, the plasmids pL2Cas9 (SpyCas9), pL2Cas9-49 (SpyCas9 + spacer targeting *orf49*) from a previous study (Lemay et al. 2017) and pTRKL2 (empty vector) were transformed into *L. lactis* MG1363-pKO49 as controls. All the strains were confirmed by plasmid analysis. The next step was to infect the five *L. lactis* MG1363-pKO49 derivatives, each transformed with one of the five abovementioned pTRKL2 constructs, with phage p2 using a double-layer plaque assay. The experiment is visualised in Figure 13 and the results are presented in Table 10. Efficiency of plaquing (EOP) values were calculated by dividing the titres obtained for those strains harbouring the p2-specific spacer (SmutCas9-49 and pL2Cas9-49) by those titres of the strains harbouring their non-specific equivalents (SmutCas9 and pL2Cas9). This generated an EOP value of 0.85 for SmutCas9-49/SmutCas9 and 0.05 for pL2Cas9-49/pL2Cas9.

The *orf49* of 100 plaques of phage p2, obtained on each of the five *L. lactis* MG1363-pKO49 strains harbouring pTRKL2 constructs, were analysed by PCR. PCR samples were loaded onto a gel and based on the observed band or bands each sample was identified as WT (band of 1803 bp), deletion mutant (1671 bp) or a mixed population of the WT and deletion mutants (both bands of 1803 bp and 1671 bp). Results can be found in Table 10. With the control containing pL2Cas9-49, all plaques had the truncated version of *orf49*. On the other hand, *L. lactis* MG1363 containing pSmutCas9-49, all plaques contained a mixture of the wild-type *orf49* and the truncated version. These differences are likely due to the level of phage resistance provided by each of the plasmids, with pL2Cas9-49 providing a stronger selective pressure than pTRKL2-SmutCas9-49.

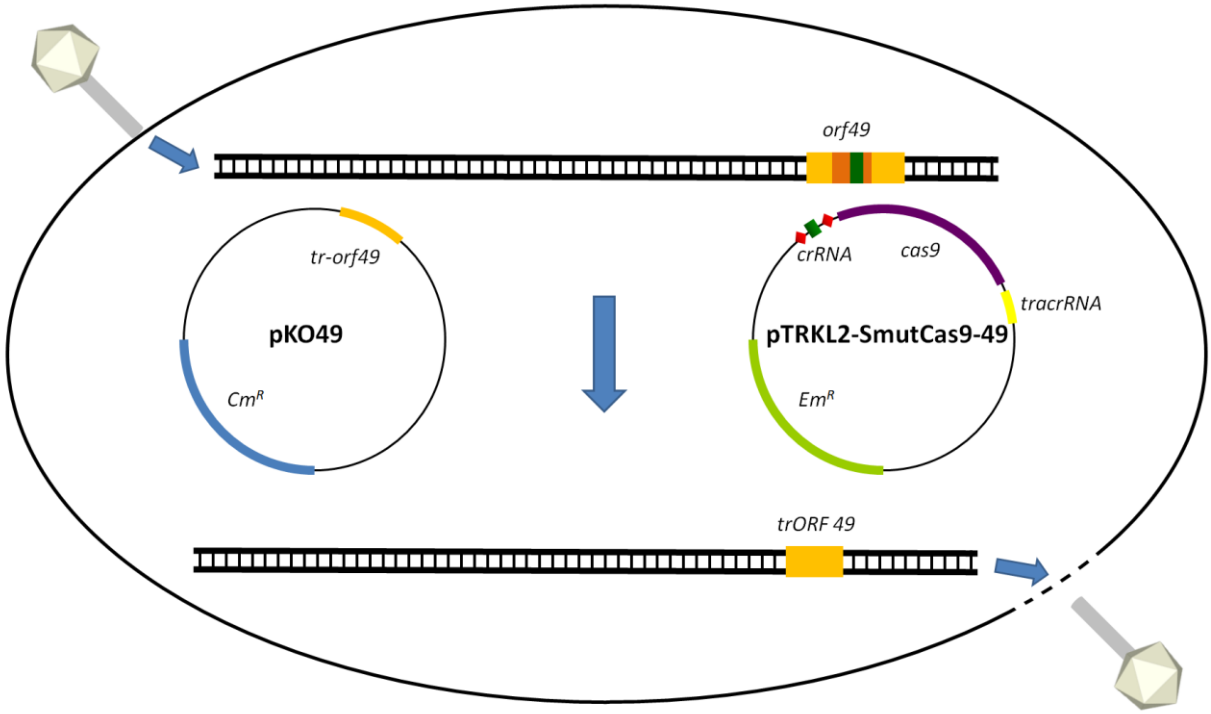


Figure 13: Editing of the *orf49* of the lactococcal phage p2 using SmutCas9. On the top, the injection of linear phage p2 genome into a cell of *L. lactis* MG1363-pKO49-SmutCas9-49 is illustrated. The *orf49* of phage p2 is highlighted in yellow in the genome. The part of the gene to be deleted is in orange and the proto-spacer is green. The plasmid pTRKL2-SmutCas9-49 contains *crRNA* with a spacer (green) that targets the proto-spacer found in *orf49*. SmutCas9 will cut the genome at this position. The plasmid pKO49 contains a truncated version of *orf49* (*tr-orf49*). Due to homologous recombination between pKO49 and the p2 genome, *orf49* will be exchanged *by tr-orf49*. The edited phage will then lyse the cell as this phage gene is non-essential for phage replication in *L. lactis* MG1363 under laboratory conditions.

Table 10: Titres of phage p2 on various *L. lactis* strains and analysis of *orf49* in the resulting phage plaques. In the first column the *L. lactis* MG1363-pKO49 variants are listed identified by the pTRKL2-based construct that was transformed into them. The titre of p2 on each of these strains can be seen in the second column and the EOPs in the third. In the fourth, fifth and sixth column the result of the ORF 49 PCR is shown. The 100 tested plaques have been classified as WT, deletion mutant or mixed population.

Strains	Titre (PFUs/mL)	EOPs	WT	Deletion mutant	Mixed population
TRKL2	$3.0 \pm 0.5 \times 10^8$	N/A	96	0	4
SmutCas9	$3.4 \pm 0.7 \times 10^8$	N/A	94	0	6
SmutCas9-49	$2.9 \pm 0 \times 10^8$	0.85	0	0	100
L2Cas9	$3.9 \pm 0.7 \times 10^8$	N/A	98	0	2
L2Cas9-49	$1.9 \pm 0 \times 10^7$	0.05	0	100	0

In order to obtain pure phage populations, phage plaques were re-suspended in buffer and used to re-infect strain pTRKL2-SmutCas9-49. The resulting lysate was once

more applied in a double-layer plaque assay, which resulted in plaques with either wild-type phage p2 or the phage p2 *orf49* deletion mutant (p2Δ49) being readily obtained.

Of note, a few mixed populations were also obtained in the controls without the selective pressure provided by Cas9. As for pTRKL2-Cas9-49, phage plaques from these mixed populations were analysed and wild-type phage p2 or p2Δ49 were recovered. These data indicate that homologous recombination likely occurred between the incoming phage genome and the repair template even in absence of Cas9 and a targeting spacer.

Discussion

The biological activities of the type II-A CRISPR-Cas system of *S. mutans* P42S were previously investigated (Mosterd and Moineau, 2020). Among others, it was shown that it can interfere with plasmid transformation and that Cas9 recognises different PAMs 5'-NAA-3' and 5'-NGA-3'. Here, we evaluated the potential of SmutCas9 in genome editing.

The genome editing tool CRISPR-Cas9 requires two components, Cas9 and a guide RNA targeting the genomic region to be edited. However, the sequence of the RNA components of the type II-A CRISPR-Cas system of *S. mutans* P42S had not been experimentally confirmed. RNA sequencing revealed the presence of four of the five expected crRNAs in the whole cell RNA extracts of *S. mutans* P42S. Three out of these four crRNAs were 43 nt in length, consisting of the last 19 nucleotides of the spacer and the first 24 nucleotides of the repeat. A fourth crRNA, corresponding to the last spacer of the native array, was 39 nt in length. In previous studies, the size of the processed crRNAs of *S. mutans* UA159 was found to be 42 nucleotide long (Chylinski et al., 2013; Deltcheva et al., 2011).

The tracrRNA of *S. mutans* P42S turned out to be 88 or 89 nt long. This was 4 or 5 nt shorter than previously estimated based on sequence analysis of the putative full unprocessed tracrRNA (Mosterd and Moineau, 2020). The first 4 nt of the putative tracrRNA were missing in the RNA sequencing reads. By the use of Poly(A) Polymerase, a poly-A tail was added to the RNA molecules in order to determine the 3' end of the

tracrRNA. The predicted full unprocessed tracrRNA has an adenine as final nucleotide, and therefore it could not be determined whether the adenine at position 89 was the last nucleotide of the tracrRNA or the first nucleotide of the poly-A tail. The size of the tracrRNA of the strain *S. mutans* UA159 was previously determined to be between 80 nt (processed) and 102 nt (non processed) (Chylinski et al., 2013; Deltcheva et al., 2011). The unprocessed tracrRNA of *S. mutans* P42S is shorter than the tracrRNA of UA159 due to two small deletions (Mosterd and Moineau, 2020).

The tracrRNA of *S. mutans* UA159 has slightly more complementary nt to the repeat regions of its crRNAs than the tracrRNA of P42S. However, there is still a region of 25 nt where there is 24 nt of complementarity between the repeat region of the crRNA and the tracrRNA of P42S, which is essential for the tracrRNA to direct the maturation of crRNA (Deltcheva et al., 2011). The complementarity between the crRNA and tracrRNA can be observed in Figure 14. In addition, based on the 3' ends of the crRNAs, the processing sites of the crRNA could be predicted, also highlighted in Figure 17.

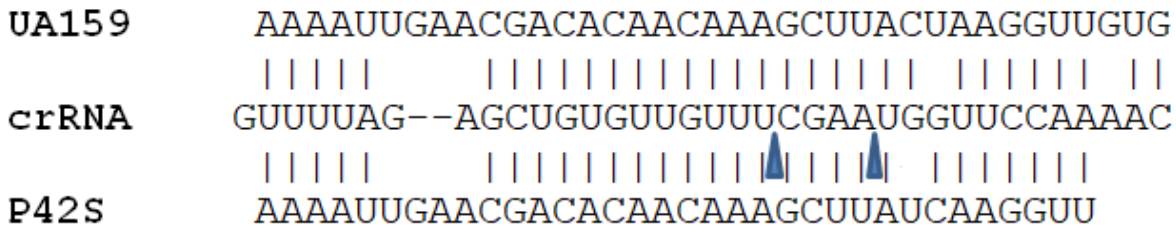


Figure 14: Complementarity between crRNA and tracrRNA of *S. mutans* P42S and UA159. Predicted processing sites of crRNAs are highlighted by arrowheads.

pTRKL2-SmutCas9 was prepared to evaluate the potential of Cas9 from *S. mutans* P42S in viral genome editing. A first indicator was the phage resistance level provided by this plasmid construct as measured by the EOPs. Based on the EOP value no effect on phage resistance could be contributed to pTRKL2-Cas9. In the case of pL2Cas9, the EOP indicated a clear increase in phage resistance due to the CRISPR-Cas system encoded by the plasmid. This suggests that SpyCas9 is more efficient than SmutCas9 in this heterologous host. The next question was whether pTRKL2-SmutCas9 and pL2Cas9 allow the selection of phage mutants.

Plaques of p2 that infected the five different *L. lactis* MG1363-pKO49 strains were analysed by PCR for deletions. Out of the 100 plaques that infected the strain pTRKL2-SmutCas9-49, remarkably all 100 plaques contained a mixed population consisting of both the WT phage p2 and p2Δ49. The phage p2Δ49 could be easily recovered in a subsequent round of phage purification. Interestingly, the phage p2Δ49 mutant was also recovered at low levels from the control pTRKL2-SmutCas9, even though the spacer in this construct did not target the phage. Likely these mutants arose due to homologous recombination between the wild-type genome and the repair tempate. Evolution rates of viruses are higher than those of bacteria (Kupczok et al., 2018; Sanjuán et al., 2010), and therefore the presence of repair template in some cases appears to be sufficient for recombination to occur in phage p2. Recombination hotspots in viral genomes are known to exist, and even without the DNA damage caused by Cas9, these recombination events may occur in such a hotspot (Doan et al., 2001). The seemingly high level of homologous recombination (6%) is intriguing and requires further investigation. Nonetheless, the presence of the spacer does strongly increase the number of phage mutants (100%), demonstrating the activity of the construct.

As expected from a previous study (Lemay et al., 2017), the pL2Cas9-49 construct (with SpyCas) led to the rapid isolation of p2Δ49 as all the 100 plaques analysed by PCR contained a dominant population of p2Δ49. Interestingly as for pTRKL2-SmutCas9, a small percentage (2%) of p2Δ49 were detected in the control pL2Cas9 despite the absence of selective pressure against *orf49*. This was also the case for control pTRKL2, where the percentage was 4%. Again, these events are likely due to recombination with the repair template offered by pKO49.

Taken altogether, both pL2Cas9 (SpyCas9) and pTRKL2-SmutCas9 can be used for viral genome editing. The two tools make use of different Cas9 proteins, which recognise different PAMs. The 5'-NGG-3' PAM recognised by SpyCas9 is found 1655 times in the genome of p2, while the 5'-NAA-3' PAM recognised by SmutCas9 is found as much as 6823 times in the p2 genome of 27,595 bp, which has a AT content of 65.3% explaining the higher number of SmutCas9 target sites. In addition, 5'-NGAA-3' is present 1296 times in the p2 genome. Current genome editing technology heavily relies on SpyCas9 and its

recognised PAM is one of the limitations of this technology (Gleditzsch et al., 2019). Possible targets within AT-rich sequences such as those found in the human genome (Romiguier et al., 2010), could be increased by the use of Cas9 proteins recognising AT-rich PAMs.

Materials and Methods

Bacterial strains, phages and growth conditions

S. mutans P42S as well as *L. lactis*MG1363 and its derivatives were obtained from the Félix d'Hérelle Reference Center for Bacterial Viruses (www.phage.ulaval.ca). *S. mutans* P42S was grown in Brain Heart Infusion (BHI) medium at 37°C with 5% CO_2. For growth on plates, 1.25% agar was added to BHI. *L. lactis* MG1363-pKO49 is a derivative of *L. lactis* MG1363 that harbours the plasmid pKO49, which contains a template to repair the lactococcal phage p2 gene *orf49* (Lemay et al., 2017). The *L. lactis* strains were grown in M17 medium, supplemented with 0.5% glucose (GM17). For maintenance of pKO49, 5 µg/ml of chloramphenicol was added to the medium. For those strains transformed with pTRKL2, pTRKL2-SmutCas9, pL2Cas9 or derivatives, 5 µg/ml of erythromycin was added to the medium.

The lactococcal phage p2 was amplified using an exponentially growing culture of *L. lactis* MG1363. During amplification, the GM17 medium was supplemented with 10 mM $CaCl_2$. Phage-infected cultures were incubated at 30 °C until lysis. The resulting lysate was then filtered (0.45 µm) and stored at 4˚C until use. Phage titration was performed using the double layer plaque assay on solid GM17 medium with 1.25% agar and top agar with 0.75% agar, both of which were supplemented with 10 mM $CaCl_2$.

RNA extraction and sequencing

All equipment and reagents used were RNase-free whenever possible. *S. mutans* P42S was grown until an OD_{600} of 0.8. RNA was extracted from 1 ml aliquots of this culture. The aliquots were spun down for 1 minute at maximum speed in a table centrifuge and supernatant was removed. The pellets were then treated with 50 µL of a solution

containing 20% sucrose and 6% lysozyme for 10 minutes at 37 ºC. After this, the pellet was re-suspended in 1 ml of TRIzol and left at room temperature for 5 minutes. Next, 200 µL of chloroform was added, the sample was mixed, incubated at room temperature for 2 minutes and centrifuged at 12,000 x g and 4 ºC for 15 minutes. The top fraction was transferred to a clean eppendorf vial and mixed with 500 µL of cold iso-propanol. After 10 minutes of incubation at room temperature, the sample was centrifuged at 12,000 x g and 4 ºC for 10 minutes. The pellet was washed with 75% ethanol three times and the dried pellet was re-suspended in water. DNA was removed by using TURBO DNase (Thermo Fisher Scientific) according to the instructions of the manufacturer, with the addition of RNase inhibitor. The RNA sample was finally cleaned using the RNeasy MinElute Cleanup Kit (Qiagen). RNA samples were used to prepare a cDNA library and sequenced by Vertis Biotechnology AG on an Illuma NextSeq 500 system using a 75-bp read length.

To obtain the full sequence of the tracrRNA, RNA samples were treated with *E. coli* Poly(A) Polymerase according to the manufacturer's protocol (New England Biolabs) to add a Poly-A tail to the 3' end of the RNA. A cDNA library was prepared using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) and their protocol, with the modification of replacing the random primers by a Poly-T primer. Then, a PCR was performed using primers tracrRNA_F (5'-AACTATTCGAAACAACACAG-3') and PolyT_R (5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3') and the resulting PCR product was sequenced at the Plateforme de Séquençage et de Génotypage des Génomes at the CHUL Centre (Québec, Canada).

Construction of pTRKL-SmutCas9

All oligos used for cloning can be found in Table S4 of Supplementary Materials. Plasmid pUC57-SmutCas9 was synthesized by BioBasic Inc. This plasmid contains *tracrRNA*, *cas9*, and a single repeat-spacer-repeat unit of *S. mutans* P42S cloned between the EcoRV and SmaI restriction sites on pUC57. Also, the 235-bp region upstream of the CRISPR locus in *S. mutans* P42S was included in the construct to allow expression of the crRNA from its natural promoter. This resulted in a plasmid of 7,478-bp, of which 4,790-bp consists of the insert (SmutCas9) harbouring *tracrRNA*, *cas9* and the repeat-spacer-

repeat unit. The SmutCas9 insert was then cloned into the low copy shuttle vector pTRKL2 (6,478-bp) encoding an erythromycin resistance gene and that is transformable in both *E. coli* and *L. lactis* (O'Sullivan and Klaenhammer, 1993).

Both pTRKL2 and SmutCas9 were PCR amplified with 30 bp overhang (pTRKL2) and 15 bp overhang (SmutCas9) to each other to allow for Gibson Assembly (Gibson et al., 2009). For pTRKL2 the primers used were pTRKL2_F and pTRKL2_R and for SmutCas9 these were SmutCas9_F and SmutCas9_R. PCR samples were loaded onto agarose gels and the bands of appropriate size (6,384 bp for pTRKL2 and 4,814 bp for SmutCas9) were excised from the gel and purified using the QIAquick Gel Extraction kit (Qiagen). The two fragments were then joined together by Gibson Assembly (New England Biolabs) at a 1:1 molar ratio with 200 ng of total DNA used. The assembly mixture was transformed into competent *E. coli* NEB-5α (New England Biolabs), which were plated on solid BHI containing 150 µg/ml of erythromycin. Transformants were then grown in liquid BHI supplemented with 150 µg/ml erythromycin and pTRKL2-SmutCas9 was extracted from overnight cultures using the QIAprep Spin Miniprep Kit (Qiagen). pTRKL2-SmutCas9 has a size of 11,110-bp.

To exchange the existing spacer of pTRKL2-SmutCas9 for one targeting *orf49* of phage p2, the presence of two BsaI restriction sites within the existing spacer sequence was exploited. pTRKL2-SmutCas9 was cut by BsaI and dephosphorylated at the 5'-end of the DNA with Antarctic Phosphatase (New England Biolabs). Next, the resulting DNA was precipitated by mixing with a 1/10 volume of 3M sodium acetate (pH 5.2) and 2 volumes of 100% ethanol and incubated on ice for 15 minutes. The sample was centrifuged for 15 minutes at 16,000 x g, the pellet washed with 70% ethanol and dried before being re-suspended in sterile distilled water. In parallel, a spacer targeting *orf49* was prepared (spacer49_F and spacer49_R for *orf49*). Each primer (100 pmole) was phosphorylated by the T4 Polynucleotide Kinase (ThermoFisher Scientific). Annealing took place by incubating the samples at 95 ºC for 5 minutes, after which the temperature was lowered at the rate of −0.1 °C/s until 25 °C was reached and incubated for 10 min at 25 °C afterwards (Martel and Moineau, 2014). The cut plasmid and spacer were then ligated overnight at 16 ºC and at a final DNA quantity of 1 µg the ligation product was transformed into competent

E. coli NEB-5α. The resulting pTRKL2-SmutCas9-49 was extracted from overnight cultures using the QIAprep Spin Miniprep Kit. This procedure is explained in Figure 15.

As controls, pL2Cas9 and pL2Cas9-49 (Lemay et al., 2017) were also used. pL2Cas9 is a pTRKL2-based construct harbouring the *tracrRNA*, *cas9* and a repeat-spacer-repeat unit from *Streptococcus pyogenes*. pL2Cas9-49 is the derivative that carries a spacer targeting *orf49*.

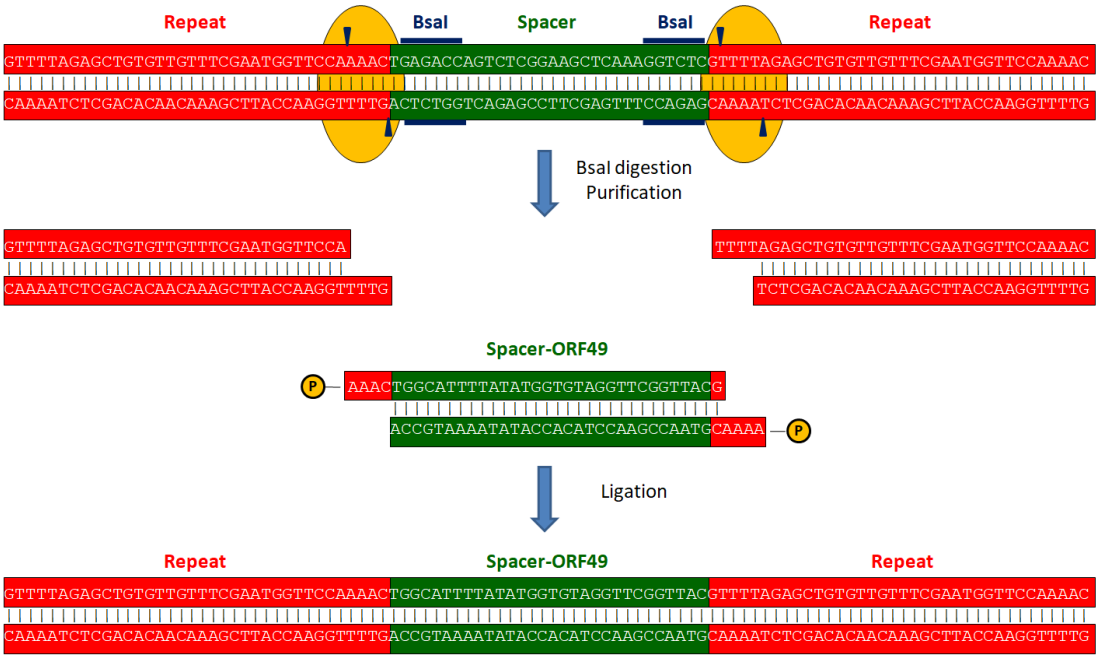


Figure 15: Exchanging the spacer of pTRKL2-SmutCas9. On the top of the figure, the repeat-spacer-repeat unit is shown as found in pTRKL2-SmutCas9. To exchange the spacer for one that targets *orf49*, the first step is the removal of the existing spacer. In blue, the 6-bp recognition sites of BsaI are indicated. The blue arrowheads indicate the exact positions where BsaI (yellow) is cut. After purification, a linearised plasmid results that lacks the spacer. Spacer-ORF49 was annealed, phosphorylated and ligated into linearised pTRKL2-SmutCas9 to obtain pTRKL2-SmutCas9-49.

Genome editing of phage p2

Abovementioned plasmids (pTRKL2, pTRKL2-SmutCas9, pTRKL2-SmutCas9-49, pL2Cas9 and pL2Cas9-49) were transformed into an electro-competent *L. lactis* MG1363-pKO49 strain (Holo and Nes, 1989). The transformants were then infected with phage p2 and plated using the double layer plaque assay. The resulting phage plaques were analysed by PCR amplification of *orf49*. Primers CB13.42 and p2.27 were used to amplify *orf49*. The primer CB13.42 anneals within *orf46* of phage p2 while the primer p2.27 anneals in the non-coding region downstream of *orf49*. Both primers anneal within the p2 genome, but

not on the repair template pKO49. The resulting PCR product sizes would be of 1,803-bp for the wild-type *orf49* and 1,671-bp for the deleted version. An overview of the genome editing protocol is presented in Figure 13, in the results section. Efficiency of plaquing (EOP) values were calculated by dividing the titres obtained for those strains harbouring the p2-specific spacer (SmutCas9-49 and pL2Cas9-49) by those titres of the strains harbouring their non-specific equivalents (SmutCas9 and pL2Cas9) using the double-layer plaque assay.

Acknowledgments

We thank Denise Tremblay for the Illumina sequencing. This work was funded by the Natural Sciences and Engineering Research Council of Canada (Discovery program). S.M. holds a T1 Canada Research Chair in Bacteriophages.

References

Anders, C., Niewoehner, O., Duerst, A., Jinek, M., 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513, 569-573.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709-1712.

Bolotin, A., Quinquis, B., Sorokin, A., Dusko Ehrlich, S., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551-2561.

Cameron, P., Coons, M.M., Klompe, S.E., Lied, A.M., Smith, S.C., Vidal, B., Donohoue, P.D., Rotstein, T., Kohrs, B.W., Nyer, D.B., Kennedy, R., Banh, L.M., Williams, C., Toh, M.S., Irby, M.J., Edwards, L.S., Lin, C.H., Owen, A.L.G., Künne, T., van der Oost, J., Brouns, S.J.J., Slorach, E.M., Fuller, C.K., Gradia, S., Kanner, S.B., May, A.P., Sternberg, S.H., 2019. Harnessing type I CRISPR–Cas systems for genome engineering in human cells. Nat. Biotechnol. 37, 1471-1477.

Chylinski, K., Le Rhun, A., Charpentier, E., 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol. 10, 726-737.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., Charpentier, E., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-607.

Deveau, H., Labrie, S.J., Chopin, M.C., Moineau, S., 2006. Biodiversity and classification of lactococcal phages. Appl. Environ. Microbiol. 72, 4338-4346.

Doan, P.L., Belanger, K.G., Kreuzer, K.N., 2001. Two types of recombination hotspots in bacteriophage T4: One requires DNA damage and a replication origin and the other does not. Genetics 157, 1077-1087.

Fricke, T., Smalakyte, D., Lapinski, M., Pateria, A., Weige, C., Pastor, M., Kolano, A., Winata, C., Siksnys, V., Tamulaitis, G., Bochtler, M., 2020. Targeted RNA knockdown by a type III CRISPR-Cas complex in zebrafish. Crispr J. 3, 299-313.

Garneau, J.E., Dupuis, M.È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., Moineau, S., 2010. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67-71.

Geagea, H., Labrie, S.J., Subirade, M., Moineau, S., 2018. The tape measure protein is involved in the heat stability of *Lactococcus lactis* phages. Appl. Environ. Microbiol. 84, e02082-17.

Gibson, D., Young, L., Chuang, R., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343-345.

Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., Randau, L., 2019. PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. RNA Biol. 16, 504-517.

Gong, T., Tang, B., Zhou, X., Zeng, J., Lu, M., Guo, X., Peng, X., Lei, L., Gong, B., Li, Y., 2018. Genome editing in *Streptococcus mutans* through self-targeting CRISPR arrays. Mol. Oral Microbiol. 33, 440-449.

Hidalgo-Cantabrana, C., Goh, Y.J., Pan, M., Sanozky-Dawes, R., Barrangou, R., 2019. Genome editing using the endogenous type I CRISPR-Cas system in *Lactobacillus crispatus*. Proc. Natl. Acad. Sci. U. S. A. 116, 15774-15783.

Holo, H., Nes, I.F., 1989. High-frequency transformation, by electroporation, of *Lactococcus lactis* subsp. *cremoris* grown with glycine in osmotically stabilized media. Appl. Environ. Microbiol. 55, 3119-3123.

Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z., Liu, D.R., 2018. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature 556, 57-63.

Hynes, A.P., Rousseau, G.M., Lemay, M.L., Horvath, P., Romero, D.A., Fremaux, C., Moineau, S., 2017. An anti-CRISPR from a virulent streptococcal phage inhibits *Streptococcus pyogenes* Cas9. Nat. Microbiol. 2, 1374-1380.

Jansen, R., Embden, J.D.A. van, Gaastra, W., Schouls, L.M., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. 43, 1565-1575.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V. V., Nguyen, N.T., Zheng, Z., Gonzales, A.P.W., Li, Z., Peterson, R.T., Yeh, J.R.J., Aryee, M.J., Joung, J.K., 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523, 481-485.

Kok, J., van Gijtenbeek, L.A., de Jong, A., van der Meulen, S.B., Solopova, A., Kuipers,

O.P., 2017. The evolution of gene regulation research in *Lactococcus lactis*. FEMS Microbiol. Rev. 41, S220-243.

Kupczok, A., Neve, H., Huang, K.D., Hoeppner, M.P., Heller, K.J., Franz, C.M.A.P., Dagan, T., 2018. Rates of mutation and recombination in *Siphoviridae* phage genome evolution over three decades. Mol. Biol. Evol. 35, 1147-1159.

Lemay, M.L., Tremblay, D.M., Moineau, S., 2017. Genome engineering of virulent lactococcal phages using CRISPR-Cas9. ACS Synth. Biol. 6, 1351-1358.

Mahony, J., Murphy, J., Van Sinderen, D., 2012. Lactococcal 936-type phages and dairy fermentation problems: from detection to evolution and prevention. Front. Microbiol. 3, 335.

Makarova, K.S., Grishin, N. V., Shabalina, S.A., Wolf, Y.I., Koonin, E. V., 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct 1, 7.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Scott, D., Shah, S.A., Siksnys, V., Terns, M.P., Venclovas, Č., White, M.F., Yakunin, A.F., Yan, W., Zhang, F., Garrett, R.A., Backofen, R., van der Oost, J., Barrangou, R., Koonin, E. V, 2019. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. 18, 67-83.

Martel, B., Moineau, S., 2014. CRISPR-Cas: An efficient tool for genome engineering of virulent bacteriophages. Nucleic Acids Res. 42, 9504-9513.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., Soria, E., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 60, 174-182.

Mosterd, C., Moineau, S., 2020. Characterization of a type II-A CRISPR-Cas system in *Streptococcus mutans*. mSphere 5, e00235-20.

Mosterd, C., Rousseau, G.M., Moineau, S., 2020. A short overview of the CRISPR-Cas adaptation stage. Can. J. Microbiol.doi: 10.1139/cjm-2020-0212.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O.O., Gootenberg, J.S., Mori, H., Oura, S., Holmes, B., Tanaka, M., Seki,

M., Hirano, H., Aburatani, H., Ishitani, R., Ikawa, M., Yachie, N., Zhang, F., Nureki, O., 2018. Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science 361, 1259-1262.

O'Sullivan, D.J., Klaenhammer, T.R., 1993. High- and low-copy-number *Lactococcus* shuttle cloning vectors with features for clone screening. Gene 137, 227-231.

Pourcel, C., Salvignol, G., Vergnaud, G., 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151, 653-663.

Romiguier, J., Ranwez, V., Douzery, E.J.P., Galtier, N., 2010. Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. Genome Res. 20, 1001-1009.

Sanjuán, R., Nebot, M.R., Chirico, N., Mansky, L.M., Belshaw, R., 2010. Viral Mutation Rates. J. Virol. 84, 9733-9748.

Wegmann, U., O'Connell-Motherway, M., Zomer, A., Buist, G., Shearman, C., Canchaya, C., Ventura, M., Goesmann, A., Gasson, M.J., Kuipers, O.P., Van Sinderen, D., Kok, J., 2007. Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. J. Bacteriol. 189, 3256-3270.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J., Regev, A., Koonin, E. V., Zhang, F., 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759-771.

Zhang, X.H., Tee, L.Y., Wang, X.G., Huang, Q.S., Yang, S.H., 2015. Off-target effects in CRISPR/Cas9-mediated genome engineering. Mol. Ther. - Nucleic Acids 4, e264.

Conclusion and perspectives

The first repeat sequences separated by variable sequences were reported in 1987 (Ishino et al., 1987) and it took 20 years before its role as an adaptive immune system, including as antiviral defence mechanism, was experimentally determined (Barrangou et al., 2007). Ever since, the field of CRISPR research has expanded dramatically. New types and subtypes of CRISPR-Cas systems, including novel *cas* genes, have been discovered and even existing ones have been renamed or reclassified (Makarova et al., 2011, 2019). There is no doubt that more CRISPR-Cas system subtypes await discovery, thanks to the equaly expanding microbial genomic efforts. The different functional stages of the CRISPR-Cas systems have been explored and additional molecular details are either known or emerging. However, despite this seemingly ever-increasing interest in CRISPR-Cas biology, the adaptation stage is understudied. The diversity in the spacer content of the CRISPR arrays between closely related strains suggests frequent new spacer acquisition events. Yet, spacer acquisition is rarely observed under laboratory conditions. In this thesis, *S. mutans* P42S was found to possess a CRISPR-Cas system in which spacer acquisition was observed. This led us to characterise the CRISPR-Cas system of *S. mutans* P42S and to determine its role in antiviral defence.

First, a complete type II-A system was found in strain P42S, consisting of a tracrRNA region, four *cas* genes (*cas9, cas1*, *cas2, csn2*) and a CRISPR locus consisting of six repeats and five spacers. The tracrRNA of *S. mutans* P42S was found to be shorter than the tracrRNA found in the reference strain *S. mutans* UA159 (Chylinski et al., 2013; Deltcheva et al., 2011). Still, the tracrRNA of *S. mutans* P42S possesses the 24/25 nt stretch of identity with the crRNA, essential for the RNA hybrid formation during interference. Whereas *cas1*, *cas2* and *csn2* shared high sequence identity with their equivalents in other *S. mutans* strains, there was a high degree of diversity in *cas9*. More specifically, there was a noticeable diversity in the PAM-interacting domain of Cas9. This implies that within the same species, Cas9 proteins with different PAM interacting domains can be found. This was experimentally demonstrated during BIM assays. The PAMs recognised by Cas9 from *S. mutans* P42S turned out to be NAA and NGAA, whereas the PAM recognised by Cas9 from *S. mutans* UA159 was already confirmed to be NGG (Fonfara et al., 2014). This divergence in the PAM interacting domain was also very recently experimentally demonstrated for Cas9 of various *S. thermophilus* strains (Agudelo et al, 2020).

The NAA sequence is present five times more often than the NGAA sequence on the phage M102AD genome. Nevertheless this PAM sequence was used significantly more often (100 times) during new spacer acquisition events (375/393 for NAA and 4/393 for NGAA). In fact, there was a clear preference for the TAAAT and TAAAA PAM, both in absolute spacer acquisition frequency and in relative frequency. In the plasmid interference assays, the CAAAT PAM led to the strongest interference activity followed by GAAAT (Figure 16). To determine whether the NAA PAM is preferred significantly over the NGAA PAM during interference, this experiment should be repeated with other proto-spacer and PAM sequences. Considering that PAM recognition during both the adaptation and interference phase occurs through the PAM-interacting domain of Cas9 in type II-A systems, this difference is interesting. Although, these two steps of the CRISPR-Cas biology employ different molecular mechanisms. It has been noted previously that the prefered PAM could vary between the adaptation and interference stages. The acronym SAM (spacer acquisition motif) was even proposed to denote the preferred PAM during spacer acquisition as well as TIM (target interference motif) for the interference-related PAM (Shah et al., 2014).

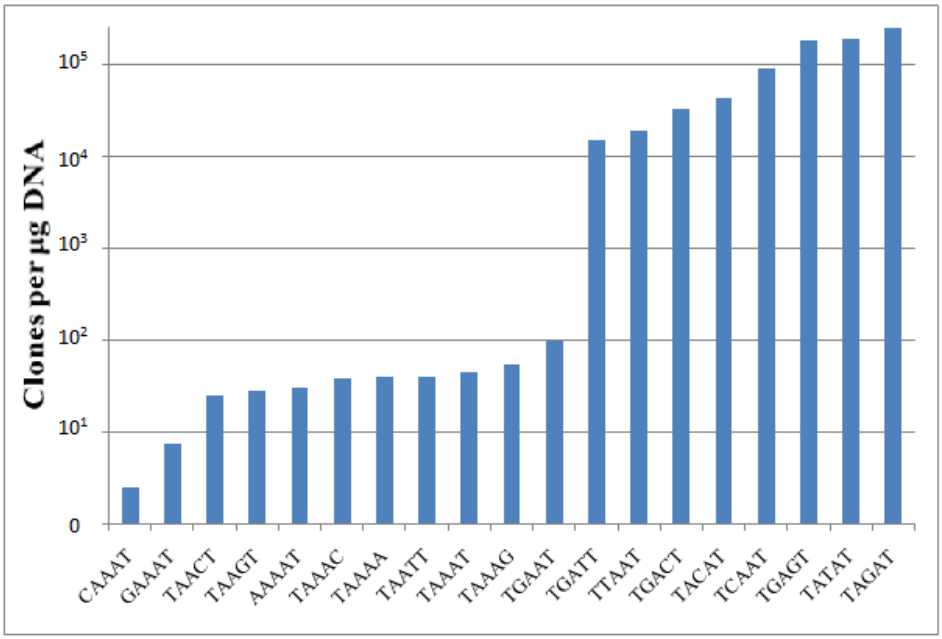


Figure 16: Transformability of plasmids containing protospacers flanked by different PAMs. The PAMs were organised, with the PAM leading to the strongest interference at the left and the weakest PAM at the right. See also data in Table 4.

Another interteresting observation in this thesis was the relatively frequent (12%) acquisition of spacers that harboured mismatches with their targets. These mismatches were

often in the 5'-region of the spacer and it was experimentally demonstrated that when the 17-bp at the 3'-end of the spacer was still identical to their target the spacer kept functionality. These observations agree with the sequence of the crRNAs of *S. mutans* P42S, which we determined to include the last 19 nucleotides of the spacer. It also points out the presence of a seed sequence in this region as observed with other CRISPR-Cas systems. The fact that 12% of the acquired spacers contain mismatches possibly indicates an error-prone spacer acquisition mechanism, unless these spacers were acquired from phage mutants in the lysates. Nevertheless only 3% of all acquired spacers, possessed mismatches within the seed sequence and therefore as much as 97% of the spacers acquired by *S. mutans* P42S likely provide protection against phage M102AD. The activity of spacers with mutations in the seed sequence was not analysed here as done by others (Martel & Moineau, 2014) and could be further investigated in the future. In addition, the level of interference may differ between spacers with full identity towards their targets and spacers with mismatches outside the seed sequence. This could be tested with the plasmid interference assay as described in this study, using pNZ-constructs harbouring targets with or without mismatches to the spacer.

BIMs that acquired spacers containing mismatches often acquired multiple spacers. Yet, spacers with mismatches were not preferentially acquired as the first novel spacer in the CRISPR array. Moreover, BIMs that acquired spacers with full identity also acquired multiple spacers in numerous occasions. This suggests that multiple spacer acquisition is not due to the acquisition of spacers with mismatches. It has been proposed that the acquisition of multiple spacers relies on the strength of the resistance provided by first spacer. Phages can bypass CRISPR interference by a point mutation in the seed sequence or the PAM (Deveau et al. 2008) and perhaps the presence of these phage mutants in the phage population favours the acquisition of a second spacer (Pyenson and Marraffini, 2020). Interestingly, during our study, the 68 BIMs that acquired a single spacer were all flanked on the phage genome by NAA, indicating that this could be the strongest PAM and therefore does not require the acquisition of further spacers.

The BIM that acquired the most spacers extended its CRISPR array from five to 15 spacers. Considering the small initial size of the CRISPR locus in the WT strain, this is a significant increase in the number of spacers. Previously it was already described that multiple spacer acquisition is more commonly observed on solid growth medium than in liquid medium.

It has been proposed that this is due to re-infection occurring on the plate (Pyenson and Marraffini, 2020). Considering that all the BIMs obtained in this study were fully resistant to phages, including those with only one acquired spacer, it is likely that the multiple spacer acquisition events are not due to re-infection of BIMs on solid medium. Another explanation would be that the CRISPR-Cas system of *S. mutans* provides weak phage resistance and other antiviral defence mechanisms are at play in order to complement. It would be of interest to engineer a strain with a phage-targeting spacer by genetic engineering (CRISPR programming) and without a phage challenge (Hynes et al., 2016), and then test the level of phage resistance provided by such a spacer.

One of possible explanations for the multiple spacer acquisition events in *S. mutans* P42S is priming. Evidence for primed spacer acquisition in type II-A systems has been presented recently (Nussenzweig et al., 2019; Pyenson and Marraffini, 2020). Similarly, our data showed a preference to acquire additional spacers from the region within the close vicinity of the first acquired spacer. This effect was strongest when analysing the BIMs that acquired two spacers. In addition, spacers appear to be acquired preferentially close from a genomic region of phage M102AD that partially matches native spacer 3 of *S. mutans* P42S. Spacer acquisition has not been analysed in liquid cultures in this thesis. Analysis of spacer acquisition in liquid culture would allow to determine whereas primed multiple spacer acquisition occurs more commonly on solid media compared to spacer acquisition in liquid media, as described elsewhere (Pyenson and Marraffini, 2020).

Whereas CRISPR-Cas is by far the most dominant phage resistance mechanism in *S. thermophilus*, the situation seems more complex in *S. mutans*. As indicated above, all the CRISPR-BIMs were fully phage resistant, even if they acquired spacers with mismatches or spacers with low identity to phage M102AD genome. Moreover, many of these BIMs displayed reduced phage adsorption. Whereas the rate of spacer acquisition in *S. mutans* P42S was rather constant at different MOIs, the MOI has a major influence on the emergence of adsorption-resistant BIMs. At low MOI, very few CRISPR BIMs also had probable mutation(s) in the unknown phage receptors of *S. mutans* P42S. At higher MOIs, the fraction of CRISPR BIMs that developed additional adsorption resistance increases significantly. Adsorption resistance has been described in the literature as a costly form of antiviral defence. It generally results in a

fitness costs in the absence of virus (Westra et al., 2015). Perhaps this suggests that at low MOIs this mechanism is too costly. It would be of interest to identify those phage receptors, inactivate them (if not-lethal for the bacterial) and test the fitness costs. The fitness of these receptor-mutants could also be compared with the spacer-engineered strains mentioned above.

Still, other antiviral defence mechanisms are likely involved that complement CRISPR-Cas activity in *S. mutans*. The arsenal of antiviral mechanisms used by bacteria to defend against viral predators is very diverse and we are still discovering new mechanisms (Doron et al., 2018). Sequencing the genome of a large number of these *S. mutans* BIMs would likely reveal a variety of mutations that could be investigated for antiviral properties. The annex A of this thesis is a very good example of such an approach. A mutation in the methionine aminopeptidase (*metAp*) gene turned out to generate phage resistance in *S. mutans* P42S (Labrie et al., 2019). We also did sequence the whole genome of one *S. mutans* P42S BIM (BIM5). In addition to the expansion of the CRISPR array, we noticed at least two mutations in the BIM compared to the WT strain (unpublished data). One of the two mutations was in a non-coding region. The second mutation was in the *rgpC* gene (serotype c-specific rhamnose-glucose polysaccharide). RgpC has been linked to adsorption of closely related phage M102 to its host (Shibata et al., 2009). The *rgpC* gene was amplified in several other BIMs but this mutation was found only in BIM5, indicating that other genes are likely involved as well. To confirm the role of the *rgpC*, BIM5 should be complemented with the wild-type gene to see if a phage sensitivity phenotype would be restored to some extent. As the presence of the novel spacers could interfere with the complementation assay, we could also genetically engineer the same mutation in the wild-type strain and then test for phage resistance.

Our second objective was to determine the potential of Cas9 from *S. mutans* P42S in genome editing, specifically to edit the genome of the virulent lactococcal phage p2. The 5'-NAA-3' PAM recognised by SmutCas9 could potentially provide an alternative to the 5'-NGG-3' PAM of SpyCas9. Natural Cas9 proteins recognising AT-rich PAMs have been described in the literature, but their potential in genome editing has not been tested (Briner and Barrangou, 2014) or their efficiency was low (Chatterjee et al., 2020; Esvelt et al., 2013). Phage p2 and its host *L. lactis* MG1363 had previously been edited using SpyCas9 (Lemay et al., 2017). Unlike SpyCas9, the introduction of SmutCas9 into *L. lactis* did not result in a signficant increase of

phage resistance. Yet, SmutCas9 was shown to strongly increase the number of phage mutants deleted in *orf49* as designed. The genome of these phage mutants has yet to be analysed to determine if off-target issues occurred during this genome editing experiments. It will be interesting to see if SmutCas9 can also be used for genome editing experiments in human cell lines.

The CRISPR-Cas system of *S. mutans* P42S was introduced into *L. lactis* through low copy vector pTRKL2. Previously, expression of SpyCas9 from a high copy vector appeared to be toxic to *L. lactis*, whereas expression from a low copy vector was not (Lemay et al., 2017). Unlike SmutCas9, SpyCas9 increases phage resistance in *L. lactis*. This could indicate that the activity of SpyCas9 is higher than the activity of SmutCas9 in this heterologous host. Therefore, this may also account for its toxicity.

Overall, this thesis provides new insights into phage-bacterial interactions in *S. mutans*. Characterisation of the type II-A CRISPR-Cas system of *S. mutans* P42S revealed that the system was capable of acquiring from the phage M102AD genome as well as interfere with targeted plasmid DNA. The precise level of phage resistance directly provided by the CRISPR-Cas system in this strain could not be determined due to the presence of other antiviral systems. However, the discovery of the new Cas9 protein recognising an AT-rich PAM uncovered a new candidate for genome editing purposes.

References

Adriaenssens, E.M., Cowan, D.A., 2014. Using signature genes as tools to assess environmental viral ecology and diversity. Appl Environ Microbiol. 80, 4470–4480.

Ajdić, D., McShan, W.M., McLaughlin, R.E., Savić, G., Chang, J., Carson, M.B., Primeaux, C., Tian, R., Kenton, S., Jia, H., Lin, S., Qian, Y., Li, S., Zhu, H., Najar, F., Lai, H., White, J., Roe, B.A., Ferretti, J.J., 2002. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc. Natl. Acad. Sci. U. S. A. 99, 14434-14439.

Agudelo, D., Carter S., Velimirovic., M., Duringer, A., Levesque, S., Rivest,J.-F., Loehr,J., Mouchiroud, M., Cyr, D., Waters, P.J., Laplante, M., Moineau, S., Goulet, A., Doyon, Y. 2020.Versatile and robust genome editing with *Streptococcus thermophilus* CRISPR1-Cas9. Genome Research. 30, 107-117.

Bachrach, G., Leizerovici-Zigmond, M., Zlotkin, A., Naor, R., Steinberg, D., 2003. Bacteriophage isolation from human saliva. Lett. Appl. Microbiol. 36, 50-53.

Banas, J.A., 2004. Virulence properties of *Streptococcus mutans*. Front. Biosci. 9, 1267.

Banas, J.A., Fountain, T.L., Mazurkiewicz, J.E., Sun, K., Margaret Vickerman, M., 2007. *Streptococcus mutans* glucan-binding protein-A affects *Streptococcus gordonii* biofilm architecture. FEMS Microbiol. Lett. 267, 80-88.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709-1712.

Barrangou, R., Oost, J., 2015. Bacteriophage exclusion, a new defense system. EMBO J. 34, 134-135.

Bender, G.R., Sutton, S.V.W., Marquis, R.E., 1986. Acid tolerance, proton permeabilities, and membrane ATPases of oral streptococci. Infect. Immun. 53, 331-338.

Bernheim, A., Sorek, R., 2020. The pan-immune system of bacteria: antiviral defence as a community resource. Nat. Rev. Microbiol. 18, 113-119.

Biswas, I., Drake, L., Erkina, D., Biswas, S., 2008. Involvement of sensor kinases in the stress tolerance response of *Streptococcus mutans*. J. Bacteriol. 190, 68-77.

Breitbart, M., Rohwer, F., 2005. Here a virus, there a virus, everywhere the same virus? Trends Microbiol. 13, 278-284.

Briner, A.E., Barrangou, R., 2014. *Lactobacillus buchneri* genotyping on the basis of clustered

regularly interspaced short palindromic repeat (CRISPR) locus diversity. Appl. Environ. Microbiol. 80, 994-1001.

Chatterjee P., Lee J., Nip L., Koseki S.R.T., Tysinger E., Sontheimer E.J., Jacobson J.M., Jakimo N.2020. A Cas9 with complete PAM recognition for adenine dinucleotides. Nat. Comm., 11, 2474.

Chen, P.M., Chen, H.C., Ho, C.T., Jung, C.J., Lien, H.T., Chen, J.Y., Chia, J.S., 2008. The two-component system ScnRK of *Streptococcus mutans* affects hydrogen peroxide resistance and murine macrophage killing. Microbes Infect. 10, 293-301.

Chylinski, K., Le Rhun, A., Charpentier, E., 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol. 10, 726-737.

Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G., Sorek, R., 2019. Cyclic GMP–AMP signalling protects bacteria against viral infection. Nature 574, 691-695.

d'Herelle, F., 1917. Sur un microbe invisible antagoniste des bacilles dysentériques. CR Acad.Sci.Paris. 165, 373-375.

Dalmasso, M., De Haas, E., Neve, H., Strain, R., Cousin, F.J., Stockdale, S.R., Ross, R.P., Hill, C., 2015. Isolation of a novel phage with activity against *Streptococcus mutans* biofilms. PLoS One 10, e0138651.

Dashper, S.G., Reynolds, E.C., 1992. pH Regulation by *Streptococcus mutans*. J. Dent. Res. 71, 1159-1165.

Dashper, S.G., Reynolds, E.C., 1996. Lactic acid excretion by *Streptococcus mutans*. Microbiology. 142, 33-39.

De Soet, J.J., Nyvad, B., Kilian, M., 2000. Strain-related acid production by oral streptococci. Caries Res. 34, 486-490.

Del Rio, B., Binetti, A.G., Martín, M.C., Fernández, M., Magadán, A.H., Alvarez, M.A., 2007. Multiplex PCR for the detection and identification of dairy bacteriophages in milk. Food Microbiol. 24, 75-81.

Del Rio, B., Martín, M.C., Martínez, N., Magadán, A.H., Alvarez, M.A., 2008. Multiplex fast real-time PCR for quantitative detection and identification of cos- and pac-type *Streptococcus thermophilus* bacteriophages. Appl. Environ. Microbiol. 74, 4779-4781.

Delisle, A.L., Guo, M., Chalmers, N.I., Barcak, G.J., Rousseau, G.M., Moineau, S., 2012.

Biology and genome sequence of *Streptococcus mutans* phage M102AD. Appl. Environ. Microbiol. 78, 2264-2271.

Delisle, A.L., Rostkowski, C.A., 1993. Lytic bacteriophages of *Streptococcus mutans*. Curr. Microbiol. 27, 163-167.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., Charpentier, E., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-607.

Deng, D.M., Liu, M.J., Ten Cate, J.M., Crielaard, W., 2007. The VicRK system of *Streptococcus mutans* responds to oxidative stress. J. Dent. Res. 86, 606-610.

Domingo-Calap, P., Delgado-Martínez, J., 2018. Bacteriophages: Protagonists of a post-antibiotic era. Antibiotics 7, 66.

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., Sorek, R., 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359, eaar4120.

Dufour, D., Cordova, M., Cvitkovitch, D.G., Lévesque, C.M., 2011. Regulation of the competence pathway as a novel role associated with a streptococcal bacteriocin. J. Bacteriol. 193, 6552-6559.

Dutilh, B.E., 2014. Metagenomic ventures into outer sequence space. Bacteriophage 4, e979664.

Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S., Church, G.M., 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat. Methods 10, 1116-1121.

Fitzgerald, R.J., Adams, B.O., Sandham, H.J., Abhyankar, S., 1989. Cariogenicity of a lactate dehydrogenase-deficient mutant of *Streptococcus mutans* serotype c in gnotobiotic rats. Infect. Immun. 57, 823-826.

Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lécrivain, A.L., Bzdrenga, J., Koonin, E. V., Charpentier, E., 2014. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res. 42, 2577-2590.

Fu, T., Fan, X., Long, Q., Deng, W., Song, J., Huang, E., 2017. Comparative analysis of prophages in *Streptococcus mutans* genomes. PeerJ. 5, e4057.

Fuhrman, J.A., 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399, 541-548.

Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G., Sorek, R., 2015. BREX is a novel phage resistance system widespread in microbial genomes. EMBO J. 34, 169-183.

Grimes, D.J., Atwell, R.W., Brayton, P.R., Palmer, L.M., Rollins, D.M., Roszak, D.B., Singleton, F.L., Tamplin, M.L., Colwell, R.R., 1986. The fate of enteric pathogenic bacteria in estuarine and marine environments. Microbiol. Sci. 3, 324-329.

Hahn, K., Faustoferri, R.C., Quivey, R.G., 1999. Induction of an AP endonuclease activity in *Streptococcus mutans* during growth at low pH. Mol. Microbiol. 31, 1489-1498.

Hamada, S., Ooshima, T., 1975. Inhibitory spectrum of a vacteriocinlike substance (mutacin) produced by some strains of *Streptococcus mutans*. J. Dent. Res. 54, 140-145.

Håvarstein, L.S., Hakenbeck, R., Gaustad, P., 1997. Natural competence in the genus *Streptococcus*: Evidence that streptococci can change pherotype by interspecies recombinational exchanges. J. Bacteriol. 173, 372-381.

Hendrix, R.W., 2003. Bacteriophage genomics. Curr. Opin. Microbiol. 6, 506-511.

Hillman, J.D., Chen, A., Snoep, J.L., 1996. Genetic and physiological analysis of the lethal effect of L-(+)- lactate dehydrogenase deficiency in *Streptococcus mutans*: complementation by alcohol dehydrogenase from *Zymomonas mobilis*. Infect. Immun. 64, 4319-4323.

Hossain, M., Biswas, I., 2012. An extracelluar protease, SepM, generates functional competence-stimulating peptide in *Streptococcus mutans* UA159. J. Bacteriol. 194, 5886-5896.

Hynes, A.P., Labrie, S.J., Moineau, S., 2016. Programming native CRISPR arrays for the generation of targeted immunity. mBio 7, e00202-16.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., Nakata, A., 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J. Bacteriol. 169, 5429-5433.

Johnson, C.P., Gross, S.M., Hillman, J.D., 1980. Cariogenic potential in vitro in man and in vivo in the rat of lactate dehydrogenase mutants of *Streptococcus mutans*. Arch. Oral Biol. 25, 707-713.

Kawabata, S., Hamada, S., 1999. Studying biofilm formation of mutants streptococci. Methods Enzymol. 310, 513-523.

Kim, M.S., Bae, J.W., 2018. Lysogeny is prevalent and widely distributed in the murine gut microbiota. ISME J. 12, 1127-1141.

Kobayashi, H., 1985. A proton-translocating ATPase regulates pH of the bacterial cytoplasm. J. Biol. Chem. 257, 1885-1889.

Kobayashi, H., Suzuki, T., Unemoto, T., 1986. Streptococcal cytoplasmic pH is regulated by changes in amount and activity of a proton-translocating ATPase. J. Biol. Chem. 261, 627-630.

Koonin, E. V., Makarova, K.S., Wolf, Y.I., 2017. Evolutionary genomics of defense systems in Archaea and bacteria. Annu. Rev. Microbiol. 71, 233-261.

Koonin, E. V., Yutin, N., 2020. The crAss-like phage group: how metagenomics reshaped the human virome. Trends Microbiol. 28, 349-359.

Kreth, J., Merritt, J., Shi, W., Qi, F., 2005. Co-ordinated bacteriocin production and competence development: A possible mechanism for taking up DNA from neighbouring species. Mol. Microbiol. 57, 392-404.

Kuhnert, W.L., Zheng, G., Faustoferri, R.C., Quivey, R.G., 2004. The F-ATPase operon promoter of *Streptococcus mutans* is transcriptionally regulated in response to external pH. J. Bacteriol. 186, 8524-8528.

Labrie, S., Moineau, S., 2000. Multiplex PCR for detection and identification of lactococcal bacteriophages. Appl. Environ. Microbiol. 66, 987-994.

Labrie, S.J., Mosterd, C., Loignon, S., Dupuis, M.È., Desjardins, P., Rousseau, G.M., Tremblay, D.M., Romero, D.A., Horvath, P., Fremaux, C., Moineau, S., 2019. A mutation in the methionine aminopeptidase gene provides phage resistance in *Streptococcus thermophilus*. Sci. Rep. 9, 13816.

Lemay, M.L., Tremblay, D.M., Moineau, S., 2017. Genome engineering of virulent lactococcal phages using CRISPR-Cas9. ACS Synth. Biol. 6, 1351-1358.

Lemos, J.A., Burne, R.A., 2008. A model of efficiency: Stress tolerance by *Streptococcus mutans*. Microbiology. 154, 3247-3255.

Lévesque, C., Mair, R.W., Perry, J.A., Lau, P.C.Y., Li, Y.H., Cvitkovitch, D.G., 2007. Systemic inactivation and phenotypic characterization of two-component systems in expression of *Streptococcus mutans* virulence properties. Lett. Appl. Microbiol. 45, 398-404.

Li, Y.H., Lau, P.C.Y., Lee, J.H., Ellen, R.P., Cvitkovitch, D.G., 2001. Natural genetic transformation of *Streptococcus mutans* growing in biofilms. J. Bacteriol. 183, 897-908.

Li, Y.H., Lau, P.C.Y., Tang, N., Svensäter, G., Ellen, R.P., Cvitkovitch, D.G., 2002a. Novel two-

component regulatory system involved in biofilm formation and acid resistance in *Streptococcus mutans*. J. Bacteriol. 184, 6333-6342.

Li, Y.H., Tang, N., Aspiras, M.B., Lau, P.C.Y., Lee, J.H., Ellen, R.P., Cvitkovitch, D.G., 2002b. A quorum-sensing signaling system essential for genetic competence in *Streptococcus mutans* is involved in biofilm formation. J. Bacteriol. 184, 2699-2708.

Loesche, W.J., 1986. Role of *Streptococcus mutans* in human dental decay. Microbiol. Rev. 50, 353-380.

Ly-Chatain, M.H., Durand, L., Rigobello, V., Vera, A., Demarigny, Y., 2011. Direct quantitative detection and identification of lactococcal bacteriophages from milk and whey by real-time PCR: Application for the detection of lactococcal bacteriophages in goat's raw milk whey in France. Int. J. Microbiol. 2011, 594369.

Lynch, D.J., Fountain, T.L., Mazurkiewicz, J.E., Banas, J.A., 2007. Glucan-binding proteins are essential for shaping *Streptococcus mutans* biofilm architecture. FEMS Microbiol. Lett. 268, 158-165.

Makarova, K., Aravind, L., Wolf, Y., Koonin, E., 2011. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol. Direct 6, 38.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Scott, D., Shah, S.A., Siksnys, V., Terns, M.P., Venclovas, Č., White, M.F., Yakunin, A.F., Yan, W., Zhang, F., Garrett, R.A., Backofen, R., van der Oost, J., Barrangou, R., Koonin, E.V., 2020. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. 18, 67-83.

Mashburn-Warren, L., Morrison, D.A., Federle, M.J., 2010. A novel double-tryptophan peptide pheromone controls competence in *Streptococcus* spp. via an Rgg regulator. Mol. Microbiol. 78, 589-606.

Matsumoto-Nakano, M., 2018. Role of *Streptococcus mutans* surface proteins for biofilm formation. Jpn. Dent. Sci. Rev. 54, 22-29.

Mohamedhussein, N.M., Foley, J.I., 2020. Screening of dental plaque for presence of *Streptococcus mutans* bacteriophages. Saudi J. Oral Dent. Res.

Moisan, M., Moineau, S., 2012. Multilocus sequence typing scheme for the characterization of 936- like phages infecting *Lactococcus lactis*. Appl. Environ. Microbiol. 78, 2264-2271.

Monchois, V., Willemot, R.M., Monsan, P., 1999. Glucansucrases: Mechanism of action and structure-function relationships. FEMS Microbiol. Rev. 23, 131-151.

Munro, C., Michalek, S., Macrina, F., 1995. Sucrose-derived exopolymers have site-dependent roles in *Streptococcus mutans*-promoted dental decay. FEMS Microbiol. Lett. 15, 327-332.

Nakayama, N., Tsuge, T., Asakawa, S., Kimura, M., 2009. Morphology, host range and phylogenetic diversity of *Sphingomonas* phages in the floodwater of a Japanese paddy field. Soil Sci. Plant. Nutr. 55, 53-64.

Nussenzweig, P.M., McGinn, J., Marraffini, L.A., 2019. Cas9 cleavage of viral genomes primes the acquisition of new immunological memories. Cell Host Microbe 26, 515-526.e6.

Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., Sorek, R., 2018. DISARM is a widespread bacterial defence system with broad anti-phage activities. Nat. Microbiol. 3, 90-98.

Oliveira, P.H., Touchon, M., Rocha, E.P.C., 2014. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res. 42, 10618-10631.

Ooshima, T., Matsumura, M., Hoshino, T., Kawabata, S., Sobue, S., Fujiwara, T., 2001. Contributions of three glucosyltransferases to sucrose-dependent adherence of *Streptococcus mutans*. J. Dent. Res. 80, 1672-1677.

Perry, J.A., Jones, M.B., Peterson, S.N., Cvitkovitch, D.G., Lévesque, C.M., 2009. Peptide alarmone signalling triggers an auto-active bacteriocin necessary for genetic competence. Mol. Microbiol. 72, 905-917.

Petersen, F.C., Assev, S., Van der Mei, H.C., Busscher, H.J., Scheie, A.A., 2002. Functional variation of the antigen I/II surface protein in *Streptococcus mutans* and *Streptococcus intermedius*. Infect. Immun. 70, 249-256.

Pyenson, N.C., Marraffini, L.A., 2020. Co-evolution within structured bacterial communities results in multiple expansion of CRISPR loci and enhanced immunity. Elife 9, e53078.

Qi, F., Kreth, J., Lévesque, C.M., Kay, O., Mair, R.W., Shi, W., Cvitkovitch, D.G., Goodman, S.D., 2005. Peptide pheromone induced cell death of *Streptococcus mutans*. FEMS Microbiol. Lett. 251, 321-326.

Qi, F., Merritt, J., Lux, R., Shi, W., 2004. Inactivation of the ciaH gene in *Streptococcus mutans* diminishes mutacin production and competence development, alters sucrose-dependent

biofilm formation, and reduces stress tolerance. Infect. Immun. 72, 4895-4899.

Quivey, R.G., Faustoferri, R., Monahan, K., Marquis, R., 2000. Shifts in membrane fatty acid profiles associated with acid adaptation of *Streptococcus mutans*. FEMS Microbiol. Lett. 189, 89-92.

Rathsam, C., Eaton, R., Simpson, C., Browne, G., Berg, T., Harty, D., Jacques, N.A., 2005. Up-regulation of competence- but not stress-responsive proteins accompanies an altered metabolic phenotype in *Streptococcus mutans* biofilms. Microbiology 151, 1823-1837.

Russell, R.R.B., 1979. Glucan-binding proteins of *Streptococcus mutans* serotype c. J. Gen. Microbiol. 112, 197-201.

Samson, J.E., Moineau, S., 2013. Bacteriophages in food fermentations: New frontiers in a continuous arms race. Annu. Rev. Food Sci. Technol. 4, 347-368.

Samson, J.E., Spinelli, S., Cambillau, C., Moineau, S., 2013. Structure and activity of AbiQ, a lactococcal endoribonuclease belonging to the type III toxin-antitoxin system. Mol. Microbiol. 87, 756-768.

Sato, Y., Yamamoto, Y., Kizaki, H., 1997. Cloning and sequence analysis of the gbpC gene encoding a novel glucan- binding protein of *Streptococcus mutans*. Infect. Immun. 65, 668-675.

Senadheera, M.D., Guggenheim, B., Spatafora, G.A., Huang, Y.C.C., Choi, J., Hung, D.C.I., Treglown, J.S., Goodman, S.D., Ellen, R.P., Cvitkovitch, D.G., 2005. A VicRK signal transduction system in *Streptococcus mutans* affects gtfBCD, gbpB, and ftf expression, biofilm formation, and genetic competence development. J. Bacteriol. 187, 4064-4076.

Shah, S.A., Erdmann, S., Mojica, F.J.M., Garrett, R.A. 2014. Protospacer recognition motifs. RNA Biology 10:891–899.

Shah, D.S.H., Russell, R.R.B., 2004. A novel glucan-binding protein with lipase activity from the oral pathogen *Streptococcus mutans*. Microbiology 150, 1947-1956.

Sheng, J., Marquis, R.E., 2007. Malolactic fermentation by *Streptococcus mutans*. FEMS Microbiol. Lett. 272, 196-201.

Shibata, Y., Yamashita, Y., Van Der Ploeg, J.R., 2009. The serotype-specific glucose side chain of rhamnose-glucose polysaccharides is essential for adsorption of bacteriophage M102 to *Streptococcus mutans*. FEMS Microbiol. Lett. 294, 68-73.

Smith, D.J., Akita, H., King, W.F., Taubman, M.A., 1994. Purification and antigenicity of a

novel glucan-binding protein of *Streptococcus mutans*. Infect. Immun. 62, 2545-2552.

Spoering, A.L., Gilmore, M.S., 2006. Quorum sensing and DNA release in bacterial biofilms. Curr. Opin. Microbiol. 9, 133-137.

Sun, X., Göhler, A., Heller, K.J., Neve, H., 2006. The *ltp* gene of temperate *Streptococcus thermophilus* phage TP-J34 confers superinfection exclusion to *Streptococcus thermophilus* and *Lactococcus lactis*. Virology 350, 146-157.

Suttle, C.A., 2005. Viruses in the sea. Nature 437, 356-361.

Svensäter, G., Larsson, U.B., Greif, E.C.G., Cvitkovitch, D.G., Hamilton, I.R., 1997. Acid tolerance response and survival by oral bacteria. Oral Microbiol. Immunol. 12, 266-273.

Swarts, D.C., Jore, M.M., Westra, E.R., Zhu, Y., Janssen, J.H., Snijders, A.P., Wang, Y., Patel, D.J., Berenguer, J., Brouns, S.J.J., Van Der Oost, J., 2014. DNA-guided DNA interference by a prokaryotic Argonaute. Nature 507, 258-261.

Torsvik, V., Goksoyr, J., Daae, F.L., 1990. High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. 56, 782-787.

Torsvik, V., Øvreås, L., 2002. Microbial diversity and function in soil: From genes to ecosystems. Curr. Opin. Microbiol. 5, 240-245.

Twort, F.W., 1915. An investigation on the nature of ultra-microscopic viruses. Lancet 186, 1241-1243.

Van der Ploeg, J.R., 2009. Analysis of CRISPR in *Streptocccus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. Microbiology 155, 1966-1976.

Van Der Ploeg, J.R., 2005. Regulation of bacteriocin production in *Streptococcus mutans* by the quorum-sensing system required for development of genetic competence. J. Bacteriol. 187, 3980-3989.

Van Der Ploeg, J.R., 2007. Genome sequence of *Streptococcus mutans* bacteriophage M102. FEMS Microbiol. Lett. 275, 130-138.

van Houte, S., Buckling, A., Westra, E.R., 2016. Evolutionary ecology of prokaryotic immune mechanisms. Microbiol. Mol. Biol. Rev. 80, 745-763.

Walker, P.J., Siddell, S.G., Lefkowitz, E.J., Mushegian, A.R., Adriaenssens, E.M., Dempsey, D.M., Dutilh, B.E., Harrach, B., Harrison, R.L., Hendrickson, C., Junglen, S., Knowles, N.J., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Nibert, M., Orton, R.J., Rubino, L.,

Sabanadzovic, S., Smith, D.B., Varsani, A., Zerbini, F.M., Davison, A.J., 2020. Changes to virus taxonomy and the statutes ratified by the International Committee on Taxonomy of Viruses (2020). Arch Virol. 165, 2737-2748.

Westra, E.R., Van houte, S., Oyesiku-Blakemore, S., Makin, B., Broniewski, J.M., Best, A., Bondy-Denomy, J., Davidson, A., Boots, M., Buckling, A., 2015. Parasite exposure drives selective evolution of constitutive versus inducible defense. Curr. Biol. 25, 1043-1049.

Woese, C.R., Kandler, O., Wheelis, M.L., 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. U. S. A. 87, 4576-4579.

Yoshida, A., Kuramitsu, H.K., 2002. *Streptococcus mutants* biofilm formation: Utilization of a gtfB promoter-green fluorescent protein (PgtfB::gfp) construct to monitor development. Microbiology 148, 3385-3394.

Zeng, L., Wen, Z.T., Burne, R.A., 2006. A novel signal transduction system and feedback loop regulate fructan hydrolase gene expression in *Streptococcus mutans*. Mol. Microbiol. 62, 187-200.

Annex A: Article 4

A mutation in the *methionine aminopeptidase* gene provides phage resistance in *Streptococcus thermophilus*

Simon J. Labrie[1,2,6], Cas Mosterd[1,2], Stéphanie Loignon[2,3], Marie-Ève Dupuis[1,2], Philippe Desjardins[1], Geneviève M. Rousseau[1,2], Denise M. Tremblay[2,3], Dennis A. Romero[4], Philippe Horvath[5], Christophe Fremaux[5] and Sylvain Moineau[1,2,3]*

[1] Département de biochimie, de microbiologie, et de bio-informatique, Faculté des sciences et de génie, Université Laval, Québec City, QC, G1V 0A6, Canada

[2] Groupe de recherche en écologie buccale, Faculté de médecine dentaire, Université Laval, Québec City, QC, G1V 0A6, Canada

[3] Félix d'Hérelle Reference Center for Bacterial Viruses, Faculté de médecine dentaire, Université Laval, Québec City, QC, G1V 0A6, Canada

[4] DuPont Nutrition and Biosciences, 3329 Agriculture Dr, Madison, WI 53716 USA

[5] DuPont Nutrition and Biosciences, BP10, Dangé-Saint-Romain, 86220 France

[6] Present Address: SyntBioLab Inc., 4820-250, rue de la Pascaline, Lévis, G6W 0L9 Canada

* Corresponding author. Tel: +1 418 656 3712; Email: Sylvain.Moineau@bcm.ulaval.ca

Résumé

Streptococcus thermophilus est une bactérie lactique largement utilisée par l'industrie laitière pour la fabrication de yogourts et de fromages de spécialité. Les systèmes CRISPR-Cas sont l'un des mécanismes de résistance aux phages les plus répandus chez *S. thermophilus*. Peu d'informations sont disponibles sur les autres facteurs hôtes impliqués dans la réplication des phages chez cette espèce bactérienne. Nous avons utilisé la souche modèle *S. thermophilus* SMQ-301 et son phage virulent DT1, hébergeant la protéine anti-CRISPR AcrIIA6, pour montrer qu'un gène hôte codant pour une méthionine aminopeptidase (metAP) est nécessaire au phage DT1 pour compléter son cycle lytique. Une seule mutation dans metAP confère à *S. thermophilus* SMQ-301 une forte résistance contre le phage DT1. La mutation empêche une étape tardive du cycle lytique puisque l'adsorption du phage, la réplication de l'ADN et l'expression des protéines n'ont pas été affectées. Lorsque la souche mutée a été complétée avec la version de type sauvage du gène, le phénotype sensible au phage a été restauré. Lorsque cette mutation a été introduite dans d'autres souches de *S. thermophilus*, elle a fourni une résistance contre les phages de type *cos* (genre *Sfi21dt1virus)* mais la réplication des phages de type *pac* (*genre Sfi11virus*) n'a pas été affectée. La mutation du gène codant pour le MetAP induit un changement d'acide aminé dans un domaine catalytique conservé dans de nombreuses espèces bactériennes. L'introduction de la même mutation chez *Streptococcus mutans* a également fourni un phénotype résistant aux phages, ce qui suggère l'importance de la méthionine aminopeptidase dans la réplication des phages.

Abstract

Streptococcus thermophilus is a lactic acid bacterium widely used by the dairy industry for the manufacture of yogurt and specialty cheeses. CRISPR-Cas systems are one of the most prevalent phage resistance mechanisms in *S. thermophilus*. Little information is available about other host factors involved in phage replication in this food-grade streptococcal species. We used the model strain *S. thermophilus* SMQ-301 and its virulent phage DT1, harbouring the anti-CRISPR protein AcrIIA6, to show that a host gene coding for a methionine aminopeptidase (*metAP*) is necessary for phage DT1 to complete its lytic

cycle. A single mutation in *metAP* provides *S. thermophilus* SMQ-301 with strong resistance against phage DT1. The mutation impedes a late step of the lytic cycle since phage adsorption, DNA replication, and protein expression were not affected. When the mutated strain was complemented with the wild-type version of the gene, the phage sensitive phenotype was restored. When this mutation was introduced into other *S. thermophilus* strains it provided resistance against *cos*-type (*Sfi21dt1virus* genus) phages but replication of *pac*-type (*Sfi11virus* genus) phages was not affected. The mutation in the gene coding for the MetAP induces amino acid change in a catalytic domain conserved across many bacterial species. Introducing the same mutation in *Streptococcus mutans* also provided a phage resistant phenotype, suggesting the wide-ranging importance of the host methionine aminopeptidase in phage replication.

Abbreviations

acr	anti-CRISPR
BHI	Brain Heart Infusion
BIM	bacteriophage-insensitive mutant
CDM	chemically defined medium
CRISPR	clustered regularly interspaced short palindromic repeats
CSP	Competence Stimulating Peptide
DNA	deoxyribonucleic acid
DTT	DL-Dithiothreitol
LB	Luria Broth
LC-MS	liquid chromatography–mass spectrometry
LM17	M17 medium supplemented with 0.5% lactose
MOI	multiplicity of infection
MS	mass spectrometry
nt	nucleotide
OD_{600nm}	optical density at 600 nanometres
PCR	polymerase chain reaction
Sie	superinfection exclusion

Introduction

Bacteriophages are one of the most important drivers of bacterial evolution as they exert a constant selective pressure (Cobián Güemes et al., 2016). Bacteria inevitably evolve to acquire phage resistance, which is often associated with a fitness cost (Avrani and Lindell, 2015; Koskella and Brockhurst, 2014; Lennon et al., 2007; Vale et al., 2015). This evolutionary arms race underscores the complex network of phage-host interactions (Buckling and Brockhurst, 2012; Labrie et al., 2010; Samson et al., 2013). Comprehensive knowledge of how phages interact with their host components is available for some *E. coli* phage-host systems but the literature is very sparse for other bacterial species.

To protect themselves against these viral invaders, bacterial hosts have acquired and evolved numerous phage resistance systems (Buckling and Brockhurst, 2012; Labrie et al., 2010; Samson et al., 2013). These anti-phage systems are generally classified based on the step of the infectious cycle they interfere with, for example, inhibition of phage adsorption, preventing DNA entry, DNA degradation, and abortive infection, among others (Labrie et al., 2010). However, our knowledge of the bacterial arsenal against phages is partial as new defence systems are still being discovered (Barrangou et al., 2007; da Silva Duarte et al., 2019; Doron et al., 2018; Kronheim et al., 2018).

Streptococcus thermophilus is a lactic acid bacterium extensively used in milk fermentation, second only to *Lactococcus lactis* in its widespread usage. Phages infecting *S. thermophilus* have been historically divided into two large groups, the so-called *cos*- and *pac*-type phages, also named the *Sfi21dt1virus* and *Sfi11virus* genera, respectively. This grouping is based on comparative genome analyses and DNA packaging strategy (Le Marrec et al., 1997). In 2011, *S. thermophilus* phage 5093, which did not share the common features of either group, was characterised, prompting the creation of the eponymous taxon (Mills et al., 2011). Five years later, four members of a new *S. thermophilus* phage taxon (987) were reported (McDonnell et al., 2016). Still the *cos*- and *pac*-type phages are by far the most common *S. thermophilus* phages worldwide and are responsible for most milk fermentation failures (Achigar et al., 2017; Lavelle et al., 2018ab; McDonnell et al., 2017).

Only a few phage resistance mechanisms have been identified in *S. thermophilus*. Among the most common phage defence systems encoded by *S. thermophilus* genomes there are restriction-modification systems, as reflected by the number of different entries in REBASE (http://rebase.neb.com). Two superinfection exclusion (Sie) mechanisms encoded by prophages were also identified in *S. thermophilus* (Ali et al., 2014; Sun et al., 2006). Sie systems interfere with the phage genome ejection into the host cytoplasm. However, it is the CRISPR-Cas systems that seem to be the most dominant defence mechanism in this bacterial species (Barrangou et al., 2007; Horvath et al., 2008). CRISPR-Cas systems are bacterial adaptive immune systems that protect the cell from invading nucleic acids via targeted cleavage (Garneau et al., 2010). The bacterial cell accumulates short sequences, named spacers, from the phage genomes into CRISPR arrays that are composed of these spacers interspersed by short repeats (Barrangou et al., 2007; Deveau et al., 2008). These spacers act as the cell's memory of previous encounters and serve as guides for specific cleavage of invading DNA. The adaptive nature and efficiency of this system may explain why only a few other phage resistance mechanisms have been identified in *S. thermophilus*. However, the recent discovery of anti-CRISPR proteins (AcrIIA5 and AcrIIA6) in *S. thermophilus* virulent phages illustrates the ongoing armsrace between phages and their hosts, and suggests that additional mechanisms are at play (Hynes et al., 2017; Hynes et al., 2018).

Recently, phage-immunity systems in *S. thermophilus* that are linked to cryptic, non-inducible prophages were discovered (da Silva Duarte et al., 2019). Another study demonstrated that transient inactivation of the CRISPR-Cas system allows isolation of non-CRISPR-mediated bacteriophage insensitive mutants (BIMs) (McDonnell et al., 2018). Characterisation of these non-CRISPR BIMs offers the possibility of identifying new bacterial factors that can be mutated to reduce phage infection such as the newly described mutations in cell wall glycans (Szymczak et al., 2018).

Here, we used the *cos*-type virulent phage DT1 (Tremblay and Moineau, 1999) and its host strain SMQ-301 to identify an additional host factor involved in phage infection. This phage encodes an anti-CRISPR protein which is naturally disabling for most of the CRISPR1-Cas system of *S. thermophilus* (Hynes et al., 2018), one of the two type II-A

systems active in *S. thermophilus* SMQ-301. Using this phage-host system, spontaneous BIMs selected after phage challenge either arise from acquisition of new spacers within the CRISPR3 array (Deveau et al., 2008; Magadán et al., 2012) or from unknown mechanism(s). Sequencing the genome of *S. thermophilus* SMQ-301 BIMs that had not acquired new spacers led to the identification of the methionine aminopeptidase (MetAP) as a key host factor for the replication of *S. thermophilus cos*-type phages.

Materials and Methods

Bacterial growth and phage propagation

Table 11: Plasmids and bacterial strains used in this study.

Name	Characteristics	Reference
Plasmids		
pNZ123	*E. coli, L. lactis* and *S. thermophilus* shuttle vector	(De Vos, 1987)
pNZ123:*metAP*	pNZ123 with *metAP* gene cloned in XbaI site	This study
pNZ123:*metAP*^{H206Q}	pNZ123 with *metAP*^{H206Q} gene cloned in XbaI site	This study
pNZ123:*metAP*Smut	pNZ123 with *metAP* of *S. mutans* HER1503 gene cloned in XbaI site	This study
Bacterial strains		
S. thermophilus SMQ-301		(Labrie et al., 2015)
S. thermophilus SMQ-301 BIM #2		This study
S. thermophilus SMQ-301 BIM #3		This study
S. thermophilus SMQ-301 BIM #5		This study
S. thermophilus SMQ-301:*metAP*^{H206Q}		This study
S. thermophilus SMQ-301:*metAP*^{H206Q} + pNZ123: *metAP*		This study
S. thermophilus DGCC7710		(Barrangou et al., 2007)
S. thermophilus DGCC7710:*metAP*^{H206Q}		This study
S. thermophilus DGCC7710:*metAP*^{H206Q} + pNZ123: *metAP*		This study
S. thermophilus DGCC7796		This study
S. thermophilus DGCC7796:*metAP*^{H206Q}		This study
S. thermophilus DGCC7796:*metAP*^{H206Q} + pNZ123:*metAP*		This study
S. thermophilus DGCC782		This study
S. thermophilus DGCC782:*metAP*^{H206Q}		This study
S. thermophilus DGCC782:*metAP*^{H206Q} + pNZ123:*metAP*		This study
*S. mutans*P42S		(Delisle et al., 2012)
*S. mutans*P42S:*metAP*^{H206Q}		This study
*S. mutans*P42S:*metAP*^{H206Q} + pNZ123:*metAP*Smut		This study

Bacterial strains, phages and plasmids used in this study are listed in Table 11 and Table 12. *S. thermophilus* was grown at 37 °C for the pre-cultures or 42 °C for the assays in M17 medium (Oxoid) supplemented with 0.5% lactose (LM17). When needed, chloramphenicol (Sigma) was added to a final concentration of 5 µg/mL for growth and selection of *S. thermophilus* strains containing pNZ123 (De Vos, 1987) and derivatives. Agar was added to a final concentration of 1% for solid media. *Escherichia coli* was grown with agitation at 37 °C in Luria Broth (LB) supplemented with 20 µg/mL chloramphenicol for selection. *Streptococcus mutans*P42S was grown at 37 °C and 5% CO_2 in Brain Heart Infusion (BHI, Difco), supplemented with 10 µg/mL of chloramphenicol when selecting for cells carrying pNZ123. Phages were propagated as previously described (Hynes et al., 2017). Phage adsorption assays were performed as described elsewhere (Garvey et al., 1996) and phage DNA replication assays were performed as reported (Boucher et al., 2000). For the growth assays, bacterial strains were grown overnight at 37 °C. M17 medium supplemented with 0.5% lactose (LM17) were inoculated at 1% with the pre-cultures and incubated at 42 °C. The OD_{600nm} was taken every 30 min for 8 hours. Generation time was calculated based on three biological triplicates and from the linear section of the semi-logarithm graph of the OD_{600nm} as a function of time.

Table 12: Phages used in this study.

	Host range					
Phages	*S. thermophilus* SMQ-301	*S. thermophilus* DGCC7796	*S. thermophilus* DGCC7710	*S. thermophilus* DGCC782	*S. mutans*P42S	References
S. thermophilus *cos*-type phages:						
DT1	X					(Tremblay & Moineau, 1999)
MD2	X					(Duplessis & Moineau, 2001)
D4090		X				This study
D4807		X				This study
D5821		X				This study
D5691			X			This study
D5913			X			This study
D6037			X			This study
D6215			X			This study
N1032				X		This study
N1117				X		This study
N1119				X		This study
N1169				X		This study
				X		This study

N1358				X		This study
N3782						This study
S. thermophilus *pac*-type phages:						
D2765		X				This study
D4274		X				This study
M5876		X				This study
D5787		X				This study
858			X			(Barrangou et al., 2007)
2972			X			(Lévesque et al., 2005)
D4259			X			This study
D939			X			This study
D3288			X			This study
D4752			X			This study
D4754			X			This study
S. mutans *cos*-type phage:						
M102AD					X	(Delisle et al., 2012)

Bacteriophage-insensitive mutant isolation

BIMs were isolated using the soft agar overlay assay as detailed elsewhere (Hynes et al., 2017). Briefly, the wild-type host SMQ-301 was challenged with the lytic phage DT1. The bacterial host was grown in LM17 at 42 °C until the optical density at 600 nm (OD_{600nm}) reached 0.6. Then, 300 µL of the culture was added to 3 mL of molten LM17 soft agar supplemented with calcium chloride (10 mM) and phage DT1 was added to achieve a multiplicity of infection (MOI) of 0.1. Plates were incubated at 42 °C for 16 h. The resulting colonies were streaked on LM17 + agar and screened for spacer acquisition in CR1 and CR3 loci using primers CR1-fwd and CR1-revLong for the CR1 locus, and primers CR3-fwd and CR3-rev for the CR3 locus (Hynes et al., 2017) (Table S5). When spacer acquisition was not detected in CRISPR arrays, BIMs were conserved for genome sequencing.

DNA isolation, sequencing and bioinformatics analysis

The genomic DNA of the selected BIMs was extracted as previously described (Bissonnette et al., 2000). The genomes were sequenced using the Illumina MiSeq platform. The libraries were prepared with the Nextera XT DNA sample preparation kit according to the manufacturer's instructions and sequenced using MiSeq reagents (2 × 250

nt paired-end). The average coverage ranged from 6.8- to 80.1-fold. The DNA reads obtained for the genome of nine BIMs were aligned on the *S. thermophilus* SMQ-301 wild-type genome (Labrie et al., 2015) using Novoalign (http://www.novocraft.com) with the default settings. The mutations were extracted from the alignment file using SAMTools (Li et al., 2009). In-house Python scripts were used to map the mutations in the bacterial genome and to determine their impact in coding sequences. The mutations with the highest score as provided by SAMTools were considered first.

Complementation assays

The BIM *S. thermophilus* SMQ-301:metAP^{H206Q} was complemented with the wild-type *metAP*. First, *metAP* was cloned into pNZ123 using Gibson assembly (Gibson et al., 2009). The vector pNZ123 was linearised with XbaI (Roche) according to the manufacturer's instructions. The insert was amplified with the primers SJL154 and SJL155 (Table S5) with Q5 high-fidelity DNA polymerase (NEB). Both primers had 30-nt extensions complementing the 3′- and 5′-ends, respectively, of the linearised vector. An insert/vector ratio of 3:1 was used for assembly. The master mixture for Gibson assembly was prepared as described previously (Gibson et al., 2009). The resulting construction was transformed into *E. coli* NEB5α according to the manufacturer's instructions and the clones were selected on LB medium supplemented with agar and 20 µg/mL chloramphenicol. One clone was confirmed by Sanger sequencing (ABI 3730xl) at the Plateforme de séquençage et de génotypage des génomes at Université Laval. Plasmid DNA was extracted using a QIAprep Spin Miniprep kit (Qiagen) and transformed into *S. thermophilus* (Hynes et al., 2017). The clones were selected by spreading the transformation mixture on LM17 supplemented with agar and 5 µg/mL of chloramphenicol.

Proteomic analysis of the phage-infected *S. thermophilus* DGCC7796 cells

Phage infection and protein analyses were performed as previously described (Scaltrlti et al., 2009). Briefly, *S. thermophilus* DGCC7796 and *S. thermophilus* DGCC7796:metAP^{H206Q} were each grown to an OD_{600nm} of 0.5 at 42 °C. The cultures were infected with *cos*-type phage D4090 or *pac*-type phage M5876 at a MOI of 5. The infections were arrested after 20 min by harvesting the infected cells by centrifugation and

the cell pellet was flash-frozen at −80 °C. The cell pellet was quickly thawed and resuspended in 200 µL of lysis buffer (0.5% sodium deoxycholate, 50 mM ammonium bicarbonate, 5 mM DL-Dithiothreitol (DTT), Complete Protease Inhibitor Cocktail (1 µL for 12.5 mL of buffer, Roche)). An equal volume of acid-washed glass beads was added to the cell suspension. The mixture was vortexed in a Mini-Beadbeater-8 cell (BioSpec Products) five times for 1 min with 1 min intervals on ice. The final lysate was centrifuged and the soluble fraction was kept at −80 °C until analysis. The LC-MS/MS analysis was done using a QTRAP 6500 at the Centre de Protéomique de l'Est du Québec. The spectra were analysed using the complete peptide database, a peptide database that contained only the N-terminal peptides, and a database that contained only the N-terminal peptides without the N-terminal methionine. The latter two databases allowed the distinction of post-translational processed proteins from native proteins that still had the N-terminal methionine. The final analysis was done with Scaffold v4.

Directed and random mutagenesis

The *metAP* gene with the desired mutation was amplified by PCR using the primers SJL128 and SJL130 (Table S5). At least 10 PCR reactions of 50 µL were done for each assay to obtain enough DNA for natural transformation. All reactions were pooled and precipitated by adding 1.5 mL of 95% ethanol and 75 µL of sodium acetate 3 M (pH 5.2). The tubes were centrifuged at 25,000 × g for 20 min at 4 °C. The DNA pellet was washed twice with 1 mL of 70% ethanol and let dry for 10 min at room temperature to remove traces of ethanol. Finally, the pellets were dissolved in 150 µL of water. The linear DNA fragments were introduced into *S. thermophilus* by natural transformation (Fontaine et al., 2010; Gardan et al., 2009). Briefly, *S. thermophilus* strains SMQ301, DGCC7710, DGCC7796 and DGCC782 were grown overnight at 37 °C in 1 mL of LM17. The cultures were centrifuged at 17,000 × g for 1 min, and the bacterial cells were washed twice with chemically defined medium (CDM) (Fontaine et al., 2010; Gardan et al., 2009) and recovered in 1 mL of CDM. The final cultures were diluted 30-fold to obtain an OD_{600nm} of 0.05 and incubated 75 min at 37 °C before storing at −20 °C. For the transformation, 1 µM of the ComS peptide and 1 µg of linear DNA were added to 300 µL of naturally competent cells. The mixture was incubated for 3 h at 37 °C. The cells were serially diluted and spread

on non-selective LM17 agar to obtain isolated colonies. The colonies were screened using a PCR approach specifically designed to detect clones with the desired mutation. One of the primers used in this protocol included the desired mutation at its 3′-end. Thus, the PCR amplification was only positive if the mutation was present in the clone screened. Primers SJL150 and SJL151 (Table S5) were used for screening the transformants. A similar approach was used for random mutagenesis, but we amplified the wild-type *metAP* gene using an error-prone PCR amplification. The proof-reading activity of the DNA polymerase was reduced by adding 10 µM of $MnSO_4$. Phage DT1 was used for the selection of resistant clones.

Directed *metAP* mutagenesis in *Streptococcus mutans*

The *metAP* gene was amplified from *S. mutans* P42S with two sets of primers containing the desired mutation ($MetAP^{H206Q}$). The two sets of primers (SJL160 and SJL161 as well as SJL162 and SJL163; Table S5) generated two overlapping amplicons, with the mutation in the overlapping region. The two fragments were purified using the QIAquick PCR purification kit (Qiagen) and fused by Gibson Assembly (Gibson et al., 2009). A PCR was then performed on the Gibson Assembly sample using primers SJL160 and SJL163 (Table S5) to amplify the fused fragments. After purification using the QIAquick PCR purification kit, the mutated *metAP* gene was transformed into *S. mutans* P42S by natural transformation. The Competence Stimulating Peptide (CSP) ordered from Biomatik was added to 500 µl of an exponentially growing culture of *S. mutans*P42S (OD_{600nm} of 0.1) at a concentration of 1 µM along with 1 µg of purified mutated *metAP* amplicon (Dufour et al., 2011). The culture was incubated overnight at 37 °C with 5% CO_2, then mixed with the virulent phage M102AD (Delisle et al., 2012) in BHI supplemented with 0.7% agar and plated onto BHI supplemented with agar. The *metAP* gene in the surviving BIMs was amplified by PCR using primers SJL160 and SJL163 (Table S5) and sequenced to confirm the mutation.

For the complementation of the *S. mutans* BIM P42S:$metAP^{H206Q}$, the wild-type *S. mutans metAP* was cloned into pNZ123. pNZ123 was linearized with XhoI and EcoRI to remove the 24 bp between the two sites and the resulting plasmid was purified using the

QIAquick Gel Extraction kit. The *metAP* gene was amplified from *S. mutans* P42S using primers CM145 and CM146 (Table S5) to generate the gene, flanked by XhoI and EcoRI restriction sites. The PCR sample was purified using the QIAgen PCR purification kit, cut by XhoI and EcoRI and ligated into the linearised pNZ123 using T4 DNA ligase. The molar insert/vector ratio used during the ligation was 3:1. The ligated plasmid was transformed into *E. coli* NEB5α and plated onto solid LB medium supplemented with 20 µg/mL chloramphenicol. Amplification of the *metAP* gene by primers CM145 and CM146 and sequencing confirmed the sequence cloned in the complementation plasmid. Plasmid DNA was extracted using a QIAprep Spin Miniprep kit and transformed into the *S. mutans*P42S:metAP^{H206Q}. The strain was grown in BHI until an OD_{600nm} of 0.1 was reached, an aliquot of 500 µL was exposed to 1 µM CSP and transformed with 1 µg of the complementation plasmid. After a 2.5-hour incubation period at 37 °C with 5% CO_2, the culture was plated onto BHI + agar supplemented with 10 µg/mL of chloramphenicol.

Mutation stability test

To test the stability of the mutation H206Q, we inoculated (1%) the strains *S. thermophilus* SMQ-301 and DGCC7796 and their respective mutants in 10 mL of M17 medium and in 10% reconstituted non-fat dry milk. Strains were incubated at 42 °C during the day. Then, an inoculum of 1% was transferred to fresh media or milk and incubated at 37 °C for the night. After 9 transfers (5 days, 4 nights) corresponding to approximately 60 generations, we randomly selected 10 colonies from each of the four strains and tested them for sensitivity to phage DT1 (*cos*) for *S. thermophilus* SMQ-301 and *S. thermophilus* SMQ-301:metAP^{H206Q}, and phages D4090 (*cos*) and D4274 (*pac*) for *S. thermophilus* DGCC7796 and *S. thermophilus* DGCC7796:metAP^{H206Q}. Sequencing of the *metAP* gene and CR1 was done for SMQ-301 and SMQ-301:metAP^{H206Q} colonies with primers SJL128/SJL130 and CR1-fwd/CR1-revLong, respectively, to confirm both the presence of the mutation after the transfers and the identity of the strains.

Results and discussion

A mutation in the gene coding for the methionine aminopeptidase provides phage resistance

We isolated BIMs of *S. thermophilus* SMQ-301 that were resistant to phage DT1. Most of the BIMs had acquired new spacers targeting the phage DT1 genome in their CRISPR3 locus (data not shown). Since we were interested in the discovery of other host factors involved in phage infection, we selected nine SMQ-301 BIMs that had not acquired any new spacer, conjecturing they had a mutation elsewhere in their chromosome to provide phage resistance. We sequenced the genomes of these nine non-CRISPR BIMs. By comparing these sequences with that of the wild-type strain SMQ-301 (accession CP011217.1) (Labrie et al., 2015) we found several mutations (data not shown). Although many of these mutations may potentially be involved in phage infection, the same transversion at position 1,426,980 occurred in the genomes of three of the nine BIMs (designated BIM #2, #3, and #5). This mutation resides within the gene coding for methionine aminopeptidase (MetAP; locus_tag SMQ301_1544, T618G) and induces an amino acid substitution from a histidine to a glutamine at position 206 (H206Q) in the protein. BLAST searches indicated that this histidine is conserved in all MetAP sequences analysed, including those deduced from all *S. thermophilus* sequences currently available in GenBank (29 complete and 27 draft genomes; data not shown).

While the MetAP of *S. thermophilus* has not been previously studied, it shares 89% identity and 96% similarity with the MetAP of *Streptococcus pneumoniae* TIGR4 for which the structure has been determined (PDB 4KM3)(Arya et al., 2013). Although the histidine at position 206 was not annotated as part of the active site (Fig. 17), this residue was shown to be a substrate-binding site in the MetAP of *E. coli*(Lowther et al., 1999). While this site is not essential for the catalytic activity of *E. coli* MetAP, replacing the corresponding histidine at position 178 with an alanine led to an impairment of the catalytic activity by 70–190 fold (Copik et al., 2003).

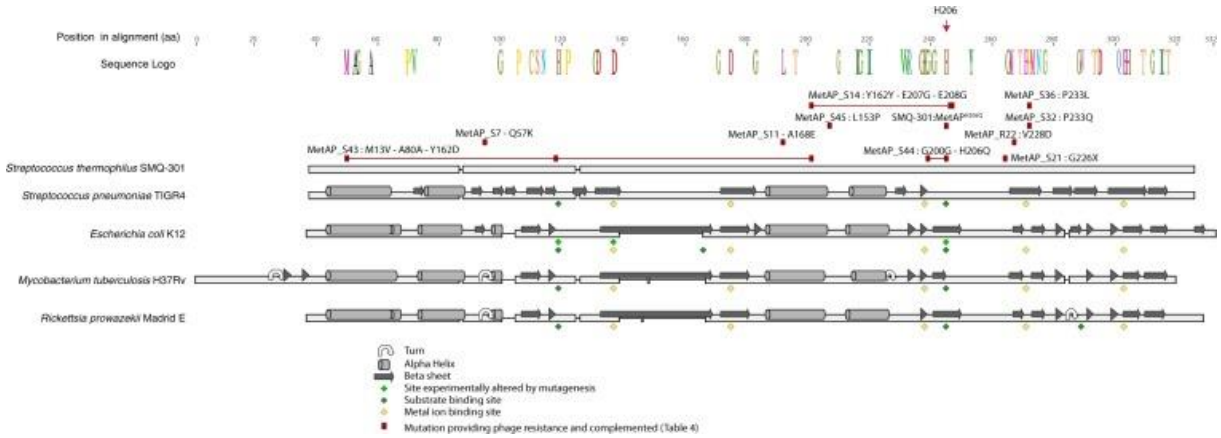


Figure 17: Protein alignment of *S. thermophilus* SMQ-301 MetAP with four other MetAP for which the structure is available. The WebLogo was only kept when the residue was conserved in all sequences. The secondary structures are represented over each sequence and diamonds indicate active sites and substrate binding sites. Red boxes highlight the mutations that provide phage resistance listed in Table 14. When linked, they occurred in the same MetAP mutant of *S. thermophilus* SMQ-301. Uniprot accession number of the protein sequences: *E. coli* K12 (P0AE18), *Rickettsia prowazekii* Madrid E (Q9ZCD3), *S. pneumonia* TIGR4 (B2IQ22) and *M. tuberculosis* H37Rv(P9WK19).

A mutation in the *metAP* gene confers phage resistance in *S. thermophilus*

To confirm that the phage resistance phenotype was due to the MetAP^{H206Q} mutation – and not to other mutations in the genome – the mutated gene *metAP*^{H206Q} was amplified and transformed into wild-type strain SMQ-301 using natural competence. Selecting with phage DT1 is very efficient, but it also increases the odds of selecting CRISPR BIMs as well as other mutations that confer phage resistance. Instead, we designed a PCR strategy to detect the transformants that integrated the PCR product with the mutation into their genome, in the absence of phage selective pressure. We used the primers SJL150 in combination with SJL151 with the desired mutation at its 3′-end (Table S5) to specifically detect the mutation in the bacterial genome. To avoid false positives due to the potential presence of residual linear DNA from the transformation inside the bacterial cytoplasm, we designed the primer SJL150 that matched the flanking genomic region of the transformed DNA fragment. A total of 282 clones were screened for the desired mutation and four positive clones were obtained (MetAP^{H206Q}). We randomly selected one of these clones (designated SMQ-301:metAP^{H206Q}) and tested its phage resistance phenotype. No plaque was visible on the bacterial lawn of SMQ-301:metAP^{H206Q} using various titres of the phage DT1, indicating a very strong phage resistance phenotype (Table 13). It was possible to observe a lysis zone at high phage titres, likely due to lysis from without or to the

endolysin activity in the phage lysate. We also tried to isolate DT1 phage mutants that would overcome the effect of the MetAP^{H206Q} mutation using various experimental conditions, but to no avail. We examined incubation temperature, anaerobic conditions, added different concentrations of glycine to the medium to weaken the bacterial cell wall, replaced agar with different concentrations of agarose and even using liquid medium in case the lysis plaque were not visible enough. Overall, we were unable to obtain phage DT1-derivatives able to propagate on SMQ-301:metAP^{H206Q}.

Table 13: Effect of MetAP mutations on phage efficiency of plaquing (EOP) and adsorption. ND = Not determined; DT1 initial titre = 6×10^9 PFU/ml; M102AD initial titre = 1×10^9 PFU/ml. PFU = Plaque forming unit.

		Strain + pNZ123		**Strain + pNZ123:*metAP* or pNZ123:*metAPSmut***	
Strain	**Phage**	***EOP***	***Adsorption %***	***EOP***	***Adsorption %***
S. thermophilus SMQ-301	DT1	1	93.8 ± 2.2	1	90.9 ± 3.3
S. thermophilus SMQ-301 BIM #2	DT1	1.7×10^{-8}	86.2 ± 3.2	1.7	86.0 ± 2.1
S. thermophilus SMQ-301 BIM #3	DT1	1.7×10^{-8}	83.2 ± 7.1	0.02	89.1 ± 2.0
S. thermophilus SMQ-301 BIM #5	DT1	1.7×10^{-8}	80.4 ± 7.7	2.4	89.0 ± 4.4
S. thermophilus SMQ-301:metAP^{H206Q}	DT1	1.7×10^{-8}	68.0 ± 3.8	1	94.9 ± 1.1
S. mutans P42S	M102AD	1	ND	ND	ND
S. mutans P42S:metAP^{H206Q}	M102AD	$<1 \times 10^{-7}$	ND	1.1	ND

Complementation with the wild-type allele restores phage sensitivity

We cloned the wild-type *metAP* gene from strain SMQ-301 into the expression vector pNZ123. The resulting construct was transformed into *S. thermophilus* SMQ-301:metAP^{H206Q} to complement the mutation in *trans*. Sensitivity to phage DT1 was completely restored for the two of the three spontaneous BIMs complemented with plasmid pNZ123:metAP (Table 13), confirming that the metAP^{H206Q} mutation was solely responsible for the phage resistance phenotype. Of note, the sensitivity to phage DT1 was not completely restored (EOP = 0.02) for the spontaneous BIM #3, suggesting that other

mutations are at play (Table 13). The mutated SMQ-301:metAP^{H206Q}*metAP* gene was also cloned into pNZ123 and transformed into the wild-type strain SMQ-301. There was no difference in EOP between the strain with pNZ123 when compared to the strain with pNZ123:metAP^{H206Q}, suggesting that the mutation does not produce a dominant phenotype.

The MetAP mutation has a broad range of action against cos-type phages

To determine if the resistance provided by the MetAP^{H206Q} mutation is phage-dependent, we tested the *cos*-type phage MD2 (Duplessis and Moineau, 2001), which is also capable of infecting strain SMQ-301. Phage MD2 was also severely inhibited by the mutated MetAP (EOP $<1 \times 10^{-6}$). To show that this mutation is not bacterial strain-dependent, we introduced the MetAP^{H206Q} mutation into several *S. thermophilus* strains using directed mutagenesis. We introduced the mutation into strain DGCC782, which is sensitive to *cos*-type phages N1032, N1117, N1119, N1169, N1358 and N3782 (Table 12). The mutation was also introduced into the strain *S. thermophilus* DGCC7710, which is sensitive to *cos*- (D5691, D5913, D6037 and D6215) and *pac*-type phages (858, 2972, D3288, 4259, D4752, D4754 and D939) (Table 12). Finally, the mutation was introduced into *S. thermophilus* DGCC7796, which is also sensitive to *cos*- (D4090, D5821 and D4807) and *pac*-phages (D2765, D4274, D5787 and M5876) (Table 12). Although the mutation was less efficient against all four *cos*-type phages infecting *S. thermophilus* DGCC7710 (EOP 10^{-4}), the mutation provided resistance against all *cos*-type phages tested. The introduction of MetAP^{H206Q} into *S. thermophilus* DGCC7796 and DGCC782 led to a reduction in the EOP of at least 10^{-6} for all *cos*-type phages (Table 12), which corresponded to the limit of detection. No phage mutant could be recovered from these assays. However, none of the *S. thermophilus pac*-type phages were affected by the mutation, suggesting that MetAP is not key for the replication of these phages.

To determine if the MetAP^{H206Q} mutation would provide resistance to another streptococcal species, we introduced it into the *S. mutans* strain P42S, which is sensitive to the *cos*-type virulent phage M102AD (Delisle et al., 2012). The methionine aminopeptidase gene of *S. mutans*P42S has the same length (861 bp) as that of *S. thermophilus*. The two genes share 74% identity while the two MetAP share 85% identity. The histidine residue

found at position 206 is present in *S. mutans*. The MetAP^{H206Q} mutation inserted into *S. mutans* provided complete resistance to phage M102AD (EOP $<1 \times 10^{-7}$). Complementation of the mutated strain with the wild-type *metAP* restored phage sensitivity. Our data suggest a relatively widespread role of MetAP in phage replication.

Other mutations in MetAP affect replication of phage DT1

To determine if other mutations in the *metAP* gene provide phage resistance, we used an error-prone PCR amplification approach. The resulting PCR products were transformed into *S. thermophilus* SMQ-301 using natural competence. Since we were not introducing an antibiotic resistance marker, we used phage DT1 to select the clones. The transformation assay with the positive control (MetAP^{H206Q} mutation) resulted in many more BIMs than with the negative control (no DNA) (Table S5). Fewer BIMs were obtained with the PCR product containing random mutations than with the positive control (Table S5).

Table 14: Random mutagenesis of the metAP gene of *S. thermophilus* SMQ-301 and resistance to DT1.

Name	Position in the gene	Mutation	Codon change	Amino acid change	EOP	EOP after complementation
MetAP_S7	205	C>A	CAG>AAG	Q57K	$<\times 10^{-6}$	5×10^{-1}
MetAP_S11	538	C>A	GCG>GAG	A168E	$<\times 10^{-6}$	3×10^{-3}
MetAP_S14	523 657 660	T>C A>G A>G	TAT>TAC GAG>GGA GAG>GGA	Y162Y E207G E208G	$<\times 10^{-6}$	5×10^{-2}
MetAP_S21	711	G>T	GGA>TGA	G226X	$<\times 10^{-6}$	3×10^{-1}
MetAP_S32	698	C>A	CCA>CAA	P233Q	$<\times 10^{-6}$	7×10^{-1}
MetAP_S36	698	C>T	CCA>CTA	P233L	$<\times 10^{-6}$	6×10^{-2}
MetAP_S43	37 240 484	A>G A>G T>G	ATG>GTG GCA>GCG TAT>GAT	M13V A80A Y162D	$<\times 10^{-6}$	2×10^{-1}
MetAP_S44	600 618	A>G G>A	GGA>GGG CAG>CAA	G200G H206Q	$<\times 10^{-6}$	1
MetAP_S45	458	T>C	CTT >CCT	L153P	$<\times 10^{-6}$	1.1
MetAP_R22	684	T>A	GTC>GAC	V228D	$<\times 10^{-6}$	1

The *metAP* gene was sequenced in 10 randomly selected clones and all of them had various mutations. Seven mutants had a single mutation in the *metAP* gene resulting in an

amino acid substitution. We tested phage sensitivity of these mutants and they were all resistant to phage DT1. We complemented the 10 resistant strains with pNZ123:metAP, and as previously observed, the wild-type MetAP restored phage sensitivity, confirming that other mutations in the MetAP can confer phage resistance (Table 14; Fig. 17).

Phage adsorption and phage DNA replication are not affected

To understand the molecular mechanism underlying the phage resistance provided by this mutation, we dissected the lytic cycle of a phage infecting a mutated strain. First, we tested if phage DT1 could still adsorb to *S. thermophilus* SMQ-301:metAP^{H206Q} (Table 13). The viral particles adsorbed at a slightly higher level on the wild-type phage-sensitive host (93%) than on the spontaneous BIMs #2, #3, and #5 (80% to 86%). The adsorption of DT1 was more reduced on *S. thermophilus* SMQ-301:metAP^{H206Q} (68%). However, this reduced adsorption cannot explain the strong phage resistance phenotype (EOP of 10^{-8}), suggesting that another step of the lytic cycle is impeded by the MetAP^{H206Q} mutation.

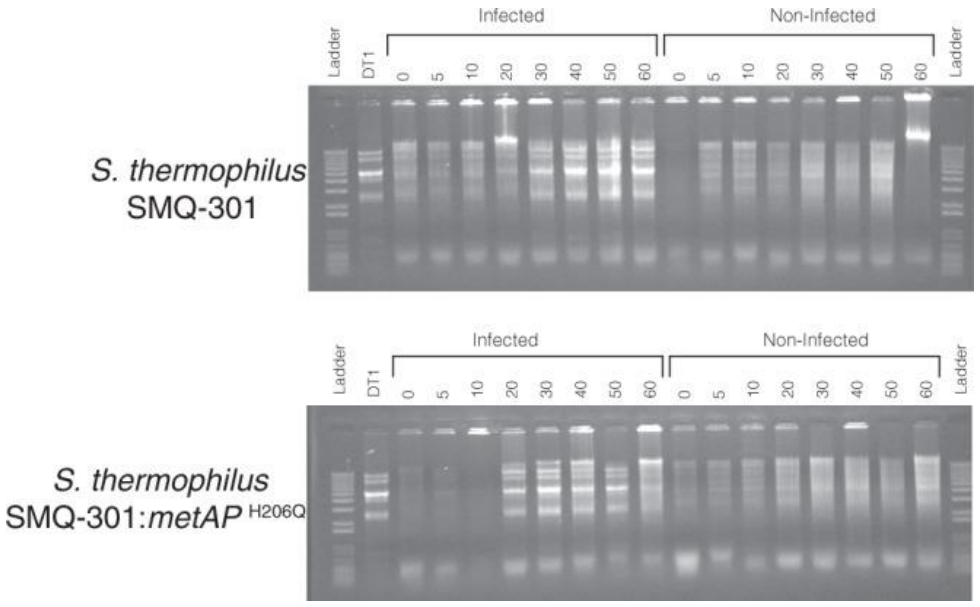


Figure 18: DNA replication of DT1 in the wild-type *S. thermophilus* SMQ-301 and *S. thermophilus* SMQ-301:metAP^{H206Q}. Values above electrophoresis gels represent time points (in min) at which total DNA was extracted from cells grown in the presence or absence of phage DT1.

Since phage DT1 can adsorb to the BIM surface, we tested whether phage DNA was replicated inside the bacterial cell. Both *S. thermophilus* SMQ-301 and SMQ-

301:metAP^{H206Q} were infected with phage DT1 using a MOI of 5 and total DNA was extracted from the infected cells at different time points, digested with a restriction enzyme and migrated on an agarose gel (Fig. 18).

The DNA fragments corresponding to DT1 digested genome were visible after 20 min of infection in both the MetAP^{H206Q} mutant and the wild-type cells, indicating that the phage DNA entered the bacterial cell and was replicated (Fig. 18). Lysis of the culture was only observed with *S. thermophilus* SMQ-301.

The MetAP^{H206Q} mutation affects N-terminal methionine processing

Since the MetAP^{H206Q} mutation does not affect phage DNA replication, we investigated whether protein expression was affected by the mutation. To compare the activity of MetAP, we used *S. thermophilus* DGCC7796, a strain sensitive to *cos*- and *pac*-type phages. Using mass spectrometry (LC-MS/MS), we analysed the complete proteome of the wild-type and mutant DGCC7796 after 20 min of infection with phages D4090 (*cos*-type) or M5876 (*pac*-type) using a MOI of 5.

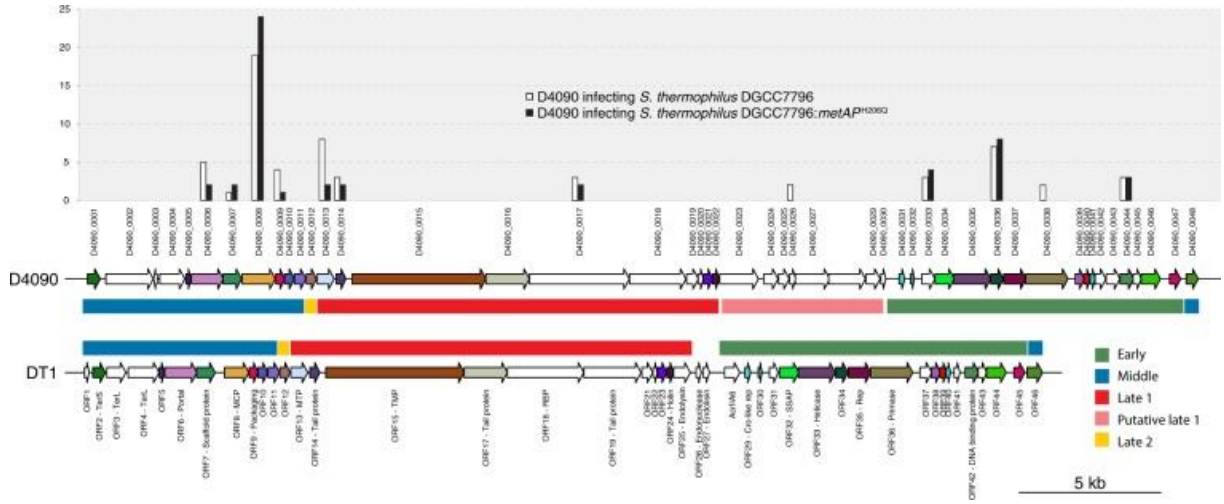


Figure 19: Genetic alignment of phages D4090 and DT1. Each protein-coding gene is represented by an arrow. When two deduced proteins share 70% identity or more, they are represented with the same color, otherwise they are shown in white. The boxes between the two genomes represent the expression modules. Phage D4090 expression modules were extrapolated from DT1 experimental data. The graph above the alignment represents the relative abundance of the proteins when phage D4090 infects the wild-type strain DGCC7796 and mutant DGCC7796:metAP^{H206Q}.

During the replication of phage D4090, although we observed a slight difference in protein expression, there was no significant trend in protein expression between mutant and wild-type strains (Fig. 19, Table S5). Peptides of 10 and 12 phage proteins were detected in the wild-type and mutant strains, respectively, and their relative abundance was similar in both strains. The difference in the number of proteins detected can be explained by the very low level of detection of the corresponding proteins in the wild-type strain.

We then used the proteomic datasets to verify the activity of the mutant and wild-type MetAP. We generated two distinct databases of peptides to search against. The first database was composed of only N-terminal peptides after tryptic digestion. The second database was essentially the same as the first one but we removed the N-terminal methionine of each peptide. This allowed us to determine if the protein underwent N-terminal post-translational processing. The complete database included the 2,051 proteins encoded by the host genome and the 41 proteins encoded by the phage genome. Following in silico tryptic digestion, the N-terminal peptide database included 579 peptides of 8–30 amino acids that have a molecular mass that could be detected with the setting used to conduct the mass spectrometry analysis. Using LC-MS/MS, we detected a total of 134 from 579 N-terminal peptides across all samples analysed (Table S6). We did not observe any difference in post-translational processing of the N-terminal methionine between the uninfected and infected samples nor between the *cos*- and *pac*-type phage-infected wild-type cells, suggesting that phage infection does not affect MetAP activity. Then, we compared the data with the MetAP mutated cells and classified the results into four groups: A) 28 proteins were unprocessed by MetAP in both wild-type and mutant strains, B) 34 proteins were processed by MetAP in both the wild-type and mutant strains, C) 22 proteins were processed by MetAP in the wild-type strain only, and D) only one protein was processed by MetAP in the mutated strain (Fig. 20).

E. coli methionine aminopeptidase is more specific for protein with a penultimate amino acid (a.a.) with small side chain (e.g. Ala, Gly, Pro, Ser, Thr, or Val)(Hirel et al., 1989; Xiao et al., 2010). Thus, it is not surprising that most proteins (Thr is the only exception out of 28) found in group A (non-processed) have penultimate a.a. with long side chains (Ile, Asn, Leu, Glu, Tyr, Gln and Asp). Most of the penultimate a.a. of proteins

found in groups B and C have short side chains (Ala, Gly, Pro, Ser, Thr, or Val) with only two exceptions found in group C (Cys and Lys). Since there are no significant differences in the penultimate a.a. of groups B and C, the findings that some proteins were not processed by the mutated MetAP (group C) suggest that its activity is impaired to some degree. Analysing functions of the above proteins and metabolic pathways in which they are involved did not reveal why phage infection was blocked in the presence of the mutated MetAP (Table S6).

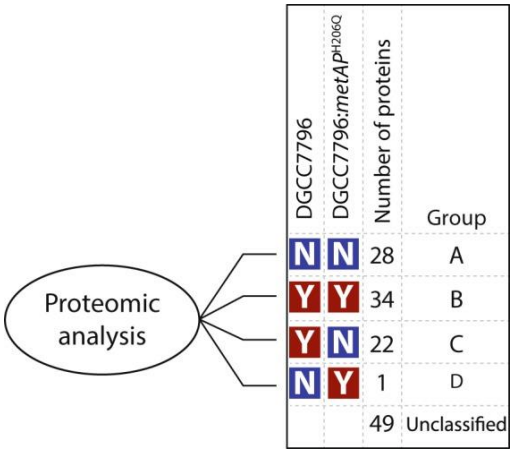


Figure 20: Proteomic analysis of the N-terminal peptides from the proteome of the wild-type *S. thermophilus* strain DGCC7796 and mutant DGCC7796:metAP^{H206Q}. Post-translational processing of the N-terminal methionine was detected in both proteomes but it was less abundant in the mutant strain. N represents unprocessed peptides while Y stands for processed peptides.

The MetAP^{H206Q} mutation affects growth of the bacterial strains

Growth of the mutant strains was also evaluated (Fig. 21). Surprisingly, the MetAP^{H206Q} mutation had a different impact depending *S. thermophilus* strains. The growth rate difference between the wild-type strains and their mutants was not significant, except for *S. thermophilus* DGCC7796 and its mutant for which the growth rates were respectively 36.2 ± 3.6 min and 53.5 ± 3.3 min (Table 15). We also observed a latent period longer with *S. thermophilus* DGCC7796:metAP^{H206Q} than with the other strains tested (Fig. 21).

The MetAP^{H206Q} mutation is stable

One of the important features for industrial fermentation is the stability of the phage resistance phenotype. Thus, we verified the stability of the mutation MetAP^{H206Q} over 60 generations of *S. thermophilus* SMQ-301 and DGCC7796 as well as their mutants in milk and LM17. While all selected colonies of the wild-type strains remained phage sensitive, all 10 colonies from each mutant remained phage resistant, indicating that the MetAP^{H206Q} mutation is stable.

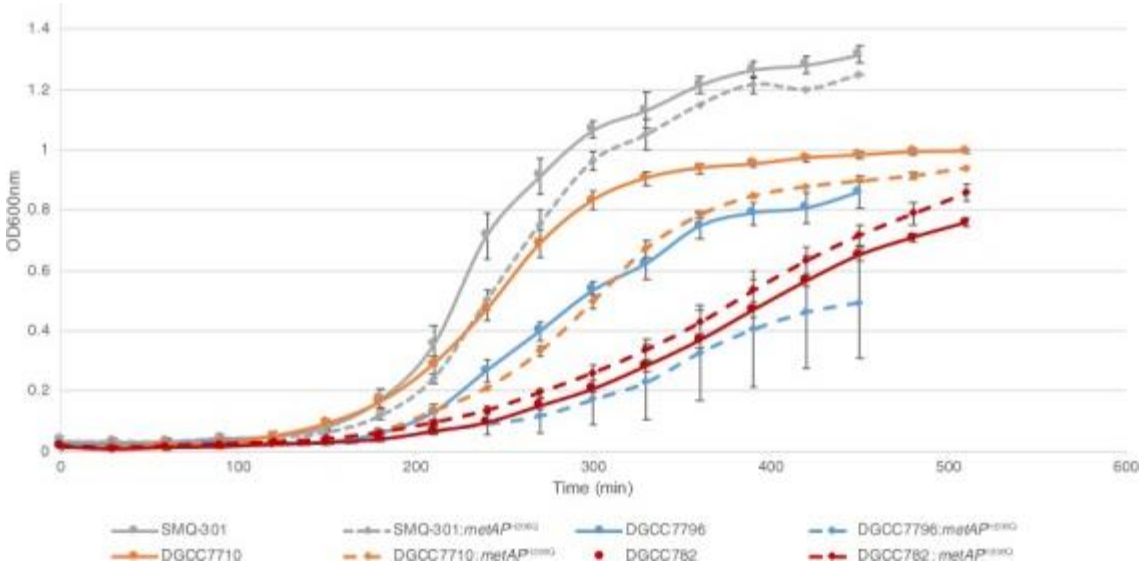


Figure 21: Growth curves of the wildtype and mutant *S. thermophilus* strains.

Table 15: Generation time of *S. thermophilus* wild-type and mutant strains.

***S. thermophilus* strain**	**Generation time (min)**
SMQ-301	36.7 ± 1.8
SMQ-301:metAP^{H206Q}	38.1 ± 1.1
DGCC7796	36.2 ± 3.6
DGCC7796:metAP^{H206Q}	53.5 ± 3.3
DGCC7710	38.1 ± 2.1
DGCC7710:metAP^{H206Q}	38.6 ± 1.7
DGCC782	49.7 ± 1.8
DGCC782:metAP^{H206Q}	53.4 ± 4.4

Conclusion

While cell wall proteins and polysaccharides were identified as host factors needed for phage infection, the literature is sparse about cytoplasmic proteins. MetAP are cytoplasmic enzymes found in all living organisms(Giglione et al., 2004). They are a unique class of proteases that remove the N-terminal residue from nascent proteins and play a central role in the synthesis and maturation of proteins in prokaryotes and eukaryotes (Giglione et al., 2015). The essentiality of this post-translational processing is underscored by the lethality of *metAP* gene inactivation in *E. coli* (Chang et al., 1989), *Salmonella typhimurium* (Miller et al., 1989), and *Saccharomyces cerevisiae* (Zuo et al., 1995). This dependence now extends to some of their parasites.

In this study, we showed that phage resistance can be acquired through mutations in the *metAP* gene of *S. thermophilus* and *S. mutans*. We were unable to isolate phage mutants that overcame MetAP^{H206Q} mutation, confirming that such mutation can provide a robust phage resistance. Early steps (adsorption, DNA replication, protein expression) of the phage lytic cycle occurred in phage-infected *S. thermophilus* SMQ-301:metAP^{H206Q}, suggesting that this mutation acts lately during the lytic cycle, perhaps at the virion assembly step. Since the catalytic activity of MetAP is at the core of living cells, we expected that the MetAP^{H206Q} mutation provided a broad phage resistance. Although the replication of *cos*-type phages was affected by MetAP mutation, the *pac*-type phages were not. As the enzymatic activity of MetAP:H206Q was less effective than the wild-type MetAP, it appears that *cos*-type phages require efficient cleavage of the N-terminal methionine of some of their proteins and/or of their host to complete their lytic cycle.

Acknowledgements

We would like to thank Barbara-Ann Conway (Medical Writer & Editor) for editorial assistance. S.M. acknowledges funding from the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Program. S.M. holds a Tier 1 Canada Research Chair in Bacteriophages.

Author contributions

S.M. headed the project. S.J.L. and S.M. wrote the manuscript and designed the project. S.J.L. executed the experiments. M.E.D. isolated the natural non-CRISPR BIM, C.M. conducted the experiments with *S. mutans* and S.L. executed the adsorption tests and growth curves. G.M.R. performed the mutation stability test and growth curves. D.T. sequenced the genomes. P.D. assisted S.J.L. for the MAP experiments. D.A.R., P.H. and C.F. helped with the design of the project. All the authors commented on the manuscript. S.J.L., D.A.R., P.H., C.F. and S.M. are co-inventors on patent applications.

References

Achigar, R., Magadán, A.H., Tremblay, D.M., Julia Pianzzola, M., Moineau, S., 2017. Phage-host interactions in *Streptococcus thermophilus*: Genome analysis of phages isolated in Uruguay and ectopic spacer acquisition in CRISPR array. Sci. Rep. 7, 43438.

Ali, Y., Koberg, S., Heßner, S., Sun, X., Rabe, B., Back, A., Neve, H., Heller, K.J., 2014. Temperate *Streptococcus thermophilus* phages expressing superinfection exclusion proteins of the Ltp type. Front. Microbiol. 5, 98.

Arya, T., Kishor, C., Saddanapu, V., Reddi, R., Addlagatta, A., 2013. Discovery of a new genetic variant of methionine aminopeptidase from streptococci with possible post-translational modifications: biochemical and structural characterization. PLoS One 8, e75207.

Avrani, S., Lindell, D., 2015. Convergent evolution toward an improved growth rate and a reduced resistance range in *Prochlorococcus* strains resistant to phage. Proc. Natl. Acad. Sci. U. S. A. 112, E2191-2200.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709-1712.

Bissonnette, F., Labrie, S., Deveau, H., Lamoureux, M., Moineau, S., 2000. Characterization of mesophilic mixed starter cultures used for the manufacture of aged cheddar cheese. J. Dairy Sci. 83, 620-627.

Boucher, I., Émond, E., Dion, E., Montpetit, D., Moineau, S., 2000. Microbiological and molecular impacts of AkiK on the lytic cycle of *Lactococcus lactis* phages of the 936 and P335 species. Microbiology 146, 445-453.

Buckling, A., Brockhurst, M., 2012. Bacteria-virus coevolution. Adv. Exp. Med. Biol. 751, 347-370.

Chang, S.Y.P., McGary, E.C., Chang, S., 1989. Methionine aminopeptidase gene of *Escherichia coli* is essential for cell growth. J. Bacteriol. 171, 4071-4072.

Cobián Güemes, A.G., Youle, M., Cantú, V.A., Felts, B., Nulton, J., Rohwer, F., 2016. Viruses as winners in the game of life. Annu. Rev. Virol. 3, 197-214.

Copik, A.J., Swierczek, S.I., Lowther, W.T., D'Souza, V.M., Matthews, B.W., Holz, R.C.,

2003. Kinetic and spectroscopic characterization of the H178A methionyl aminopeptidase from *Escherichia coli*. Biochemistry 42, 6283-6292.

Da Silva Duarte, V., Giaretta, S., Campanaro, S., Treu, L., Armani, A., Tarrah, A., Oliveira de Paula, S., Giacomini, A., Corich, V., 2019. A cryptic non-inducible prophage confers phage-immunity on the *Streptococcus thermophilus* M17PTZA496. Viruses 11, 7.

De Vos, W.M., 1987. Gene cloning and expression in lactic streptococci. FEMS Microbiol. Lett. 46, 281-295.

Delisle, A., Guo, M., Chalmers, N., Barcak, G., Rousseau, G., Moineau, S., 2012. Biology and genome sequence of *Streptococcus mutans* phage M102AD. Appl. Environ. Microbiol. 78, 2264-2271.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., Moineau, S., 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J. Bacteriol. 190, 1390-1400.

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., Sorek, R., 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359, 6379.

Dufour, D., Cordova, M., Cvitkovitch, D.G., Lévesque, C.M., 2011. Regulation of the competence pathway as a novel role associated with a streptococcal bacteriocin. J. Bacteriol. 193, 6552-6559.

Duplessis, M., Moineau, S., 2001. Identification of a genetic determinant responsible for host specificity in *Streptococcus thermophilus* bacteriophages. Mol. Microbiol. 41, 325-336.

Fontaine, L., Dandoy, D., Boutry, C., Delplace, B., De Frahan, M.H., Fremaux, C., Horvath, P., Boyaval, P., Hols, P., 2010. Development of a versatile procedure based on natural transformation for marker-free targeted genetic modification in *Streptococcus thermophilus*. Appl. Environ. Microbiol. 76, 7870-7877.

Gardan, R., Besset, C., Guillot, A., Gitton, C., Monnet, V., 2009. The oligopeptide transport system is essential for the development of natural competence in *Streptococcus thermophilus* strain LMD-9. J. Bacteriol. 191, 4647-4655.

Garneau, J.E., Dupuis, M.È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P.,

Fremaux, C., Horvath, P., Magadán, A.H., Moineau, S., 2010. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67-71.

Garvey, P., Hill, C., Fitzgerald, G.F., 1996. The lactococcal plasmid pNP40 encodes a third bacteriophage resistance mechanism, one which affects phage DNA penetration. Appl. Environ. Microbiol. 62, 676-679.

Gibson, D.G., Young, L., Chuang, R., Venter, J.C., Hutchison, C., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods. 6, 343-345.

Giglione, C., Boularot, A., Meinnel, T., 2004. Protein N-terminal methionine excision. Cell. Mol. Life Sci. 61, 1455-1474.

Giglione, C., Fieulaine, S., Meinnel, T., 2015. N-terminal protein modifications: Bringing back into play the ribosome. Biochimie 114, 134-146.

Hirel, P.H., Schmitter, J.M., Dessen, P., Fayat, G., Blanquet, S., 1989. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. Proc. Natl. Acad. Sci. U. S. A. 86, 8247-8251.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., Barrangou, R., 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J. Bacteriol. 190, 1401-1412.

Hynes, Alexander P., Lemay, M.L., Trudel, L., Deveau, H., Frenette, M., Tremblay, D.M., Moineau, S., 2017. Detecting natural adaptation of the *Streptococcus thermophilus* CRISPR-Cas systems in research and classroom settings. Nat. Protoc. 12, 547-565.

Hynes, A.P., Rousseau, G.M., Lemay, M.-L., Horvath, P., Romero, D.A., Fremaux, C., Moineau, S., 2017. An anti-CRISPR from a virulent streptococcal phage inhibits *Streptococcus pyogenes* Cas9. Nat Microbiol. 2, 1374-1380.

Hynes, A.P., Rousseau, G.M., Agudelo, D., Goulet, A., Amigues, B., Loehr, J., Romero, D.A., Fremaux, C., Horvath, P., Doyon, Y., Cambillau, C., Moineau, S., 2018. Widespread anti-CRISPR proteins in virulent phages inhibit a range of Cas9 proteins. Nat. Commun. 9, 2919.

Koskella, B., Brockhurst, M.A., 2014. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. FEMS Microbiol. Rev. 38, 916-931.

Kronheim, S., Daniel-Ivad, M., Duan, Z., Hwang, S., Wong, A.I., Mantel, I., Nodwell, J.R., Maxwell, K.L., 2018. A chemical defence against phage infection. Nature 564, 283-286.

Labrie, S.J., Samson, J.E., Moineau, S., 2010. Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8, 317-327.

Labrie, S.J., Tremblay, D.M., Plante, P.L., Wasserscheid, J., Dewar, K., Corbeil, J., Moineau, S., 2015. Complete genome sequence of *Streptococcus thermophilus* SMQ-301, a model strain for phage-host interactions. Genome Announc. 3, e00480-15.

Lavelle, K., Martinez, I., Neve, H., Lugli, G.A., Franz, C.M.A.P., Ventura, M., Dal Bello, F., van Sinderen, D., Mahony, J., 2018a. Biodiversity of *Streptococcus thermophilus* phages in global dairy fermentations. Viruses 10, 577.

Lavelle, K., Murphy, J., Fitzgerald, B., Lugli, G., Zomer, A., Neve, H., Ventura, M., Franz, C., Cambillau, C., van Sinderen, D., Mahony, J., 2018b. A decade of *Streptococcus thermophilus* phage evolution in an Irish dairy plant. Appl. Environ. Microbiol. 84, e02855.

Le Marrec, C., Van Sinderen, D., Walsh, L., Stanley, E., Vlegels, E., Moineau, S., Heinze, P., Fitzgerald, G., Fayard, B., 1997. Two groups of bacteriophages infecting *Streptococcus thermophilus* can be distinguished on the basis of mode of packaging and genetic determinants for major structural proteins. Appl. Environ. Microbiol. 63, 3246-3253.

Lennon, J.T., Khatana, S.A.M., Marston, M.F., Martiny, J.B.H., 2007. Is there a cost of virus resistance in marine cyanobacteria? ISME J. 1, 300-312.

Lévesque, C., Duplessis, M., Labonté, J., Labrie, S., Fremaux, C., Tremblay, D., Moineau, S., 2005. Genomic organization and molecular analysis of virulent bacteriophage 2972 infecting an exopolysaccharide-producing *Streptococcus thermophilus* strain. Appl. Environ. Microbiol. 71, 4057-4068.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

Lowther, W.T., Orville, A., Madden, D., Lim, S., Rich, D.H., Matthews, B.W., 1999. *Escherichia coli* methionine aminopeptidase: Implications of crystallographic analyses

of the native, mutant, and inhibited enzymes for the mechanism of catalysis. Biochemistry 38, 7678-7688.

Magadán, A.H., Dupuis, M.È., Villion, M., Moineau, S., 2012. Cleavage of phage DNA by the *Streptococcus thermophilus* CRISPR3-Cas system. PLoS One 7,e40913.

McDonnell, B., Mahony, J., Hanemaaijer, L., Kouwen, T.R.H.M., van Sinderen, D., 2018. Generation of bacteriophage-insensitive mutants of *Streptococcus thermophilus* via an antisense RNA CRISPR-Cas silencing approach. Appl. Environ. Microbiol. 84, e01733-17.

McDonnell, B., Mahony, J., Hanemaaijer, L., Neve, H., Noben, J.P., Lugli, G.A., Ventura, M., Kouwen, T.R., van Sinderen, D., 2017. Global survey and genome exploration of bacteriophages infecting the lactic acid bacterium *Streptococcus thermophilus*. Front. Microbiol. 8, 1754.

McDonnell, B., Mahony, J., Neve, H., Hanemaaijer, L., Noben, J.P., Kouwen, T., van Sinderen, D., 2016. Identification and analysis of a novel group of bacteriophages infecting the lactic acid bacterium *Streptococcus thermophilus*. Appl. Environ. Microbiol. 82, 5153-5165.

Miller, C.G., Kukral, A.M., Miller, J.L., Movva, N.R., 1989. pepM is an essential gene in *Salmonella typhimurium*. J. Bacteriol. 171, 5215-5217.

Mills, S., Griffin, C., O'Sullivan, O., Coffey, A., McAuliffe, O.E., Meijer, W.C., Serrano, L.M., Ross, R.P., 2011. A new phage on the "Mozzarella" block: Bacteriophage 5093 shares a low level of homology with other *Streptococcus thermophilus* phages. Int. Dairy J. 21, 963-969.

Samson, J.E., Magadán, A.H., Sabri, M., Moineau, S., 2013. Revenge of the phages: Defeating bacterial defences. Nat. Rev. Microbiol. 11, 675-687.

Scaltrlti, E., Tegoni, M., Rivetti, C., Launay, H., Masson, J.Y., Magadan, A.H., Tremblay, D., Moineau, S., Ramoni, R., Lichlère, J., Campanacci, V., Cambillau, C., Ortiz-Lombardia, M., 2009. Structure and function of phage p2 0RF34P2, a new type of single-stranded DNA binding protein. Mol. Microbiol. 73, 1156-1170.

Sun, X., Göhler, A., Heller, K.J., Neve, H., 2006. The *ltp* gene of temperate *Streptococcus thermophilus* phage TP-J34 confers superinfection exclusion to *Streptococcus thermophilus* and *Lactococcus lactis*. Virology 350, 146-157.

Szymczak, P., Filipe, S.R., Covas, G., Vogensen, F.K., Neves, A.R., Janzen, T., 2018. Cell wall glycans mediate recognition of the dairy bacterium *Streptococcus thermophilus* by bacteriophages. Appl. Environ. Microbiol. 84, e01847-18.

Tremblay, D.M., Moineau, S., 1999. Complete genomic sequence of the lytic bacteriophage DT1 of *Streptococcus thermophilus*. Virology 255, 63-76.

Vale, P.F., Lafforgue, G., Gatchitch, F., Gardan, R., Moineau, S., Gandon, S. 2015. Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. Proc. Biol. Sci. 282, 20151270.

Xiao, Q., Zhang, F., Nacev, B.A., Liu, J.O., Pei, D., 2010. Protein N-terminal processing: Substrate specificity of *Escherichia coli* and human methionine aminopeptidases. Biochemistry 49, 5588-5599.

Zuo, S., Guo, Q., Ling, C., Chang, Y.H., 1995. Evidence that two zinc fingers in the methionine aminopeptidase from *Saccharomyces cerevisiae* are important for normal growth.Mol. Gen. Genet. 246, 247-253.

Annex B: Review

A short overview of the CRISPR-Cas adaptation stage

Cas Mosterd[1,2], Geneviève M. Rousseau[1,2] and Sylvain Moineau[1,2,3]*

[1] Département de biochimie, de microbiologie et de bio-informatique, Faculté des sciences et de génie, Université Laval, Québec, Canada, G1V 0A6

[2] Groupe de recherche en écologie buccale, Faculté de médecine dentaire, Université Laval, Québec, Canada, G1V 0A6

[3] Félix d'Hérelle Reference Center for Bacterial Viruses, Faculté de médecine dentaire, Université Laval, Québec, Canada, G1V 0A6

Correspondence should be addressed to:
Tel: +1 418 656 3712; Email: Sylvain.Moineau@bcm.ulaval.ca

Canadian Journal of Microbiology 2020. doi: 10.1139/cjm-2020-0212

Keywords: CRISPR-Cas, bacteriophage, plasmid, antiviral, defence, resistance

Résumé

La recherche sur CRISPR a débuté il y a plus de 30 ans avec la découverte fortuite d'un arrangement inhabituel de nucléotides dans le génome d'*Escherichia coli*. Il a fallu 20 ans pour découvrir la fonction principale des systèmes CRISPR-Cas comme mécanisme de défense adaptative contre des acides nucléiques envahisseurs et notre connaissance de leur biologie s'est accrue régulièrement depuis. En parallèle, le nombre d'applications dérivées des systèmes CRISPR-Cas a augmenté de manière spectaculaire. L'outil d'édition du génome est sans doute l'application la plus passionnante tant en recherche fondamentale qu'en recherche appliquée. Récemment, la recherche sur CRISPR-Cas s'est partiellement orientée vers l'aspect le moins connu de sa biologie : la capacité des systèmes CRISPR-Cas à acquérir de nouvelles immunités durant l'étape dite d'adaptation. À ce jour, le système naturel le plus efficace pour acquérir facilement de nouveaux espaceurs est le système de type II-A de la bactérie à Gram positif *Streptococcus thermophilus.* La découverte de systèmes additionnels capables d'acquérir de nouveaux espaceurs devrait attirer l'attention sur cette étape de la biologie de CRISPR-Cas. Cette revue se concentre sur les récentes percées sur l'adaptation et sur les questions à résoudre.

Abstract

CRISPR research began over 30 years ago with the incidental discovery of an unusual nucleotide arrangement in the *Escherichia coli* genome. It took twenty years to find the main function of CRISPR-Cas systems as an adaptive defence mechanism against invading nucleic acids and our knowledge of their biology has steadily increased ever since. In parallel, the number of applications derived from CRISPR-Cas systems has risen spectacularly. The CRISPR-based genome editing tool is arguably the most exciting application in both basic and applied research. Lately, CRISPR-Cas research has partially shifted to the least understood aspect of its biology: the ability of CRISPR-Cas systems to acquire new immunity during the so-called adaptation step. To date, the most efficient natural system to readily acquire new spacers is the type II-A system of the gram-positive dairy bacterium *Streptococcus thermophilus*. The discovery of additional systems able to acquire new spacers will hopefully draw more attention to this step of CRISPR-Cas

biology. This review focuses on the breakthroughs that have helped to unravel the adaptation phase and on questions that remain to be answered.

Abbreviations

aca	anti-CRISPR associated
acr	anti-CRISPR
BIM	bacteriophage-insensitive mutant
bp	base pair
cas	CRISPR-associated
Cascade	CRISPR-associated complex for antiviral defence
CEM	CRISPR escaping mutant
Chi	Crossover hotspot instigator
Cmr	CRISPR RAMP module
CRISPR	clustered regularly interspaced short palindromic repeats
cpf	CRISPR-Cas substype as in *Prevotella* and *Francisella*
crRNA	CRISPR-RNA
csf	CRISPR-Cas subtype as in *Acidithiobacillus ferrooxidans*
DNA	deoxyribonucleic acid
DSBs	double stranded DNA breaks
dsDNA	double stranded DNA
iap	isozyme alkaline phosphatase
PAM	protospacer adjacent motif
RNA	ribonucleic acid
ssDNA	single stranded DNA
ssRNA	single stranded RNA
tracrRNA	trans-activating crRNA

Introduction

Clustered regularly interspaced short palindromic repeats (CRISPRs) were first described in bacteria in 1987, during the analysis of the nucleotide sequence of the *iap* gene responsible for the isozyme conversion of alkaline phosphatase (and its flanking region) in *Escherichia coli*. A series of five repeated palindromic nucleotide sequences flanking the 3' end of the *iap* gene were separated by variable sequences of uniform length, with no homology to each other (Ishino et al., 1987). Six years later, another research group described a similar nucleotide sequence pattern in *Archaea* (Mojica et al., 1993). During these early days, multiple groups studying these mysterious repeats used different names and acronyms, resulting in confusion within the field. In 2002, the term CRISPR was introduced and the first CRISPR-associated (*cas*) genes were described (Jansen et al., 2002). These genes were found to flank CRISPR loci and were often absent in CRISPR-negative strains. The importance of the variable sequences, named spacers, between the repeats remained a mystery until 2005. Homologies were discovered between some spacers within a few CRISPR loci, and virus and plasmid sequences (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). Several hypotheses were proposed as the mode of action, one of them being that CRISPRs may play a role in the defence against foreign genetic material (Mojica et al., 2005; Pourcel et al., 2005). In 2006, bioinformatics analyses of various CRISPR loci and *cas* genes suggested the cooperation between CRISPRs and Cas proteins in defence against plasmids and phages, by using a mechanism comparable to eukaryotic RNA interference systems (Makarova et al., 2006).

The first experimental demonstration of the function of CRISPR-Cas was published in 2007 in the dairy bacterial species *Streptococcus thermophilus*. Upon exposure to virulent (also known as lytic) streptococcal phages, the vast majority of the phage-infected cells died but a small number of cells survived phage infection. Analysis of the CRISPR array of the surviving bacterial cells, revealed that a new spacer was integrated into the CRISPR locus along with the addition of a repeat (Barrangou et al., 2007). The Eureka moment soon followed when it was realised that the new spacers originated from the genome of the virulent phage used in the assay. Later, the exact same sequence present in the phage genome was termed a "protospacer" to differentiate it from the bacterial spacer

found in the array (Deveau et al., 2008). Integrating a novel spacer identical to a protospacer resulted in resistance towards the phage carrying this sequence. The removal of the spacer sequence restored the phage-sensitive phenotype (Barrangou et al., 2007). Interference activity against plasmids was demonstrated in 2008 (Marraffini and Sontheimer, 2008). It was then shown in *E. coli*, that the CRISPR locus is transcribed and later processed by Cas proteins to serve as small guide RNAs that allow a complex of Cas proteins and CRISPR-RNA (crRNA) to interfere with an invasion by recognised foreign DNA (Brouns, 2008). In 2010, the CRISPR-Cas system of *S. thermophilus* was shown to precisely cleave target phage and plasmid DNA within the protospacer sequence *in vivo*. This provided a clear explanation for the phage resistance and plasmids interference (Garneau et al., 2010). Although multiple Cas proteins are involved in the interference step in *E. coli*, Cas9 is the only Cas protein required to cut the target DNA in *S. thermophilus* (Garneau et al., 2010; Gasiunas et al., 2012) and *Streptococcus pyogenes* (Jinek et al., 2012). This paved the way for the development of the CRISPR-Cas9-based genome editing tool (Jinek et al., 2012).

General principle of CRISPR-Cas function

CRISPR-Cas systems are divided into two classes, as well as in six types and several subtypes. The differences between the CRISPR-Cas systems are based on the composition of the *cas* genes (Koonin et al., 2017). Differences in repeat sequences and number of spacer-repeat units also occur between CRISPR-Cas systems of the same subtype and within the same species, respectively (Díez-Villaseñor et al., 2010; Horvath et al., 2008). Although on a molecular level, the mechanisms used by CRISPR-Cas systems inside the cell differ depending on the type of system, the general principle consists of three steps that all systems appear to follow.

The first step is the adaptation phase, during which a novel spacer is acquired from an invading nucleic acid sequence. In some types of CRISPR-Cas systems, adaptation occurs only when a sequence motif flanking the protospacer, named the protospacer adjacent motif (PAM), is recognised and the protospacer is inserted into the CRISPR locus (Deveau et al., 2008; Horvath et al., 2008). The duplication of a repeat sequence indicates

that the expansion of the CRISPR locus by one novel repeat-spacer unit is completed. Cas1 and Cas2 are conserved within almost all CRISPR-Cas types and form a complex involved in the adaptation step (Nuñez et al., 2014). The adaptation phase is followed by the expression phase, during which the CRISPR locus is transcribed into one long strand of pre-crRNA while additional *cas* genes are expressed. Some of these expressed Cas proteins are involved in the processing of pre-crRNA into mature short crRNAs (Charpentier et al., 2015). These crRNAs then form a complex with one or more Cas proteins. Finally, the third and final step is called the interference phase. Upon the subsequent invasion by nucleic acids harbouring a protospacer, the sequence is recognised and cleaved by the crRNA-Cas complexes (Brouns, 2008). Whereas the Cas1 and Cas2 proteins of the adaptation phase are ubiquitous among the different system types, there is more diversity in the Cas proteins involved in interference (Koonin et al., 2017). Proteins involved in the interference phase can also sometimes play a role in the initial adaptation step (Heler et al., 2015; Wei et al., 2015b).

Classification

CRISPR-Cas systems have been found in approximately 40% of the bacterial genomes (Makarova et al., 2019) and their classification is being re-evaluated and updated on a regular basis (Koonin et al., 2017; Makarova et al., 2019, 2018, 2015, 2011). As indicated above, two classes have been defined. Systems that belong to class 1 use complexes made up of multiple Cas proteins that mediate interference, whereas class 2 systems use one single Cas multidomain effector protein to cut nucleic acids. An overview of the different classes, types and subtypes is presented in Figure 22.

Class 1 CRISPR-Cas systems comprise types I, III and IV, with each type having its unique signature gene. In type I, the gene is *cas3*, which encodes a protein with both helicase and nuclease domains (Brouns, 2008; Sinkunas et al., 2011). For type III, the signature gene is *cas10*, a gene encoding a multidomain protein with nuclease activity involved in the interference step (Ramia et al., 2014). The *cas10* gene is missing in subtype III-E (Makarova et al., 2019). The proposed signature gene for type IV is *csf1* (Koonin et al., 2017; Makarova et al., 2015, 2018), although it is missing in subtype IV-C (Makarova

et al., 2019). Types I and III systems both make use of effector complexes consisting of several Cas proteins and crRNA. The effector complex of type I systems is the Cascade complex, whereas for type III systems, it is the Cmr complex. These effector complexes differ in protein composition and share little sequence identity. Nevertheless, the architecture and function of these two complexes are similar, since they evolved from a common ancestor (Jackson and Wiedenheft, 2015; Makarova et al., 2011). A major functional difference between the Cascade and Cmr complexes is that the Cascade complex targets DNA (Brouns et al., 2008; Sinkunas et al., 2011), while the Cmr complex target both RNA (Hale et al., 2009) and DNA (Samai et al., 2015). Type IV is different from the two other types in that it does not contain *cas1* and *cas2* genes. Moreover, its *cas* genes are usually not found close to a CRISPR locus, but the associated proteins are nevertheless predicted to form an effector complex (Koonin et al., 2017; Makarova et al., 2018, 2015).

The class 2 CRISPR-Cas systems include types II, V and VI. Although class 1 CRISPR-Cas systems are more common in bacteria, class 2 has received significantly more attention owing to its application in genome editing. Type II CRISPR-Cas systems, the first of the class 2 to be described, are the simplest in terms of the number of genes. The signature gene *cas9* encodes a protein that fulfils the roles of the multiple proteins found in the class 1 effector complexes. Cas9 contains two nuclease domains, RuvC and HNH, and is the only Cas protein required to cleave the target DNA (Gasiunas et al., 2012; Jinek et al., 2012). Besides Cas1 and Cas2, Cas9 has also been shown to play a role in the adaptation step (Heler et al., 2015; Wei et al., 2015b). In addition to the *cas* genes and CRISPR locus, trans-activating small RNAs (tracrRNAs) and the host RNaseIII are also essential components of type II systems required for the maturation of pre-crRNA into crRNAs (Chylinski et al., 2013; Deltcheva et al., 2011). Types V and VI were later additions to class 2. As in other systems, Cas1 and Cas2 are involved during adaptation in types V and VI systems (Makarova et al., 2015). Type V was proposed after the discovery of the *cas12* gene and it was predicted that Cas12 (also known as Cpf1) is the only protein involved in the interference stage (Strecker et al., 2019; Swarts and Jinek, 2019). Cas12 contains only a RuvC nuclease domain, thereby lacking the HNH domain (Swarts and Jinek, 2019; Zetsche et al., 2015). No tracrRNA is needed for the maturation of the crRNAs (Swarts and Jinek, 2019; Zetsche et al., 2015). The exceptions are subtypes V-F1 and V-G

which do require a tracrRNA. The *cas12* gene is not found in subtype V-U (Makarova et al., 2019). The discovery of *cas13* gave rise to type VI and Cas13 encodes a tracrRNA-independent nuclease, which targets RNA (Abudayyeh et al., 2016; Shmakov et al., 2015).

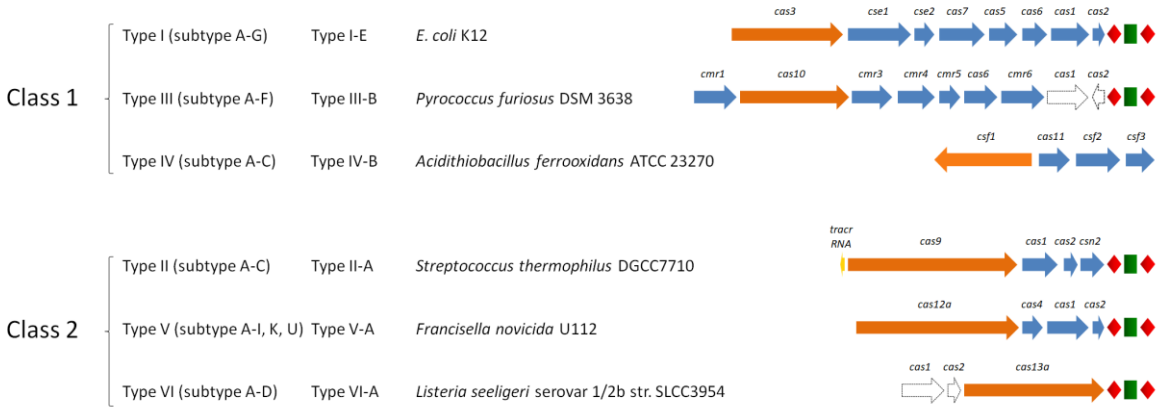


Figure 22: Classification of CRISPR-Cas systems. In this figure an example is given for each CRISPR system type. The signature genes are highlighted in orange. The other genes are blue. The type III-B and type VI-A systems do not possess their own *cas1* and *cas2* genes, but use those of other systems and these genes are therefore presented in white with dotted lines. The CRISPR locus is illustrated in the form of red diamonds (repeats) and green squares (spacers).

Adaptation

In recent years, steady progress has been made in our understanding of CRISPR-Cas systems with the interference phase attracting the most attention. Many molecular details remain poorly understood in the spacer acquisition process. During this phase, a stretch of DNA (prespacer) is first captured by Cas proteins (Wilkinson et al., 2019). Then, the Cas proteins and the bound prespacer are directed to the CRISPR locus, a repeat sequence is duplicated and the prespacer is integrated between the two repeats as a novel spacer. The *cas1* and *cas2* genes are essential and specific to the adaptation phase and both are relatively well conserved in almost all CRISPR-Cas systems (Brouns, 2008). Other universal elements of the acquisition process are the leader sequence (discussed below) and the first repeat of the CRISPR locus (Amitai and Sorek, 2016), suggesting that some processes are likely conserved in all types (Makarova et al., 2015).

Cas1 is a metal-dependent endonuclease that cleaves single-stranded DNA (ssDNA) (Babu et al., 2011), double-stranded DNA (dsDNA) (Wiedenheft et al., 2009), and single-stranded RNA (ssRNA) (Kim et al., 2013). Like Cas1, Cas2 is also a metal-dependent endonuclease. The Cas2 protein was first described as having activity against ssRNA (Beloglazova et al., 2008). However, depending on the CRISPR-Cas system, Cas2 may have activity against dsDNA (Nam et al., 2012). Cas1 and Cas2 form a complex consisting of a Cas2 dimer flanked by two Cas1 dimers ($Cas1_4$-$Cas2_2$). Within the complex, only the nuclease activity of Cas1 is essential for adaptation. Cas1 has a strong binding affinity for the CRISPR locus, as it is involved in the integration of new spacers. In the absence of Cas2, Cas1 is incapable of binding to the CRISPR locus, making both proteins essential during adaptation (Nuñez et al., 2014). Cas2 functions as an adaptor protein, bridging the Cas1 proteins as well as binding and stabilising the prespacer DNA (Wang et al., 2015).

In addition, a CRISPR repeat and a leader sequence are required for the integration of a novel spacer. The AT-rich leader sequence is typically located upstream of the CRISPR locus, which often contains the promoter that directs transcription of the CRISPR locus into pre-crRNA (Pougach et al., 2010; Pul et al., 2010). Repressor proteins silence transcription in certain species (Medina-Aparicio et al., 2011; Pougach et al., 2010; Pul et al., 2010), whereas in other species expression of the CRISPR operon can be constitutive and up-regulated during phage infection (Agari et al., 2010; Quax et al., 2013; Young et al., 2012). In addition to its function in transcription, the leader sequence has another role within the integration process as the integration of novel spacers usually occurs at the leader end of the CRISPR locus (Barrangou et al., 2007; Díez-Villaseñor et al., 2013; Yosef et al., 2012). Indeed, the leader sequence and the flanking repeat sequence harbour recognition signals with the ability to direct the Cas1-Cas2 complex to this position. These recognition sites are located within 10 bp of the integration site in both the leader and repeat sequences (Wei et al., 2015b).

The method by which sequences are integrated into the CRISPR locus suggests the probable acquisition of spacers that also target the bacterial chromosome (self-DNA) instead of foreign DNA. However, there appears to be a bias towards acquiring new spacers from foreign DNA compared with self-DNA. The number of PAM sequences present on

bacterial self-DNA is similar to that of foreign DNA and therefore, the reason cannot be related to PAM prevalence. It has been shown that one of the sources of protospacers for the Cas1-Cas2 complex are degradation products of the bacterial RecBCD complex, which processes double stranded DNA breaks (DSBs), although this does not apply to all the CRISPR-Cas system types. The large majority of these DSBs occur at replication forks during DNA replication and more are found on plasmids than on the bacterial chromosome. The RecBCD complex unwinds and degrades the DNA, starting at the DSB until it reaches a crossover hotspot instigator (Chi) site. Compared with plasmid and phage DNA, bacterial genomic DNA is rich in Chi sites and, therefore, is indirectly protected from spacer acquisition. Because phage DNA enters the cell as linear DNA, contains few Chi sites, and is highly replicative, it is presumably an easier target than chromosomal DNA for the RecBCD complex. This may explain why the CRISPR-Cas system preferentially acquires spacers from foreign DNA sources rather than its own (Levy et al., 2015). In type I systems, the degradation products from the activity of Cas3 during primed adaptation (see below) have also been demonstrated to function as prespacers (Künne et al., 2016; Staals et al., 2016). While the RecBCD complex is found only in Gram-negative bacteria, Gram-positive bacteria possess a highly similar mechanism in the form of the AddAB repair machinery. As with the RecBCD complex, the function of the AddAB machinery has been demonstrated to have a large impact on spacer acquisition (Modell et al., 2017).

Depending on the type of CRISPR-Cas system, additional factors are required to process the prespacers into spacers ready for integration. In type I-E systems, Cas2 is fused to a DnaQ domain, which degrades prespacers at the 3'-end to generate suitable spacers for integration (Drabavicius et al., 2018; Kim et al., 2020). A similar role has been proposed for Cas4, present in types I, II, and V systems. Cas4 cuts the 3' overhangs of prespacer DNA until a bound $Cas1_4$-$Cas2_2$ complex is encountered to generate spacers of specific length and a correct PAM. As such, Cas4 prevents the integration of non-functional spacers with an inappropriate length or PAM (Almendros et al., 2019; Kieper et al., 2018; Lee et al., 2019, 2018; Rollie et al., 2018; Shiimori et al., 2018). The function of Cas4 and DnaQ can also be performed by non-CRISPR nucleases in other systems (Yoganand et al., 2019).

Spacer acquisition under laboratory conditions

The large diversity of spacers found in strains possessing CRISPR loci indicates that spacer acquisition clearly occurs in numerous natural ecosystems. Surprisingly, spacer acquisition in bacterial strains has rarely been observed under laboratory conditions. The first case of spacer acquisition was observed with the dairy bacterium *S. thermophilus* and correlated with phage resistance (Barrangou et al., 2007) and plasmid interference (Garneau et al., 2010). Spacer acquisition was also demonstrated in *Streptococcus mutans* following phage infection, although in this species, other phage resistance mechanisms appear to be at play (Mosterd and Moineau, 2020; van der Ploeg, 2009). Strains of *S. thermophilus* and *S. mutans* possess type II-A CRISPR-Cas systems. Spacer acquisition against plasmids in a type II-C system has also been described in *Riemerella anatipestifer* (He et al., 2018). The type II-C system of *Campylobacter jejuni* was also able to acquire novel spacers into its CRISPR array, but only after infection by a phage expressing a Cas4-like protein (Hooton and Connerton, 2015).

Spacer acquisition in type I CRISPR-Cas systems has also been described (Datsenko et al., 2012; Li et al., 2014; Richter et al., 2014; Swarts et al., 2012). In the *E. coli* K12 type I-E system, spacer acquisition required the upregulation of *cas* genes or the deletion of a repressor. With other CRISPR-Cas systems, a fitness cost was associated with constitutive expression of the *cas* genes and to a lesser extent, during the acquisition of novel spacers, which may explain the presence of these repressors (Vale et al., 2015). The first evidence of spacer acquisition through priming was also observed in *E. coli*, although both primed and naïve spacer acquisition were observed (Datsenko et al., 2012; Swarts et al., 2012). Whereas naïve spacer acquisition refers to spacer acquisition from a virus not previously encountered, priming refers to spacer acquisition following the previous acquisition of a spacer from the same foreign DNA. Spacer acquisition was also observed in the CRISPR locus of the type I-B system of the archeon *Haloarcula hispanica*, although pre-existing spacers partially matching the infecting viral genome were required for adaptation (Li et al., 2014). These partially matching spacers did not provide resistance against the virus, but still increase spacer acquisition events during new encounters with the same virus. Similar results were obtained with the type I-F system of *Pectobacterium*

atrosepticum (Richter et al., 2014). Spacer acquisition was also observed in the archaea *Sulfolobus solfataricus* and *Sulfolobus islandicus*, although only during viral infection combined with the transformation of a plasmid (Erdmann and Garrett, 2012).

Growth conditions have been reported to be a major influence on the capacity of CRISPR-Cas to acquire novel spacers. *Pseudomonas aeruginosa* strains harbouring an active CRISPR-Cas predominantly became phage-resistant through surface receptor modification under nutrient-rich growth conditions, whereas under low-nutrient conditions derivatives became phage-resistant due to the acquisition of new spacers (Westra et al., 2015). CRISPR-Cas appears to provide a fitness advantage during growth in low-nutrient conditions (Chabas et al., 2016; Westra et al., 2015). The fitness disadvantage of surface receptor modification was further increased in a co-culture of multiple bacterial species, increasing the prevalence of CRISPR resistance (Alseth et al., 2019). A possible explanation for this is that the type IV pilus that functions as a phage receptor in *P. aeruginosa* has a role in the competition between species (Alseth et al., 2019; Westra et al., 2015).

While the acquisition of a single new spacer is the norm, the acquisition of multiple spacers has also been observed under laboratory conditions, particularly on solid media. Depending on the capacity of the first acquired spacer to protect the bacterium against an invading virus, a second spacer can be acquired to increase the overall phage resistance. It has been proposed that the acquisition of the second spacer is driven by priming during a single infection (Nussenzweig et al., 2019; Pyenson and Marraffini, 2020) or by the re-infection of the virus (Pyenson and Marraffini, 2020).

To identify systems that acquire novel spacers, tactics or factors that promote spacer acquisition can be employed. In the priming phenomenon described above the mismatches between the interference complex (crRNA) and the incoming phage genome do not allow for a strong interaction but still favour the Cas adaptation module to acquire additional spacers (Datsenko et al., 2012; Jackson et al., 2019; Swarts et al., 2012). Priming has been mainly associated with type I CRISPR-Cas systems. Evidence for primed acquisition has not been demonstrated in type II systems, nonetheless, it has been noted that upon

acquisition of a second spacer, this spacer is typically acquired from the vicinity of the first spacer (Nussenzweig et al., 2019). Bioinformatics analysis revealed that acquisition of newer spacers does not appear to be random but in the proximity of presumed older spacers (Nicholson et al., 2019). In *S. thermophilus*, it was demonstrated that prior phage exposure to UV light increased the number of spacer acquisition events. Specifically, the UV light generated defective phage particles that retained the capacity to inject their genome into a bacterial host but lost the capacity to generate infectious virions (Hynes et al., 2014). Presumably, it allows more time to select and integrate protospacers within the CRISPR locus, compared with an infection by an infective virulent phage where the adaptation module must be expressed and a spacer acquired before the virus lyses the cell (Hynes et al., 2014). A similar observation was made when the invading phage DNA is cleaved by restriction enzymes. Cleavage of phage DNA by restriction enzymes neutralises infection and provides prespacers for the CRISPR-Cas system (Dupuis et al., 2013).

Type II-A CRISPR-Cas system of *S. thermophilus* as a model for spacer acquisition

As indicated previously, the first observation of spacer acquisition under laboratory conditions was in *S. thermophilus*, specifically with the dairy industrial strain DGCC7710 (Barrangou et al., 2007). This unique strain is used worldwide in the production of yoghurt because of its production of exopolysaccharides, which leads to the desired organoleptic properties of the final fermented milk product. The dairy industry relies on DGCC7710 strains that are naturally resistant to phages; therefore, this resulted in CRISPR research with this specific bacterial strain.

In the initial experimental studies on CRISPR-Cas systems (Barrangou et al., 2007; Deveau et al., 2008; Garneau et al., 2010), *S. thermophilus* DGCC7710 was challenged with the virulent phages 2972 or 858 or both, and the CRISPR loci of the resulting bacteriophage-insensitive mutants (BIMs) were analysed (Barrangou et al., 2007; Deveau et al., 2008). Novel spacers were acquired from the phage genome in response to phage exposure and these new spacers came from the phage genome. These BIMs were resistant not only to the phages used in the assay but also to other phages that carry the identical

protospacer sequence. Furthermore, an overall stronger phage resistance phenotype was observed when multiple spacers were acquired from the same phage (Barrangou et al., 2007). Interestingly, some phages were also found to evade CRISPR immunity (CRISPR escaping mutants, or CEMs). Genomic analysis of CEMs revealed point mutations in protospacers or the flanking PAMs, which allowed the circumvention of the interference complexes. Subsequently, when exposed to CEMs, former BIMs were also able to establish renewed phage resistance by acquiring additional spacers (Deveau et al., 2008). This adaptive immunity is thus comparable to adding a new vaccine to your immunisation record and the spacer content of a bacterial strain can be seen as a molecular archive of past infections by foreign nucleic acids.

Despite its adaptability, any given bacterial strain can also carry multiple CRISPR-Cas systems and *S. thermophilus* is no exception. Four of these systems have been reported in this food-grade bacterium. CRISPR1 and CRISPR3 are both type II-A systems and have been experimentally shown to actively acquire novel spacers, with the large majority of the acquisition events taking place in the CRISPR1 locus (Horvath et al., 2008; Magadán et al., 2012). The two systems have the capacity to cooperate and increase the overall phage resistance within the same cell (Magadán et al., 2012). In addition, the CRISPR-Cas systems of *S. thermophilus* are compatible with other bacterial defence mechanisms, including restriction-modification systems (Dupuis et al., 2013). Furthermore, a type III-A system (CRISPR2) and a type I-E system (CRISPR4) can be found in some *S. thermophilus* strains (Magadán et al., 2012). No *in vivo* activity has been observed in the streptococcal type I-E system, but the isolated Cascade complex has been shown to interfere with targeted plasmid DNA (Sinkunas et al., 2013).

The two type II-A systems of *S. thermophilus* consist of four *cas* genes. In addition to the type II system signature gene *cas9* and the conserved *cas1* and *cas2* genes, a fourth gene called *csn2* distinguishes type II-A systems from other type II systems (Makarova et al., 2015). The *csn2* gene is essential for adaptation, although its precise role is unknown (Barrangou et al., 2007). Although the two type II-A systems have the same gene composition, there is a significant variation at the amino acid levels between the *cas* gene products. The percent identity is as low as 24.4% for Cas9, 32.5% for Cas1, 38.7% for

Cas2, and 21.8% for Csn2 (Figure 23). All four proteins and distinct tracrRNAs are needed for spacer acquisition, with Cas9 providing PAM specificity (Heler et al., 2015; Ka et al., 2018).

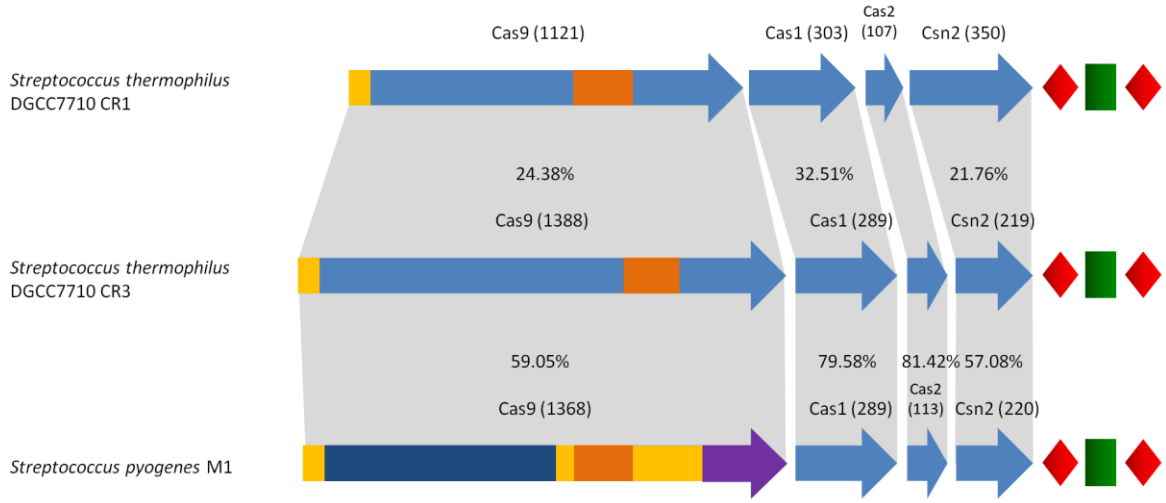


Figure 13: Comparison of Cas9 from *S. thermophilus* and *Streptococcus pyogenes*. The percent identity of the Cas proteins from CR1 of *S. thermophilus* DGCC7710 and *S. pyogenes* M1 compared to the Cas proteins of CR3 of *S. thermophilus* DGCC7710 are shown in the figure. Within Cas9, the RuvC domains are highlighted in yellow, the recognition lobe in marine blue, the HNH domains in orange and the PAM recognition domain in purple.

Structural analysis of the Cas proteins of the CRISPR3 type II-A system of *S. thermophilus* revealed a complex consisting of $Cas1_8$-$Cas2_4$-$Csn2_8$. Although not detected in the complex, Cas9 was hypothesised to interact with Csn2 through DNA (Wilkinson et al., 2019). Csn2 has been shown to bind dsDNA ends (Ellinger et al., 2012; Lee et al., 2012; Nam et al., 2012). It was proposed that the $Cas1_8$-$Cas2_4$-$Csn2_8$ complex binds the free ends of prespacers generated by the AddAB repair machinery through Csn2. Then, the complex slides down the DNA until it encounters Cas9 that is bound to a PAM, and then, the spacer may be cut from the DNA (Wilkinson et al., 2019). This process is highlighted in Figure 24. In the interference stage, Cas9 is the sole protein required in both CRISPR1 and CRISPR3 to cut the target DNA within the protospacer, three nucleotides upstream of the PAM (Garneau et al., 2010).

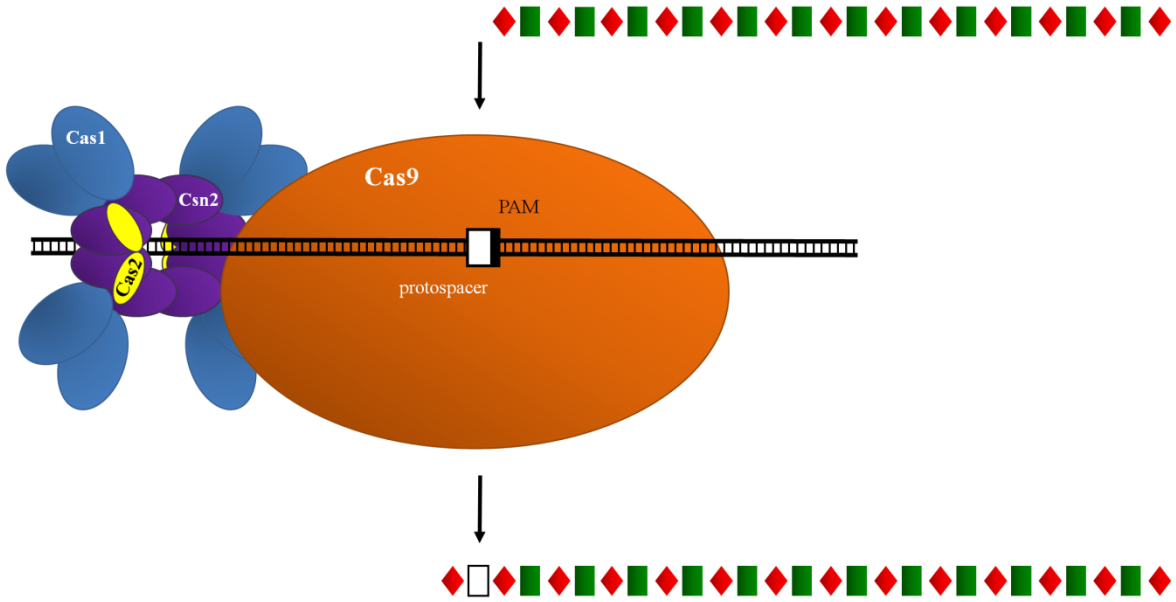


Figure 24: Spacer acquisition in CRISPR3 (type II-A) of *S. thermophilus*. The CRISPR locus of *S. thermophilus* DGCC7710 CRISPR3 consists of 13 repeats (red diamonds) and 12 spacers (green squares). The acquisition complex consists of 4 Cas1 dimers (blue), 2 Cas2 dimers (yellow) and 2 Csn2 tetramers (purple) and likely slides on the viral DNA until it encounters Cas9 (orange) bound to a PAM sequence (black). The flanking protospacer (white) is somehow cut from the viral DNA and is ready for integration into the CRISPR locus as a novel spacer.

Spacer acquisition events have typically been observed at the 5' end of the CRISPR locus, flanking the leader sequence. However, ectopic spacer acquisition has also been observed in *S. thermophilus* (Achigar et al., 2017). This phenomenon was first described in *S. pyogenes* and refers to the acquisition of spacers at positions other than the 5'-end (McGinn and Marraffini, 2016). The leader sequence in strains that acquired ectopic spacers was compared with that of strains that had not acquired ectopic spacers. A deletion of one nucleotide was discovered within the leader sequence directly upstream of the CRISPR locus in the strains that acquired ectopic spacers (Achigar et al., 2017). This region of the leader sequence is where the Cas proteins involved in adaptation interact with the CRISPR locus, likely explaining the reduced ability of these strains to integrate spacers at this position (Wei et al., 2015a). Furthermore, an AG motif was found at the 5' end of the spacer upstream of the newly acquired ectopic spacers. This AG motif may play a role in guiding the adaptation module to this specific position (Achigar et al., 2017).

Anti-CRISPR

Phages have evolved several ways to cope with a wide range of antiviral defence systems. This on-going evolutionary battle has been called an arms race as every evolutionary step, from either bacteria or phages, drives a new evolutionary step for the other (Safari et al., 2020). While escaping CRISPR immunity can be facilitated by a single mutation or a deletion within the protospacer or PAM sequence as mentioned previously (Deveau et al., 2008), it is more challenging to replicate in a diverse population of cells globally carrying a large number of different spacers. It has been hypothesised that this limitation led to the evolution of a more general method to disable the CRISPR immunity in the form of anti-CRISPR proteins (ACRs) (Van Houte et al., 2016).

The first anti-CRISPR genes were discovered in temperate phages infecting *P. aeruginosa* (Bondy-Denomy et al., 2013). While integrated into the *P. aeruginosa* genome as a prophage, the phage expressed proteins that interfered with the activity of the host CRISPR-Cas system. Since these phage proteins did not interfere with the transcription of *cas* genes and the CRISPR locus, the anti-CRISPR activity was found to take place after the formation of the crRNA-Cascade complex (Bondy-Denomy et al., 2013). The activity of these anti-CRISPR proteins was against type I-F systems only, but soon after, additional anti-CRISPR proteins that exhibited activity against type I-E systems were also discovered (Pawluk et al., 2014). At the sequence level, a signature motif that allowed for the identification of anti-CRISPR genes in other phages and archaeal viruses were found. This helix-turn-helix motif was found in proteins for which the genes were present directly downstream from anti-CRISPR genes. The anti-CRISPR associated (*aca*) genes, which code for transcription regulators, helped to identify new anti-CRISPR genes for other systems, including type II using bioinformatic approaches.

The first type II anti-CRISPR genes were found in prophages of *Neisseria meningitides*. They interfered with the nuclease activity of the type II-C Cas9 in *N. meningitides* but not with the type II-A Cas9 of *S. pyogenes* (Pawluk et al., 2016). Anti-CRISPR genes interfering with type II-A CRISPR-Cas systems were found using other approaches. One such approach was to screen for prophage sequences that carry a sequence

identical to one of the spacers present in the cell. This would typically result in self-targeting in the case of an active CRISPR-Cas system, but would be prevented by the expression of an ACR protein. Using this strategy, several new *acr* genes were discovered in prophages of *Listeria monocytogenes* that were capable of inhibiting SpCas9 (Rauch et al., 2017). The first anti-CRISPR genes discovered in virulent phages were found in *S. thermophilus*. These ACR-containing phages were discovered because phage-resistant mutants were extremely rare following some phage-bacteria interactions while some phages were able to infect BIMs carrying spacers targeting their genomes (Hynes et al., 2018, 2017). Additional anti-CRISPR proteins targeting type II systems were discovered by the transformation and subsequent expression of metagenomic libraries, after which the ACR activity against *S. pyogenes* Cas9 was confirmed *in vitro* and *in vivo* (Uribe et al., 2019).

Several mechanisms for anti-CRISPR proteins have also been discovered, revealing that different ACR proteins target different Cas proteins. For example, binding of the ACR to Cas proteins prevented binding or cleaving of nucleic acids (Bondy-Denomy et al., 2015; Davidson et al., 2020; Dong et al., 2017; Hardouin and Goulet, 2020; Harrington et al., 2017; Marino et al., 2020; Shin et al., 2017; Yang and Patel, 2017). A nomenclature for ACRs has been proposed and an online spreadsheet is available to avoid names being used twice in this fast growing field (Bondy-Denomy et al., 2018). The number of anti-CRISPR genes that has been discovered has expanded dramatically and so has the number of targeted system types. To date, over 50 families of anti-CRISPR proteins have been described that inhibit various types of CRISPR-Cas systems, and the search for new anti-CRISPR proteins still continues (Davidson et al., 2020; Stanley and Maxwell, 2018).

Conclusion

A few decades have passed since the incidental discovery of a CRISPR locus. Several research groups have uncovered many of its mysteries and it has turned out to be one of the most fascinating discoveries in microbiology. Currently, six CRISPR-Cas system types have been described as well as several subtypes (Makarova et al., 2019) and many additional variants likely await discovery. The development of the ever-growing CRISPR-Cas tools for genome editing was due, in part, to a better understanding of the crRNA

synthesis and interference steps. This CRISPR frenzy as a genome editing tool resulted in more research into these two steps of CRISPR-Cas biology. As such, studies on the adaptation phase have been limited for the most part. The experimental demonstration of spacer adaptation remains elusive for several CRISPR-Cas subtypes as it still appears to be a rare event under laboratory conditions.

While a very efficient crRNA synthesis or interference stage is needed to readily protect the cell against invasion of foreign nucleic acids, a spacer acquisition process in overdrive would be detrimental to the cell as self-targeting would likely occur, thereby killing the cells. The presence of *acr* genes in the bacterial genome (as prophages) or on plasmids are also likely responsible for these apparent experimental frustrations associated with spacer acquisition assays. The use of ACR-containing virulent phages in these experiments may also have played a role. The recent analysis of spacer acquisition in a bacterial community (Alseth et al., 2019) opens new avenues and could draw attention back to the adaptation phase.

Although CRISPR-Cas is considered a defence mechanism in a cellular context, other functions have been proposed and described (Faure et al., 2019). Under certain circumstances, CRISPR-Cas has been associated with gene regulation, genome evolution, DNA repair and cooperation with other antiviral systems (Westra et al., 2014). As the number of publications on CRISPR-Cas biology increases each year, novel functions and applications will likely be discovered in the future. It will be interesting to see if Cas proteins involved in the adaptation phase also play other functions. Cas1 was found to increase the resistance of *Mycobacterium smegmatis* towards multiple anti-tuberculosis drugs (Wei et al., 2019) and has also been associated with DNA repair (Babu et al., 2011). In *Legionella pneumophila*, Cas2 has been shown to be an essential factor to facilitate the infection of amoebae (Gunderson et al., 2015; Gunderson and Cianciotto, 2013). New functions for these versatile proteins are likely waiting to be discovered.

Acknowledgements

We thank Michi Waygood and Amanda Toperoff for editorial assistance. S.M. acknowledges funding from the Natural Sciences and Engineering Research Council of

Canada (NSERC) Discovery Program. S.M. holds a Tier 1 Canada Research Chair in Bacteriophages.

References

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E.S., Koonin, E. V., and Zhang, F. 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 353, aaf5573.

Achigar, R., Magadán, A.H., Tremblay, D.M., Julia Pianzzola, M., and Moineau, S. 2017. Phage-host interactions in *Streptococcus thermophilus*: genome analysis of phages isolated in Uruguay and ectopic spacer acquisition in CRISPR array. Sci. Rep. 7, 43438.

Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S., and Shinkai, A. 2010. Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. J. Mol. Biol. 395, 270-281.

Almendros, C., Nobrega, F.L., McKenzie, R.E., and Brouns, S.J.J. 2019. Cas4-Cas1 fusions drive efficient PAM selection and control CRISPR adaptation. Nucleic Acids Res. 47, 5223-5230.

Alseth, E.O., Pursey, E., Luján, A.M., McLeod, I., Rollie, C., Westra, E.R., 2019. Bacterial biodiversity drives the evolution of CRISPR-based phage resistance. Nature 574, 549-552.

Amitai, G., Sorek, R., 2016. CRISPR-Cas adaptation: Insights into the mechanism of action. Nat. Rev. Microbiol. 14, 67-76.

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., Phanse, S., Joachimiak, A., Koonin, E. V., Savchenko, A., Emili, A., Greenblatt, J., Edwards, A.M., and Yakunin, A.F. 2011. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol. Microbiol. 79, 484-502.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709-1712.

Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., Koonin, E. V.,

Edwards, A.M., Savchenko, A., Yakunin, A.F., 2008. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J. Biol. Chem. 283, 20361-20371.

Bolotin, A., Quinquis, B., Sorokin, A., Dusko Ehrlich, S., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551-2561.

Bondy-Denomy, J., Davidson, A.R., Doudna, J.A., Fineran, P.C., Maxwell, K.L., Moineau, S., Peng, X., Sontheimer, E.J., and Wiedenheft, B. 2018. A unified resource for tracking anti-CRISPR names. Crispr J. 1, 304-305.

Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M.F., Hidalgo-Reyes, Y., Wiedenheft, B., Maxwell, K.L., Davidson, A.R., 2015. Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. Nature 526, 136-139.

Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., Davidson, A.R., 2013. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature 493, 429-432.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J.,Makarova, K.S., Koonin, E.V., van der Oost J., 2008. Small CRISPR RNAs guide antiviraldefense in prokaryotes. Science 321, 960-964.

Chabas, H., Van Houte, S., Høyland-Kroghsbo, N.M., Buckling, A., Westra, E.R., 2016. Immigration of susceptible hosts triggers the evolution of alternative parasite defence strategies. Proc. R. Soc. B Biol. Sci. 283, 20160721.

Charpentier, E., Richter, H., van der Oost, J., White, M.F., 2015. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. FEMS Microbiol. Rev. 39, 428-441.

Chylinski, K., Le Rhun, A., Charpentier, E., 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol. 10, 726-737.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., Semenova, E., 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. 3, 945.

Davidson, A.R., Lu, W.-T., Stanley, S.Y., Wang, J., Mejdani, M., Trost, C.N., Hicks, B.T., Lee, J., and Sontheimer, E.J. 2020. Anti-CRISPRs: Protein inhibitors of CRISPR-Cas systems. Annu. Rev. Biochem. 89, 309-332.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., Charpentier, E., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-607.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J. Bacteriol. 190, 1390-1400.

Díez-Villaseñor, C., Almendros, C., García-Martínez, J., and Mojica, F.J.M. 2010. Diversity of CRISPR loci in *Escherichia coli*. Microbiology 156, 1351-1361.

Díez-Villaseñor, C., Guzmán, N.M., Almendros, C., García-Martínez, J., and Mojica, F.J.M. 2013. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. RNA Biol. 10, 792-802.

Dong, D., Guo, M., Wang, Sihan, Zhu, Y., Wang, Shuo, Xiong, Z., Yang, J., Xu, Z., Huang, Z., 2017. Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. Nature 546, 436-439.

Drabavicius, G., Sinkunas, T., Silanskas, A., Gasiunas, G., Venclovas, Č., Siksnys, V., 2018. DnaQ exonuclease-like domain of Cas2 promotes spacer integration in a type I-E CRISPR-Cas system. EMBO Rep. 19, e45543.

Dupuis, M.È., Villion, M., Magadán, A.H., Moineau, S., 2013. CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. Nat. Commun. 4, 2087.

Ellinger, P., Arslan, Z., Wurm, R., Tschapek, B., MacKenzie, C., Pfeffer, K., Panjikar, S., Wagner, R., Schmitt, L., Gohlke, H., Pul, Ü., Smits, S.H.J., 2012. The crystal structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*. J. Struct. Biol. 178, 350-362.

Erdmann, S., Garrett, R.A., 2012. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Mol. Microbiol. 85, 1044-1056.

Faure, G., Makarova, K.S., and Koonin, E. V. 2019. CRISPR–Cas: complex functional networks and multiple roles beyond adaptive immunity. J. Mol. Biol. 431, 3-20.

Garneau, J.E., Dupuis, M.È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P.,

Fremaux, C., Horvath, P., Magadán, A.H., Moineau, S., 2010. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67-71.

Gasiunas, Giedrius, Barrangou, R., Horvath, P., Siksnys, V., 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U. S. A. 109, E2579-2586.

Gasiunas, G., Barrangou, R., Horvath, P., Siksnys, V., 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. 109, E2579-E2586.

Gunderson, F.F., and Cianciotto, N.P. 2013. The CRISPR-associated gene cas2 of *Legionella pneumophila* is required for intracellular infection of amoebae. mBio. 4, e00074-13.

Gunderson, F.F., Mallama, C.A., Fairbairn, S.G., and Cianciotto, N.P. 2015. Nuclease activity of *Legionella pneumophila* Cas2 promotes intracellular infection of amoebal host cells. Infect. Immun. 83, 1008-1018.

Hale, C., Zhao, P., Olson, S., Duff, M., Graveley, B., Wells, L., Terns, R., and Terns, M.P. 2009. RNA-Guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 139, 945-956.

Hardouin, P., Goulet, A., 2020. Diversity of molecular mechanisms used by anti-CRISPR proteins: the tip of an iceberg? Biochem. Soc. Trans. 48, 507-516.

Harrington, L.B., Doxzen, K.W., Ma, E., Liu, J.J., Knott, G.J., Edraki, A., Garcia, B., Amrani, N., Chen, J.S., Cofsky, J.C., Kranzusch, P.J., Sontheimer, E.J., Davidson, A.R., Maxwell, K.L., and Doudna, J.A. 2017. A broad-spectrum inhibitor of CRISPR-Cas9. Cell 170, 1224-1233.e15.

He, Y., Wang, M., Liu, M., Huang, L., Liu, C., Zhang, X., Yi, H., Cheng, A., Zhu, D., Yang, Q., Wu, Y., Zhao, X., Chen, S., Jia, R., Zhang, S., Liu, Y., Yu, Y., and Zhang, L. 2018. Cas1 and Cas2 from the Type II-C CRISPR-Cas system of *Riemerella anatipestifer* are required for spacer acquisition. Front. Cell. Infect. Microbiol. 8, 195.

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., Marraffini, L.A., 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature 519, 199-202.

Hooton, S.P.T., Connerton, I.F., 2015. *Campylobacter jejuni* acquire new host-derived

CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. Front. Microbiol. 5, 744.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., Barrangou, R., 2008. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J. Bacteriol. 190, 1401-1412.

Hynes, A.P., Rousseau, G.M., Lemay, M.-L., Horvath, P., Romero, D.A., Fremaux, C., Moineau, S., 2017. An anti-CRISPR from a virulent streptococcal phage inhibits *Streptococcus pyogenes* Cas9. Nat. Microbiol. 2, 1374,1380.

Hynes, A.P., Rousseau, G.M., Agudelo, D., Goulet, A., Amigues, B., Loehr, J., Romero, D.A., Fremaux, C., Horvath, P., Doyon, Y., Cambillau, C., Moineau, S., 2018. Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9 proteins. Nat. Commun. 9, 2919.

Hynes, A.P., Villion, M., Moineau, S., 2014. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. Nat. Commun. 5, 4399.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., Nakata, A., 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J. Bacteriol. 169, 5429-5433.

Jackson, R.N., and Wiedenheft, B. 2015. A conserved structural chassis for mounting versatile CRISPR RNA-guided immune responses. Mol. Cell. 58, 722-728.

Jackson, S.A., Birkholz, N., Malone, L.M., and Fineran, P.C. 2019. Imprecise spacer acquisition generates CRISPR-Cas immune diversity through primed adaptation. Cell Host Microbe 25, 250-264.e4.

Jansen, R., Embden, J.D.A. van, Gaastra, W., Schouls, L.M., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. 43, 1565-1575.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821.

Ka, D., Jang, D.M., Han, B.W., Bae, E., 2018. Molecular organization of the type II-A CRISPR adaptation module and its interaction with Cas9 via Csn2. Nucleic Acids Res. 46, 9805-9815.

Kieper, S.N., Almendros, C., Behler, J., McKenzie, R.E., Nobrega, F.L., Haagsma, A.C.,

Vink, J.N.A., Hess, W.R., and Brouns, S.J.J. 2018. Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. Cell Rep. 22, 3377-3384

Kim, S., Loeff, L., Colombo, S., Jergic, S., Brouns, S.J.J., Joo, C., 2020. Selective loading and processing of prespacers for precise CRISPR adaptation. Nature 579, 141-145.

Kim, T.Y., Shin, M., Huynh Thi Yen, L., Kim, J.S., 2013. Crystal structure of Cas1 from *Archaeoglobus fulgidus* and characterization of its nucleolytic activity. Biochem. Biophys. Res. Commun. 441, 720-725.

Koonin, E. V., Makarova, K.S., Zhang, F., 2017. Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. 37, 67-78.

Künne, T., Kieper, S.N., Bannenberg, J.W., Vogel, A.I.M., Miellet, W.R., Klein, M., Depken, M., Suarez-Diez, M., and Brouns, S.J.J. 2016. Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. Mol. Cell 63, 852-864.

Lee, H., Dhingra, Y., Sashital, D.G., 2019. The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation. Elife 8, e44248.

Lee, H., Zhou, Y., Taylor, D.W., and Sashital, D.G. 2018. Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays. Mol. Cell. 70, 48-59.e5.

Lee, K.H., Lee, S.G., Eun Lee, K., Jeon, H., Robinson, H., Oh, B.H., 2012. Identification, structural, and biochemical characterization of a group of large Csn2 proteins involved in CRISPR-mediated bacterial immunity. Proteins Struct. Funct. Bioinforma. 80, 2573-2582.

Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., Sorek, R., 2015. CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature 520, 505-510.

Li, M., Wang, R., Zhao, D., Xiang, H., 2014. Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. Nucleic Acids Res. 42, 2483-2492.

Magadán, A.H., Dupuis, M.È., Villion, M., Moineau, S., 2012. Cleavage of phage DNA by the *Streptococcus thermophilus* CRISPR3-Cas system. PLoS One 7, e40913.

Makarova, K., Aravind, L., Wolf, Y., Koonin, E., 2011. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol.

Direct 6, 38.

Makarova, K.S., Grishin, N. V., Shabalina, S.A., Wolf, Y.I., Koonin, E. V., 2006. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct 1, 7.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.J., Charpentier, E., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Terns, R.M., Terns, M.P., White, M.F., Yakunin, A.F., Garrett, R.A., Van Der Oost, J., Backofen, R., Koonin, E. V., 2015. An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. 13, 722-736.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Scott, D., Shah, S.A., Siksnys, V., Terns, M.P., Venclovas, Č., White, M.F., Yakunin, A.F., Yan, W., Zhang, F., Garrett, R.A., Backofen, R., van der Oost, J., Barrangou, R., Koonin, E. V, 2019. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. 18, 67-83.

Makarova, K.S., Wolf, Y.I., and Koonin, E. V. 2018. Classification and nomenclature of CRISPR-Cas systems: where from here? Crispr. J. 1, 325-336.

Marino, N.D., Pinilla-Redondo, R., Csörgő, B., Bondy-Denomy, J., 2020. Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. Nat. Methods. 17, 471-479.

Marraffini, L.A., Sontheimer, E.J., 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322, 1843-1845.

McGinn, J., and Marraffini, L.A. 2016. CRISPR-Cas systems optimize their immune response by specifying the site of spacer integration. Mol. Cell 64, 616-623.

Medina-Aparicio, L., Rebollar-Flores, J.E., Gallego-Hernández, A.L., Vázquez, A., Olvera, L., Gutiérrez-Ríos, R.M., Calva, E., Hernández-Lucas, I., 2011. The CRISPR/Cas immune system is an operon regulated by LeuO, H-NS, and leucine-responsive regulatory protein in *Salmonella enterica* serovar Typhi. J. Bacteriol. 193, 2396-2407.

Modell, J.W., Jiang, W., Marraffini, L.A., 2017. CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. Nature 544, 101-104.

Mojica, F.J., Juez, G., Rodríguez-Valera, F., 1993. Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified *Pst*I sites. Mol. Microbiol. 9, 613-621.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., Soria, E., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 60, 174-182.

Mosterd, C., Moineau, S., 2020. Characterization of a Type II-A CRISPR-Cas System in *Streptococcus mutans*. mSphere 5, e00235-20.

Nam, K.H., Ding, F., Haitjema, C., Huang, Q., DeLisa, M.P., and Ke, A. 2012. Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. J. Biol. Chem. 287, 35943-35952.

Nicholson, T.J., Jackson, S.A., Croft, B.I., Staals, R.H.J., Fineran, P.C., Brown, C.M., 2019. Bioinformatic evidence of widespread priming in type I and II CRISPR-Cas systems. RNA Biol. 16, 566-576.

Nuñez, J.K., Kranzusch, P.J., Noeske, J., Wright, A. V., Davies, C.W., Doudna, J.A., 2014. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. Nat. Struct. Mol. Biol. 21, 528-534.

Nussenzweig, P., McGinn, J., and Marraffini, L.A. 2019. Cas9 cleavage of viral genomes primes the acquisition of new immunological memories. Cell Host Microbe. 26, 515-526.e6.

Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheimer, E.J., Maxwell, K.L., and Davidson, A.R. 2016. Naturally occurring off-switches for CRISPR-Cas9. Cell 167, 1829-1838.e9.

Pawluk, A., Bondy-Denomy, J., Cheung, V.H.W., Maxwell, K.L., and Davidson, A.R. 2014. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. mBio 5, e00896.

Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L., Severinov, K., 2010. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. Mol. Microbiol. 77, 1367-1379.

Pourcel, C., Salvignol, G., Vergnaud, G., 2005. CRISPR elements in *Yersinia pestis* acquire

new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151, 653-663.

Pul, Ü., Wurm, R., Arslan, Z., Geißen, R., Hofmann, N., Wagner, R., 2010. Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. Mol. Microbiol. 75, 1495-1512.

Pyenson, N.C., Marraffini, L.A., 2020. Co-evolution within structured bacterial communities results in multiple expansion of CRISPR loci and enhanced immunity. Elife 9, e53078.

Quax, T.E.F., Voet, M., Sismeiro, O., Dillies, M.-A., Jagla, B., Coppee, J.-Y., Sezonov, G., Forterre, P., van der Oost, J., Lavigne, R., and Prangishvili, D. 2013. Massive activation of archaeal defense genes during viral infection. J. Virol. 87, 8419-8428.

Ramia, N.F., Tang, L., Cocozaki, A.I., and Li, H. 2014. *Staphylococcus epidermidis* Csm1 is a 3′-5′ exonuclease. Nucleic Acids Res. 42, 1129-1138.

Rauch, B., Silvis, M., Hultquist, J., Waters, C., McGregor, M., Krogan, N., and Bondy-Denomy, J. 2017. Inhibition of CRISPR-Cas9 with bacteriophage proteins. Cell 168, 150-158.e10.

Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N.J., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H.J., Fineran, P.C., 2014. Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Res. 42, 8516-8526.

Rollie, C., Graham, S., Rouillon, C., White, M.F., 2018. NAR breakthrough article: Prespacer processing and specific integration in a type I-A CRISPR system. Nucleic Acids Res. 46, 1007-1020.

Safari, F., Sharifi, M., Farajnia, S., Akbari, B., Karimi Baba Ahmadi, M., Negahdaripour, M., Ghasemi, Y., 2020. The interaction of phages and bacteria: the co-evolutionary arms race. Crit. Rev. Biotechnol. 40, 119-137.

Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A., and Marraffini, L.A. 2015. Co-transcriptional DNA and RNA cleavage during type III CRISPR-cas immunity. Cell 161, 1164-1174.

Shiimori, M., Garrett, S.C., Graveley, B.R., and Terns, M.P. 2018. Cas4 nucleases define the PAM, length, and orientation of DNA fragments integrated at CRISPR loci. Mol.

Cell. 70, 814-824.e6.

Shin, J., Jiang, F., Liu, J.J., Bray, N.L., Rauch, B.J., Baik, S.H., Nogales, E., Bondy-Denomy, J., Corn, J.E., Doudna, J.A., 2017. Disabling Cas9 by an anti-CRISPR DNA mimic. Sci. Adv. 3, e1701620.

Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., Zhang, F., and Koonin, E. V. 2015. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. Mol. Cell 60, 385-397.

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., Siksnys, V., 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J. 30, 1335-1342.

Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. 2013. In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. EMBO J. 32, 385-394.

Staals, R.H.J., Jackson, S.A., Biswas, A., Brouns, S.J.J., Brown, C.M., Fineran, P.C., 2016. Interference dominates and amplifies spacer acquisition in a native CRISPR-Cas system. Nat. Commun. 23, 127-135.

Stanley, S.Y., and Maxwell, K.L. 2018. Phage-encoded anti-CRISPR defenses. Annu. Rev. Genet. 52, 445-464.

Strecker, J., Jones, S., Koopal, B., Schmid-Burgk, J., Zetsche, B., Gao, L., Makarova, K.S., Koonin, E. V., Zhang, F., 2019. Engineering of CRISPR-Cas12b for human genome editing. Nat. Commun. 10, 212.

Swarts, D.C., and Jinek, M. 2019. Mechanistic insights into the cis- and trans-acting DNase activities of Cas12a. Mol. Cell. 73, 589-600.e4.

Swarts, D.C., Mosterd, C., van Passel, M.W.J., Brouns, S.J.J., 2012. CRISPR interference directs strand specific spacer acquisition. PLoS One 7, e35888.

Uribe, R. V., van der Helm, E., Misiakou, M.-A., Lee, S.-W., Kol, S., and Sommer, M.O.A. 2019. Discovery and characterization of Cas9 inhibitors disseminated across seven bacterial phyla. Cell Host Microbe 25, 233-241.e5.

Vale, P.F., Lafforgue, G., Gatchitch, F., Gardan, R., Moineau, S., Gandon, S. 2015. Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. Proc. Biol. Sci.

282, 20151270.

van der Ploeg, J.R., 2009. Analysis of CRISPR in *Streptocccus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. Microbiology 155, 1966-1976.

Van Houte, S., Ekroth, A.K.E., Broniewski, J.M., Chabas, H., Ashby, B., Bondy-Denomy, J., Gandon, S., Boots, M., Paterson, S., Buckling, A., Westra, E.R., 2016. The diversity-generating benefits of a prokaryotic adaptive immune system. Nature 532, 385-388.

Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., and Wang, Y. 2015. Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. Cell 163, 840-853.

Wei, J., Lu, N., Li, Z., Wu, X., Jiang, T., Xu, L., Yang, C., and Guo, S. 2019. The *Mycobacterium tuberculosis* CRISPR-Associated Cas1 involves persistence and tolerance to anti-tubercular drugs. Biomed Res. Int. 2019, 7861695.

Wei, Y., Chesne, M.T., Terns, R.M., and Terns, M.P. 2015a. Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*. Nucleic Acids Res. 43, 1749-1758.

Wei, Y., Terns, R.M., Terns, M.P., 2015b. Cas9 function and host genome sampling in type II-A CRISPR–Cas adaptation. Genes Dev. 29, 356-361.

Westra, E.R., Buckling, A., Fineran, P.C., 2014. CRISPR-Cas systems: Beyond adaptive immunity. Nat. Rev. Microbiol. 12, 317-326.

Westra, E.R., Van Houte, S., Oyesiku-Blakemore, S., Makin, B., Broniewski, J.M., Best, A., Bondy-Denomy, J., Davidson, A., Boots, M., Buckling, A., 2015. Parasite exposure drives selective evolution of constitutive versus inducible defense. Curr. Biol. 25, 1043-1049.

Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W., and Doudna, J.A. 2009. Structural Basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure 17, 904-912.

Wilkinson, M., Drabavicius, G., Silanskas, A., Gasiunas, G., Siksnys, V., Wigley, D.B., 2019. Structure of the DNA-bound spacer capture complex of a Type II CRISPR-Cas system. Mol. Cell. 75, 90-101.e5.

Yang, H., and Patel, D.J. 2017. Inhibition mechanism of an anti-CRISPR suppressor AcrIIA4 targeting SpyCas9. Mol. Cell 67, 117-127.e5.

Yoganand, K.N., Muralidharan, M., Nimkar, S., Anand, B., 2019. Fidelity of prespacer capture and processing is governed by the PAM-mediated interactions of Cas1-2 adaptation complex in CRISPR-Cas type I-E system. J. Biol. Chem. 294, 20039-20053.

Yosef, I., Goren, M.G., and Qimron, U. 2012. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Res. 40, 5569-5576.

Young, J.C., Dill, B.D., Pan, C., Hettich, R.L., Banfield, J.F., Shah, M., Fremaux, C., Horvath, P., Barrangou, R., and Verberkmoes, N.C. 2012. Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in *Streptococcus thermophilus*. PLoS One. 7, e38077.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J., Regev, A., Koonin, E. V., and Zhang, F. 2015. Cpf1 Is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759-771.

Annex C: Supplementary data

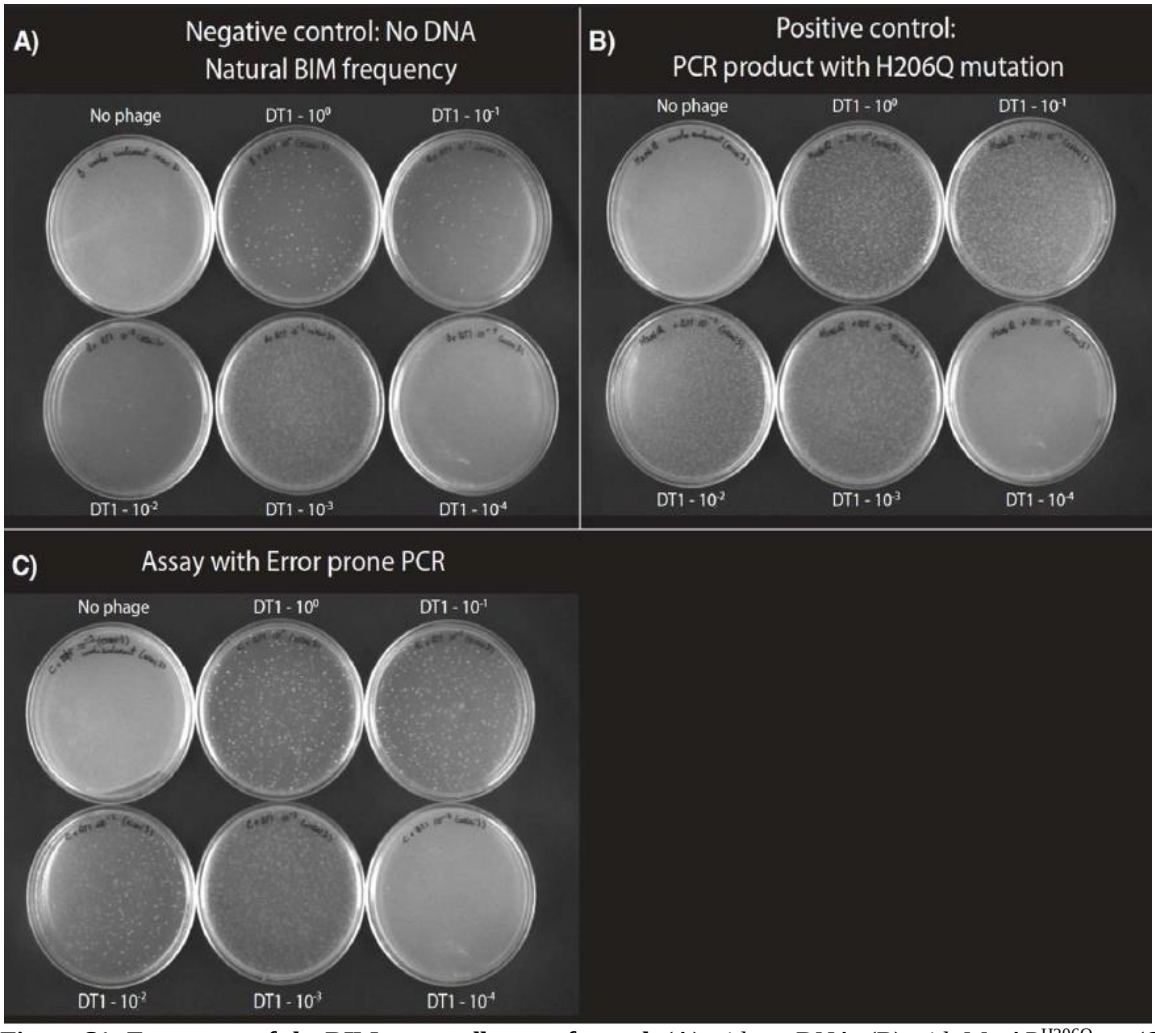


Figure S1: Frequency of the BIMs naturally transformed. (A) without DNA; (B) with MetAP^{H206Q}; or (C) with a MetAP obtained following an error prone PCR.

Table S1: List of spacers acquired by different BIMs of *S. mutans.* Mismatches to the M102AD genome are shown in *italics* in the proto-spacersequence. Spacers that are not 100% identical to M102AD are indicated by an asterisk in the spacercolumn.

Spacer	Position in phage genome (5′-3′)	Size (nt)	Times acquired	Proto-spacer sequence (5′-3′)	PAM
1	24306-24335	30	19	CTCTTTTAGCAATTGTGAAAGGACGTAATT	TAAAT
2	18743-18714	30	1	TTTTGGTCTAAAATTCTCAGGAATTTCACC	TAAAT
3	14311-14284	28	1	GCTCTGATTTCGTGTTTGTGTTATCGCC	TAAGT
4	25547-25518	30	4	GTCAAGCATACGTATATATGCTTGTGCACT	CAAAT
5	24368-24338	31	4	TGATGAAGTAAACCTCTTTTGTGAAAGGATT	TAAAT
6	16891-16922	32	4	TTAGCGCGAGTGATGATGGGTTGGTAATTGCC	AAAAA
7	14862-14891	30	4	CTCTAGCTTTATCTATTTTGATAAAGACAC	AAAAA
8	7587-7616	30	5	GTACACTCTGCAACTAACCCATCGGCACCA	CAAAT
9*	X	30	4		
10	14314-14284	31	2	AACGCTCTGATTTCGTGTTTGTGTTATCGCC	TAAGT
11	27729-27700	30	5	AAATTTTATAGCATATGCGAATATTGTTGT	TAAAA
12	27896-27867	30	1	GTTAACCGCAAGCGTAAAGTTTGCATATGC	TAAGT
13	15737-15708	30	1	AGAATTTTTCCATTCTTGCTCTTGGTTGGT	TAAAT
14	3074-3103	30	1	AGATGATAGTGACTTGTTTGCGGTAATTAA	CAAAT
15	15938-15967	30	1	AACTCTAACACTGGCTATTACTGATAAGAC	TAAAA
16	15736-15707	30	1	GAATTTTTCCATTCTTGCTCTTGGTTGGTT	AAATC
17	24343-24314	30	1	AGGATTTAAATTACGTCCTTTCACAATTGC	TAAAA
18	26698-26669	30	3	TGGTTTGCACATTTTTTTTCCTTCCTTTTT	TAAAT
19	22856-22827	30	4	TAAGATTACATTTTGCAAGTAATCTTTCTT	TAAAT
20	24402-24431	30	1	GCAAAGAAGCGATTATGAAGCGTGCAGAAG	AAAAT
21	294-265	30	1	GAATTGGGTTTTCCACAGTAGTAGCAAAGA	TAAAG
22	7700-7671	30	1	GTATGCGCCTTCTGCATTTTCGACATATCC	TAAAA
23	23484-23513	30	1	CTGTACAGTAGGATTTTTAAAGGCTCTACT	AAAAA
24*	23555-23584	31	1	GATAAAATTTTAATGCGACAAACAGAAATC	AAAAA
25	8860-8889	30	1	CAATCAAACAGCAGAAAGCCTTGCTAAAGC	AAAAT
26	12350-12320	31	1	CAAGATGGCTGAAGAGACCGGGGTTAACGCC	TAAAT
27	27519-27490	30	1	TAACGCGTTCATAATGTTGAGCAAGCCAGC	TAAAT
28	4067-4097	31	1	TTTGCGGGCGCCGAGATATACAGCGAACTTA	AAGCT
29	15738-15709	30	1	CAGAATTTTTCCATTCTTGCTCTTGGTTGG	TTAAA
30*	X	30	2		
31	19453-19424	30	2	ATACTAATAATTTCAGGCTTTAACAAATCA	TAAAT
32	15108-15078	31	3	TAGTAAAGTCAAAAGTAACTGCATTAGTCTT	TAAAT
33	24486-24456	31	1	CCTTCTTTAGTTTCAACTACAATTCCAGCCT	CAAAT
34	15510-15481	30	1	CTAGCGTATTGTCTTTATCTGTTTTTAACT	CAAAT
35	10673-10644	30	2	CTGTAATTAGACCGCCCAAGCCATGCTTGT	TAAAT
36	24485-24456	30	1	CTTCTTTAGTTTCAACTACAATTCCAGCCT	CAAAT
37	23729-23759	31	1	GAAGAATTGCAAGCAAAAATTGAAGAGGTGT	TAAAT
38	25233-25262	30	1	GCTAGTGACGTTAAAGATTTTGATGATAAT	CAAAA
39	10138-10167	30	1	TAAACACAAAGAAGCTTTGCAAGCCGTCGG	CAAGG
40	8584-8614	31	1	GCAGACAAAAGCTAAACAAGCTCTTGACTAT	CAAAA
41	9878-9907	30	1	GCTACTCGTATGTTGGATGTTATCGACGCC	AAAAT
42	5035-5065	31	1	TGAACGATTTGGGAAAAATAAAAAAAGCTCT	TAAAA
43	26558-26587	30	5	ACTTAACTCATTATAAAAAGCGTTACAGAG	AAAAT

44	27509-27538	30	1	TGAACGCGTTATATATACGTATAACGTTAA	GAAAT
45	680-709	30	1	GTAATGGCGGTTGAAAATATTCGCAGAGCT	TACAG
46	15454-15483	30	1	TTGCAAGTCAATTTGATTGTGAATTTGAGT	TAAAA
47*	23269-23240	31	1	TCGACTACATTCACAAACGCCGCTGTCTTA	TCGCC
48	21368-21339	30	1	ACATAGCCAGTGTGACCGAAAGGGTGGCCG	TAAAG
49*	24306-24335	31	1	CTCTTTTAGCAATTGTGAAAGGACGTAATT	TAAAT
50	24040-24011	30	1	GCCAACGCACTCATGTCTTTGTGCCGTTAA	AAAAT
51*	24367-24338	32	5	GATGAAGTAAACCTCTTTTGTGAAAGGATT	TAAAT
52	4463-4492	30	5	CTAAAACCGCAAGACACAGAGCCACAGGCT	CAAAA
53	21422-21452	31	1	ATGACAGCAATTTTTACACTGGCACAGTAGC	TAAAT
54	196-167	30	2	TTCTTTACTAGTAAAACTTCTGTACATTTA	TAAAA
55*	26770-26799	31	2	ATTAAAAGCTATGCGCAATAAGGACTATGT	TAAAT
56	22183-22213	31	1	GCCTTTAGACGAATGTATCCAAAATGTATCC	AAAGT
57	5500-5471	30	1	GTTTTGGCTGTAACGTCTTTGACAACGCCG	TAAAA
58	16567-16538	30	1	TTTTGTAACTGCGTATCATCAGCGCTCGAG	TAAGT
59	6897-6927	31	1	CTTGCGATGTGGACAAATTGGGGCACGGTCA	AAATG
60	4783-4754	30	1	TAAAGCGTTTTGGATTAACTGCGCTTTAGC	AAAAA
61	23905-23935	31	1	CGCATAGAGTTTTGAGAGGTGAAGAATGTTT	AAAAT
62	27259-27288	30	1	GCAGAAAAAGAAGAACTGATAAGCGATATT	AAAAA
63	24244-24274	31	1	TGGAAATACTTTGCGAGATTTTGTAGACCCG	CAAAA
64*	7587-7616	31	1	GTACACTCTGCAACTAACCCATCGGCACCA	CAAAT
65	28219-28190	30	1	ATTTTAAACATATTTTCTCACCACCTTTTT	TAAAA
66	27520-27490	31	2	ATAACGCGTTCATAATGTTGAGCAAGCCAGC	TAAAT
67	26598-26627	30	1	AGTGAGAGTAGCTACAGTTGCAAGAGCTGA	TAAAT
68	26670-26700	31	1	AAAAGGAAGGAAAAAAAATGTGCAAACCAAT	CAAAA
69	25601-25572	30	1	CTCTAAGTCATAGTCTGCTTTAATTTGCAT	TAAAA
70	6185-6214	30	2	AGCTGCTACTATCTATGAGTACAGACAAGC	AAAAT
71	6654-6625	30	1	TATTTTCAGCCATGTTTTTCGGTTTCCTTA	CAAAT
72	13969-13998	30	1	GCGCCAAAAGCAAAAACGGAGCCAGTAATC	AAAAT
73	10069-10098	30	1	CGTTTTCAAGGCCTTCAGTGGTGACAAGAG	TAAAA
74	29484-29513	30	1	GATGACATCGCATTAATTACAATCAGTTTA	TAAGT
75	10674-10645	30	1	CCTGTAATTAGACCGCCCAAGCCATGCTTG	TTAAA
76	19742-19713	30	1	TGTCTGTTGCGTTAGTTATTAAGCTATCAA	TAAGT
77	21056-21027	30	1	GATACAAACAATAAACTAGCTGACAAACCT	AAAAT
78	23097-23067	31	1	TTTGCGCGTGCGTTTTAGCTTCAAAGTTATC	AAAAC
79	9369-9340	30	3	GCTTCGTCAGTTTGCCCAAAAGCATCTTGC	AAAGT
80	4742-4771	30	1	GATAATGTTTTTGCTAAAGCGCAGTTAATC	CAAAA
81	9113-9084	30	1	TGATTTTGCTAAGTGCACCACGAACAAGAT	TAAGC
82	10549-10578	30	2	CAAAACACTACAAGTTGCTTTTGCCCCCCT	CAAAT
83*	20065-20095	31	1	AAGCCA CAGAAACAACTAC GCACCAGCA	GAAAC
84	5182-5211	30	1	ATACGAAACTTTCGAAATTAATGACAGTTG	TAAAT
85	21258-21287	30	1	ACGTGCAGGCTATGAGATTGTGCCAGCTAA	TAAAC
86*	24044-24015	31	1	GTGCGCCAACGCACTCATGTCTTTGTGCCG	TTAAA
87	29686-29716	31	1	TCATTATGTTTACACAATATTACATTAGTAT	TAAGT
88	18567-18538	30	1	CCATATCTAGAATATTTTTCCCGTTGAGGT	AAAAG
89	7577-7548	30	1	GTTACCGTCGATACGTTCGATGGAACCTTC	TAAGT
90	28701-28730	30	1	GAAATTAGAAGCAGAAGAGATTTTCTGTGA	TAAAA
91	27384-27355	30	2	ATAAGTTTGAGCGACTTTTTTCGAAACCTT	TAAAT
92	26049-26020	30	1	CTCTGCGCGCTACTATCGCAATATCATGAT	TAAAA
93	30543-30514	30	1	AAAAAAATCACATAAAGGTAGTATTGTAGA	AAAAT
94	23836-23807	30	1	GCATAGCTTTTGTGCTATCACTGATTTTCC	AAAAG

95	9783-9753	31	1	CGCTTAAAACGCGCTTCTATCAGCGGGTCCG	AAACC
96	17625-17596	30	1	CGTTGCCAAAACTATATCTGATATCATGCC	TAAAA
97	28220-28249	30	1	CGAAAAATATAGACCTTTCCGAATTAATTA	CAAAT
98	13968-13998	31	1	AGCGCCAAAAGCAAAAACGGAGCCAGTAATC	AAAAT
99	22318-22348	31	1	GGAAAAGCAATTGTTAAGCGTCGAAGGAATT	AAAGG
100	23531-23501	31	1	TTTGTACACATCATTTTTAGTAGAGCCTTTA	AAAAT
101	27730-27700	31	1	AAAATTTTATAGCATATGCGAATATTGTTGT	TAAAA
102	14469-14441	29	1	CTTTGAACCAATGAACTTGTAAGGAGTCA	CAAAG
103	15498-15469	30	1	CTTTATCTGTTTTTAACTCAAATTCACAAT	CAAAT
104	30423-30452	30	1	ATGTCACATTATTAGAATACTAGAATATTC	TAATA
105	18539-18570	32	1	CCTCAACGGGAAAAATATTCTAGATATGGTAA	AAGTG
106	5019-4989	31	1	TAGCTTCTGCCGGGTCTGAAGTAGATAGCCC	AAAAT
107	27731-27760	30	1	CAGACGGATACCTGGCGTTATGTTATGCAG	CAAAT
108	100-71	30	1	ATTTCGCTTAATCCGATTGAATATCCGGTT	TAAAT
109	19150-19121	30	1	GGTTAATCCGCTGACTTGAATAGTTGCACT	CAAAT
110	22555-22584	30	1	GTATTTAAGACCAGAAACACTTTTCAGTAA	TAAAT
111*	9430-9459	32	1	CAGACATGTTGTCCATCATCAATACATTT CC	TAAAA
112	3522-3492	31	1	TGGCAAATGTATATAGTTTTCGTTTAAGTTC	TAAAT
113	21490-21461	30	1	AACCTAACATGTAATCGCGTGTGCGGTGCA	TAAAT
114*	29177-29148	31	1	AACCATCTGTTAGTTTTCTCATTAAATTT	CAAAA
115	24367-24338	30	7	GATGAAGTAAACCTCTTTTGTGAAAGGATT	TAAAT
116	9920-9949	30	1	GTGGCAAAAGTGATAGCTAGGAACATGCCT	CAAAT
117	12260-12231	30	1	GGTCAGCTCCTGCGTCTTTAGCTTTGTCTT	TAAAG
118	1220-1249	30	1	AGATAAGAAAGACGATAATCTAAGATTTAA	TAAAA
119	29184-29155	30	1	TTTCGAAAAACCATCTGTTAGTTTTCTCAT	TAAAT
120*	30034-30004	32	1	CTGCGCTTCTCGAGTTTTATCGATAGACT CA	AAAAT
121	27692-27721	30	2	TTGTTTTAACAACAATATTCGCATATGCTA	TAAAA
122	19103-19132	30	1	CGTAGCTTCAACTATTTGAGTGCAACTATT	CAAGT
123	17928-17899	30	1	CGGCAGTTTTAGCACTTGCGGTAACTTGAC	TAAAG
124	12349-12320	30	2	AAGATGGCTGAAGAGACCGGGGTTAACGCC	TAAAT
125	5097-5068	30	1	CAATGTCGTTTTGCACAAAGTTTTGAGTTT	TAAGA
126	29057-29086	30	1	ACGCTTGATGCGCGCTTGGTTGTTTGATTA	CAAAA
127	14150-14179	30	1	CAGCTAAAAATGTTGCAATCAACACGCTTC	CAAAT
128	13903-13932	30	1	GGTCACGACAGCGATTATGGAACCATCACC	AAAAA
129	24102-24132	31	1	CCGAGTCTAGATTAACTATCAGTGTAGCGCT	AAAGT
130	4611-4641	31	1	ACAACGACGACGAGCAAGACGAAAAGCCTGA	TAAAG
131*	X	30	2		
132	24305-24335	31	1	GCTCTTTTAGCAATTGTGAAAGGACGTAATT	TAAAT
133	5165-5195	31	1	TGAACGTGTGCGTGACTATACGAAACTTTCG	AAATT
134	20127-20156	30	1	GCGCAGAGGTACCATCTGCAAACGAGATTT	TAAAA
135	22694-22723	30	1	GTGGCCGTTAAAAAGCAAGGAAGAATACGT	TAAAT
136	26541-26570	30	2	GCTAGCACGTTGAAAACACTTAACTCATTA	TAAAA
137	21423-21452	30	1	TGACAGCAATTTTTACACTGGCACAGTAGC	TAAAT
138	29338-29367	30	1	CAAAGTCTTTAAGCACAAACGAGCGACTGA	CAAAT
139	26203-26173	31	1	CAAAATAAAAAAGCTTATAGAAATAAGCTTT	AAAAT
140	4074-4045	30	1	CCCGCAAAATAAGAGCCACCGCCGGAGTTC	AAAAT
141	19201-19230	30	1	AAACAAGAAAATCTTTGTTCAGCTACTTCC	AAAAT
142*	27729-27700	31	1	AAATTTTATAGCATATGCGAATATTGTTGT	TAAAA
143	9523-9552	30	1	TGAAGAGTATACAAAAATGCAGTCGGAAGG	CAAAA
144	10070-10099	30	1	GTTTTCAAGGCCTTCAGTGGTGACAAGAGT	AAAAT

145*	20010-20039	30	1	TTAAAGACGAAAATTCACGCGG ACTGTAG	TGAAA
146	6754-6725	30	1	GCGAGCTCAGCAAACTTCTTTGCGCCTGCG	TAAGT
147	25859-25888	30	1	TGAAAGAAAGGTTAAGCTAATGCTAAAACT	AAAAG
148	25628-25599	30	5	TTTTTCTTTTTTCACTTTCTGCGCTTTCTC	TAAGT
149*	24367-24338	31	3	GATGAAGTAAACCTCTTTTGTGAAAGGA TT	TAAAT
150	26090-26120	31	3	ACTGGAAAACGGGCTCAAGGAGTTTAAACAG	CAAAT
151	10581-10611	31	2	AATTAATTGGCACTTTAGTTGTCCTACCTTT	TAAAA
152	26863-26892	30	1	TATACGTTATTTCAACTGGAAAATTTGCAA	CAAAT
153	18921-18892	30	1	CCAATTTCTAACATATTCTCACCCCCTTTC	GAAAT
154	26142-26171	30	1	CGACTGAGAAAATTGATGTCAATAGAATTT	TAAAG
155*	9920-9949	31	1	GTGGCAAAAGTGATAGCTAGGAACATGC CT	CAAAT
156	10923-10952	30	1	AACAAACACAAAAAGACATACAGAAGATGA	TAAAA
157*	19006-18977	30	1	TCCACTAATTT GTCATCACTAAAATCAAC	TAAAG
158	3061-3031	31	1	CGTGTCGTCAATGCAAATGTCGACGTAAAAT	CAAAC
159	19971-20000	30	1	AACAAATTAGTGACTTAATTGAGTCTGCCG	TAAAT
160	29822-29793	30	1	CAAAGCCTTTAGGCTTTGATATTTACTTGC	TAAAT
161	9430-9459	30	1	AGACATGTTGTCCATCATCAATACATTTCC	TAAAA
162	1296-1325	30	1	GCTAGCGTGCAGATTGATTATGATTTTCTT	ATAAA
163*	5314-5343	31	1	TTGACGACTTAATCCAACTCGACTTTAA CA	TCAAA
164	21281-20310	30	1	AATCAAACTGACTGAGTCGACTGGTTATTA	TAACA
165	7685-7714	30	1	TGCAGAAGGCGCATACGAACTTCAAGAGGT	TAAAC
166	14911-14882	30	1	AACTCAAATCTAATGTTTTTGTGTCTTTAT	CAAAA
167	26285-26315	31	1	GGTAAAAACCATGAAAACAATCGCAAAATTA	CAAAA
168	24307-24335	29	1	TCTTTTAGCAATTGTGAAAGGACGTAATT	TAAAT

Table S2: Construct insert sequences. Proto-spacer sequences with flanking PAM in *italic*.

Constructname	**Insert sequence (5'-3')**
pNZ123-sp1	AAATATTTGAAAATTGTTTTTCACTAGATA-*TAAAT*
pNZ123-sp2	CTCTTTTAGCAATTGTGAAAGGACGTAATT-*TAAAT*
pNZ123-sp3	TTTTGGTCTAAAATTCTCAGGAATTTCACC-*TAAAT*
PAM1TA	AAATATTTGAAAATTGTTTTTCACTAGATA-*AAAAT*
PAM1TC	AAATATTTGAAAATTGTTTTTCACTAGATA-*CAAAT*
PAM1TG	AAATATTTGAAAATTGTTTTTCACTAGATA-*GAAAT*
PAM2AC	AAATATTTGAAAATTGTTTTTCACTAGATA-*TCAAT*
PAM2AG	AAATATTTGAAAATTGTTTTTCACTAGATA-*TGAAT*
PAM2AT	AAATATTTGAAAATTGTTTTTCACTAGATA-*TTAAT*
PAM3AC	AAATATTTGAAAATTGTTTTTCACTAGATA-*TACAT*
PAM3AG	AAATATTTGAAAATTGTTTTTCACTAGATA-*TAGAT*
PAM3AT	AAATATTTGAAAATTGTTTTTCACTAGATA-*TATAT*
PAM4AC	AAATATTTGAAAATTGTTTTTCACTAGATA-*TAACT*
PAM4AG	AAATATTTGAAAATTGTTTTTCACTAGATA-*TAAGT*
PAM4AT	AAATATTTGAAAATTGTTTTTCACTAGATA-*TAATT*
PAM5TA	AAATATTTGAAAATTGTTTTTCACTAGATA-*TAAAA*
PAM5TC	AAATATTTGAAAATTGTTTTTCACTAGATA-*TAAAC*
PAM5TG	AAATATTTGAAAATTGTTTTTCACTAGATA-*TAAAG*
PAM2AG4AT	AAATATTTGAAAATTGTTTTTCACTAGATA-*TGATT*
PAM2AG4AC	AAATATTTGAAAATTGTTTTTCACTAGATA-*TGACT*
PAM2AG4AG	AAATATTTGAAAATTGTTTTTCACTAGATA-*TGAGT*

Table S3: Phage adsorption percentages in all strains tested in this study. BIMs are identified by the MOI at which they were obtained, followed by the colony number. The second column indicates the adsorption of phage M102AD to the strain. The third column lists the number of assays the results are based on.

Strain	Adsorption percentage	Number of assays
WT	97.4 ± 2.2	18
BIM 5	75.1 ± 17	17
0.07 – 27	98.8 ± 0.4	6
0.07 – 33	98.2 ± 1.3	4
0.07 – 35	97.5 ± 0.5	2
0.07 – 37	98 ± 1.9	6
0.07 – 45	96.8 ± 2.5	6
0.07 – 47	94 ± 0	2
0.07 – 51	96 ± 0	2
0.07 – 59	97.5 ± 0.5	2
0.14 – 22	98.5 ± 0.5	2
0.14 – 35	98 ± 0	2
0.68 – 26	96 ± 0	1
0.68 – 49	92 ± 4	2
0.68 – 50	97 ± 1	2
0.68 – 51	94.5 ± 1.5	2
0.68 – 58	44 ± 11	2
0.68 – 59	44.5 ± 1.5	2
1.4 – 9	26 ± 6	2
1.4 – 15	98 ± 1	2
1.4 – 16	0 ± 0	2
1.4 – 17	97.5 ± 0.5	2
1.4 – 28	98.5 ± 0.5	2
2.2 – 6AB	12.5 ± 6.5	2
2.2 – 6AC	23.5 ± 14.5	2
2.2 – 6B	98.5 ± 0.5	2
2.2 – 7	99 ± 0	2
2.2 – 11	67.5 ± 4.5	2
2.2 – 12	70.5 ± 1.5	2
2.2 – 17	19 ± 11	2
2.2 – 20A	65 ± 3	2
2.2 – 20C	62 ± 5	2
2.2 – 29B	54 ± 6	2
2.2 – 29C	46 ± 2	2
2.2 – 40A	24.5 ± 23.5	2
2.2 – 40C	47.5 ± 1.5	2
2.2 – 42A	59.5 ± 0.5	2
2.2 – 43CA	93 ± 1	2
2.2 – 44C	57 ± 0	2
2.2 – 45B	83 ± 0	2
2.2 – 45C	69.5 ± 2.5	2
2.2 – 46	98 ± 0	2
2.2 – 48	74 ± 2	2
4.6 – 8	15 ± 10	2

4.6 – 9B	98.5 ± 0.5	2
4.6 – 9C	65 ± 2	2
4.6 – 10	97.5 ± 0.5	2
4.6 – 13	76 ± 6	2
4.6 – 14	26 ± 13	2
4.6 – 15	20.5 ± 9.5	2
4.6 – 16	98.5 ± 0.5	2
4.6 – 18	48 ± 11	2
4.6 – 23	49.5 ± 3.5	2
4.6 – 25	97 ± 0	2
4.6 – 27	42 ± 3	2
4.6 – 29C	98.5 ± 0.5	2
4.6 – 34	84 ± 2	2
4.6 – 35	34 ± 3	2
4.6 – 36	40 ± 3	2
4.6 – 37C	14 ± 12	2
4.6 – 40	26 ± 8	2
4.6 – 41	97.5 ± 0.5	2
6.8 – 8	98 ± 1	2
6.8 – 14	23 ± 1	2
6.8 – 16	56 ± 11	2
6.8 – 18	98.5 ± 0.5	2
6.8 – 26	27.5 ± 27.5	2
6.8 – 28	58.5 ± 0.5	2
6.8 – 33	97.5 ± 0.5	2
6.8 – 38	95.5 ± 0.5	2
9.2 – 1A	27 ± 3	2
9.2 – 3	96.5 ± 2.5	2
9.2 – 5	54 ± 2	2
9.2 – 6	45 ± 6	2
9.2 – 8	68 ± 11	2
9.2 – 8A	98.5 ± 0.5	2
9.2 – 8C	98 ± 1	2
9.2 – 10	37.5 ± 17.5	2
9.2 – 13	38.5 ± 16.5	2
9.2 – 14	43.5 ± 6.5	2
9.2 – 15	99 ± 0	2
9.2 – 16	99 ± 1	2
9.2 – 20	19 ± 3	2
9.2 – 21	99 ± 0	2
9.2 – 22	94.5 ± 2.5	2
9.2 – 27	98.5 ± 0.5	2
9.2 – 31	15.5 ± 15.5	2
9.2 – 43	25.5 ± 9.5	2
9.2 – 44	99 ± 0	2
9.2 – 46	42.5 ± 9.5	2
18.3 – 4	82 ± 2	2
18.3 – 8	58 ± 9	2
18.3 – 12	92.5 ± 1.5	2

18.3 – 14	97.5 ± 1.5	2
18.3 – 15	33.5 ± 1.5	2
26.7 – 18A	61 ± 2	2
26.7 – 25B	97.5 ± 0.5	4
26.7 – 27A	0 ± 0	2
26.7 – 27C	15.3 ± 20.3	4
26.7 – 31A	95.5 ± 1.5	2
26.7 – 50B	97 ± 1	2
26.7 – 52A	29 ± 8	2
26.7 – 52B	27 ± 7	2
26.7 – 56A	1.5 ± 0.5	2
26.7 – 67A	12.5 ± 12.5	2
26.7 – 67B	4.5 ± 4.5	2
26.7 – 70B	26 ± 27.2	4
26.7 – 73C	95.5 ± 2.5	2
26.7 – 78A	0 ± 0	2
26.7 – 82A	93 ± 0	2
26.7 – 85A	93 ± 0	2
26.7 – 85B	91 ± 2	2
26.7 – 85C	93.5 ± 2.5	2
27.5 – 3A	25 ± 6	2
27.5 – 19	42 ± 4	2
27.5 – 25A	63 ± 0	1
53.3 – 4B	96.5 ± 1.5	2
53.3 – 4C	97.5 ± 1.5	2
53.3 – 28A	27.5 ± 9.4	4
53.3 – 54A	96.5 ± 3.5	2
53.3 – 57A	92.5 ± 1.5	2
53.3 – 57B	97 ± 0	2
53.3 – 65A	92.5 ± 4.5	2
53.3 – 65C	96.5 ± 0.5	2
53.3 – 76B	96.8 ± 2.5	4
53.3 – 79F	92.5 ± 4.5	2
53.3 – 82B	29.8 ± 18.8	4
53.3 – 83A	24 ± 8	2
53.3 – 83B	26 ± 5	2
53.3 – 83C	5 ± 1	2
53.3 – 85B	14 ± 5	2
53.3 – 87A	1± 0	2
53.3 – 90A	27 ± 8	2
53.3 – 90B	25.5 ± 9.5	2
53.3 – 90C	24.5 ± 2.5	2

Table S4: List of primers in Article 3. Restriction sites for BsaI are in *italics*.

Oligo	**Sequence (5' -> 3')**
CB13.42	GCAAATGACAGAAGAACAGC
p2.27	GCACAACCTATTGTAAAACC
pTRKL2_F	AAATATAGAAATATTTCTGTATTTTTTGGGCCAGTGAATTCCCGGGGATC

pTRKL2_R	AAGTGTCTTTTATGGGATTTTCTTTAAATCTTCTATTTAATCACTTTGAC
SmutCas9_F	GTCAAAGTGATTAAATAGAAGATTTAAAGAAAATCCCATAAAA
SmutCas9_R	GATCCCCGGGAATTCACTGGCCCAAAAAATACAGAAATATTTC
spacer49_F	*AAAC*CATCTATCTTATTGGTAGTGGCTGGAGTAT*G*
spacer49_R	*AAAAC*ATACTCCAGCCACTACCAATAAGATAGATG

Table S5: List of primers Article 4.

Primer	**Sequence 5'-3'**	**Function**	**Reference**
SJL128	TCAATCTACTCAAGGTATGAATCA	Natural transformation of *metAP*	This study
SJL130	GTCAGTAGTAGTGGTCAAGA	Natural transformation of *metAP*	This study
SJL150	AGTTCCTGATAGGTCGCATT	Detection of the mutation in*metAP*	This study
SJL151	GAGTTGGACCAACAATGCAG	Detection of the mutation in*metAP*	This study
SJL154	CAGCAGCGGCCTGGTGCCGCGCGGCAGC CAAATGATTACACTGAAATCAGCACGTG	Cloning *metAP* in pNZ123	This study
SJL155	GCCGGATCTCAGTGGTGGTGGTGGTGGT GCTTAATAAGTTCTTTCTTCCCCTTGAG	Cloning *metAP* in pNZ123	This study
SJL160	ATTACAGCTCCAGATCCAGTACTGAATT CTTGAGCCTGCTATGATTGACTCTGCA	Cloning the mutated *S. mutans metAP* in pNZ123	This study
SJL161	ATTGGGTTCTTCCTGCATGGTTGG	Cloning the mutated *S. mutans metAP* in pNZ123	This study
SJL162	CCAACCATGCAGGAAGAACCCAAT	Cloning the mutated *S. mutans metAP* in pNZ123	This study
SJL163	GAAAATATGCACTCGAGAAGCTTGAGCT CTCCGAAGGTGGACAACATAATAGC	Cloning the mutated *S. mutans metAP* in pNZ123	This study
CM_145	GAATTCGAATTCAAAGGCTGTTGTGACA GCAA	Cloning the *S. mutans metAP* in pNZ123	This study
CM_146	CTCGAGCTCGAGTTAATAAGTCCCTTCT TGACCC	Cloning the *S. mutans metAP* in pNZ123	This study
CR1-fwd	TGCTGAGACAACCTAGTCTCTC	CR1 locus screening	Hynes et al., 2017
CR1-revLong	TAAACAGAGCCTCCCTATCC	CR1 locus screening	Hynes et al. 2017
CR3-fwd	CTGAGATTAATAGTGCGATTACG	CR3 locus screening	Hynes et al. 2017
CR3-rev	GCTGGATATTCGTATAACATGTC	CR3 locus screening	Hynes et al. 2017

Table S6: Analysis of the post-translational processing of the N-terminal methionine. Y = processed, N = not processed. Groups A-D refer to the classification of Figure 22.

Group	Accession number for DGCC7796 (except other wise stated)	Putative function	TIGR Family	Link	Role	Subrole	DGCC7796	DGCC7796 + D4090	DGCC7796:MetAP	DGCC7796:MetAP	DGCC7796 + M5876	DGCC7796:MetAP
A	0002	DNA polymerase III beta subunit	00663	132	DNA metabolism	Degradation of DNA	N	N	N	N	N	N
A	0312	COG1399 in cluster with ribosomal L32p	00092	157	Unknown function	Conserved			N		N	N
A	0406	1-phospho-fructokinase					N	N	N	N	N	
A	0563	Mercuric ion reductase					N	N	N	N	N	N
A	0581	Threonyl-tRNA synthetase	00745	83	Synthesis of cofactors, prosthetic groups, and carriers	Molybdopterin			N		N	N
A	0620							N		N		
A	0641	Catabolite control protein A	00115	95	Protein fate	Toxin production and resistance		N	N		N	N
A	07022	Haloalkanoic acid dehalogenase						N	N		N	
A	1159							N		N	N	N
A	1159	LSU ribosomal protein L19p	01979	76	Synthesis of cofactors, prosthetic groups, and carriers	Serine family	N	N	N	N	N	N
A	1250	FIG137478: Hypothetical protein YbgI						N	N		N	N
A	1295	Phenylalanyl-tRNA synthetase alpha chain						N	N		N	N
A	1361	oxidoreductase of aldo/keto reductase family, subgroup 1					N	N	N	N		
A	1424						N	N	N	N	N	N
A	1657	Thioredoxin reductase						N		N		
A	1741	Sucrose-6-phosph. hydrolase	00040	703	Unknown function	Conjugation	N	N	N	N	N	N
A	1757	SSB					N	N	N	N	N	N

A	1861	Pyrroline-5-carboxylate reductase	00232	117	Energy metabolism	Glycolysis Gluconeogenesis	N	N	N	N	N	N
A	1880	Tyrosyl-tRNA synthetase					N	N	N	N	N	
A	1905	Hypothetical protein ywlG					N	N	N	N		N
A	1922	Hypothetical protein	01136	75	Amino acid biosynthesis	Pyruvate family			N		N	
A	1932						N	N	N	N		N
A	1956	Preprotein translocase secY subunit						N		N		
A	2046	ABC transporter ATP-binding protein uup	03168				N	N	N	N	N	N
A	0282							N	N			
A	M5876_0033		00061	158	Protein synthesis	General					N	N
A	M5876_0036		00496	169	Protein synthesis	tRNA & rRNA basemodification					N	N
A	1173										N	N
B	0004	GTP-binding and nucleic acid-binding protein YchF									N	N
B	0034							Y		Y		
B	0072		00962	111	Energy metabolism	Anaerobic		Y		Y		
B	0093	2-dehydropantoate 2-reductase	01216	111	Energy metabolism	Anaerobic	Y	Y	Y	Y	Y	Y
B	0117							Y		Y		
B	0129	Cell division trigger factor	00855	158	Protein synthesis	General	Y	Y	N	Y	Y	Y
B	DGCC7796_0229	Aminopeptidase C	01891	138	Protein fate	tRNA aminoacylation	Y	Y	Y	N	Y	Y
B	0242							Y		Y		
B	0366	Cysteine synthase	01173	89	Cell envelope	Other	Y	N	N	N	Y	Y
B	0371		00418	137	Protein synthesis	Other		Y		Y		
B	0392		01123	74	Amino acid biosynthesis	Glutamate family		Y		Y		
B	0460							Y		Y		
B	0481							Y		Y		
B	0486	ATP synthase alpha chain	01481	261	Regulatory functions	Chlorophyll & bacteriochlorphyl	Y	Y	Y	Y	Y	Y
B	0489	ATP synthase epsilon chain	01060	116	Energy metabolism	Fermentation		Y			Y	Y
B	0542	LSU ribosomal	00499	137	Protein	Other	Y	Y	Y	Y	Y	Y

		protein L7/L12 (P1/P2)			synthesis							
B	0648	Enolase	01549	703	Unknown function	Conjugation	Y	Y	Y	Y	Y	Y
B	0747	Cell division protein FtsZ	00065	93	Cellular processes	Other		Y	Y		Y	Y
B	0752	Cell division initiation protein DivIVA	03544				N	Y	Y	N	Y	Y
B	0837	Tripeptide aminopeptidase					Y		Y		Y	Y
B	0982	Orotate phospho-ribosyltransferase					Y	Y	Y		Y	Y
B	1299	Aspartate-semialdehyde dehydrogenase	00536	140	Protein fate	Degradation of proteins,peptides, glycopeptides	Y	Y		Y	Y	Y
B	1359	NADH dehydrogenase	01882	138	Protein fate	tRNA aminoacylation	Y	Y	Y	Y		N
B	1399	UDP-glucose 4-epimerase	01886	138	Protein fate	tRNA aminoacylation	Y	Y	Y	Y	Y	Y
B	1624	Nicotinate-nucleotide adenylyltransferase	00107	127	Purines, pyrimidines, nucleosides, nucleotides	Pyrimidine ribonucleotide biosynthesis			Y		Y	Y
B	1782	Glutamine synthetase type I					Y	Y	Y	Y	Y	Y
B	1797	SSU ribosomal protein S12p (S23e)	02093	105	Energy metabolism	Polyamine biosynthesis	Y	Y	Y	Y		
B	1881	Ketol-acid reductoisomerase	00972	143	Transport binding proteins	Amino acids, peptides and amines	Y	Y	Y	Y	Y	Y
B	1940		00336	126	Purines, pyrimidines, nucleosides, nucleotides	Purine ribonucleotide biosynthesis		Y		Y		
B	1941		01870	719				Y		Y		
B	1944	SSU ribosomal protein S8p (S15Ae)					Y	Y	Y	Y	Y	Y
B	0921										Y	Y
B	1890										Y	Y
B	1987		00884	125	Purines, pyrimidines, nucleosides, nucleotides	Nucleotide and nucleoside interconversions					Y	Y
C	0201		00686	157	Unknown function	Conserved		Y		N		

C	0257	Phosphoesterase	01135	89	Cell envelope	Other	Y	Y	N	N	Y	N
C	0308	Transketolase					Y	Y	N	Y	Y	N
C	0331	FIG01115503: hypothetical protein						Y	N			N
C	0409		01024	158	Protein synthesis	General		Y		N		
C	0602	Branched-chain amino acid aminotransferase	02651	166	Transcription	Transcription factors	Y	Y	N	N	Y	
C	0759	3'-to-5' oligoribonuclease A					Y	Y	N	N	Y	N
C	0848	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase						Y	N	N	Y	N
C	0941	Maltodextrin phosphorylase	00486	157	Unknown function	Conserved	Y	Y	N	N	Y	N
C	1052	GMP synthase, ATP pyrophosphatase subunit					Y	Y	N	N	Y	N
C	1064	Glucosamine--fructose-6-phosphate aminotransferase						Y	N		Y	
C	1214	Cystathionine beta-lyase	00468	137	Protein synthesis	Other	Y	Y	N	N	Y	
C	1410	Glutathione biosynthesis bifunctional protein gshF	01296	71	Amino acid biosynthesis	Other	Y		N	N	Y	N
C	1525	Phosphoserine aminotransferase					Y	Y	N	N	Y	N
C	1643	Universal stress protein family					Y	Y	N	N	Y	N
C	1714	Gamma-glutamyl phosphate reductase	01179	119	Energy metabolism	Pyruvate dehydrogenase			N		Y	N
C	1744	Alkaline shock protein	01435	86	Synthesis of cofactors, prosthetic groups, carriers	Riboflavin, FMN, and FAD	Y		N			
C	1813	Putative transport accessory protein					Y	Y	N	N	Y	N
C	1939							Y		Y		
C	1964	Adenylosuccinate synthetase	01214	90	Cell envelope	Biosynthesis and degradation of murein sacculus and peptidogly	Y	Y		N	Y	

						can						
C	2024		01364	75	Amino acid biosynthesis	Pyruvate family		Y		N		N
C	0760										Y	N
D	0699	Lysyl-tRNA synthetase (class II)	00125				N		Y			
P	1476	dTDP-4-dehydrorhamnose reductase					N	N	N	N	Y	N
P	1828	Carbonic anhydrase	01292	112	Energy metabolism	ATP-proton motive force interconversion	Y	Y	N	N	Y	Y
P	1937	Adenylate kinase	00407	73	Amino acid biosynthesis	Aspartate family	N	N	N	N	Y	N
U	0130	DNA-directed RNA polymerase delta subunit	01322	105	Energy metabolism	Polyamine biosynthesis	N	N	N	N	Y	N
	0155	Amino acid transport ATP-binding protein							N			
	0162	Cysteine desulfurase, SufS subfamily	00621	132	DNA metabolism	Degradation of DNA			N			
	0202		00653	73	Amino acid biosynthesis	Aspartate family		Y		Y		
	0236							Y				
	0302	Methionine ABC transporter permease protein	00981	158	Protein synthesis	General			N			
	0417	LSU ribosomal protein L21p							N			N
	0440	Ribosome recycling factor							N	N		N
	0541	LSU ribosomal protein L10p (P0)							N	N		N
	0560	Aminoacylase/N-acyl-L-amino acid amidohydrolase/hippurate hydrolase	00112	73	Amino acid biosynthesis	Aspartate family			N	N		N
	0570	Glucosamine-1-phosphate N-acetyltransferase	00234	137	Protein synthesis	Other	Y	Y			Y	
	0607		00465	74	Amino acid biosynthesis	Glutamate family		Y				
	0753		01440	156	Hypothetical proteins	Transposon functions				N		
	0764	N(5)-glutamine methyltransferase PrmC							N			
	0861	Purine nucleoside phosphorylase	00564	70	Amino acid biosynthesis	Other				N		
	0921	Dihydrolipoamide acetyltransferase	00167	116	Energy metabolism	Fermentation	Y	Y	N	Y		

	0948	Phosphate transport ATP-binding protein PstB							N		Y	
	0988	CRISPR-associated, Csm2 family	01351	124	Purines, pyrimidines, nucleosides, nucleotides	2'-Deoxyribonucleotide metabolism			N			
	1002	Fibronectin/fibrinogen-binding protein							N			
	1032	FIG01119309: hypothetical protein							N			
	1037							Y				
	1063	Alkylphosphonate utilization operon protein PhnA							N			N
	1147	Phosphoglucomutase							N	N		N
	1213	Ribonuclease								N		
	1245	dTDP-4-dehydrorhamnose 3,5-epimerase								N		N
	1267							Y				
	1272							Y				
	1469	Rhamnose-containing polysacharide translocation permease							N			
	1535	Putrescine transport ATP-binding protein PotA	00184	125	Purines, pyrimidines, nucleosides, nucleotides	Nucleotide and nucleoside interconversions		Y				
	1794							Y				
	1814							N				
	1917	Hypothetical protein								N		
	1925	Fructose-bisphosphate aldolase class II							N	N		N
	1942							Y				
	1947							N				
	1952							N				
	1953									N		N
	1955							Y				
	0031											N
	0095											N
	0330										Y	
	0388											N
	0626											N
	0649											N
	1461											N
	M587										N	

	6_0027											

Annex D: Sequence pTRKL2-SmutCas9.

SmutCas9 is in *italics*.

```
  1         10         20         30         40         50
  |         |          |          |          |          |
  AAGGAGAACAGCTGAATGAATATCCCTTTTGTTGTAGAAACTGTGCTTCA
  TGACGGCTTGTTAAAGTACAAATTTAAAAATAGTAAAATTCGCTCAATCA
  CTACCAAGCCAGGTAAAAGCAAAGGGGCTATTTTTGCGTATCGCTCAAAA
  TCAAGCATGATTGGCGGTCGTGGTGTTGTTCTGACTTCCGAGGAAGCGAT
  TCAAGAAAATCAAGATACATTTACACATTGGACACCCAACGTTTATCGTT
  ATGGAACGTATGCAGACGAAAACCGTTCATACACGAAAGGACATTCTGAA
  AACAATTTAAGACAAATCAATACCTTCTTTATTGATTTTGATATTCACAC
  GGCAAAAGAAACTATTTCAGCAAGCGATATTTTAACAACCGCTATTGATT
  TAGGTTTTATGCCTACTATGATTATCAAATCTGATAAAGGTTATCAAGCA
  TATTTTGTTTTAGAAACGCCAGTCTATGTGACTTCAAAATCAGAATTTAA
  ATCTGTCAAAGCAGCCAAAATAATTTCGCAAAATATCCGAGAATATTTTG
  GAAAGTCTTTGCCAGTTGATCTAACGTGTAATCATTTTGGTATTGCTCGC
  ATACCAAGAACGGACAATGTAGAATTTTTTGATCCTAATTACCGTTATTC
  TTTCAAAGAATGGCAAGATTGGTCTTTCAAACAAACAGATAATAAGGGCT
  TTACTCGTTCAAGTCTAACGGTTTTAAGCGGTACAGAAGGCAAAAAACAA
  GTAGATGAACCCTGGTTTAATCTCTTATTGCACGAAACGAAATTTTCAGG
  AGAAAAGGGTTTAATAGGGCGTAATAACGTCATGTTTACCCTCTCTTTAG
  CCTACTTTAGTTCAGGCTATTCAATCGAAACGTGCGAATATAATATGTTT
  GAGTTTAATAATCGATTAGATCAACCCTTAGAAGAAAAAGAAGTAATCAA
  AATTGTTAGAAGTGCCTATTCAGAAAACTATCAAGGGGCTAATAGGGAAT
  ACATTACCATTCTTTGCAAAGCTTGGGTATCAAGTGATTTAACCAGTAAA
  GATTTATTTGTCCGTCAAGGGTGGTTTAAATTCAAGAAAAAAAGAAGCGA
  ACGTCAACGTGTTCATTTGTCAGAATGGAAAGAAGATTTAATGGCTTATA
  TTAGCGAAAAAAGCGATGTATACAAGCCTTATTTAGTGACGACCAAAAAA
  GAGATTAGAGAAGTGCTAGGCATTCCTGAACGGACATTAGATAAATTGCT
  GAAGGTACTGAAGGCGAATCAGGAAATTTTCTTTAAGATTAAACCAGGAA
  GAAATGGTGGCATTCAACTTGCTAGTGTTAAATCATTGTTGCTATCGATC
  ATTAAAGTAAAAAAAGAAGAAAAAGAAAGCTATATAAAGGCGCTGACAAA
  TTCTTTTGACTTAGAGCATACATTCATTCAAGAGACTTTAAACAAGCTAG
  CAGAACGCCCTAAAACGGACACACAACTCGATTTGTTTAGCTATGATACA
  GGCTGAAAATAAAACCCGCACTATGCCATTACATTTATATCTATGATACG
  TGTTTGTTTTTTCTTTGCTGTTTAGCGAATGATTAGCAGAAATATACAGA
  GTAAGATTTTAATTAATTATTAGGGGGAGAAGGAGAGAGTAGCCCGAAAA
  CTTTTAGTTGGCTTGGACTGAACGAAGTGAGGGAAAGGCTACTAAAACGT
  CGAGGGGCAGTGAGAGCGAAGCGAACACTTGATTTTTTAATTTTCTATCT
  TTTATAGGTCATTAGAGTATACTTATTTGTCCTATAAACTATTTAGCAGC
  ATAATAGATTTATTGAATAGGTCATTTAAGTTGAGCATATTAGAGGAGGA
  AAATCTTGGAGAAATATTTGAAGAACCCGATTACATGGATTGGATTAGTT
  CTTGTGGTTACGTGGTTTTTAACTAAAAGTAGTGAATTTTTGATTTTTGG
  TGTGTGTGTCTTGTTGTTAGTATTTGCTAGTCAAAGTGATTAAATAGAA*G*
  *ATTTAAAGAAAATCCCATAAAAGACACTTGAATTATTGCTTCAAGTGTCT*
  *TTTGCTTATTTAAAAATCTTTCGCATAATGTGTTTTAGTAAATCAGTAAC*
  *TAACATGGCTCTAAAACATGTGCCGCCTTTACGGATTATATGTGCCAGTT*
  *TTGGAACCATTCGAAACAACACAGCTCTAAAACGAGACCTTTGAGCTTCC*
  *GAGACTGGTCTCAGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACC*
  *TCGTGGACCAATTTTCGTCACGACATTTCGTAATCGCGCCATTCATCTCA*
  *GCTAGAATTCAGCTTCAAATAATTCCTTTTTAAAAGATTTTAATTAATTA*
  *TATTATACCTAAAAAAGTGCTAGAAATCAATCTTTCTAGCGCTTTTCTTG*
  *TTTTATAATTGTAGTTTCCTTTCCAAAGCTTTTGTTAATACTTCAGAATA*
  *ATTCACTTGATCACGATCTGCTAATTGTACTAGATTAGTCTCCTCCTAAC*
  *TTACTGAGATCAATCCGCGTTTCGTAAAGACCGGTGATGGATTGGTGGAT*
  *GAGGGTAGCCTCCATACATGGTAAAAGATAATCTTTCTTACCTGTATATT*
```

GTTTTTGATTAAGTTTTATCCCTAAAAAGCTAAATGGAGAGGTCGCTCCC
AATTGAGTAAATCCTAATAATTTAATAAAACCATCAGCTAATTCCTCAAT
TGAAGCAGAATCCCGCTCCTTCGAATAAGCTTCTTCAATTTTTTGAATGT
GCTCTTTCCCTAGTTTACATTTTTTAGAAAAAGAAGTAATTTCATTAAAC
AAAATATCAAATTGTTGATAATGATTTTTAACATATTCATTACTATAATC
GTTGTCTAAATGATGAGCATGATAAAGTAAATCTTGAGATTTTTTAGACA
CAACCAATTGATTTCCTTTATGTACTTCTTTTGAGCTAGCCCATAAACGT
TTAATGCCATTGCCAATATCTACTAGAGTGTATTTAGGTAATTTAACAAA
GTTATTCTTTTCTATCTGTTGATAACCTTTATCTTTCAAAAATTTAACAG
GATTTTGTTCAAATCGTTTCTTATCCATTACTGAAATTGGAATTAACTGT
TTTGTATCTACTATTAATAAAATTGGATAAGCTGTTGTTGGTTTTTGATA
GCCTCCATATTTCTCAGGATTTAATGCTTTTTTTAGTGGAACAAGCTTAC
TAGGTATAACCTCTAATGGCGATTTAGGATTATCATCAAAAAGCCCCCCA
TTTTGTCCAATGGTTTGTTCCTCCACTTTCTTAACAATATTGACTTGTGG
ATAAGAAAGAACTTTTTTGACAGTAGTGATATGCTTTGTTTTATCCCAAG
CAATTTCTCCTGTTTCATCATTTACTTCAATCATCGGACGTTCGACAATT
TGATCATCTGCCAATTTGACTTTGGACTTGAACATATTCATGATATTAGA
GTAGAAAAACATCTTTTGCGTTGCTGTTTTACGCTCACGATAGCTATTGT
ATTTTGGATATTCACCGTAAACAAATTCTGGTTCCAATTTTGGATATTTT
ACGAGAAGAGCTTTGACAACAACCGCGTTAAGATAAGCATCATGTGCGTG
GTGATAATCGTTAATTTCACGGACTTTATAAAGTTCAAATTCTTTTCGGA
AAGATGACACAAGATTTGATTTCAAAGTGACAATTTTGACCCTGCGAATA
CGTTTATTATTGTCATCAAATTCTTTATTAAAGCGTTCATCAAGCATACG
TGCCACGTGTTTAGTGATTTGTCTGGTTTCAACCAATTGGCGTTTAATAA
AGCCAGCCTTATCATCTGGTGTCAATTCTTTCTTGGTTAGATTGTTATAT
TTCCGTTGAGAAATCAAGCCACTACTCAACAACTTATTCCAAAACGGACG
CATCCTATTGACAACATCTTCACTTGGTACATCATCCGATTTTCCACGAT
TTTCCTTTGAGCTAGTCAATACTCTATTATCAATAGAATTATCCTTGATA
AAAGCTTGCGGGATAATATGGTCTATATCATACTGGCTTAGATAATCAAT
ATCCAATTCTTCTCCAGTATACATATCTCTGCCGTTTTGTAAATAATATA
GAAACAATCTATCATTTTGTAACTGTGAATTCTCAACCGGATGTTCTTTA
AGAATTTGACTTCCAAATTCTTTAATAGAATCTGTCAAACCTTTCAAACG
CTGCTGTGAATTTCGTCGTCCCTGATTGGTAAACTGGTTTTCACGCGCCA
TCTCCACGACGATATTTTCAGGTTGATGTCCCATAATTTTGACAAGCTCA
TCAACAATCTTCAAGCTTTGTAAAATTCCTTTTTTAATAGCAGGGCTGCC
AGCAATATCACTAACAACTTGATTTAGATTGTCTGTTTCTCCAATAACTT
GTGCCTTAGCAATCTCTTCTTTGAAAGAAAGAGCATCATCGTTAATCAGT
TGCATAAAGTTACGATTGCTATTGCCATCATCAATGAGATAATCAAGAAT
TGTTTTTCTGCTTTCTTTGTTGCGAATACCATGAATTAACTCAGCTGATA
ATCTTCCCCAACCAGTATAATGACGTCTTTCCAGCTTTTTCACTTGTTCT
TTGGTCAATAAATCACTGTAATTTTCTAGACGTTTTCTAATCATTTCTCT
ATCTTCAAATAACGTTAAGGTCAACACAATATCTTCTAAAATCTTTTCAT
TCTTTGAATTATCGAGAAAATCTTTATCTAAAATTTTACGCAAATCATGA
TAAGTTCCATAAGAAGCGTTAAATGCTTTATTTTCTTTATCCAGACCTGT
TAAATCAACAATACGAAATTCATCAAATTCTTTTTCAAGGAAATCCATTA
ATTTATCTTTAGTTACTTTTCGATAAACCTTAAATACGCCATCAAAGATT
TCTTGCTTCATATTGGCATCAAAAAATGCTGTTTTTCCTTGCTCTGTTTT
ATATTTAACCTTTGTTAATTCATTGTAAACAGTAAATTTTTCGTATAATA
AACTATGTTTAGGAAGAACTTTTTGATTTGGCAAGTACAAATCATAATTT
GTCATACGATTGATAAAAGCTTCTGCAGAGGATTCTTTATCAACGATTTC
ATCAAAATTCCATGGTGTAATTTTATCAGCCGATTTCCGACTTAACCAAG
CAAAATCACTTTTTCCGCGCGCTAATGGACCAACATAGTAGGGAATACGG
AAAGTCAATATTTTCTCAATCCTATCTTGATTGTCTGCTAAAAACGGATA
AAATTCAGCCTGTCTACGAATGATAGCACGCATTTCTTGAAGATGAATCT
GATGTGGAATAGAGCCATTGTCAAAGGTACGTTGCTTTCTTAGAAAATCT
TCACGCTCAATTTTATCAAGGAAATAGCCACTTCCCTCAATCTTATTTAA
TAGACCTTTAAGGTATTTATAAAAAGCTTCTTGATTTGTTTTCCCATCAA

TATAACCCGCATAGCCGTCTTTTGAAACATCAGAAAAAACTTCGTTATAT
TTATCTGATAATTTCTGACGAATGAATTGTTTAAGCTGAGCTAAATCCAT
CTGATGTTCATTATATCGCTGAATCATCGAAGCAGATAAAGGCGCTTTGG
TACTAACATCAGTAACTGTTAAAATCCCTGATAAAAGGATACTATCATAC
AGTTTCTTTGCTGATAAAAAGAGCTCTGCGTAATTATCTCCAATTTGAGC
TAATAGTACTTCTAACTCTTCTTCATAAGTATCTTTAGAAAATTGCAATG
GTGTTTTCTCTTCTAATTCAAAATGCTTTTTAAAATCAGCTTGATTACCA
ACAATTAGTTTTAGAAATTCTGCAAAGCGGCCATTAGACTTTTCATTAGG
AAAAAGTTTCAAAACTCTATCTTTCTTAGCAGACTTACTGATTTTATCAG
TCAGAATTTCCTCAACTTGAACATTTTGCTCCTGAAGCGAACTATTCTCA
AAAGTATTATCATAGACTGCTAAAAATTCTTGAAACAGTCTTTGTACATC
ATTATTGCGTGTATCAAACTTTCCTTCAATTAAAAAATGGCCTCTAAACT
TAATTATATGTGCCAAAGCCAAATAAACTAAACGCAAATCAACCTTTTCT
GGATTATCCGCAAGATATTGCCGCAAATGATAAATGGTTGGAAAATTTTC
ATGATACTTAACTTCTTCTTCAAGATTCCCAAAAATGGGATGGCGCTCTC
CTCGTTTATCCTCAGTAACAAGAAAAGAATCCTCTAAACGATGAAAGAAA
CTATCATCTACCTTGCTCATTTCTTCTGCAAAAATCTCTTGCAAATATAA
AATACGATTTCTACGACGTGTATAACGACGGCGAGCTGTACGCTTTAAAC
GTCTATCTGCTGCAGTATTCCCGCTATCAAATAATAAAGCGCCAAGCAAA
TTTTTCTTGATATGACTTTTATCTGTATTTCCCAGAACCTTCATCTTCTT
AGCAGGAACCTTGTAGTCATCTGTCACAACAGCCCAACCAACAGAATTGG
TTCCAATATCAAGTCCAATAGAGTAAGGTTTTTTCATAAATGTCTCCTTT
GAAAAAATGTCCTTGCAAAAAAGCAATAAGTATTATACAATAATTGTGTT
GGAACTATTCGAAACAACACAGCAAGTTAAAATAAGGTTTATCCGTATTC
AACTTGAAAAAGTGCGCACCGATTCGGTGCTTTTTTATTTGCTTTACTGT
AAGCGTTTCTATTATACCACTATCTACGAAATATTTTAACAATATAAACG
*AAAAAATATAGAAATATTTCTGTATTTTTTGGG*CCAGTGAATTCCCGGGG
ATCCGTCGACCTGCAGCCAAGCTTTCGCGAGCTCGAGATCTAGATATCGA
TGAATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCG
CTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTG
GGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGC
CCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGC
CAACGCGCGGGGAGAGGCGGTTTGCGTATTGGAATTCCGGATGAGCATTC
ATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATT
TTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGT
TATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGA
TGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCAT
TTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCG
GTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGA
TCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAAC
AGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTA
TTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGCCAACTTACTGATTTAG
TGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGCT
GTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAAAGCACC
GCCGGACATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTGGCACT
GATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCA
CCGGTGCGTCAGCAGAATATGTGATACAGGATATATTCCGCTTCCTCGCT
CACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTT
ACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGG
GAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCT
GACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGAC
AGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCT
CTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCG
CGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCT
CCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCC
TTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGC
ACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAG

TCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGC
TCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCT
TCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATT
ACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAA
ATATTTCTAGACTCGAGAGCTCTCGAGTCTAGAATCGATACGATTTTGAA
GTGGCAACAGATAAAAAAAAGCAGTTTAAAATTGTTGCTGAACTTTTAAA
ACAAGCAAATACAATCATTGTCGCAACAGATAGCGACAGAGAAGGCGAAA
ACATTGCCTGGTCGATCATTCATAAAGCAAATGCCTTTTCTAAAGATAAA
ACGTATAAAAGACTATGGATCAATAGTTTAGAAAAAGATGTGATCCGTAG
CGGTTTTCAAAATTTGCAACCAGGAATGAATTACTATCCCTTTTATCAAG
AAGCGCACAAAAAGAAAAACGAAATGATACACCAATCAGTGCAAAAAAAG
ATATAATGGGAGATAAGACGGTTCGTGTTCGTGCTGACTTGCACCATATC
ATAAAAATCGAAACAGCAAAGAATGGCGGAAACGTAAAAGAAGTTATGGA
AATAAGACTTAGAAGCAAACTTAAGAGTGTGTTGATAGTGCAGTATCTTA
AAATTTTGTATAATAGGAATTGAAGTTAAATTAGATGCTAAAAATTTGTA
ATTAAGAAGGAGTGATTACATGAACAAAAATATAAAATATTCTCAAAACT
TTTTAACGAGTGAAAAAGTACTCAACCAAATAATAAAACAATTGAATTTA
AAAGAAACCGATACCGTTTACGAAATTGGAACAGGTAAAGGGCATTTAAC
GACGAAACTGGCTAAAATAAGTAAACAGGTAACGTCTATTGAATTAGACA
GTCATCTATTCAACTTATCGTCAGAAAAATTAAAACTGAATACTCGTGTC
ACTTTAATTCACCAAGATATTCTACAGTTTCAATTCCCTAACAAACAGAG
GTATAAAATTGTTGGGAGTATTCCTTACCATTTAAGCACACAAATTATTA
AAAAAGTGGTTTTTGAAAGCCATGCGTCTGACATCTATCTGATTGTTGAA
GAAGGATTCTACAAGCGTACCTTGGATATTCACCGAACACTAGGGTTGCT
CTTGCACACTCAAGTCTCGATTCAGCAATTGCTTAAGCTGCCAGCGGAAT
GCTTTCATCCTAAACCAAAAGTAAACAGTGTCTTAATAAAACTTACCCGC
CATACCACAGATGTTCCAGATAAATATTGGAAGCTATATACGTACTTTGT
TTCAAAATGGGTCAATCGAGAATATCGTCAACTGTTTACTAAAAATCAGT
TTCATCAAGCAATGAAACACGCCAAAGTAAACAATTTAAGTACCGTTACT
TATGAGCAAGTATTGTCTATTTTTAATAGTTATCTATTATTTAACGGGAG
GAAATAATTCTATGAGTCGCTTTTGTAAATTTGGAAAGTTACACGTTACT
AAAGGGAATGTAGATAAATTATTAGGTATACTACTGACAGCTTCCAAGGA
GCTAAAGAGGTCCCTAGCGCTTAGAATCGCTTTAGGAAACACGATCCAGT
CCAATAATCGTCGATAAAAACTTTTGAAAAAGGTTGGTGAAATTACCTAC
TTTTGGAATAATCACAAATCACAAGTGATTAATCACAAATCACAAGTGAT
TAATCACTTGTTTATTAAGATATTAAAAGCTATAATTTAAATAAAGCGTG
AATTTTATTACACAAAAAGAGGGGGGAGAAACTTGGAACTAGCATTTAGA
GAAAGCTTAAAAAAGATGAGAGGTACCAAATCAAAAGAAAAATTCTCCCA
AGAATTAGAAATGAGTAGATCAAATTATTCACGAATAGAATCAGGAAAAT
CAGATCCAACCATAAAAACACTAGAACAAATTGCAAAGTTAACTAACTCA
ACGCTAGTAGTGGATTTAATCCCAAATGAGCCAACAGAACCAGAACCAGA
AACAGAATCAGAACAAGTAACATTGGATTTAGAAATGGAAGAAGAAAAAA
GCAATGACTTCGTGTGAATAATGCACGAAATCGTTGCTTATTTTTTTTAA
AAGCGGTATACTAGATATAACGAAACAACGAACTGAATAGAAACGAAAAA
AGAGCCATGACACATTTATAAAATGTTTGACGACATTTTATAAATGCATA
GCCCGATAAGATTGCCAAACCAACGCTTATCAGTTAGTCAGATGAACTCT
TCCCTCGTAAGAAGTTATTTAATTAACTTTGTTTGAAGACGGTATATAAC
CGTACTATCATTATATAGGGAAATCAGAGAGTTTTCAAGTATCTAAGCTA
CTGAATTTAAGAATTGTTAAGCAATCAATCGGAAATCGTTTGATTGCTTT
TTTTGTATTCATTTATAGAAGGTGGAGTTTGTATGAATCATGATGAATGT
AAAACTTATATAAAAAATAGTTTATTGGAGATAAGAAAATTAGCAAATAT
CTATACACTAGAAACGTTTAAGAAAGAGTTAGAAAAGAGAAATATCTACT
TAGAAACAAAATCAGATAAGTATTTTTCTTCGGAGGGGGAAGATTATATA
TATAAGTTAATAGAAAATAACAAAATAATTTATTCGATTAGTGGAAAAAA
ATTGACTTATAAAGGAAAAAAATCTTTTTCAAAACATGCAATATTGAAAC
AGTTGAATGAAAAAGCAAACCAAGTTAATTAAACAACCTATTTTATAGGA
TTTATAGGAA