JYU DISSERTATIONS 279

Ville Hoikkala

Memoirs of a Fish Pathogen

How Crispr-Cas Captures Phage Encounters in *Flavobacterium Columnare*

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF MATHEMATICS
AND SCIENCE

JYU DISSERTATIONS 276

Ville Hoikkala

Memoirs of a Fish Pathogen

How Crispr-Cas Captures Phage Encounters in *Flavobacterium Columnare*

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella julkisesti tarkastettavaksi syyskuun 26. päivänä 2020 kello 12.

Academic dissertation to be publicly discussed, by permission of the Faculty of Mathematics and Science of the University of Jyväskylä, on September 26, at 12 o'clock noon.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2020

Editors
Varpu Marjomäki
Department of Biological and Environmental Science, University of Jyväskylä
Timo Hautala
Open Science Centre, University of Jyväskylä

.

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-8264-5

ISBN 978-951-39-8264-5 (PDF)
URN:ISBN:978-951-39-8264-5
ISSN 2489-9003

ABSTRACT

Hoikkala, Ville
Memoirs of a fish pathogen: how CRISPR-Cas captures phage encounters in *Flavobacterium columnare*
Jyväskylä: University of Jyväskylä, 2020, 66 p.
(JYU Dissertations
ISSN 2489-9003; 276)
ISBN 978-951-39-8264-5
Diss.

Bacteria are in constant interaction with their viruses, bacteriophages (phages). To prevent or abort phage infections, a variety of defence mechanisms have evolved. CRISPR-Cas, the only known adaptive bacterial immune system, targets intracellular phage genomes by utilizing genetic memories of past infections. A memory is formed during CRISPR adaptation when a fragment of a phage genome is integrated into a CRISPR array on the bacterial genome. This fragment, called a spacer, is later used in the interference phase to recognize and cleave phage genomes with matching sequences. While these core principles are shared by most CRISPR-Cas systems, many subtypes have not been thoroughly explored, especially in their native hosts or in natural environments. In this thesis, I characterize type II-C and VI-B CRISPR-Cas systems in the fish pathogen *Flavobacterium columnare* in association with its virulent phage. The first study describes how aquaculture settings can be harnessed for coevolutionary studies in semi-natural settings using bacteria that carry CRISPR-Cas loci. The second study shows *F. columnare* and its virulent phages evolving for several years in aquaculture, where coevolutionary dynamics were reflected by spacer acquisition in bacteria and genomic and host-range changes in phages. The third study examines the adaptation process of the II-C and VI-B CRISPR-Cas systems in the laboratory. The RNA-targeting VI-B locus was dependent on the spacer acquisition machinery of the II-C locus, leading to characteristic interference patterns for both loci. The fourth study shows how the presence of eukaryotic host signals accelerates spacer acquisition, suggesting that environmental determinants play important roles in phage defence strategies. Together, these studies show that type II-C and VI-B CRISPR-Cas systems are active in natural and laboratory conditions, driving coevolution between *F. columnare* and its virulent phages. Understanding native functioning of CRISPR-Cas is also important for practical applications such as phage therapy.

Keywords: Adaptation; bacteria; bacteriophage; coevolution; CRISPR-Cas; immunity.

Ville Hoikkala, University of Jyväskylä, Department of Biological and Environmental Science, P.O. Box 35, FI-40014 University of Jyväskylä, Finland

TIIVISTELMÄ

Hoikkala, Ville
Kalapatogeenin muistelmat: CRISPR-Cas -immuunijärjestelmän rooli *Flavobacterium columnaren* ja sen faagien vuorovaikutuksessa.
Jyväskylä: Jyväskylän yliopisto, 2020, 66 s.
(JYU Dissertations
ISSN 2489-9003; 276)
ISBN 978-951-39-8264-5
Diss.

Bakteerit ovat jatkuvassa vuorovaikutuksessa niitä tartuttavien virusten, bakteriofagien (faagien) kanssa. Faagitartuntojen ehkäisemiseksi on kehittynyt laaja valikoima puolustusmekanismeja. CRISPR-Cas on adaptiivinen immuunijärjestelmä, jonka toiminta perustuu menneistä tartunnoista kerättyihin muistijälkiin. Muistijäljet kehittyvät immuunivasteen adaptaatiovaiheessa, jossa solun sisälle päässeen faagin perimästä leikataan osa, joka asetetaan CRISPR-lokukseen niin kutsuttuna välijaksona. Tätä välijaksoa hyödynnetään myöhemmin saman faagiperimän tunnistamiseen ja tuhoamiseen. Erilaiset CRISPR-Cas -järjestelmät perustuvat jaettuihin perusperiaatteisiin, mutta eroavat yksityiskohtaisilta mekanismeiltaan, eikä monia vähemmän tunnettuja järjestelmiä ole tutkittu lainkaan niiden alkuperäisessä isäntäbakteerissa tai luonnollisissa olosuhteissa. Tässä väitöskirjassa selvitän CRISPR-Cas alatyyppien II-C ja VI-B toimintaa kalapatogeeni *Flavobacterium columnaren* ja sen faagien yhteydessä. Ensimmäisessä tutkimuksessa kartoitan millä tavoin kalanviljelylaitoksia voidaan hyödyntää yhteisevoluutiotutkimuksessa ja mitä CRISPR-Cas-järjestelmiä niissä esiintyvät bakteerilajit kantavat. Toisessa tutkimuksessa selvitän kuinka *F. columnare* ja sen faagit kehittyvät näissä olosuhteissa usean vuoden ajan. Kolmannessa tutkimuksessa CRISPR-Cas:n evoluutiota seurattiin laboratorio-olosuhteissa, mikä paljasti molemmille lokukselle omaleimaiset adaptaatiomallit sekä riippuvuussuhteen näiden kahden lokuksen välille. Neljännessä tutkimuksessa osoitin pitkällä evoluutiokokeella, kuinka bakteerin isännästä viestivät signaalit vaikuttavat puolustusjärjestelmien valintaan faagien läsnä ollessa. Tyypin II-C ja VI-B CRISPR-Cas -järjestelmät ovat toiminnallisia sekä luonnossa että laboratoriossa ja vaikuttavat olennaisesti *F. columnaren* ja faagien yhteisevoluutioon. Näiden vuorovaikutusten ymmärtäminen on tärkeää esimerkiksi faagien lääkekäytön, eli faagiterapian näkökulmasta.

Avainsanat: Adaptaatio; bakteeri; bakteriofagi; CRISPR-Cas; faagi; immuunipuolustus; yhteisevoluutio.

Ville Hoikkala, Jyväskylän yliopisto, Bio- ja ympäristötieteiden laitos PL 35, 40014 Jyväskylän yliopisto

Author

MSc. Ville Hoikkala
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland
ville.hoikkala@jyu.fi

Supervisors

Assoc. prof. Lotta-Riina Sundberg
Department of Biological and Environmental Science
P.O. Box 35
FI-40014 University of Jyväskylä
Finland

University lecturer Janne Ravantti
Molecular and Integrative Biosciences
P.O. Box 65
FI-00014 University of Helsinki
Finland

Reviewers

Prof. Peter Fineran
Department of Microbiology and Immunology
P.O. Box 56
NZ-9054 University of Otago
New Zealand

Asst. prof Kimberley Seed
Department of Molecular Biology & Microbiology
US-94720 University of California, Berkeley
United States

Opponent

Prof. Edze Westra
Environment and Sustainability Institute
UK-TR10 9FE University of Exeter
United Kingdom

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The thesis is based on the following original papers, which will be referred to in the text by their Roman numerals I–IV.

I Ville Hoikkala, Gabriel De Freitas Almeida, Elina Laanto & Lotta-Riina Sundberg 2019. Aquaculture as a source of empirical evidence for coevolution between CRISPR-Cas and phage. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374: 20180100.

II Elina Laanto, Ville Hoikkala, Janne Ravantti & Lotta-Riina Sundberg 2017. Long-term genomic coevolution of host-parasite interaction in the natural environment. *Nature Communications* 8: 931.

III Ville Hoikkala, Janne Ravantti, César Díez-Villaseñor, Marja Tiirola, Rachel A. Conrad, Mark J. McBride, Lotta-Riina Sundberg 2020. Cooperation between CRISPR-Cas types enables adaptation in an RNA-targeting system. Submitted manuscript.

IV Gabriel Almeida[1], Ville Hoikkala[1], Janne Ravantti & Lotta-Riina Sundberg 2020. Ecological determinants of phage defence strategy in an opportunistic pathogen. Manuscript.

[1] Shared first authorship

Responsibilities of the author:

I Conceptualization, literature research, writing, characterizing CRISPR-Cas profiles of selected pathogens

II Study design, characterizing CRISPR-Cas loci and spacer sequences, illustrating results, statistics, adsorption laboratory work, writing

III Study design, creating spacer acquisition protocol, performing experiments, writing software and analysis pipelines, statistics, illustrating results, writing

IV Data analysis, statistics, illustration, writing

ABBREVIATIONS

BAM	Bacteriophage adherence to mucus model
crRNA	CRISPR-RNA
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Cas	CRISPR-associated protein
PAM	Protospacer adjacent motif
PFS	Protospacer flanking site
SAM	Spacer acquisition motif
TIM	Target interference motif

1 INTRODUCTION

Prokaryotes are a diverse group of micro-organisms that occupy even the most inhospitable niches on Earth. Almost anywhere they settle, they are accompanied by a faithful companion: the bacteriophage. Bacteriophages, or phages, are prokaryotic viruses that outnumber their hosts up to 15-fold (Suttle 2007), making them the most abundant biological entity on the planet. During their roughly four-billion years of coexistence (Cole 2016), a staggering amount of interactions between prokaryotes and their viral parasites has occurred. It is no wonder then, that much of prokaryotic evolution is shaped by phages, especially in the immunological defense systems that keep phages at bay. An example a defense is surface modification, where phage entry is prevented by mutating of cell surface receptors (Hampton *et al.* 2020). This and most other defense mechanisms are "innate", where pathogen recognition is based on characteristics common to many invaders. Recognition and destruction of specific invaders, adaptive immunity, was long thought to be exclusive to eukaryotes. In 2005, however, a DNA-based memory system that enables the cell to identify and destroy "memorized" phage genomes was discovered (Mojica *et al.* 2005, Pourcel *et al.* 2005, Bolotin *et al.* 2005). This memory system was termed Clustered Regularly Interspaced Short Palindromic Repeats, and was also found to be associated with nearby *cas*-genes (Jansen *et al.* 2002). Together, these components were termed CRISPR-Cas – an acronym that now embodies an entire field of bacterial immunology with strong connotations in gene editing technology.

While the last 15 years of CRISPR-Cas research has revealed the workings of this multi-faceted immune system in astonishing detail, many questions still lie unanswered. Surprisingly few prokaryotic species have allowed investigation of their native CRISPR-Cas systems, as many species that carry these systems are unculturable or lack established protocols in the laboratory (Hynes *et al.* 2017). In this thesis, I explore CRISPR-Cas systems in the fish pathogen *Flavobacterium columnare* in association with its virulent phages. This pathogen has been extensively studied due to its important role as a causative agent of columnaris disease in freshwater fish, but its CRISPR-Cas systems have

received little attention. Owing to the large body of literature and pre-established laboratory protocols, the study of type II-C and VI-B CRISPR-Cas in this native host became possible. Here, I investigate these systems with evolutionary, ecological as well as mechanistic viewpoints and draw conclusions on both general and *F. columnare*-specific levels. I hope this thesis to lay foundation for further studies on these intriguing phage defenses.

2 REVIEW OF THE LITERATURE

2.1 Prokaryotes and their viral parasites

Bacteriophages (phages) are obligate prokaryotic viruses that lack the metabolic machinery required for self-contained replication. The outside-host (extracellular) form of a phage, the virion, is an immobile protein capsule that encloses the DNA or RNA phage genome (Guttman *et al.* 2005). Upon encountering its host, the virion attaches to a specific protein (or other structure such as a lipopolysaccharide) on the cell surface in a process called adsorption. Following adsorption, the phage inserts its genome into the cytoplasm of the host (Guttman *et al.* 2005).

Phages are generally categorized as virulent or temperate. Virulent phages hijack the host's cellular machinery to produce copies of its genome as well as the structural components necessary for assembling new virions (Guttman *et al.* 2005). After producing enough progeny, the host cell is typically lysed, releasing virions that begin the search for a new host. In contrast, a temperate phage does not begin immediate replication, but instead integrates into the host genomes as a prophage. A host cell that carries a prophage is called a lysogen. Prophages often provide advantages to the host by, for example, increasing their virulence, providing phage resistance or lysing competitor cells (Howard-Varona *et al.* 2017). However, some cues such as starvation may trigger prophages to switch to a lytic cycle, eventually leading to phage proliferation and host death (Guttman *et al.* 2005).

The discovery of phages more than a hundred years ago caused a paradigm shift in microbiology (Twort 1915, d'Herelle 1917). The medical potential of these "invisible microbes" in treating bacterial disease was realized soon after (d'Hérelle 1921). Using phages to treat bacterial infections, phage therapy, became a revolutionary treatment that prospered until the 1940's (Almeida and Sundberg 2020). However, the discovery of penicillin began another revolution in western medicine and the simplicity of antibiotics (among other factors, see Fruciano and Bourne 2007) steered clinical and scientific

interest away from phage therapy. However, increasing antibiotic resistance in bacteria during the last few decades (Theuretzbacher 2013) has led to resurgence of phage therapy and basic phage research. As antibiotics are losing their potency, it is now perhaps more crucial than ever to understand how bacteria and phages interact and coevolve, specifically from the viewpoint of phage resistance.

2.1.1 Bacterial defence mechanisms

Bacteria have evolved to hamper or abort phage infections using a variety of defence mechanisms (Hampton *et al.* 2020), many of which are just being uncovered (Ofir *et al.* 2018, Doron *et al.* 2018). The first line of defence occurs in the extracellular space on the bacterial cell wall, where phages need to adsorb to cell surface proteins to initiate the infection (Fig. 1). Mutating, removing or masking these receptors, collectively known as surface modification (SM), may effectively prevent adsorption. Examples of SM in bacteria include (and is not limited to) the fish pathogen *Flavobacterium psychrophilum* (Castillo *et al.* 2015), the human pathogens *Staphylococcus aureus* (Nordström and Forsgren 1974) and *Vibrio cholerae* (Seed *et al.* 2014).

However, for reasons enclosed in the next chapter, phages may still gain entry to the cytoplasm. Once injected inside, the phage genome receives an unwelcoming reception from intracellular defences, such as the restriction modification system (RM) (Fig. 1). RM is a widespread immune system comprised of a methyltransferase and a restriction endonuclease, which methylate the self-genome and cleave any unmethylated nucleic acids, such as the phage genome (Oliveira *et al.* 2014). Due to their capability to recognize and cleave specific unmethylated sequences, RM has also become an invaluable laboratory tool in molecular biology.

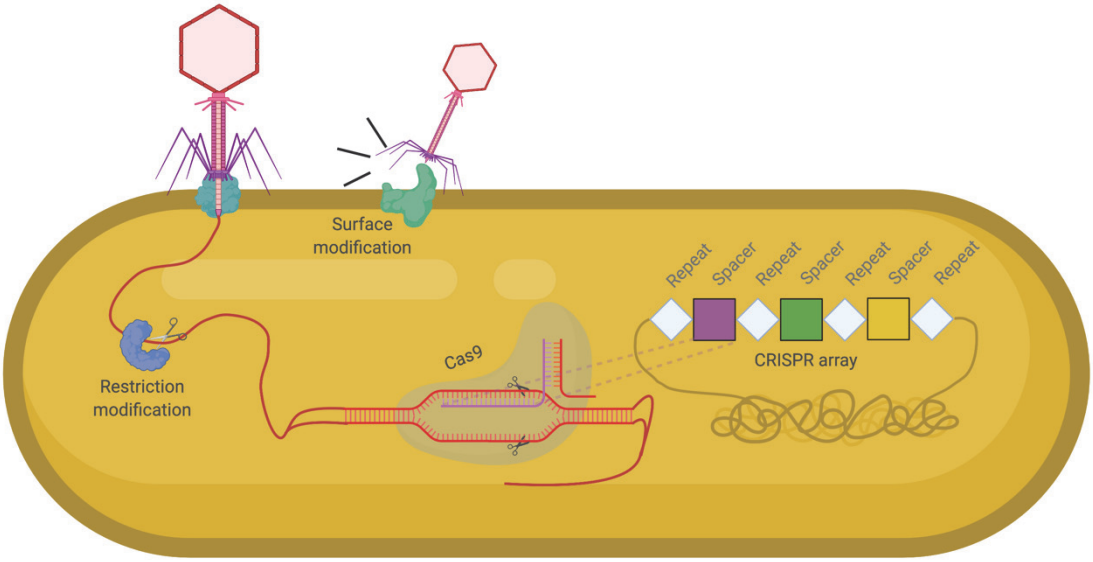

Figure 1. Bacterial defense mechanisms. Three mechanisms that bacteria use to fight phage infections. Surface modification prevents phage entry, while restriction modification and CRISPR-Cas are intracellular defenses.

Intracellular defences include several additional systems, many of which are still poorly understood (Hampton *et al.* 2020). Immune systems are categorized as either adaptive or innate, the latter represented by SM, RM and most other prokaryotic defence mechanisms. Adaptive immunity in prokaryotes is represented solely by a system known as CRISPR-Cas, which is discussed in detail in chapter 2.2, and illustrated in a simplified manner in Fig. 1. It is clear that prokaryotic immune mechanisms are sophisticated and highly diverse. The necessity of these systems is best understood through an evolutionary viewpoint.

2.1.2 Coevolution between phage and bacteria

Evolution results from change of allele frequencies in populations over time, with selection favoring certain configurations over others. When species share ecological interaction that leads to reciprocal adaptation and counter-adaptation, the term coevolution is used (Janzen 1980). Few groups fit this description better than prokaryotes and phages, who have coexisted and interacted for billions of years, diversifying into the variety of groups we see today.

Bacteria generally benefit from avoiding infections by phages, while phages are reliant on bacteria for replication. This relationship has resulted in the evolution of the defence mechanisms in bacteria and in corresponding coevolutionary adaptations in phages. The resulting coevolution can escalate to a seemingly endless arms race in which both parties evolve to overcome one another (Red Queen Hypothesis, see Van Valen 1973). However, the potential of bacteria and phages to evolve is asymmetrical (Lenski 1984, Lenski and Levin 1985). For example, mutational routes to SM in bacteria are numerous, while

phages are required to adapt to the one exact mutation that actually occurs. This asymmetry in evolutionary potential has been predicted to lead to the eventual extinction of the phage (Lenski 1984, Lenski and Levin 1985). Despite these gloomy prospects, phages are still the most abundant organism on the planet, killing 15-40% of oceanic bacteria daily (Suttle 2007). What, then, causes phages not only to persevere, but thrive?

The escalation to an arms race, that should eventually end in the phage extinction, is dampened by fitness loss associated with resistance (Koskella and Brockhurst 2014, Hampton *et al.* 2020). It is crucial to realize that phage adsorption is a secondary, unwanted function of a prokaryotic surface receptor whose primary function could include, for example, movement or metabolism (Koskella and Brockhurst 2014). It is likely no coincidence that phages tend to target receptors imperative to bacterial survival, as any modifications in such structures result in fitness loss for the bacterium (see e.g. Li *et al.* 2017). Therefore, resistance mutations may revert back to their ancestral form in the absence of phage. However, this reversal also increases potential for re-infection by the ancestral phage. The resulting evolutionary cycles are embodied by fluctuating Red Queen dynamics (Brockhurst *et al.* 2014), which predicts that phenotypes emerge, disappear and re-emerge in nature over time. A related and partially overlapping theory called "kill-the-winner" (Winter *et al.* 2010) predicts that predators evolve to infect the most common or active phenotype of prey (the "winner"). This negative frequency-dependent selection thus favours less-common or less active, perhaps resistant, phenotypes. These re-occurring selective sweeps enable the rise and fall of different phenotypes in a fluctuating manner, perpetuating the existence of both predator and prey. Examples of trade-offs between fitness and phage resistance can be found across the prokaryotic biosphere. One example is the human pathogen *Vibrio cholerae,* in which phage resistance causes lowered virulence and transmission capability (Seed *et al.* 2014) (although this pathogen also benefits from prophage-induced increase in virulence, see Waldor and Mekalanos 1996).

2.1.3 Opportunism, virulence and phage resistance

Bacteria that infect other organisms, pathogens, are generally divided into obligates and opportunists (Brown *et al.* 2012). Obligates are dependent on their host organism and are thus unable to survive in outside-host environments. Opportunists, on the other hand, have adapted to environment both outside and inside their focal host (Brown *et al.* 2012). When inhabiting outside-host environments, opportunists are expected to conform to the "sit-and-wait" strategy (Walther and Ewald 2004). According to this hypothesis, opportunists maintain continuous high virulence as they must be prepared to colonize a host upon an encounter, which may be a rare event. An example of such an opportunist is *F. columnare* that is able to survive in lake water for at least five months while still maintaining virulence (Sundberg *et al.* 2014).

The tendency of prokaryotic pathogens to infect their host has spawned its own selection of innate and adaptive immune mechanisms in their multicellular

hosts, metazoans. While metazoan immune systems are well characterized, one overlooked layer of immunity involves the cooperation of phages and metazoans (Barr *et al.* 2013). The surfaces of metazoans are partly or completely covered in a layer called mucus that is made up of glycoproteins, DNA and cellular debris (Cone 2009). Mucus forms a robust defensive layer that acts as a physical barrier against pathogens while also providing habitat and nutrients for commensal bacteria (Cone 2009). According to the bacteriophage adherence to mucus (BAM) model, mucus also actively attracts and binds phages (Barr *et al.* 2013). This binding occurs through the interaction between mucosal glycoproteins and Ig-like receptors on phage capsids (Barr *et al.* 2013), essentially coating the animal with phage particles. Through this arrangement, the metazoan is protected from pathogens by phages, who in turn increase the odds of encountering their own prokaryotic host (Barr *et al.* 2013).

The concentration of phages on metazoan surfaces has important consequences on prokaryotic ecology. To colonize a host, a pathogen must pass the mucus barrier while retaining their capacity to infect the host (i.e. their virulence) (Cone 2009). Virulence is usually associated with specific receptors on the bacterial cell surface, but as stated before, these receptors are often also utilized by phages for adsorption. A conflict emerges when metazoan host signals endorse upholding virulence-associated structures, while phage presence may endorse their modification or downregulation. This conflict may be especially pronounced in opportunistic pathogens, whose chances of encountering their host is small (Sundberg *et al.* 2014). Indeed, in the fish pathogen *Flavobacterium columnare*, it was shown that the presence of mucus during phage exposure makes the cells more virulent and more susceptible to phage infection (Almeida *et al.* 2019).

How bacteria alter their phage resistance strategies when colonization and virulent phage infections overlap has not been thoroughly investigated. Certain intracellular systems, such as RM or CRISPR-Cas, may play a heightened role during colonization, as they might not reduce the pathogen's virulence to the extent that SM does. A recent study tested this idea in the context of bacterial competition and found that CRISPR-Cas is favoured in phage defence when *Pseudomonas aeruginosa* is competing with other species during colonization. Favouring of CRISPR-Cas was attributed to the amplification of SM-based fitness loss during competition (Alseth *et al.* 2019).

2.2 The adaptive immune system CRISPR-Cas

2.2.1 A brief history of CRISPR-Cas

The story of CRISPR-Cas begins in the late eighties with the observation of 29-nt repetitive sequences interspaced by 32-nt spacing sequences in an *Escherichia coli* genome (Ishino *et al.* 1987). Similar repetitive regions with different sequence than those in *E. coli* were observed later in *Halofelax mediterranei*

(Mojica *et al.* 1993, 1995) and other species (Mojica *et al.* 2000). In some species, the spacing sequences between the repeats (later termed spacers) were found to be useful in distinguishing strains that were otherwise genetically indistinguishable. This method of differentiation was termed spoligotyping and became an essential tool in identifying strains of the human pathogen *Mycobacterium tuberculosis* (Groenen *et al.* 1993, Demay *et al.* 2012).

Despite the common prevalence of these repetitive regions, and the presence of genes that seemed to co-occur with them (Jansen *et al.* 2002) (later termed *crispr associated*, or *cas* genes), their core functions remained unknown. Three separate studies published in 2005 found that spacers originate from phages or plasmids, and likely represent a prokaryotic immune system that can prevent the proliferation of such mobile genetic elements (MGEs) (Mojica *et al.* 2005, Pourcel *et al.* 2005, Bolotin *et al.* 2005). Two years later, the immunological capabilities of CRISPR-Cas was confirmed in a landmark study showing *Streptococcus thermophilus* becoming phage resistant through the acquisition of phage-matching spacer sequences (Barrangou *et al.* 2007). Soon after, it was shown that CRISPR-Cas can also block horizontal gene transfer when targeting plasmids (Marraffini and Sontheimer 2008). What followed was an avalanche of research that aimed to unravel the biochemical, genetic, ecological and evolutionary features of CRISPR-Cas. The increasing interest in CRISPR-Cas is partly fuelled by biotechnological tools derived from *cas* genes or other CRISPR components. These tools range from gene editing (Jinek *et al.* 2012, Cong *et al.* 2013) to diagnosis of disease-causing agents (Kellner *et al.* 2019).

2.2.2 Components of CRISPR-Cas

CRISPR-Cas loci are composed of several genetic components, including genes that code for Cas proteins, and various coding and non-coding regions. A defining feature of a CRISPR-Cas locus is the repeat-spacer array. This region is composed of identical sequences called repeats interspaced by unique sequences called spacers. Repeats are 20-47 nt long (Grissa *et al.* 2007) and usually palindromic in sequence; spacers are roughly similarly sized and form the basis of immunological memory in CRISPR-Cas (Mojica *et al.* 2005, Pourcel *et al.* 2005, Bolotin *et al.* 2005). New spacers are generally added to one end of the array, which is characterized by its adjacency to an AT-rich sequence called the leader (Yosef *et al.* 2012). Flanking the repeat-spacer array are CRISPR associated (*cas)* genes, the identity of which depend on the type of CRISPR-Cas system. These genes code different proteins that facilitate the acquisition of new spacers, process the repeat-spacer array transcripts (crRNA) or degrade incoming nucleic acids. Some CRISPR-Cas systems also contain tracrRNA – non-coding RNA that is essential for functional interference (Deltcheva *et al.* 2011). An example of a subtype II-C CRISPR-Cas locus is illustrated in Fig. 2. How these components contribute to CRISPR-Cas defense will be discussed below.

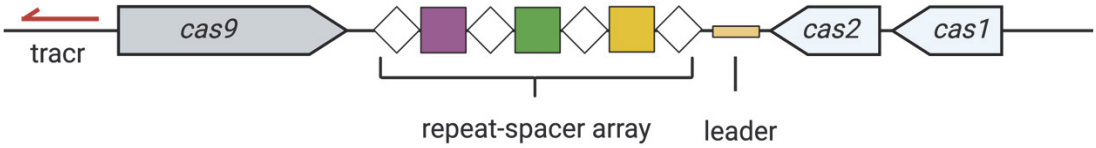

Figure 2. Components of a type II-C CRISPR-Cas locus. This particular arrangement shows the type II-C CRISPR-Cas locus of *Flavobacterium columnare* (characterized in study II). Protein-coding genes are depicted by arrows. Annotations not to scale.

2.2.3 Classification and phylogeny of CRISPR-Cas

Classification of CRISPR-Cas systems is challenging due to the lack of universal markers, the fast evolution of these loci (both horizontally and vertically) and constant discovery of new systems (Makarova *et al.* 2019). An effort to categorize newly discovered systems and rearrange previous discoveries is published every few years (Makarova *et al.* 2011, 2015, 2019). Classification relies on type-specific proteins, conserved sequences such as repeats and leaders, and sequence-based phylogeny of the only universally conserved Cas protein, Cas1. Currently CRISPR-Cas systems branch into more than 35 subtypes (Makarova *et al.* 2019).

On the top level, CRISPR-Cas are divided into classes 1 and 2. Class 1 systems are characterized by complicated multi-subunit complexes that facilitate interference, whereas class 2 systems rely on a single effector protein. The classes are divided into types that are marked by roman numerals (I, II, III, IV, V and VI). Class 1 systems contain types I, III and IV, and class 2 contains types II, V and VI. Types are further divided into subtypes denoted by a letter following the numeral (e.g. II-C) (Makarova *et al.* 2019).

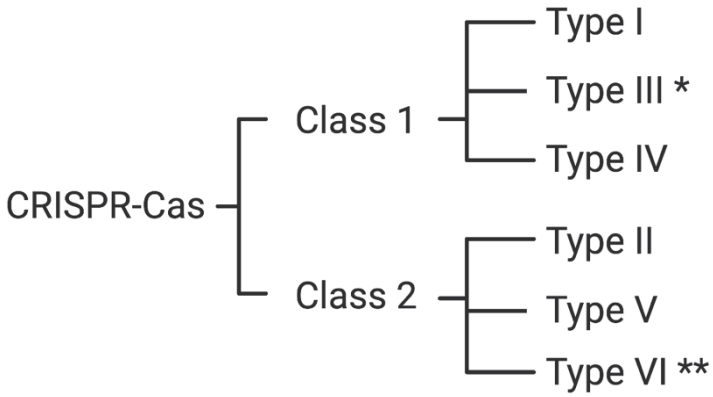

Figure 3. CRISPR-Cas classification. * DNA/RNA-targeting. ** RNA-targeting. Tree based on Makarova *et al.* 2019. Subtypes are not shown.

2.2.4 CRISPR-Cas adaptation

Adaptation (or spacer acquisition) allows a CRISPR-Cas system to update its immunological memory. This process is crucial in maintaining a functioning CRISPR-Cas system, as previously targeted phage genomes may mutate (e.g. Paez-Espino *et al.* 2015), spacers may also be lost over time (Horvath *et al.* 2008, Lopez-Sanchez *et al.* 2012) and the migration or recombination of phages may introduce new malicious genetic material (e.g. Andersson and Banfield 2008).

The first experimental evidence of CRISPR adaptation in laboratory conditions was found in the lactic acid bacterium *S. thermophilus* (Barrangou *et al.* 2007). Upon exposure to a virulent phage, *S. thermophilus* cells that acquired phage-targeting spacers became phage-resistant. The potency of a phage-targeting spacer in this bacterium was strong enough to provide purely CRISPR-Cas based resistance (instead of e.g. SM) and such colonies were thus easy to isolate, paving the way for this bacterium to become a model species for CRISPR-Cas research.

The most central proteins of spacer acquisition, Cas1 and Cas2, are found in most CRISPR-Cas loci (Makarova *et al.* 2019). For functional adaptation, Cas1/2 are often joined by helper proteins and the exact composition of the resulting adaptation complex varies across different CRISPR-Cas types (although Cas1/2 can be enough, see Nuñez *et al.* 2014, He *et al.* 2018). For example, type I-D systems require Cas4 for selecting spacers with compatible PAM sequences (see chapter 2.2.6) (Kieper *et al.* 2018). Type II-A adaptation in *S. thermophilus* recruits Csn2 and Cas9, the latter facilitating acquisition next to proper PAMs (PAMs are discussed in chapter 2.2.6) (Wei *et al.* 2015, Heler *et al.* 2015).

The integration of spacers in the CRISPR array occurs in a polarized fashion, so that new spacers are typically added to the leader-adjacent end of the array (Yosef *et al.* 2012). The integration of the processed pre-spacer into the array occurs through two nucleophilic attacks performed by Cas1 against the leader-adjacent repeat's both ends (Nuñez *et al.* 2015, Rollie *et al.* 2015). This results in separation of the double-stranded repeat into two single-stranded repeats that flank the newly inserted spacer. The single-stranded repeats are then used as templates to synthesize complete repeats on both sides of the spacer using an unknown mechanism (Nuñez *et al.* 2015). In type I systems the integration also requires an integrated host factor (IHF) (Nuñez *et al.* 2016). Due to the intricacy of this process, integration is dependent on conserved motifs in leader and repeat sequences. In *S. thermophilus*, the first ten repeat-adjacent nucleotides in the leader sequence are essential for spacer integration (Wei *et al.* 2015). The *E. coli* type I-E system on the other hand requires conservation of the first 41 to 43 nucleotides (Yosef *et al.* 2012, Díez-Villaseñor *et al.* 2013), whereas the I-D system of this bacterium requires leader motifs that are distal from the repeat-leader junction (Kieper *et al.* 2019). Crucially, Cas1/2 seem to coevolve with their respective host locus, constraining their capacity to work with "foreign" repeat and leader sequences (Shah and Garrett 2011, Garrett *et al.* 2011, Alkhnbashi *et al.* 2016). Furthermore, spacer integration relies on precise

ruler-mechanisms that ensure the maintenance of correctly sized repeats (Díez-Villaseñor *et al.* 2013, Goren *et al.* 2016, Kim *et al.* 2019), thus constraining Cas1/2 to function with repeats of specific size.

Adaptation can occur through two different pathways, termed naive and primed adaptation. Naive adaptation occurs when new spacers are acquired from a target genome by the acquisition complex with no influence from possible pre-existing spacers (Yosef *et al.* 2012). This is generally considered an inefficient adaptation method, with a rate of 10^{-7} to 10^{-6} new spacers acquired per cell according to one estimate in the type II-A *Staphylococcus aureus* system (Heler *et al.* 2017, Pyenson and Marraffini 2020). By contrast, primed adaption occurs when a pre-existing spacer influences the acquisition of a new spacer and is drastically more efficient, with one study reporting a 500-fold increase in acquisition events (Staals *et al.* 2016). In the face of mutating phages, priming is useful in enabling faster reconfiguration of the immune system. Most studies on primed adaptation have concentrated on class 1 systems, in which partial complementarity with the mismatched target becomes the trigger for primed adaptation (Swarts *et al.* 2012, Datsenko *et al.* 2012). Imperfect binding with the target during interference leads to the acquisition of new spacers, mediated by the helicase/nuclease activity of Cas3 (Redding *et al.* 2015, Künne *et al.* 2016, Dillard *et al.* 2018). Later studies showed that fully matching spacers may also stimulate the acquisition of new ones in type I systems (Semenova *et al.* 2016, Staals *et al.* 2016).

A recent bioinformatic study showed that many class 2 systems may also utilize primed adaptation (Nicholson *et al.* 2019). These predictions were experimentally validated on the type II-A system of *Streptococcus pyogenes* and *S. thermophilus,* in which efficient acquisition of new spacers next to a pre-existing one was observed (Nussenzweig *et al.* 2019). Type II priming strictly requires fully matching spacers, leading to the production of free dsDNA-ends produced by Cas9-mediated cleavage on the original target. Such free dsDNA ends were previously shown to be the preferred material for spacer acquisition in the same system (Modell *et al.* 2017).

As high spacer diversity on the population has been shown to correlate with higher rates of phage extinction and mutation rates (Paez-Espino *et al.* 2015, van Houte *et al.* 2016b), acquiring multiple spacers, even in the presence of matching ones, is likely beneficial. From an evolutionary point of view, CRISPR-Cas has been speculated to flip the evolutionary potential in favour phages, since bacteria need to acquire specific spacers for functional defence (Koskella and Brockhurst 2014). However, priming may rebalance this scale, as multiple spacers require multiple escape mutations in the phage genomes; accumulating multiple escape mutations is likely more costly for the phage than is the acquisition of the spacers for the bacterium (Levin *et al.* 2013, van Houte *et al.* 2016b, Common *et al.* 2019).

2.2.5 CRISPR-Cas expression

Expression of the repeat-spacer array is the first stage in the process that ultimately leads to cleaving incoming nucleic acids. The repeat-spacer array is generally transcribed starting from the leader sequence, producing precursor CRISPR RNA (pre-crRNA). This long transcript is processed to yield mature crRNA that is comprised of a single repeat and a spacer (Brouns *et al.* 2008). Cleaving of the pre-crRNA is facilitated by different proteins depending on CRISPR-Cas type. Types I and III use specialized proteins such as Cas6 or Cas5d for this purpose (Brouns *et al.* 2008, Carte *et al.* 2008). Type II systems recruit the conserved endonuclease RNase III, the interference proteins Cas9 and a short transcript called tracrRNA (Deltcheva *et al.* 2011, Chylinski *et al.* 2013). Partial complementarity enables the tracrRNA to bind repeat sequences that are then cleaved by RNAse III in the presence of Cas9. Subtype II-C systems have also been shown not drive array transcription from the leader, but from within repeats towards the leader (Zhang *et al.* 2013, Mir *et al.* 2017). Finally, types V and VI code for CRISPR-Cas interference proteins (Cas12 and Cas13, respectively) that have the intrinsic ability to process their own pre-crRNA (Fonfara *et al.* 2016, East-Seletsky *et al.* 2016).

2.2.6 CRISPR-Cas interference

In interference, the spacers acquired during adaptation are used to defend the host against phages. In a typical interference response, the crRNA produced during the expression stage is joined with endonucleases. The resulting interference complex, guided by the sequence in the crRNA, navigates to a matching sequence in the target nucleic acid (Hale *et al.* 2009). The nuclease domains of the interference proteins are activated upon target recognition and the invading nucleic acid is cleaved within or near the target sequence (Gasiunas *et al.* 2012).

The specifics of interference are the major diversifiers between CRISPR-Cas types. Class 1 interference occurs through a multi-protein complex called the Cascade (CRISPR-associated complex for antiviral defense), whereas class 2 utilizes single effector proteins for the same purpose (Makarova *et al.* 2019). This chapter will concentrate on class 2 systems due to their central role in this thesis.

Type II (of class 2) interference is based on the Cas9 endonuclease which has been famously harnessed for gene editing purposes. Division of type II into its three subtypes (A, B and C) is based on Cas9 homology and the presence or absence additional proteins. Subtype II-A contains Csn2 and subtype II-B contains Cas4 (Shmakov *et al.* 2015). By contrast, subtype II-C is characterized by the absence of both and is comprised only of Cas9 and the acquisition proteins Cas1 and Cas2 (Mir *et al.* 2017). Cas9 contains two distinct endonuclease domains: a RuvC-like domain that is split into three subdomains I, II and III, and a single HNH domain (Jinek *et al.* 2014). Type V interference

relies on the endonuclease Cas12 (previously known as Cpf1) that produces staggered cuts on its target sequence (Zetsche *et al.* 2015).

Type VI CRISPR-Cas loci are divergent from other types, as they exclusively target RNA instead of DNA (although type III-B and III-A systems from class 1 have dual roles in DNA and RNA targeting (Samai *et al.* 2015)). First signs of this CRISPR-Cas type was discovered through bioinformatic approaches (Shmakov *et al.* 2015), and experimental validation soon showed how Cas13 (then called C2c2) exclusively degrades RNA targets (Abudayyeh *et al.* 2016). Type VI was later divided into subtypes A, B, D and E (Makarova *et al.* 2019). The ribonuclease that facilitates RNA-interference in type VI is Cas13 (a-d). In subtype VI-B, Cas13b is often accompanied by two regulatory proteins termed Csx27 and Csx28 (Smargon *et al.* 2017). When targeting dsDNA phages, a crucial requirement for Cas13 activity is that the crRNA produced from the array must be complementary to an RNA target. The capacity of a spacer to bind mRNA is dictated by the orientation in which dsDNA spacer was initially inserted into the array (Shmakov *et al.* 2014). Intriguingly, once activated by the interaction between the crRNA and its target, Cas13 becomes a non-selective RNase that degrades both phage and host transcripts (Meeske *et al.* 2019).

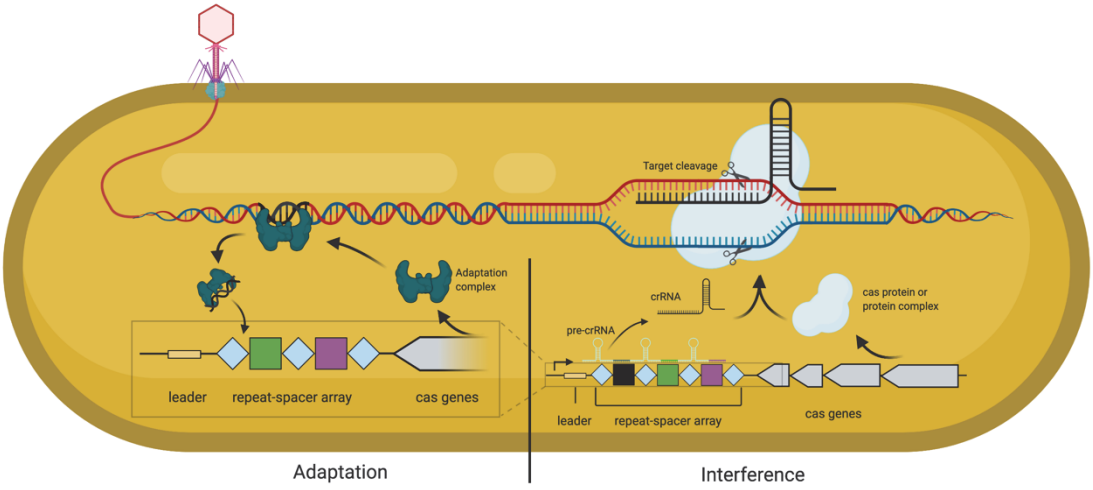

Figure 4. Simplified models of CRISPR-Cas adaptation and interference. Adaptation facilitates the insertion of fragments of invading nucleic acids to the CRISPR array as spacers. The newly inserted spacer is used in interference to guide endonucleases to the complementary sequence on the target genome (different position for adaptation and interference shown here for convenience).

2.2.7 Crosstalk between CRISPR-Cas systems

CRISPR-Cas loci are generally "self-sufficient" in facilitating adaptation, expression and interference (excluding common host factors such as RNase III in type II systems (Deltcheva *et al.* 2011). However, some functional CRISPR-Cas systems have been found to lack essential components. This has led to speculation that components from co-occurring CRISPR-Cas loci may function *in trans* to compensate for their absence. Such crosstalk between CRISPR-Cas

loci has so far been largely overlooked and may have important implications for prokaryotic immunology.

Evidence for crosstalk between CRISPR-Cas systems comes mostly from archaea, in which functional modules (i.e. interference, expression and adaptation modules) undergo frequent exchange between strains and species (Garrett *et al.* 2011). Perhaps owing to the modularity of archaeal CRISPR-Cas systems, interaction between distinct loci has also been observed directly or indirectly in all three phases of CRISPR-Cas. In *Sulfolobus solfataricus*, the processing of crRNA of a type III-A system strictly requires the activity of Cas6a from a co-occurring I-A locus (Deng *et al.* 2013). In the bacterium *Marinomonas mediterranea*, type III-B interference proteins can utilize crRNA from a horizontally acquired type I-F system (Silas *et al.* 2017).

Probably the most interesting aspect of crosstalk is related to acquisition, in which spacer acquisition machinery from one locus is used to insert spacers into another (Fig. 5). Due to the constraints mentioned in chapter 2.2.4, acquisition proteins are highly specific to their associated CRISPR array, arguing against sharing of such an arrangement. However, despite their lack of spacer acquisition machinery (Erdmann *et al.* 2013), archaeal type III systems have been shown to acquire spacers in laboratory conditions upon phage exposure (Erdmann and Garrett 2012). Moreover, new spacers are sampled next to PAM sequences, which is not a requirement for type III systems. This bias towards PAMs suggested that this locus utilize acquisition machinery from the co-occurring type I CRISPR-Cas system, whose repeat sequences deviate only by 1bp from the type III repeats (Erdmann and Garrett 2012). The recently discovered type IV systems that are largely encoded on plasmid have also been predicted to utilize type I spacer acquisition machinery (Pinilla-Redondo *et al.* 2020). This connection is based on co-occurrence of these loci and high conservation in repeat sequences. Using a similar strategy, another study found higher-than-expected correlation between CRISPR-Cas subtypes that contain and lack acquisition machinery, suggesting epistatic interactions (Bernheim *et al.* 2020). However, the *in trans* spacer acquisition hypothesis still awaits experimental verification.

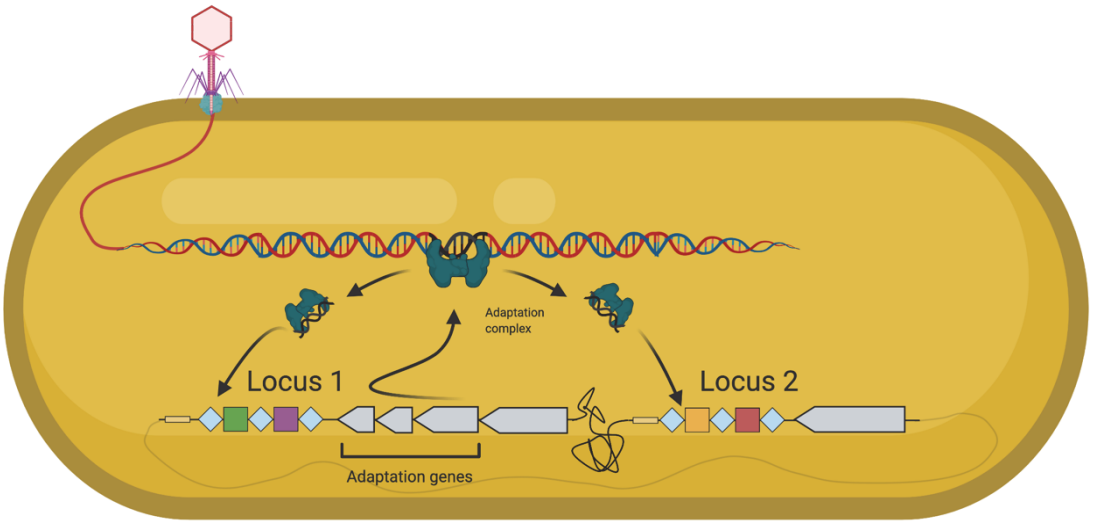

Figure 5. Conceptualization of shared spacer acquisition machinery. Locus 2 lacks adaptation genes (e.g. *cas1* and *cas2*) and is thus dependent on locus 1 for adaptation.

2.2.8 The significance of CRISPR-Cas

The sophistication and diversity of CRISPR-Cas systems suggests that their role in bacterial defense is important (Westra *et al.* 2016). Yet, it has been estimated that the prevalence of this system in bacterial species ranges from 40% (Grissa *et al.* 2007) to as low as 10% (Burstein *et al.* 2016). Why do most bacteria not carry CRISPR-Cas?

The costs of carrying a CRISPR-Cas system are not entirely clear. A common premise is that as an inducible system, CRISPR-Cas should impose relatively little fitness costs to the host as it is only activated upon infection. Experiments with *P. aeruginosa* supported this view (Westra *et al.* 2015), but in *S. thermophilus*, CRISPR-Cas was found to impose both inducible costs resulting from its activation as well as constitutive costs from carrying the system (Vale *et al.* 2015). Another fitness cost may arise in species that benefit from horizontal gene transfer, as CRISPR-Cas can effectively block plasmids and other mobile genetic elements (Marraffini and Sontheimer 2008). Autoimmunity that results from targeting the bacterium's own chromosome is also likely to reduce the likelihood of carrying these systems (e.g. Vercoe *et al.* 2013). CRISPR-Cas has also been documented to disappear after a bacterium switched its host species (Delaney *et al.* 2012), suggesting that a change in the bacterium's ecology no longer favored the maintenance of a CRISPR-Cas system (although whether this was adaptive or coincidental is not entirely clear).

To obtain a comprehensive view of whether a certain system is beneficial or not, it is crucial to consider the full ecological spectrum of these organisms. Biotic interactions such as the frequency of phage encounters, the presence of a metazoan host, predation by other microbes or any relationships with a third party can affect the arrangement of defences systems (van Houte *et al.* 2016a). Likewise, abiotic factors such as oxygen levels may influence the occurrence of

CRISPR-Cas (Weissman *et al.* 2019). In the archaeon *Sulfolobous,* spacer acquisition was even found to be inducible with stress from freeze-thaw cycles (Erdmann *et al.* 2013).

When inhabiting diverse environments, a wide repertoire of defenses is indeed predicted to be beneficial (Hamilton *et al.* 2008). In *P. aeruginosa*, the nutrient and phage concentrations play a major role in choosing a defense strategy between an inducible one (CRISPR-Cas) and a constitutive one (surface modification, SM) (Westra *et al.* 2015). Rare phage encounters favor CRISPR-Cas, whereas constant bombardment of phages (occurring through high-nutrient mediated growth of the bacterial and phage populations) favor SM. It was speculated that upon crossing a certain threshold, the inducible cost of CRISPR-Cas becomes higher than the constitutive cost elicited by SM (Westra *et al.* 2015). It has also been shown that spacer acquisition in *S. thermophilus* is affected by whether the cells were grown in liquid media or on solid agar surfaces (Pyenson and Marraffini 2020). It is also important to recognize the role of CRISPR-Cas not only on the individual bacterial level, but also on the population level, as herd immunity through diversified spacer sequences is effective in curtailing the coevolving phage population (van Houte *et al.* 2016b). Taken together, ecology seems to be a crucial determinant of the relative importance of CRISPR-Cas (or any other immune system) in phage defence.

2.2.9 CRISPR-Cas in nature

It is common for studies on micro-organisms to utilize simplified laboratory environments. This approach may be beneficial or absolutely crucial for isolating specific phenomena, where the experimental setup benefits from minimized interactions with other biotic or abiotic factors. However, as pointed out in the previous sub-chapter, the ecological environment of a bacterium is an important determinant of phage defence. To obtain a holistic understanding on CRISPR-Cas, it is important to also extend these studies to natural environments.

Metagenomic approaches have played an important role in monitoring CRISPR spacer content in natural communities. In biofilm populations of acid mine dwelling *Leptospirillum* bacteria, highly dynamic spacer acquisition and loss was observed within a short time period (Tyson and Banfield 2008). This study was to first to show that spacer turnover in nature is rapid and can diversify otherwise nearly clonal bacterial populations. Later, similar methods and models were used to study the association of metagenomic spacer sequences with co-occurring phage genomes (Andersson and Banfield 2008). This study found that only the most recent spacers targeted the contemporary phage population and that phage genomes evaded CRISPR-Cas targeting primarily through recombination. Another study, combining six years of metagenomic sampling with mathematical models, found that despite extensive shuffling of spacers in the leader-adjacent end of CRISPR arrays, the back-bone of older spacers remained remarkably conserved (Weinberger *et al.* 2012). The conservation of "trailer-end" of the leader was attributed to fluctuating

emergence of previously encountered viral genotypes. These studies show that CRISPR-Cas dynamics in nature may be more complicated than escalating arms race dynamics observed in laboratory conditions (Paez-Espino *et al.* 2015, Common *et al.* 2019).

2.3 *Flavobacterium columnare* and its phages

A member of the *Flavobacteriaceae* family, the fish pathogen *Flavobacterium columnare* (previously known as *Bacillus columnaris, Chondrococcus columnaris, Cytophaga columnaris* or *Flexibacter columnaris* (Bernardet *et al.* 1996)) is a gram-negative, rod-shaped opportunistic bacterium that causes columnaris disease in freshwater fish (Declercq *et al.* 2013). Epidemics in fish farms usually result from the inflow of environmental strains of *F. columnare* that become enriched in these facilities (Kunttu *et al.* 2012), also leading to selection of more virulent strains in these settings (Sundberg *et al.* 2016). Based on genetic differences, *F. columnare* has been divided into three genomovars and in several genotypes that differ e.g. growth rates and virulence (Suomalainen 2005).

F. columnare has the ability to move on surfaces using its gliding motility machinery (Bernardet 1989). The capacity for gliding motility is also linked with *F. columnare's* morphological appearance that can be rhizoid, rough or soft (Fig. 6) (Bernardet 1989, Kunttu *et al.* 2009). Only rhizoid is considered motile and virulent (Kunttu *et al.* 2009, 2011, Laanto *et al.* 2012, Sundberg *et al.* 2014). Many genes that enable gliding motility are encoded by the type 9 secretion system (T9SS) (McBride and Nakane 2015), and mutations in T9SS genes correlate with different morphologies (Li *et al.* 2017). The switch in morphology can sometimes be reversed (Laanto *et al.* 2012) possibly due to changes in expression instead of genetic mutations (Penttinen *et al.* 2018).

When isolates are sampled from semi-natural fish farm environments, only rhizoid colonies are found (Kunttu *et al.* 2012, Laanto *et al.* 2014). The diversification of morphologies in the laboratory results from exposure to phage, starvation or extended sub-culturing (Kunttu *et al.* 2009, 2011, Laanto *et al.* 2012). The links between virulence, phage resistance and morphology change suggested that gliding motility related proteins enable bacterial virulence but are also used by phages for adsorption. It was indeed shown that mutations in T9SS genes lead to morphology change decreasing virulence and increasing phage resistance (Fig. 6) (Li *et al.* 2017). The presence of mucin has also been shown to increase *F. columnare*'s virulence, while also making it more susceptible to phages (Almeida *et al.* 2019). These trade-offs are in line with the theoretical framework aiming to explain phage-bacterium coexistence (chapter 2.1.3).

F. columnare outbreaks in fish farms have enabled the collection of both bacterial and phage isolates during epidemics. Most *F. columnare* phages discovered so far belong to the Myoviridae, Podoviridae or Siphoviridae families (Laanto *et al.* 2011). The phages are tailed with an icosahedral capsid

and their genomes are dsDNA (Laanto *et al.* 2011). Typically, *F. columnare* phages have narrow host-ranges, being able to infect only a specific *F. columnare* genotypes (Laanto *et al.* 2011). This host-specificity allows for tracking coevolution between specific phage-bacterium pairs also in natural environments (Laanto *et al.* 2011). These phages have also been shown to have potential for phage therapy in aquaculture settings (Laanto *et al.* 2015).

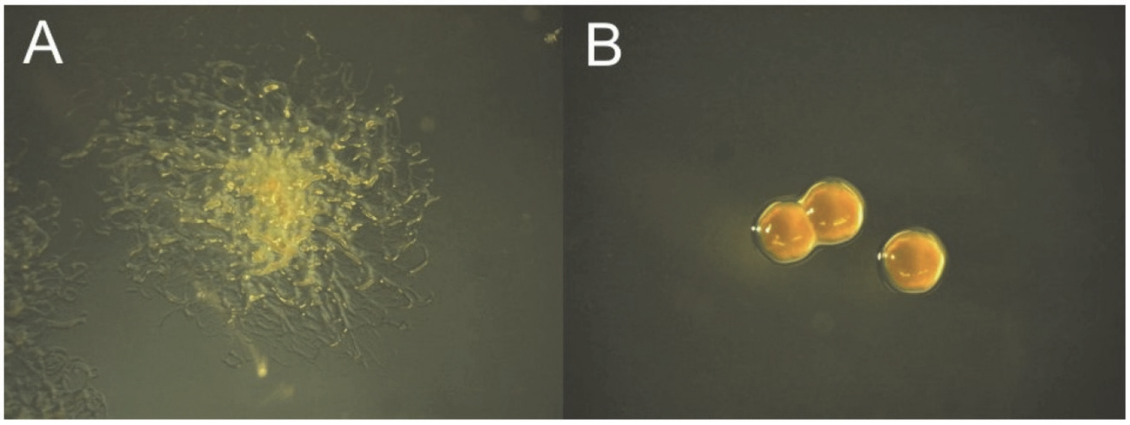

Figure 6. *F. columnare* colony morphologies. A) Rhizoid (virulent, susceptible to phage). B) Rough (nonvirulent, phage resistant). Figure from Laanto *et al.* 2012 (CC BY 4.0 license).

3 AIMS OF THE STUDY

CRISPR-Cas research has often relied on the transfer of CRISPR-Cas systems from less-known species to model organisms such as *E. coli,* leading to pivotal discoveries of this adaptive immune system. However, monitoring CRISPR-Cas activity in native hosts is also important as the interaction of CRISPR-Cas with other host defense mechanisms, the unique effects of native phages and other biotic and abiotic factors cannot be fully replicated outside the native host. This approach has been complicated by difficulties in culturing such species or repeatedly sampling them from nature, as well as by challenges involved with triggering the CRISPR-Cas immune response and the lack of fully sequenced phage or bacterial isolates (Hynes *et al.* 2017). Native bacterial hosts that can be used to study the increasing catalogue of CRISPR-Cas systems are therefore needed. *F. columnare*'s role as an important fish pathogen has motivated extensive sampling and culturing of this bacterium, resulting in vast collections of fully sequenced bacterial isolates and virulent phages. Tools that allow for genetic editing of this bacterium have also been developed, allowing for gene knock-out experiments.

The overarching goal of this thesis was to establish a phage-bacterium system for studying how CRISPR-Cas types II-C and VI-B function in their native host. Furthermore, I wanted to contrast naturally evolved phage and bacterial populations with experimental evolution in the laboratory.

More specific goals were:

1. Understanding phage-bacterium coevolution in artificial and semi-natural settings, highlighting aquaculture as a model system
2. Establishing *F. columnare* as a model for native type II-C and VI-B CRISPR-Cas research
3. Characterizing the prerequisites and nature of CRISPR-Cas type II-C and VI-B spacer acquisition

4 MATERIALS AND METHODS

These broadly categorized methods show the laboratory techniques utilized in each study. Details on these techniques are found in the materials and methods sections of each manuscript. Figures 7 and 8 on the next page show the experimental setups of studies III and IV.

Method	Study I	Study II	Study III	Study IV
PCR		X	X	X
Phage titration		X	X	X
Bacterial culturing		X	X	X
Adsorption experiments		X		
Gene deletion			X	
Infection experiments		X	X	X
Experimental evolution			X	X
Simulations			X	
Sanger sequencing		X	X	X
Next-gen sequencing			X	X
Transcriptomics			X	
Bioinformatics	X	X	X	X
Custom software			X	
Database mining	X		X	
Statistics		X	X	X

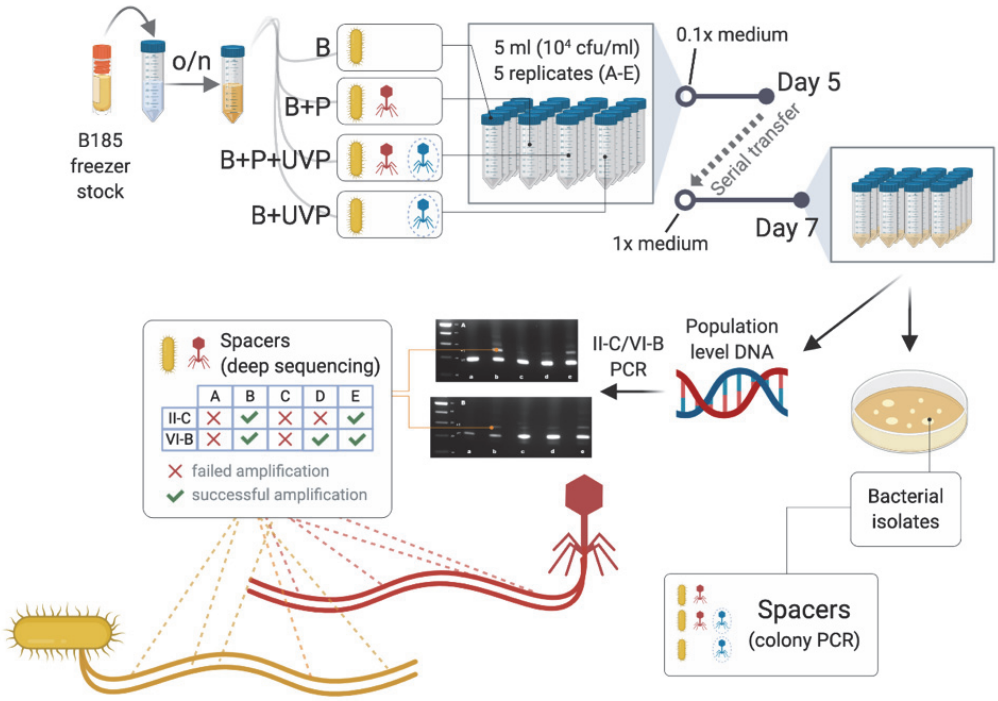

Figure 7. Experimental flow of study III. *F. columnare* strain B185 and its virulent phage FCL-2 (irradiated or non-irradiated) were co-cultured in laboratory settings for seven days, after which new CRISPR-Cas spacers were screened on the population level and individual colonies and mapped on the target genomes.

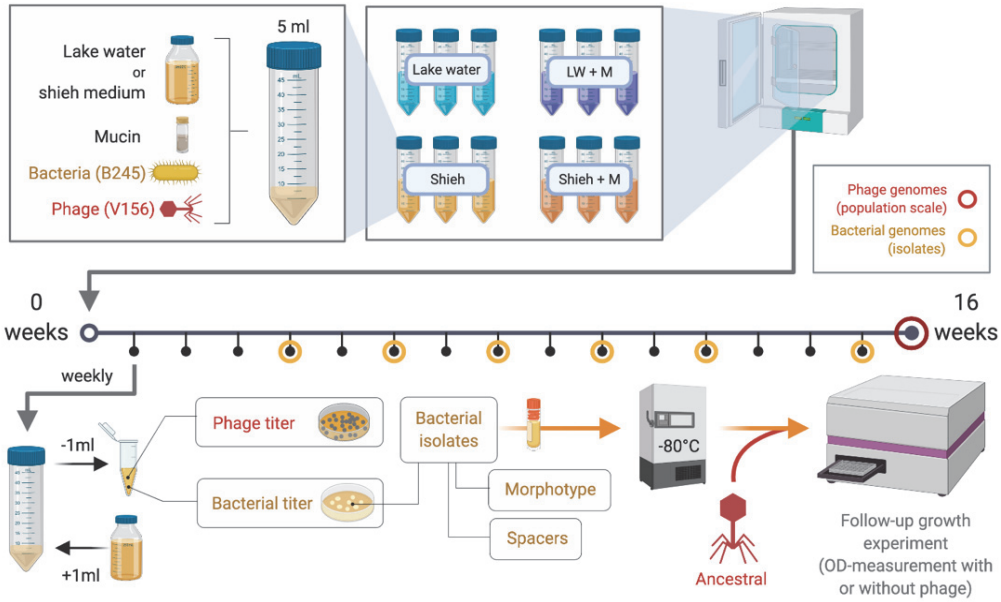

Figure 8. Experimental flow of study IV. *F. columnare* strain B245 and its virulent phage V156 were cocultured in lake water (LW) and Shieh medium in the absence or presence of mucin. During the 16-week experiment, bacterial isolates were stored and their CRISPR arrays sequenced. Several bacterial genomes were also sequenced during the experiment, as well as phage population-scale metagenomes at the end of the experiment. The bacterial isolates were also subjected to growth experiments with the ancestral phage to determine their phage resistance levels.

5 RESULTS AND DISCUSSION

5.1 Aquaculture provides framework for coevolutionary studies (study I)

Experimental evolution often relies on simplified laboratory environments. However, any extrapolations made from these settings to natural ones may overlook additional biotic and abiotic factors that are present in nature (Koskella and Brockhurst 2014). Thus, mirroring laboratory-constrained results with those obtained from natural settings is crucial in understanding the full complexity of any biological phenomenon. However, at least in the case of bacteria and phages, natural settings that allow for the repeated sampling of specific species or strains over long time periods, are difficult to find. In study I, intensive aquaculture (fish farming) is proposed as a solution to this problem. A crucial benefit of these systems is the possibility for repeated sampling over several years, made possible by re-occurring epidemics caused by fish pathogens. As these epidemics occur through the enrichment of bacteria from natural waters, aquaculture provides a semi-natural environment for repeated sampling. It is also important to notice that sampling does not merely satisfy the curious mind of a biologist but is also a crucial tool for monitoring epidemics and thus benefits the industry. Also, due to the commercial and health aspects related to aquaculture, governmental agencies across the globe have likely established collections of frozen aquaculture pathogens over decades, which could already be utilized in evolutionary studies.

The ecology of aquaculture provides diverse habitats for natural microbes that flow in from natural waters and become enriched in these facilities due to the high density of prey and other nutrients. The distribution of microbes in aquaculture is heterogeneous, with cells congregating in biofilms on tank surfaces, floating freely in water columns or feeding on their metazoan hosts, fish. As hotspots for bacterial enrichment, aquaculture also attracts phages. The presence of both predator (phage) and prey (fish) of the bacterium make aquaculture an intriguing environment for studying themes ranging from the

BAM model to long coevolutionary studies in bacterial defence and phage resistance.

While the benefit of aquaculture for sampling naturally occurring bacteria and phages has already been widely documented (e.g. Suomalainen 2005, Suomalainen *et al.* 2006, Kunttu *et al.* 2009, 2012), literature on CRISPR-Cas in these settings was lacking. To this end, 23 pathogens that colonize fish farms were listed and their CRISPR-Cas content determined using publicly available genome data. Out of the 183 analyzed strains, roughly 40% had at least one CRISPR-Cas system. Two genera stood out as rich in different CRISPR-Cas systems: *Vibrio* and *Flavobacteriia*.

Variability among *Vibrio* strains was high, but four of the five investigated species had at least one strain with CRISPR-Cas. Many *Vibrio* species have already been used as models in CRISPR-Cas studies (e.g. Sun *et al.* 2015, Kalatzis *et al.* 2017) and have potential for further CRISPR-Cas class 1 research, especially in aquaculture settings. It is worth mentioning that *Vibrio cholerae* was not included in the list due to its less severe (or even beneficial) effect on fish (Halpern and Izhaki 2017). However, *V. cholerae* is already an important model species for CRISPR-Cas, especially when associated with its human host (e.g. Seed *et al.* 2014, Box *et al.* 2016, Bin Naser *et al.* 2017). Intriguingly, *V. cholerae* phages have also been found to carry CRISPR-Cas systems (Seed *et al.* 2013).

The genus Flavobacterium was represented by three species, all of which carried class 2 systems. Natural metagenomic samples of *F. psychrophilum* show spacer variation in its type II-C locus (Lopatina *et al.* 2016), but an experiment to trigger spacer acquisition in the laboratory did not result in expanded arrays (Castillo *et al.* 2015). *F. branchiophilum* was the most versatile bacterium of this group, carrying II-C, VI-B and V-A systems, but to our knowledge no further CRISPR-Cas research has been done on this bacterium. The CRISPR-Cas systems of *F. columnare* (type II-C and VI-B) are discussed in studies II to IV.

The main motivation for determining CRISPR-Cas content in aquaculture pathogens was to encourage researchers that work or aim to work with these bacteria to consider their potential for natural studies of CRISPR-Cas. These settings also allow for investigating how CRISPR-Cas systems may contribute to bacterial virulence, as shown in type II systems (Louwen *et al.* 2014). Aquaculture is also a potential future target for phage therapy applications (Laanto *et al.* 2015), which makes understanding phage-bacterium coevolution in these settings crucial. Screening for CRISPR spacers may also help in developing effective phage cocktails or engineered phages that bypass pre-existing CRISPR-Cas defence. During a phage therapy intervention, spacer screening can also be used to assess if bacteria are forming resistance to the administered phages.

5.2 *F. columnare* coevolves with its phages in nature (study II)

Aquaculture as a tool for coevolutionary studies was put to practice in study II. As material, we utilized a large collection of *F.* columnare bacterial and phage isolates sampled from a fish farm between 2007 and 2014. To investigate if bacteria evolve resistance and if phages evolve virulence over time, we cross-infected bacteria from all time points with phages from all time points. This so-called time-shift experiment (Gaba and Ebert 2009) revealed that bacteria isolated a later time point were resistant to phages from a previous time point (Fig. 9B). Likewise, when phages were "from the future" and bacteria "from the past", the bacteria were sensitive. This pattern suggested that not only do bacteria evolve resistance, but phages also evolve to overcome this resistance. Similar results were previously obtained in a six-month experiment that sampled natural communities of bacteria and phage from host trees (Koskella 2013), as well as in a study utilizing water fleas and its parasites trapped in layers of pond-sediments (Decaestecker *et al.* 2007).

We then looked for causes behind these cross-infection patterns by sequencing roughly half of the phage isolates' genomes. Genome alignment revealed that the phage genomes remained remarkably similar and trackable over the years, similar to previous screenings of *Vibrio cholerae* phage isolates (Seed *et al.* 2011, Angermeyer *et al.* 2018). However, the *F. columnare* phages also accumulated mutations and new genetic material, mostly concentrating into one end of the phage genome (Fig. 9D). Intriguingly, growth in genome size and accumulation of mutations correlated positively with the phages' host range, providing a direct link between the phenotypic cross-infection patterns and genomic organisation.

I then investigated the CRISPR-Cas content of *F. columnare.* The genomes of these bacteria contained two CRISPR-Cas loci: a type II-C locus with *cas9*, *cas1*, *cas2* and a repeat-spacer array, and a second mysterious locus that had a CRISPR array but no known *cas* genes in its vicinity. A large ORF upstream of the array was widespread in other species and invariably occurred next to a CRISPR array. Moreover, it lacked previously known Cas protein motifs but did include a conserved RNAseL motif also found in the eukaryotic Ire1 enzyme (Lee *et al.* 2008). Suspicions of this being an RNA-targeting Cas protein were confirmed soon, as Cas13 was described around this time (Abudayyeh *et al.* 2016, Smargon *et al.* 2017), giving rise to CRISPR-Cas type VI. This was a unique opportunity to examine both RNA and DNA-targeting CRISPR-Cas loci in a natural setting. I also made note that the VI-B locus was missing spacer acquisition machinery (the central theme of study III).

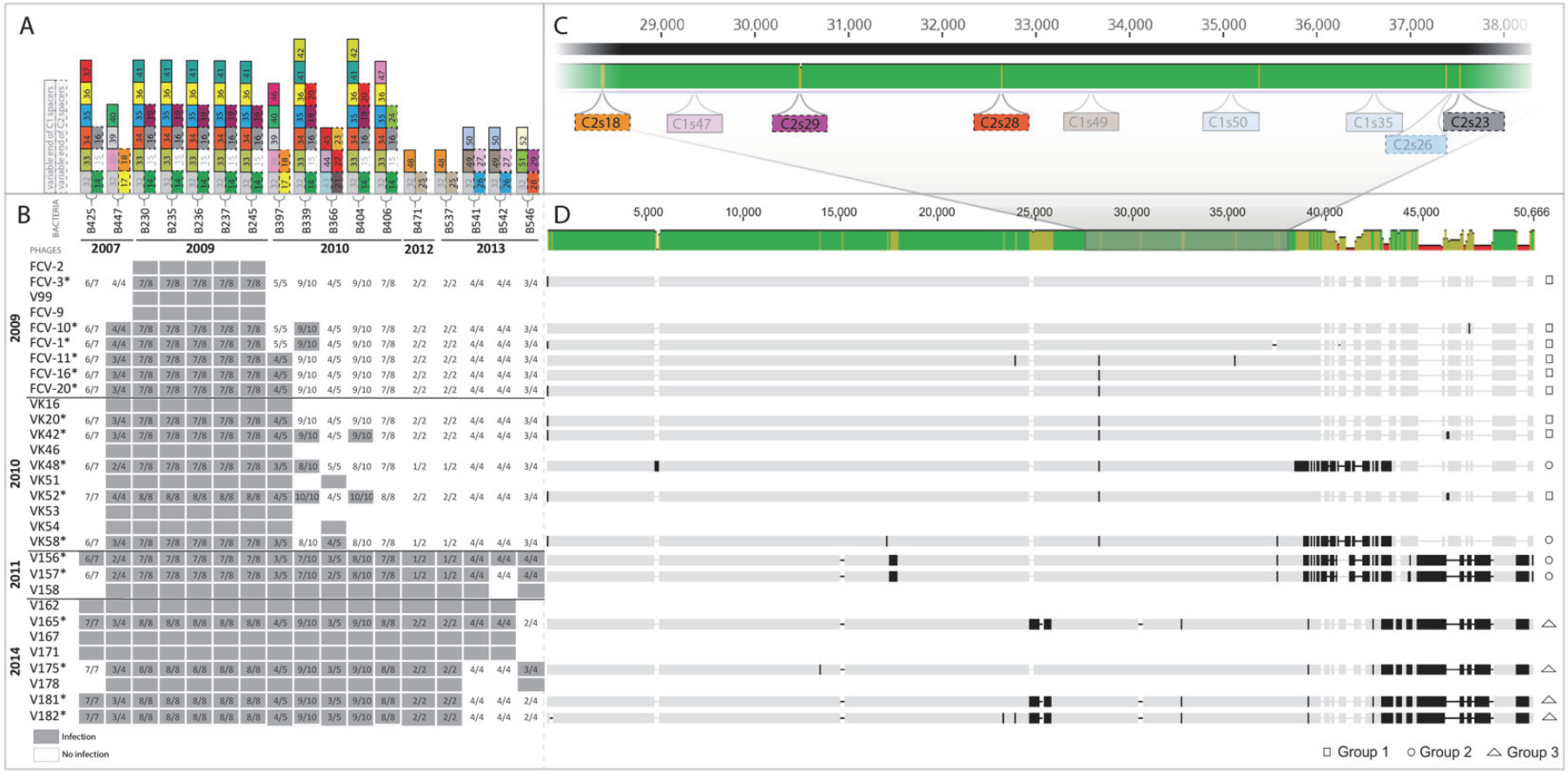

Figure 9. Main results from study II. A) Variable ends of both CRISPR arrays in different bacteria isolates. The conserved backbone is not shown. B) Cross-infection experiment, where phages are displayed on rows and bacteria in columns. Each cell indicates the outcome of an infection experiment, with grey cells indicating infection and white cells indicating resistance. The numbers in each cell denote the number of II-C/VI-B spacers targeting the phage in the current phage-bacterium pair. C) Point mutations in several VI-B protospacers on the phage genome in the zoomed in region. D) Alignment of phage genomes from selected isolates, corresponding to rows in panel B. Black regions denote variation. Figure initially published in Nature Communications.

To examine the spacer content of the II-C and VI-B loci, I sequenced the arrays of both loci in all bacterial strains (Fig. 6A). All isolates shared the same conserved locus-specific backbone of spacers, with variation only in the leader-adjacent ends of the arrays similar to a previous metagenomic study (Weinberger *et al.* 2012). After attributing a target for each spacer, I discovered that most spacers in the variable ends were targeting the co-occurring phage genomes, providing evidence that these phages and bacteria are in constant interaction in the aquaculture environment. Protospacers on the phage genomes mostly congregated on the end of the genome that also exhibits most variation over time. A similar end-targeting pattern was found in *S. aureus*, where the end of the phage genome that enters the cell first was targeted the most (Modell *et al.* 2017). Furthermore, on several occasions the phage genomes had small deletions or base substitutions exactly on the protospacers, as seen in experimental evolution of CRISPR-Cas (Paez-Espino *et al.* 2015, Common *et al.* 2019). Intriguingly, at least some of these protospacer mutations occurred after the corresponding spacer had appeared in the bacterial population. Most such mutations were associated with VI-B spacers but not II-C spacers. However, as the phage genome end was highly variable, it was challenging to pinpoint which mutations occurred in protospacer regions through random drift and which were signals of CRISPR-Cas evasion.

Knowing that the Cas13b of the VI-B locus is an RNA-targeting endoribonuclease, I examined the capability of the VI-B spacers to target predicted mRNA on the phage genomes. If the crRNA from these spacers were not complementary to the coding strand of an ORF (and thus mRNA), they were deemed unable to bind RNA (this presumption disregards possible phage antisense transcription, in which case binding to the opposite strand would also enable RNA-targeting). Strikingly, all VI-B spacers fulfilled this mRNA-targeting criterion. In contrast, the same analysis done for the II-C spacers revealed a 1:1 ratio in mRNA and non-RNA targeting spacers. These results suggested that mRNA-targeting spacers become enriched in the VI-B locus, perhaps through acquisition of such spacers directly from mRNA (Silas *et al.* 2016) or through positive selection of RNA-targeting spacers from a mixture of both mRNA and DNA-targeting spacers (a matter that was investigated further in study IV).

I then investigated if the number of spacers correlated with increased bacterial resistance. However, bacterial resistance was found to be completely unaffected by the number of phage-matching spacers. This was a surprising finding, as the high turnover of phage-targeting spacers over time, coupled with protospacer mutations on the phage genomes, strongly suggested that CRISPR-Cas plays an important role in *F. columnare* phage defence. The most likely explanation for this discrepancy stems from the experimental setup of measuring phage resistance. To determine resistance, these standard laboratory experiments rely on extremely high concentrations of phages and nutrients compared to those found in natural habitats of *F. columnare.* Previous studies have indicated that CRISPR-Cas is likely not an effective defence in the face of high phage pressure (Westra *et al.* 2015). Continuous barrage of phage

infections increases the relative cost of CRISPR-Cas systems, in which case surface modification -based defences, that prevent phage entry altogether, are more likely preferred. Our experimental setup therefore likely highlights the evolution of resistance based on defences other than CRISPR-Cas. However, as we did not fully sequence the bacterial genomes, we do not know what these mutations are (this aspect was explored later in study IV).

This study showed that *F. columnare* and its phages are in a coevolutionary relationship in nature and possible signatures of both fluctuating and/or arms race dynamics were detected. Supporting fluctuating dynamics are the appearance and disappearance of spacers in the bacterial isolates, as well as the appearance and disappearance of corresponding protospacer on the phage genomes. Similar results were obtained in metagenomic sampling of CRISPR-Cas, where spacer diversity in the variable end changed over time while the trailer end remained conserved (Weinberger *et al.* 2012). Furthermore, while bacterial resistance towards "past" phages increased over time, the earliest (2007) bacterial isolate B425 exhibited strong resistance to "future phages", resembling resistance patterns of the most recent bacterial isolates. If the 7-year snapshot of coevolution captured by this study contained a single "wave" of fluctuating dynamics, it is possible that the earliest isolate represented a survivor of a previous fluctuation in a kill-the-winner scheme. The growth of the phage genomes, on the other hand, was more reminiscent of an escalating arms race. However, as the phage genome size is tightly constrained by the capsid size, longer sampling might have revealed corresponding shortening of the genomes as expected in fluctuating dynamics.

It is worth mentioning that we only sampled a specific genotype (genotype C) of *F. columnare* and thus any analysis of evolutionary dynamics is constrained within that genotype. However, multiple genotypes of *F. columnare* may cause epidemics simultaneously (Suomalainen 2005). As the phages of *F. columnare* generally only infect one genotype, it is possible that the densities of different genotypes are controlled by negative frequency dependent selection, following kill-the-winner dynamics. This should be addressed in future studies.

5.3 *F. columnare* CRISPR-Cas loci acquire spacers in specific laboratory conditions (study III)

Inspired by the high turnover of spacers in natural populations (study II), I took a closer look at spacer acquisition in *F. columnare* in the laboratory. The aim was to seek conditions that enable phage-borne spacer acquisition without the need for artificial expression in another host nor artificial overexpression of acquisition genes. Upon successful execution, *F. columnare* would become one of the few species which allows for native spacer acquisition in the laboratory (Hynes *et al.* 2017) and so far the only species for which native type VI acquisition has been demonstrated.

Previous studies showed that *F. columnare* reacts rapidly to the presence of phage by changing its morphotype from rhizoid to rough, thereby preventing phage entry while simultaneously losing infectivity towards fish (Laanto *et al.* 2012, 2014, 2020). Indeed, morphotype change dominated over CRISPR-Cas in *F. columnare* in initial attempts to induce spacer acquisition. As speculated in study II, nutrient conditions (and the resulting growth in phage pressure) may affect the defence response on phage predation. I therefore experimented with different media, dilutions and phage/bacterial concentrations (unpublished data). Finally, a condition that allowed for spacer acquisition involved an initial growth period of five days in diluted (0.1x) Shieh medium with the phage (alpha phase). The alpha phase was followed by serial transfer to 1x Shieh medium, from which expanded CRISPR arrays could be detected after a couple of days (beta phase). As this study did not specifically focus on the ecological factors of spacer acquisition, but rather on the spacers themselves, I did not perform further experiments to dissect the ecological factors that lead to spacer acquisition. However, it is interesting to speculate why sequential growth in alpha and beta-phases induces spacer acquisition, while growing in either phase alone does not. It is likely that during growth in alpha-phase, low cell density leads to less frequent phage encounters, promoting CRISPR-Cas as shown by Westra et al. (Westra *et al.* 2015). After this five-day "vaccination", the bacterial and phage population is enriched in the undiluted medium of beta-phase. During explosive growth in beta-phase, cells with previously activated CRISPR-Cas defences likely become positively selected. However, prolonged incubation in beta-phase lowered the odds of observing spacer acquisition (data not shown), suggesting that morphotype change eventually dominates in this environment.

In the actual experiment, illustrated by Fig. 7, *F. columnare* strain B185 was grown either alone (B), with the wild-type FCL2 phage (B+P), with UV-treated phage (B+UVP) or with a mixture of both (B+P+UVP). The UV-treated phage was included since previous studies using *S. thermophilus* had shown increased spacer acquisition in the presence of UV-treated phage, likely owing to their hampered replication capacity (Hynes *et al.* 2014). After the seven-day acquisition assay was complete, the bacteria were plated and CRISPR arrays of individual colonies were assessed. The B+P+UVP treatment contained the largest proportion of colonies with expanded arrays, while the B+P had less and B or the B+UVP treatments had none. Serial plating of colonies with expanded arrays also revealed reversion back to wild-type on multiple occasions, suggesting that cells within colonies may be heterogenous in their spacer content as shown in Pyenson and Marraffini 2020.

To gain a comprehensive view on spacer contents, I amplified both arrays from population-scale DNA extractions taken directly from the liquid cultures. This approach revealed similar efficiencies of acquisition in the B+P and B+P+UVP treatments, while the B and B+UVP treatments did not show any acquisition. Intriguingly, while the efficiency of acquisition seemed to vary between the five replicates per treatments, the efficiency was relatively similar between the two loci within each replicate.

I then subjected the population scale B+P samples to amplicon sequencing using the Ion Torrent PGM platform. The resulting data was processed using custom scripts and mapped against the phage and bacterial genomes. I divided the analysis into unique or non-unique spacers: in the unique spacer dataset, each spacer sequence appears only once, while the non-unique does not limit the occurrence of any sequence. An imbalance in the targeting preferences between the loci was apparent: II-C was predominantly acquiring unique spacers that target the FCL2 phage genome, while the VI-B locus targeted both genomes relatively evenly (Fig. 10A), with more pronounced trends in total spacer counts (Fig. 10B). The size distribution of spacers in both loci centred around 30 bp (Fig. 10C) and most (58%) unique phage-targeting spacers between the loci were identical (Fig. 10D). Simulating similarly sized spacer sets with random sampling across the phage genome produced considerably lower similarities, suggesting that the loci share a bias towards certain spacers. Regions surrounding the protospacers on the phage genome exhibited similar sequences, albeit in reverse-complement and different localisations (upstream in VI-B and downstream II-C) (Fig. 10E).

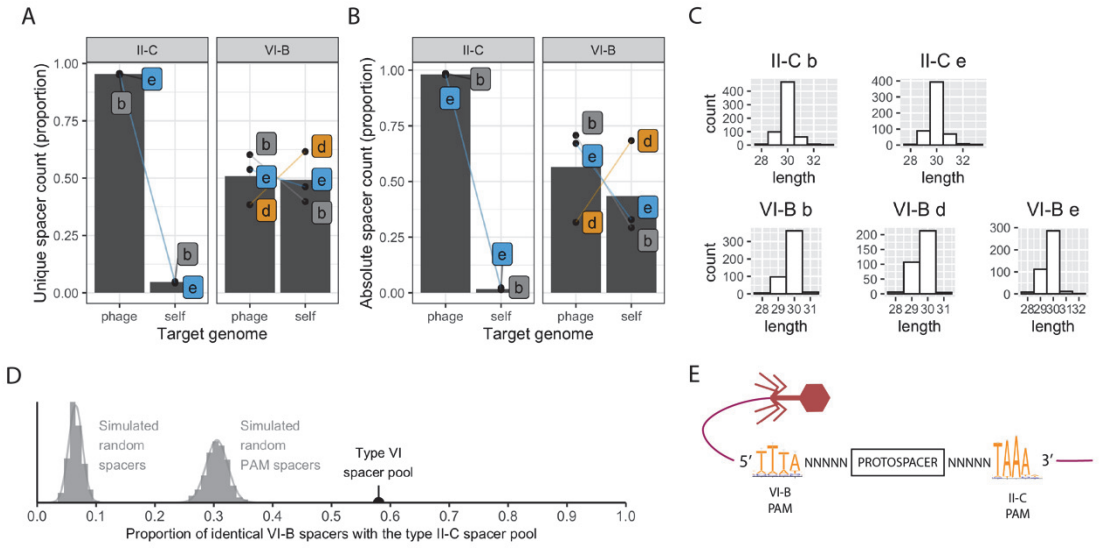

Figure 10. Deep sequenced spacers. Unique (A) or total spacer counts (B) of II-C and VI-B spacers targeting the phage or bacterial genome. C) Spacer size distribution across the replicate cultures D) The proportion of shared spacers between the II-C spacer pool and simulated data (the distributions) or between II-C locus and VI-B spacers pool. E) PAM sequences for both loci.

In line with study II, both loci acquired spacers mostly from the end of the phage genome (Fig. 11). When considering the type II-C non-unique spacer targets, a trend not visible in the unique spacer dataset formed on the phage genome: two asymmetrical targeting peaks surrounded a pre-existing type II-C spacer in the end of the genome (Fig. 11A). This trend was reminiscent of primed adaptation, in which a CRISPR-Cas system relies on pre-existing spacers to propagate spacer acquisition in nearby regions. In the type II-A CRISPR-Cas system, priming is dependent on a perfect match between a spacer and its protospacer (Nussenzweig *et al.* 2019), which is also what occurs in the current study. If the targeting peaks are due to priming, it remains unclear why

the three additional pre-existing II-C spacers did not induce similar patterns. One reason could be that they are located on the "wrong" end of the phage genome, if the priming spacer targets the end of the genome that first enters the cell. The end of the phage genome to first enter the cell has indeed been shown to attract most spacer acquisition events (Modell *et al.* 2017). A priming trend was not visible in the VI-B spacers (Fig. 9), although their distribution was similarly drawn towards the end of the phage genome. The possible priming effect in II-C spacers should be studied further with modified CRISPR arrays, where priming spacers are removed, as well as using different *F. columnare* strains.

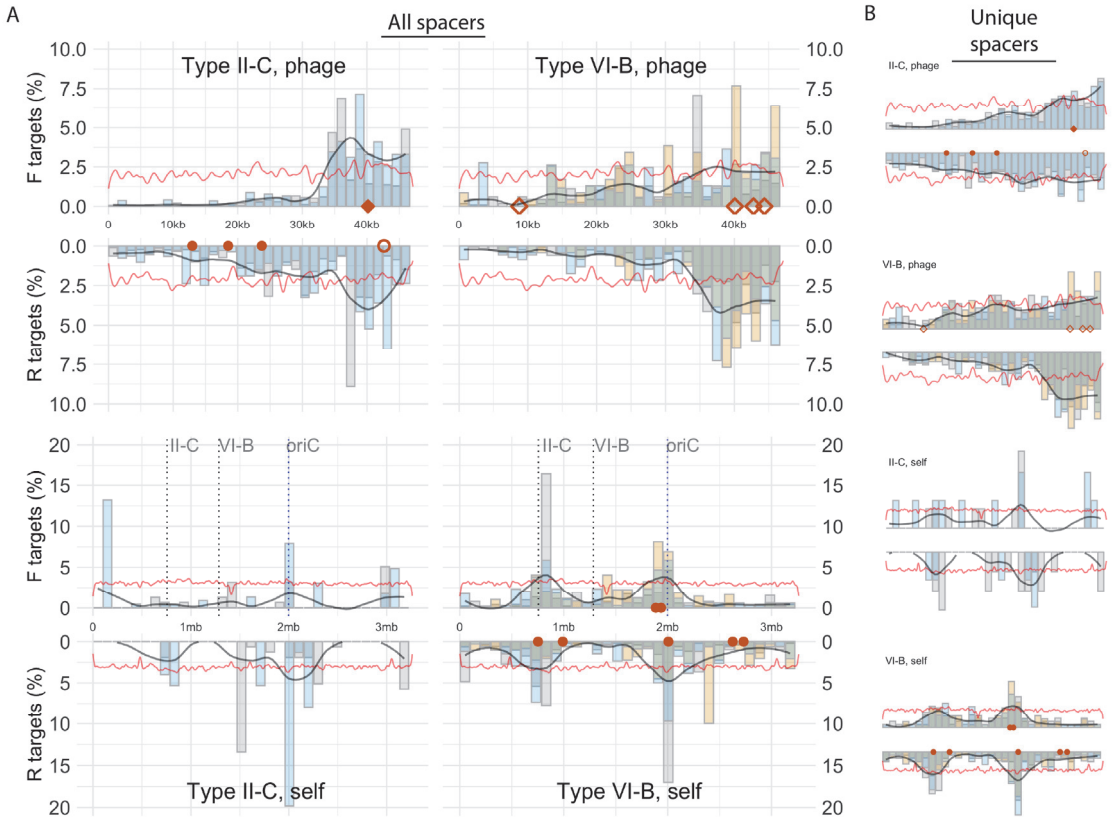

Figure 11. Phage targets on phage and the bacterial genomes. A) Total spacer counts B) Unique spacer counts. Pre-existing spacers in both panels are shown by red markers: diamonds are mRNA-complementary spacers and circles are not mRNA-targeting. Non-filled markers indicate 1-3bp mismatches with the target.

On the bacterial genome, both loci targeted two specific hotspots: the origin of replication of the chromosome (oriC) and the II-C CRISPR-Cas locus. These hotspots also exhibited asymmetrical targeting peaks on the opposing strands, similar to the non-unique II-C spacers on the phage genome. The oriC and a CRISPR array had previously been found to attract spacer acquisition in *E. coli* (Levy *et al.* 2015). The authors found that spacer acquisition was concentrated on double-stranded breaks occurring at stalled genome replication forks. Such forks were overrepresented in the oriC, which serves as the site where genome replication begins, whereas the CRISPR array stalls replication due to insertion

of new spacers (Levy *et al.* 2015). Further studies are needed to assess if this is also the case in *F. columnare*.

Orientation of a spacer on the array dictates the capability of crRNA to bind mRNA. The orientation of type VI-B spacers was of special interest, given the previous discovery that all VI-B spacers in natural isolates are mRNA-targeting (study II). However, deep sequencing the spacers in this study revealed that less than half of spacers acquired in laboratory conditions target mRNA from phage ORFs. This indicated that mRNA-targeting VI-B spacers are likely positively selected in nature from a pool of mixed spacers. It is worth noting that this analysis is based on the assumption that mRNA constitutes all phage-based RNA, not taking into account possible antisense transcription. While the role of antisense RNA in these *myoviridae* phages is unknown, the proportion of possible antisense RNA in the total transcriptome is likely minor (Mojardín and Salas 2016). We therefore considered our assumptions to reasonably reflect the RNA-targeting capabilities of the spacers. This is also supported by the enrichment of mRNA-targeting VI-B spacers in nature (study II). However, alternative functions of non-mRNA-binding VI-B spacers cannot be ruled out.

5.4 Type VI-B locus uses II-C spacer acquisition machinery *in trans* (study III)

In addition to characterizing the spacer contents of the II-C and VI-B CRISPR-Cas loci of *F. columnare*, I wanted to understand how the type VI-B locus acquires spacers despite the lack of necessary machinery to do so. This question had been asked before (e.g. Sternberg *et al.* 2016, Smargon *et al.* 2017), as Cas1 and Cas2 are generally considered necessary for spacer acquisition. However, indirect evidence from systems other than type VI (e.g. Pinilla-Redondo *et al.* 2020, Bernheim *et al.*) suggested that "borrowing" Cas1/2 proteins from other loci is likely a common phenomenon in systems that lack an adaptation module. What made these claims problematic, however, were presumptions on how Cas1/2 coevolve with their respective leader and repeat sequences (Díez-Villaseñor *et al.* 2013, Nuñez *et al.* 2014, Rollie *et al.* 2015, Kieper *et al.* 2019, Kim *et al.* 2019). Thus, deviations from these motifs, which are likely to arise when comparing different types of CRISPR-Cas systems, should abrogate or weaken spacer acquisition. In *F. columnare*, I found that the leaders and repeats were not identical but did contain certain shared characteristics that might enable the operation of II-C Cas1/2 on the VI-B array. Indirect evidence for *in trans* adaptation came from their shared spacer acquisition motifs (SAMs) despite the PAM-independence of type VI, similar to a previous archaeal study (Erdmann and Garrett 2012). As SAMs are considered unique to specific acquisition complexes, it was very likely that acquisition occurred in both loci using the same machinery. I also looked at other species that share II-C and VI-B CRISPR-

Cas loci and found that their repeats differ in length, whereas *F. columnare* has similarly sized repeats in both loci. Indeed, a previous study showed spacer acquisition in a II-C locus of *Riemerella anatipestifer*, but not in the VI-B locus (He *et al.* 2018).

To investigate if *in trans* acquisition occurs in *F. columnare*, we deleted *cas1* from the II-C locus and repeated the spacer acquisition experiment. No spacers were inserted into either locus, while the simultaneously grown wild-type controls exhibited efficient acquisition. Further studies should investigate the plasticity of this acquisition machinery and specifically determine the importance of specific leader/repeat motifs. Given the differently sized repeats in other species with similar CRISPR-Cas configurations, it is possible that *F. columnare* is the only sequenced species for which VI-B *in trans* acquisition occurs. Positive results in these other species would reveal interesting aspects of Cas1/2 mechanics that typically rely on specifically sized repeats and spacers (Díez-Villaseñor *et al.* 2013, Goren *et al.* 2016, Kim *et al.* 2019).

This study showed the first case of native type VI adaptation and first experimental proof of *in trans* adaptation between CRISPR-Cas types. Spacers were acquired in both loci with differing biases in phage and self-targeting. Type II-C spacers displayed possible evidence of primed adaptation, while VI-B did not. Surprisingly, most VI-B (or II-C) spacers did not target phage mRNA. Despite successful adaptation in the laboratory, spacer acquisition was still not very effective, and not all replicate cultures could be deep sequenced. We therefore considered what ecological factors could intensify spacer acquisition even more. To this end, in study IV, we turned to the bacteriophage adherence to mucus (BAM) model.

5.5 Chemical host signals intensify spacer acquisition (study IV)

After investigating long-term coevolution in natural settings (study II) and short-term spacer acquisition in the laboratory (study III), it was tempting to see how *F. columnare* and its phages coevolve in controlled laboratory conditions over a longer time period, and in the presence or absence of host-signals. Three-partite relations (phage, bacterium and the eukaryotic host of the bacterium) have not been extensively studied in the context of CRISPR-Cas and *F. columnare* was a promising model for such an investigation.

A recent study by Almeida et al. had shown that phage virulence increases when *F. columnare* is grown in the presence of a mucus component called mucin (Almeida *et al.* 2019). Mucin also increased *F. columnare* virulence, presumably signalling the bacterium of host presence, while simultaneously reducing the bacterium's resistance towards phages. Phage resistance and bacterial virulence are negatively correlated, because both rely partly on the same surface receptors in *F. columnare* (Li *et al.* 2017). The presence of mucin is therefore expected to suppress morphotype change, as only the rhizoid morphotype enables colonization of the host. Remaining rhizoid, however, provides an opportunity

for phages to infect the cells, as the gliding motility machinery is likely used as an entryway into the cell (Zhang *et al.* 2014, Li *et al.* 2017). We hypothesized that the addition of mucin may thus favor intracellular defences, namely CRISPR-Cas (Fig. 12).

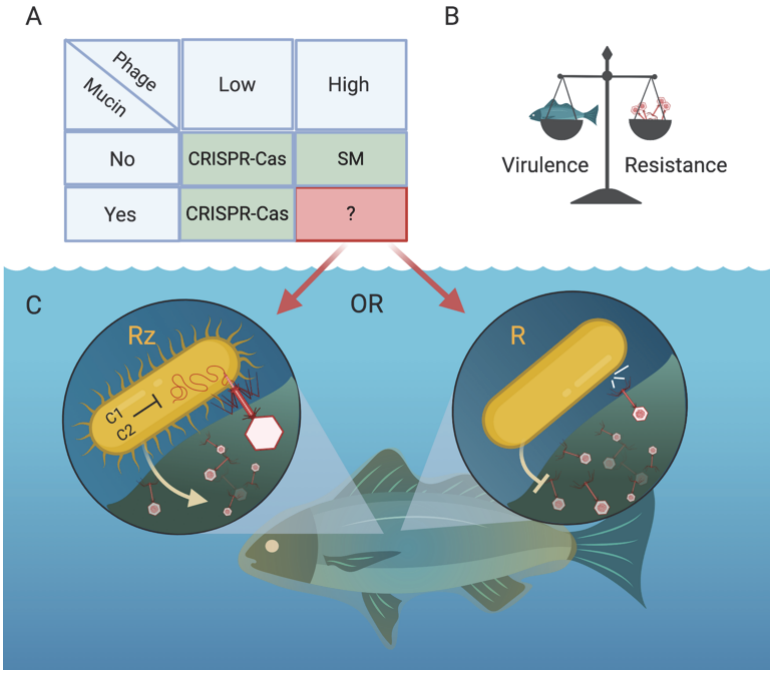

Figure 12. Conceptualization of the effect of host-signals on phage-defense strategy. A) Expected outcomes of different ecological settings. B) Phage resistance may cause a decline in virulence C) If this tradeoff occurs during colonization, a conflict emerges. One solution may be to switch defense strategies into ones with less tradeoffs with virulence. Rz = rhizoid, R = rough.

We grew *F. columnare* in the presence or absence of mucin in either sterilized lake water (LW) or Shieh medium. During this 16-week experiment, we did weekly screening of the abundance of bacteria and phage and sequenced several bacteria genomes as well as population-scale phage genomes (Fig. 8).

Bacteria and phage both survived until the end of the 16-week experiment in all conditions (Fig. 13). The addition of mucin in LW or Shieh increased the bacterial concentrations in all replicates. In LW, mucin decreased phage titers, while in Shieh the addition of mucin increased phage titers. Variation between replicates were generally higher in Shieh than in LW treatments. These dynamics indicate that the addition of mucin, especially in the low-nutrient LW treatment, has important effects on both bacterial and phage growth. Moreover, bacterial and phage titers were often negatively correlated, indicating the development of bacterial resistance and phage infectivity during the experiment (Fig. 13).

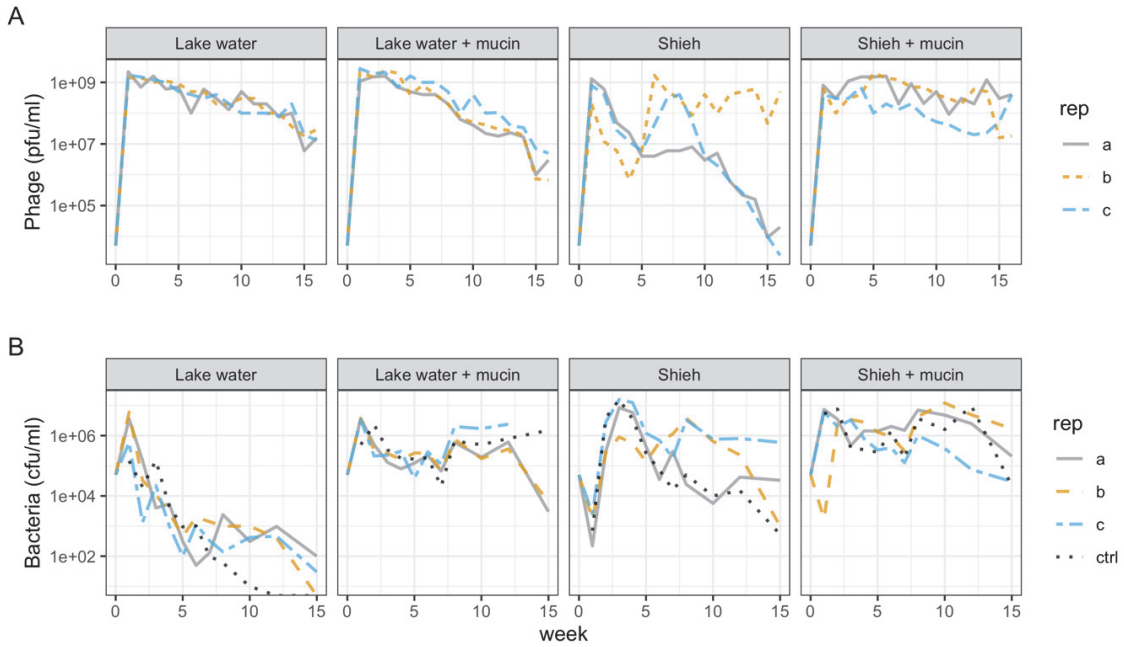

Figure 13. Phage (A) and bacterial (B) titers over the 16-week experiment. Replicates are shown by separate lines. Dashed line indicates control with no phage.

The addition of mucin significantly enhanced spacer acquisition, especially when mixed with lake water: after a few weeks, over 60% of colonies in this treatment (LW+M) contained at least one new spacer (Fig. 14A). Spacer count in individual colonies was also high, with up to six II-C spacers and four VI-B spacers in a single isolate (Fig. 14B). Interestingly, morphotype did no correlate with spacer acquisition: isolates with new spacers in LW+M were both rhizoid and rough (Fig. 14C). In the LW-treatment, the cells had invariably turned to the rough morphotype (Fig. 14C), but still had a small proportion of colonies with expanded arrays. In the Shieh treatment, no spacers were acquired, but with added mucin, up to 25% of colonies had expanded arrays while also displaying mixed morphologies (Fig. 14B-C). These results show that mucin increases CRISPR-Cas acquisition, especially in low-nutrient environments. The accumulation of several spacers in individual isolates suggests that acquisition is very efficient, perhaps accelerated by primed adaptation as speculated in study III.

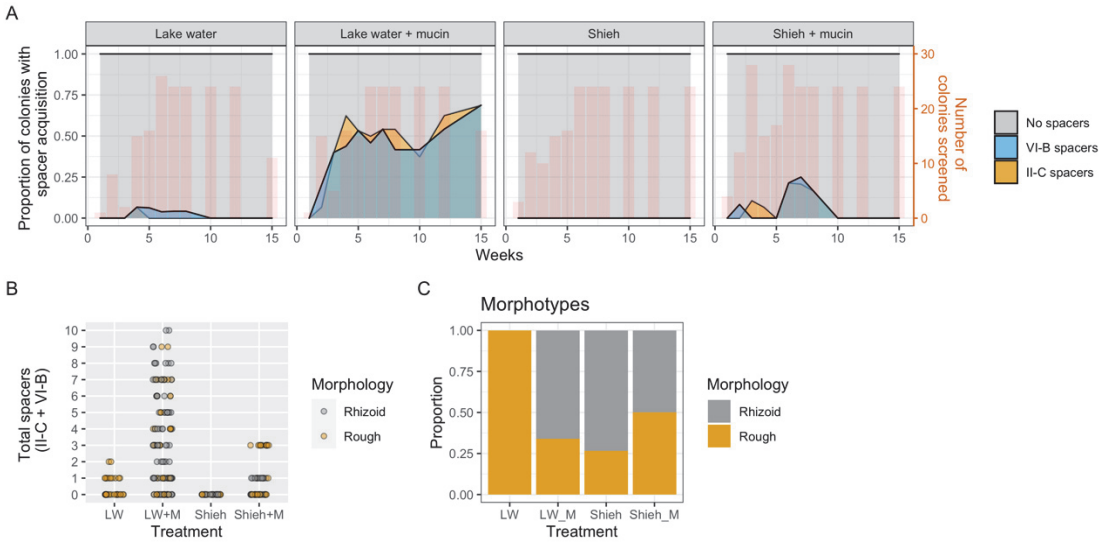

Figure 14. Spacer acquisition and morphology change in different treatments. A) Proportion of colonies that had acquired spacers throughout the 16-week experiment. Red bars indicate the number of colonies examined each week (right y-axis). B) The accumulation of spacers in single isolates from all time points, both loci combined. C) Morphology distributions of samples from all time points.

Throughout the 16-week experiment, isolates were frozen for a subsequent follow-up growth experiment. In this experiment, the isolates were grown with or without the ancestral phage under constant cell density (OD) measurement in liquid Shieh medium. While growth of the ancestral B245 control was efficiently suppressed by the phage, the samples that had coevolved with the phage previously did not show difference in maximum OD (Fig. 15A). Apart from one, all control isolates (that had coevolved in one of the four treatments without the phage) showed a similar response to the presence of phage as the ancestral B245 (Fig. 15B). The exception was the control from Shieh+M: isolates from this treatment grew significantly worse in the absence of phage, but significantly better in the presence of phage as compared to ancestor. This unexpected result suggested that prolonged incubation in Shieh+M had preconditioned the bacterium for phage resistance. Reasons for this are unknown, but an abundance of mutations that may explain these results were found on this control's genome (see below).

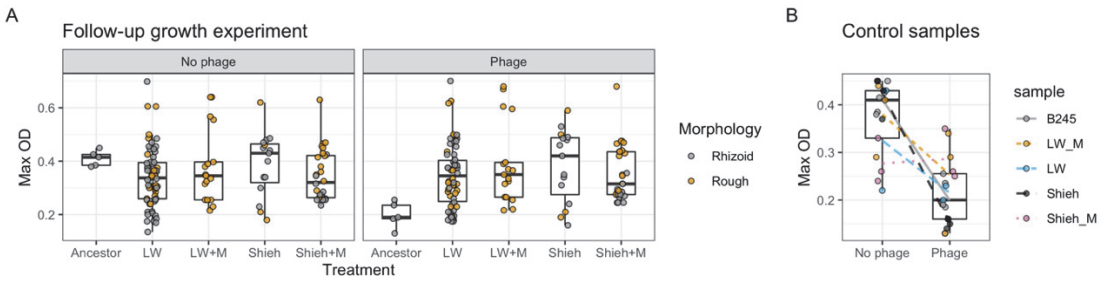

Figure 15. Maximum bacterial densities reached in the follow-up growth experiment in the presence or absence of phage, with ancestral B245 as control. A) Samples that had coevolved with phage B) Control samples that had evolved in the same treatments, but without phage.

To assess the role of spacers and morphology in the follow-up growth experiment, we looked at the spacer-rich LW+M and Shieh+M samples more closely. In LW+M, colonies with rough morphologies and zero new spacers grew significantly better in the presence of phage than rhizoid colonies with no new spacers (Fig. 16A). Strikingly, spacers had a minor yet significant negative effect on rough isolates and a strong positive effect on rhizoid isolates. In contrast, rhizoid isolates from the Shieh+M treatment showed no response to spacer acquisition while rough isolates gained a slight growth advantage (Fig. 16B). These results indicated that the effect of spacers may depend on the previous culturing condition and the current morphotype of the bacterium. Differences in spacer targets may also explain some of these dynamics, but we did not investigate them further in this study.

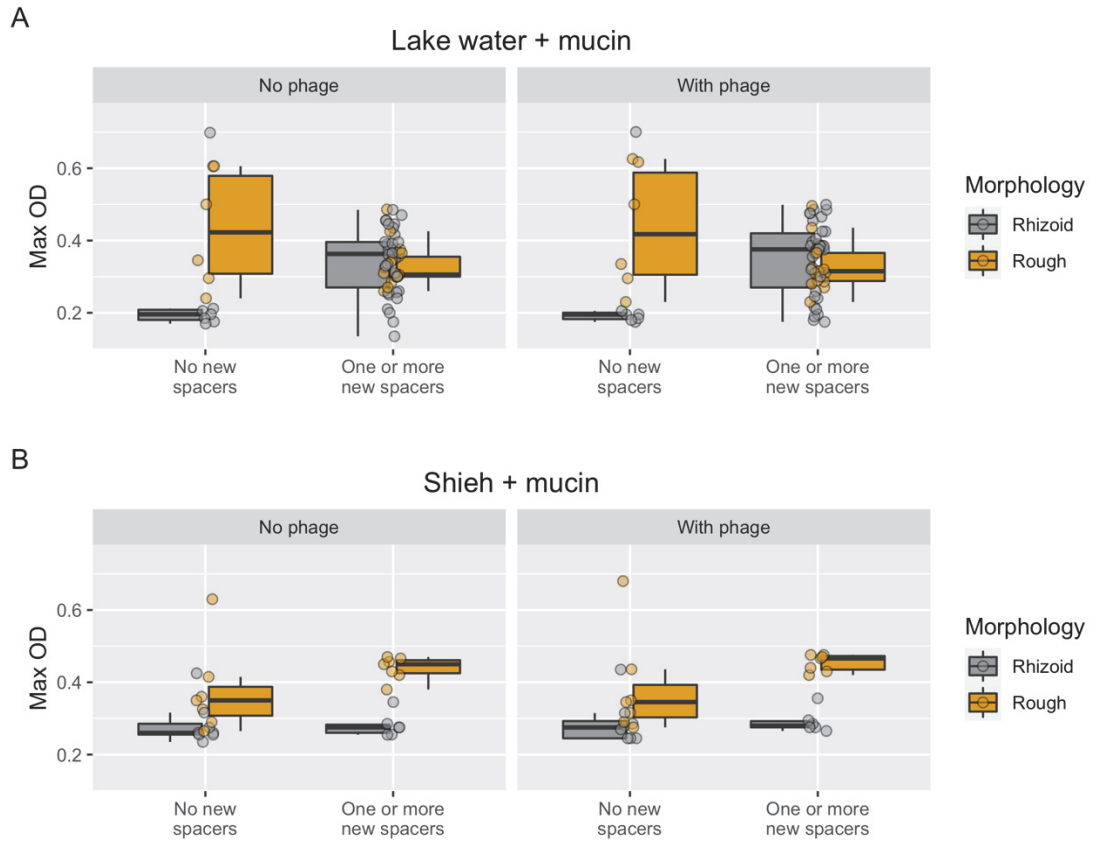

Figure 16. The effect of spacers in A) LW+M or B) Shieh+M. Isolates are pooled into having no new spacers or having at least one new spacer in either locus.

Sequencing of the bacterial genomes revealed mutations in gliding motility genes in all conditions. These mutations were either single nucleotide polymorphisms (SNP) resulting in non-synonymous codon changes or generation of premature stop codons, or larger deletions. Downregulation of some of these genes were previously reported in *F. columnare* (Penttinen *et al.* 2018). Phage immunity from disabling these genes were also shown in *F. johnsoniae* (McBride and Nakane 2015) and *F. psychrophilum* (Castillo *et al.* 2015). As these mutations were also prevalent in settings with efficient spacer

acquisition (the LW+M treatment), it is clear that both surface modification and CRISPR-Cas are utilized in phage defence and can overlap. The Shieh+M control that showed surprising resistance against phage had accumulated by far most mutations. These non-synonymous mutations included mostly metabolism-related genes (*Lon*, *rpoB*, *surE* and a histidine kinase component coding gene) whose contribution to phage resistance could not be directly assessed. A putative gene encoding a type VI secretion system -like protein also had a non-synonymous mutation. As type VI secretion systems have been associated with bacterial competition and antagonism in *Flavobacterium johnsoniae*, this mutation may be related to increased bacterial competition in the presence of host-signals (Russell *et al.* 2014). However, the causes and effects of these mutations, especially in the absence of phage, are only speculative and require further research.

The phage metapopulations at the end of the experiment were also sequenced from each replicate. Mutations were searched by comparing the reads to the ancestral V156 genome using Breseq (Deatherage and Barrick 2014). This analysis revealed few mutations that could be discerned with confidence. Replicate *b* from the Shieh treatment showed a non-synonymous mutation in an unknown gene located next to structural genes. Interestingly, this replicate was the only Shieh replicate where phage titers remained high, while the other two replicates went almost extinct. Replicate *c* (from Shieh + M treatment) showed a similar mutation in a baseplate protein and was also associated with a high phage titer (although the other two replicates in this treatment were also high). These two mutations may facilitates phage binding to bacterial surface receptors (Guttman *et al.* 2005), but require further verification.

The mutational analysis suggests that in these settings the evolutionary potential of bacteria is higher than of phages: phage mutations were only detected in two replicates, while almost all bacterial isolates had a mutation in a gliding motility related gene. It is also possible that rare phage mutations were missed by our approach that relies on amplification of the phage population using the ancestral host before DNA extraction. For this reason, any mutations that confer fitness costs in the ancestral host are likely to be underrepresented in the final DNA extraction. However, considering the high similarity between natural phage samples over a seven-year time period in study II, it is unlikely that 16 weeks could result in considerable changes in these *F. columnare* phage genomes. The slow tendency to mutate is in contrast with studies done using *S. thermophilus*, in which phage mutations occur considerably faster (Paez-Espino *et al.* 2015, Common *et al.* 2019).

This study identified a dramatic increase in spacer acquisition in the presence of mucin, especially in the low-nutrient lake water treatment. These results suggest that CRISPR-Cas activity (as measured by spacer acquisition) is increased in the presence of host-signals. This increase likely stems from the trade-off that is associated with SM-based phage resistance, as the fitness loss due to SM is amplified in the presence of host. Colonization may thus be a priority, especially for an opportunistic pathogen, that may only rarely meet its host (Brown *et al.* 2012). The results resonate with a study done on *Pseudomonas*

aeruginosa, in which CRISPR-Cas was favoured over SM when host colonization was disturbed by phages and competing species (Alseth *et al.* 2019). In the absence of competing species, the role of SM increased. While this study differs from ours by the presence of competing species, both studies highlight the fact that deployment of different defence strategies is affected by ecological factors, with the overarching goal of maximizing fitness in a given setting.

5.6 Is CRISPR-Cas active in *F. columnare*?

The studies in this thesis show that both CRISPR-Cas loci in *F. columnare* have functional acquisition in nature and in the laboratory. Phage and self-targeting spacers are inserted in both loci upon suitable conditions through the communal Cas1/2 spacer acquisition complex (and possibly helper proteins such as Cas9). What these studies did not directly address, was the efficiency of interference in either locus. If anything, study II showed that spacers have no effect on phage resistance and study IV suggested that spacers may have a positive an effect only on rhizoid colonies in specific conditions. Taken together, these results are not conclusively showing effective *F. columnare* CRISPR-Cas interference. However, several indirect lines of evidence suggest that interference is still active.

First, self-targeting by II-C spacers is strongly selected against (but permitted in the case of non-canonical PAM sequences), which suggests that the endonuclease activity of Cas9 is functional and dependent on the PAM sequence (study III). As mentioned previously, it is likely that the experiments used to assess phage resistance in study II or IV did not reflect CRISPR-Cas but rather other defences that are deployed upon strong phage pressure. It is also possible that the phages code for anti-CRISPR proteins that block Cas9 activity when targeting phages (Pawluk *et al.* 2017).

Second, assuming that spacer acquisition does not discriminate between spacers that do or do not target mRNA, the complete saturation of mRNA-targeting VI-B spacers in natural conditions is likely achieved through positive selection of such spacers (study II). For such positive selection to occur, the VI-B locus must be functional in interference. Study II also showed that in nature, phages accumulate small mutations precisely in VI-B protospacers, suggesting that this system imposes selection on the phage population.

Third, spacer acquisition itself is a rare phenomenon: the frequency of spacer acquisition in *S. pyogenes* was calculated at 1 in 10^7 cells (Heler *et al.* 2015). The rarity of acquisition suggests that expanded arrays would not be observable if they were associated with neutral selection. However, study IV showed that up to 60% of colonies had acquired at least one new spacer, suggesting that spacer acquisition is linked with positive selection, likely through interference, in *F. columnare*.

Fourth, on a broader scale, the conservation of CRISPR-Cas across different genotypes of *F. columnare* coupled with variable spacer content suggests an evolutionary advantage to maintaining these systems.

To summarize, bacterial defense mechanisms are likely to have different optimal environments and *F. columnare's* immune system may be a prime example of this. In *S. pyogenes*, CRISPR-Cas is likely to function in low-nutrient conditions where phage predation is rare in contrast to high-nutrient conditions that increase phage encounters (Westra *et al.* 2015). In our assays (studies II and IV), phage infectivity was measured in high nutrient media, therefore likely shifting bacterial defenses towards constitutive mechanisms such as SM. In *F. columnare*, this is observed through morphology change from rhizoid form to rough (Kunttu *et al.* 2009, Laanto *et al.* 2012, Zhang *et al.* 2014). In nature, however, *F. columnare* is an opportunistic pathogen that, in addition to inhabiting fish surfaces or biofilms that are high in nutrient, may also float freely in the low-nutrient water columns. In this latter condition, phage encounters are likely rare, which may also favor the alternative strategy of CRISPR-Cas despite the absence of the host.

5.7 Broader view

This thesis focused on basic, non-applied research on CRISPR-Cas biology in a native host. However, like all basic research, these studies may also have implications on applied fields. Phage therapy utilizes bacteriophages to treat bacterial diseases and shows promise in solving the many problems associated with antibiotics. However, our understanding of phage-bacterium coevolution is still lacking, especially in natural populations. The studies presented here show how *F. columnare* coevolves with its phages both by accumulating spacers in the CRISPR-Cas loci and by mutating cell surface receptors. Both (and likely many more) forms of defense must be taken into account in planning and executing phage therapy interventions. For example, simply screening spacer content of a fish farm may reveal genomic evidence of types of phages the pathogen interacts with. This information can be used in planning interventions accordingly and possibly evolving or editing phage strains that can circumvent CRISPR-Cas defense. Spacer profiling can also be utilized in tracking of epidemics and for differentiating genetically similar strains of pathogens, as speculated in study I. We have already used spacer content to identify highly clonal strains of *F. columnare* in competition experiments performed in the laboratory (Ashrafi et al., unpublished). Additionally, understanding how different CRISPR-Cas systems function in nature may enable the development of new biotechnological tools.

5.8 Conclusions

In thesis I presented four studies that characterize the CRISPR-Cas immune defense system in the fish pathogen *F. columnare*. I began by describing how aquaculture may be used for coevolutionary studies between phage and bacteria and characterized putative pathogens that can be used for this purpose (study I). The next study was a manifestation of such a study: *F. columnare* and its phages were shown coevolve in aquacultures settings over several years (study II). The two remaining studies examined the *F. columnare* CRISPR-Cas system in more detail: study III characterized spacer acquisition in detail, while study IV investigated the effect of mucin on long-term phage-bacterium coevolution in the laboratory.

I argue that CRISPR-Cas is an integral part of phage defense in *F. columnare*. This is supported by the prevalence of CRISPR-Cas systems across different *F. columnare* strains, the highly dynamic spacer content in natural populations and efficient spacer acquisition during laboratory experiments. Various observations also indicate that both loci are likely active in their interference. However, it is also clear that CRISPR-Cas only functions in a narrow range of conditions, and that other defense mechanisms, namely morphotype change through surface modification, are important. *F. columnare* has potential to become an important model species for CRISPR-Cas research and it will be interesting to see further studies done this bacterium and its phages.

From the broader perspective of CRISPR-Cas research, these studies provide information on native functioning of the II-C and VI-B systems. Both have rarely been observed in native hosts, and in interaction with native virulent phages. One of the key discoveries was that type VI-B loci are able to utilize acquisition machinery from adjacent loci despite different leader and repeat sequences.

Results inferred from laboratory-studies often overlook the diverse biotic and abiotic factors present only in natural environments, while natural sampling lack the benefits of isolating certain phenomena. When striving to attain complete understanding of a biological system, I believe both approaches are necessary. An important aim of this thesis was to contrast natural settings (study II) with experimental evolution in the laboratory (study III and IV) in the context of CRISPR-Cas. Using both viewpoints revealed aspects of this adaptive immune system that would not have been apparent in either approach alone.

Acknowledgements

There are many people (and one dog) without whom this project would not have been possible. First, I wish to thank my superb supervisors Lotta-Riina and Janne. You gave me just the right amount of guidance and support whenever needed. Many hypotheses were tested, and while most of them didn't lead anywhere, I'm grateful you gave me the opportunity to explore each one of them. This gave me important lessons in planning and executing scientific research, while knowing you always had my back. Elina L., Heidi, Gabriel, César and Marja: I asked you a million (stupid?) questions during this project and you were always ready to help without hesitation. Thank you Rogi, Anniina, Kati, Miika, Matti, Pilvi, Reetta and Elina V. (and others) for all the help and company. Thanks to Tero, Katie, Veera and Matilde for being awesome students and colleagues, and to CMBS1104 course students who participated in optimizing some of the CRISPR protocols. Petri and Alli, I don't know how the labs could ever run without you. Thanks for your help and patience. Emily and Maija, thank you for the pleasant support group meetings. Jaana, thank you for having me in the lab – I wish I could share these discoveries with you. Sylvain and rest of the phage bunker: thank you for the warm welcome in Quebec City and for teaching me your techniques. It was thrilling to visit and work in the place where many pivotal discoveries in this field were made. Mark and Rachel, thank you so much for working with us and sharing your protocols and materials. My original study cohort, aka Lampaat or the sheep-people: it was great doing the bachelor and master's period with you and I'm happy we still get keep getting together. Special thanks to Antti for sharing the flat for five years and still exchanging weird messages and questionnaires this day. Lyseon jengi, you are unique. Board game nights, Sysmä, gaming and other adventures have offered much needed contrast to this work. Hunttipäiväs with a pint of Bishop's Finger were essential during the last two years of the project. Joni, thank you for always being there for everything. My Unzymian friends, the band has been an important artistic outlet and gigs have been great experiences. I'm also grateful you gave me the time I needed to finish the PhD. Mum and dad: I was not exactly planning on following your footsteps, but somehow that just happened and here we are. Thank you for your continuous support, encouragement and inspiration. Discussions with you during the PhD were essential and provided important new perspectives. Tomppa and Timppa (+ family) it's been nice explaining my research to you and I'm lucky to have you as brothers. Finally, dear Noora: I can't remember a single time when you weren't there to support me when I needed it. You always listened when things didn't go as planned and celebrated whole-heartedly when they did. I'm happy I get to share this project and my life with you. Aada, you have shown me that dog-petting is often the key to scientific insight.

My deepest gratitude to Kone Foundation, the Academy of Finland and Foundation for Research on Viral Diseases for funding the research.

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. & Zhang F. 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science (New York, N.Y.)* 353: aaf5573.

Alkhnbashi O.S., Shah S.A., Garrett R.A., Saunders S.J., Costa F. & Backofen R. 2016. Characterizing leader sequences of CRISPR loci. *Bioinformatics* 32: i576–i585.

Almeida G.M. de F. & Sundberg L.-R. 2020. The forgotten tale of Brazilian phage therapy. *The Lancet infectious diseases* 20: e90–e101.

Almeida G.M.F., Laanto E., Ashrafi R. & Sundberg L.-R. 2019. Bacteriophage Adherence to Mucus Mediates Preventive Protection against Pathogenic Bacteria. *mBio* 10.

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.

Andersson A.F. & Banfield J.F. 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science (New York, N.Y.)* 320: 1047–1050.

Angermeyer A., Das M.M., Singh D.V. & Seed K.D. 2018. Analysis of 19 Highly Conserved Vibrio cholerae Bacteriophages Isolated from Environmental and Patient Sources Over a Twelve-Year Period. *Viruses* 10.

Barr J.J., Auro R., Furlan M., Whiteson K.L., Erb M.L., Pogliano J., Stotland A., Wolkowicz R., Cutting A.S., Doran K.S., Salamon P., Youle M. & Rohwer F. 2013. Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proceedings of the National Academy of Sciences* 110: 10771–10776.

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 (New York, N.Y.)* 315: 1709–1712.

Bernardet J.F. 1989. Flexibacter-Columnaris - 1st Description in France and Comparison with Bacterial Strains From Other Origins. *Diseases of Aquatic Organisms* 6: 37–44.

Bernardet J.F., Segers P., Vancanneyt M., Berthe F., Kersters K. & Vandamme P. 1996. Cutting a gordian knot: Emended classification and description of the

genus Flavobacterium, emended description of the family Flavobacteriaceae, and proposal of Flavobacterium hydatis nom nov (basonym, Cytophaga aquatilis Strohl and Tait 1978). *International journal of systematic bacteriology* 46: 128–148.

Bernheim A., Bikard D., Touchon M. & Rocha E.P.C. 2020. Atypical organizations and epistatic interactions of CRISPRs and cas clusters in genomes and their mobile genetic elements. *Nucleic Acids Research* 48: 748–760.

Bernheim A., Bikard D., Touchon M., bioRxiv E.R.2019. Co-occurrence of multiple CRISPRs and cas clusters suggests epistatic interactions. *biorxiv.org*.

Bin Naser I., Hoque M.M., Nahid M.A., Tareq T.M., Rocky M.K. & Faruque S.M. 2017. Analysis of the CRISPR-Cas system in bacteriophages active on epidemic strains of Vibrio cholerae in Bangladesh. *Scientific Reports* 7: 1–10.

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

Box A.M., McGuffie M.J., O'Hara B.J. & Seed K.D. 2016. Functional Analysis of Bacteriophage Immunity through a Type I-E CRISPR-Cas System in Vibrio cholerae and Its Application in Bacteriophage Genome Engineering. *Journal of bacteriology* 198: 578–590.

Brockhurst M.A., Chapman T., King K.C., Mank J.E., Paterson S. & Hurst G.D.D. 2014. Running with the Red Queen: the role of biotic conflicts in evolution. *Proceedings. Biological sciences / The Royal Society* 281.

Brouns S.J.J., Jore M.M., Lundgren M., Westra E.R., Slijkhuis R.J.H., Snijders A.P.L., Dickman M.J., Makarova K.S., Koonin E.V. & van der Oost J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science (New York, N.Y.)* 321: 960–964.

Brown S.P., Cornforth D.M. & Mideo N. 2012. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. *Trends in microbiology* 20: 336–342.

Burstein D., Sun C.L., Brown C.T., Sharon I., Anantharaman K., Probst A.J., Thomas B.C. & Banfield J.F. 2016. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nature Communications* 7: 10613.

Carte J., Wang R., Li H., Terns R.M. & Terns M.P. 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & Development* 22: 3489–3496.

Castillo D., Christiansen R.H., Dalsgaard I., Madsen L. & Middelboe M. 2015. Bacteriophage Resistance Mechanisms in the Fish Pathogen Flavobacterium psychrophilum: Linking Genomic Mutations to Changes in Bacterial Virulence Factors Björkroth J. (ed.). *Applied and Environmental Microbiology* 81: 1157–1167.

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

Cole L.A. 2016. *Biology of Life*. Academic Press.

Common J., Morley D., Westra E.R. & van Houte S. 2019. CRISPR-Cas immunity leads to a coevolutionary arms race between Streptococcus thermophilusand lytic phage. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374: 20180098–11.

Cone R.A. 2009. Barrier properties of mucus. *Advanced Drug Delivery Reviews* 61: 75–85.

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 (New York, N.Y.)* 339: 819–823.

d'Hérelle F. 1921. *Le bactériophage; son rôle dans l'immunité*. Masson et Cie, Paris.

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. *Nature Communications* 3: 945–947.

Deatherage D.E. & Barrick J.E. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. In: *Methods in Molecular Biology*, pp. 165–188.

Decaestecker E., Gaba S., Raeymaekers J.A.M., Stoks R., Van Kerckhoven L., Ebert D. & De Meester L. 2007. Host-parasite 'Red Queen' dynamics archived in pond sediment. *Nature* 450: 870–873.

Declercq A.M., Haesebrouck F., Van den Broeck W., Bossier P. & Decostere A. 2013. Columnaris disease in fish: a review with emphasis on bacterium-host interactions. *Veterinary research* 44: 27–17.

Delaney N.F., Balenger S., Bonneaud C., Marx C.J., Hill G.E., Ferguson-Noel N., Tsai P., Rodrigo A. & Edwards S.V. 2012. Ultrafast evolution and loss of CRISPRs following a host shift in a novel wildlife pathogen, Mycoplasma Gallisepticum. *PLoS Genetics* 8: e1002511.

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.

Demay C., Liens B., Burguière T., Hill V., Couvin D., Millet J., Mokrousov I., Sola C., Zozio T. & Rastogi N. 2012. SITVITWEB – A publicly available international multimarker database for studying Mycobacterium tuberculosis genetic diversity and molecular epidemiology. *Infection, Genetics and Evolution* 12: 755–766.

Deng L., Garrett R.A., Shah S.A., Peng X. & She Q. 2013. A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. *Molecular microbiology* 87: 1088–1099.

Dillard K.E., Brown M.W., Johnson N.V., Xiao Y., Dolan A., Hernandez E., Dahlhauser S.D., Kim Y., Myler L.R., Anslyn E.V., Ke A. & Finkelstein I.J. 2018. Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. *Cell* 175: 934–946.e15.

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 Biology* 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 (New York, N.Y.)* 3: eaar4120–18.

d'Herelle F. 1917. An invisible microbe that is antagonistic to the dysentery bacillus. *Comptes rendus Acad. Sci. Paris* 165: 373–375.

East-Seletsky A., O'Connell M.R., Knight S.C., Burstein D., Cate J.H.D., Tjian R. & Doudna J.A. 2016. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*: 1–17.

Erdmann S. & Garrett R.A. 2012. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Molecular microbiology* 85: 1044–1056.

Erdmann S., Shah S.A. & Garrett R.A. 2013. SMV1 virus-induced CRISPR spacer acquisition from the conjugative plasmid pMGB1 in Sulfolobus solfataricus P2. *Biochemical Society Transactions* 41: 1449–1458.

Fonfara I., Richter H., Bratovič M., Le Rhun A. & Charpentier E. 2016. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* 532: 517–521.

Fruciano E. & Bourne S. 2007. Phage as an antimicrobial agent: d'Herelle's heretical theories and their role in the decline of phage prophylaxis in the West. In: Hindawi, pp. 19–26.

Gaba S. & Ebert D. 2009. Time-shift experiments as a tool to study antagonistic coevolution. *Trends in ecology & evolution* 24: 226–232.

Garrett R.A., Vestergaard G. & Shah S.A. 2011. Archaeal CRISPR-based immune systems: Exchangeable functional modules. *Trends in microbiology* 19: 549–556.

Gasiunas G., Barrangou R., Horvath P. & Siksnys V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* 109: E2579–E2586.

Goren M.G., Doron S., Globus R., Amitai G., Sorek R. & Qimron U. 2016. Repeat Size Determination by Two Molecular Rulers in the Type I-E CRISPR Array. *Cell Reports* 16: 2811–2818.

Grissa I., Vergnaud G. & Pourcel C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8: 172.

Groenen P.M.A., Bunschoten A.E., van Soolingen D. & van Errtbden J.D.A. 1993. Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. *Molecular microbiology* 10: 1057–1065.

Guttman B., Raya R. & Kutter E. 2005. Basic phage biology. In: Kutter E. & Sulakvelidze A. (eds.), *Bacteriophages: biology and applications*, CRC press, Boca Raton, FL, pp. 26–66.

Hale C.R., Zhao P., Olson S., Duff M.O., Graveley B.R., Wells L., Terns R.M. & Terns M.P. 2009. RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* 139: 945–956.

Halpern M. & Izhaki I. 2017. Fish as Hosts of Vibrio cholerae. *Frontiers in microbiology* 8: 282.

Hamilton R., Siva-Jothy M. & Boots M. 2008. Two arms are better than one: parasite variation leads to combined inducible and constitutive innate immune responses. *Proceedings. Biological sciences / The Royal Society* 275: 937–945.

Hampton H.G., Watson B.N.J. & Fineran P.C. 2020. The arms race between bacteria and their phage foes. *Nature* 577: 1–10.

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. *Frontiers in Cellular and Infection Microbiology* 8: 7884–13.

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.

Heler R., Wright A.V., Vucelja M., Bikard D., Doudna J.A. & Marraffini L.A. 2017. Mutations in Cas9 Enhance the Rate of Acquisition of Viral Spacer Sequences during the CRISPR-Cas Immune Response. *Molecular Cell* 65: 168–175.

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. *Journal of bacteriology* 190: 1401–1412.

Howard-Varona C., Hargreaves K.R., Abedon S.T. & Sullivan M.B. 2017. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *The ISME Journal* 11: 1511–1520.

Hynes A.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. *Nature protocols* 12: 547–565.

Hynes A.P., Villion M. & Moineau S. 2014. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. *Nature Communications* 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. *Journal of bacteriology* 169: 5429–5433.

Jansen R., Embden J.D.A.V., Gaastra W. & Schouls L.M. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* 43: 1565–1575.

Janzen D.H. 1980. When is it coevolution? *Evolution* 34: 611–612.

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 (New York, N.Y.)* 337: 816–821.

Jinek M., Jiang F., Taylor D.W., Sternberg S.H., Kaya E., Ma E., Anders C., Hauer M., Zhou K., Lin S., Kaplan M., Iavarone A.T., Charpentier E., Nogales E. & Doudna J.A. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (New York, N.Y.)* 343: 1247997–1247997.

Kalatzis P., Rørbo N., Castillo D., Mauritzen J., Jørgensen J., Kokkari C., Zhang F., Katharios P. & Middelboe M. 2017. Stumbling across the Same Phage: Comparative Genomics of Widespread Temperate Phages Infecting the Fish Pathogen Vibrio anguillarum. *Viruses* 9: 122–19.

Kellner M.J., Koob J.G., Gootenberg J.S., Abudayyeh O.O. & Zhang F. 2019. SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nature protocols* 14: 2986–3012.

Kieper S.N., Almendros C. & Brouns S.J.J. 2019. Conserved motifs in the CRISPR leader sequence control spacer acquisition levels in Type I-D CRISPR-Cas systems. *FEMS Microbiology Letters* 366.

Kieper S.N., Almendros C., Behler J., McKenzie R.E., Nobrega F.L., Haagsma A.C., Vink J.N.A., Hess W.R. & Brouns S.J.J. 2018. Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Reports* 22: 3377–3384.

Kim J.G., Garrett S., Wei Y., Graveley B.R. & Terns M.P. 2019. CRISPR DNA elements controlling site-specific spacer integration and proper repeat length by a Type II CRISPR-Cas system. *Nucleic Acids Research*.

Koskella B. 2013. Phage-Mediated Selection on Microbiota of a Long-Lived Host. *Current biology* 23: 1256–1260.

Koskella B. & Brockhurst M.A. 2014. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews* 38: 916–931.

Kunttu H.M.T., Jokinen E.I., Valtonen E.T. & Sundberg L.R. 2011. Virulent and nonvirulent Flavobacterium columnare colony morphologies: Characterization of chondroitin AC lyase activity and adhesion to polystyrene. *Journal of applied microbiology* 111: 1319–1326.

Kunttu H.M.T., Sundberg L.-R., Pulkkinen K. & Valtonen E.T. 2012. Environment may be the source of Flavobacterium columnare outbreaks at fish farms. *Environmental Microbiology Reports* 4: 398–402.

Kunttu H.M.T., Suomalainen L.-R., Jokinen E.I. & Valtonen E.T. 2009. Flavobacterium columnare colony types: Connection to adhesion and virulence? *Microbial Pathogenesis* 46: 21–27.

Künne T., Kieper S.N., Bannenberg J.W., Vogel A.I.M., Miellet W.R., Klein M., Depken M., Suarez-Diez M. & Brouns S.J.J. 2016. Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Molecular Cell* 63: 852–864.

Laanto E. Exploring phage-bacterium interactions: new ways to combat a fish pathogen. University of Jyväskylä.

Laanto E., Bamford J.K.H., Laakso J. & Sundberg L.-R. 2012. Phage-Driven Loss of Virulence in a Fish Pathogenic Bacterium Brockhurst M.A. (ed.). *PloS one* 7: e53157–e53158.

Laanto E., Bamford J.K.H., Ravantti J.J. & Sundberg L.-R. 2015. The use of phage FCL-2 as an alternative to chemotherapy against columnaris disease in aquaculture. *Frontiers in microbiology* 6: 829.

Laanto E., Mäkelä K., Hoikkala V., Ravantti J.J. & Sundberg L.-R. 2020. Adapting a Phage to Combat Phage Resistance. *Antibiotics* 9.

Laanto E., Penttinen R.K., Bamford J.K.H. & Sundberg L.-R. 2014. Comparing the different morphotypes of a fish pathogen--implications for key virulence factors in Flavobacterium columnare. *BMC Microbiology* 14: 170.

Laanto E., Sundberg L.R. & Bamford J.K.H. 2011. Phage Specificity of the Freshwater Fish Pathogen Flavobacterium columnare. *Applied and Environmental Microbiology* 77: 7868–7872.

Lee K.P.K., Dey M., Neculai D., Cao C., Dever T.E. & Sicheri F. 2008. Structure of the Dual Enzyme Ire1 Reveals the Basis for Catalysis and Regulation in Nonconventional RNA Splicing. *Cell* 132: 89–100.

Lenski R.E. 1984. Coevolution of bacteria and phage: are there endless cycles of bacterial defenses and phage counterdefenses? *Journal of theoretical biology* 108: 319–325.

Lenski R.E. & Levin B.R. 1985. Constraints on the Coevolution of Bacteria and Virulent Phage - a Model, Some Experiments, and Predictions for Natural Communities. *The American Naturalist* 125: 585–602.

Levin B.R., Moineau S., Bushman M. & Barrangou R. 2013. The Population and Evolutionary Dynamics of Phage and Bacteria with CRISPR–Mediated Immunity Hughes D. (ed.). *PLoS Genetics* 9: e1003312–e1003317.

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 N., Zhu Y., LaFrentz B.R., Evenhuis J.P., Hunnicutt D.W., Conrad R.A., Barbier P., Gullstrand C.W., Roets J.E., Powers J.L., Kulkarni S.S., Erbes D.H., García J.C., Nie P. & McBride M.J. 2017. The type IX secretion system is required for virulence of the fish pathogen Flavobacterium columnare. *Applied and Environmental Microbiology* 83.

Lopatina A., Medvedeva S., Shmakov S., Logacheva M.D., Krylenkov V. & Severinov K. 2016. Metagenomic Analysis of Bacterial Communities of Antarctic Surface Snow. *Frontiers in microbiology* 7: 255–13.

Lopez-Sanchez M.-J., Sauvage E., Da Cunha V., Clermont D., Ratsima Hariniaina E., Gonzalez-Zorn B., Poyart C., Rosinski-Chupin I. & Glaser P. 2012. The highly dynamic CRISPR1 system of Streptococcus agalactiae controls the diversity of its mobilome. *Molecular microbiology* 85: 1057–1071.

Louwen R., Staals R.H.J., Endtz H.P., van Baarlen P. & van der Oost J. 2014. The Role of CRISPR-Cas Systems in Virulence of Pathogenic Bacteria. *Microbiology and Molecular Biology Reviews* 78: 74–88.

Makarova K.S., Haft D.H., Barrangou R., Brouns S.J.J., Charpentier E., Horvath P., Moineau S., Mojica F.J.M., Wolf Y.I., Yakunin A.F., van der Oost J. & Koonin E.V. 2011. Evolution and classification of the CRISPR–Cas systems. *Nature reviews. Microbiology* 9: 467–477.

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. *Nature Publishing Group* 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 A.X.E., White M.F., Yakunin A.F., Yan W., Zhang F., Garrett R.A., Backofen R., Oost J., Barrangou R. & Koonin E.V. 2019. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Publishing Group* 18: 1–17.

Marraffini L.A. & Sontheimer E.J. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science (New York, N.Y.)* 322: 1843–1845.

McBride M.J. & Nakane D. 2015. Flavobacterium gliding motility and the type IX secretion system. *Current opinion in microbiology* 28: 72–77.

Meeske A.J., Nakandakari-Higa S. & Marraffini L.A. 2019. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature* 570: 241–245.

Mir A., Edraki A., Lee J. & Sontheimer E.J. 2017. Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS chemical biology* 13: 357–365.

Modell J.W., Jiang W. & Marraffini L.A. 2017. CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* 544: 1–18.

Mojardín L. & Salas M. 2016. Global Transcriptional Analysis of Virus-Host Interactions between Phage ϕ29 and Bacillus subtilis. *Journal of virology* 90: 9293–9304.

Mojica F.J.M., Díez-Villaseñor C.S., García-Martínez J.S. & Soria E. 2005. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution* 60: 174–182.

Mojica F.J.M., Ferrer C., Juez G. & Rodríguez-Valera F. 1995. Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Molecular microbiology* 17: 85–93.

Mojica F.J.M., Juez G. & Rodríguez-Valera F. 1993. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Molecular microbiology* 9: 613–621.

Mojica F.J., Díez-Villaseñor C., Soria E. & Juez G. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology* 36: 244–246.

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 Biology* 16: 566–576.

Nordström K. & Forsgren A. 1974. Effect of protein A on adsorption of bacteriophages to Staphylococcus aureus. *Journal of virology* 14: 198–202.

Nuñez J.K., Bai L., Harrington L.B., Hinder T.L. & Doudna J.A. 2016. CRISPR Immunological Memory Requires a Host Factor for Specificity. *Molecular Cell*: 1–11.

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. *Nature Publishing Group* 21: 528–534.

Nuñez J.K., Lee A.S.Y., Engelman A. & Doudna J.A. 2015. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519: 193–198.

Nussenzweig P.M., McGinn J. & Marraffini L.A. 2019. Cas9 Cleavage of Viral Genomes Primes the Acquisition of New Immunological Memories. *Cell Host and Microbe* 26: 515–526.e516.

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. *Nature Microbiology* 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 Research* 42: 10618–10631.

Paez-Espino D., Sharon I., Morovic W., Stahl B., Thomas B.C., Barrangou R. & Banfield J.F. 2015. CRISPR Immunity Drives Rapid Phage Genome Evolution in Streptococcus thermophilus. *mBio* 6: e00262–15–9.

Pawluk A., Davidson A.R. & Maxwell K.L. 2017. Anti-CRISPR: discovery, mechanism and function. *Nature Publishing Group* 16: 12–17.

Penttinen R., Hoikkala V. & Sundberg L.-R. 2018. Gliding Motility and Expression of Motility-Related Genes in Spreading and Non-spreading Colonies of Flavobacterium columnare. *Frontiers in microbiology* 9: 12139–12.

Pinilla-Redondo R., Mayo-Muñoz D., Russel J., Garrett R.A., Randau L., Sørensen S.J. & Shah S.A. 2020. Type IV CRISPR-Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Research* 48: 2000–2012.

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.

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.

Redding S., Sternberg S.H., Marshall M., Gibb B., Bhat P., Guegler C.K., Wiedenheft B., Doudna J.A. & Greene E.C. 2015. Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell* 163: 854–865.

Rollie C., Schneider S., Brinkmann A.S., Bolt E.L. & White M.F. 2015. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *eLife* 4.

Russell A.B., Wexler A.G., Harding B.N., Whitney J.C., Bohn A.J., Goo Y.A., Tran B.Q., Barry N.A., Zheng H., Peterson S.B., Chou S., Gonen T., Goodlett D.R., Goodman A.L. & Mougous J.D. 2014. A type VI secretion-related pathway in Bacteroidetes mediates interbacterial antagonism. *Cell Host and Microbe* 16: 227–236.

Samai P., Pyenson N., Jiang W., Goldberg G.W., Hatoum-Aslan A. & Marraffini L.A. 2015. Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* 161: 1164–1174.

Seed K.D., Bodi K.L., Kropinski A.M., Ackermann H.-W., Calderwood S.B., Qadri F. & Camilli A. 2011. Evidence of a dominant lineage of Vibrio cholerae-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. *mBio* 2: e00334–10.

Seed K.D., Lazinski D.W., Calderwood S.B. & Camilli A. 2013. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* 494: 489–491.

Seed K.D., Yen M., Shapiro B.J., Hilaire I.J., Charles R.C., Teng J.E., Ivers L.C., Boncy J., Harris J.B. & Camilli A. 2014. Evolutionary consequences of intra-patient phage predation on microbial populations. *eLife* 3: e03497.

Semenova E., Savitskaya E., Musharova O., Strotskaya A., Vorontsova D., Datsenko K.A., Logacheva M.D. & Severinov K. 2016. Highly efficient primed spacer acquisition from targets destroyed by the Escherichia colitype I-E CRISPR-Cas interfering complex. *Proceedings of the National Academy of Sciences of the United States of America*: 201602639–6.

Shah S.A. & Garrett R.A. 2011. CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. *Research in microbiology* 162: 27–38.

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. & Koonin E.V. 2015. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular Cell* 60: 385–397.

Shmakov S., Shmakov S., Savitskaya E., Savitskaya E., Semenova E., Semenova E., Logacheva M.D., Logacheva M.D., Datsenko K.A., Datsenko K.A., Severinov K. & Severinov K. 2014. Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Research* 42: 5907–5916.

Silas S., Lucas-Elio P., Jackson S.A., Aroca-Crevillén A., Hansen L.L., Fineran P.C., Fire A.Z. & Sanchez-Amat A. 2017. Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. *eLife* 6: aaf5573.

Silas S., Mohr G., Sidote D.J., Markham L.M., Sanchez-Amat A., Bhaya D., Lambowitz A.M. & Fire A.Z. 2016. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science (New York, N.Y.)* 351: aad4234–aad4234.

Smargon A.A., Cox D.B.T., Pyzocha N.K., Zheng K., Slaymaker I.M., Gootenberg J.S., Abudayyeh O.A., Essletzbichler P., Shmakov S., Makarova K.S., Koonin E.V. & Zhang F. 2017. Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Molecular Cell* 65: 618–630.e7.

Staals R.H.J., Jackson S.A., Biswas A., Brouns S.J.J., Brown C.M. & Fineran P.C. 2016. Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nature Communications* 7: 1–13.

Sternberg S.H., Richter H., Charpentier E. & Qimron U. 2016. Adaptation in CRISPR-Cas Systems. *Molecular Cell*: 1–12.

Sun H., Li Y., Shi X., Lin Y., Qiu Y., Zhang J., Liu Y., Jiang M., Zhang Z., Chen Q., Sun Q. & Hu Q. 2015. Association of CRISPR/cas evolution with vibrio parahaemolyticus virulence factors and genotypes. *Foodborne Pathogens and Disease* 12: 68–73.

Sundberg L.-R., Ketola T., Laanto E., Kinnula H., Bamford J.K.H., Penttinen R. & Mappes J. 2016. Intensive aquaculture selects for increased virulence and interference competition in bacteria. *Proceedings of the Royal Society B: Biological Sciences* 283: 20153069.

Sundberg L.-R., Kunttu H.M.T. & Valtonen E.T. 2014. Starvation can diversify the population structure and virulence strategies of an environmentally transmitting fish pathogen. *BMC Microbiology* 14: 1–6.

Suomalainen L.R., Kunttu H., Valtonen E.T., Hirvelä-Koski V. & Tiirola M. 2006. Molecular diversity and growth features of Flavobacterium columnare strains isolated in Finland. *Diseases of Aquatic Organisms* 70: 55–61.

Suomalainen L.-R. 2005. Flavobacterium columnare in Finnish fish farming : characterisation and putative disease management strategies. University of Jyväskylä.

Suttle C.A. 2007. Marine viruses--major players in the global ecosystem. *Nature reviews. Microbiology* 5: 801–812.

Swarts D.C., Mosterd C., van Passel M.W.J. & Brouns S.J.J. 2012. CRISPR Interference Directs Strand Specific Spacer Acquisition Mokrousov I. (ed.). *PloS one* 7: e35888–7.

Theuretzbacher U. 2013. Global antibacterial resistance: The never-ending story. *Journal of Global Antimicrobial Resistance* 1: 63–69.

Twort F.W. 1915. An investigation on the nature of ultramicroscopic viruses. *The Lancet* 186: 1241–1243.

Tyson G.W. & Banfield J.F. 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environmental Microbiology* 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. *Proceedings. Biological sciences / The Royal Society* 282: 20151270–20151368.

van Houte S., Buckling A. & Westra E.R. 2016a. Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiology and Molecular Biology Reviews* 80: 745–763.

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. 2016b. The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* 532: 385–388.

Van Valen L. 1973. A new evolutionary law. *Evolutionary Theory* 1: 1–30.

Vercoe R.B., Chang J.T., Dy R.L., Taylor C., Gristwood T., Clulow J.S., Richter C., Przybilski R., Pitman A.R. & Fineran P.C. 2013. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genetics* 9: e1003454.

Waldor M.K. & Mekalanos J.J. 1996. Lysogenic Conversion by a Filamentous Phage Encoding Cholera Toxin. *Science (New York, N.Y.)* 272: 1910–1914.

Walther B.A. & Ewald P.W. 2004. Pathogen survival in the external environment and the evolution of virulence. *Biological Reviews of the Cambridge Philosophical Society* 79: 849–869.

Wei Y., Terns R.M. & Terns M.P. 2015. Cas9 function and host genome sampling in Type II-A CRISPR–Cas adaptation. *Genes & Development* 29: 356–361.

Weinberger A.D., Sun C.L., Pluciński M.M., Denef V.J., Thomas B.C., Horvath P., Barrangou R., Gilmore M.S., Getz W.M. & Banfield J.F. 2012. Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS computational biology* 8: e1002475.

Weissman J.L., Laljani R.M.R., Fagan W.F. & Johnson P.L.F. 2019. Visualization and prediction of CRISPR incidence in microbial trait-space to identify drivers of antiviral immune strategy. *The ISME Journal* 13: 2589–2602.

Westra E.R., Dowling A.J., Broniewski J.M. & van Houte S. 2016. Evolution and Ecology of CRISPR. *Annual Review of Ecology, Evolution, and Systematics* 47: 307–331.

Westra E.R., van Houte S., Oyesiku-Blakemore S., Ben Makin, 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. *Current biology : CB* 25: 1043–1049.

Winter C., Bouvier T., Weinbauer M.G. & Thingstad T.F. 2010. Trade-offs between competition and defense specialists among unicellular planktonic organisms: the 'killing the winner' hypothesis revisited. *Microbiology and Molecular Biology Reviews* 74: 42–57.

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

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 J., Laakso J., Mappes J., Laanto E., Ketola T., Bamford J.K.H., Kunttu H. & Sundberg L.-R. 2014. Association of colony morphotypes with virulence, growth and resistance against protozoan predation in the fish pathogen Flavobacterium columnare. *FEMS Microbiology Ecology* 89: 553–562.

Zhang Y., Heidrich N., Ampattu B.J., Gunderson C.W., Seifert H.S., Schoen C., Vogel J. & Sontheimer E.J. 2013. Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. *Molecular Cell* 50: 488–503.

ORIGINAL PAPERS

I

AQUACULTURE AS A SOURCE OF EMPIRICAL EVIDENCE FOR COEVOLUTION BETWEEN CRISPR-CAS AND PHAGE

by

Ville Hoikkala, Gabriel De Freitas Almeida, Elina Laanto &
Lotta-Riina Sundberg 2019

Philosophical Transactions of the Royal Society B:
Biological Sciences 374: 20180100

https://doi.org/10.1098/rstb.2018.0100

PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Review

Cite this article: Hoikkala V, Almeida GMF, Laanto E, Sundberg L-R. 2019 Aquaculture as a source of empirical evidence for coevolution between CRISPR-Cas and phage. *Phil. Trans. R. Soc. B* **374**: 20180100.
http://dx.doi.org/10.1098/rstb.2018.0100

Accepted: 11 November 2018

One contribution of 17 to a discussion meeting issue 'The ecology and evolution of prokaryotic CRISPR-Cas adaptive immune systems'.

Subject Areas:
ecology, evolution, molecular biology, genetics, microbiology

Keywords:
CRISPR, aquaculture, ecology, coevolution, bacteria, phage

Author for correspondence:
Lotta-Riina Sundberg
e-mail: lotta-riina.sundberg@jyu.fi

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.4395956.

THE ROYAL SOCIETY PUBLISHING

Aquaculture as a source of empirical evidence for coevolution between CRISPR-Cas and phage

Ville Hoikkala, Gabriel M. F. Almeida, Elina Laanto and Lotta-Riina Sundberg

Centre of Excellence in Biological Interactions, Department of Biological and Environmental Science and Nanoscience Center, University of Jyvaskyla, PO Box 35, 40014 Jyvaskyla, Finland

VH, 0000-0003-3995-3215; GMFA, 0000-0003-2317-5092; EL, 0000-0003-4172-3128; L-RS, 0000-0003-3510-4398

So far, studies on the bacterial immune system CRISPR-Cas and its ecological and evolutionary effects have been largely limited to laboratory conditions. While providing crucial information on the constituents of CRISPR-Cas, such studies may overlook fundamental components that affect bacterial immunity in natural habitats. Translating laboratory-derived predictions to nature is not a trivial task, owing partly to the instability of natural communities and difficulties in repeated sampling. To this end, we review how aquaculture, the farming of fishes and other aquatic species, may provide suitable semi-natural laboratories for examining the role of CRISPR-Cas in phage/bacterium coevolution. Existing data from disease surveillance conducted in aquaculture, coupled with growing interest towards phage therapy, may have already resulted in large collections of bacterium and phage isolates. These data, combined with premeditated efforts, can provide empirical evidence on phage–bacterium dynamics such as the bacteriophage adherence to mucus hypothesis, phage life cycles and their relationship with CRISPR-Cas and other immune defences. Typing of CRISPR spacer content in pathogenic bacteria can also provide practical information on diversity and origin of isolates during outbreaks. In addition to providing information of CRISPR functionality and phage–bacterium dynamics, aquaculture systems can significantly impact perspectives on design of phage-based disease treatment at the current era of increasing antibiotic resistance.

This article is part of a discussion meeting issue 'The ecology and evolution of prokaryotic CRISPR-Cas adaptive immune systems'.

1. Introduction

Bacteriophages (phages), the obligate viral parasites of bacteria, pose a constant threat of infection on their hosts. By consequence, a wide range of prokaryotic defence mechanisms have evolved. Phage infection and replication can be blocked in various ways, including preventing phage entry by modifying relevant surface receptors (surface modification, SM) or by degrading an intracellular phage genome using restriction modification (see review [1]). These well-known examples of innate defences, while central to prokaryotic immunity, lack the ability to store and update immunological memories of genetic invaders. Adaptive immunity in prokaryotes is represented by loci known as *c*lustered *r*egularly *i*nterspaced *p*alindromic *r*epeats and associated *cas* genes (CRISPR-Cas). CRISPR-Cas loci can be reprogrammed by the acquisition of distinct genetic sequences from invading nucleic acids, thereby preparing them for recognition and termination of upcoming infections with similar sequences [2–5]. CRISPR-Cas research is central to microbiology and biotechnology, but its ecological and evolutionary consequences are still surprisingly poorly understood. While several studies have addressed CRISPR-Cas and phage genomics in natural populations (see below), studies linking genetic and phenotypic data from repeatedly

sampled isolates are largely missing [6]. We believe this is partly owing to the generally low number of relatively stable phage–bacterium systems in the environment, which could be repeatedly monitored over longer timescales.

Intensive farming systems are considered as hotspots for the evolution of pathogens, as high densities of susceptible hosts promote pathogen transmission and virulence [7–9]. This feature of the farming environment is also likely to extend to biological interactions between pathogenic bacteria and their phages, owing to increased (pathogenic) bacterial population sizes and phage–bacterium interactions. Aquaculture, the farming of fishes and other aquatic species is one of the fastest growing sectors in food production, providing high-quality protein for human consumption [10]. The enrichment of naturally occurring pathogens in aquaculture environments opens an attractive opportunity to study phage–bacterium coevolution and to observe the functioning of CRISPR-Cas in semi-natural conditions.

Existing national disease surveillance projects for aquaculture-related bacteria and increasing efforts to isolate phages against these pathogens can provide essential empirical information for understanding phage–bacterium coevolution. To estimate whether generalizations derived from simplified laboratory coevolution studies reflect interaction dynamics in natural, semi-natural and clinical settings, more empirical evidence on (i) phage–bacterium interactions at wider taxonomic scale and (ii) different CRISPR-Cas systems are needed. In this review, we discuss the general properties of aquaculture as a domain for phage–bacterium coevolution and examine the position of CRISPR-Cas in these settings. From a practical point of view, it is important to understand the effects of adaptive bacterial immunity for applications such as phage therapy in aquaculture.

2. CRISPR-Cas, the adaptive immune system of prokaryotes

CRISPR-Cas operation is divided into two main stages. During the adaptation stage, a protein-complex (comprising of at least Cas1 and Cas2) excises a short sequence from an invasive phage genome. This sequence, known as a spacer, is integrated into the CRISPR array of the host's CRISPR-Cas locus [11]. In the interference phase of CRISPR-Cas, an infection leads to transcription of the CRISPR array. The transcript, called pre-crRNA (CRISPR-RNA), is processed to smaller fragments and used to guide endonucleases, such as Cas9, to corresponding positions in the phage genomes (protospacers). This results in cleavage of invading genetic material and termination of the infection. In addition to phage genomes, CRISPR-Cas systems can target plasmids and other mobile genetic elements (MGEs) and, as such, act as a barrier for horizontal gene transfer [12]. Other roles for CRISPR-Cas have also been reported, including biofilm formation, sporulation, DNA repair and regulation of virulence (see review [13]). Phages may evade CRISPR-Cas by modifying their protospacers [14] or by producing anti-CRISPR proteins [15].

The unique ability of CRISPR-Cas to store genetic information from infections, often in chronological order, opens exciting opportunities for microbial evolution research. Past infections are revealed by the CRISPR arrays, in which the oldest spacers may date back hundreds of thousands of years [16], while the acquisition of novel spacers may be monitored in almost real-time (e.g. [17]). Multiple studies have addressed spacer dynamics in natural populations [18–20] and corresponding changes in phage genomes [6,21,22]. While showing that CRISPR-Cas is active and adaptive in nature, these studies also demonstrate that the diversity of spacer content varies drastically between species, reflecting both the extent of interactions with phages and the relative importance of CRISPR-Cas amidst other defence mechanisms (see review [23]). In addition to acquiring and losing individual spacers, reassortment of CRISPR-Cas loci may be important in shaping spacer profiles in nature [20]. Spacer diversity and corresponding changes in phage protospacers may also be used as a metric for the level of asymmetry in evolutionary potential between bacteria and phages [24]. These dynamics have direct ecological and evolutionary effects in hotspots of pathogen emergence, such as aquaculture settings, but have not been thoroughly investigated with long-term sampling.

3. Aquatic and aquaculture environments as domains for phage–bacterium interaction

The complexities of natural habitats contrast the simplified settings of laboratory experiments. Dynamics of bacterial immunity observed *in vitro*, shielded from diverse biotic and abiotic factors, may therefore not be directly translatable to natural environments [24]. To clarify this separation, we review the distinct features of aquatic environments that may contribute to different ecological and evolutionary outcomes in natural settings.

(a) The ecology of aquatic environments

The probability of phage–bacterium interactions is dictated not only by the abundance, but also by the distribution of bacterial cells and phage particles. Although the numbers of bacteria and their phages in aquatic environments are enormous (e.g. [25]), their distribution is asymmetric. Bacteria congregate largely in biofilms [26,27], but the physical characteristics of water allow transmission of microbes even for long distances. Free-floating (planktonic) cells may drift in currents or actively move towards attractants or from repellents [28]. Furthermore, aquatic microbial communities are composed of several species and the abundance of each varies between microhabitats, also depending on biotic and abiotic factors [29,30].

Phages, on the other hand, have no capacity for active movement, and thus are either drifting randomly by Brownian movement in the water column or associated with organic matter or sediments [29]. A chance encounter between a phage and a bacterium may appear to be a rare event when considering their small sizes and asymmetric distribution, the large volumes of water and the near atomic distances needed for interaction. However, as phage–host interaction is a strong evolutionary driver, it is likely that mechanisms to increase encounter rates have been favoured by evolution. Firstly, and evidently, the high number of phage particles in the environment increase likelihood of encounters with the hosts. Second, a broader host range for the phage (polyvalency) may increase the chance of successful infections. While most phage isolates investigated in laboratory conditions are host-specific, polyvalency has been suggested to be prevalent in natural communities [31]. Third, although phage infections

are often considered harmful for bacteria, they may also benefit bacterial populations and promote selection for mechanisms that favour encounters resulting in relationships with mutualistic phages. Phages may contribute to bacterial pathogenicity by providing virulence factors [32], by protecting the bacterium against other phages via superinfection exclusion mechanisms (e.g. [33]), or by restructuring communities through 'killing the winner' dynamics [29].

(b) Aquatic metazoans provide territory for phage–bacterium coevolution

In aquaculture, aquatic animals (metazoans) exist in confined, high-density populations. Metazoans are covered in mucus, which provides a physical and an immunological barrier for the animal. Skin mucosal surfaces are also one of the most nutrient-rich surfaces available for aquatic microbes. By eliciting positive chemotaxis stimuli, they attract both beneficial microbes as well as pathogens, as exemplified by molecular data from European eels [34]. In this species, bacteria selected by mucus were shown to have heightened resistance against host immunity, metals, antibiotics and amoebas. Additionally, these species were abundant in genes related to biofilm formation, bacterial communication and displayed evidence of horizontal gene transfer. Mucosal surfaces may provide a natural habitat for pathogen evolution and emergence, acting as an intermediate niche between water and host that selects microbes best adapted to survive and colonize mucus.

Metazoan mucus layers have also been found to be enriched with phages [35]. This finding has led to the proposal of the bacteriophage adherence to mucus (BAM) model, which predicts an important yet so far overlooked symbiosis between metazoans and phages. Phages are concentrated by weak interactions with mucus components, creating a ubiquitous non-host derived immunity against bacterial invaders during the mucus colonization process [35]. In addition, interaction with mucins leads to subdiffusive motion patterns (in contrast to expected Brownian movement) and promotes phage persistence inside the mucosal layer despite continuous mucus shedding [36]. This would favour phage–bacteria interactions since any bacterial invader, be it pathogenic or not, would end up finding its phage when colonizing the animal. On the other hand, by interacting with mucus components, phages may solve the problem of finding hosts in an open water system by concentrating themselves on the substrate favoured by bacteria.

The enrichment of phages in mucus, coupled with the constant influx of bacteria trying to colonize this environment, makes metazoan mucosal surfaces a hotspot for phage–bacterium interactions. While these interactions can take place in the free water column or in the sediments, it has been suggested that mucus-based encounters are favoured by evolution [36]. The implications of phage–bacterium interaction on metazoan mucosal surfaces may therefore be of great importance to coevolution and to phage therapy, and has been so far overlooked. This is especially important for aquaculture systems, as fish skin and gills are covered with a mucus layer and often targeted by bacterial pathogens. The dynamics of phages and bacteria in the mucus can even be more complex since spatial structuring of mucus has been speculated to have a role in phage replication strategies [37].

(c) Effects of aquatic niches on bacterial immunity

The environmental heterogeneity of aquatic settings is likely to have consequences for bacterial immunity, as changes in phage and nutrient abundance may select for specific defence strategies. The efficiency of CRISPR-Cas has been predicted to decrease under increasing viral diversity using both theoretical models [38,39] and practical experiments [40]. As diversity correlates with mutation rates and population sizes, higher abundances of phages are predicted to result in immunity mediated by mechanisms other than CRISPR-Cas, such as SM [39]. Simply translating to aquatic settings, water columns with low phage concentration can therefore be predicted to favour CRISPR-Cas, whereas SM would be promoted by phage-rich mucosal surfaces [35]. However, opposing selective forces arise from the fact that SM often compromises the pathogens' ability to colonize their host (see review [41]). Trade-offs associated with SM can therefore also be expected to promote alternative, less costly defences in niches where colonization is prioritized. The resulting defence strategies are further complicated by other abiotic and biotic factors such as migration from the environment and the use of antibiotics (see below). Empirical evidence of alternating defence strategies in aquaculture have been found in the fish pathogen *Flavobacterium columnare*. Its CRISPR-Cas loci are active in these settings [6], but upon exposure to a high titre of phage in laboratory settings the colony morphotype changes to a phage-resistant and non-virulent one [42].

4. CRISPR-Cas and phage life cycles

Maybe surprisingly, the majority of lysogens (bacteria carrying temperate phages) also carry CRISPR-Cas systems [43]. Among aquaculture-related bacterial species, at least *Flavobacterium psychrophilum* [44] and *Vibrio anguillarum* [45] carry temperate phages that can be induced into the lytic cycle. However, the interaction between CRISPR-Cas and phage life cycles has remained poorly understood. A central dilemma arises when spacers target an integrated phage genome, as self-targeting is generally lethal. Some prophages overcome this problem by coding for anti-CRISPR proteins that suppress the immune system [15].

The phage–bacteria–metazoan mucus interactions also play a role in lysis–lysogeny switches [37], which may have significant implications for phage–CRISPR-Cas coevolution in aquaculture settings. Lysogeny may be favoured in lower mucus concentrations (outer layer) and lysis in higher concentrations (inner layers). This piggyback the winner (PtW) model would allow bacteria containing the phages to enter the mucosal layer and, when deep enough, undergo a lytic infection and release more phages [37]. Whereas the BAM model [35] can benefit phages by favouring encounters with the hosts, the PtW model [37] may benefit the bacterial hosts by favouring lysogeny. The metazoan that provides the mucosal environment benefits from both by becoming protected from invaders.

During unfavourable conditions for the host, phages can also establish alternative lifestyles such as pseudolysogeny [46]. Here, the phage chromosome is not integrated nor replicated, but inherited by one of the two daughter cells. Pseudolysogeny has been suggested to increase the effective lifespan of phage genomes by keeping it safe from outside host conditions [46]. The role of CRISPR-Cas in initiation and maintenance of this life cycle is unknown. Pseudolysogeny

Table 1. CRISPR-Cas systems in aquaculture pathogens. (Dark grey cells indicate that a feature (either no CRISPR-Cas or a CRISPR-Cas subtype) is present in all analysed strains of the given species. Light grey cells indicate presence of a feature in some of the strains. The number of strains with the specified feature is displayed in each individual cell and the total number of analysed strains is displayed in parentheses after the species. The table was compiled using CRISPRdisco [47]. Only complete genomes in NCBI's database with CRISPR-Cas loci containing both *cas* genes and CRISPR arrays were considered. Non-pathogenic subspecies were excluded from analysis. Putative CRISPR-Cas systems (type IV and V-U) were excluded from analysis similar to Crawley *et al*. [47]. For details and complete list of genomes see the electronic supplementary material, table S1.)

		class 1				class 2				
species	none	I-C	I-E	I-F	III-C	II-A	II-B	II-C	V-A	VI-B
Aeromonas salmonicida (6)[a]	6									
Edwardsiella ictaluri (3)[a]	3									
Edwardsiella piscicida-like species (2)[a]	1		1							
Edwardsiella tarda (4)[a]	4									
Flavobacterium branchiophilum (1)								1	1	1
Flavobacterium columnare (5)[a]								5		5
Flavobacterium psychrophilum (8)[a]	5							3		
Francisella noatunensis subsp. orientalis (7)	7									
Lactococcus garvieae (3)[a]	3									
Moritella viscosa (1)				1						
Mycobacterium marinum (2)	2									
Neorickettsia helminthoeca (1)	1									
Photobacterium damselae subsp. Piscicida (1)[a]	1									
Piscirickettsia salmonis (19)	19									
Pseudomonas plecoglossicida (1)[a]	1									
Renibacterium salmoninarum (1)	1									
Streptococcus agalactiae (49)[a]	2	15				47	17			
Tenacibaculum maritimum (1)	1									
Vibrio alginolyticus (14)[a]	13			1						
Vibrio anguillarum (13)[a] (see reference [45])	13									
Vibrio harveyi (4)[a]	3		1							
Vibrio parahaemolyticus (22)[a]	21			1						
Vibrio vulnificus (15)[a]	13			1	1					
Yersinia ruckeri (5)[a]	5									

[a]Species for which phages have been isolated.

might also be a prevalent life cycle in aquaculture-related phage–bacterium systems, as farming settings are often subjected to seasonal changes that cause variability in phage life cycles and phage–host interactions.

5. CRISPR-Cas in relevant aquaculture pathogens

Although there are many potential aquaculture-associated phage–bacterium systems that could elaborate how CRISPR-Cas functions in these environments, such studies are few in numbers. To inspire further research, we compiled the most important sequenced aquaculture pathogens and examined their CRISPR content using publicly available complete genome assemblies and existing publications. We determined CRISPR-Cas types of 24 aquaculture-relevant species. Eleven (approx. 46%) were found to carry a CRISPR-Cas locus in at least one strain (table 1). This analysis reveals that datasets for further CRISPR-phage coevolutionary studies already exist, as phages have been isolated against many of these bacterial species (electronic supplementary material, table S1). Since our analysis was limited to complete genomes, additional CRISPR-Cas systems are likely to arise in species that have not yet been thoroughly sequenced. Below, we highlight two groups of important aquatic pathogens derived from our analysis.

(a) *Vibrio* species

Vibrio species are abundant in aquatic environments [49,50] and many of these species are associated with diseases of farmed fishes and shrimp, known as vibriosis [51,52]. While most *Vibrio* strains are devoid of CRISPR-Cas, individual strains carrying CRISPR-Cas were found in four out of five *Vibrio* species (table 1). In addition, a recent study on *V. anguillarum* showed that the bacterium shares evolutionary history with a H20-like prophage, and that CRISPR spacers targeting this prophage are widespread across many *Vibrio* species [45]. Another study, using multiple strains of *Vibrio parahaemolyticus*, found

positive correlation between the occurrence of virulence factors and CRISPR-Cas elements [53]. *Vibrio parahaemolyticus* and its phages are also the source of the recently discovered anti-CRISPR protein acrF9 [54].

(b) *Flavobacterium* species

The genus *Flavobacterium* is comprised of 130 species, some of which infect freshwater fishes and cause major economic losses in fish farming across the globe [48]. The most important aquaculture pathogens are *F. psychrophilum*, *F. columnare* and *Flavobacterium branchiophilum*, which all carry class 2 CRISPR-Cas loci (table 1).

The *F. branchiophilum* genome contains three CRISPR-Cas loci: a type V-A locus associated with the Cas12 (Cpf1) nuclease, a type VI-B locus with the RNA-targeting ribonuclease Cas13b and a type II-C locus with Cas9. Interestingly, *F. columnare* also carries the II-C and VI-B systems and *F. psychrophilum* the II-C system. Recurrence of these class 2 systems in these species may be owing to shared evolutionary history or recent horizontal gene transfer, which is known to promote transmission of CRISPR loci across species [55,56].

Flavobacteria and their phages isolated from fish farms have already contributed to our understanding of phage–bacterium coevolution in semi-natural settings. Repeated long-term (2007–2014) sampling of *F. columnare* and its phages from an aquaculture site revealed temporal dynamics of the CRISPR-phage coevolutionary arms race [6]. Over time, the bacterial host incorporated novel, phage-matching spacers in both type II-C and type VI-B CRISPR-Cas loci. Genome sequencing of the phages revealed cases where the presence of CRISPR spacers in the host population was followed by changes in the corresponding phage protospacer regions, and even subsequent loss of spacers in the host population. However, this study also demonstrated that in addition to CRISPR-Cas, innate resistance mechanisms are also important drivers of genomic and phenotypic evolution in the phage population, which may eventually lead to a broader host range and higher infectivity of the phage.

Studies on the fish pathogen *F. psychrophilum* have demonstrated variance in the number of CRISPR-Cas loci in different strains [57]. Whereas a previous laboratory experiment suggested that CRISPR-Cas may not be active in this species [58], metagenomic samples showing high spacer diversity [59] suggest that CRISPR-Cas defence may be effective under natural conditions. Furthermore, comparison of CRISPR spacers with phage genomes has revealed that especially prophage 6H and its close relatives are ubiquitous companions of *F. psychrophilum*, with a worldwide distribution [57].

6. Practical aspects

(a) CRISPR-Cas in strain typing

CRISPR arrays may reflect previous phage infections, the rejection of plasmids or genetic matter of unknown origin [60]. In some species, spacer content is highly conserved (suggesting a lesser role for CRISPR-Cas in coevolutionary interactions), while in others the spacer profile constitutes a fingerprint that is often unique enough to distinguish otherwise nearly clonal strains of the same species (see review [23]). In fact, spacer-based typing (spoligotyping) was developed long before CRISPR's role as an immune system was uncovered [61] and has played an important role in typing strains of *Mycobacterium tuberculosis* [62]. CRISPR-typing has since been applied to many clinically relevant species, often in combination with other typing-methods [63].

If phage and MGE populations in different aquaculture facilities imprint unique spacer profiles on CRISPR-Cas positive species, the resulting diversity could be used for epidemiological and surveillance purposes. This would require establishing databases of bacterial strains, their respective spacer profiles and isolation sources, and could prove useful in tracking the spread of epidemics and characterizing bacterial diversity during outbreaks. While relevant aquaculture pathogens have not been studied from this point of view, correlations between geographical location and CRISPR arrays have been shown in other bacterial species [64–66].

(b) CRISPR-Cas and phage therapy

Phage therapy is considered to be an alternative or complement to antibiotic use, and has been used successfully in aquaculture-relevant settings related to mollusc, fish and crustacean diseases [67–69]. Understanding the role of CRISPR-Cas in phage–bacterium interactions may be central to the success of phage therapy aimed towards CRISPR-Cas positive species in aquaculture (table 1). By principle, phage therapy will significantly increase phage–bacterium interaction rates, which will promote evolution of bacterial resistance, both via innate mechanisms and CRISPR-Cas. This will have consequences for the success of phage therapy, but also for leakage of both the phage and the resistant bacterial strains into the environment. Circulation of phage-targeting CRISPR spacers in the environment has been suggested to cause corresponding evolutionary change in the phage population [6].

Maintaining up-to-date CRISPR spacer profiles of bacterial pathogens at fish farms may support phage therapy interventions. As phages are often strain-specific [70,71], they must be chosen carefully to target the prevalent bacterial community. Monitoring spacers and protospacers could also aid in experimental selection or genetic engineering of infective phages. While bacterial resistance is less likely to arise in cases where phage cocktails are used [72], monitoring the emergence of novel spacers may be used in designing new cocktail combinations.

(c) Use of antibiotics in aquaculture

Fish bacterial diseases are treated with antibiotics, which are usually given in feed. It has been estimated that 30–80% of antibiotics leak into the water owing to excretion and uneaten pellets [73]. While the antibiotic load in fishes is likely to stay at a clinical dose (thus preventing bacterial infections), minor levels of unabsorbed antibiotics are likely to affect phage–bacterium coevolution outside the host. The presence of antibiotics may increase antimicrobial resistance genes carried in the phage genomes [74,75] and contribute to prophage induction [76,77]. Both antibiotics [78,79] and phage infections [71,80] individually increase bacterial mutation rate and fitness. Interestingly, simultaneous exposure of bacteria to antibiotics and phages have been shown to increase resistance to both [81]. However, how exposure to antibiotics influence phage–bacterium coevolution and CRISPR-Cas based resistance outside laboratory conditions has remained less understood. The use of antibiotics in aquaculture and the tendency of CRISPR-Cas to target any incoming MGEs [12]

may have important consequences for the spread of antibiotic resistance. Mutants with deprecated CRISPR-Cas systems (or species with no CRISPR-Cas to begin with) may undergo positive selection during antibiotic exposure, thereby increasing the proportion of antibiotic resistant strains with increased phage sensitivity [82–85].

7. Conclusion

Aquaculture provides semi-natural and relatively stable habitats for microbial communities, enabling repeated sampling over long time periods. While most aquaculture-related pathogens, as bacteria in general, lack CRISPR-Cas, there are still many tractable species that harbour this immune system (table 1). Individual research groups and national disease surveillance laboratories have undoubtedly already collected numerous pathogenic bacteria and phage isolates from these settings over decades. While only a fraction of the collected phage and bacterial isolates have been sequenced and submitted to databases, it is evident that isolates already exist to conduct studies that can reveal details of phage-CRISPR-Cas dynamics in nature. Coupling isolate-based approaches with metagenomics may be the most effective method for scoping natural bacterial and phage communities, as this would strengthen the (often weak or missing) link between genotype and phenotype. The preferred bacterial resistance mechanisms are likely to vary across planktonic and mucosal environments in accordance with nutrient availability and phage pressure [38–40]. Therefore, models such as BAM [35] and PtW [37] need to be integrated into phage–bacterium interaction studies with fishes, other eukaryotes and mucosal surfaces. Datasets collected from aquaculture environments may also be useful in studying other cellular functions of CRISPR-Cas, such as virulence or biofilm forming capacity.

Data accessibility. This article has no additional data.

Authors' contributions. All authors participated in planning and writing the manuscript. V.H. compiled CRISPR-Cas data for aquaculture pathogens. All authors gave final approval for publication.

Competing interests. We have no competing interests

Funding. This work was supported by the Finnish Centre of Excellence Program of the Academy of Finland; the CoE in Biological Interactions 2012–2017 (no. 252411), by the Academy of Finland grant nos 266879, 304615 and 314939, by the Jane and Aatos Erkko Foundation, and by the Kone Foundation.

References

1. Labrie SJ, Samson JE, Moineau S. 2010 Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317–327. (doi:10.1038/nrmicro2315)
2. Bolotin A. 2005 Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561. (doi:10.1099/mic.0.28048-0)
3. Mojica FJM, Díez-Villaseñor CS, García-Martínez J, Soria E. 2005 Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182. (doi:10.1007/s00239-004-0046-3)
4. 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. (doi:10.1099/mic.0.27437-0)
5. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007 CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712. (doi:10.1126/science.1138140)
6. Laanto E, Hoikkala V, Ravantti J, Sundberg L-R. 2017 Long-term genomic coevolution of host-parasite interaction in the natural environment. *Nat. Commun.* **8**, 111. (doi:10.1038/s41467-017-00158-7)
7. Galvani AP. 2003 Epidemiology meets evolutionary ecology. *Trends Ecol. Evol. (Amst.)* **18**, 132–139. (doi:10.1016/S0169-5347(02)00050-2)
8. Schrag SJ, Wiener P. 1995 Emerging infectious disease: what are the relative roles of ecology and evolution? *Trends Ecol. Evol. (Amst.)* **10**, 319–324. (doi:10.1016/S0169-5347(00)89118-1)
9. Peeler EJ, Feist SW. 2011 Human intervention in freshwater ecosystems drives disease emergence. *Freshw. Biol.* **56**, 705–716. (doi:10.1111/j.1365-2427.2011.02572.x)
10. FAO. 2016 *The state of world fisheries and aquaculture 2016. Contributing to food security and nutrition for all*. Rome, Italy: FAO.
11. Barrangou R, Marraffini LA. 2014 CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell* **54**, 234–244. (doi:10.1016/j.molcel.2014.03.011)
12. Marraffini LA, Sontheimer EJ. 2008 CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845. (doi:10.1126/science.1165771)
13. Ratner HK, Sampson TR, Weiss DS. 2015 I can see CRISPR now, even when phage are gone. *Curr. Opin. Infect. Dis.* **28**, 267–274. (doi:10.1097/QCO.0000000000000154)
14. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. 2008 Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400. (doi:10.1128/JB.01412-07)
15. Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. 2013 Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432. (doi:10.1038/nature11723)
16. Touchon M, Rocha EPC. 2010 The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*. *PLoS ONE* **5**, e11126-14. (doi:10.1371/journal.pone.0011126)
17. Hynes AP, Lemay M-L, Trudel L, Deveau H, Frenette M, Tremblay DM, Moineau S. 2017 Detecting natural adaptation of the *Streptococcus thermophilus* CRISPR-Cas systems in research and classroom settings. *Nat. Protoc.* **12**, 547–565. (doi:10.1038/nprot.2016.186)
18. Tyson GW, Banfield JF. 2008 Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* **10**, 200–207. (doi:10.1111/j.1462-2920.2007.01444.x)
19. Held NL, Herrera A, Cadillo-Quiroz H, Whitaker RJ. 2010 CRISPR associated diversity within a population of *Sulfolobus islandicus*. *PLoS ONE* **5**, e12988–9. (doi:10.1371/journal.pone.0012988)
20. Held NL, Herrera A, Whitaker RJ. 2013 Reassortment of CRISPR repeat-spacer loci in *Sulfolobus islandicus*. *Environ. Microbiol.* **15**, 3065–3076. (doi:10.1111/1462-2920.12146)
21. Weinberger AD *et al.* 2012 Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput. Biol.* **8**, e1002475-16. (doi:10.1371/journal.pcbi.1002475)
22. Andersson AF, Banfield JF. 2008 Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**, 1047–1050. (doi:10.1126/science.1157358)
23. England WE, Whitaker RJ. 2013 Evolutionary causes and consequences of diversified CRISPR immune profiles in natural populations. *Biochem. Soc. Trans.* **41**, 1431–1436. (doi:10.1042/BST20130243)
24. Koskella B, Brockhurst MA. 2014 Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **38**, 916–931. (doi:10.1111/1574-6976.12072)
25. Suttle CA. 2005 Viruses in the sea. *Nature* **437**, 356–361. (doi:10.1038/nature04160)
26. Dunne WM. 2002 Bacterial adhesion: seen any good biofilms lately? *Clin. Microbiol. Rev.* **15**, 155–166. (doi:10.1128/CMR.15.2.155-166.2002)
27. Matz C, McDougald D, Moreno AM, Yung PY, Yildiz FH, Kjelleberg S. 2005 Biofilm formation and

phenotypic variation enhance predation-driven persistence of *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **102**, 16 819–16 824. (doi:10.1073/pnas.0505350102)

28. Stocker R, Seymour JR. 2012 Ecology and physics of bacterial chemotaxis in the ocean. *Microbiol. Mol. Biol. Rev.* **76**, 792–812. (doi:10.1128/MMBR.00029-12)
29. Weinbauer MG. 2004 Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* **28**, 127–181. (doi:10.1016/j.femsre.2003.08.001)
30. González JM, Suttle CA. 1993 Grazing by marine nanofiagellates on viruses and virus-sized particles: ingestion and digestion. *Mar. Ecol. Prog. Ser.* **94**, 1–10. (doi:10.3354/meps094001)
31. Chibani-Chennoufi S, Bruttin A, Dillmann M-L, Brüssow H. 2004 Phage-host interaction: an ecological perspective. *J. Bacteriol.* **186**, 3677–3686. (doi:10.1128/JB.186.12.3677-3686.2004)
32. Vica Pacheco S, García González O, Paniagua Contreras GL. 1997 The lom gene of bacteriophage lambda is involved in *Escherichia coli* K12 adhesion to human buccal epithelial cells. *FEMS Microbiol. Lett.* **156**, 129–132. (doi:10.1016/S0378-1097(97) 00415-1)
33. Vostrov AA, Vostrukhina OA, Svarchevsky AN, Rybchin VN. 1996 Proteins responsible for lysogenic conversion caused by coliphages N15 and phi80 are highly homologous. *J. Bacteriol.* **178**, 1484–1486. (doi:10.1128/jb.178.5.1484-1486.1996)
34. Carda-Diéguez M, Ghai R, Rodríguez-Valera F, Amaro C. 2017 Wild eel microbiome reveals that skin mucus of fish could be a natural niche for aquatic mucosal pathogen evolution. *Microbiome* **5**, 162. (doi:10.1186/s40168-017-0376-1)
35. Barr JJ *et al*. 2013 Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl Acad. Sci. USA* **110**, 10 771–10 776. (doi:10.1073/pnas.1305923110)
36. Barr JJ *et al.* 2015 Subdiffusive motion of bacteriophage in mucosal surfaces increases the frequency of bacterial encounters. *Proc. Natl Acad. Sci. USA* **112**, 13 675–13 680. (doi:10.1073/pnas.1508355112)
37. Silveira CB, Rohwer FL. 2016 Piggyback-the-winner in host-associated microbial communities. *NPJ Biofilms Microbiomes* **2**, 16010. (doi:10.1038/npjbiofilms.2016.10)
38. Weinberger AD, Wolf YI, Lobkovsky AE, Gilmore MS, Koonin EV. 2012 Viral diversity threshold for adaptive immunity in prokaryotes. *MBio* **3**, e00456-12. (doi:10.1128/mBio.00456-12)
39. Iranzo J, Lobkovsky AE, Wolf YI, Koonin EV. 2013 Evolutionary dynamics of the prokaryotic adaptive immunity system CRISPR-Cas in an explicit ecological context. *J. Bacteriol.* **195**, 3834–3844. (doi:10.1128/JB.00412-13)
40. Westra ER *et al.* 2015 Parasite exposure drives selective evolution of constitutive versus inducible defense. *Curr. Biol.* **25**, 1043–1049. (doi:10.1016/j.cub.2015.01.065)
41. León M, Bastías R. 2015 Virulence reduction in bacteriophage resistant bacteria. *Front. Microbiol.* **06**, 1–7. (doi:10.3389/fmicb.2015.00343)
42. Laanto E, Bamford JKH, Laakso J, Sundberg L-R. 2012 Phage-driven loss of virulence in a fish pathogenic bacterium. *PLoS ONE* **7**, e53157-8. (doi:10.1371/journal.pone.0053157)
43. Touchon M, Bernheim A, Rocha EP. 2016 Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J.* **10**, 2744–2754. (doi:10.1038/ismej.2016.47)
44. Castillo D, Espejo R, Middelboe M. 2014 Genomic structure of bacteriophage 6H and its distribution as prophage in *Flavobacterium psychrophilum* strains. *FEMS Microbiol. Lett.* **351**, 51–58. (doi:10.1111/1574-6968.12342)
45. Kalatzis P, Rørbo N, Castillo D, Mauritzen J, Jørgensen J, Kokkari C, Zhang F, Katharios P, Middelboe M. 2017 Stumbling across the same phage: comparative genomics of widespread temperate phages infecting the fish pathogen *Vibrio anguillarum*. *Viruses* **9**, 122. (doi:10.3390/v9050122)
46. Ripp S, Miller RV. 1997 The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. *Microbiology* **143**, 2065–2070. (doi:10.1099/00221287-143-6-2065)
47. Crawley AB, Henriksen JR, Barrangou R. 2018 CRISPRdisco: an automated pipeline for the discovery and analysis of CRISPR-Cas systems. *CRISPR J.* **1**, 171–181. (doi:10.1089/crispr.2017.0022)
48. Bernardet J-F, Bowman JP. 2006 The genus Flavobacterium. In *The prokaryotes* (eds M Dworkin, S Falkow, E Rosenberg, K-H Schleifer, E Stackedbrandt), pp. 481–531. New York, NY: Springer. (doi:10.1007/0-387-30747-8_17)
49. Thompson FL, Iida T, Swings J. 2004 Biodiversity of vibrios. *Microbiol. Mol. Biol. Rev.* **68**, 403–431. (doi:10.1128/MMBR.68.3.403-431.2004)
50. Takemura AF, Chien DM, Polz MF. 2014 Associations and dynamics of Vibrionaceae in the environment, from the genus to the population level. *Front. Microbiol.* **5**, 38. (doi:10.3389/fmicb.2014.00038)
51. Toranzo AE, Magariños B, Romalde JL. 2005 A review of the main bacterial fish diseases in mariculture systems. *Aquaculture* **246**, 37–61. (doi:10.1016/j.aquaculture.2005.01.002)
52. Austin B, Austin D, Sutherland R, Thompson F, Swings J. 2005 Pathogenicity of vibrios to rainbow trout (*Oncorhynchus mykiss*, Walbaum) and *Artemia nauplii*. *Environ. Microbiol.* **7**, 1488–1495. (doi:10.1111/j.1462-2920.2005.00847.x)
53. Sun H *et al.* 2015 Association of CRISPR/cas evolution with vibrio parahaemolyticus virulence factors and genotypes. *Foodborne Pathog. Dis.* **12**, 68–73. (doi:10.1089/fpd.2014.1792)
54. Pawluk A, Staals RHJ, Taylor C, Watson BNJ, Saha S, Fineran PC, Maxwell KL, Davidson AR. 2016 Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat. Microbiol.* **1**, 16085. (doi:10.1038/nmicrobiol.2016.85)
55. Chakraborty S, Snijders AP, Chakravorty R, Ahmed M, Tarek AM, Hossain MA. 2010 Comparative network clustering of direct repeats (DRs) and cas genes confirms the possibility of the horizontal transfer of CRISPR locus among bacteria. *Mol. Phylogenet. Evol.* **56**, 878–887. (doi:10.1016/j.ympev.2010.05.020)
56. Godde JS, Bickerton A. 2006 The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J. Mol. Evol.* **62**, 718–729. (doi:10.1007/s00239-005-0223-z)
57. Castillo D, Christiansen RH, Dalsgaard I, Madsen L, Espejo R, Middelboe M. 2016 Comparative genome analysis provides insights into the pathogenicity of *Flavobacterium psychrophilum*. *PLoS ONE* **11**, e0152515-18. (doi:10.1371/journal.pone.0152515)
58. Castillo D, Christiansen RH, Dalsgaard I, Madsen L, Middelboe M. 2015 Bacteriophage resistance mechanisms in the fish pathogen *Flavobacterium psychrophilum*: linking genomic mutations to changes in bacterial virulence factors. *Appl. Environ. Microbiol.* **81**, 1157–1167. (doi:10.1128/AEM.03699-14)
59. Lopatina A, Medvedeva S, Shmakov S, Logacheva MD, Krylenkov V, Severinov K. 2016 Metagenomic analysis of bacterial communities of Antarctic surface snow. *Front. Microbiol.* **7**, 398. (doi:10.3389/fmicb.2016.00398)
60. Shmakov S *et al.* 2017 Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182. (doi:10.1038/nrmicro.2016.184)
61. Groenen PMA, Bunschoten AE, van Soolingen D, van Errtbden JDA. 1993 Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol. Microbiol.* **10**, 1057–1065. (doi:10.1111/j.1365-2958.1993.tb00976.x)
62. Demay C *et al*. 2012 SITVITWEB—a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect. Genet. Evol.* **12**, 755–766. (doi:10.1016/j.meegid.2012.02.004)
63. Louwen R, Staals RHJ, Endtz HP, van Baarlen P, van der Oost J. 2014 The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiol. Mol. Biol. Rev.* **78**, 74–88. (doi:10.1128/MMBR.00039-13)
64. Wietz M, Millán-Aguiñaga N, Jensen PR. 2014 CRISPR-Cas systems in the marine actinomycete *Salinispora*: linkages with phage defense, microdiversity and biogeography. *BMC Genomics* **15**, 936. (doi:10.1186/1471-2164-15-936)
65. Cui Y *et al.* 2008 Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS ONE* **3**, e2652-10. (doi:10.1371/journal.pone.0002652)
66. Sorokin VA, Gelfand MS, Artamonova II. 2010 Evolutionary dynamics of clustered irregularly interspaced short palindromic repeat systems in the ocean metagenome. *Appl. Environ. Microbiol.* **76**, 2136–2144. (doi:10.1128/AEM.01985-09)
67. Doss J, Culbertson K, Hahn D, Camacho J, Barekzi N. 2017 A review of phage therapy against bacterial pathogens of aquatic and terrestrial organisms. *Viruses* **9**, 50. (doi:10.3390/v9030050)

68. Laanto E, Bamford JKH, Ravantti JJ, Sundberg L-R. 2015 The use of phage FCL-2 as an alternative to chemotherapy against columnaris disease in aquaculture. *Front. Microbiol.* **6**, 829. (doi:10.3389/fmicb.2015.00829)
69. Richards GP. 2014 Bacteriophage remediation of bacterial pathogens in aquaculture: a review of the technology. *Bacteriophage* **4**, e975540. (doi:10.4161/21597081.2014.975540)
70. Laanto E, Sundberg LR, Bamford JKH. 2011 Phage specificity of the freshwater fish pathogen *Flavobacterium columnare*. *Appl. Environ. Microbiol.* **77**, 7868–7872. (doi:10.1128/AEM.05574-11)
71. Holmfeldt K, Middelboe M, Nybroe O, Riemann L. 2007 Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their *Flavobacterium* hosts. *Appl. Environ. Microbiol.* **73**, 6730–6739. (doi:10.1128/AEM.01399-07)
72. Górski A *et al.* 2016 Phage therapy: combating infections with potential for evolving from merely a treatment for complications to targeting diseases. *Front. Microbiol.* **7**, 1878–1879. (doi:10.3389/fmicb.2016.01515)
73. Cabello FC, Godfrey HP, Tomova A, Ivanova L, Dölz H, Millanao A, Buschmann AH. 2013 Antimicrobial use in aquaculture re-examined: its relevance to antimicrobial resistance and to animal and human health. *Environ. Microbiol.* **15**, 1917–1942. (doi:10.1111/1462-2920.12134)
74. Modi SR, Lee HH, Spina CS, Collins JJ. 2013 Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**, 219–222. (doi:10.1038/nature12212)
75. Lekunberri I, Subirats J, Borrego CM, Balcázar JL. 2017 Exploring the contribution of bacteriophages to antibiotic resistance. *Environ. Pollut.* **220**, 981–984. (doi:10.1016/j.envpol.2016.11.059)
76. Allen HK, Looft T, Bayles DO, Humphrey S, Levine UY, Alt D, Stanton TB. 2011 Antibiotics in feed induce prophages in swine fecal microbiomes. *MBio* **2**, 1–9. (doi:10.1128/mBio.00260-11)
77. Maiques E, Ubeda C, Campoy S, Salvador N, Lasa I, Novick RP, Barbé J, Penadés JR. 2006 Beta-lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. *J. Bacteriol.* **188**, 2726–2729. (doi:10.1128/JB.188.7.2726-2729.2006)
78. Perron GG, Hall AR, Buckling A. 2010 Hypermutability and compensatory adaptation in antibiotic-resistant bacteria. *Am. Nat.* **176**, 303–311. (doi:10.1086/655217)
79. Melnyk AH, Wong A, Kassen R. 2014 The fitness costs of antibiotic resistance mutations. *Evol. Appl.* **8**, 273–283. (doi:10.1111/eva.12196)
80. Buckling A, Rainey PB. 2002 Antagonistic coevolution between a bacterium and a bacteriophage. *Proc. R. Soc. Lond. B* **269**, 931–936. (doi:10.1098/rspb.2001.1945)
81. Cairns J, Frickel J, Jalasvuori M, Hiltunen T, Becks L. 2017 Genomic evolution of bacterial populations under coselection by antibiotics and phage. *Mol. Ecol.* **26**, 1848–1859. (doi:10.1111/mec.13950)
82. Palmer KL, Gilmore MS. 2010 Multidrug-resistant enterococci lack CRISPR-Cas. *MBio* **1**, 297. (doi:10.1128/mBio.00227-10)
83. Price VJ, Huo W, Sharifi A, Palmer KL. 2016 CRISPR-Cas and restriction-modification act additively against conjugative antibiotic resistance plasmid transfer in *Enterococcus faecalis*. *mSphere* **1**, e00064-16-13. (doi:10.1128/mSphere.00064-16)
84. 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**, e1003844-13. (doi:10.1371/journal.pgen.1003844)
85. Bikard D, Hatoum-Aslan A, Mucida D, Marraffini LA. 2012 CRISPR interference can prevent natural transformation and virulence acquisition during *in vivo* bacterial infection. *Cell Host Microbe* **12**, 177–186. (doi:10.1016/j.chom.2012.06.003)

II

LONG-TERM GENOMIC COEVOLUTION OF HOST-PARASITE INTERACTION IN THE NATURAL ENVIRONMENT

by

Elina Laanto, Ville Hoikkala, Janne Ravantti & Lotta-Riina Sundberg 2017

Nature Communications 8: 931

https://doi.org/10.1038/s41467-017-00158-7

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-017-00158-7 OPEN

Long-term genomic coevolution of host-parasite interaction in the natural environment

Elina Laanto [1], Ville Hoikkala[1], Janne Ravantti[2] & Lotta-Riina Sundberg [1]

Antagonistic coevolution of parasite infectivity and host resistance may alter the biological functionality of species, yet these dynamics in nature are still poorly understood. Here we show the molecular details of a long-term phage–bacterium arms race in the environment. Bacteria (*Flavobacterium columnare*) are generally resistant to phages from the past and susceptible to phages isolated in years after bacterial isolation. Bacterial resistance selects for increased phage infectivity and host range, which is also associated with expansion of phage genome size. We identified two CRISPR loci in the bacterial host: a type II-C locus and a type VI-B locus. While maintaining a core set of conserved spacers, phage-matching spacers appear in the variable ends of both loci over time. The spacers mostly target the terminal end of the phage genomes, which also exhibit the most variation across time, resulting in arms-race-like changes in the protospacers of the coevolving phage population.

[1] Centre of Excellence in Biological Interactions, Department of Biological and Environmental Science and Nanoscience Center, University of Jyvaskyla, PO box 35, FI-40014 Jyvaskyla, Finland. [2] Department of Biosciences and Institute of Biotechnology, Viikinkaari 9, University of Helsinki, FI-00014 Helsinki, Finland. Correspondence and requests for materials should be addressed to L.-R.S. (email: lotta-riina.sundberg@jyu.fi)

One of the fundamental questions in evolutionary biology and medicine is how host immunity drives the evolution of pathogens, as all living organisms are exposed to infections that can cause significant harm and selection in the host population. The antagonistic coevolutionary arms race of parasite infectivity and host resistance leads to adaptations and counter-adaptations in the coevolving partners[1–4], and also has a central role in the evolution of host–parasite relationships in the microbial world. This arms race has been shown especially under experimental settings, where lethal infections by bacterial viruses, (bacterio)phages, shape the diversity and dynamics of the coevolving host bacterial populations[1, 2, 4, 5], whereas the phages have the capacity to rapidly overcome host immunity[1, 5–9]. It has been suggested, however, that the adaptive potential of phage and high costs in phage resistance in bacteria can limit phage–bacterium coevolution[10]. Yet, the arms-race dynamics observed in laboratory experiments substantially differ from the real-life dynamics under the complex web of surrounding interactions present in the environment, which influences the ecology and evolution of both phages and their hosts[11, 12].

CRISPR (clustered regularly interspaced short palindromic repeats) and associated *cas* genes form the bacterial CRISPR-Cas adaptive immune system against phages and plasmids. CRISPR-Cas protects bacteria from infections by cleaving invading nucleic acid sequences (protospacers) that are identical or nearly identical to the spacers in the bacterial CRISPR repeat-spacer array[13, 14]. Protospacer adjacent motifs (PAMs) are used to distinguish self from non-self and are crucial in most CRISPR systems[15], while protospacer flanking sites (PFSs) determine the efficiency of interference in RNA-targeting CRISPR systems but have no known autoimmunity-related functions[16]. Over time, novel spacers accumulate in one end of the array and may therefore confer resistance to multiple phages. However, in accordance with antagonistic coevolution theory, phages can counter-adapt to host immunity via mutations in the protospacer or PAM/PFS sequences[5, 17, 18], or by using anti-CRISPR proteins[19–22].

Following evolutionary change in natural communities is challenging, and linking genomic change data with the coevolutionary dynamics of phages and bacteria is one of the key challenges in microbiology. Therefore, a comprehensive view of host–parasite coevolutionary dynamics in the environment, surrounded by a network of other trophic interactions, is missing. In fact, previous studies have concentrated either on phenotypic changes[7, 23] or CRISPR genetics using metagenomic approaches[24–26]. Particularly, empirical evidence on the relative importance of constitutive and adaptive resistance (mutations preventing successful phage life cycle and CRISPR loci insertions, respectively[27]) in bacteria and the corresponding changes in the parasitic phage[28] supported by phenotypic data from natural settings is missing. Constitutive and adaptive resistance mechanisms are predicted to be important in different ecological conditions[29–31], therefore teasing apart the relative role of different bacterial resistance mechanisms is essential for understanding microbial community ecology. From an applied perspective, such information is also crucial for the development phage therapy applications[32], where evolution of resistance is expected to limit the functionality of the treatment over time.

The importance and volume of aquaculture is growing steadily to meet the increasing need for high-quality protein for human consumption[33], and new methods such as phage therapy, to control and manage bacterial diseases are under vigorous research[34, 35]. Here, we characterize the coevolution of populations of bacteriophages and their bacterial hosts (i.e., the fish pathogen *Flavobacterium columnare*[36]) at the phenotypic and genetic level, in a flow-through aquaculture setting during the period 2007–2014. We sampled for bacterial and phage isolates from the fish farming facilities or immediate surroundings and performed all-against-all cross-infections to analyze bacterial resistance patterns. To link the phenotypic patterns with molecular changes, we sequenced phage genomes to understand the determinants of the host range and characterized the bacterial CRISPR loci to understand the role of adaptive immunity in driving genome evolution and host range in the phage population. We report the patterns of arms race coevolution in a natural community and demonstrate that whereas bacteria evolve resistance, phages evolve a broader host range over time, which is associated with increase in genome size. We also observed evolutionary change in phage genomes in response to bacterial adaptive (CRISPR) and constitutive immunity, exemplifying the importance of both resistance mechanisms in natural bacterial populations. We also demonstrate a type VI-B CRISPR system functioning in its natural host and in a natural setting.

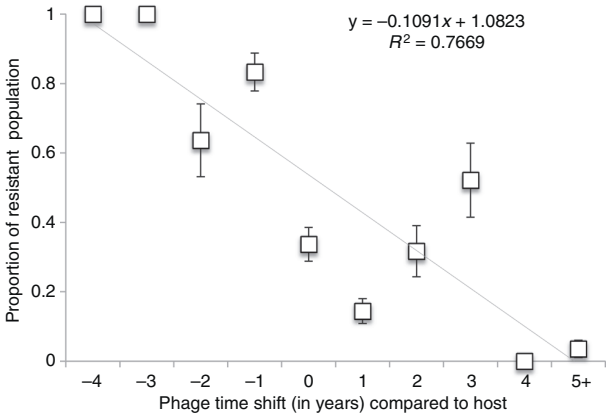

Fig. 1 Time shift of phage–bacterium coevolution in the fish farming environment. Phage–bacterium coevolution measured as mean proportion (± S.E.) of resistant host population, described as annual relative change in host resistance when exposed to phages from the past (−4 to −1 years), contemporary (0), and future (+1 to +5 years). The bacteria are in general resistant against phages from the past but susceptible to infection by phages from contemporary and future time points (24% and 18% resistant, respectively; GLMM, $F_{(2,\ 507)} = 15.099$, $p < 0.001$)

Results

Phage and bacterial isolates. *F. columnare* phages are genotype-specific, each infecting strains of only one genotype[37]. This specificity allowed us to analyze both phage and host isolates over long time scales in locally adapting populations. The phage and bacteria were collected from a fish farm in Central Finland mainly during the warm water season when outbreaks occur. During the period 2007–2013, we isolated 17 *F. columnare* isolates that belong to the genetic group C (Supplementary Table 1). Furthermore, we isolated 30 dsDNA phages (belonging to the family Myoviridae) during 2009–2014, which infect specifically this bacterial host group (Supplementary Table 2).

Coevolution of phage infectivity and bacterial resistance. First, we analyzed the phage host range by cross-infecting all 30 phage isolates with all 17 bacterial isolates. The most recently isolated phages had the widest host range, being able to infect nearly all bacterial hosts. Analyzed using a time-shift approach[38], the phage isolation time point had a significant effect on infectivity compared to the bacterial host. The bacteria were in general resistant against phages from the past but susceptible to infection by phages from contemporary and future time points (24% and 18%

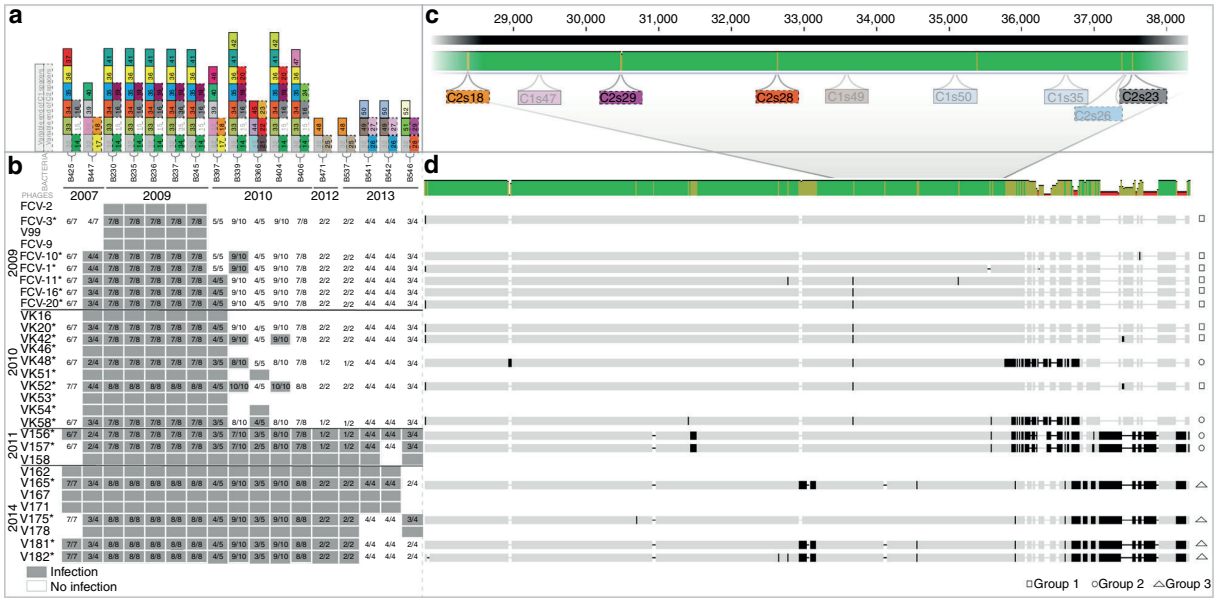

Fig. 2 Arms race coevolution of phage infectivity and host resistance at genetic level. The effect of long-term genomic coevolution on phage infectivity and genome, and on host bacterium CRISPR content. **a** Spacer diversity in the variable ends of both CRISPR loci (C1 and C2) is indicated by numbered and colored *rectangles* above each bacterial strain (each number and color referring to a specific spacer within a locus). C1 spacers are marked with *solid lines* (left spacer column) and C2 spacers with *dotted lines* (right spacer column). Numbers and colors are locus-specific and only variable spacers are shown. All spacers except C1s32, C1s38, and C2s15 are targeting the phages. **b** Host range of phages (rows) infecting *Flavobacterium columnare* isolates (columns). *Gray rectangles* indicate infection (presence of plaques) and *white rectangles* indicate resistance. The first number within these *rectangles* indicates the number of CRISPR spacers with an identical match in the phage genome and the second number the maximum number of phage-targeting spacers the bacterium has. **c** Close-up: CRISPR protospacer regions mapped to a ~11 kb portion of the multiple sequence alignment, where the four highlighted protospacers indicate specific, possibly CRISPR-driven, changes in these areas. **d** Patterns of molecular evolution in the phage genomes in response to host range. Each sequenced phage genome follows the corresponding host range row from **b**. *Green color* in the consensus sequence (above) indicates identical sequence, *yellow* 30%, and *red* < 30% identity. Note that also the bar height changes accordingly. *Black* indicates nucleotide differences and insertion elements

resistant, respectively; generalized linear mixed model (GLMM), $F_{(2,\ 507)} = 15.099$, $p < 0.001$). The phage isolation year included as a random factor in this analysis had a significant effect (Wald $Z = 2.959$, $p = 0.003$), while bacterial isolation place within the farm (inlet water, outlet water, fish) or phage enrichment host did not ($Z = 1.36$, $p = 0.174$, $Z = 0.586$, $p = 0.558$, respectively). All bacteria were resistant to phages from 3–4 years from the past, and, on the other hand, bacterial resistance was below 0.5% against phages from 4–5 years in the future (Fig. 1). In addition, enrichment host did not affect the phage host range (number of infected hosts, Kruskal–Wallis $\chi^2 = 11.7$, df = 7, $p = 0.103$).

We tested phage adsorption efficiency of eight phages on nine bacteria isolated in different years (Supplementary Fig. 1) to see whether the role of cell surface modifications over other resistance mechanisms could be detected. Adsorption efficiency of the phages differed from 0 to 91%. In some cases, adsorption was less efficient in all or some of the resistant strains indicating the possibility of resistance via surface modifications and hence the lower adsorption efficiency for the phage. Also, increased adsorption was detected in some of the resistant strains (e.g., VK42 to B471 and VK58 to B406), which suggest post-adsorption resistance mechanisms in these pairs. In addition, the efficiency of plating (EOP), adjusted to the titers on strain B230 that was infected by all phages (Supplementary Table 3), did not relate to the observed adsorption.

When analyzed using generalized linear mixed models with adsorption percentage as the response variable, resistance was found to have no significant effect in adsorption. Phages originating from a previous time point (as compared to the bacterium) were predicted to have reduced adsorption percentage (approximately one-third of that of contemporary phages, GLMM, $F_{(2,\ 212)} = 6.506$, $Z = -2.722$, $p = 0.006$). No significant difference in adsorption was observed between contemporary and future phages.

Host resistance drives phage genome evolution. To understand phenotypic resistance at the genomic level, we sequenced genomes of 17 phage isolates from 2009 to 2014. This resulted in complete genomes with lengths ranging from 46,481 to 49,084 bp (Fig. 2; Supplementary Table 2; Supplementary Fig. 2). Phage genomes from different years were highly similar, enabling us to study differences at the nucleotide level. From the predicted 74 (2009) to 76 (2014) open reading frames (ORFs), 52 were 100% identical between all genomes, including putative terminase, portal protein, and several structural proteins (e.g., major capsid protein). However, we observed non-synonymous differences in putative tail and structural proteins (two phages from 2011 in ORFs 27, 28, and three phages from 2014 in ORFs 36 and 37, Supplementary Fig. 2). This could indicate a conventional arms race between phage infectivity and host resistance via receptor mutation in natural settings (see refs [8, 9] for similar experimental evolution results).

In multiple sequence alignment, the phage genomes clustered in three major groups (Fig. 2, marked with symbols). Genomes of phages in Group 1 (seven phages from 2007 to 2009) were

identical. In Group 2 phages (four phages from 2010 to 2011), genomic changes were located at 38,922–43,395 bp (consensus sequence, in phage VK48 starting from 38,421 bp). According to Blastp analysis, this genomic region might contain ORFs associated with phage replication. Furthermore, these differences were missing in the Group 1 and Group 3 phages (Supplementary Discussion).

The broadest host range phages (Group 3 with four phages; isolated in 2014) had changes in putative structural proteins (ORFS 36 and 37, see above). Interestingly, they, as well as two phages isolated in 2011 from Group 2 had extra DNA at the terminal end of their genomes, coding additional predicted ORFs. All these additional ORFs are located upstream of ORF66, which has a putative ICE (integrative and conjugative element) domain. Together with changes in the bacterial CRISPR spacer content (see below), these genomic differences seem to correlate with a broader phage host range. The additional ORFs were assigned as hypothetical proteins. One of these ORFs (no. 67.1) has a conserved domain belonging to the family of N-acyltransferases (Supplementary Fig. 2; Supplementary Table 4). Another ORF (no 74.1) is a putative YopX protein, having also mobile and extrachromosomal element and prophage functions.

In addition, we found indications that phage infectivity may significantly differ in the three genomic groups (Groups 1, 2, and 3, indicated in Fig. 2). Infectivity was interpreted from EOP values (Supplementary Table 3; Kruskal–Wallis, $\chi^2 = 38.571$, df = 2, $p < 0.001$, Supplementary Fig. 3). However, it should be noted that here we were not able to test the plaque production against contemporary bacterial isolates, which may influence the results.

Characterization of the CRISPR loci and PAM sequences. Next, using a previously published *F. columnare* genome sequence[39] and our own sequence data, we identified two CRISPR loci found in all 17 *F. columnare* isolates used in this study, designated here as C1 (CRISPR1, ATCC49512: 391567–394479) and C2 (CRISPR2, ATCC49512: 1680008–1680571) (Supplementary Fig. 4; Supplementary Table 5). C1 contains Cas9 followed by a repeat-spacer array and two smaller *cas* genes (Cas1 and Cas2) in opposing directions (Supplementary Fig. 4). On the basis of *cas* gene composition[14] and Cas9 homology[40], C1 is a type II-C CRISPR locus. C2 represents the recently discovered type VI-B CRISPR system with its large RNA-targeting single-component endonuclease Cas13b. We confirm that Csx27 and Csx28, small *cas* gene regulators associated with some type VI-B systems, are missing from *F. columnare* as noted by Smargon et al.[41]. Two putative transposases were also discovered immediately downstream of this locus.

Repeat-spacer arrays from both loci were sequenced and corresponding protospacers searched in the phage genomes. Across all bacterial isolates, a total of 52 unique C1 spacers were discovered, 18 of which were targeting phage genomes in this study on both coding and non-coding strands (8 and 10 spacers, respectively), and both intergenic and predicted ORF regions (5 and 13 spacers, respectively). A total of 29 unique spacers were discovered in C2. Interestingly, all 15 C2 spacers are targeting predicted ORFs on the phage coding strand, supporting the notion that type VI-B systems may be targeting viral transcripts[41].

Protospacer distribution in the phage genomes was biased toward the terminal end of the genomes (Supplementary Fig. 5) and differed significantly from an expected uniform distribution in both loci (one-tailed permutation test, C1 $p = 0.001$, C2 $p = 0.016$). Roughly half of C1 protospacers were located in the fourth quadrant of the phage genomes, with C2 protospacers following a similar yet less pronounced trend.

Analyzing regions surrounding C1 protospacers revealed a putative downstream PAM sequence of NNNNNTAAAA (Supplementary Table 6). A similar analysis of C2 protospacer PFSs supported the findings of a previous study[41] and also implied possible involvement of C1 acquisition machinery in C2 spacer acquisition (Supplementary Table 7; Supplementary Discussion).

Specificity of the phage–bacterium coevolution. Over time, all bacterial strains displayed variation in spacer content while maintaining a large set of conserved spacers shared among all isolates. The variable ends did not only vary in sequence, but also in the number of spacers, with the surprising finding that the most recent bacteria had generally fewer spacers than older isolates (Fig. 2a). Whereas none of the conserved spacers were targeting phages, the variable ends of the arrays consisted almost solely of phage-targeting spacers (Fig. 2a; Supplementary Data 1), exemplifying how exclusively CRISPR evolution is driven by the phages in close proximity in time and space. Individual bacteria often contained multiple spacers matching individual phages (Supplementary Fig. 5), demonstrating the specificity of the host–parasite interaction. Furthermore, the incorporation of novel DNA in the ends of the Group 2 and 3 phage genomes was followed by the appearance of two CRISPR spacers targeting these areas (C2s22, C1s52). Surprisingly, however, one spacer (C2s16) was already present in the bacteria (Fig. 2a) before the emergence of a novel protospacer-containing sequence in the phages.

In addition to establishing that new bacterial spacers mostly originate from phages, we also considered possible coevolutionary dynamics between the observed spacer and protospacer sequences. Specifically, we wanted to see if phage protospacer or PAM/PFS sequences vary after the appearance of phage-targeting spacers in the bacterial population. Overall, 6 of the 18 phage-targeting C1 spacers and 9 of the 15 phage-targeting C2 spacers had altered protospacer sequences in phages isolated after the appearance of the corresponding spacer in the bacterial population (Supplementary Fig. 5). Some protospacer alterations resulted from large re-organizations in the terminal ends of the genomes, while some were due to very specific single-nucleotide polymorphisms or deletions within the protospacer sequences. Most specific sequence variations took place on a roughly 11 kbp area (27,352–38,174 in the consensus sequence). This area is targeted by four C1 and five C2 spacers (Fig. 2c; Supplementary Table 8), and includes six mutational sites (1–12 bp in length). Four of these mutational sites are located within the C2 protospacers (spacers C2s18, C2s23, C2s28, and C2s29) and their appearance correlates chronologically with the introduction of the corresponding spacers to the bacterial population (Supplementary Fig. 5). Following the appearance of these alternative protospacers, these spacers were no longer detected in later bacterial isolates. Interestingly, the ancestral protospacer sequence (targeted by spacers C2s18 and C2s23) reappeared in later phage isolates, which could suggest that bacterial strains harboring these spacers no longer co-exist.

Currently, knowledge on class 2 affiliated anti-CRISPR proteins is very limited. On the basis on the findings of a recent study[22], we searched for homologs of the anti-CRISPR proteins AcrIIA1, AcrIIA2, AcrIIA3, and AcrIIA4 from all the phage genomes in our data set. No significant hits were recovered.

Despite the apparent coevolutionary dynamics displayed by the CRISPR arrays and phage genomes, linking CRISPR data with phenotypic infection patterns did not, in most cases, reveal clear correlation between bacterial resistance and spacer content (Fig. 2b).

Discussion

Our results highlight the long-term arms-race coevolution of hosts and parasites in an aquaculture environment down to the molecular level. In accordance with previous findings[1, 7–9], the bacterial population was generally resistant against phages from past time points and susceptible to infections by (broad host range) phages from the future. Further molecular characterization pinpointed genetic changes in both host and phage populations that fit the assumption of arms-race coevolution. This allows estimations of the role of constitutive and adaptive immunity as drivers of parasite infectivity in environmental populations. These results are especially important for economically relevant conditions that might be suitable for phage therapy approaches.

The evolution of generalist parasites has been shown to arise in response to host resistance in experimental settings[1, 42]. Combined with phenotypic resistance patterns, our data suggest that the evolution of host immunity may have caused directional selection on phage infectivity and host range via genomic expansion (Fig. 2; Supplementary Fig. 3). In correlation with higher infectivity (Supplementary Fig. 1), the genomes of the most recently isolated phages had larger genome size, increasing from 46,448 bp (in 2009) to 49,121 bp (in 2014). Phage host range was independent of the enrichment host used in the original phage isolation. An earlier isolated enrichment host was always used (except for first five phages in 2009 (Supplementary Table 2), however, similar phages could be isolated even when using different enrichment hosts (e.g., phage pairs VK156 and VK157, and FCV-10 and FCV-11).

All additional DNA was concentrated in the terminal regions of the genomes. Although no clear functions for the three additional ORFs were predicted, structural similarity to a YopX-like protein was found. Although this protein is an outer membrane protein of *Yersinia*, it has been suggested to be linked with the phage life cycle[43]. Together with the other additional ORFs (and deletion of less efficient ones), it could provide an improved infection cycle. Indeed, when studied over all bacterial isolates, the host generalist phages isolated most recently were most infective (Supplementary Fig. 3). This likely results from a combination of phage characteristics, but further genetic studies are needed for confirmation. From an evolutionary perspective, parasite host range is expected to correlate negatively with fitness[44]. Our finding of parallel increases in host range and infectivity demonstrates that the evolutionary trajectories of the phage–bacterium interactions are not yet fully understood.

Genome and gene family expansions have been shown to be the starting points of evolutionary innovations in eukaryotic parasites[45]. These gene family expansions are often located in the recombinogenic parts of the genome, which could also apply to the phage genomes in our data set, as the additional ORFs are found upstream of ORF66, which contains a putative ICE domain. ICEs integrate in chromosomes and mediate horizontal gene transfer in bacteria[46], but their possible functionality in viruses has not been characterized. Nevertheless, these phage genome expansions indicate the role of the surrounding environment as a source of novel DNA and highlight the importance of long-term environmental sampling in understanding phage evolution. Yet, as the physical characteristics of the capsid strongly constrain phage genome size, our data raise new questions on what limits the phage host range, virulence, and genome evolution, and what the associated trade-offs are.

The presence of CRISPR-Cas systems has been found in ~40% of studied bacterial species[40], and the molecular mechanisms of different CRISPR systems are under active research[47]. Some repeat-spacer arrays remain relatively constant over millennia[48], while in some species no two identical CRISPR loci are found even in cells living in close proximity[25]. However, previous CRISPR studies have mainly concentrated on unraveling the mechanisms of CRISPR immunity in laboratory conditions, whereas the role of CRISPR immunity for microbial evolutionary ecology in nature is poorly understood.

Studies on a related bacterium, *Flavobacterium psychrophilum*, have discovered inactive CRISPR systems across multiple strains[49], but also high variability between metagenomic samples from nature[50]. In *F. columnare*, we report two active CRISPR loci: C1, a type II-C system and C2, a recently identified RNA-targeting type VI-B system. Two ORFs downstream of C2, whose translations show possible transposase motifs, may be remnants from possible horizontal transfer of this locus, see e.g., ref. [40].

Both of the CRISPR loci accumulate novel (mostly phage-targeting) spacers extending the backbone of conserved (not phage-targeting) spacers (Fig. 2a). On several occasions, the introduction of CRISPR spacers to the bacterial population was followed by the appearance of phage isolates with modifications in the corresponding protospacer regions. This is consistent with studies demonstrating the appearance of CRISPR-escape mutants due to altered protospacers[5, 17] or PAM regions[51]. In the four most specific protospacer alterations (Fig. 2c), the changes were non-synonymous, leading to alternative amino-acid sequences in the predicted ORFs. While promoting evasion of host CRISPR immunity, these mutations are also likely to have a fitness cost for the phage. Assuming these non-synonymous mutations arose before synonymous ones by chance, a possible explanation for their prevalence may be the large benefit gained from CRISPR-evasion. In this scenario, the costs of these mutations would only be unmasked after the disappearance of CRISPR-based selection pressure, leading to the re-emergence of the ancestral protospacer sequence. However, our data in this respect are limited and these predictions require further investigation.

Spacer acquisition is still one of the least known aspects of CRISPR[52]. We found an unequal distribution of protospacers in phage genomes, indicating that CRISPR defense targets the terminal end of the phage genomes. A recent study showed that in the type II-A CRISPR system of *Streptococcus pyogenes*, spacer acquisition begins while phage dsDNA is being injected inside the cell[53]. This results in polarization of spacer acquisition, with most spacers originating from the phage genome end that first enters the cell. This has been suggested to enrich spacers targeting this region during selection and less efficient interference by spacers targeting the opposite end. Typically, in tailed phages (excluding T4) the end of the packaged genome is the first end to be ejected[54]. While the packaging and injecting direction of the phage genomes in our data set is unknown, protospacer distribution indicates similar polarization as in *S. pyogenes*. Interestingly, this genome end also contains most of the variation among the phages, providing a possible link between the proposed model of spacer acquisition and the observed patterns of phage genome reorganization.

The size of the variable ends of the CRISPR loci varied between the years (Fig. 2a). Surprisingly, the most recent and resistant bacterial isolates (originating from rainbow trout) contained the least number of phage-targeting spacers. While the bacterial isolation source did not significantly affect the resistance patterns (Fig. 2b), it is possible that these bacteria have experienced more phage encounters during bacterial disease outbreaks, which would influence the resistance mechanisms and patterns. It is also possible that the absolute number of spacers is not a decisive factor in determining resistance and that spacers vary in effectiveness based on their target sequence (see e.g., ref. [41]).

Despite the coevolutionary dynamics observed between spacer content and phage genomes, CRISPRs could not completely explain the resistance patterns of the host, possibly due to high phage pressure in the experimental setup[55]. However, under low-

phage selection (such as natural aquatic settings) CRISPRs may still be preferred over mutation-based immunity due to their minimal effects on overall fitness of the host[30], at least when not constantly expressed[56]. Consistent with a previous suggestion[57], the presence of the high diversity of CRISPR spacers (e.g., in 2010) could have reduced phages' ability to overcome bacterial resistance by point mutations, and may have caused larger shifts in the structure of the phage population, observed here as phages with increased genome size. In addition, adsorption percentages in infectious and non-infectious phage–bacterium pairs showed no clear correlation, suggesting that adsorption does not guarantee infection. Considering the costs related to maintaining broad resistance against phages, combining constitutive immunity and diverse CRISPR immunity may be the most successful strategy for a natural bacterial population[58], as the phage have the potential to rapidly overcome these bacterial defense mechanisms to ensure their own persistence. The expanding host range driven by increased genome size is likely to be restricted by the physical capacity of the capsid. Nevertheless, other factors, such as novel anti-CRISPR functions, that limit the evolutionary potential of this specific phage–bacterium interaction have yet to be discovered. Furthermore, the expansion of the host range may trade-off with phage infectivity[44], which might be the point at which the adaptive CRISPR defense alone would be sufficient to overcome those infections.

The bacterial and phage isolates used in this study originate from a fish farming environment, a semi-natural system with a constant flow of incoming environmental water. It is possible that although attempts to isolate *F. columnare* phages from waters outside fish farming have been mainly unsuccessful[37], phages and bacteria may have already interacted in nature, and the farming system enriches the phage population sizes to detectable levels. However, as a man-made environment, the phage–bacterium interaction is subjected to selection by farming practices, especially high fish densities and the use of medication, which may select for the most virulent and antibiotic tolerant *F. columnare* strains[59]. Thus, by providing essential information on the phage–bacterium interactions under multifactorial selection in these settings, our data and study system are directly relevant for sustainable use of phage therapy applications in field conditions. For example, in laboratory conditions, eliciting phage resistance in *F. columnare* has been shown to cause reversible morphotypic changes that can maintain constitutive resistance[60, 61]. These resistant morphotypes are, however, missing from natural isolates, probably because of their impaired growth, tolerance to protozoan predation and capacity to infect fish[60, 62]. The absence of resistant morphotypes indicates higher benefits of CRISPR immunity under multifactorial selection in field conditions, although our data show that both constitutive and adaptive bacterial resistance mechanisms are likely to be active and important in natural settings.

Methods

Phage and bacterial isolates used in this study. Phage and bacteria were isolated from a private fish farm (from fish and tank water) in Central Finland and from its immediate surroundings from inlet and outlet water (Supplementary Tables 1 and 2). The isolates were originally collected for other purposes, therefore we do not have isolates from each year, and from same source, and also different enrichment hosts were used for phage isolation over the years. The interaction between *F. columnare* and its phages is genotype-specific (each phage type infecting strains of only one bacterial genotype[37]), which allowed us to monitor the evolution of phage–bacterium relationship in this study using a specific host genotype.

Bacterial strains were originally isolated from water samples and from the fish by plating 100 µl of water or gill swabs on Shieh agar[63] or Shieh agar supplemented with tobramycin[64], as previously described[65]. The plates were incubated in room temperature for 48 h, after which individual yellow rhizoid colonies exhibiting the typical colony morphology of *F. columnare* were picked, pure-cultured, and stored frozen at −80 °C, with 10% fetal calf serum and 10% glycerol. Among the collected bacterial isolates over the years from this farm, bacteria belonging to genetic group C were chosen based on ARISA genotyping using the methodology described previously[65, 66]. In brief, the intergenic spacer region (between the 16S and 23S ribosomal genes) of *F. columnare* isolates was amplified by PCR. The products were analyzed using ABI 3130*xl* Genetic Analyzer to assign the isolates into previously determined genetic groups.

Phages were isolated from water samples by enrichment. A filtered water sample (pore size 0.45 µm, Nalgene) was used to dilute five-fold Shieh medium, and an overnight-grown enrichment host (belonging to genetic group C) was inoculated in this medium. Cultures were grown at room temperature (23 °C), at 110 r.p.m. on a benchtop shaker (New Brunswick Scientific) until they turned turbid. A double agar layer method was used for detecting phages: 300 µl of turbid sample with 3 ml of 0.7% soft Shieh agar (tempered to 47 °C) was applied on solid Shieh agar. After incubation (23 °C) for 24–48 h (depending on the bacterial growth) plaques were picked. Three rounds of plaque purification were performed for each phage isolate. For further characterization, phage lysates were prepared adding 5 ml of culture media on a plate with confluent lysis, and shaken at 8 °C, at 95 r.p.m. on a benchtop shaker for 6 h. Phage isolates were stored frozen in −80 °C with 20% glycerol. Phages were characterized by genome restriction profiles using prior EcoRI experiments, and for this study we used all the phage isolates infecting the bacterial genetic group C.

Phage host range. For analyzing phage infectivity and host range, a fresh phage lysate was prepared using host strain C4, which is a non-virulent mutant of *F. columnare*[67]. An overnight-grown bacterial culture (in Shieh medium, 23 °C, 110 r.p.m.) of the test strains was plated using the double agar overlay method by mixing bacterial culture 1:10 with 0.7% soft Shieh agar tempered to 47 °C and pouring this on agar plates. The phage was diluted until million fold (10^{-6}), and 10 µl samples of each dilution was spotted on top of the bacterial lawn. All bacterial strains were infected with all phages in a pairwise manner. After incubation of 48 h, the plates were checked for single plaques. Plaque formation was considered as a positive result for phage infectivity.

Time-shift experiment and statistical analysis. Bacterial resistance in response to phages from past, contemporary, and future time points was analyzed with generalized linear mixed models using SPSS 22.0. This was done using the infection data presented in Fig. 2, and we calculated the proportion of resistance of bacteria in relation to the temporal difference of isolation of each phage. In the statistical analysis, the response variable of phage infection was encoded as a binary trait, therefore binomial distribution with logit link transformation was used. Time shift was used as a fixed factor and phage isolation year, bacterial isolation place and phage enrichment host as random factors, with bacterial identity as a blocking factor (subject). Only phages able to produce plaques in the host bacteria were considered infective in the analysis.

Adsorption assay. Adsorption efficiency of eight phages to nine bacterial hosts (Supplementary Fig. 1) during 20 min was measured following the method by Kropinski 2009[68] with modifications. In brief, bacteria were grown to exponential phase and diluted to ~3×10^8 colony forming units ml^{-1}. A total volume of 3 ml was used in the assay: 0.3 ml of phage suspension (adjusted to 1–3×10^5 plaque forming units (PFU) ml^{-1}) was added to 2.7 ml of bacterial culture. Phage added to Shieh medium without bacteria was used as a control. The assays were done in three replicates. After 20 min a sample of 50 µl was diluted to 950 µl of cold growth media supplemented with three drops of chloroform. PFU ml^{-1} of the free phage particles was determined by plating 100 µl of the sample with C4 strain using the double agar overlay method. Adsorption data were analyzed using a generalized linear mixed model with adsorption percentage as the response variable (binomial distribution and logit link) using R 3.3.3 and RStudio 1.0.136. Time shift (past, contemporary, and future phage) and bacterial resistance were used as explanatory variables and replicate identities as random factors. In the few cases where the observed number of plaques in VK20 exceeded the control values due to measurement error due to instrumental or observational error (i.e., having 110 plaques resulting from a bacterial adsorption tube, when only 100 plaques were extracted from the control tube), the adsorption percentage was fixed to 0.

Phage genome sequencing. Phages subjected to genome sequencing were chosen based on their infection profiles (Fig. 2). Phage DNA was isolated using a protocol described by Santos[69] with slight modifications. Briefly, after RNase (10 µg ml^{-1}) and DNase (1 µg ml^{-1}) treatment phages were precipitated with 40 mM $ZnCl_2$ and incubated for 5 min followed by pelleting (15,000 × *g*, 5 min) and resuspended in TES buffer (0.1 M Tris-HCl, pH 8; 0.1 M EDTA; 0.3% SDS). DNA was purified using a GeneJET Genomic DNA isolation kit column (Fermentas). About 100 ng of DNA was subjected to genome sequencing with Illumina MiSeq in Institute for Molecular Medicine Finland (FIMM), apart from isolate FCV-1, which was sequenced using Roche 454 at LGC Genomics, Germany.

Genome assembly. The phage genomes were assembled using Velvet-assembler (v. 1.2.10)[70] with optimal k-mers per genome (average coverage ~×1600). Differences appearing in only one genome were checked and corrected if needed. Single

insertions and nucleotide differences were checked by sequencing directly from the genome using specific primers (Supplementary Table 7). ORF's were predicted using GeneMarkS (v. 4.28)[71] and Glimmer (v. 3.02b)[72]. Protein Homology/analogY Recognition Engine V 2.0 (PHYRE^2)[73] was used for predicting protein structures.

CRISPR sequencing and analyses. Bacterial DNA was extracted from overnight-grown turbid cultures using the GeneJET Genomic DNA isolation kit (Fermentas). We used the previously published *F. columnare* ATCC 49512 genome[39] and our own unpublished sequence data as references in designing primers for amplifying the CRISPR loci. NetPrimer (Premier Biosoft) was used for the analysis. Primers (Supplementary Table 9) were provided by Sigma. CRISPR loci from the extracted DNA were amplified in a PCR reaction (PIKO Thermal cycler by Finnzymes) using the Phusion Flash II polymerase in a 20 µl reaction with a template concentration of 0.5 ng µl^{-1}. The protocol was completed following the manufacturer's instructions. The annealing temperature was 65 °C for CRISPR1 and 67 °C for CRISPR2, and the elongation step was 30 s for CRISPR1 and 27 s for CRISPR2. The reactions were purified for sequencing using the QIAquick PCR Purification Kit (Qiagen). The amplified PCR products were sequenced with the Sanger method. Raw data were transformed using SequenceAnalysis 6 basecalling. Two replicates were done for each read and the consensus sequences were manually determined using Geneious 9.1.4 (Biomatters Ltd.).

The determination of the orientations of the CRISPR loci (namely, the transcription direction of the repeat-spacer arrays) was based on analyses using CRISPRDetect[74], comparisons with other similar CRISPR systems and observations of the polarization of spacer acquisition. Unlike in most CRISPR systems, the crRNA in type II-C systems is transcribed starting from the conserved end[75]. By assuming this transcription direction, we could detect a conserved PAM sequence downstream of C1 protospacers. As all type II systems described so far exhibit downstream PAMs, we assume that the C1 repeat-spacer array is also transcribed starting from the conserved end. Very low confidence values associated with C1 direction prediction by CRISPRDetect led us to not use this tool with this locus. However, CRISPRDetect was able to produce confident direction prediction values for the C2 repeat-spacer arrays, which are in line with those of a previous study[31] and which place the transcription initiation point at the variable end, as in most CRISPR systems. This direction also enables the spacers to target messenger RNA (mRNA) from predicted viral ORFs and produces PFSs in line with those of a previous study[31].

All spacers across all isolates were pooled and labeled according to their CRISPR locus, creating global spacer pools for both loci. Spacers were numbered starting from the conserved end of the oldest strain, B425 (in which the oldest C1 spacer is called C1s1 and the most recent one C1s37). After labeling all B425 C1 spacers, we proceeded to the next strain (B447) and searched for any novel spacers, labeling them by adding to the global C1 numerical sequence (e.g., C1s38). Eventually, each C1 and C2 spacer in the total spacer pool had a locus-specific ID. Each spacer may be isolate-specific (e.g., C2s29) or shared by two or more isolates (e.g., C1s48). In C1, a 114 bp area containing degenerate repeats between spacers 4 and 5 was skipped, and numbering was resumed downstream of this area.

The C1 PAMs are reported using the guide-centric approach[76]. Due to the RNA-targeting nature of C2, the PFSs reported here are obtained from the single-stranded mRNA of the predicted ORFs using the target-centric approach.

Protospacer location analysis. Possible bias in the distribution of protospacers on the phage genomes was analyzed by comparing the actual protospacer location means against simulated sets of protospacers distributed randomly and uniformly over a 50,666 bp long genome (the size of the consensus sequence). The mean protospacer positions (18 spacers for C1, 15 spacers for C2) of each of these permutations (100,000 simulations) was used as the statistic for creating a null distribution. The observed means from the collection of 18 C1 (mean position 35706.61) and 15 C2 (mean position 33342.33) protospacers were compared to this distribution and *p*-values calculated as the proportion of the simulated means exceeding the observed ones over the total number of simulations. These analyses were conducted with R 3.3.3 and RStudio 1.0.136.

It must be noted that the possible effects of primed spacer adaptation have not been taken into account in these analyses. Priming has not been shown to affect spacer acquisition beyond class I systems[52] and by current knowledge requires the cascade complex and the Cas3 nuclease[77], which are missing from the CRISPR loci of *F. columnare*.

Analysis. Geneious (v. 9.1.4) was used for the annotation and analysis of the CRISPR sequence data, and phage PAM and PFS extractions. WebLogo (v. 3.5.0) was used to create sequence logos of the PAM and PFS sequences.

Data availability. Phage genome sequences are available in GenBank (accession numbers KY951963, KY951964, KY979235-KY979247, KY992519, KY992520). All other data are included in the supplementary files and available from the authors.

Received: 23 January 2017 Accepted: 2 June 2017
Published online: 24 July 2017

References

1. Buckling, A. & Rainey, P. Antagonistic coevolution between a bacterium and a bacteriophage. *Proc. R. Soc. B Biol. Sci.* **269**, 931–936 (2002).
2. Brockhurst, M. A. et al. Running with the Red Queen: the role of biotic conflicts in evolution. *Proc. R. Soc. B Biol. Sci.* **281**, 20141382 (2014).
3. Bohannan, B. J. M. & Lenski, R. E. Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecol. Lett.* **3**, 362–377 (2000).
4. Brockhurst, M. A., Buckling, A. & Rainey, P. B. The effect of a bacteriophage on diversification of the opportunistic bacterial pathogen, *Pseudomonas aeruginosa*. *Proc. R. Soc. B Biol. Sci.* **272**, 1385–1391 (2005).
5. Paez-Espino, D. et al. CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. *MBio.* **6**, e00262–15 (2015).
6. Gómez, P. & Buckling, A. Bacteria-phage antagonistic coevolution in soil. *Science* **332**, 106–109 (2011).
7. Koskella, B. Phage-mediated selection on microbiota of a long-lived host. *Curr. Biol.* **23**, 1256–1260 (2013).
8. Paterson, S. et al. Antagonistic coevolution accelerates molecular evolution. *Nature* **464**, 275–U154 (2010).
9. Scanlan, P. D., Hall, A. R., Lopez-Pascua, L. D. C. & Buckling, A. Genetic basis of infectivity evolution in a bacteriophage. *Mol. Ecol.* **20**, 981–989 (2010).
10. Lenski, R. E. & Levin, B. R. Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. *Am. Nat.* **125**, 585–602 (1985).
11. Betts, A., Rafaluk, C. & King, K. C. Host and parasite evolution in a tangled bank. *Trends Parasitol.* **32**, 863–873 (2016).
12. Parratt, S. R. & Laine, A.-L. The role of hyperparasitism in microbial pathogen ecology and evolution. *ISME J.* **10**, 1815–1822 (2016).
13. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
14. Makarova, K. S. et al. An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
15. Leenay, R. T. et al. Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. *Mol. Cell* **62**, 137–147 (2016).
16. Abudayyeh, O. O. et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016).
17. Levin, B. R., Moineau, S., Bushman, M. & Barrangou, R. The population and evolutionary dynamics of phage and bacteria with CRISPR–mediated immunity. *PLoS Genet.* **9**, e1003312 (2013).
18. Deveau, H. et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
19. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013).
20. Bondy-Denomy, J. et al. Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature* **526**, 136–139 (2015).
21. Pawluk, A. et al. Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat. Rev. Microbiol.* **1**, 16085 (2016).
22. Rauch, B. J. et al. Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell* **168**, 150–158.e10 (2017).
23. Decaestecker, E. et al. Host–parasite 'Red Queen' dynamics archived in pond sediment. *Nature* **450**, 870–873 (2007).
24. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**, 1047–1050 (2008).
25. Tyson, G. W. & Banfield, J. F. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* **10**, 200–2007 (2007).
26. Weinberger, A. D. et al. Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput. Biol.* **8**, e1002475 (2012).
27. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317–327 (2010).
28. Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defences. *Nat. Rev. Microbiol.* **11**, 675–687 (2013).
29. Hamilton, R., Siva-Jothy, M. & Boots, M. Two arms are better than one: parasite variation leads to combined inducible and constitutive innate immune responses. *Proc. R. Soc. B Biol. Sci.* **275**, 937–945 (2008).
30. Westra, E. R. et al. Parasite exposure drives selective evolution of constitutive versus inducible defense. *Curr. Biol.* **25**, 1043–1049 (2015).
31. van Houte, S., Buckling, A. & Westra, E. R. Evolutionary ecology of prokaryotic immune mechanisms. *Microbiol. Mol. Biol. Rev.* **80**, 745–763 (2016).
32. Scanlan, P. D., Buckling, A. & Hall, A. R. Experimental evolution and bacterial resistance: (co)evolutionary costs and trade-offs as opportunities in phage therapy research. *Bacteriophage* **5**, e1050153 (2015).
33. Food and Agriculture Organization of the United Nations. *The state of world fisheries and aquaculture. Opportunities and Challenges* (FAO, 2014).
34. Romero, J., Feijoo, C. G. & Navaterre, P. *Health and Environment in Aquaculture*, 159–198 (InTech Europe, 2014).

35. Richards, G. P. Bacteriophage remediation of bacterial pathogens in aquaculture: a review of the technology. *Bacteriophage* **4**, e975540 (2014).
36. Declercq, A. M., Haesebrouck, F., Van den Broeck, W., Bossier, P. & Decostere, A. Columnaris disease in fish: a review with emphasis on bacterium-host interactions. *Vet. Res.* **44**, 27 (2013).
37. Laanto, E., Sundberg, L.-R. & Bamford, J. K. H. Phage specificity of the freshwater fish pathogen *Flavobacterium columnare*. *Appl. Environ. Microbiol.* **77**, 7868–7872 (2011).
38. Gaba, S. & Ebert, D. Time-shift experiments as a tool to study antagonistic coevolution. *Trends Ecol. Evol.* **24**, 226–232 (2009).
39. Tekedar, H. C. et al. Genome sequence of the fish Pathogen *Flavobacterium columnare* ATCC 49512. *J. Bacteriol.* **194**, 2763–2764 (2012).
40. Burstein, D. et al. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* **7**, 10613 (2016).
41. Smargon, A. A. et al. Cas13b Is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx28. *Mol. Cell* **65**, 618–630.e (2017).
42. Hall, A. R., Scanlan, P. D. & Buckling, A. Bacteria-phage coevolution and the emergence of generalist pathogens. *Am. Nat.* **177**, 44–53 (2011).
43. Yasmin, A. et al. Comparative genomics and transduction potential of *Enterococcus faecalis* temperate bacteriophages. *J. Bacteriol.* **192**, 1122–1130 (2010).
44. Leggett, H. C., Buckling, A., Long, G. H. & Boots, M. Generalism and the evolution of parasite virulence. *Trends Ecol. Evol.* **28**, 592–596 (2013).
45. Leckenby, A. & Hall, N. Genomic changes during evolution of animal parasitism in eukaryotes. *Curr. Opin. Genet. Dev.* **35**, 86–92 (2015).
46. Johnson, C. M. & Grossman, A. D. Integrative and conjugative elements (ICEs): what they do and how they work. *Annu. Rev. Genet.* **49**, 577–601 (2015).
47. Amitai, G. & Sorek, R. CRISPR–Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* **14**, 67–76 (2016).
48. Savitskaya, E. et al. Dynamics of *Escherichia coli* type I-E CRISPR spacers over 42,000 years. *Mol. Ecol.* **26**, 2019–2026 (2016).
49. Castillo, D., Christiansen, R. H., Dalsgaard, I., Madsen, L. & Middelboe, M. Bacteriophage resistance mechanisms in the fish pathogen *Flavobacterium psychrophilum*: linking genomic mutations to changes in bacterial virulence factors. *Appl. Environ. Microbiol.* **81**, 1157–1167 (2015).
50. Lopatina, A. et al. Metagenomic analysis of bacterial communities of antarctic surface snow. *Front. Microbiol.* **7**, 255 (2016).
51. Sun, C. L. et al. Phage mutations in response to CRISPR diversification in a bacterial population. *Environ. Microbiol.* **15**, 463–470 (2013).
52. Sternberg, S. H., Richter, H., Charpentier, E. & Qimron, U. Adaptation in CRISPR-cas systems. *Mol. Cell* **61**, 797–808 (2016).
53. Modell, J. W., Jiang, W. & Marraffini, L. A. CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* **544**, 101–104 (2017).
54. Molineux, I. J. & Panja, D. Popping the cork: mechanisms of phage genome ejection. *Nature* **11**, 194–204 (2013).
55. McGinn, J. & Marraffini, L. A. CRISPR-cas systems optimize their immune response by specifying the site of spacer integration. *Mol. Cell* **64**, 616–623 (2016).
56. Vale, P. F. et al. Costs of CRISPR-Cas-mediated resistance in *Streptococcus thermophilus*. *Proc. R. Soc. B Biol. Sci.* **282**, 20151270 (2015).
57. van Houte, S. et al. The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532**, 385–388 (2016).
58. Vale, P. F. & Little, T. J. CRISPR-mediated phage resistance and the ghost of coevolution past. *Proc. R. Soc. B Biol. Sci.* **277**, 2097–2103 (2010).
59. Sundberg, L.-R. et al. Intensive aquaculture selects for increased virulence and interference competition in bacteria. *Proc. R. Soc. B Biol. Sci.* **283**, 20153069 (2016).
60. Laanto, E., Bamford, J. K. H., Laakso, J. & Sundberg, L.-R. Phage-driven loss of virulence in a fish pathogenic bacterium. *PLoS ONE* **7**, e53157 (2012).
61. Laanto, E., Bamford, J. K. H., Ravantti, J. J. & Sundberg, L.-R. The use of phage FCL-2 as an alternative to chemotherapy against columnaris disease in aquaculture. *Front. Microbiol.* **6**, 829 (2015).
62. Zhang, J. et al. Association of colony morphotypes with virulence, growth and resistance against protozoan predation in the fish pathogen *Flavobacterium columnare*. *FEMS Microbiol. Ecol.* **89**, 553–562 (2014).
63. Song, Y. L., Fryer, J. L. & Rohovec, J. S. Comparison of six media for the cultivation of *Flexibacter columnaris*. *Fish Pathol.* **23**, 81–94 (1988).
64. Decostere, A., Haesebrouck, F. & Devriese, L. Shieh medium supplemented with tobramycin for selective isolation of *Flavobacterium columnare* (*Flexibacter columnaris*) from diseased fish. *J. Clin. Microbiol.* **35**, 322–324 (1997).
65. Kunttu, H. M. T., Sundberg, L.-R., Pulkkinen, K. & Valtonen, E. T. Environment may be the source of *Flavobacterium columnare* outbreaks at fish farms. *Environ. Microbiol. Rep.* **4**, 398–402 (2012).
66. Ashrafi, R., Pulkkinen, K., Sundberg, L.-R., Pekkala, N. & Ketola, T. A multilocus sequence analysis scheme for characterization of *Flavobacterium columnare* isolates. *BMC Microbiol.* **15**, 243–10 (2015).
67. Kunttu, H. M. T., Suomalainen, L.-R., Jokinen, E. I. & Valtonen, E. T. *Flavobacterium columnare* colony types: connection to adhesion and virulence? *Microb. Pathog.* **46**, 21–27 (2009).
68. Kropinski, A. M. Measurement of the rate of attachment of bacteriophage to cells. *Methods Mol. Biol.* **501**, 151–155 (2009).
69. Santos, M. A. An improved method for the small scale preparation of bacteriophage DNA based on phage precipitation by zinc chloride. *Nucleic Acids Res.* **19**, 5442 (1991).
70. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
71. Besemer, J., Lomsadze, A. & Borodovsky, M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**, 2607–2618 (2001).
72. Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673–679 (2007).
73. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
74. Biswas, A., Staals, R. H. J., Morales, S. E., Fineran, P. C. & Brown, C. M. CRISPRDetect: a flexible algorithm to define CRISPR arrays. *BMC Genomics* **17**, 356 (2016).
75. Zhang, Y. et al. Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. *Mol. Cell* **50**, 488–503 (2013).
76. Leenay, R. T. & Beisel, C. L. Deciphering, communicating, and engineering the CRISPR PAM. *J. Mol. Biol.* **429**, 177–191 (2016).
77. Li, M., Wang, R., Zhao, D. & Xiang, H. Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res.* **42**, 2483–2492 (2014).

Acknowledgements

This work was supported by the Finnish Centre of Excellence Program of the Academy of Finland; the CoE in Biological Interactions 2012–2017 (#252411), by the Academy of Finland (grants #266879 and #304615), and by the Jane and Aatos Erkko Foundation. We would like to thank MSc Jenni Marjakangas, MSc Katja Neuvonen, and Mr. Petri Papponen for skillful help in the laboratory, Dr. Heidi Kunttu for kindly donating genotyped bacterial strains, Dr. Reetta Penttinen for help in bacterial isolation, and Dr. Andrés López-Sepulcre for help in statistical analyses. In memory of Prof. Jaana Bamford.

Author contributions

E.L.: Performed the host range experiments and phage genome analyses. V.H.: Sequenced and characterized the bacterial CRISPR loci. J.R.: Assembled the phage genome data and supervised bioinformatics analyses. L.-R.S.: Performed the time shift and statistical analyses and wrote the first draft of the manuscript. All authors participated in data analyses and in writing the manuscript.

Additional information

Supplementary Information accompanies this paper at doi:10.1038/s41467-017-00158-7.

Competing interests: The authors declare no competing financial interests.

III

COOPERATION BETWEEN CRISPR-CAS TYPES ENABLES ADAPTATION IN AN RNA-TARGETING SYSTEM

by

Ville Hoikkala, Janne Ravantti, César Díez-Villaseñor, Marja Tiirola, Rachel A. Conrad, Mark J. McBride, Lotta-Riina Sundberg 2020

Submitted manuscript

Cooperation between CRISPR-Cas types enables adaptation in an RNA-targeting system

Authors: Ville Hoikkala[1], Janne Ravantti[2], César Díez-Villaseñor[1], Marja Tiirola[1], Rachel A. Conrad[3], Mark J. McBride[3], Lotta-Riina Sundberg[1]*.

Affiliations:

[1] University of Jyväskylä Dept. Biol. and Environmental Science, Nanoscience Center

[2] University of Helsinki

[3] University of Wisconsin-Milwaukee

* Correspondence to: lotta-riina.sundberg@jyu.fi

Abstract: CRISPR-Cas immune systems adapt to new threats by acquiring spacers from invading nucleic acids such as phage genomes. However, some CRISPR-Cas loci lack genes necessary for spacer acquisition, despite apparent variation in spacer content between strains. It has been suggested that such loci may use acquisition machinery from co-occurring CRISPR-Cas systems. Here, using a virulent dsDNA phage, we observe spacer acquisition in the native host *Flavobacterium columnare* that carries an acquisition-deficient subtype VI-B locus and a complete subtype II-C locus. We show that the VI-B locus acquires spacers from both the bacterial and phage genomes, while II-C spacers mainly targets the latter. Both loci target the terminal end of the phage genome, with priming-like patterns around pre-existing II-C protospacers. Through gene deletion we show that the RNA-cleaving VI-B locus acquires spacers *in trans* using acquisition machinery from the DNA-cleaving II-C locus. Our observations reinforce the concept of crosstalk between CRISPR-Cas systems and raise further questions regarding plasticity of adaptation modules.

Main Text

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) arrays consist of multiple sequence repeats separated by unrepeated spacer sequences (1). The repeats and spacers, together with Cas (CRISPR associated) proteins (2), constitute CRISPR-Cas systems that protect bacteria and archaea against infections by bacteriophages (phages) (3). Immunity operates under three main phases: adaptation, expression and interference. In the adaptation (or acquisition) phase, a fragment from the phage genome (protospacer (4)), next to a protospacer adjacent motif (PAM) (5), is inserted into a CRISPR array as a spacer. New spacers are inserted into the array in a polarized fashion, so that one end of the array (called the variable end) accumulates the most recent spacers. The variable end is adjacent to the AT-rich leader region (2-4, 6). In the expression phase, the array is transcribed (usually initiated from promoters in the leader sequence and progressing towards the array) and processed to produce crRNA. These short RNA fragments, consisting of a single spacer and repeat, guide endonucleases to complementary sequences on the invading phage genome in the interference phase (7-9). Interference generally targets DNA, but type VI CRISPR-Cas systems target RNA and type III systems can target both (10). Upon binding to the target and recognizing the PAM sequence (or the shorter protospacer flanking site, PFS, in type VI systems (11)), the target is cleaved, curing the host of infection.

CRISPR-Cas systems are divided into six types and several subtypes on the basis of *cas* gene composition (10). Adaptation is almost universally (10) mediated by a complex consisting of Cas1 and Cas2, which are joined by additional proteins depending on the subtype. For example, types I-E and I-F require Cas3 for adaptation (12), while type II-A recruits the interference protein Csn2 and Cas9 for successful recognition of the spacer acquisition motif (SAM) (13-15). The SAM is used to denote the PAM during acquisition, as it may differ from the

interference-related PAM called the target interference motif (TIM) (13). Some subtype III-B systems can also acquire spacers from RNA using a Cas1 fused with a reverse transcriptase (16, 17), which has also been predicted to occur in some type VI-A systems (18).

Spacer acquisition generally occurs through either naive or primed adaptation. Naive adaptation occurs when no pre-existing spacers target the incoming genome (19), while primed adaptation benefits from partially or fully matching pre-existing spacers (20, 21). Priming has been generally studied in class I systems, but recent evidence shows that priming is likely widespread across type II systems (22, 23) and has experimentally been shown in a subtype II-A system (24). Type II-A priming relies on Cas9-mediated cleavage through a perfect match between the spacer and its target, leading to production of free DNA ends that are used as material for spacer acquisition (24, 25).

Regardless of the composition of the adaptation complex, the insertion of new spacers requires interaction between the complex and the CRISPR array's leader-repeat junction (19). This interaction relies on conserved sequences in leaders and repeats: in the *Streptococcus thermophilus* subtype II-A CRISPR-Cas locus, the conservation of the first ten nucleotides on both sides of the leader-repeat junction are essential for spacer integration (26) as are the first 41 to 43 nucleotides of the leader in the *E. coli* subtype I-E locus (19, 27). Subtype I-D leaders also contain essential motives that are more distant from the leader-repeat junction (28). As repeats and leaders vary in sequence and in length, their respective adaptation proteins co-evolve accordingly (13, 29-32). Therefore, individual CRISPR-Cas loci co-occurring within a genome have distinct versions of Cas1 and Cas2 (and possibly other acquisition related proteins), restraining their capacity to function on other leaders and repeats. Surprisingly, some CRISPR-Cas loci lack adaptation modules but still appear to have variable spacer content in

different bacterial isolates (13, 33-36) and it has been suggested that such loci may utilize acquisition machinery from other intragenomic CRISPR-Cas loci (33, 37-41). Crosstalk between CRISPR-Cas loci has been found in the highly modular CRISPR-Cas systems of Archaea (31) in the expression (42) phase, and interference phase in bacteria (16). Many archaeal type III systems also lack acquisition machinery (31), and the acquisition-deficient subtype III-B locus in *Sulfolobus solfataricus* acquires spacers with PAMs matching those of a co-occurring type I system (35). Further support for *in trans* acquisition comes from the plasmid-encoded subtype IV-A3 systems that lack Cas1/2 (36). These loci have higher-than-expected co-occurrence with host-encoded subtype I-E systems and share similarities with their leader, repeat and spacer-associated PAM sequences (36). *In trans* mediated spacer acquisition between CRISPR-Cas systems therefore seems wide-spread, but experimental evidence to verify this is still lacking.

The recently discovered type VI CRISPR-Cas systems (37, 39, 43) often lack Cas1/2, especially in subtype VI-B and VI-A (10), and have thus been proposed to acquire spacers *in trans* (37, 39-41). However, as type VI loci exclusively target RNA, obtaining spacers from dsDNA (the *modus operandi* for most acquisition complexes) is not optimal, as only half of potential spacers may functionally target the mRNA strand. Relying on acquisition without an associated reverse transcriptase also prevents immunization towards RNA phages. Adaptation in type VI systems has not been experimentally demonstrated and it is unknown if new spacers are acquired through the *in trans* adaptation model, and if so, how this affects their capacity to target RNA.

The genome of the fish pathogen *Flavobacterium columnare* has subtype II-C and VI-B CRISPR-Cas loci (Fig. 1). The II-C locus contains the adaptation proteins Cas1 and Cas2, and

100 the endonuclease Cas9 that cleaves dsDNA (although at least Cas9 of *Neisseria meningitidis*
101 and *Campylobacter jejuni* can also cleave ssRNA (44, 45)). The II-C locus also contains genes
102 encoding a predicted ParDE type II toxin-antitoxin system between *cas9* and the repeat-spacer
103 array, which may contribute to the CRISPR-Cas immune response (46, 47). Previously studied
104 II-C systems differ from other CRISPR-Cas systems in that array expression is not initiated
105 from the leader, but from repeat-encoded promoters (44, 48). The subtype VI-B of *F.*
106 *columnare* only encodes for Cas13b and does not contain adaptation genes nor Csx27 or Csx28
107 that are often associated with type VI-B systems (10). Cas13 exclusively cleaves RNA and is
108 guided to its target by its bound crRNA, similar to other Cas proteins. However, once activated
109 by this primary target, Cas13 becomes a promiscuous RNase that also cleaves non-
110 complementary phage and host transcripts, (49), potentially leading to cellular dormancy or
111 death (43). Previously, we showed that natural isolates of *F. columnare* vary in their phage-
112 targeting spacer content in both loci, driving the evolution of sympatric phages (34). Only the
113 phage mRNA strand was targeted in VI-B viral protospacers (n=15), while this was true for
114 only half of the II-C protospacers (34). While accumulation of mRNA-targeting spacers in the
115 VI-B locus seems clear, it is unknown if this bias stems from a selective acquisition process or
116 from positive selection due to successful interference. To further characterize spacer
117 acquisition in these loci, we co-cultured *F. columnare* with a virulent dsDNA phage in
118 controlled laboratory conditions (Fig. 2). Our results reveal details of subtype VI-B and II-C
119 adaptation in a native host and show the first experimental evidence of *in trans* adaptation
120 between CRISPR-Cas types, despite striking differences in leader and repeat sequences.

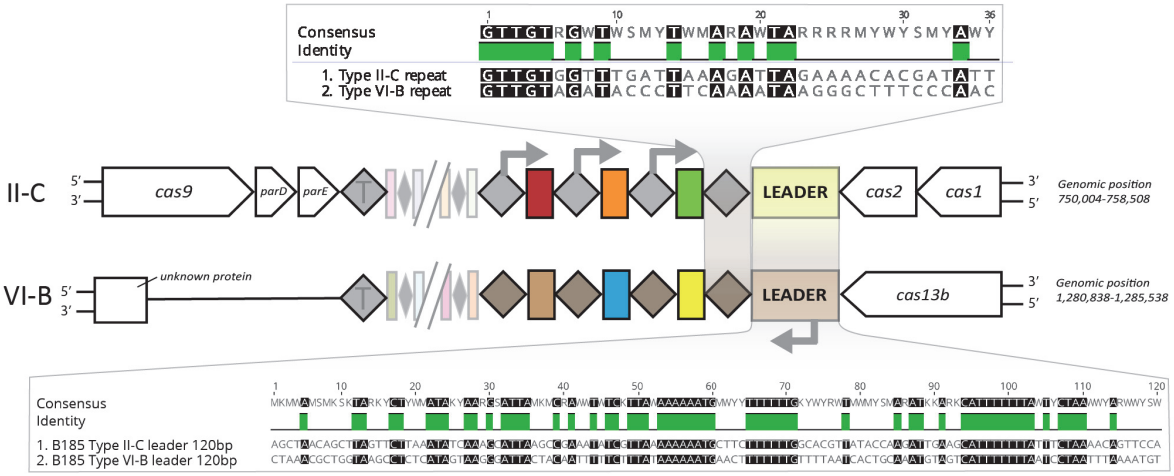

Figure 1 | Comparison of the arrangements of II-C and VI-B CRISPR-Cas loci in *F. columnare* B185. Colored boxes represent spacers and diamonds represent repeats (T denotes the terminal repeat). Repeats and leaders are aligned with no gaps and presented in 5' to 3' direction (base 36 in repeats is leader-proximal). The grey arrows show the directions of transcription that are inferred from previous studies, the presence of putative promoters and our transcriptome data. The II-C array also contains the toxin/antitoxin genes *parD* and *parE*. The VI-B locus has been flipped from its relative genomic orientation for convenient comparison.

Phage-induced acquisition in both loci

To study the adaptation phase of both subtype II-C and VI-B loci in a controlled environment, we cultured *F. columnare* strain B185 (50, 51) (genome acc.no. CP010992) with its virulent dsDNA phage FCL2 (NC 027125) (51) in liquid medium (B+P treatment, Fig. 2). We also included treatments with UV-irradiated (52) FCL-2 (B+UVP) and a mixture of irradiated and non-irradiated phage (B+P+UVP). The II-C CRISPR array of this bacterium contain four pre-existing FCL-2 targeting spacers and four VI-B spacers with partial FCL-2 matches (Supplementary Table S3A-B). After a week of co-culturing, we plated the liquid cultures to screen for expanded CRISPR arrays in individual colonies. Only bacteria that were exposed to

the phage mixture (B+P+UVP) had acquired new spacers, although most of these colonies still had unexpanded arrays. We did not observe spacer acquisition in bacteria that were not exposed to phage (B), or were exposed to either phage type alone (B+P or B+UVP) (Table S1 and S2). Inclusion of irradiated phage enhances spacer acquisition in *S. thermophilus*, presumably due to their reduced replication capability (52), which may also explain the results observed here.

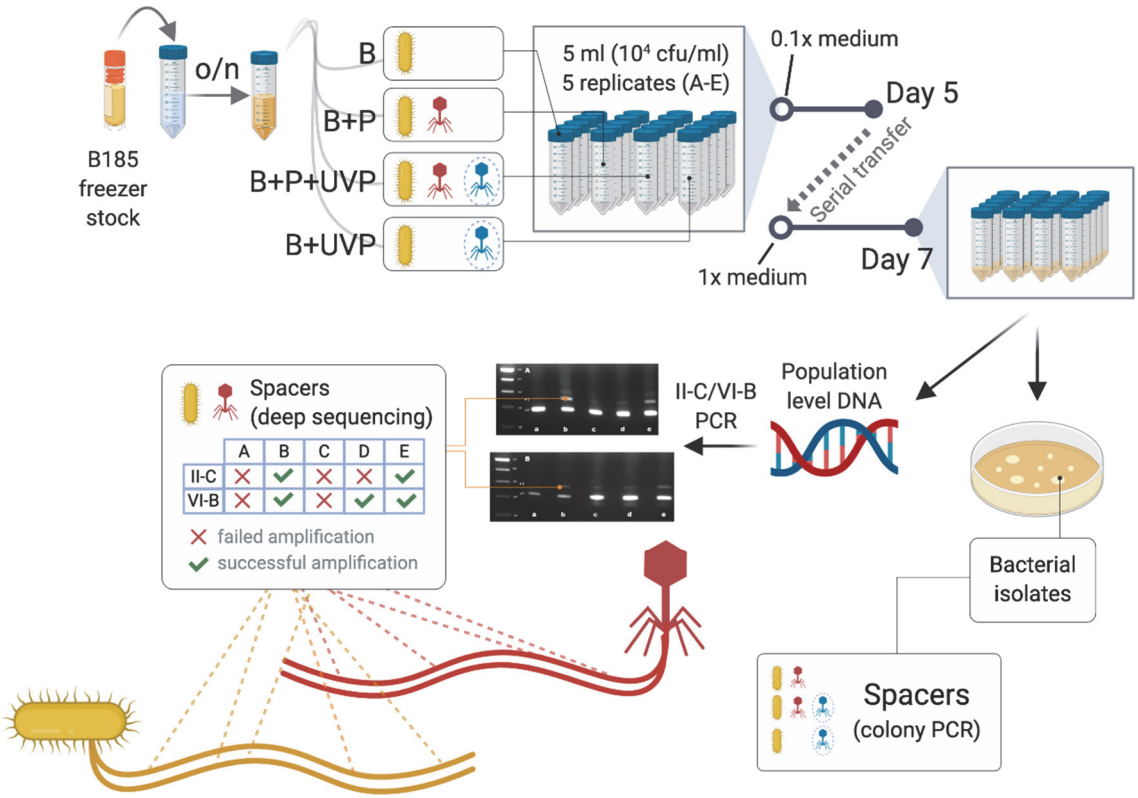

Figure 2 | Overview of the experiment. Spacer acquisition in *F. columnare* strain B185 was investigated using a two-step growth method in diluted and undiluted medium. Spacers were examined in individual colonies in all three phage-based treatments as well as by deep sequencing the CRISPR arrays in the non-irradiated phage treatment (B+P).

Next, we focused on CRISPR adaptation on the bacterial population scale by amplifying the variable ends of both arrays directly from the liquid cultures. We observed spacer acquisition in the phage mixture treatment (B+P+UVP) (data not shown) and in the non-irradiated phage

(B+P) treatments (Supplementary Figure S1). All downstream analyses were done on the B+P samples, as this treatment is a closer approximation of a real-life scenario. Within each replicate of B+P treatment (identified by letters a to e), the relative efficiency of spacer acquisition was similar between the two arrays (Fig. S1). To identify the diversity and origin of the new CRISPR spacers in the bacterial population, we deep sequenced the population-scale variable ends of the arrays. We obtained data on new II-C spacers in replicates b and e, and VI-B spacers in b, d and e (other replicates did not allow for deep sequencing due to low DNA yields). We also deep sequenced the wild-type amplicons in the control cultures to ensure the stability of the arrays (no variance in these sequences were observed).

Spacer lengths were between 28 and 32 bp, with the majority at 30 bp in both loci (Fig. 3C). We divided the spacers into unique and absolute sets: in the unique spacer dataset, each spacer sequence was counted only once, whereas the absolute set allowed for repeated observations of the same spacer. The absolute set reflects biases towards specific spacers, while the unique set captures the overall distribution of different spacer sequences. Most unique II-C spacers targeted the phage genome, with a minority targeting the bacterial genome (Fig. 3A). VI-B spacers targeted both genomes relatively evenly, although the replicates had some differences (Fig 3A). All differences were accentuated when spacer count was viewed as absolute, but the overall trend remained the same (Fig 3B). The different targeting preferences between the loci are likely explained by negative selection of autoimmunity resulting from Cas9-mediated self-targeting (e.g. (53)). The abundance of self-targeting VI-B spacers is surprising, given that such spacers should induce dormancy even in the absence of phage and thus be selected against (54). The observed pattern therefore indicates that, at least during this seven-day experiment, self-targeting by VI-B is not particularly harmful.

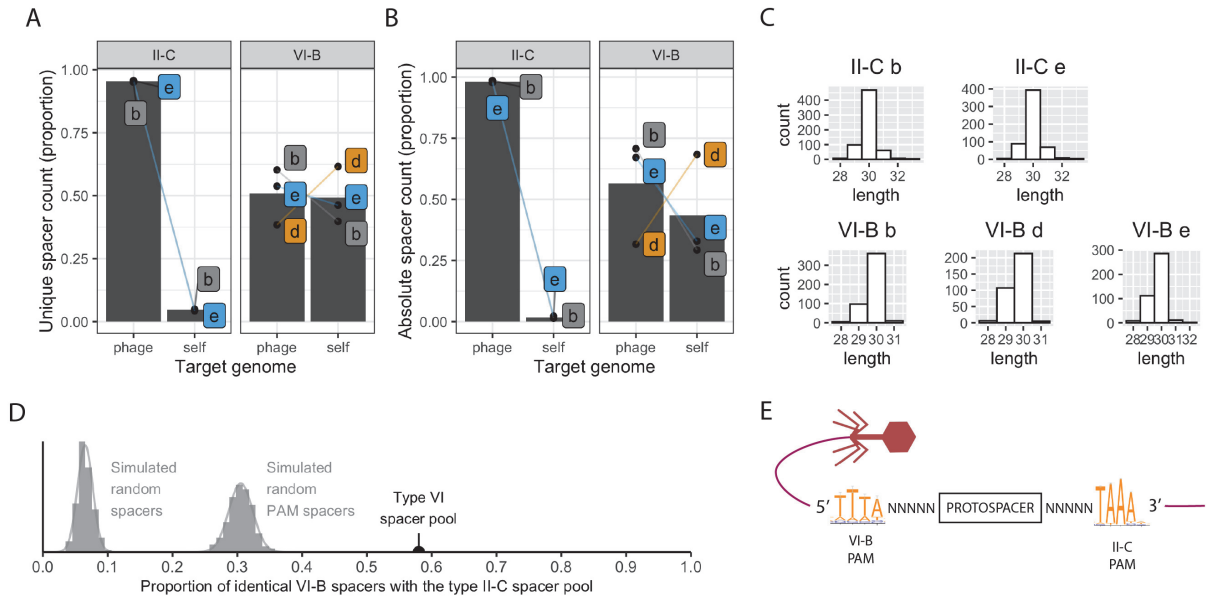

Figure 3 | Analysis of acquired spacers in the co-culture experiment. A) Proportions of unique subtype II-C or VI-B spacers targeting the phage or bacterial (self) genome. Dots show the exact counts of spacers in a replicate (b, e or d) and bars show their respective mean. B) Similar to A but shows proportions using absolute spacer counts C) Size distribution of unique spacers in different replicates. D) The proportion of simulated and observed VI-B spacers that are identical with the II-C spacer pool. E) Phage-targeting II-C and VI-B PAM sequences (this example from replicate e; for all replicates see Supplementary Figure S2A).

To calculate the proportion of identical new spacers between the loci, we pooled unique phage-targeting spacers from both loci in all replicates. Remarkably, 58% of VI-B spacers were also found in the larger II-C spacer pool, suggesting that the loci share predisposition towards a limited set of spacers. For statistical comparison, we simulated spacer acquisition across the phage genome by sampling random positions either freely or next to predicted PAM sites. Both simulations were run 1000 times, each run sampling 430 spacer positions (the size of the VI-B spacer pool). We calculated the proportion of shared spacers between each run and the observed II-C spacers, and found that 6.3% (SD 1.2%) of the random spacers and 29.6% (SD 1.85%) of

196 the PAM-adjacent spacers had a match in the subtype II-C spacer set (Fig. 3D). The similarity
197 between the observed VI-B and II-C spacers is significantly higher than for either simulated
198 pool ($P < 10^{-10}$, one-tailed test), which suggests that new spacers are sampled from a limited
199 protospacer pool that is shared by both loci. It is also possible that acquisition itself is uniform
200 and random across the genome, but these specific spacers become selected through
201 interference. This analysis was not performed on the absolute spacer dataset, as unique spacers
202 better capture the overall variance of spacers sequences.
203
204 **Shared acquisition motifs suggest *in trans* adaptation**
205 To further characterize the protospacers, we investigated PAM sequences for both loci. We
206 used the guide-oriented approach that determines PAMs on the target genome around the
207 sequence that matches a particular spacer's crRNA (55). It is necessary to establish the
208 direction of crRNA transcription before evaluating the PAM sequence using the guide-oriented
209 approach, as the transcription direction determines the resulting crRNA strandedness. Previous
210 studies have shown that unlike in other CRISPR-Cas systems, subtype II-C arrays are
211 expressed from within the array towards the variable end using repeat-encoded promoters (5'-
212 TAAAT-3') (44, 48). The *F. columnare* II-C repeats lack this sequence but include a TTTG
213 motif 33 bp upstream of each spacer. As the sequence TTG is an established -33 promoter in
214 *Flavobacterium hibernum* (56), this putative promoter may drive II-C array transcription in *F.*
215 *columnare.* While this is only speculative, our transcriptome data also supports within-array
216 transcription that proceeds towards the leader (albeit with only eight reads mapping on the
217 array). The subtype VI-B array, on the other hand, shows an abundance of transcripts occurring
218 in the leader-to-array direction as previously observed in these systems (39, 43) and no putative
219 repeat-encoded promoters.
220

In the subtype II-C locus, the previously reported PAM (34, 57) NNNNNTAAA was observed downstream of most (~63%) phage-targeting spacers (Fig. 3E and S2A). In the few self-targeting subtype II-C spacers, the canonical PAM was almost always either absent or had extended or shortened N-regions, suggesting that both the sequence TAAA and the length of the N-region in the PAM play a role subtype II-C interference in *F. columnare* (Fig. S2B). The subtype VI-B had the same PAM sequence in reverse complement and upstream of the protospacers (TTTANNNNN) (Fig. 3E), with slightly lower conservation compared to the subtype II-C protospacers (~50%) (Fig. S2B).

These results show that when using the guide-oriented approach, the PAMs of the II-C and VI-B loci are complementary to each other and on opposite sides of their spacer targets. In both loci, the PAM-adjacent ends of the protospacers are oriented towards the leader. Therefore, while the target interference motif (TIM) of the loci is seemingly different, the spacer acquisition motif (SAM) is shared by the loci. We further note that the term TIM is not strictly applicable to the RNA-targeting type VI systems, and that their interference has only been shown to be affected by protospacer flanking sites (PFSs) (39, 43). No PFSs were discovered around the VI-B targets in *F. columnare*. Overall, these results strongly suggest that the arrays share spacer acquisition machinery, as SAMs are generally hallmarks of specific acquisition complexes (12, 58).

Spacer target distributions are non-uniform and show possible II-C priming

We investigated which regions of the target genomes are targeted by II-C and VI-B spacers. To do this, we examined both unique and absolute spacer counts separately. Despite the relatively even distribution of predicted PAM sequences across the phage and bacterial genomes, spacer targets were not uniformly spread. On the phage genome, spacer targets were

246 concentrated on one end of the genome (Fig. 4). Similar patterns were previously observed in
247 natural isolates of *F. columnare* (34) and in the subtype II-A CRISPR-Cas locus of
248 *Staphylococcus aureus* (25). Furthermore, the two strands of the dsDNA phage genome were
249 targeted unevenly, with stronger differences at the end of the genome (downstream of ~30 kbp),
250 especially in the unique spacer set. Intriguingly, the non-unique II-C spacers congregated
251 strongly around a pre-existing (fully-matching) II-C spacer on the phage genome (Fig. 4A) – a
252 trend not visible in the unique spacer set (Fig. 4B). The distribution of spacers around this
253 hotspot follow an asymmetric pattern that is characteristic to primed adaptation (59, 60). As
254 the targeting pattern in the current study flanks a perfect spacer match on the phage genome, it
255 is possible that priming plays a role also in subtype II-C adaptation. However, it is unclear why
256 the three additional phage-targeting II-C spacers with intact PAM sequences do not give rise
257 to similar patterns (Fig. 4A). One possibility is that targeting the end of the phage genome that
258 first enters the cell is more efficient, as shown in a *Staphylococcus pyogenes* subtype II-A
259 system (25). Whether this end of the phage genome first enters the *F. columnare* cell has not
260 been experimentally investigated.

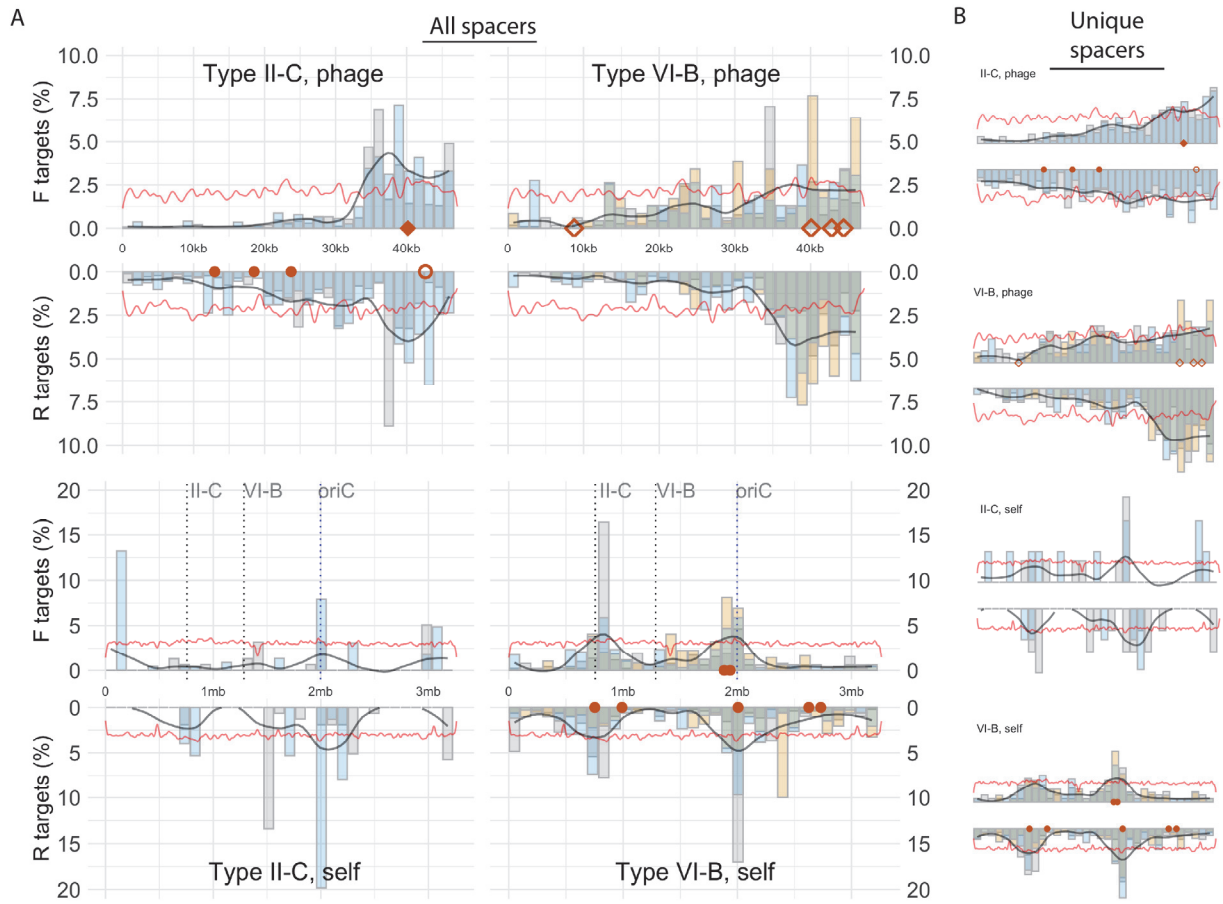

Figure 4 | Distribution of spacers on the phage and bacterial genomes. The genomes are divided into bins (spanning 3% of the genomes) and the proportion of spacer targets in each bin per replicate is shown by overlapping bars. The black line indicates smoothed average of targeting across replicates. The red line shows the relative frequency of the putative PAM sequence. Red markers are spacers that pre-exist in strain B185: diamonds target mRNA and circles do not. Filled markers are spacers with perfect matches and PAMs, while non-filled markers are spacers with 1-3 bp mismatches with their target sequence. A) All spacers mapped on both genomes. B) Unique spacers mapped on both genomes.

On the bacterial genome, spacers from both loci congregated on two hotspots (Fig. 4). The first was centered on the subtype II-C CRISPR-Cas locus, while the other was in the predicted origin

of replication (oriC). A previous study reported an acquisition hotspot on the *E. coli* oriC and consequently discovered that acquisition is increased in regions where replication of the host chromosome is stalled due to action of RecBD (61). Stalled replication forks are overrepresented at oriC, since this region is the first to be replicated and thus becomes enriched in a replicating cell. Similar to us, Levy et al. and Staals et al. (62) also found an acquisition peak on a CRISPR array. This was speculated to arise from frequent nicking of DNA during spacer acquisition, as nicking of the chromosome would also stall the replication fork leading to increased spacer acquisition (61). However, given that this is the cause for concentration of spacers on the II-C locus, why spacers do not congregate on the *F. columnare* VI-B locus is unclear, as this locus also acquires spacers.

The spacer targeting patterns observed here suggest that priming may play a role in II-C adaptation, while similar patterns were not visible for the VI-B spacers. This is surprising, since the production of free dsDNA by Cas9 cleavage is expected to provide spacer material with no intrinsic bias towards any specific array. Therefore, it is likely that the dissimilarity between the non-unique II-C and VI-B spacer target patterns arises through positive selection of primed spacers in II-C but not in VI-B. However, this evidence is indirect and requires further investigation.

Most subtype VI-B spacers are not mRNA-targeting

We investigated if the new spacers reflect the different target-requirements of the loci (DNA vs RNA). We examined the capability of each spacer crRNA to bind mRNA by assessing their complementarity to predicted ORFs from both target genomes. Spacers from all replicates within a locus were also pooled together for an overall estimate. For statistical context, we calculated the probabilities of mRNA-targeting proportions in both loci with a binomial

distribution that uses sample sizes matching those of the pooled spacers and assumes an equal probability of targeting each strand (the binomial tests were not included in the absolute spacer counts due to pseudoreplication that results from multiple instances of the same spacers).

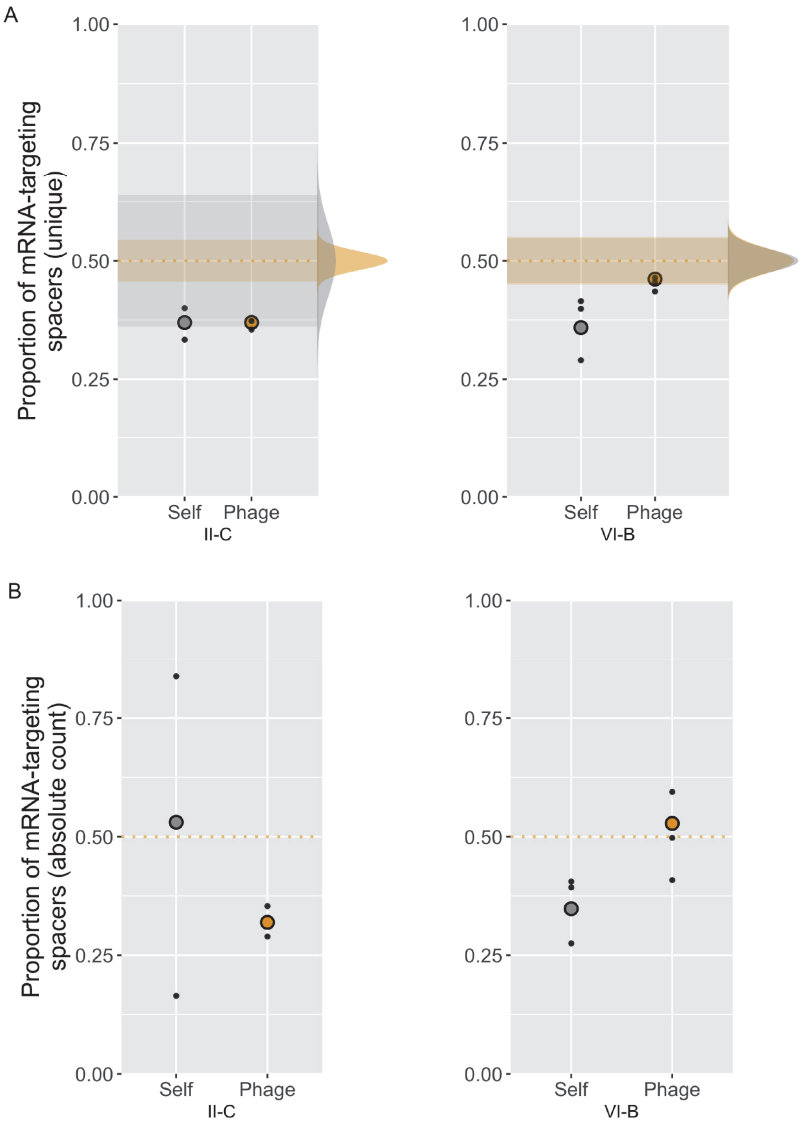

Figure 5 | The proportions of ORF targeting spacers that are complementary to mRNA. Dots are replicate cultures and circles are pooled replicates. A) Unique spacers. Distributions on the right show the spread of expected values adjusted for the size of each spacer pool (yellow = self, grey = phage). B) Absolute spacer counts. The binomial distributions were omitted from absolute counts.

When considering unique spacers, the proportion of phage mRNA-targeting VI-B spacers was slightly less than half (0.46). Chance of randomly obtaining a proportion at least this low was 7.4% (P=0.074, one-tailed test), suggesting a weak or non-existent bias towards non-mRNA targeting spacers. mRNA-targeting on the bacterial genome, however, was more rare and significantly below expected at 0.359 ($P < 10^{-9}$, one-tailed test) (Fig. 5A). For total spacer counts, the proportion of pooled phage mRNA-targeting reaches 0.52, suggesting that such spacers are slightly more represented than in the unique spacer pool (Fig. 5B). Self-mRNA-targeting VI-B spacers remained close to the unique counts at 0.348 (Fig. 5B).

Subtype II-C proportions in the unique spacer set were the same for both target genomes (0.37 and 0.37). Given their vastly different sample sizes, the proportion of mRNA-targeting spacers on the phage genome was significantly below expected ($P<10^{-10}$, one-tailed test), while on the self-genome the proportion was only weakly biased (P=0.052, one-tailed test) (Fig. 5A). In the total spacer counts, the self-targeting II-C proportions and divergence may seem radical (Fig. 5B) but stem from the very small sample sizes of self-targeting II-C spacers, where slight overrepresentation of a spacer may disproportionally skew the result by chance. Absolute phage-targeting II-C transcripts were slightly lower than in unique counts, at 0.319 (Fig. 5B).

The minimal phage mRNA-targeting in subtype II-C spacers is unexpected. One potential explanation would be a strongly uneven distribution of PAM sequences on the two strands, but this is not the case in either genome (B185 has 102,630 PAMs with 50.4%/49.6% strand distribution and FCL-2 has 1672 PAMs with 50.7%/49.3% distribution). Another possibility is the inability of RNA-bound Cas9 to degrade DNA, as competition between RNA and DNA

targets for Cas9 association would impact subtype II-C immunity negatively. However, this hypothesis would not hold if *F. columnare* Cas9 is also able to degrade RNA similar to II-C Cas9 of *Neisseria meningitidis* or *Campylobacter jejuni* (44, 45).

The lack of VI-B self-mRNA-targeting spacers is expected due to their possible dormancy-inducing effects (54) that likely transpire even in the absence of phage. This is also in agreement with the lack of pre-existing self-mRNA-targeting VI-B in B185 (n = 7, Fig. 3D, Supplementary Table S3B). However, this counter-selection is much weaker than for self-targeting subtype II-C spacers (Fig. 3A-B), which argues for a lesser effect of subtype VI-B autoimmunity in this experimental setting. Saturation of phage-mRNA-targeting VI-B spacers is not shown in the current study likely due to the short time span of the experiment, but may also be affected by the pre-existing, fully matching II-C spacers (Supplementary Table S3A). If the pre-existing spacers already provide CRISPR-Cas based immunity, the need to acquire new ones is decreased and their selective advantage is also weaker. Why these pre-existing II-C spacers do not provide full immunity ("CRISPR-immunized" bacteria are still infected by the phages (63)) is unknown, but could be related decreased benefit of CRISPR-Cas in these high nutrient environments (64) or anti-CRISPR proteins encoded in phage genomes (65).

Overall, the reasons for unexpectedly low levels of mRNA-targeting in both loci remain speculative. In this analysis we relied on predicted ORFs and do not take into account possible antisense transcription or intergenic transcripts – our estimate of RNA-targeting may therefore slightly underestimate the actual RNA-targeting proportions. However, comparing these results to our previous study on natural *F. columnare* isolates suggests that the complete saturation of phage mRNA-targeting VI-B spacers in natural isolates (n=19) (including the

pre-existing four partially spacers in the wild-type array of B185, Supplementary Table S3B) are likely the result of positive selection towards mRNA-targeting spacers during interference. Selection acting on VI-B targeting is also supported by the fact that none of the pre-existing VI-B spacers in strain B185 are complete matches with their protospacers on the phage genome, and by the previous observation that VI-B protospacers accumulate mutations in VI-B protospacers over time (63). Future studies are needed to assess how RNA-based VI-B immunity evolves over time and what constitutes functional targets for VI-B interference in native systems.

The VI-B array uses II-C Cas1 *in trans*

Identical acquisition efficiencies, similar protospacer localization patterns and shared PAM sequences supported the hypothesis that the *F. columnare* subtype II-C and VI-B CRISPR-Cas loci share acquisition machinery. To verify this, we deleted *cas1* from the subtype II-C locus from another *F. columnare* strain where genetic manipulation is currently possible (B245). Using the resulting B245-Δ*cas1* strain, we performed a prolonged adaptation experiment with ten phage-exposed (phage V156) and two bacteria-only replicates. As a control, we used a "reversion wild-type" strain (rev-wt) that resulted from an alternative outcome of the mutation process and thus contains an intact *cas1*. We did not observe spacer acquisition in either array in any of the 12 Δ*cas1* cultures (Fig. S5), all of which survived until the end of the experiment. Of all 12 rev-wt cultures, only two phage-exposed replicates survived, both showing spacer acquisition in their II-C and VI-B arrays (Fig. S5). These results demonstrate the dependence of the VI-B locus on the II-C acquisition machinery.

The inability of B245 rev-wt to survive in the presence of phage is in contrast with the previous experiment performed with strain B185 and phage FCL2. This difference may simply stem

from the experimental setups, as B245 and B185 were cultivated in different media and for three weeks or one week, respectively (see material and methods). Also, B245 has four pre-existing, fully matching phage mRNA-targeting VI-B spacers in the array's variable end (Supplementary Table S3C-D), whereas B185 only has incomplete VI-B spacer matches with its target, all located in the array's non-variable end. It is therefore possible that the B245 phage-targeting VI-B spacers induce dormancy during the CRISPR-Cas response (54), whereas B185's lack of fully matching pre-existing spacers does not. If this is the case, the survival of B245 Δ*cas1* is surprising as the inability to acquire new spacers should not abrogate interference using pre-existing ones. Further experiments to explore the effect of spacer diversity and role of different *cas* genes in these systems are required.

Comparison of II-C/VI-B leaders and repeats across species

In *F. columnare,* the subtype VI-B and II-C have similarly sized repeats (36 bp), with sequence similarity mostly in the leader-distal ends (5'-GTTGT-3', Fig. 1). The lack of similarity in the leader-adjacent end of the repeat is surprising, given the importance of conservation in this region during spacer insertion at least in the *S. thermophilus* subtype II-A system (26). The *F columnare* leaders also lack similarity in the leader-repeat junctions but share similarly positioned poly-A and poly-T regions up to 100 bp from the junction (Fig. 1). To see if II-C and VI-B CRISPR-Cas loci display similar patterns in other species, we extracted the repeats and leaders of nine species that carry these loci (lacking *cas1* and *cas2* in their VI-B locus). II-C and VI-B repeats shared the same leader-distal motifs across the species, but *F. columnare* was the only species with similarly sized II-C and VI-B repeats (II-C repeat in the other species were on average 11 bp longer) (Fig. S6). Leaders, clustered by Cas1-similarity for meaningful alignment, had similar leader-distal poly-A and poly-T motifs as *F. columnare* (Fig. S7), which may act as binding or signaling sites for spacer integration. As the generation of new repeats is

based on precise ruler-mechanisms in the acquisition complex (15, 27, 66), it remains unknown how the other species with differing repeat lengths generated their subtype VI-B arrays. It is therefore possible that all or some of the subtype VI-B CRISPR-Cas systems in these species are inactive and that *F. columnare* is one of the few or the only species maintaining crosstalk between VI-B and II-C systems. Indeed, subtype II-C adaptation was recently shown in one of these species, *Riemerella anatipestifer*, but its co-occurring VI-B locus (denoted as an orphan array) did not acquire new spacers (67).

F. columnare **CRISPR-Cas model**

Based on our observations and previous biochemical studies, we propose a model for *F. columnare* spacer acquisition and interference. During acquisition, new spacers are bound to the Cas1/2 acquisition complex (possibly aided by Cas9) in a specific conformation relative to the PAM (68, 69). This pre-spacer is inserted into the array (15, 69-71), with the PAM end of the spacer facing the leader (5, 23, 72). Leader recognition is probably mediated by shared motifs in the subtype II-C and VI-B leaders and repeats that are distal from the leader-repeat junctions. During interference, spacers oriented this way produce a downstream interference-PAM (TIM) in subtype II-C, as is generally observed in type II systems (73). Since the arrays are transcribed in opposing directions in respect to the variable end of the array (Fig. S4), the TIM location is switched for subtype VI-B spacers during interference. Acquisition events relying on an identical SAM can therefore lead to two complementary interference patterns, dictated by the array in which the spacer is inserted (Fig. 6).

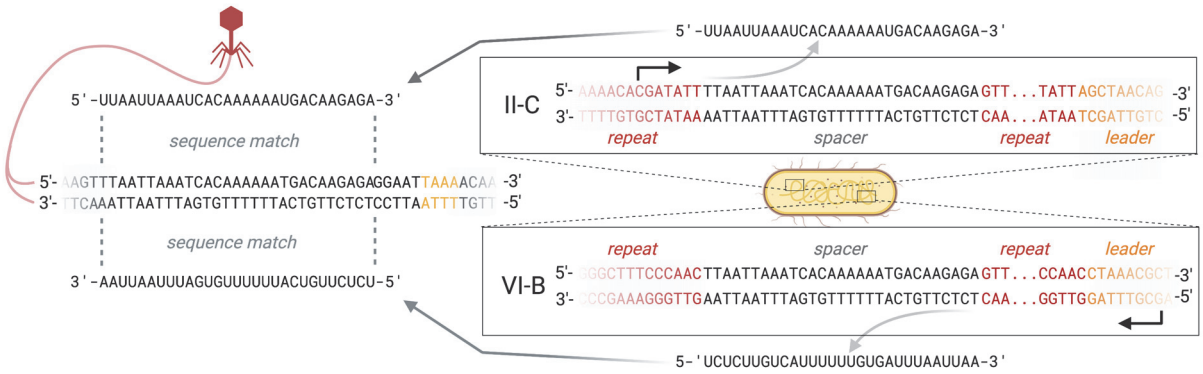

Figure 6 | A model for subtype II-C and VI-B in *F. columnare*. In this scenario, the first spacer of the II-C array has also been inserted in the VI-B array. However, due to opposite transcription directions, the crRNA becomes mirrored between the loci. Mapping the crRNA on the phage genome (using the guide-oriented approach) reveals a downstream TIM for II-C and an upstream "TIM" for VI-B (golden letters on the phage genome), although the SAM is the same for both. All sequences shown here correspond to actual sequences (except the VI-B spacer, which is the same as the II-C spacer for illustration purposes).

Conclusions

We demonstrate that a native subtype VI-B CRISPR-Cas locus acquires spacers from a virulent dsDNA phage and from the host genome. Despite differences in repeat and leader sequences, shared characteristics of the newly acquired spacers with the co-adapting subtype II-C locus supported the model of *in trans* spacer acquisition in the subtype VI-B locus, which was confirmed by analysis of a *cas1* deletion mutant. This arrangement provides the RNA-targeting VI-B array with an abundance of spacers that do not target mRNA. Other species with similar CRISPR-Cas loci had more divergent repeat sequences, suggesting that *in trans* acquisition may not occur in these species or may utilize a highly plastic acquisition complex.

448 This study also highlights *F. columnare* as one of the few known species (74) that can be
449 induced to acquire spacers from virulent phages in laboratory conditions. While this study
450 focused on adaptation in *F. columnare*, some remarks can also be made on its CRISPR-Cas
451 functionality in interference. It is clear that pre-existing spacers do not grant *F. columnare* total
452 phage immunity (63) as they do in e.g. *Streptococcus thermophilus* (3). One reason for this
453 may be an experimental setup that favours constitutive defences (morphology change) over
454 CRISPR-Cas (64), exemplified by the current study in which the cells were required to be
455 initially grown in diluted medium to observe spacer acquisition (pre-experiments revealed that
456 growing *F. columnare* in undiluted medium lead to morphology change with no spacer
457 acquisition). While CRISPR-Cas interference in *F. columnare* may often be masked by non-
458 optimal conditions, several observations suggest that interference is active in both loci. First,
459 self-targeting II-C spacers were virtually absent in this study, suggesting lethality in self-
460 targeting by Cas9. Second, both mRNA and DNA-targeting VI-B spacers are acquired in the
461 laboratory, but only mRNA-targeting spacers are observed in natural samples, suggesting
462 positive selection of such spacers and implying that VI-B is active in interference. Also, none
463 of the pre-existing VI-B spacers of B185 (Supplementary Table S3B) are perfect matches with
464 their targets, suggesting that phages have previously escaped VI-B targeting as also suggested
465 by our previous study (63). Third, *F. columnare* fulfils several criteria that can be used to assess
466 the importance of a CRISPR-Cas system (75): both loci coexist in all *F. columnare* strains
467 sequenced so far (76), with high diversity in natural spacers that target sympatric phages (63),
468 and with high diversity of spacers across *F. columnare* isolates (our unpublished observations).
469
470 We expect future studies to reveal the extent of crosstalk in type VI systems and the
471 biochemical basis for the plasticity of the acquisition machinery when interacting with different

leader and repeat sequences. The observation of possible primed adaptation in the subtype II-C locus provokes further research in this area.

Acknowledgments: We thank Sylvain Moineau for comments on the manuscript, Petri Papponen and Elina Laanto for conducting the RNA sequencing, and Elina Virtanen for Ion Torrent preparation.

Funding: This work received funding from Kone Foundation, Jane and Aatos Erkko Foundation and the Academy of Finland (grant #314939). This work resulted from the BONUS Flavophage project supported by BONUS (Art 185), funded jointly by the European Union (EU) and the Academy of Finland, and was financially supported, in part, by United States Department of Agriculture-ARS CRIS project 5090-31320-004-00D, cooperative agreement #5090-31320-004-03S to MJM and RAC. The views contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Government. Mention of trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply its approval to the exclusion of other products that may be suitable. This manuscript is submitted for publication with the understanding that the United States Government is authorized to reproduce and distribute reprints for governmental purposes. The USDA is an equal opportunity employer.

Author contributions: Study and experimental design by VH, JR and L-RS. Cas1-mutant planned and executed by RAC and MJM and finalized by VH and CDV. Ion Torrent sequencing planned and assisted by MT. Other laboratory work by VH. Data and statistical analysis by VH with assistance from CDV, JR and LR-S. VH wrote the manuscript with edits and comments from all authors.

Competing interests: Authors declare no competing interests.

Data and materials availability: The genomes of *F. columnare* strain B185 (NZ_CP010992.1) and phage FCL2 (NC_027125.1) are available in Genbank. Raw Ion Torrent reads from the spacer acquisition experiment are available at https://bit.ly/2tVWGhW (for review only. C1 refers to II-C and C2 to VI-B locus). The B245 genome is currently available at https://bit.ly/37X6Uyx and its assembly is described in study IV of this doctoral thesis. Custom code and instructions for analyzing the raw data and reproducing all figures in the manuscript are in GitHub: https://github.com/vihoikka/spacerAQ_vh

Supplementary Materials:

Materials and Methods

Figures S1-S8

Tables S1-S3

References

1. **Mojica FJ**, **Díez-Villaseñor C**, **Soria E**, **Juez G**. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol Microbiol **36**:244–246.

2. **Jansen R**, **Embden JDAV**, **Gaastra W**, **Schouls LM**. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol **43**:1565–1575.

3. **Barrangou R**, **Fremaux C**, **Deveau H**, **Richards M**, **Boyaval P**, **Moineau S**, **Romero DA**, **Horvath P**. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315**:1709–1712.

4. **Horvath P**, **Romero DA**, **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.

5. **Mojica FJM**, **Díez-Villaseñor C**, **García-Martínez J**, **Almendros C**. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology **155**:733–740.

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

7. **Hale CR**, **Zhao P**, **Olson S**, **Duff MO**, **Graveley BR**, **Wells L**, **Terns RM**, **Terns MP**. 2009. RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. Cell **139**:945–956.

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

9. **Gasiunas G**, **Barrangou R**, **Horvath P**, **Siksnys V**. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proceedings of the National Academy of Sciences **109**:E2579–86.

10. **Makarova KS**, **Wolf YI**, **Iranzo J**, **Shmakov SA**, **Alkhnbashi OS**, **Brouns SJJ**, **Charpentier E**, **Cheng D**, **Haft DH**, **Horvath P**, **Moineau S**, **Mojica FJM**, **Scott D**, **Shah SA**, **Siksnys V**, **Terns MP**, **Venclovas AXE**, **White MF**, **Yakunin AF**, **Yan W**, **Zhang F**, **Garrett RA**, **Backofen R**, **Oost J**, **Barrangou R**, **Koonin EV**. 2019. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nature Publishing Group **18**:1–17.

11. **Abudayyeh OO**, **Gootenberg JS**, **Konermann S**, **Joung J**, **Slaymaker IM**, **Cox DBT**, **Shmakov S**, **Makarova KS**, **Semenova E**, **Minakhin L**, **Severinov K**, **Regev A**, **Lander ES**, **Koonin EV**, **Zhang F**. 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science **353**:aaf5573–11.

12. **Kieper SN**, **Almendros C**, **Behler J**, **McKenzie RE**, **Nobrega FL**, **Haagsma AC**, **Vink JNA**, **Hess WR**, **Brouns SJJ**. 2018. Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. Cell Rep **22**:3377–3384.

13. **Shah SA**, **Erdmann S**, **Mojica FJM**, **Garrett RA**. 2013. Protospacer recognition motifs: Mixed identities and functional diversity. RNA Biology **10**:891–899.

14. **Heler R**, **Samai P**, **Modell JW**, **Weiner C**, **Goldberg GW**, **Bikard D**, **Marraffini LA**. 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature **519**:199–202.

15. **Kim JG**, **Garrett S**, **Wei Y**, **Graveley BR**, **Terns MP**. 2019. CRISPR DNA elements controlling site-specific spacer integration and proper repeat length by a Type II CRISPR-Cas system. Nucleic Acids Research.

16. **Silas S**, **Lucas-Elio P**, **Jackson SA**, **Aroca-Crevillén A**, **Hansen LL**, **Fineran PC**, **Fire AZ**, **Sanchez-Amat A**. 2017. Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. Elife **6**:aaf5573.

17. **Silas S**, **Mohr G**, **Sidote DJ**, **Markham LM**, **Sanchez-Amat A**, **Bhaya D**, **Lambowitz AM**, **Fire AZ**. 2016. Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. Science **351**:aad4234–aad4234.

18. **Toro N**, **Mestre MR**, **Martínez-Abarca F**, **González-Delgado A**. 2019. Recruitment of Reverse Transcriptase-Cas1 Fusion Proteins by Type VI-A CRISPR-Cas Systems. Front Microbiol **10**:2160.

19. **Yosef I**, **Goren MG**, **Qimron U**. 2012. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Research **40**:5569–5576.

20. **Swarts DC**, **Mosterd C**, **van Passel MWJ**, **Brouns SJJ**. 2012. CRISPR Interference Directs Strand Specific Spacer Acquisition. PLoS ONE **7**:e35888–7.

21. **Datsenko KA**, **Pougach K**, **Tikhonov A**, **Wanner BL**, **Severinov K**, **Semenova E**. 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nature Communications **3**:945–7.

22. **Nicholson TJ**, **Jackson SA**, **Croft BI**, **Staals RHJ**, **Fineran PC**, **Brown CM**. 2019. Bioinformatic evidence of widespread priming in type I and II CRISPR-Cas systems. RNA Biology **16**:566–576.

23. **Jackson SA**, **McKenzie RE**, **Fagerlund RD**, **Kieper SN**, **Fineran PC**, **Brouns SJJ**. 2017. CRISPR-Cas: Adapting to change. Science **356**:eaal5056–11.

24. **Nussenzweig PM**, **McGinn J**, **Marraffini LA**. 2019. Cas9 Cleavage of Viral Genomes Primes the Acquisition of New Immunological Memories. Cell Host and Microbe **26**:515–526.e6.

25. **Modell JW**, **Jiang W**, **Marraffini LA**. 2017. CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. Nature **544**:1–18.

26. **Wei Y**, **Chesne MT**, **Terns RM**, **Terns MP**. 2015. Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. Nucleic Acids Research **43**:1749–1758.

27. **Díez-Villaseñor C**, **Guzmán NM**, **Almendros C**, **García-Martínez J**, **Mojica FJM**. 2013. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of Escherichia coli. RNA Biology **10**:792–802.

28. **Kieper SN**, **Almendros C**, **Brouns SJJ**. 2019. Conserved motifs in the CRISPR leader sequence control spacer acquisition levels in Type I-D CRISPR-Cas systems. FEMS Microbiology Letters **366**.

29. **Shah SA**, **Garrett RA**. 2011. CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. Res Microbiol **162**:27–38.

30. **Lillestøl RK**, **Shah SA**, **Brügger K**, **Redder P**, **Phan H**, **Christiansen J**, **Garrett RA**. 2009. CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. Mol Microbiol **72**:259–272.

31. **Garrett RA**, **Vestergaard G**, **Shah SA**. 2011. Archaeal CRISPR-based immune systems: Exchangeable functional modules. Trends Microbiol **19**:549–556.

32. **Alkhnbashi OS**, **Shah SA**, **Garrett RA**, **Saunders SJ**, **Costa F**, **Backofen R**. 2016. Characterizing leader sequences of CRISPR loci. Bioinformatics **32**:i576–i585.

33. **Makarova KS**, **Haft DH**, **Barrangou R**, **Brouns SJJ**, **Charpentier E**, **Horvath P**, **Moineau S**, **Mojica FJM**, **Wolf YI**, **Yakunin AF**, **van der Oost J**, **Koonin EV**. 2011. Evolution and classification of the CRISPR–Cas systems. Nat Rev Microbiol **9**:467–477.

34. **Laanto E**, **Hoikkala V**, **Ravantti J**, **Sundberg L-R**. 2017. Long-term genomic coevolution of host-parasite interaction in the natural environment. Nature Communications **8**:931.

35. **Erdmann S**, **Garrett RA**. 2012. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. Mol Microbiol **85**:1044–1056.

36. **Pinilla-Redondo R**, **Mayo-Muñoz D**, **Russel J**, **Garrett RA**, **Randau L**, **Sørensen SJ**, **Shah SA**. 2020. Type IV CRISPR-Cas systems are highly diverse and involved in competition between plasmids. Nucleic Acids Research **48**:2000–2012.

37. **Shmakov S**, **Abudayyeh OO**, **Makarova KS**, **Wolf YI**, **Gootenberg JS**, **Semenova E**, **Minakhin L**, **Joung J**, **Konermann S**, **Severinov K**, **Zhang F**, **Koonin EV**. 2015. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Molecular Cell **60**:385–397.

38. **Sternberg SH**, **Richter H**, **Charpentier E**, **Qimron U**. 2016. Adaptation in CRISPR-Cas Systems. Molecular Cell 1–12.

39. **Smargon AA**, **Cox DBT**, **Pyzocha NK**, **Zheng K**, **Slaymaker IM**, **Gootenberg JS**, **Abudayyeh OA**, **Essletzbichler P**, **Shmakov S**, **Makarova KS**, **Koonin EV**, **Zhang F**. 2017. Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. Molecular Cell **65**:618–630.e7.

40. **Barrangou R**, **Gersbach CA**. 2017. Expanding the CRISPR Toolbox: Targeting RNA with Cas13b. Molecular Cell **65**:582–584.

41. **Zhu Y**, **Klompe SE**, **Vlot M**, **van der Oost J**, **Staals RHJ**. 2018. Shooting the messenger: RNA-targetting CRISPR-Cas systems. Biosci Rep **38**.

42. **Deng L**, **Garrett RA**, **Shah SA**, **Peng X**, **She Q**. 2013. A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. Mol Microbiol **87**:1088–1099.

43. **Abudayyeh OO**, **Gootenberg JS**, **Konermann S**, **Joung J**, **Slaymaker IM**, **Cox DBT**, **Shmakov S**, **Makarova KS**, **Semenova E**, **Minakhin L**, **Severinov K**, **Regev A**, **Lander ES**, **Koonin EV**, **Zhang F**. 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science **353**:aaf5573.

44. **Dugar G**, **Leenay RT**, **Eisenbart SK**, **Bischler T**, **Aul BU**, **Beisel CL**, **Sharma CM**. 2018. CRISPR RNA-Dependent Binding and Cleavage of Endogenous RNAs by the Campylobacter jejuni Cas9. Molecular Cell **69**:893–905.e7.

45. **Rousseau BA**, **Hou Z**, **Gramelspacher MJ**, **Zhang Y**. 2018. Programmable RNA Cleavage and Recognition by a Natural CRISPR-Cas9 System from Neisseria meningitidis. Molecular Cell **69**:906–914.e4.

46. **Maikova A**, **Peltier J**, **Boudry P**, **Hajnsdorf E**, **Kint N**, **Monot M**, **Poquet I**, **Martin-Verstraete I**, **Dupuy B**, **Soutourina O**. 2018. Discovery of new type I toxin-antitoxin systems adjacent to CRISPR arrays in Clostridium difficile. Nucleic Acids Research **46**:4733–4751.

47. **Koonin EV**, **Zhang F**. 2016. Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. Bioessays **39**:e201600186–9.

48. **Zhang Y**, **Heidrich N**, **Ampattu BJ**, **Gunderson CW**, **Seifert HS**, **Schoen C**, **Vogel J**, **Sontheimer EJ**. 2013. Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. Molecular Cell **50**:488–503.

49. **Smargon A**, **Cox DBT**, **Pyzocha N**, **Zheng K**, **Slaymaker IM**, **Gootenberg JS**, **Abudayyeh OA**, **Essletzbichler P**, **Shmakov S**, **Makarova KS**, **Koonin EV**, **Zhang F**. 2016. Cas13b is a Type VI-B CRISPR-associated RNA-Guided RNase differentially regulated by accessory proteins Csx27 and Csx28 1–49.

50. **Ravantti JJ**, **Laanto E**, **Papponen P**, **Sundberg L-R**. 2019. Complete Genome Sequence of Fish Pathogen Flavobacterium columnare Strain B185, Originating from Finland. Microbiol Resour Announc **8**.

51. **Laanto E**, **Sundberg L-R**, **Bamford JKH**. 2011. Phage specificity of the freshwater fish pathogen Flavobacterium columnare. Appl Environ Microbiol **77**:7868–7872.

52. **Hynes AP**, **Villion M**, **Moineau S**. 2014. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. Nature Communications **5**:4399.

53. **Stern A**, **Keren L**, **Wurtzel O**, **Amitai G**, **Sorek R**. 2010. Self-targeting by CRISPR: gene regulation or autoimmunity? Trends in Genetics **26**:335–340.

54. **Meeske AJ**, **Nakandakari-Higa S**, **Marraffini LA**. 2019. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. Nature **570**:241–245.

55. **Leenay RT**, **Beisel CL**. 2016. Deciphering, communicating, and engineering the CRISPR PAM. Journal of Molecular Biology 1–45.

56. **Chen S**, **Bagdasarian M**, **Kaufman MG**, **Walker ED**. 2007. Characterization of strong promoters from an environmental Flavobacterium hibernum strain by using a green fluorescent protein-based reporter system. Appl Environ Microbiol **73**:1089–1100.

57. **Lopatina A**, **Medvedeva S**, **Shmakov S**, **Logacheva MD**, **Krylenkov V**, **Severinov K**. 2016. Metagenomic Analysis of Bacterial Communities of Antarctic Surface Snow. Front Microbiol **7**:255–13.

58. **Heler R**, **Samai P**, **Modell JW**, **Weiner C**, **Goldberg GW**, **Bikard D**, **Marraffini LA**. 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature **519**:199–202.

59. **Staals RHJ**, **Jackson SA**, **Biswas A**, **Brouns SJJ**, **Brown CM**, **Fineran PC**. 2016. Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR–Cas system. Nature Communications **7**:1–13.

60. **Richter C**, **Dy RL**, **McKenzie RE**, **Watson BNJ**, **Taylor C**, **Chang JT**, **McNeil MB**, **Staals RHJ**, **Fineran PC**. 2014. Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. Nucleic Acids Research **42**:8516–8526.

61. **Levy A**, **Goren MG**, **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.

62. **Staals RHJ**, **Jackson SA**, **Biswas A**, **Brouns SJJ**, **Brown CM**, **Fineran PC**. 2016. Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR–Cas system. Nature Communications **7**:1–13.

63. **Laanto E**, **Hoikkala V**, **Ravantti J**, **Sundberg L-R**. 2017. Long-term genomic coevolution of host-parasite interaction in the natural environment. Nature Communications **8**:111.

64. **Westra ER**, **van Houte S**, **Oyesiku-Blakemore S**, **Ben Makin**, **Broniewski JM**, **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.

65. **Pawluk A**, **Davidson AR**, **Maxwell KL**. 2017. Anti-CRISPR: discovery, mechanism and function. Nature Publishing Group **16**:12–17.

66. **Goren MG**, **Doron S**, **Globus R**, **Amitai G**, **Sorek R**, **Qimron U**. 2016. Repeat Size Determination by Two Molecular Rulers in the Type I-E CRISPR Array. CellReports **16**:2811–2818.

67. **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**:7884–13.

68. **Wang J**, **Li J**, **Zhao H**, **Sheng G**, **Wang M**, **Yin M**, **Wang Y**. 2015. Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. Cell **163**:840–853.

69. **Nuñez JK**, **Kranzusch PJ**, **Noeske J**, **Wright AV**, **Davies CW**, **Doudna JA**. 2014. Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. Nature Publishing Group **21**:528–534.

70. **Rollie C**, **Schneider S**, **Brinkmann AS**, **Bolt EL**, **White MF**. 2015. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. Elife **4**.

71. **Arslan Z**, **Hermanns V**, **Wurm R**, **Wagner R**, **Pul Ü**. 2014. Detection and characterization of spacer integration intermediates in type I-E CRISPR–Cas system. Nucleic Acids Research **42**:7884–7893.

72. **Shmakov S**, **Shmakov S**, **Savitskaya E**, **Savitskaya E**, **Semenova E**, **Semenova E**, **Logacheva MD**, **Logacheva MD**, **Datsenko KA**, **Datsenko KA**, **Severinov K**, **Severinov K**. 2014. Pervasive generation of oppositely oriented spacers during CRISPR adaptation. Nucleic Acids Research **42**:5907–5916.

73. **Chylinski K**, **Makarova KS**, **Charpentier E**, **Koonin EV**. 2014. Classification and evolution of type II CRISPR-Cas systems. Nucleic Acids Research **42**:6091–6105.

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

75. **Westra ER**, **van Houte S**, **Gandon S**, **Whitaker R**. 2019. The ecology and evolution of microbial CRISPR-Cas adaptive immune systems. Phil Trans R Soc B **374**:20190101–8.

76. **Hoikkala V**, **Almeida GMF**, **Laanto E**, **Sundberg L-R**. 2019. Aquaculture as a source of empirical evidence for coevolution between CRISPR-Cas and phage. Phil Trans R Soc B **374**:20180100.

IV

ECOLOGICAL DETERMINANTS OF PHAGE DEFENCE STRATEGY IN AN OPPORTUNISTIC PATHOGEN

by

Gabriel Almeida, Ville Hoikkala, Janne Ravantti &
Lotta-Riina Sundberg 2020

Manuscript

Manuscript: Ecological determinants of phage defense strategies in an opportunistic pathogen

Gabriel Almeida, Ville Hoikkala, Janne Ravantti, Lotta-Riina Sundberg

Abstract

Host-parasite interactions have resulted in the emergence of a wide variety of immune systems in all domains of life. Prokaryotes utilize multiple different defenses to prevent infections by their viruses, bacteriophages (phages) [1]. However, it is unclear what environmental stimuli lead to the deployment of specific defenses [2]. Here, we studied this question in a three-partite setting, where the bacterial fish pathogen *Flavobacterium columnare* was exposed to its virulent phages in the presence of eukaryotic host signals (mucin). Mucin exposure has been shown to increase virulence in *F. columnare* while making the bacterium susceptible to phage, as virulence and phage resistance operate on the same cell surface receptors [3,4]. We examined if the lack of surface modification in the presence of mucin is compensated by heightened activity in the bacterium's type II-C and VI-B CRISPR-Cas loci. We co-cultured the bacterium and phage for 16 weeks in low or high nutrient media and with or without added mucin. Bacteria and phages persisted through the 16-week experiment, bacteria becoming resistant to the ancestral phage in all conditions. Mucin triggered dramatic increase in CRISPR spacer acquisition, especially in low nutrient conditions where over 60% of colonies had obtained at least one new spacer. The effect of new spacers on bacterial growth depended on the treatment and bacterial morphology. Genomic analyses showed that despite active spacer uptake, bacterial surface protein mutations occurred in all treatments, while phage genomes showed only few or no mutations. These results suggest that environmental factors are important in determining phage defense strategies, and that the concentration of phages on metazoan surfaces may select for the diversification of bacterial immune systems.

Intro

One of the central questions in host-parasite interactions is to understand why there is such a diversity of defense strategies, and why certain defense mechanisms are favored over others [2,5-7]. It has been speculated that the variety of different defenses exists for the reason of plasticity: one defense strategy may be useful in one setting but inefficient, or costly, in another [2]. Having a repertoire of defenses therefore expands the ecological niche, allowing the organism to shift between suitable defenses depending on the current environmental stimuli. However, the ecological stimuli that lead to selective deployment of different immune strategies are still poorly understood. Evolutionary pressures behind antimicrobial defense mechanisms might be even more complex in systems where multiple trophic levels are involved, as pathogens to certain organisms may be hosts to others. Examples of this are pathogenic bacteria that infect eukaryotic hosts, but are also parasitized by their viruses, bacteriophages (phages). This three-partite relationship involves two layers of defenses: one between the bacterium and the phage, and another between the metazoan and the bacterium. The interplay of these layers is still poorly understood [8], and in some instances evolutionary interactions have been described even between the phage and the metazoan [3,9,10].

Bacterial defense mechanisms against phage infections are numerous and have evolved towards almost all phases of phage life cycles [1]. Despite decades of research, novel mechanisms are still being found [11,12]. Surface modification (SM) is an extracellular defense response, in which the mutation, downregulation or deletion of cell surface proteins prevents phages from attaching to the cell [1]. Within the cell, restriction modification and CRISPR-Cas systems interrogate invading phage genomes

by distinguishing them from host nucleic acids and targeting them for destruction. CRISPR-Cas is an adaptive immune system that stores genetic memories of phage genomes (spacers) into a repeat-spacer array in a CRISPR locus. Upon a subsequent infection, the array is expressed to CRISPR-RNA (crRNA) [13]. The resulting repeat-spacer oligos guide endonucleases, such as Cas9, to the spacer-complementary sequence in the phage genome that is then cut by the endonuclease [14,15].

While defense mechanisms can be effective at curtailing phage infections, they usually come with a cost. Since phage receptor proteins often play an important role in bacterial nutrient intake, secretion, motility or virulence, disruptions in these proteins generally impose fitness costs [1,16,17]. It has been suggested that these tradeoffs are important for maintaining phage-bacterium coexistence [17]. Costs may be permanent or short-lived: constitutive defences, such as SM, impose continuous costs, whereas inducible defences, such as CRISPR-Cas, minimize their costs by being activated only under specific stimuli. The relative investment to inducible or constitutive defences depends on their relative costs in a given environment [6]. For example, surface receptors required for virulence can have major fitness costs if they are mutated during host colonization [18].

Despite the sophistication of CRISPR-Cas, less than half of all bacterial species carry CRISPR-Cas loci [19], implying they may carry fitness costs or be useful in limited circumstances. Ecological variables such as temperature and oxygen levels are indeed important factors that correlate with CRISPR-Cas occurrence [20]. Interestingly, type II CRISPR-Cas systems seem to be enriched in pathogenic bacteria [20], and CRISPR-Cas has been suggested to regulate bacterial virulence in some bacterial species [21], e.g. in *Francicella novicida* [22] and *Campylobacter jejuni* [23,24]. However, as multiple defenses may be beneficial [2], bacteria with CRISPR-Cas do not only rely on this defense. The relative costs of CRISPR-Cas have been found to increase with phage concentration, so that above a certain threshold, these costs exceed those of SM, leading to a shift in defense strategy [6]. Furthermore, presence of competing bacteria can select for increased CRISPR-Cas based resistance due to an amplification in SM-based costs [25]. The factors that influence defense strategies are still largely unknown and they may have important effects on phage evolution (e.g. via genomic mutations or the evolution of anti-CRISPR proteins) and population size. Indeed, while surface modifications prevent phage adsorption and thus do not influence population density, CRISPR-Cas may act as a sink that decreases phage population size [26].

The colonization of eukaryotic hosts is central to pathogenic bacteria, and likewise, eukaryotic hosts utilize multiple innate and adaptive immune systems to combat bacterial infections. The outer layer of many metazoans is covered with mucus that provides an immunological barrier against bacterial infections [27]. Tailed phages have evolved a symbiotic relationship with eukaryotes by binding to mucin glycoproteins on the mucosal surfaces [9]. This enhances phages' probability of encounter with bacterial hosts, while providing the host external mucosal immunity against invading bacteria [3,9,10]. The abundance of phages on the mucosal layer, combined with potential to colonize the eukaryotic host, may result in signals that promote both the upregulation of virulence factors and the triggering of phage defense in bacterial pathogens. If phage defenses incur tradeoffs in virulence, conflicts may emerge. Thus, phage defense strategies that minimize virulence trade-offs during colonization may be favored.

We hypothesized that the mucosal interface may be a tipping point between CRISPR-Cas based defense that likely maintains virulence and SM that may reduce virulence (Figure 1). We investigated this idea using the opportunistic fish pathogen *Flavobacterium columnare* that causes columnaris disease in freshwater fish [28]. This bacterium has type II-C and VI-B CRISPR-Cas loci, both active in natural and laboratory environments [29,30]. Exposure to high concentration of phage in laboratory conditions elicits SM defense in *F. columnare*, causing loss of virulence and motility associated with

colony morphotype change [31]. Low phage pressure and low nutrient level, on the other hand, can be used to trigger CRISPR-Cas spacer acquisition [30]. Exposure to mucin has been shown to increase virulence and chemoattraction towards fish in *F. columnare,* while also making the bacteria more susceptible to phage infection [3]. A transcriptome study also revealed the upregulation of several virulence genes, as well as the CRISPR-Cas adaptation gene *cas2* (in the absence of phage) [32]. How simultaneous exposure to host signals (mucin) and phage predation affects the choice of phage defense strategy between CRISPR-Cas and SM in *F. columnare* has not been explored.

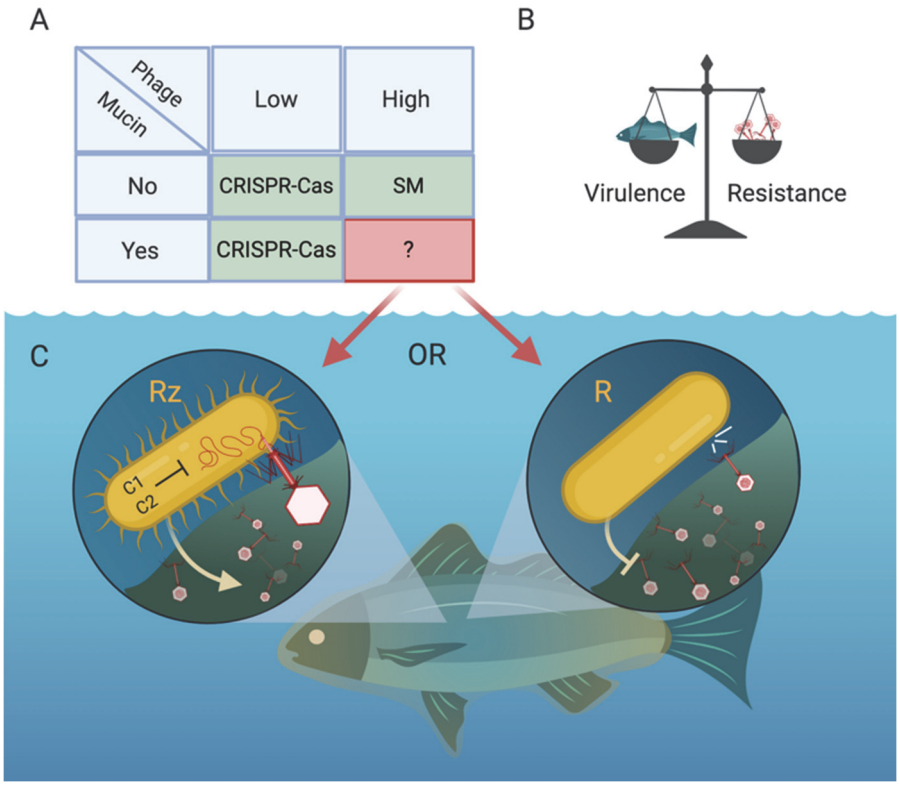

Figure 1. Optimal defense strategies may be defined by the environment. A) The selection of a specific immune response can be affected by the force of phage infection and presence of host-signals (mucin). B) The choice of immunity is affected by the relative trade-offs associated with phage resistance and bacterial virulence. C) The presence of host-signals promotes bacterial virulence factors, yet simultaneous high phage pressure promotes surface modification that reduces virulence. Intracellular defenses, such as CRISPR-Cas, may enable colonization in the presence of phage.

We co-cultured *F. columnare* strain B245 and its virulent phage V156 in different low-nutrient growth conditions consisting of sterile lake water or diluted growth medium, with or without purified mucin. Here, we show that bacterial surface modification is common across the treatments in response to phage exposure. The presence of mucin, however, increases CRISPR-Cas spacer uptake to remarkably high levels, where more than 60% of the bacterial population acquire new spacers and up to ten new spacers in a single isolate. Our study provides a fresh outlook into the ecological factors that shape immune defense strategies in the face of phage predation and host colonization.

Materials and methods

Phage and bacteria

Flavobacterium columnare strain B245 was isolated from the same fish farm in Central Finland in 2009 as its phage V156 [33]. Conventional culturing of B245 was made using Shieh medium [34] without glucose. V156 stocks were produced by harvesting confluent soft-agar layers from double-agar plates,

adding 4ml of media, centrifuging (10.000 rpm, 10 minutes, Sorvall RC34 rotor) and filtering the supernatant through 0.22 µm filters.

Long-term culturing conditions and sampling

The effect of four different nutritional conditions on phage resistance mechanisms were studied: autoclaved lake water alone, lake water supplemented with 0.1% purified porcine mucin, 0.1x Shieh media alone 0.1x Shieh media supplemented with 0.1% purified porcine mucin. The Lake water was collected from Lake Jyvasjarvi (Jyvaskyla, Finland) on 13.2.2018 and autoclaved. The water was analyzed by Eurofins Scientific and contained N: 790 µg/l, P: 18 µg/l and Fe: 540 µg/l. Shieh media was diluted in ultrapure sterile water. Autoclaved 2% w:v solution of purified porcine mucin (Sigma-Aldrich, catalog no. M1778) was used as stock.

The initial inoculum was $5x10^4$ cfu of B245 and $5x10^3$ pfu of V156 (multiplicity of infection of 0.1). As non-infected controls, one culture of each condition was made with only B245 without phage. Each condition was tested in triplicates, in a final volume of five milliliters, and incubated at 26 degrees under 120 rpm. Every week after day zero, one milliliter of each culture was removed and replaced with one milliliter of the corresponding culturing media (autoclaved lake water or 0.1x Shieh, supplemented or not with 0.1% mucin). An overview of the experiment setup can be seen in Figure 2.

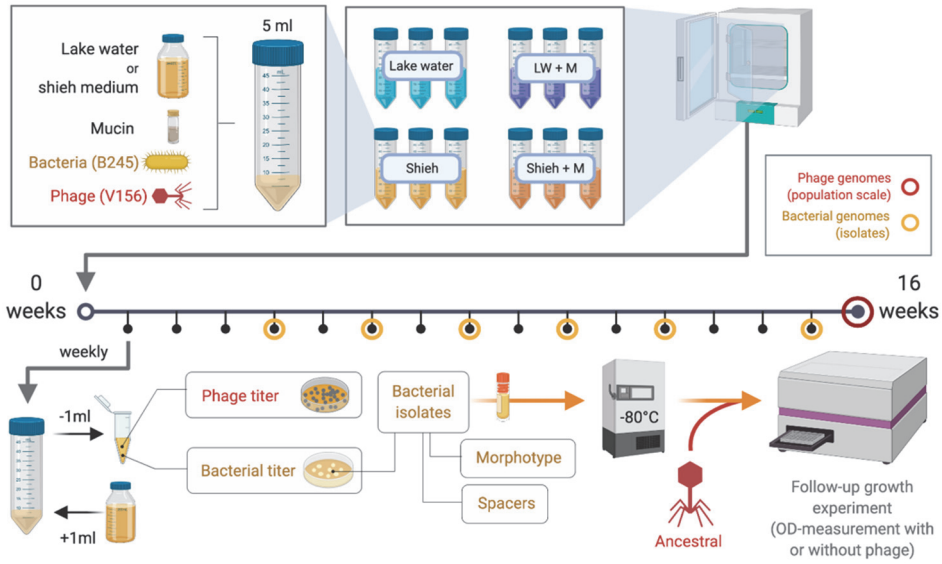

Figure 1. Overview of the experimental setup. The 16-week experiment (denoted by the horizontal line) contained four culturing conditions, which were sampled and restocked with fresh media weekly. These isolates were characterized in their morphotype and CRISPR spacer content, and used later in the follow-up growth experiment with or without phage. Phage genomes were extracted at week 16 while bacterial genomes were sequenced throughout the experiment. Figure made in ©BioRender - biorender.com

Bacteria and phage titrations

Immediately after sampling, each sample was serially diluted on Shieh-agar plates for analysis of bacterial population size. Chloroform (10% v:v) was added to the remaining sample to kill bacterial cells. Serial dilutions of the chloroform-treated supernatants were used for titrating phages with the double-agar layer method (**Adams 1959**) using the parental B245 as host. Bacteria and phage plates were incubated at room temperature for three days, followed by the enumeration of bacterial colonies and viral plaques. Phages were titrated every week during the experiment, while bacteria were titrated in weeks 1 to 8, 10, 12 and 15.

Detecting the acquisition of new CRISPR spacers (both loci)

We aimed at collecting an equal number of rough and rhizoid colonies per sample whenever possible. After counting, random colonies were picked and transferred to 50 microliters of Shieh on 96 well plates. Two microliters of each resuspended colony were used as template for PCR reactions designed to detect the insertion of new spacers on both CRISPR loci. Reactions were made with DreamTaq polymerase (Thermo Fisher), in 20 µl reactions containing 0.5mM of DNTPs and 0.5 µM of each primer using previously published primers for strain B245 [30]. Cycling conditions were 95 degrees for 3 minutes followed by 30 cycles of 95 degrees for 30 seconds, 60 degrees for 30 seconds, 72 degrees for one minute and a final extension step of 72 degrees for 15 minutes. PCR reactions were resolved in 2% agarose gels and the addition of spacers verified by the size of each amplicon.

Bacterial growth characteristics

Representative bacterial isolates, considering colony morphology and CRISPR loci sizes, were chosen every week for further analysis. Following PCR, the remaining volume of resuspended colonies was transferred to 5ml of Shieh media. After overnight growth (120 rpm, 26 degrees) the isolates were frozen at -80 degrees for future use. For testing the immunity of individual colonies collected along the experiment against the parental phage, we quantified bacterial using Bioscreen C® (Growth curves Ltd, Helsinki, Finland). Stored isolates were taken from the freezer and grown overnight. Then 10^3 cfu/ml of each isolate were added to Bioscreen plates, in triplicates (200 microliters per well). Each isolate was tested in the presence and in the absence of the parental V156 virus (10^3pfu/ml, moi 1). Parental B245 was included on every plate as control. Optical density measurements were made every 10 minutes for a period of 4 days. Plates were kept without agitation at 27 degrees for the whole time. Minor differences in individual plates were included as random effects in statistical models when appropriate. However, the differences between the plates was generally minimal (Supplementary Figure 1).

Genomic DNA extraction and sequencing

Population-level phage DNA at week 16 and clonal DNA of selected bacterial isolates from different time points were sequenced with Illumina (BGI). For bacterial genomic DNA extraction, isolates were taken from the freezer and grown overnight. DNA of turbid was extracted using the GeneJet Genomic DNA Purification Kit (Thermo Fisher). For phage DNA extraction, lysates from week 16 were used to infect B245. Confluent soft-agar bacterial lawns were collected, mixed with 4ml of Shieh media, centrifuged (10.000 rpm, 10 minutes, Sorvall RC34) and filtered. Phage precipitation was made with ZnCl2 and cleaning of host DNA with nucleases, as described in **Santos 1991**. After Protease K treatment, the material was mixed with Guanidine:Ethanol and the extraction finished using the kit mentioned above. All samples were sequenced using 150PE BGISEQ platform at BGI Group. We were unable to obtain phage sequence data from sample LW+M b due to technical problems.

Construction of the B245 reference genome

The ancestral B245 genome was assembled from Illumina reads using Spades 3.14.1 (--isolate mode) [35]. The resulting 487 contigs were combined to a single contig relying on a previously compiled complete *F. columnare* genome FCO-F2 (accession number CP051861, not yet online) as reference using RagOO 1.1 [36]. The genome was annotated for the purpose of mutational analysis using dFast 1.2.3 [37]. This unpublished B245 genome is available for review purpose at https://bit.ly/37X6Uyx.

Mutation analysis

We used Breseq 0.35.1 [38] to analyze mutations occurring in the phage and bacterial genomes. Since the phage samples represented mixed phage populations, Breseq was run in --polymorphism-prediction mode. Bacterial samples were run in default mode. As references we used the ancestral B245 bacterial genome or the previously published v156 phage genome [29]. To reduce noise in phage mutation, we used a value of 0.05 for --polymorphism-bias-cutoff. During analysis of the phage genomes, we discovered the existence of a large number of identical polymorphisms in all or some samples. As it is very unlikely that such mutations occur independently, we discarded any identical mutations that co-occur in two or more samples as false positives. It is possible that these mutations arise from an undiscovered prophage or pseudolysogen – however, no trace of phage emerged when the control samples grown in similar conditions without phage were assayed for plaques. When investigating mutations in unknown putative phage genes, we used HHpred to predict protein function [39].

Statistics/Data analysis

All data analysis and statistic were done in R 3.5.3 using RStudio 1.1.463. Model comparisons and inclusion of random effects was aided by Akaike's information criterion (AIC) comparison when applicable [40].

Effect of treatment of phage resistance

In the follow-up growth experiment, we used the R library glmmTMB [41] to analyze how the isolates differ from the ancestral bacterium in the presence or absence of phage. We used maximum OD reached during the follow-up growth experiment (OD_{MAX}) as the response variable and the interaction between phage and treatment as the explanatory variables. We included replicate and bacterial ID as random effects and used a gaussian distribution as the response variable was normally distributed. The control samples were investigated using a linear model (lm in base-R) using the same interaction terms (the same random structure as in the phage-exposed samples could not be used due to limited sample size in the control dataset).

Differences in titers of LW and LW+M

To reveal the effect of mucin on growth in lake water, we analyzed phage and bacterial titers from these conditions using a linear model in R. Titer was set as the response variable and treatment (LW or LW+M) as the explanatory variable. For phage titers, we only considered time points after 9 weeks, as the divergence between the two treatments started at this time (Supplementary Figure 2A). As the bacterial titers diverged starting from week 1 (Supplementary Figure 2B), we included all time points in the model. Similar analysis was not done for the Shieh and Shieh + mucin due to large variation of replicate cultures in these conditions.

The effect of spacers and morphology

To analyze the effects of spacers and morphology on OD_{MAX} in the LW+M and Shieh+M treatments, we created a generalized linear mixed model (GLMM) using the R package glmmTMB [41]. Both treatments were studied separately and categorized as having no new spacers or having one or more spacers. OD_{MAX} was used as the response variable and predictors were [the presence of spacers] * [morphology]. As random effects we included plate ID (the plate on which each sample was measured on) and replicate, which both increased model-fit considerably. A gamma distribution of the response variable was favored over other distributions with AIC difference > 2.

The effect of phage on control samples

We examined if OD_{MAX} of the control samples (grown in the four treatments without phage) in the follow-up growth experiment behaved similarly to the ancestral bacterium. We used a linear model (lm) in R with OD_{MAX} as the response variable and treatment as the explanatory variable. The ancestral bacterium was used as intercept against which all treatments were compared to.

Effect of extraction time on OD_{MAX}

Using a linear mixed model, we investigated the effect of extraction time on OD_{MAX} and OD_T. Including replicate as a random factor produced the best models according to AIC. None of the samples showed a significant effect of isolation time on either response variable in the presence or absence of phage (Supplementary Figure 3).

Results

***Presence of mucin stabilizes survival of both the bacterium and the phage during* 16 weeks of co-culture**

To avoid population bottlenecks, our sampling was based on weekly collection 20% of the cultures and replacing with the same volume of fresh medium. Long term co-existence of both B245 and V156 was observed in all treatments. In lake water with (LW+M) or without mucin (LW) the phage titers remained similar until week 9, after which LW+M showed a significant decline in phage numbers compared to LW (LM, T_{46}=-2.737, P = 0.0087) with roughly a ten-fold difference at week 16 (Figure 3A, Supplementary Figure 2A). The effects on bacteria in these treatments were opposite and more dramatic: bacterial numbers in LW+M were an average of 45-fold higher across all time points after an initial spike at week 1 (LM, T_{77}=3.893, P < 0.001) (Figure 3b, Supplementary Figure 2B). Surprisingly, the no-phage control of LW+M became extinct after week 10, while no extinction occurred in the phage-containing cultures.

Cultures in Shieh medium (without mucin) had large variation between replicates in both bacterial and phage titers. Despite similar bacterial titers at week one, differences between replicates grew to more than 100-fold towards the end of the experiment (Figure 3B). Phage titers declined in Shieh for the first 4-5 weeks, after which replicates *b* and *c* recovered while replicate *a* stabilized. Around week 10, phage titers declined sharply in replicates *a* and replicate *c*, while replicate *b* remained with high titer until the end (Figure 3A). Presence of mucin in Shieh decreased variation between replicates compared to Shieh alone, with high both phage and bacterial titers (Figure 3). No statistical analysis was performed on the Shieh cultures due to high divergence in replicates.

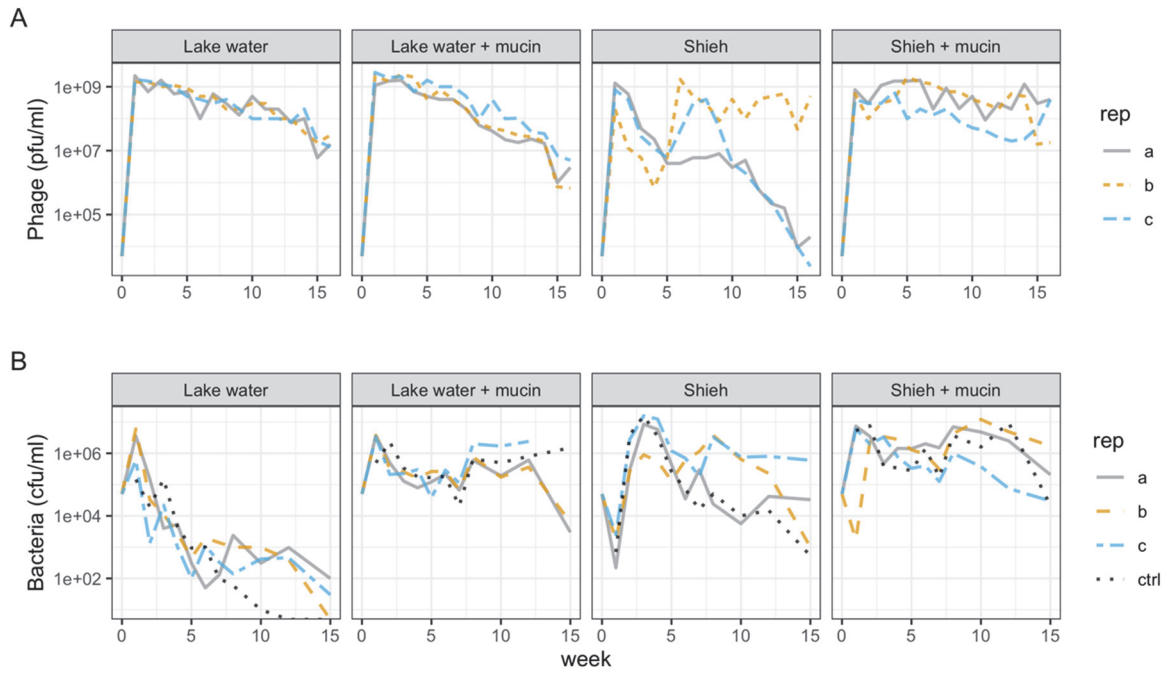

Figure 3. Phage (A) and bacterial (B) titers over the 16-week experiment in the four treatments. Each line represents one of the three replicates in each treatment. The dotted lines in (B) are control cultures without phage.

Presence of mucin determines spacer acquisition in type II-C and VI-B CRISPR-Cas loci

Culture conditions, especially presence of mucin, had a significant impact in the acquisition of new spacers (Figure 4A-B). The presence of mucin in lake water increased spacer acquisition 7.14-fold compared to plain lake water (GLMM, $Z_{3,68}$ = -3.718, P=0.0037), and the efficiency of acquisition was 2.43-fold when compared to Shieh with mucin (GLMM, $Z_{3,68}$ = -2.909, P<0.001) (Figure 4B). Shieh without mucin did not show any new spacers. The number of spacers in one locus in a single isolate was up to six spacers in LW+M, two in LW and up to three in Shieh with mucin (Figure 4B). The efficiency of spacer acquisition was roughly similar between the loci regardless of the treatment. Timewise, the maximum efficiency of acquisition was reached around week 3 in the LW+M treatment, with over 60% of LW+M colonies having acquired new spacers (Figure 4A). These data also show that spacer acquisition is not exclusive to the Rhizoid morphotype. In fact, in Shieh + mucin treatment, isolates with most spacers were Rough, indicating overlap of surface modification and spacer acquisition (Figure 4B). However, after a peak in isolates with CRISPR spacers in this treatment around week 7, newly acquired spacers disappeared on week 10. When only considering morphotype, LW had no Rhizoid colonies while other conditions had a minor bias towards them (Figure 4C).

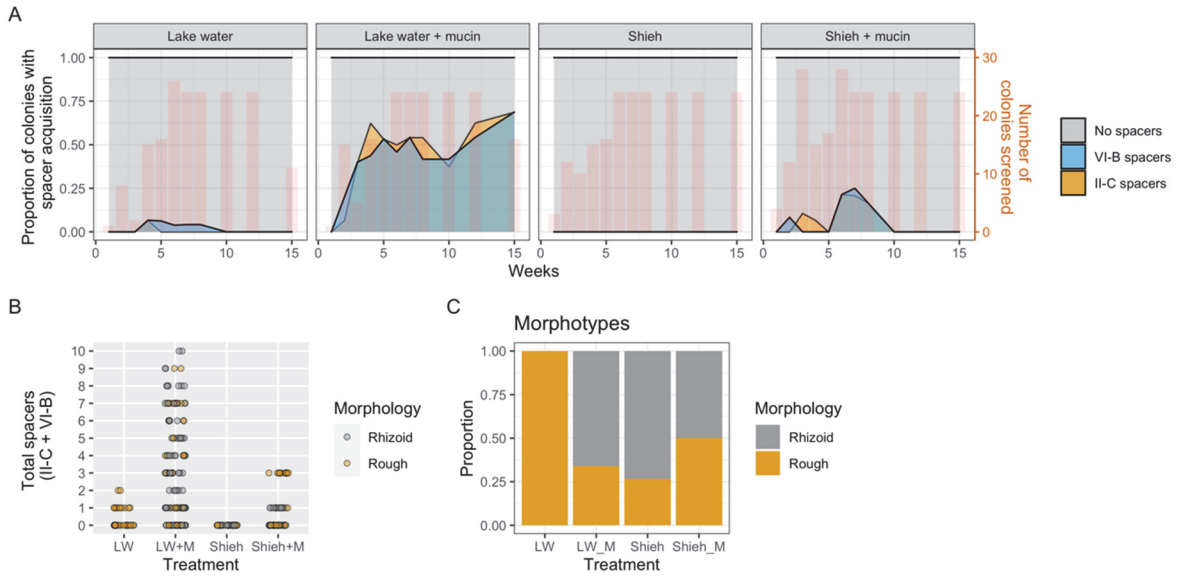

Figure 4. Dynamics CRISPR-Cas spacers and morphotypes. A) CRISPR-Cas spacer acquisition over time in both loci. The area graphs (left Y-axis) represents the proportion of colonies in which the CRISPR-Cas array expanded by at least one new spacer. Values are means from the three replicates. The red line (right Y-axis) shows the total number of colonies that were screened at each time point to obtain these proportions. Missing red bars indicate no screening on that week – the respective proportional data is therefore an interpolation. B) The absolute number of spacers from individual isolates in different treatments. New spacers in both loci have been added together C) Morphotype distribution across the treatments (isolates from all time points pooled together).

Co-culturing with phage leads to immunity

To detect the development of phage resistance and any associated costs during the 16-week co-culture experiment, we grew the isolates in the presence or absence of the ancestral phage and compared their maximum OD (OD_{MAX}) to that of ancestral B245. In the absence of phage, the OD_{MAX} of the four treatments did not significantly differ from the ancestral B245. In the presence of phage, however, the OD_{MAX} of the ancestral B245 decreased from 0.41 to 0.19 (GLMM, $Z_{4,241}$=-13.78, $P < 0.001$), while OD_{MAX} in the four treatments remained unchanged (GLMM, $Z_{4,241}$=-13.78, $P > 0.3$) (Figure 5A). When compared with the ancestor in the presence of phage, the predicted maximum ODs were significantly higher by 0.22 (LW), 0.21 (LW+M), 0.21 (Shieh) and 0.22 (Shieh + mucin) OD units (GLMM, $Z_{4,24}$=12.673, 13.590, 11.631, 13.250, respectively, $P < 0.001$ in all) (Figure 5A). These results suggest that co-culturing *F. columnare* with phage caused phage resistance to evolve in all four treatments.

We also measured isolates from the control cultures (same conditions but without phage). Isolates from all treatments behaved similarly in the presence or absence of phage, except the Shieh + mucin isolate (Figure 5B). This isolate had a significantly lower OD_{MAX} in the absence of phage compared to the ancestor (OD 0.41) with a predicted OD of 0.276 (LM, $T_{2,24}$=-3.053, $P = 0.0055$). In the presence of phage, however, this control had a predicted OD of 0.287, which is significantly higher than that of the ancestor's 0.199 (LM, $T_{2,24=}$3.565, $P = 0.0016$). These results suggest that prolonged incubation in Shieh + M in the absence of phage has equipped the cells with phage resistance, perhaps incidentally through adaptation to the nutrient-rich mucin environment.

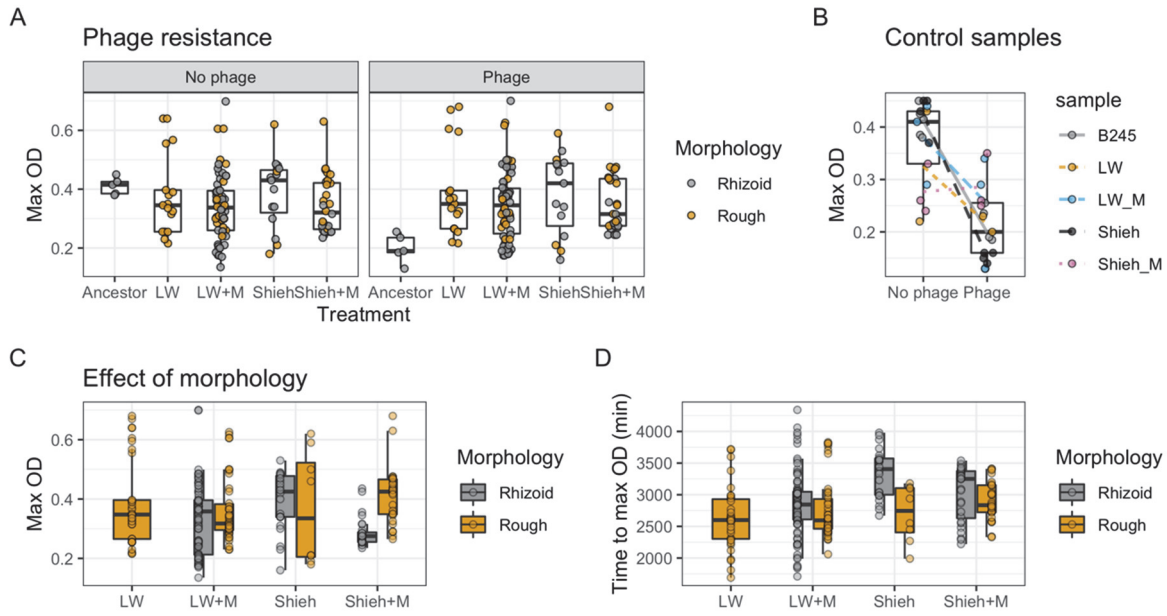

Figure 5. Effect of origin, colony morphology and CRISPR defense on bacterial growth. Each dot represents the mean of three replicate OD measurements of an isolate. A) Ancestral B245 and isolates originating from the four treatments in the presence or absence of phage. B) Control samples (grown without phage) from the same four treatments plus the ancestor in the presence or absence of phage. C) OD_{MAX} reached by isolates from different treatments in the follow-up growth experiment. Measurements with and without phage are pooled together, as the presence of phage does not affect growth of these isolates D) Same as C but showing the time to reach OD_{MAX} on the Y-axis.

As many of the isolates had changed morphotype during the experiment (Fig 4C), we next tested if different morphotypes (Rhizoid vs Rough) affected growth in the presence of phage (Figure 5C-D). Only one treatment showed a significant shift in growth: OD_{MAX} of Rough Shieh + M isolates was 0.14 units higher than the Rhizoid OD_{MAX} of 0.32 (GLMM, Z_{24}=5.69, P < 0.001). In Shieh, growth of Rough isolates did not differ from Rhizoid (OD_{MAX} 0.39, GLMM, Z_{11}=0.184, P = 0.854). Similarly, in LW+M the OD_{MAX} of Rough isolates did not differ from that of Rhizoid isolates (0.33) (GLMM, Z_{58}=0.76, P = 0.446). LW was not tested as this treatment only had Rough morphotypes (Fig 4C).

When examining time-to-max-OD (OD_T) (Figure 5D), Shieh isolates showed a significant difference between morphotypes: Rough isolates reached OD_T 20.2% faster than Rhizoid ones in the presence of phage (GLMM, Z_{11}=-2.82, P = 0.004), and 18% faster in the absence of phage (GLMM, Z_{11}=-2.82, P = 0.005). Other treatments did not show significant trends.

CRISPR adaptation can be costly if combined with surface modification

We then investigated how the spacer acquisition in different morphotypes affected growth. This analysis was only done for the LW+M and Shieh+M treatments, which produced enough isolates with expanded CRISPR arrays for statistical analysis. In the absence of phage, Rhizoid and Rough isolates with unchanged CRISPR arrays from the LW+M treatment reached predicted OD_{MAX} of 0.173 and 0.337, respectively (Figure 6A). This difference was statistically significant (GLMM, $Z_{1,50}$=6.34, P < 0.001). However, acquisition of one or more spacers significantly increased the OD_{MAX} of Rhizoid isolates from 0.127 to 0.337 (GLMM, $Z_{1,50}$= 5.56, P < 0.001). The effect of spacers was opposite for rough isolates, which had an OD_{MAX} of 0.34 (GLMM, $Z_{1,50}$=-5.326, P < 0.001). These effects were similar and significant also in the presence of phage.

In Shieh+M samples, Rough isolates had a similar effect on growth with an OD_{MAX} of 0.376 compared to the 0.285 of Rhizoid colonies (GLMM, $Z_{1,21}$=2.183, P = 0.029) (Figure 6B). New spacers did not have an effect in the absence of phage. In the presence of phage, spacers significantly increased OD_{MAX} Rough from 0.413 to 0.596 (GLMM, $Z_{1,21}$=2.314, P = 0.021). Spacers did not affect rhizoid isolates in this treatment.

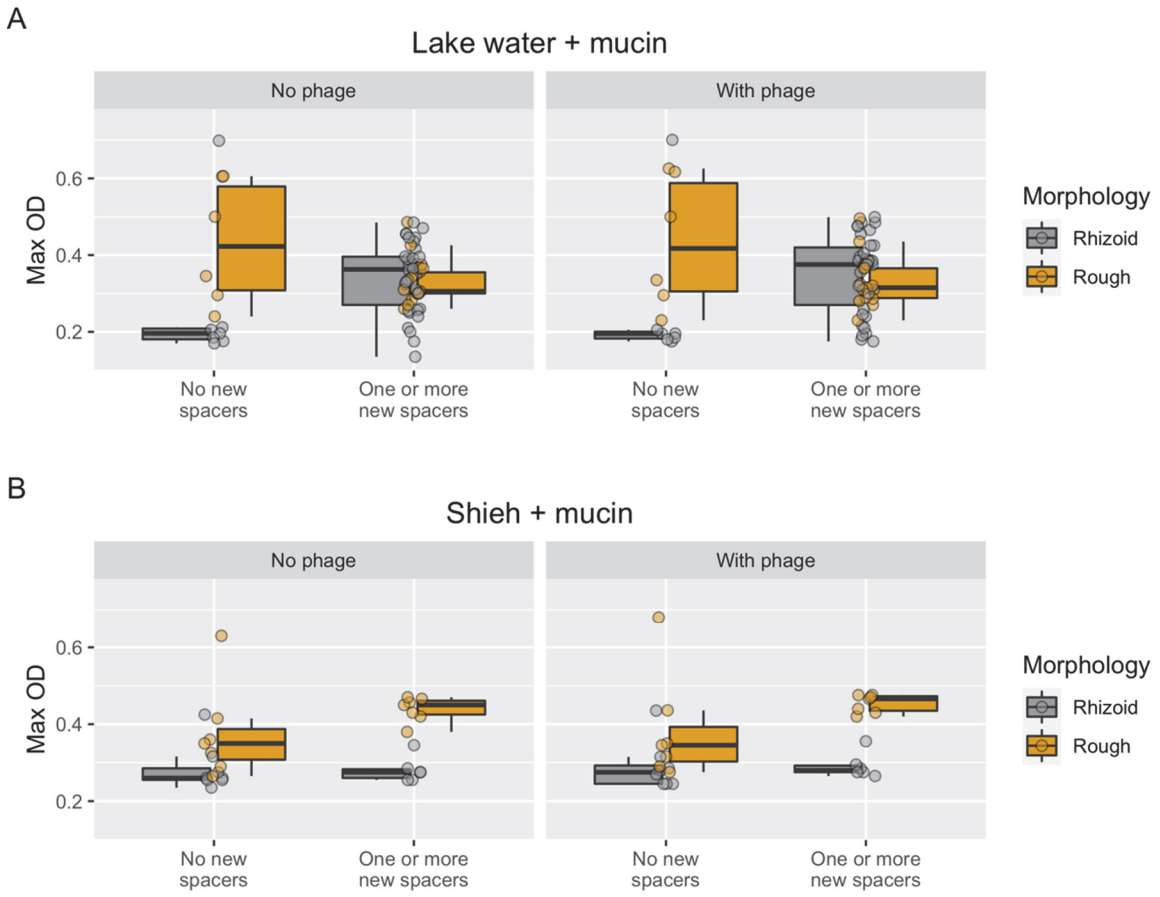

Figure 6. The effect of spacers and morphotype in LW+M and Shieh+M isolates on bacterial growth. Isolates are grouped into having no new spacers or having one or more new spacers regardless of the CRISPR-Cas locus.

Genome analysis

We sequenced phage and bacterial genomes to search for genetic variation resulting from coevolution. We also sequenced the control bacterial genomes (evolved without phage) to differentiate between mutations caused by interaction with phage and those that may stem from different culturing conditions.

Phage genomes were investigated on the population level: at the end of the 16-week experiment, phage samples of each culture (representing the variety of phages present in that culture) was used to infect the ancestral bacterium and the resulting lysate was deep sequenced. We found an abundance of shared mutations across multiple replicates. Due to the low likelihood of the same mutations occurring convergently across multiple samples, we discarded most mutations as false positives (however, all mutations are listed in Supplementary File 1). We were confidently able to recover individual mutations in only two cultures: Replicate *c* from Shieh + M treatment had a non-synonymous mutation ($A_{357}T$) in a predicted phage baseplate protein and a nonsynonymous mutation

($M_{200}V$) in a putative protein with no predicted function (Table 1). Replicate *b* from Shieh treatment had a non-synonymous mutation in an unknown protein near the aforementioned baseplate protein ($V_{300}I$), as well as a synonymous mutation ($K_{10}K$) in a predicted DNA helicase (Table 1). Common to both replicates is that they suffered a drop in phage titer after the initial spike but recovered towards the end (Figure 2A). In fact, replicate *b* in Shieh treatment was the only replicate in this treatment in which the phage titer was high at the end of the experiment.

Table 1. Mutations discovered in phage genomes on the population level. The frequency column shows the percentage of reads the mutation was found in. Position refers to the B245 genome. * = while the function of this protein could not be predicted, its genomic neighborhood suggests a role as a phage structural protein.

Treatment	Replicate	Frequency	Position	AA change	Protein prediction
Shieh + mucin	*c*	11.3 %	22,215	$A_{357}T$	Baseplate wedge protein
Shieh + mucin	*c*	6.6 %	38,183	$M_{200}V$	Unknown
Shieh	*b*	19.4 %	24,026	$V_{300}I$	Unknown*
Shieh	*b*	6.9 %	40,807	$K_{10}K$	DNA helicase

Bacterial genomes were investigated on the isolate (clonal) level. Isolates were picked from different treatments at several time points during the experiment. Evidence of surface modification was found in almost all phage-exposed samples in the flavobacterial gliding motility genes (T9SS) which are associated with type IX secretion system and whose mutations are expected to cause colony morphotype change from Rhizoid to Rough [4]. Most of these mutations caused premature stop codons or introduced frameshift mutations (Table 2). While most T9SS mutants were rough, some were also rhizoid (Table 2). Isolates from the control treatments without phage did not show mutations in gliding motility genes but had variation in other ORFs. It is therefore possible that extended growth in these conditions introduced other adaptive changes, although these mutations were not present in the phage-exposed samples. Indeed, most variation was detected in the Shieh + M control isolate, which seemed to have had developed resistance against phage even in the phage's absence (Figure 2B). This sample contained non-synonymous mutations in several metabolism related genes (e.g. Lon, rpoB, surE), and in a putative type VI secretion system -like gene (Supplementary File 2). Type VI secretion systems have previously been shown to be associated with host colonization and bacterial antagonism in *Flavobacterium johnsoniae* [42].

Table 2. Bacterial mutations in gliding motility related genes. For a detailed list of all mutations, see supplementary file 2.

Treatment (replicate)	Week	Morphotype	Gliding motility mutation	Mutation type
LW (b)	4	Rough	sprA	Stop codon
LW (b)	4	Rough	sprA	Stop codon
LW (b)	8	Rough	sprA	Stop codon
LW (c)	15	Rough	gldM	Stop codon
LW+M (a)	15	Rough	gldK	SNP nonsynonymous
LW+M (b)	12	Rhizoid	gldK (intergenic)	Intergenic poly-AT
LW+M (b)	15	Rhizoid	sprA	Single nt deletion -> frameshift middle of gene
LW+M (c)	8	Rhizoid	*none*	
Shieh (a)	12	Rough	sprA, SusC/RagA TonB-linked outer membrane protein	sprA: poly-A 8->7. SusC: 48bp deletion
Shieh (b)	8	Rough	*none*	

Shieh (c)	10	Rhizoid	*none*	
Shieh+M (a)	12	Rhizoid	ompA	SNP nonsynonymous
Shieh+M (a)	8	Rough	ompA, sprE	ompA: (CCATCA)repeat 3->2. sprE: poly-A 7->8
Shieh+M (c)	8	Rough	GldN, gliding motility related gene (LOCUS_17550)	GldN: 1bp deletion -> frameshift. LOCUS 17550: 42bp deletion
LW (ctrl)	3	Rhizoid	*none*	
LW+M (ctrl)	15	Rough	*none*	
Shieh (ctrl)	12	Rough	*none*	
Shieh+M (ctrl)	15	Rhizoid	Putative T6SS -related gene	

Discussion

Environmentally transmitted bacteria maintain long periods in the outside-host environment, which may select for reduced growth and metabolic rate [43]. Chemical signals of host presence affects bacterial metabolism, induces chemotaxis and regulates genes needed for invasion and virulence [32]. However, the ability to colonize hosts is often associated with trade-offs in phage resistance [18]. In nature, the co-occurrence of host signals and phages is common, as the points of entry on metazoan surfaces are often covered with mucus where phages are concentrated. This sets up a symbiotic relationship with the eukaryote where phages provide protection, and exploit the altered metabolism of the bacterium for replication [9]. How this arrangement affects relative investment into different phage defense mechanisms is poorly understood.

Here, we investigated this question using the opportunistic pathogen *F. columnare* that causes mucosal disease in freshwater fish [28] but is known to be resilient and able to withstand adverse conditions outside the host [43,44]. Previous studies have shown that mucin, the glycoprotein component of mucus, upregulates virulence factors in *F. columnare* [32] but as a trade-off, make the bacteria more susceptible to phage infection [3]. Our study suggests that the presence of mucin has a central role in triggering CRISPR-Cas immunity and in buffering bacterial survival.

The lake water conditions with and without mucin (LW+M, LW) used in this study are the closest approximations of natural conditions for *F. columnare*. The starved bacterial population (LW) relied almost completely on extracellular immunity via surface modifications, leading to steady decline in the bacterial population density over the 16-week experiment. As the control isolate went extinct in the LW treatment but phage-exposed replicates did not, it is possible that phages have a positive effect for the long term survival of the bacteria population during starvation, perhaps through resistant cells feeding on lysed ones or through the Cas13-induced dormancy by VI-B CRISPR-Cas locus [45] (however, this result may also be fluctuation in our experiment due to having just one control culture). Presence of mucin (LW+M), on the other hand, triggered the intracellular CRISPR-Cas systems (as measured by spacer acquisition), and also supported higher bacterial population densities. Presence of mucin also led to more dramatic decline in phage titers, possibly through active removal of phages from the environment using the CRISPR-Cas systems [26]. In rich medium (Shieh), mucin also accelerated spacer acquisition during the first half of the experiment but with much lesser effect, while medium without mucin did not lead to spacer acquisition at all.

These results suggest that the role of CRISPR-Cas in *F. columnare* may be important specifically during colonization of the metazoan, where the fitness loss associated with surface modification is amplified through reduced virulence. Colony morphotype change leads to loss of virulence [31,46,47] via mutations in the with flavobacterial gliding motility genes, associated with type 9 secretion system

[4]. Most phage-exposed bacterial isolates collected in this experiment had mutations in *gld* or *spr* genes involved in the secretion of adhesins on the cell surface. These mutations also occurred in the LW+M treatment, indicating that CRISPR-Cas is not the sole resistance mechanism even in these settings. However, the phage concentrations in our experimental system was likely higher than on natural fish surfaces, in which CRISPR-Cas mediated defense may be enough to carry the bacterium through the colonization process without compromising virulence through SM.

In general, bacteria isolated from all treatments throughout the 16-week experiment showed increased resistance to phage. Resistance was especially associated with the Rough colony morphotype. Surprisingly, the experimental conditions had an impact on the benefits of acquired CRISPR spacers. Expanded CRISPR arrays had a positive effect on bacterial growth in Rhizoid colonies (but negative effects on Rough colonies) in isolates originating from the LW+M treatment (Figure 6A). However, in isolates from Shieh+M treatment, the result was opposite: Rough isolates had higher population density than Rhizoid with and without new spacers. Why specifically rhizoid LW+M samples gained the most benefit from new spacers is unclear but may reflect an altered bacterial metabolic state induced by starvation and host-signals. This may have important real-life implications for the phage-bacterium interactions in metazoan hosts, which are not observed in standard laboratory culture. Together, these results show that the benefit from additional spacers is affected by the previous environment of the isolates, and that spacers may even have a negative effect depending on simultaneous surface modification. It is also worth noting that the growth experiment was performed in the nutrient-rich Shieh medium, which may downplay the activity of CRISPR-Cas as shown in the 16-week experiment.

Despite the selection pressure by surface modifications and CRISPR-Cas defense, phage populations showed remarkably high genetic stability through the entire experiment, as mutated phage genomes were found only in two replicates. In both cases, genetic change was observed in structural genes, indicating selection imposed by surface modification-based resistance, as seen also previously [48]. Both SM and CRISPR-driven changes in phage genomes have been observed also in the environment [29], but despite the high prevalence of new CRISPR spacers in the LW+M treatment, no phage mutations were found in this treatment. We also did not find mutations in the pre-existing B245 spacer targets in any treatment. The lack of CRISPR evasion points to either inefficient interference, or an efficient degradation of phages with no leakage of mutated phages. The evolutionary potential of these phages may also be limited, as suggested by slow emergence of mutations in natural samples over several years in this phage-bacterium system [29].

Our results support the view that phage defense strategies are influenced by ecological determinants, similar to previous studies showing the effects of nutrient and phage concentrations or the presence of competing species [2,6,25]. These results are also the first description of a bacterial evasion mechanism against non-host derived immunity provided by phages to metazoans, as described by the BAM model. In our study system, the bacterium increased intracellular defenses in the presence of metazoan host-signals, which may be an adaptive maneuver to maximize colonization. The necessity to avoid surface modification in mucosal settings may select for diversification of immune mechanisms, and partly explain why some CRISPR-Cas systems are enriched in pathogens [20]. It will be interesting to see if these results are generalizable to other phage-bacterium-metazoan systems. Furthermore, the roles of other intracellular systems besides CRISPR-Cas, such as restriction modification, should be investigated. From a practical viewpoint, understanding the interplay of bacterial virulence and phage defense during bacterial colonization is crucial for the development of phage therapy, which specifically functions in this three-partite scenario.

References

1. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. 2020 The arms race between bacteria and their phage foes. *Nature* **577**, 1–10. (doi:10.1038/s41586-019-1894-8)

2. Hamilton, R., Siva-Jothy, M. & Boots, M. 2008 Two arms are better than one: parasite variation leads to combined inducible and constitutive innate immune responses. *Proc. Biol. Sci.* **275**, 937–945. (doi:10.1073/pnas.0305096101)

3. Almeida, G. M. F., Laanto, E., Ashrafi, R. & Sundberg, L.-R. 2019 Bacteriophage Adherence to Mucus Mediates Preventive Protection against Pathogenic Bacteria. *mBio* **10**, 1782–12. (doi:10.1128/mBio.01984-19)

4. Li, N. et al. 2017 The type IX secretion system is required for virulence of the fish pathogen Flavobacterium columnare. *Appl. Environ. Microbiol.* **83**. (doi:10.1128/AEM.01769-17)

5. van Houte, S., Buckling, A. & Westra, E. R. 2016 Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiol. Mol. Biol. Rev.* **80**, 745–763. (doi:10.1128/MMBR.00011-16)

6. Westra, E. R. et al. 2015 Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense. *Curr. Biol.* **25**, 1043–1049. (doi:10.1016/j.cub.2015.01.065)

7. Schmid-Hempel, P. 2009 Immune defence, parasite evasion strategies and their relevance for 'macroscopic phenomena' such as virulence. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **364**, 85–98. (doi:10.1098/rstb.2008.0157)

8. Chatterjee, A. & Duerkop, B. A. 2018 Beyond Bacteria: Bacteriophage-Eukaryotic Host Interactions Reveal Emerging Paradigms of Health and Disease. *Front Microbiol* **9**, 1394. (doi:10.3389/fmicb.2018.01394)

9. Barr, J. J. et al. 2013 Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proceedings of the National Academy of Sciences* **110**, 10771–10776. (doi:10.1073/pnas.1305923110)

10. Barr, J. J., Youle, M. & Rohwer, F. 2013 Innate and acquired bacteriophage-mediated immunity. *Bacteriophage* **3**, e25857. (doi:10.4161/bact.25857)

11. 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. (doi:10.1038/s41564-017-0051-0)

12. 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* **3**, eaar4120–18. (doi:10.1126/science.aar4120)

13. Brouns, S. J. J. et al. 2008 Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964. (doi:10.1126/science.1159689)

14. Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M. & Terns, M. P. 2009 RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* **139**, 945–956. (doi:10.1016/j.cell.2009.07.040)

15. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. 2012 Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* **109**, E2579–86. (doi:10.1073/pnas.1208507109)

16. Mangalea, M. R. & Duerkop, B. A. 2020 Fitness trade-offs resulting from bacteriophage resistance potentiate synergistic antibacterial strategies. *Infect. Immun.* (doi:10.1128/IAI.00926-19)

17. Koskella, B. & Brockhurst, M. A. 2014 Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews* **38**, 916–931. (doi:10.1111/1574-6976.12072)

18. León, M. & Bastías, R. 2015 Virulence reduction in bacteriophage resistant bacteria. *Front Microbiol* **6**, 343. (doi:10.3389/fmicb.2015.00343)

19. Grissa, I., Vergnaud, G. & Pourcel, C. 2007 The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**, 172. (doi:10.1186/1471-2105-8-172)

20. Weissman, J. L., Laljani, R. M. R., Fagan, W. F. & Johnson, P. L. F. 2019 Visualization and prediction of CRISPR incidence in microbial trait-space to identify drivers of antiviral immune strategy. *The ISME Journal* **13**, 2589–2602. (doi:10.1038/s41396-019-0411-2)

21. Louwen, R., Staals, R. H. J., Endtz, H. P., van Baarlen, P. & van der Oost, J. 2014 The Role of CRISPR-Cas Systems in Virulence of Pathogenic Bacteria. *Microbiol. Mol. Biol. Rev.* **78**, 74–88. (doi:10.1128/MMBR.00039-13)

22. Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y.-L. & Weiss, D. S. 2013 A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* **497**, 254–257. (doi:10.1038/nature12048)

23. Louwen, R. et al. 2012 A novel link between Campylobacter jejuni bacteriophage defence, virulence and Guillain–Barré syndrome. *Eur J Clin Microbiol Infect Dis* **32**, 207–226. (doi:10.1007/s10096-012-1733-4)

24. Saha, C. et al. 2020 Guide-free Cas9 from pathogenic Campylobacter jejuni bacteria causes severe damage to DNA. *Sci Adv* **6**, eaaz4849.

25. 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. (doi:10.1038/s41586-019-1662-9)

26. Payne, P., Geyrhofer, L., Barton, N. H. & Bollback, J. P. 2018 CRISPR-based herd immunity can limit phage epidemics in bacterial populations. *Elife* **7**, e32035. (doi:10.7554/eLife.32035)

27. Cone, R. A. 2009 Barrier properties of mucus. *Advanced Drug Delivery Reviews* **61**, 75–85. (doi:10.1016/j.addr.2008.09.008)

28. Declercq, A. M., Haesebrouck, F., Van den Broeck, W., Bossier, P. & Decostere, A. 2013 Columnaris disease in fish: a review with emphasis on bacterium-host interactions. *Vet. Res.* **44**, 27–17. (doi:10.1186/1297-9716-44-27)

29. Laanto, E., Hoikkala, V., Ravantti, J. & Sundberg, L.-R. 2017 Long-term genomic coevolution of host-parasite interaction in the natural environment. *Nature Communications* **8**, 111. (doi:10.1038/s41467-017-00158-7)

30. Hoikkala, V., Ravantti, J., Díez-Villaseñor, C., Tiirola, M., Conrad, R. A., McBride, M. J. & Sundberg, L.-R. 2020 Cooperation between CRISPR-Cas types enables adaptation in an RNA-targeting system. *bioRxiv* **18**, 2020.02.20.957498. (doi:10.1101/2020.02.20.957498)

31. Laanto, E., Bamford, J. K. H., Laakso, J. & Sundberg, L.-R. 2012 Phage-Driven Loss of Virulence in a Fish Pathogenic Bacterium. *PLoS ONE* **7**, e53157–8. (doi:10.1371/journal.pone.0053157)

32. Lange, M. D., Farmer, B. D. & Abernathy, J. 2018 Catfish mucus alters the Flavobacterium columnare transcriptome. *FEMS Microbiology Letters* **365**. (doi:10.1093/femsle/fny244)

33. Laanto, E., Sundberg, L. R. & Bamford, J. K. H. 2011 Phage Specificity of the Freshwater Fish Pathogen Flavobacterium columnare. *Appl. Environ. Microbiol.* **77**, 7868–7872. (doi:10.1128/AEM.05574-11)

34. Shieh, H. S. 1980 Studies on the nutrition of a fish pathogen, Flexibacter columnaris. *Microbios Letters* **13**, 129–133.

35. Bankevich, A. et al. 2012 SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477. (doi:10.1089/cmb.2012.0021)

36. Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B. & Schatz, M. C. 2019 RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol* **20**, 224–17. (doi:10.1186/s13059-019-1829-6)

37. Tanizawa, Y., Fujisawa, T. & Nakamura, Y. 2018 DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* **34**, 1037–1039. (doi:10.1093/bioinformatics/btx713)

38. Deatherage, D. E. & Barrick, J. E. 2014 Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. In *Methods in Molecular Biology*, pp. 165–188.(doi:10.1007/978-1-4939-0554-6_12)

39. Zimmermann, L. et al. 2018 A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology* **430**, 2237–2243. (doi:10.1016/j.jmb.2017.12.007)

40. Akaike, H. 1998 Information Theory and an Extension of the Maximum Likelihood Principle. In *Selected Papers of Hirotugu Akaike* (eds E. Parzen K. Tanabe & G. Kitagawa), pp. 199–213. New York, NY: Springer New York.

41. Brooks, M., Kristensen, K., van Benthem, K., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H., Mächler, M. & Bolker, B. 2017 glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal* **9**, 378–400.

42. Russell, A. B. et al. 2014 A type VI secretion-related pathway in Bacteroidetes mediates interbacterial antagonism. *Cell Host and Microbe* **16**, 227–236. (doi:10.1016/j.chom.2014.07.007)

576 43. Sundberg, L.-R., Kunttu, H. M. T. & Valtonen, E. 2014 Starvation can diversify the population
577 structure and virulence strategies of an environmentally transmitting fish pathogen. *BMC*
578 *Microbiology* **14**, 67. (doi:10.1186/1471-2180-14-67)

579 44. Kunttu, H. M. T., Sundberg, L.-R., Pulkkinen, K. & Valtonen, E. T. 2012 Environment may be
580 the source of Flavobacterium columnare outbreaks at fish farms. *Environ Microbiol Rep* **4**,
581 398–402. (doi:10.1111/j.1758-2229.2012.00342.x)

582 45. Meeske, A. J., Nakandakari-Higa, S. & Marraffini, L. A. 2019 Cas13-induced cellular dormancy
583 prevents the rise of CRISPR-resistant bacteriophage. *Nature* **570**, 241–245.
584 (doi:10.1038/s41586-019-1257-5)

585 46. Kunttu, H. M. T., Suomalainen, L.-R., Jokinen, E. I. & Valtonen, E. T. 2009 Flavobacterium
586 columnare colony types: Connection to adhesion and virulence? *Microb. Pathog.* **46**, 21–27.
587 (doi:10.1016/j.micpath.2008.10.001)

588 47. Zhang, J., Laakso, J., Mappes, J., Laanto, E., Ketola, T., Bamford, J. K. H., Kunttu, H. &
589 Sundberg, L.-R. 2014 Association of colony morphotypes with virulence, growth and
590 resistance against protozoan predation in the fish pathogen Flavobacterium columnare. *FEMS*
591 *Microbiol Ecol* **89**, 553–562. (doi:10.1111/1574-6941.12356)

592 48. Laanto, E., Mäkelä, K., Hoikkala, V., Ravantti, J. J. & Sundberg, L.-R. 2020 Adapting a Phage to
593 Combat Phage Resistance. *Antibiotics* **9**. (doi:10.3390/antibiotics9060291)

594