

RESEARCH ARTICLE

Open Access



# Pan genome and CRISPR analyses of the bacterial fish pathogen *Moritella viscosa*

Christian Karlsen<sup>1,2\*</sup> , Erik Hjerde<sup>3</sup>, Terje Klemetsen<sup>3</sup> and Nils Peder Willassen<sup>3,4</sup>

## Abstract

**Background:** Winter-ulcer *Moritella viscosa* infections continue to be a significant burden in Atlantic salmon (*Salmo salar* L.) farming. *M. viscosa* comprises two main clusters that differ in genetic variation and phenotypes including virulence. Horizontal gene transfer through acquisition and loss of mobile genetic elements (MGEs) is a major driving force of bacterial diversification. To gain insight into genomic traits that could affect sublineage evolution within this bacterium we examined the genome sequences of twelve *M. viscosa* strains. Matches between *M. viscosa* clustered, regularly interspaced, short palindromic, repeats and associated *cas* genes (CRISPR-Cas) were analysed to correlate CRISPR-Cas with adaptive immunity against MGEs.

**Results:** The comparative genomic analysis of *M. viscosa* isolates from across the North Atlantic region and from different fish species support delineation of *M. viscosa* into four phylogenetic lineages. The results showed that *M. viscosa* carries two distinct variants of the CRISPR-Cas subtype I-F systems and that CRISPR features follow the phylogenetic lineages. A subset of the spacer content match prophage and plasmid genes dispersed among the *M. viscosa* strains. Further analysis revealed that prophage and plasmid-like element distribution were reflected in the content of the CRISPR-spacer profiles.

**Conclusions:** Our data suggests that CRISPR-Cas mediated interactions with MGEs impact genome properties among *M. viscosa*, and that patterns in spacer and MGE distributions are linked to strain relationships.

**Keywords:** *Moritella viscosa*, CRISPR-Cas, Mobile genetic element, Atlantic salmon pathogen

## Background

The genus *Moritella* comprises seven psychrophilic species associated with deep seawater and ocean sediments. *Moritella viscosa* is the only species so far associated with fish pathogenicity, being the causative agent of winter-ulcer disease in farmed salmonids [1, 2]. Outbreaks occur in salmonid aquaculture across the North Atlantic [3–7] and infected fish develop chronic skin ulcers that may be followed by terminal septicaemia [3, 6]. Two major phenotypic and genotypic clades ('typical' and 'variant') have been identified in *M. viscosa* [5]. It is suggested that phylogenetic lineages within *M. viscosa* have evolved compatibility factors that adapt typical *M. viscosa* to host-specific virulence [8].

Phenotypic and genotypic variations may originate from horizontal gene transfer (HGT) that introduces new elements through mechanisms such as conjugation, transformation and transduction through bacteriophage-mediated DNA or plasmid transfer [9]. Acquisition or loss of mobile genetic elements (MGEs) could alter virulence properties, e.g. by introducing a novel toxin or surface alteration in a strain [9]. Bacteriophage might also present a danger to the host bacteria as bacteriophages can cause bacteriolysis [10]. Temperate bacteriophages have, unlike virulent phages, the ability to integrate their DNA into the bacterium's chromosome where it enters a prophage dormant state replicating along with the host genome.

In response, bacteria have mechanisms to resist infection of MGEs. One is the clustered, regularly interspaced, short palindromic repeats (CRISPRs) flanked by CRISPR-associated (*cas*) genes. The CRISPR-Cas system is used in most archaea [11] and are widespread across diverse bacteria [12, 13] including the phylum

\* Correspondence: christian.karlsen@nofima.no

<sup>1</sup>Department of Food Safety and Infection Biology, Norwegian University of Life Sciences (NMBU), Pb 8146 Dep.0033 Oslo, Norway

<sup>2</sup>Present address: Nofima AS, Division of Aquaculture, PO Box 210, Ås N-1431, Norway

Full list of author information is available at the end of the article



Cyanobacteria [14]. The system can act against invading foreign viruses and plasmids by targeting DNA in a sequence specific manner [15]. CRISPRs consist of short (23–47 bp) highly conserved repeats separated by variable sequences called spacers. Spacers are acquired mostly independently from foreign DNA, and only a smaller subset is transmitted vertically [15]. The Cas proteins are involved in this defence mechanism, both processing, binding and targeting of foreign DNA, and integrating novel spacer units into the CRISPR locus [15]. The complementary sequence to spacers that originate from invading genetic elements are termed protospacers. Spacers incorporated into the CRISPR loci are transcribed acting as guides that anneal to the complementary protospacers of the invading genetic element. The CRISPR-Cas mechanism will then degrade the foreign nucleic acids. The invader can in turn evade this resistance by modifying the targeted DNA sequence generating CRISPR escape mutations [16]. Thus, CRISPRs are considered to be a form of acquired immunity from past infections which may provide insights into bacterial niche adaptation, evolution and phage-host dynamics that have occurred within the bacterial populations [17]. CRISPR is rapidly evolving in the genomes of some microbial pathogens and can be used to detect and genotype clinical isolates of *Mycobacterium tuberculosis* [18], *Corynebacterium diphtheria* [19] and *Salmonella enterica* subs. *enterica* [20]. However, CRISPR distribution may not always correlate to phylogenetic relationships, as independent evolution in select lineages can advance in part by HGT and environmental differences in phage predation [13].

In this study, a bioinformatics approach has been used to resolve genomic diversity between twelve *M. viscosa* isolated from different geographical locations and fish species. We analysed the CRISPR-Cas systems and the CRISPR locus organization to determine relatedness to strain origin. All *M. viscosa* spacers were then examined to establish spacer diversity and to identify the protospacers of targeted genes. In order to examine the potential function of the CRISPR-Cas system in *M. viscosa* all spacers were searched against the twelve *M. viscosa* genomes, and examined by relating the results obtained to MGE distribution in the corresponding strains. Our analyses suggest that the CRISPR-Cas system in *M. viscosa* is an important determinant of genetic transfer involved in prophage and plasmid distribution influencing the evolution of this fish pathogenic species.

## Methods

### Bacterial strains and DNA extraction

The 12 *M. viscosa* strains analysed here include representatives isolated from different fish species that span

the geographical area of occurring outbreaks of winter-ulcer disease across the North Atlantic region (Additional file 1: Table S1). The isolates include both typical and variant *M. viscosa*, which were categorized as per standard biochemical and phenotypic methods as well as sequence analysis [2, 5, 8, 21]. The complete genome of the virulent *M. viscosa* MV 0609139 [22, 23] was used as reference. Strains were cultured in Luria-Bertani broth containing 3.5% NaCl at 12 °C. DNA was extracted using the Qiagen DNeasy blood and tissue kit protocol for Gram-negative bacteria.

### Genome sequencing, assembly and annotation

Sequencing libraries for the bacterial isolates were made using the Nextera XT kit according to the manufacturer's protocol, and the fragment size distribution analysed to be 500–1000 bp using the Agilent 2100 Bioanalyzer System. The sample libraries were multiplexed and sequenced in a single run on a MiSeq machine (Illumina) using v3 reagents with 2 × 150 cycles according to the manufacturer's instructions. This yielded an average of 2.06 million reads per bacterial isolate. The twelve genomes were assembled *de novo* using CLC Genomics Workbench v6.5 (<https://www.qiagen-bioinformatics.com/>) with default parameters, not performing scaffolding and with 500 bases as minimum cutoff length for each contig. The resulting contigs were mapped against our reference genome using standard Nucmer settings with ABACAS v1.3.1 [24]. Unmapped contigs were included by appending them to the output fasta-file with the mapped contigs. This was followed by concatenation using the six-frame stop-codon "CTAGC-TAGCTAG" as separators between contigs. Glimmer v3.02 [25] was then used to identify possible protein coding genes (CDSs) on the concatenated sequences before subsequent annotation by basic local alignment search tool (BLAST), using protein-protein BLASTp (UniProt database release 01 2014) [26, 27], HMMER3 v3.1b1 (hmmscan applying Pfam database v27.0) [28, 29] and SignalP v4.0 [30]. Genome sequences are available from European Nucleotide Archive (ENA) through the study accession number PRJEB1601. Accession number for each genome is listed in Additional file 2: Table S2A.

### Orthologue identification

Clustering of orthologous genes was done by OrthoMCL v1.4 [31], with the input consisting of 12 multifasta-files containing the predicted CDSs from each sequenced strains. The parameters were set at 90 percent identity cutoff and 20 percent match cutoff for the clustering algorithm. BLAST p-value cutoff, max weight and MCL inflation were set to default.

### Pan genome analysis

A pan genome of all 12 strains was identified using the 4720 clusters determined by OrthoMCL. This was achieved by extracting each cluster separately before creating a precursory consensus sequence from each cluster using the script Consensus.pl available on Github (<https://github.com/josephhughes/Sequence-manipulation>). All consensus sequences were then amassed in a single multifasta-file in the same order as the orthoMCL output while appending the 967 unclustered (unique) genes. For the sake of clarity, cluster information, consensus sequence lengths and annotation were additionally handled in an excel spreadsheet to sort the number of genes in each cluster, from highest to lowest with the associated strains. Gene clusters present in all strains were defined as being part of the 'core' genome. Gene clusters present in all strains containing additional paralogs were defined as 'core plus,' while clusters not represented by all strains were part of the 'accessory' genome. Genes only present in single strains were defined as 'unique'. The ordered data was used to generate a pan genome diagram using Circos [32].

### Gene ontology

Annotation of Gene Ontology (GO) [33] was also performed on the predicted CDSs using InterProScan [34]. The resulting outputs were counted using the web tool WEGO [35], where GO data from the four uppermost levels of the ontologies were collected for each strain and compared in a line plot.

### Whole genome phylogenetics

Single-nucleotide polymorphisms (SNPs) were identified and a Maximum likelihood tree reconstructing the phylogenetic relationship between the isolates was performed on the core genome using the alignment free software kSNP [36]. A gene content tree was constructed from a binary pan genome cluster matrix (presence or absence of genes in each isolate relative to the other isolates) generated with GET\_HOMOLOGUES [37] using the discrete character parsimony algorithm. The tree comparison was performed with EPoS [38] with ten tanglegram computations.

### Prophage prediction

Prophages in *M. viscosa* genomes were identified using the Phage Search Tool (PHAST) webserver [39]. We further checked whether the *M. viscosa* phylogeny was linked to presence of certain prophages.

### CRISPR-Cas analysis and protospacer identification

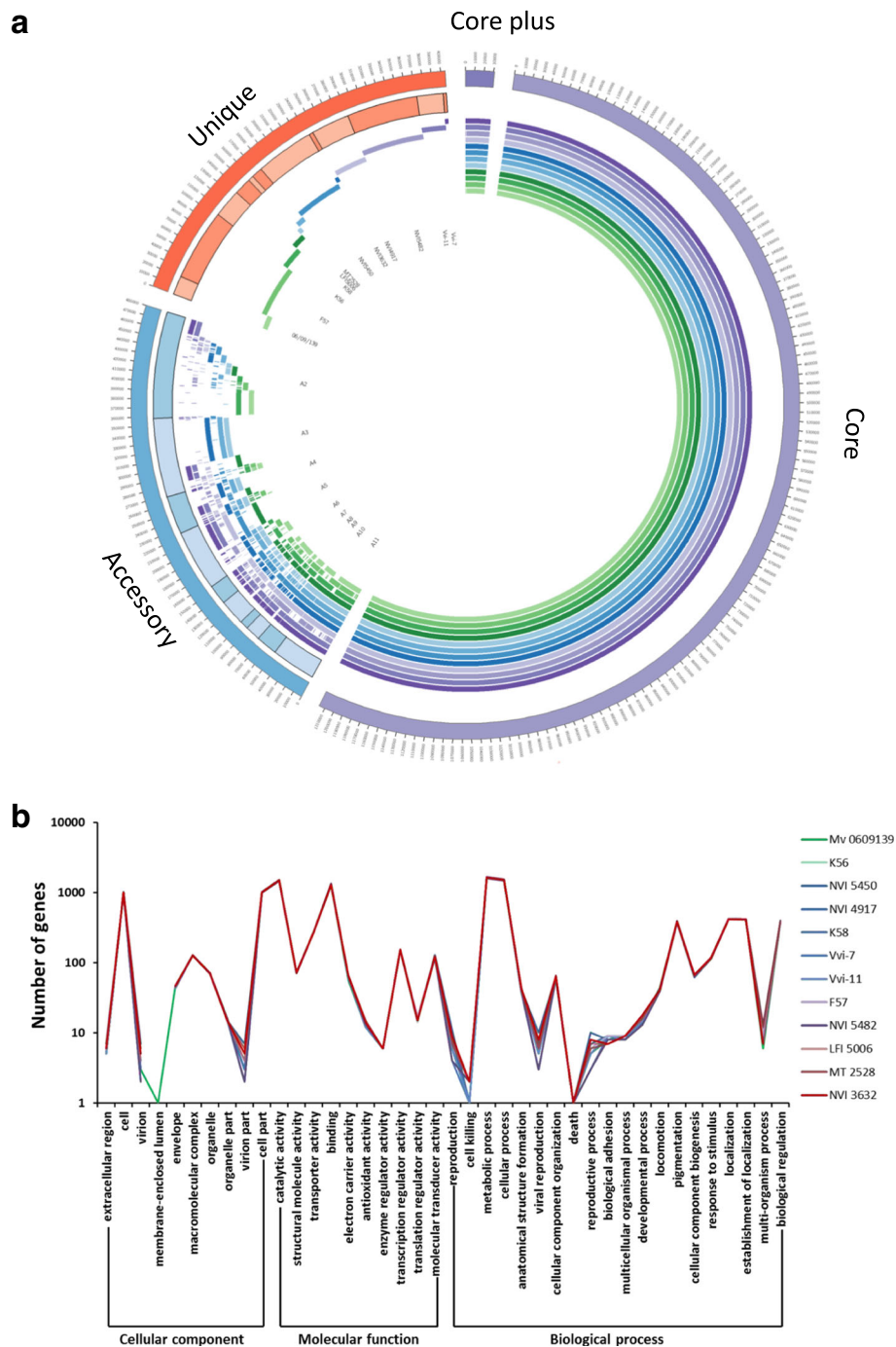
The orthologue analysis identified CRISPR related Cas genes in variant *M. viscosa*, and the genomes of all *M. viscosa* were searched for CRISPR arrays using

CRISPRfinder [40] and by BLAST searches of the identified *cas* genes in variant *M. viscosa* against a local database generated from the CDSs of all *M. viscosa* genomes in BioEdit [41]. Cas gene sequences and the deduced amino acid sequences from these genes within *M. viscosa* CRISPR type I and CRISPR type II were aligned using ClustalW. To examine the potential significance of the CRISPR-Cas system in *M. viscosa*, all *M. viscosa* spacers (Additional file 3) were searched against CRISPRTarget [42] to identify possible protospacers. A match against the GenBank-Phage or RefSeq-Plasmid databases was counted when a spacer had  $\leq 4$  SNPs over the length of 32 nucleotides. A relative measure of relatedness was calculated from BLASTn results generated from pairwise comparison of each spacers to all *M. viscosa* spacers. Spacers from one strain that matched to the spacer-array of another *M. viscosa* were defined as  $\leq 1$  SNP (31/32 nucleotides). All *M. viscosa* spacers were further utilized in BLASTn searches against the CDSs of all *M. viscosa* to identify possible protospacers or targeted genes within *M. viscosa*. The investigated *M. viscosa* strains were found to carry a range of different MGEs and the detection of protospacers were further related to the MGE distribution in the corresponding strains. The putative uncharacterized protein encoded by K56\_4594 and MT2528\_4809 in plasmid B was analysed further by utilizing the Phyre2 web portal for protein modelling, prediction and analysis [43].

## Results

### General features and comparisons and the core genome of *M. viscosa*

The comparative genome content of twelve *M. viscosa* is shown in Fig. 1. The completeness of the draft genomes were assessed by mapping onto the complete reference genome of *M. viscosa* MV 0609139 [22]. Percentage of bases mapped to the reference genome range from 61.7 to 94.8% with an average of 84.0%. Genome sizes and number of predicted genes ranged from 4.96 to 5.3 Mbp and 4532 to 4924, respectively. General genomic and sequence statistics and the numbers of CDSs shared between or being unique to *M. viscosa* strains are shown in Additional file 2: Table S2A-D. The average number of genes was 4718, with 3737 core genes found in all strains. Orthologue analysis (Fig. 1) revealed that strains share between 465 and 1028 dispensable (accessory) genes and that number of strain specific (unique) genes (in total 1888) varied between 22 to 362 genes in each strain. Grouping all functional genes from the twelve *M. viscosa* genomes identified 5589 pan genomic gene clusters. Comparing the core genes to the pan genome cluster showed that the core genome accounts for 67% of the pan genome.



**Fig. 1** Comparative genome content of twelve *Moritella viscosa*. **a** The outermost circle indicates the classification into core plus, core, accessory and unique genes in the pan genome of *M. viscosa*. Internal circles indicate gene presence (solid colour) or absence (unfilled) of each gene in each of the 12 strains examined. Genes are represented only once in the diagram, but the gene order in the different core, accessory and unique segments are discontinuous, since genes may be represented in different segments. Circles from outer to inner are representing strains in the following order. Purple; Vvi-7, Vvi-11, NVI 5482, NVI 4917. Blue; NVI 3632, NVI 5450, MT 2528, LFI 5006. Green; K58, K56, F57, MV 0609139. **b** Gene ontology (GO) term category distribution. Functional classification of genes with GO terms encoded by *M. viscosa* genomes is displayed for each *M. viscosa* isolate as a line chart plotted against a logarithmic scale. The number of genes represents the amount annotated into the corresponding term of three GO categories at level 2. The twelve *M. viscosa* isolates are indicated by colour codes

### Functional categories of predicted *M. viscosa* genes

Identified genes were categorized by GO assignments into 40 functional processes within the “cellular component”, “molecular function” and “biological process” categories at level 2 (Fig. 1b). The homology assignments revealed little discrepancy in the distribution of genes within the *M. viscosa* genomes investigated. Refining the categorization further (results not shown) revealed that in the cellular component category the largest numbers of genes grouped into the sub-category membrane or membrane part. For genes within molecular function, the largest sub-categories were nucleic acid binding, transferase and hydrolase activity. In the biological process category, genes sub-grouped into cellular-, primary-, nitrogen-, and biosynthetic-metabolic processes. The sub-categorization of the “biological process” category revealed further that most discrepancies are associated with MGEs such as prophage-associated genes.

### Relationships among the *M. viscosa* genomes

*M. viscosa* can be separated into two major phenotypically and genetically different clusters (typical and variant) by haemolytic activity, which is consistent with Western blot, plasmid profile, pulsed field gel electrophoresis and *gyrB* gene sequence analyses [5]. However, the whole-genome phylogenetic SNP analysis and the gene content tree (Fig. 2) do not separate strains into the present typical/variant classification. Strains LFI 5006, NVI 3632 and MT 2528 from Norwegian and Scottish Atlantic salmon do group into typical *M. viscosa* as previously described [5, 8]. The variant *M. viscosa* are sublineaged into three clades where both clade 2 and clade 3 form a cluster with typical *M. viscosa*. Clade 2 contains isolates from Norwegian (strain MV 0609139) and Icelandic (strain K56) farmed Atlantic salmon. Clade 3 contains isolates from farmed Norwegian cod (strain NVI 5482) and Icelandic lump sucker (strain F57). The more distantly related strains form clade 1, which contains isolates from Canadian (strain Vvi-7 and Vvi-11) and Icelandic (strain K58) farmed Atlantic salmon including Norwegian farmed trout (strain NVI 4917 and NVI 5450). While the phylogenetic tree built from SNPs in the core region of the genomes and hence represents the vertical evolution, the gene content tree counts presence and absence of genes in isolates relative to each other and hence represents the horizontal evolution of the isolates. To test whether the uptake and loss of MGEs was the main driver of the *M. viscosa* evolution, the congruency between SNP phylogeny and the gene content tree was tested. The comparison revealed that the topology of the trees was similar and that the majority of clades are congruent in both trees resulting in a Robinson Fould Distance of 0.30 [44]. This gives further support to the relationships among the divergent *M. viscosa*

lineages. Only NVI 5450 had a different placement. The comparative analysis between typical and variant *M. viscosa* revealed 231 genes shared between typical *M. viscosa* but which were not present in other variant *M. viscosa*. Of the 231 genes, 126 are annotated as putative uncharacterized proteins. A high number of the remaining predicted genes are homologues to predicted genes in other *Moritella* and *Vibrio* spp.

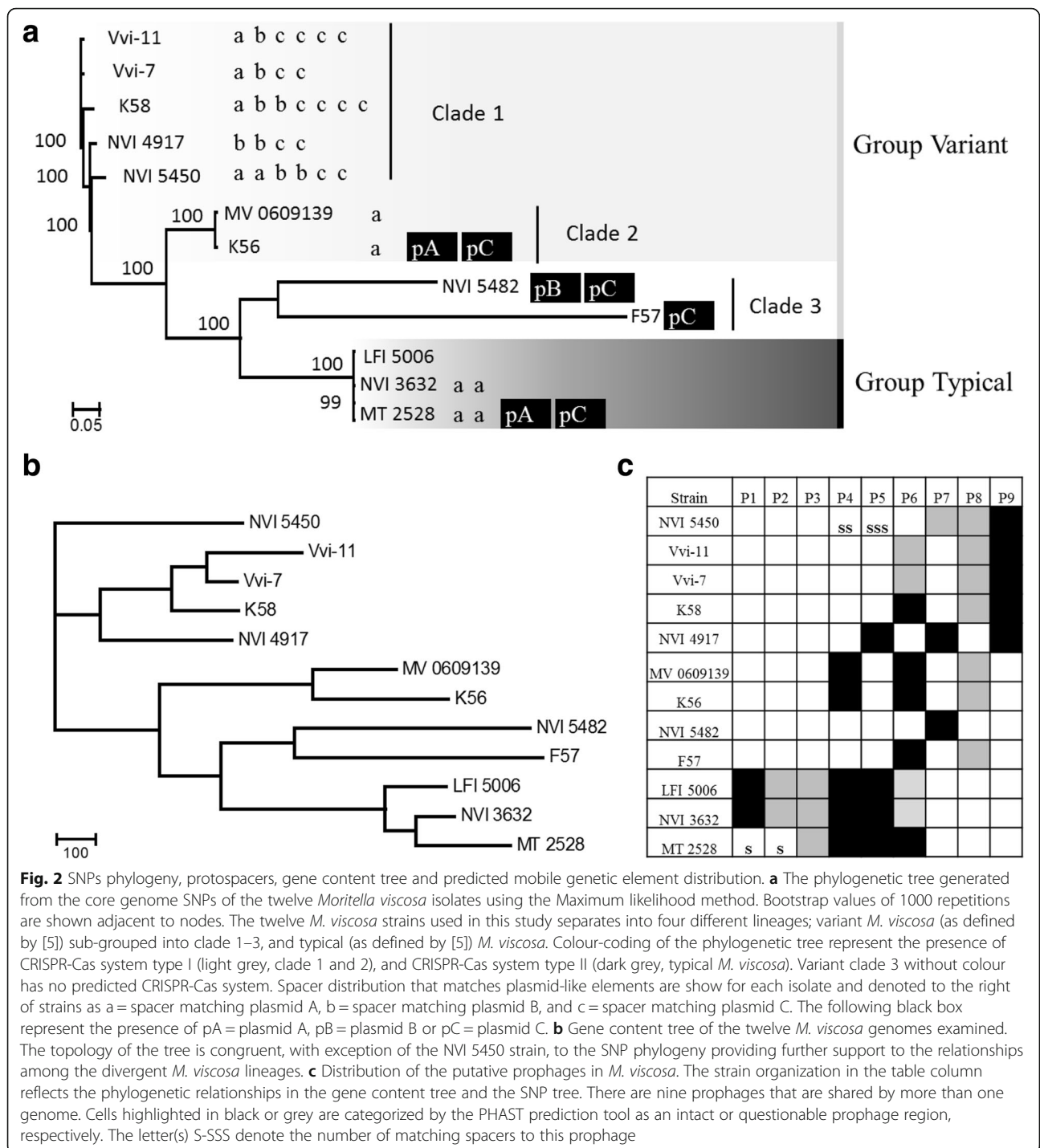
### Plasmid-like elements in *M. viscosa* and their putative encoding genes

From the comparative genomic study, we observed putative plasmid-like elements. We describe here the elements with complimentary sequences to spacers present in the CRISPR loci. One, which is present in MT 2528 (MT2528\_3989 to MT2528\_3955) and K56 (K56\_4570 to K56\_4597) is termed plasmid A. The analysis of plasmid A revealed nine genes encoding homologues to Trb proteins indicative of a P-type conjugation system. Also a putative type II-like secretion system (T2SS) protein, a hypothetical type IV (T4) pilin and a number of uncharacterized proteins were predicted indicative that the cluster encodes a T2SS or T4 pilus like transport system. The top ranking model for K56\_4594 (and equivalent MT2528\_4809) predicted by Phyre2 [43] is the *Vibrio cholerae* VesB protease (PDB template c4lk4A, model not shown). 80% of the sequence (residues 23–317) was modelled with 100.0% confidence with an N-terminal signal peptide and a C-terminal domain similar to an immunoglobulin (Ig) fold with a membrane spanning helix at the C-terminal end.

The putative plasmid B element in NVI 5482 (NVI5482\_4403 to NVI5482\_4431) contains genes encoding homologues to Tra proteins indicative of an F-type conjugation system. Blast searches of amino acid sequences to modules of the plasmid show highest identity to other marine bacteria such as *Aliivibrio salmonicida*, *Shewanella baltica*, *Aeromonas salmonicida* and *Photobacterium* sp.

The plasmid C element in MT 2528, NVI 5482, F57 and K56 (K56\_4540 to K56\_4568) is intriguing and may be remnants from a larger plasmid-like element as annotation reveals hallmarks (results not shown) for linear plasmid-like prophages reported from other Gram-negative marine bacteria [45]. The *repA* adjacent sequence stretches are not similar between Mt 2528, NVI 5482 and F57, and it is possible that the assortment of genes originates from sequence assembly difficulties. Most CDSs are annotated as uncharacterized proteins but several genes encode transposases, integrases, DNA modifying proteins, and phage related proteins.

In addition, all the predicted plasmid-like elements provided Pfam predicted relaxases using the Pfam-web tool [46].



**Prophages in *M. viscosa* genomes**

The PAST predicted prophages were separated into lineages according to their predicted similarity to known prophages and by their conserved synteny of the genomic structure. Predicted prophages (Additional file 4: Figure S1) found in two or more of the twelve sequenced *M. viscosa* genomes are presented as prophage 1–9. Prophage distribution between *M. viscosa* strains was then

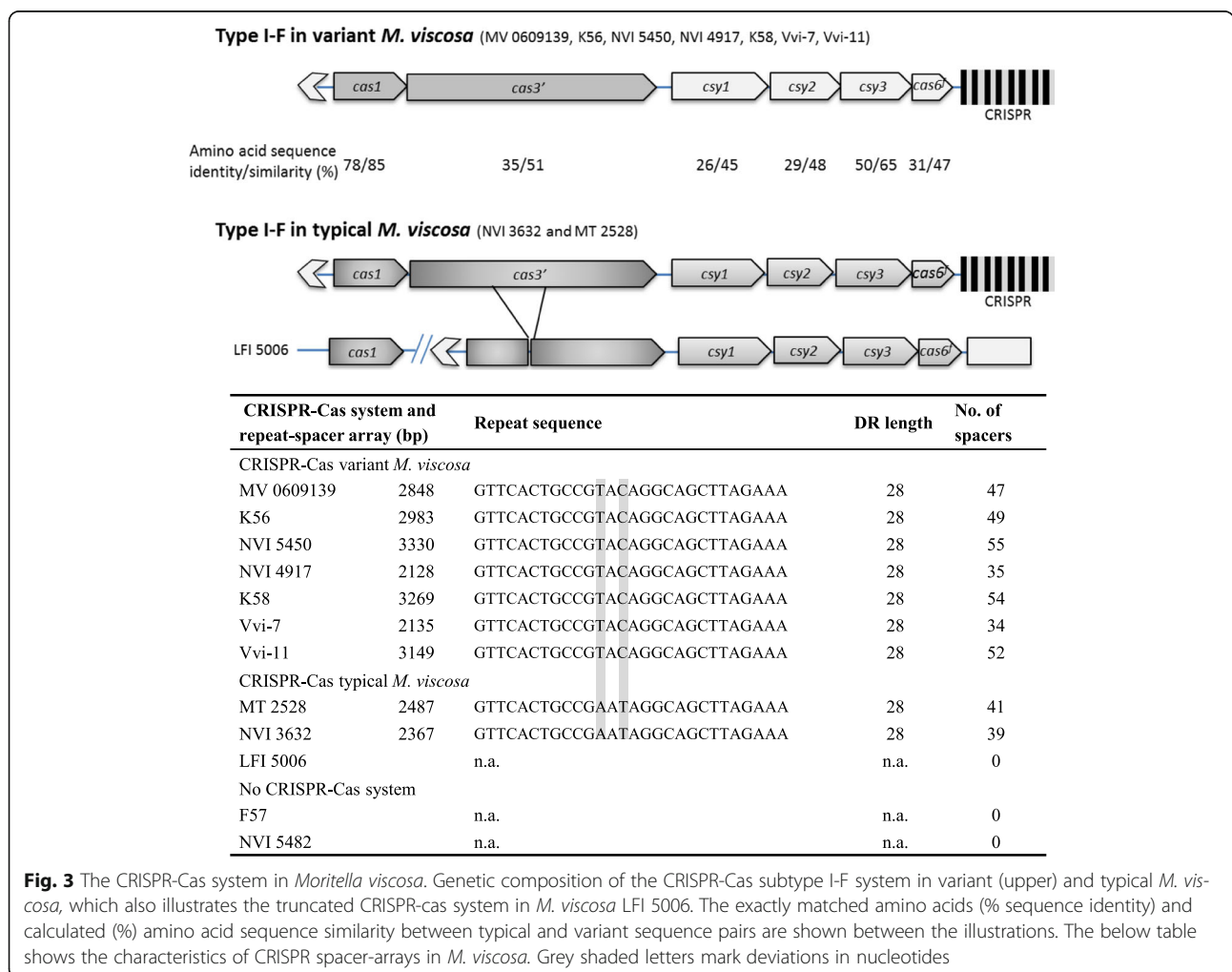
resolved by manually allocating similar structured prophages to one of the nine prophage types as shown in Fig. 2b. The topology of the SNP and gene content trees is congruent, and comparing the prophage presence to the tree topology shows that the distribution of these prophages make patterns that support an evolutionary relatedness in the *M. viscosa* genomes. Only a small number of proteins can be related to known

functions (Additional file 4: Figure S1). Genes for which function can be predicted are putative integrases, terminases and phage-structural proteins. Six of the prophage types contain integrases. A phylogenetic analysis based on the amino acid sequence of these integrases cluster in accordance to the predicted prophages supporting the allocation of these prophages to the correct prophage-type (Additional file 4: Figure S2). In addition to phage protein orthologs, attL and attR sites for site-specific integration into the genome and integrases were detected (Additional file 4: Table S3). The attachment sites are identical to the specific integrases that are phylogenetically related. All of the predicted attachment sites are repeatedly found throughout the genomes (results not shown).

**The CRISPR-Cas system in *M. viscosa***

Two distinct variants of the CRISPR-Cas system with amino acid sequence score alignments ranging between 26-78% identity were identified in *M. viscosa* (Fig. 3). They are divided between the variant *M.*

*viscosa* clade 1 and clade 2, and typical *M. viscosa* (Fig. 2a). Both systems are classified by the system of Makarova et al. 2011 to belong to subtype I-F and include six genes (Fig. 3). Nucleotide alignments of the *cas* and *csy* genes show 100% nucleotide identity between all typical *M. viscosa* isolates harbouring these genes (except the truncated version of *cas3'* in LFI 5006). The CRISPR-Cas genes were also conserved within variant *M. viscosa* (>99.9% identity) with the presence of a single conserved SNP. The *cas* operon encodes Cas1, Cas3', and the subtype specific proteins Csy1, Csy2, Csy3 and Cas6<sup>f</sup> (formerly Csy4) followed by a repeat-spacer array with the number of spacer per strain ranged from 0 to 55 (Fig. 3). However, LFI 5006 possesses a truncated *cas3'* in addition to a dispersed *cas1* gene. These genes are required for integrating new spacer sequences [47], and could explain the lack of a predicted repeat-spacer array in this strain. The partly palindromic repeat sequences differ by two nucleotide substitutions between typical and variant *M. viscosa* CRISPR-arrays (Fig. 3). The closest

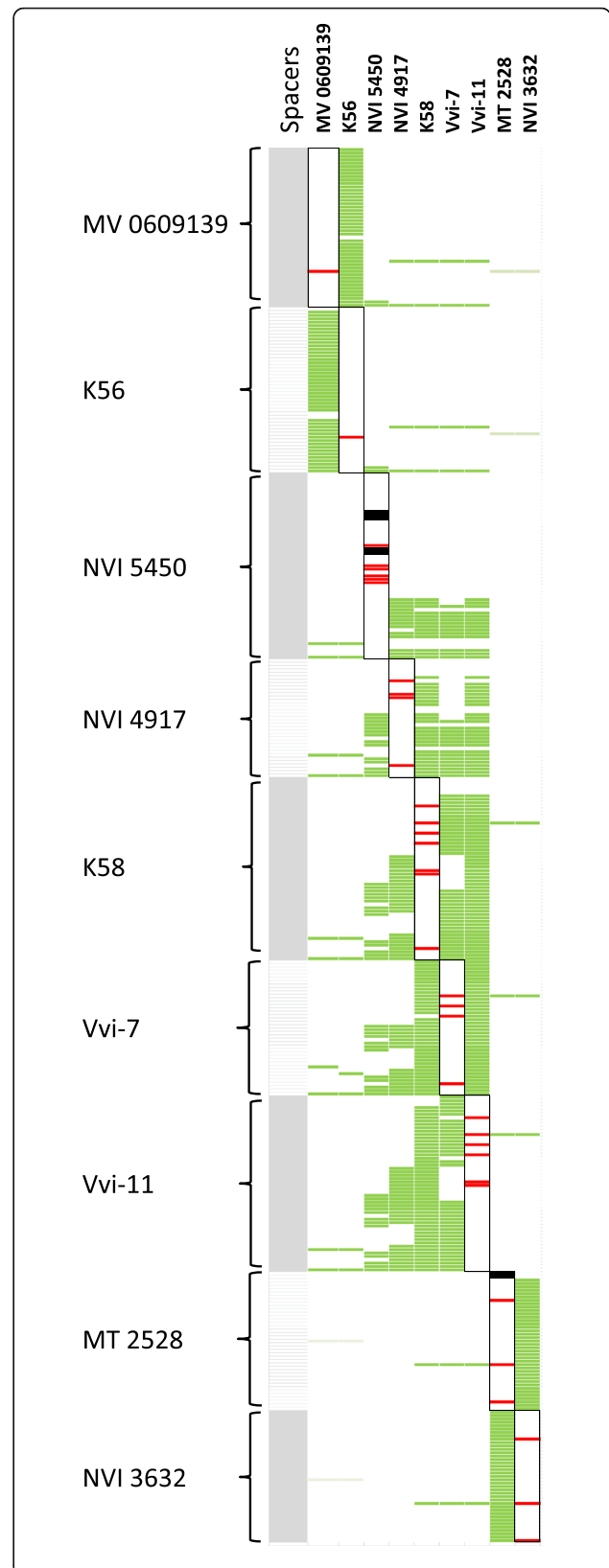


**Fig. 3** The CRISPR-Cas system in *Moritella viscosa*. Genetic composition of the CRISPR-Cas subtype I-F system in variant (upper) and typical *M. viscosa*, which also illustrates the truncated CRISPR-cas system in *M. viscosa* LFI 5006. The exactly matched amino acids (% sequence identity) and calculated (% amino acid sequence similarity) between typical and variant sequence pairs are shown between the illustrations. The below table shows the characteristics of CRISPR spacer-arrays in *M. viscosa*. Grey shaded letters mark deviations in nucleotides

experimentally validated CRISPR-Cas system to variant *M. viscosa* predicted by BLAST searches is the CRISPR-Cas system of *Pectobacterium atrosepticum* [48] (Additional file 4: Table S4). No CRISPR-Cas system could be identified in *M. viscosa* F57 and NVI 5482 (variant clade 3) using the same method. Further support for this observation was found using the flanking regions of the CRISPR-Cas. Downstream of the operons harboured an ABC transporter and a cold-shock DNA-binding domain family protein genes. Nucleotidyltransferase or a ferrous iron transport protein gene was identified upstream. Using these genes, the same regions were identified in F57 and NVI 5482 without signs of any CRISPR-Cas.

**Protospacer sequences are shared in related *M. viscosa***

In total, 412 spacers were identified among the nine CRISPR carrying *M. viscosa* (Fig. 4 and Table 1). Searches against the GenBank-Phage and RefSeq-Plasmid databases revealed only two spacer matches (defined as  $\leq 4$  SNPs = 28/32 nucleotides). Spacer 4919r6 matched *Gluconobacter oxydans* 621H plasmid pGOX1, while spacer 5450r53 matched an *Oenococcus* phage sequence. Comparing the spacers within our isolate collection identified 57 unique spacers mostly at the leader proximal end, which implies that they are the most recent spacers in terms of acquisition. The structure and similarity of the repeat-spacer arrays show a high heterogeneity of spacer content among *M. viscosa*. Overall, three main genotypes of spacer-sets could be assigned to variant clade 1, variant clade 2 and typical *M. viscosa* isolates (Fig. 4, Table 1), congruent to strain evolutionary relationships. The commonality between spacer-arrays in typical *M. viscosa* strains reflects the phylogenetic clustering of typical *M. viscosa*. The more distantly related isolates of clade 2 contain a different spacer-array set, which is conserved in synteny among clade 2 strains. The spacer-arrays in variant clade 1 *M. viscosa* is further comparable to strain evolutionary distance. Meaning that closely related isolates, e.g. K58, Vvi-7 and Vvi-11, are also displaying more similar repeat-spacer arrays, which become more variable with phylogenetic distance (compared to NVI 4917, or even further to NVI 5450). Trout isolates of clade 1 show a spacer-array pattern of similar origin but with a higher diversity in the more recent acquired spacers compared to Atlantic salmon isolates. The anchor spacer is the oldest spacer in terms of acquisition. This spacer is attained identical in all variant clade 1 and clade 2 strains. Strain K56 and MV 0609139 (clade 2) spacer-arrays are very similar in structure to each other and two spacers are identical to spacers in the arrays of the remaining variant strains. One spacer in typical *M. viscosa* is found in variant spacer-arrays.





**Fig. 4** CRISPR profiles in nine *Moritella viscosa* strains indicating matches between spacers or other genomic elements in *M. viscosa*. Each spacer in all *M. viscosa* strains are presented in numerical order (latest acquired is first) in the first row. Spacer ID's in Additional file 3: Data S1 are named by its bacterial affiliation and numbered progressively where the highest number designates the last obtained spacer. Percentage and the number of identical spacers shared between *M. viscosa* strains are shown in detail in Table 1. Green spacers indicate identical spacers found in two or more *M. viscosa* strains. Light green indicate spacers with one mismatch to spacers in other *M. viscosa* strains. Black spacers indicate spacers found in specific *M. viscosa* spacer-arrays that are identical to prophage genes identified in other *M. viscosa* strains. Red spacers indicate spacers found in specific *M. viscosa* spacer-arrays that are identical to plasmid genes identified in other *M. viscosa* strains

**Protospacer containing prophage and plasmid-like CDS in *M. viscosa***

BLAST searches of all *M. viscosa* spacers against the *M. viscosa* genomes revealed complimentary sequences (protospacers) that were part of prophage related genes (Additional file 4: Figure S3). MT 2528 spacers 2528r41 and 2528r40 are identical in sequence to prophage 1 and prophage 2, respectively. Concurrently, prophage 1 and 2 are predicted in the typical *M. viscosa* for both NVI 3632 and LFI 5006, except MT 2528 (Fig. 2). Spacer 2528r40 is also similar with one mismatch to prophage 7. NVI 5450 spacers 5450r24 and 5450r25 are identical in sequences to two genes predicted within prophage 4, while spacers 5450r13, 5450r14 and 5450r15 are similar to three genes in prophage 5.

Identical protospacers were also identified in the *M. viscosa* genomes to plasmid related genes (Additional file 5). Thirty-three spacers matched sequences within three putative *M. viscosa* plasmid designated A, B and C. Spacers within *M. viscosa* strains from variant clade 1 and clade 2

in addition to typical *M. viscosa* matched to plasmid A. Strains from variant clade 1 had spacers against plasmid B and plasmid C. Genes in plasmid A and B that contain one or more protospacers are predicted with functions that are essential to conjugative transfer. In plasmid A, *trbC*, *trbJ*, *trbL* and *traG* are targeted in addition to an uncharacterized protein gene (K56\_4586 and MT2528\_4007) and a putative serine protease (K56\_4594 and MT2528\_4809). In plasmid B, the conjugative transfer genes *traN*, *traE* and the *repA* gene encoding the putative replication protein, are targeted. Spacer K56r10 and Vvi-11r8 identical to protospacer sequence in plasmid B *repA* are also similar with three mismatches to the plasmid C *repA*, which could be caused by the sequence similarity. It is noteworthy that spacers in variant clade 1 strains repeatedly match the plasmid C *repA* gene. Both consecutive spacers, as well as spacers that are acquired at different time points (other spacers are between them) are observed.

**Discussion**

This study presents the first comparative genome analysis of *M. viscosa*. Analyses of the genome plasticity among strains revealed that vertical and horizontal evolution relationships are concurrent to each other. By predicting the function of accessory and unique genes among *M. viscosa*, it was revealed that many of the genes resulted from predicted MGEs such as prophages and plasmids. We further used genome structure characteristics to investigate if *M. viscosa* has mechanisms for acquired immunity against MGEs. Two subtypes of the CRISPR-Cas I-F system were identified. The distribution of these systems and the spacer-array variants correlate with the phylogenetic lineage pattern. The whole-genome phylogenies indicate four *M. viscosa* lineages expanding the previously suggested classification of

**Table 1** Identical spacers shared between *Moritella viscosa* strains

Strains	Variant clade 2		Variant clade 1					Typical	
	MV 0609139	K56	NVI 5450	NVI 4917	K58	Vvi-7	Vvi-11	MT 2528	NVI 3632
	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)	No. (%)
MV 0609139	47 (100)	46 (97.9)	2 (4.3)	2 (4.3)	2 (4.3)	2 (4.3)	2 (4.3)	1 (2.1)	1 (2.1)
K56	46 (93.9)	49 (100)	2 (4.1)	2 (4.1)	2 (4.1)	2 (4.1)	2 (4.1)	1 (2.0)	1 (2.0)
NVI 5450	2 (3.6)	2 (3.6)	55 (100)	14 (25.5)	14 (25.5)	12 (21.8)	14 (25.5)	0 na.	0 na.
NVI 4917	2 (5.7)	2 (5.7)	14 (40.0)	35 (100)	25 (71.4)	15 (42.9)	25 (71.4)	0 na.	0 na.
K58	2 (3.7)	2 (3.7)	14 (25.9)	25 (46.3)	54 (100)	39 (72.2)	49 (90.7)	1 (1.9)	1 (1.9)
Vvi-7	2 (5.0)	2 (5.0)	12 (30.0)	15 (37.5)	39 (97.5)	40 (100)	40 (100)	1 (2.5)	1 (2.5)
Vvi-11	2 (3.9)	2 (3.9)	14 (26.9)	25 (48.1)	49 (94.2)	40 (76.9)	52 (100)	1 (1.9)	1 (1.9)
MT 2528	1 (2.4)	0 na.	na.	na.	1 (2.4)	1 (2.4)	1 (2.4)	41 (100)	39 (95.1)
NVI 3632	1 (2.6)	0 na.	na.	na.	1 (2.6)	1 (2.6)	1 (2.6)	39 (100)	39 (100)

Grey shading indicates the three CRISPR-array genotypes in *M. viscosa*. Quantities of identical spacers are displayed as the specific numbers with the percentage of the total spacer array in brackets

typical and variant *M. viscosa* [5, 8], which might suggest that sublineage definition among *M. viscosa* needs revision. Spacer-arrays within each lineage are conserved in synteny. In contrast, little commonality is observed between each lineage. That spacer composition can be linked to *M. viscosa* population structure and evolutionary relationships is similar to other bacteria [49]. CRISPR typing can provide tracking and subtyping of pathogenic strains [18–20, 50, 51]. Strain typing and tracking of *M. viscosa* could potentially enhance our understanding of the ecological context of infectious winter-ulcer disease. However, a broader range and number of isolates are needed to establish such a method as there is no evident phylogenetic or genotypic pattern that associate *M. viscosa* subgroups to geographic distribution or with host type from the isolates used in this study.

That only two *M. viscosa* CRISPR-spacers matched to protospacer sequences of known plasmids and phages could be a result of the expected large variety of MGEs present in marine environments. Functionality of CRISPR-Cas where spacer sequences provide prophage resistance [52] and limit plasmid transfer [53] in *M. viscosa* was indicated by the correlation between the CRISPR-spacer content and the distribution of the matching MGE. That MGE-matching CRISPR-spacers are excluded is observed in MT 2528 where the two unique and most recent acquired spacers match prophages present in typical *M. viscosa*, except MT 2528. Similarly, CRISPR-spacers in NVI 5450 match to prophages present in other *M. viscosa* strains, but which are absent in NVI 5450. Supporting this model, plasmid B is absent from genomes containing matching CRISPR-spacers, but is present in NVI 5482 without any matching CRISPR-spacers. Plasmid C is similarly predicted in genomes lacking matching CRISPR-spacers.

Divergence to this model is observed between NVI 3632 and MT 2528 that both have two CRISPR-spacers directed at plasmid A. NVI 3632 is not predicted with plasmid A but MT 2528 harbours plasmid A. The CRISPR-arrays are identical except for the two most recent prophage spacers in MT 2528, which suggests a functional CRISPR-Cas in MT 2528. The possibility of CRISPR autoimmunity is rejected, as plasmid A spacers do not match any of the CRISPR-Cas gene sequences, which in addition are identical at the nucleic acid level ruling out any recent mutational effect causing inefficient or defective CRISPR-Cas system in MT 2528. The reason is not known but the escape from the CRISPR-Cas system could be caused by mutations in other sequence motifs, which is known to avoid recognition [48]. It is interesting to note that these spacers are acquired at two different time points with 30 in-between-spacers suggesting multiple interactions with this

plasmid-type. Strain K56 harbor a CRISPR-spacer with 1 bp spacer-mismatch to plasmid A, which could explain how this plasmid evade CRISPR-Cas immunity in this strain. Mutations in the targeted MGE can lead to repetitive acquisition or incorporating of new spacers to the CRISPR-array that again increase resistance against the invading MGE [54]. The repetitive acquisition of spacers, in addition to spacers that show mismatches to essential genes within MGEs predicted in this study, suggests reoccurring encounters or interaction with variants of these MGEs at previous time points as described in other marine bacteria [49]. The existence of a co-evolutionary “arms race” where CRISPR immunity drives MGE evolution [16] may also occur between *M. viscosa* CRISPR-Cas and their targets.

The CRISPR-Cas targeted prophage genes in *M. viscosa* are essential for genome integration and to a prophage life cycle. Targeted plasmid genes are essential for replication or conjugation. Essential genes are often more conserved in sequence conservation, meaning that targeting these genes would confer a more efficient immunity over an extended period. It is noteworthy that *repA* in plasmid C (but also plasmid B) is repetitively targeted by the CRISPR-Cas system in variant clade 1. This might be due to spacer acquisition preferences as CRISPR-Cas target plasmids in preferentially regions [55]. Alternatively, genetic elements could acquire escape mutations or genetic shuffling that elude the CRISPR-Cas immunity [56] and adapt to infect their environment preferential host type [57] being able to repeatedly infect the host as observed in the distance between acquired spacers in the CRISPR-array. In the *Escherichia coli* plasmid prophage N15, *repA* is the only gene necessary for replication [58]. Targeting this gene will provide defence against all variants of MGEs containing this or related *repA* with matching protospacers and could suggest that *M. viscosa* CRISPR-Cas also targets MGEs in a meticulous manner.

CRISPR-Cas mediated immunity can provide bacteria an advantage in the presence of a lytic phage [59]. It is shown that temperature may induce bacterial stress responses that activate the lysogenic switch of prophages [60]. Although, no lysis module was predicted in *M. viscosa* prophages, it cannot be excluded that prophages may play a role in the lytic switch of *M. viscosa* observed above 10 °C [61] and be a situation where CRISPR-Cas mediated immunity provide an advantage in *M. viscosa*. Targeting of conjugative plasmids is likely dependent on if plasmid genes may become a burden in particular environments or not [62, 63]. Spacers matching to conjugative transfer genes in *M. viscosa* could suggest that some conjugative plasmids impose an unwanted burden in *M. viscosa*. Targeting of unessential plasmid genes indicates additional specific genetic elements unwanted in

*M. viscosa* such as the plasmid A encoding T2SS / T4 pilus-like transport system. It is unknown if the system could affect the genomic T2SS but the complex is likely driving the translocation of the predicted trypsin-like serine protease that share structural similarities to VesB, a T2SS exoprotein in *V. cholerae* [64]. Transportation is supported by the predicted N-terminal signal peptide similar to other proteases that enters the periplasm via the Sec pathway before T2SS [65]. The protease has similar to VesB a predicted Ig-fold of unknown function [66]. Ig-like domains are found in several types of cell surface proteins involved in substrate specificity or surface recognition [67]. Expression of plasmid A genes could alter host cell adhesion and invasion properties of *M. viscosa* or alternatively result in autolysis similar to the T2SS translocated serine protease in *Vibrio vulnificus* [68].

The high population density and eutrophic environment in fish farming could have selected for and facilitated the rapid strain flow of host specific typical *M. viscosa* [8] in Norwegian Atlantic salmon aquaculture compared to the more diversification of the pathogen in other fish species and geographical areas [5]. If assumed that CRISPR-arrays are an indirect reflection of the environment, i.e. it reflects the type of MGEs encountered in the environment occupied by the bacteria, it will indicate that the different sublineages originate from different environments. However, a variety of mechanisms unrelated to CRISPR-Cas conferred immunity could affect the sensitivity to MGEs [69]. It is further postulated that CRISPR-Cas systems are lost when they confer immunity to acquired beneficial genes, and subsequently regained in environments where protection against MGEs again increase fitness [70]. The CRISPR-Cas of variant clade 3 *M. viscosa* could have similarly been lost during clade specific evolution. This lineage is the closest in relationship to typical *M. viscosa*, which has a separate CRISPR-Cas system that could have been gained in response to a different environment. Although the CRISPR genotypes are distinct, they are all found in isolates from salmonids. This could relate to a relatively isolated niche in which these strains are isolated and could indicate that CRISPR-Cas inferred immunity has a positive consequence in the eutrophic environment of fish farming.

## Conclusions

From the comparative genome analysis in this study, we describe how the genome plasticity and relationships among *M. viscosa* is reflected by MGEs. The correlation between CRISPR-spacers that matches protospacers suggests that CRISPR-Cas confer adaptive immunity against MGEs in *M. viscosa*, and is a counter-strategy acquired in multiple events. Moreover, our findings suggest

that CRISPR-Cas and their spacer-array contents originating from foreign DNA correlate with the evolutionary relationships among *M. viscosa* that could provide a new tool for evaluating diversity and strain tracking of *M. viscosa*.

## Additional files

**Additional file 1: Table S1.** *Moritella viscosa* isolates genome sequenced in this study. (DOCX 31 kb)

**Additional file 2: Table S2A.** Summary of genomic statistics of *M. viscosa*. **Table S2B.** Sequence statistics. **Table S2C.** Orthologue statistics. **Table S2D** Comparative genome analysis. (XLSX 1103 kb)

**Additional file 3: Data S1.** *M. viscosa* CRISPR spacers. (TXT 17 kb)

**Additional file 4: Figure S1.** Prophages in *M. viscosa*, **Figure S2.** Prophage integrase phylogeny, **Table S3.** Prophage att sites, **Table S4.** Homology analysis of CRISPR-Cas systems identified in *Moritella viscosa*, **Figure S3.** Prophage protospacers. (DOCX 442 kb)

**Additional file 5: Table S5A.** *M. viscosa* spacers matching prophage-like elements. **Table S5B.** *M. viscosa* spacers matching plasmid-like elements. (XLSX 18 kb)

## Abbreviations

BLAST: Basic local alignment search tool; CDS: Coding sequence; CRISPR-Cas: Clustered regularly interspaced short palindromic repeats and associated Cas proteins; HGT: Horizontal gene transfer; MGE: Mobile genetic element; PHAST: Phage search tool; SNP: Single-nucleotide polymorphism; T2SS: Type II secretion system

## Acknowledgements

We would like to thank Prof. Henning Sørum for helpful comments in the work of this manuscript.

## Funding

This work was funded by the Norwegian Research Council (grant no. HAVBRUK-216196/E40).

Nofima provided support in the form of salary for author CK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Availability of data and materials

The strains mentioned in this work are available. Send inquiries to Christian Karlsen, Nofima AS, Division of Aquaculture, PO Box 210, Ås N-1431, Norway. Email: Christian.karlsen@nofima.no.

The *M. viscosa* genomes have been deposited to the European Nucleotide Archive (ENA) under the ENA study accession number PRJEB1601 with the following genome accession numbers ERS1419585 to ERS1419596.

## Authors' contributions

CK, EH and NPW designed the study. CK did the laboratory work. EH and TK assembled, annotated and performed comparative genomics analysis. CK performed CRISPR analysis. CK, EH, TK and NPW contributed with interpretation and discussion of the results. CK, EH and TK drafted and revised the manuscript. All authors read and approved the final manuscript.

## Competing interests

Nofima is a non-profit research institution. CK is employed by Nofima. There are no patents or products in development to declare. The authors declare that they have no competing interests.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Not applicable.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>Department of Food Safety and Infection Biology, Norwegian University of Life Sciences (NMBU), Pb 8146 Dep.0033 Oslo, Norway. <sup>2</sup>Present address: Nofima AS, Division of Aquaculture, PO Box 210, Ås N-1431, Norway. <sup>3</sup>Department of Chemistry, Faculty of Science and Technology, University of Tromsø, N-9037 Tromsø, Norway. <sup>4</sup>The Norwegian Structural Biology Centre, University of Tromsø, N-9037 Tromsø, Norway.

Received: 2 February 2017 Accepted: 6 April 2017

Published online: 20 April 2017

### References

- Benediktsdottir E, Verdonck L, Sproer C, Helgason S, Swings J. Characterization of *Vibrio viscosus* and *Vibrio wodanis* isolated at different geographical locations: a proposal for reclassification of *Vibrio viscosus* as *Moritella viscosa* comb. nov. *Int J Syst Evol Microbiol*. 2000;50:479–88.
- Lunder T, Sørum H, Holstad G, Steigerwalt AG, Mowinckel P, Brenner DJ. Phenotypic and genotypic characterization of *Vibrio viscosus* sp. nov. and *Vibrio wodanis* sp. nov. isolated from Atlantic salmon (*Salmo salar*) with 'winter ulcer'. *Int J Syst Evol Microbiol*. 2000;50(2):427–50.
- Benediktsdottir E, Helgason S, Sigurjonsdottir H. *Vibrio* spp. isolated from salmonids with shallow skin lesions and reared at low temperature. *J Fish Dis*. 1998;21(1):19–28.
- Bruno DW, Griffiths J, Petrie J, Hastings TS. *Vibrio viscosus* in farmed Atlantic salmon *Salmo salar* in Scotland: field and experimental observations. *Dis Aquat Organ*. 1998;34(3):161–6.
- Grove S, Wiik-Nielsen CR, Lunder T, Tunsjø HS, Tandstad NM, Reitan LJ, Marthinussen A, Sørgaard M, Olsen AB, Colquhoun DJ. Previously unrecognised division within *Moritella viscosa* isolated from fish farmed in the North Atlantic. *Dis Aquat Organ*. 2010;93(1):51–61.
- Lunder T, Evensen O, Holstad G, Hastein T. Winter ulcer in the Atlantic salmon *Salmo salar* - Pathological and bacteriological investigations and transmission experiments. *Dis Aquat Organ*. 1995;23(1):39–49.
- Whitman KA, Backman S, Benediktsdottir E, Coles M, Johnson G. Isolation and characterization of a new *Vibrio* spp. (*Vibrio wodanis*) associated with winter ulcer disease in sea water raised Atlantic salmon (*Salmo salar* L.) in New Brunswick. Aquaculture Association Canada, SpecialPublication. 2000;4:115–7.
- Karlsen C, Ellingsen AB, Wiik-Nielsen C, Winther-Larsen HC, Colquhoun D, Sørum H. Host specificity and clade dependent distribution of putative virulence genes in *Moritella viscosa*. *Microb Pathog*. 2014;77:53–65.
- Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature*. 2000;405(6784):299–304.
- Parisien A, Allain B, Zhang J, Mandeville R, Lan CQ. Novel alternatives to antibiotics: bacteriophages, bacterial cell wall hydrolases, and antimicrobial peptides. *J Appl Microbiol*. 2008;104(1):1–13.
- Sorek R, Kunin V, Hugenholtz P. CRISPR - a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Micro*. 2008;6(3):181–6.
- He L, Fan X, Xie J. Comparative genomic structures of *Mycobacterium* CRISPR-Cas. *J Cell Biochem*. 2012;113(7):2464–73.
- Horvath P, Coûté-Monvoisin A-C, Romero DA, Boyaval P, Fremaux C, Barrangou R. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int J Food Microbiol*. 2009;131(1):62–70.
- Cai F, Axen SD, Kerfeld CA. Evidence for the widespread distribution of CRISPR-Cas system in the Phylum *Cyanobacteria*. *RNA Biol*. 2013;10(5):687–93.
- Horvath P, Barrangou R. CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science*. 2010;327(5962):167–70.
- Paez-Espino D, Sharon I, Morovic W, Stahl B, Thomas BC, Barrangou R, Banfield JF. CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. *mBio*. 2015;6(2):e00262-15.
- Stern A, Mick E, Tirosh I, Sagy O, Sorek R. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Res*. 2012;22(10):1985–94.
- Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol*. 1997;35(4):907–14.
- Mokrousov I, Limeschenko E, Vyazovaya A, Narvskaya O. *Corynebacterium diphtheriae* spoligotyping based on combined use of two CRISPR loci. *Biotechnol J*. 2007;2(7):901–6.
- Liu F, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ, Dudley EG. Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Appl Environ Microbiol*. 2011;77(6):1946–56.
- Bjornsdottir B, Gudmundsdottir T, Gudmundsdottir BK. Virulence properties of *Moritella viscosa* extracellular products. *J Fish Dis*. 2011;34(5):333–43.
- Hjerde E, Karlsen C, Sørum H, Parkhill J, Willassen NP, Thomson NR. Co-cultivation and transcriptome sequencing of two co-existing fish pathogens *Moritella viscosa* and *Aliivibrio wodanis*. *BMC Genomics*. 2015;16(1):447–59.
- Karlsen C, Vanberg C, Mikkelsen H, Sørum H. Co-infection of Atlantic salmon (*Salmo salar*), by *Moritella viscosa* and *Aliivibrio wodanis*, development of disease and host colonization. *Vet Microbiol*. 2014;171:112–21.
- Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics*. 2009;25(15):1968–9.
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*. 2007;23(6):673–9.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
- Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res*. 2004;32 suppl 1:D115–9.
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2014;42(D1):D222–D230.
- Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res*. 2013;41(12):e121.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth*. 2011;8(10):785–6.
- Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003;13(9):2178–89.
- Krzywinski M, Schein J, Birol I, Connors J, Gascayne R, Horsman D, Jones SJ, Marra MA. Circos: An information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639–45.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25–9.
- Zdobnov EM, Apweiler R. InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. 2001;17(9):847–8.
- Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, et al. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res*. 2006;34(Web Server issue):W293–7.
- Gardner SN, Hall BG. When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. *PLoS One*. 2013;8(12): e81760.
- Vinuesa P, Contreras-Moreira B. Robust identification of orthologues and paralogues for microbial pan-genomics using GET\_HOMOLOGUES: A case study of *plncA/C* plasmids. In: Mengoni A, Galardini M, Fondi M, editors. *Bacterial Pangenomics: Methods and Protocols*. New York, NY: Springer New York; 2015. p. 203–32.
- Griebel T, Brinkmeyer M, Böcker S. EPOs: a modular software framework for phylogenetic analysis. *Bioinformatics*. 2008;24(20):2399–400.
- Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: A Fast Phage Search Tool. *Nucleic Acids Res*. 2011;39(suppl\_2):W347–W352.
- Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*. 2007;35(Web Server issue):W52–7.
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999;41:95–8.
- Biswas A, Gagnon JN, Brouns SJJ, Fineran PC, Brown CM. CRISPRTarget: Bioinformatic prediction and analysis of crRNA targets. *RNA Biol*. 2013;10(5): 817–27.
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protocols*. 2015;10(6):845–58.

44. Robinson DF, Foulds LR. Comparison of phylogenetic trees. *Math Biosci.* 1981;53(1):131–47.
45. Paul JH. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J.* 2008;2(6):579–89.
46. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2015;44(D1):D279–85.
47. Yosef I, Goren MG, Qimron U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* 2012;40(12):5569–76.
48. Richter C, Fineran PC. The subtype I-F CRISPR–Cas system influences pathogenicity island retention in *Pectobacterium atrosepticum* via crRNA generation and Csy complex formation. *Biochem Soc Trans.* 2014;41:1468–74.
49. Wietz M, Millan-Aguinaga N, Jensen PR. CRISPR–Cas systems in the marine actinomycete *Salinispora*: linkages with phage defense, microdiversity and biogeography. *BMC Genomics.* 2014;15:936.
50. Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, Accou-Demartin M, de Romans S, Lim C, Roux C, Passet V, et al. CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. *PLoS One.* 2012;7(5), e36995.
51. van Belkum A, Soriaga LB, LaFave MC, Akella S, Veyrieras J-B, Barbu EM, Shortridge D, Blanc B, Hannum G, Zambardi G, et al. Phylogenetic distribution of CRISPR–Cas systems in antibiotic-resistant *Pseudomonas aeruginosa*. *mBio.* 2015;6(6):e01796–15.
52. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* 2007;315(5819):1709–12.
53. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA. *Science (New York, NY).* 2008;322(5909):1843–5.
54. Richter C, Dy RL, McKenzie RE, Watson BNJ, Taylor C, Chang JT, McNeil MB, Staals RHJ, Fineran PC. Priming in the Type I-F CRISPR–Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.* 2014;42(13):8516–26.
55. Westra ER, Staals RH, Gort G, Høgh S, Neumann S, de la Cruz F, Fineran PC, Brouns SJ. CRISPR–Cas systems preferentially target the leading regions of MOB(F) conjugative plasmids. *RNA Biol.* 2013;10(5):749–61.
56. Maniv I, Jiang W, Bikard D, Marraffini LA. Impact of different target sequences on Type III CRISPR–Cas immunity. *J Bacteriol.* 2016;198(6):941–50.
57. Kottara A, Hall JPJ, Harrison E, Brockhurst MA. Multi-host environments select for host-generalist conjugative plasmids. *BMC Evol Biol.* 2016;16:70.
58. Ravin NV, Kuprianov VV, Gilcrease EB, Casjens SR. Bidirectional replication from an internal ori site of the linear N15 plasmid prophage. *Nucleic Acids Res.* 2003;31(22):6552–60.
59. Levin BR. Nasty viruses, costly plasmids, population dynamics, and the conditions for establishing and maintaining CRISPR-mediated adaptive immunity in bacteria. *PLoS Genet.* 2010;6(10), e1001171.
60. Cochran PK, Paul JH. Seasonal abundance of lysogenic bacteria in a subtropical estuary. *Appl Environ Microbiol.* 1998;64(6):2308–12.
61. Benediktsdottir E, Heidarsdottir KJ. Growth and lysis of the fish pathogen *Moritella viscosa*. *Lett Appl Microbiol.* 2007;45(2):115–20.
62. Ghigo JM. Natural conjugative plasmids induce bacterial biofilm development. *Nature.* 2001;412(6845):442–5.
63. Rosch TC, Golman W, Hucklesby L, Gonzalez-Pastor JE, Graumann PL. The presence of conjugative plasmid pLS20 affects global transcription of its *Bacillus subtilis* host and confers beneficial stress resistance to cells. *Appl Environ Microbiol.* 2014;80(4):1349–58.
64. Zielke RA, Simmons RS, Park BR, Nonogaki M, Emerson S, Sikora AE. The Type II Secretion pathway in *Vibrio cholerae* is characterized by growth phase-dependent expression of exoprotein genes and is positively regulated by  $\alpha E$ . *Infect Immun.* 2014;82(7):2788–801.
65. Filloux A. The underlying mechanisms of type II protein secretion. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research.* 2004;1694:163–79.
66. Gadwal S, Korotkov KV, Delarosa JR, Hol WGJ, Sandkvist M. Functional and structural characterization of *Vibrio cholerae* extracellular serine protease B. *VesB The Journal of Biological Chemistry.* 2014;289(12):8288–98.
67. Bodelon G, Palomino C, Fernandez LA. Immunoglobulin domains in *Escherichia coli* and other enterobacteria: from pathogenesis to applications in antibody technologies. *FEMS Microbiol Rev.* 2013;37(2):204–50.
68. Lim MS, Kim JA, Lim JG, Kim BS, Jeong KC, Lee KH, Choi SH. Identification and characterization of a novel serine protease, VvpS, that contains two functional domains and is essential for autolysis of *Vibrio vulnificus*. *J Bacteriol.* 2011;193(15):3722–32.
69. Samson JE, Magadan AH, Sabri M, Moineau S. Revenge of the phages: defeating bacterial defences. *Nat Rev Micro.* 2013;11(10):675–87.
70. Jiang W, Maniv I, Arain F, Wang Y, Levin BR, Marraffini LA. Dealing with the evolutionary downside of CRISPR immunity: Bacteria and beneficial plasmids. *PLoS Genet.* 2013;9(9), e1003844.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
www.biomedcentral.com/submit

