

RESEARCH

Open Access



# Extensive diversity and rapid turnover of phage defense repertoires in cheese-associated bacterial communities

Vincent Somerville<sup>1,2\*†</sup>, Thibault Schowing<sup>2,3†</sup>, H el ene Chabas<sup>4</sup>, Remo S. Schmidt<sup>2</sup>, Ueli von Ah<sup>2</sup>, R emy Bruggmann<sup>3</sup> and Philipp Engel<sup>1\*</sup>

## Abstract

**Background:** Phages are key drivers of genomic diversity in bacterial populations as they impose strong selective pressure on the evolution of bacterial defense mechanisms across closely related strains. The pan-immunity model suggests that such diversity is maintained because the effective immune system of a bacterial species is the one distributed across all strains present in the community. However, only few studies have analyzed the distribution of bacterial defense systems at the community-level, mostly focusing on CRISPR and comparing samples from complex environments. Here, we studied 2778 bacterial genomes and 188 metagenomes from cheese-associated communities, which are dominated by a few bacterial taxa and occur in relatively stable environments.

**Results:** We corroborate previous laboratory findings that in cheese-associated communities nearly identical strains contain diverse and highly variable arsenals of innate and adaptive (i.e., CRISPR-Cas) immunity systems suggesting rapid turnover. CRISPR spacer abundance correlated with the abundance of matching target sequences across the metagenomes providing evidence that the identified defense repertoires are functional and under selection. While these characteristics align with the pan-immunity model, the detected CRISPR spacers only covered a subset of the phages previously identified in cheese, providing evidence that CRISPR does not enable complete immunity against all phages, and that the innate immune mechanisms may have complementary roles.

**Conclusions:** Our findings show that the evolution of bacterial defense mechanisms is a highly dynamic process and highlight that experimentally tractable, low complexity communities such as those found in cheese, can help to understand ecological and molecular processes underlying phage-defense system relationships. These findings can have implications for the design of robust synthetic communities used in biotechnology and the food industry.

**Keywords:** Bacteriophage-host interaction, Microbial ecology, Evolutionary dynamics, Shotgun metagenomics

## Background

Bacteria have evolved diverse defense systems to cope with the parasitic lifestyle of phages [19, 45]. These systems can be divided into the innate and the adaptive “prokaryotic immune system” [18]. Classical examples of innate immunity are restriction-modification [80] or abortive infection systems [39]. However, many additional innate immune mechanisms have recently been discovered highlighting the strong selective pressure imposed by phages on microbial communities [19, 38].

<sup>†</sup>Vincent Somerville and Thibault Schowing contributed equally to this work.

\*Correspondence: vincent.somerville@unil.ch; philipp.engel@unil.ch

<sup>1</sup> Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland

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



The only 'adaptive' immune system known so far is the CRISPR-Cas system (Clustered Regularly Interspaced Short Palindromic Repeat-CRISPR Associated). It is based on the incorporation of short DNA sequences of phages or other genetic elements (so-called spacers) into dedicated CRISPR arrays encoded in the bacterial genome. Upon a phage encounter, the transcribed spacers bind to the phage DNA and target it for degradation via the Cas proteins [37].

Phage defense systems are prevalent across bacteria and most bacteria harbor several systems in their genome [19, 52]. However, their distribution varies across bacteria. For example, CRISPR-Cas systems are found in 40% of all known bacterial species [14], while Viperins are restricted to a few taxonomic groups [5]. Moreover, phage defense systems are often strain-specific [78], i.e., closely related bacteria can harbor completely different arsenals of defense systems. Various factors have been discussed to influence the distribution of defense systems across bacteria [60]. Notably, the presence of CRISPR-Cas has been associated with environmental factors such as temperature, oxygen levels or phage abundance [10, 50, 82]. Also, genetic incompatibilities of defense systems with other cellular functions, including other defense systems, have been reported [4, 7].

Innate immune systems often cluster in genomic islands and are associated with mobile genetic elements [41, 46, 48] suggesting an important role of horizontal gene transfer (HGT) for defense system evolution. The ecological relevance of such genomic plasticity was demonstrated by a recent study which showed that nearly clonal isolates of *Vibrio spp.* are resistant to diverse phages due to the presence of distinctive defense islands acquired via HGT [33].

Based on the observation that phage defense mechanisms show a high extent of genetic turnover, the pan-immunity hypothesis has been proposed, which states that the effective immune system of a bacterial species is not the one encoded in a single genome, but in the pan-genome of the entire population [6]. In other words, while a single strain cannot carry all possible defense systems, the presence of nearly clonal strains with different defense systems increases the available arsenal of defensive mechanisms via HGT and thus increases the resistance of the entire population (pan-immunity).

Comparative genomics of closely related isolates combined with shotgun metagenomics provide excellent opportunities to assess the diversity of phage defense systems in natural bacterial communities and can help to advance our understanding of their evolutionary ecology [69]. Two recent shotgun metagenomic studies have looked at CRISPR spacer diversity in microbial

communities, one focusing on environmental samples from the Earth Microbiome Project [50] and another one focusing on diverse samples from the human microbiome [53]. Meaden et al. identified a positive association between CRISPR spacer and the abundance of the corresponding phage target sequences (protospacer), suggesting that there is a direct link between phage pressure and the maintenance of corresponding spacer sequences. Münch et al. identified differences in the prevalence of CRISPR spacers between different human body sites suggesting the existence of niche-specific phage populations. However, none of the two studies looked at the diversity of innate immune systems. Moreover, in both studies highly diverse communities from heterogeneous environments were compared providing limited insights into the intraspecific variation of phage defense systems and their evolutionary turnover in bacterial populations.

Here, we focused on cheese-associated bacterial communities. These communities harbor only a few bacterial species, have been propagated in relatively stable environments (i.e., cheese or milk) over generations, and are known to be exposed to diverse phages [8, 40, 44, 51]. This makes them tractable systems to study the evolution of phage defense systems in microbial communities at the strain-level [25, 75]. Indeed, our previous study of a single Swiss cheese starter culture has shown that extensive intraspecific CRISPR spacer diversity exists in these otherwise nearly clonal populations of bacteria [74].

Here, we expanded this analysis to all publicly available genomic datasets from cheese-associated communities (excluding cheese rind) comprising 26 bacterial species, 2778 genomes, and 188 metagenomes. We determined the distribution of both innate and adaptive immune systems across these datasets and quantified the diversity, abundance, turnover rate, and viral targets of all CRISPR spacers. We find that (i) cheese-associated bacteria contain an unprecedented high degree of diversity in phage resistance mechanisms across nearly identical strains, (ii) there is a strong correlation between CRISPR spacer and phage abundance, and (iii) CRISPR spacers only provide immunity to a subset of the phages identified in cheese. These results indicate highly dynamic bacteria-phage interactions driving genomic plasticity in cheese-associated environments and are compatible with the pan-immunity model of the evolutionary ecology of defense systems in microbial communities.

## Materials and methods

### Metagenomes

In order to analyze species and CRISPR diversity in cheese related samples we gathered 188 shotgun metagenomic samples and 240 16S rRNA amplicon

sequencing samples from overall 18 studies from NCBI (see Suppl. Tables 1 and 2) [1, 11, 15–17, 21, 24, 26, 35, 43, 47, 59, 67, 74, 76, 81]. We included mesophilic (cooked at lower temperatures) and thermophilic (cooked at higher temperatures) cheese starter cultures as well as samples from ripened cheese as based on the FoodMicrobionet database (8.2020), a database containing information about species found in food-associated microbial communities [56]. We excluded cheese rind samples from this analysis, because they consist of highly variable microbial communities with a more complex ecology [86].

### Metagenomic species profiling

To determine the species composition of the metagenomic samples, we used MetaPhlAn (v3.0.7-1) [70]. For the 16S rRNA gene amplicon sequencing datasets we used the extensive FoodMicrobionet (8.2020) collection and analysis pipeline (see script section) [56, 58]. While 16S rRNA community analysis samples were only used for species profiling, the metagenomic samples were also used for spacer and protospacer analysis (see further down). Overall, 185 species were identified. We selected the dominant species which were present in > 2% of samples and had a median overall abundance > 0.1%. From those species we randomly selected one genome and created a species tree with BCGtree (v1.1.0) [3].

### Genomes analysis

All genomes (or max. 500 newest genomes if more were available) for the 185 identified species were downloaded from NCBI RefSeq (17.05.2021). For the 26 dominant spe-

script (see script section). Only spacers with a high evidence level (> 4) and shorter than 50bp were retained. CRISPR-Cas subtypes were assigned with CRISPR-CasTyper (v1.4.4) [68]. Further, the remaining defense mechanisms were annotated with defense mechanisms specific HMM files. For the R-M we used previously described HMMs [54]. The search was done with hmsearch (v3.1b2) [62].

### Pairwise strains comparison

Average nucleotide identity (ANI) was calculated with fastANI (v1.3) [34] and percentage of shared spacers were computed for each pair of strains within an organism. The proportion of shared spacers was calculated as being the proportion of identical spacer clusters between two strains, divided by all spacers.

Additionally, the number of common spacers, the nucleotide diversity and the turnover rate were computed for all strain pairs as well as for each array. The nucleotide diversity was computed as *bidirectional fragment mapping* \* *fragment length (1kb)/ANI* and the turnover rate as *nucleotide diversity / unique spacers*, unique spacers meaning the number of spacers not shared between the two strains. Further, the CRISPR acquisition rate in microbial communities per generation (i.e., CRISPR turnover rate) was calculated as following (turnover rate\* cell density of  $10^7$  cells)/mutation rate. When accounting for a mutation rate of  $8.9 \times 10^{-11}$  per bp per generation [85] and a cell density of  $10^7$  cells within a community, we were able to calculate a CRISPR acquisition rate in microbial communities per generation (i.e., CRISPR turnover rate, Fig. 2B) with the following formula:

$$\text{novels pacers per generation within a community} = \frac{\left( \frac{\text{number differing SNVs}}{\text{number of differing spacers}} \right)}{\text{core genome mutation rate}} \times \text{community Size}$$

cies, we additionally integrated all available genomes from the *in-house* cheese database of Agroscope called “Dialact”. The Dialact database includes 536 genomes from 9 species isolated from a wide range of samples obtained from Swiss cheese and Swiss cheese starter cultures (Suppl. Fig. 1B). Due to limited metadata information associated with many of the genomes obtained from NCBI, we cannot exclude that some of the strains included in our analysis may have not been isolated from cheese.

### Detection of CRISPR and defense mechanisms

The genome assemblies were annotated with CRISPR-CasFinder (v4.2.20-1) [14]. The raw JSON outputs were parsed and quality filtered with custom Python and R

### Detection of CRISPR arrays in metagenomes

In order to identify CRISPR in the metagenomic samples, the raw metagenomic reads were processed using CRASS (v0.3.12) [72] with default parameters. CRISPR spacers, repeats and flanking sequences were then extracted. Each spacer was automatically annotated with coverage information as well as the spacer count per million reads for the whole sample. Spacers with a coverage of 1 were removed as well as spurious spacers of length smaller than 15 bp

### Sequences clustering

Repeats, often referred to as consensus repeats or direct repeats, are not necessarily identical within an array [68] as well as spacers, which do not need a perfect match

with their target sequence to be cleaved [71]. Thus, repeat and spacer sequences were clustered using CD-HIT-EST (v4.8.1) [42] using a 100% and an 80% identity threshold [53]. Eighty percent identity clusters were added in the database separately for genomic repeats, genomic spacers, metagenomic repeats, and metagenomic spacers. Venn diagram representations (Fig. 3) have been done with BioVenn [32] and adapted within inkscape, while strictly keeping the calculated surfaces.

#### Quantification of spacers and repeats in metagenomes

The observed number of spacers and repeats was calculated by dividing the absolute number of spacers and repeats by the total number of reads and later by the total number of bacterial species (richness). Here we also included the data previously described for the human microbiome [53].

#### Mapping spacer sequences to reads

In order to quantify the proportion of spacers and protospacers in the metagenomes, we created for every spacer a *repeat-spacer-repeat* sequence and mapped all metagenomic samples individually against this reference. Reads mapping only to the ~ 40 bp spacer area were assigned to protospacers, whereas reads mapping to the repeat and spacer area were assigned to the CRISPR array. The sum of these counts was normalized by the reads count for each sample.

#### Spacers target mapping in viral and bacterial databases

BLASTn (megablast) was performed using an e-value cutoff of 0.01, word-size of 4, keeping only 1 query-subject alignment per pair (max\_hsps = 1) and 1 aligned sequence (max\_target\_seqs = 1) (v2.5.0+). Three databases were used, IMG/VR (v.3.0) [66], nt (non-redundant nucleotide), and the PLSDB plasmid database (v2020\_11\_19) [27]. To rule out putative prophage and annotate gene targets, we further blasted to the nr (non-redundant protein) and screened for phage genes and annotated with eggno3 [31]. Further CRISPR mappings were ruled out by scanning the 2 kb up and downstream of the target site for CRISPR repeat annotation in the corresponding genbank files. The assignment to vOTUs was made on the basis of the IMG/VR (v.3.0) [66] mapping.

#### Statistics, scripts and data

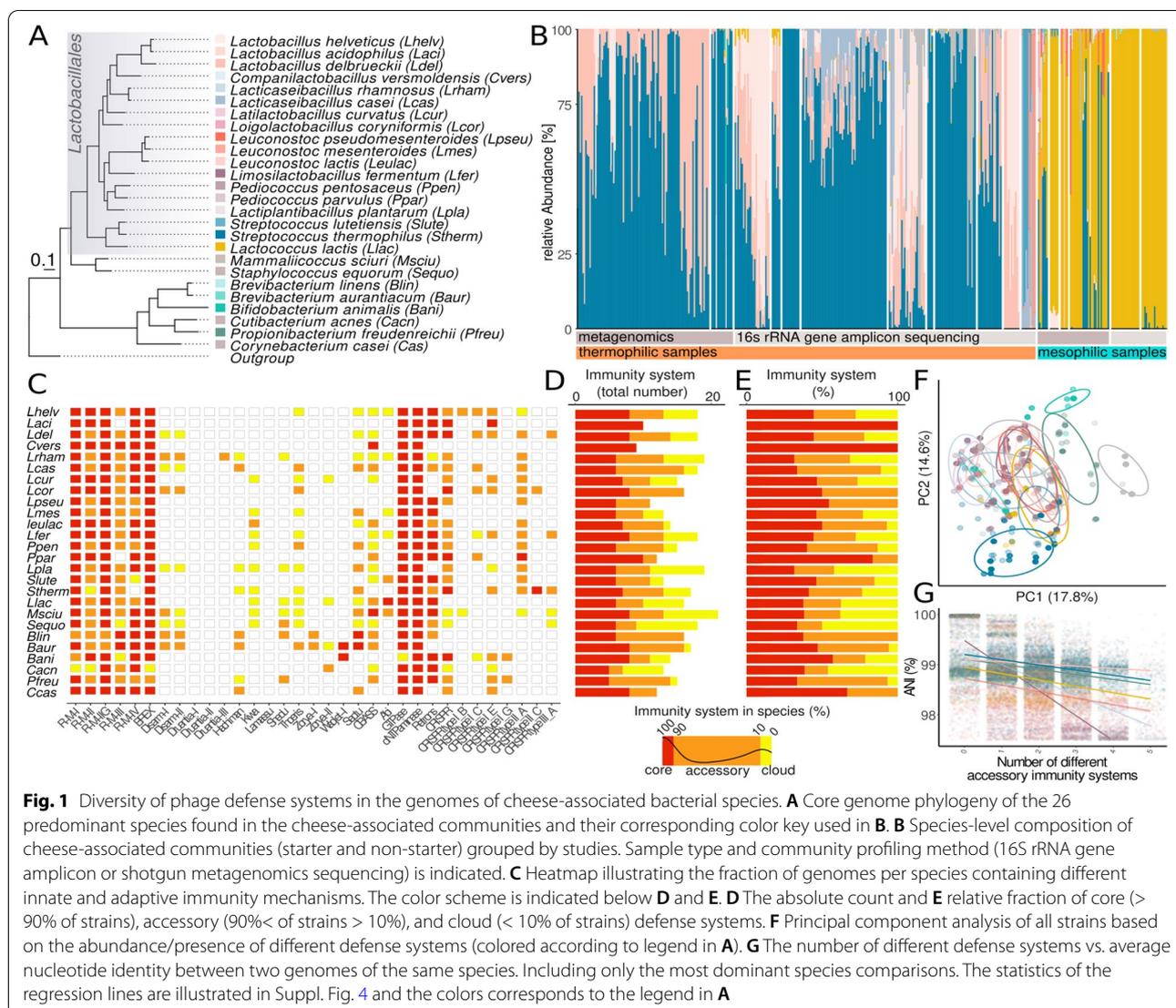
All statistics were done within R [64] and ggplot2 [84]. All code for the bioinformatics is available here (<https://github.com/ThibaultSchowing/CRISPRscope>) and the figures here (<https://github.com/Freevini/CRISPRscope>). The data is deposited on zenodo (10.5281/zenodo.6444686).

## Results and discussion

### Strain-specific immune gene arsenals across cheese-associated bacteria

In order to obtain an overview of the diversity of phage defense systems in cheese-associated bacteria, we first determined which taxonomic groups are prevalent across cheese-associated communities by profiling 480 community samples from 18 different studies (Suppl. Table 1). These included mesophilic (cooked at lower temperatures) and thermophilic (cooked at higher temperatures) cheese starter cultures as well as samples from ripened cheese selected from the FoodMicrobionet database [56]. We excluded cheese rind samples from our analysis, because they consist of highly variable microbial communities with a more complex ecology [86]. Overall, we included 322 16S rRNA gene amplicon sequencing and 158 shotgun metagenomics datasets (Fig. 1A, B). A total of 196 species were identified of which 26 were present in > 2% of all samples with a median abundance of > 0.1% (Fig. 1B and Suppl. Fig. 1). The large majority of these species were from the order *Lactobacillales* with *Lactococcus lactis* dominating mesophilic and *Streptococcus thermophilus* dominating thermophilic cheese samples (Suppl. Fig. 2). These two species make up on average 72% (sd = 25%) of the community (Suppl. Fig. 1A), The non-*Lactobacillales* species were present in less than 10% of the samples and are likely minor members, especially in non-pasteurized cheese communities (Suppl. Fig. 1A).

We next retrieved 2778 genomes of the 26 predominant species from NCBI and the *in house* genomic database of Agroscope (Suppl. Fig. 1). While the genomic data from our in-house database exclusively originates from strains isolated from cheese, the metadata associated with genomes obtained from NCBI was limited so that we cannot exclude that some strains included in our analysis may have been isolated from other environments than cheese. The genomes were screened for the presence of homologs of 25 different phage defense systems using a hmm-search approach. In total, 17,565 innate immune systems and 1'972 CRISPR-Cas systems were identified. Restriction/modification (RM) systems (Fig. 1C–E), GTPases, deaminases, and retrons were common innate immune systems across almost all species. On the contrary, only a few species harbored homologs of e.g. Abi systems (Suppl. Fig. 3). All species contained CRISPR-Cas systems with the exception of *Brevibacterium aurantiacum*, *Brevibacterium linens*, *Companilactobacillus versmoldensis*, *Lactococcus lactis*, and *Leuconostoc mesenteroides* (Fig. 1C). None of the defense systems were found to be specific to a given species. Moreover, species did not cluster by defense



systems composition (Fig. 1F) suggesting overlapping defense strategies across species. On average we found 7.5 (sd = 1.4, Fig. 1D) defense systems per genome with all species harboring more defense systems than previously described for bacteria of other environments [78]. Considering that the number of defense systems reflects the extent of phage pressure in a given environment, this supports the idea that phages are prevalent in cheese-associated communities [78].

Notably, only 49% (sd = 19%) of the defense systems detected within a species were shared among all strains of that species (core), 35% (sd = 14%) were shared between 10 and 90% of the strains (accessory), and 25% (sd=13%) were present in less than 10% of the strains (cloud) (Fig. 1D, E). Similarly, while CRISPR-Cas systems were detected in almost all species, only 49% (sd = 40%) of the strains of these species encoded CRISPR-Cas systems

in their genome (Fig. 1E). The number of shared innate immune systems decreased with increasing genomic divergence as measured by pairwise average nucleotide identity (ANI) (Fig. 1G, Suppl. Fig. 4).

No correlation was found between the presence of different innate immune systems across the analyzed genomes (Suppl. Fig. 5). However, we did find that species without CRISPR-Cas systems harbored significantly more innate defense mechanisms than species with CRISPR-Cas (unpaired Wilcoxon test,  $p$  value < 0.001, Suppl. Fig. 6). Interestingly, this pattern was reversed when comparing strains of the same species (Wilcoxon test,  $p$  value < 0.001, Suppl. Fig. 6); i.e., strains without CRISPR tended to have fewer innate defense mechanisms than strains with CRISPR. As phage defense systems are costly to maintain, the loss of such genes could be the result of extensive passaging of certain strains in phage-deprived environments,

especially as many of the sequenced genomes come from laboratory strains.

### Rapid turnover of CRISPR spacers in nearly identical strains of cheese-associated bacteria

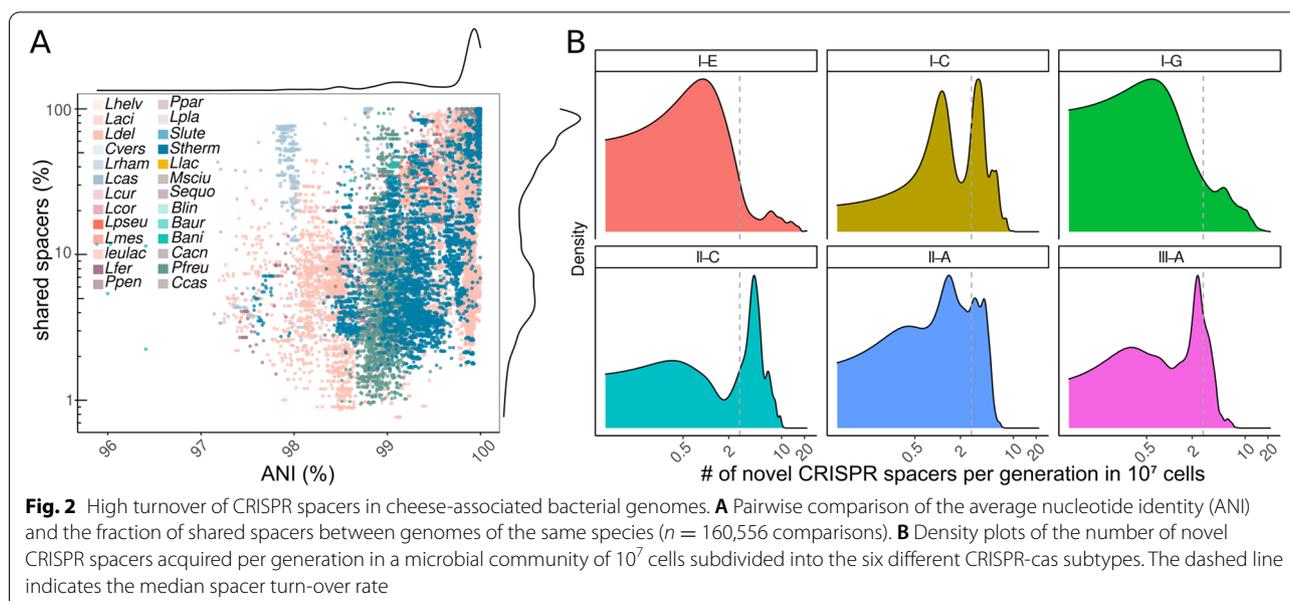
To assess if the high diversity of the innate immune system is paralleled by a similarly high extent of diversity in adaptive immunity, we identified all CRISPR spacers across all 2'778 genomes of the 26 predominant cheese-associated bacterial species. We detected a total of 1'972 CRISPR arrays containing 16,506 unique spacers (Suppl. Fig. 7). The number of spacers per array (median = 25.83, sd = 19.29) varied within and across species (Suppl. Fig. 7) and also between the different CRISPR-Cas subtypes detected in the analyzed datasets (Suppl. Fig. 8).

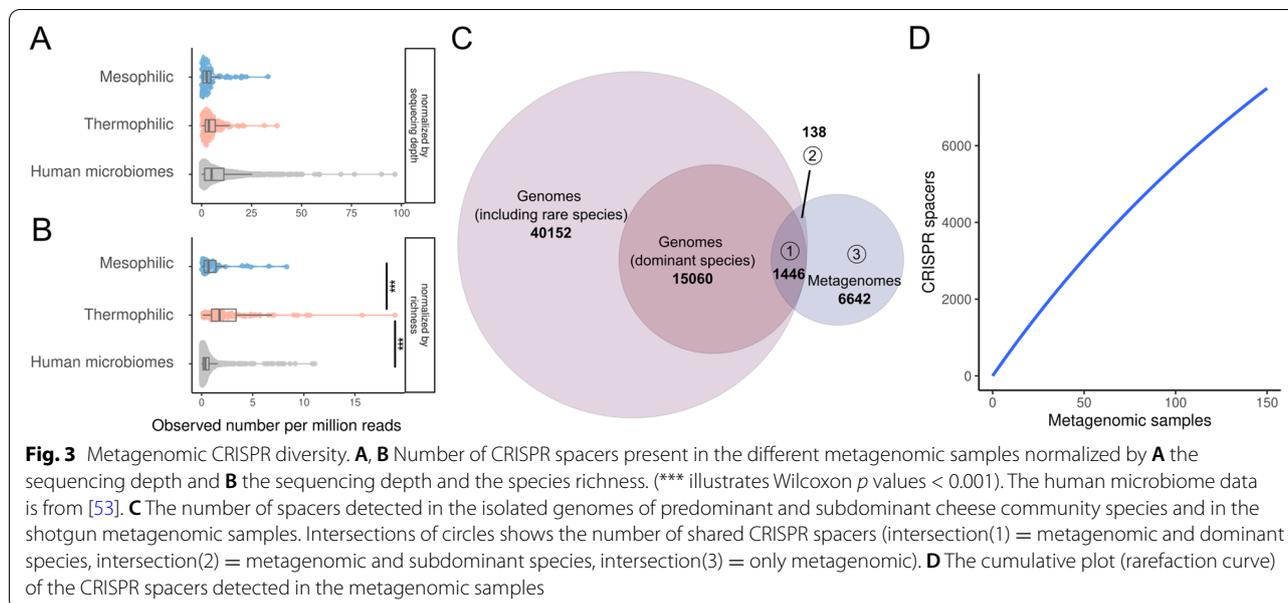
As expected, no spacers were shared between genomes belonging to different species or between arrays from different CRISPR-Cas subtypes. However, also within species (ANI > 95%), only 41% of all genome pairs shared spacers. Moreover, the fraction of shared spacers between two genomes was relatively small (median = 9.2%, sd = 33.8%). Even nearly identical strains with an ANI > 99.5% shared on median as little as 55% of their spacers (sd = 27%). The proportion of shared spacers decreased with genomic distance (as measured by decreasing pairwise ANI) (slope = 41.3,  $R^2 = 0.48$ , Fig. 2A) in all species (Suppl. Fig. 9). This is in line with the observed decrease in shared innate immune systems with increasing genetic distance between strains. Although there seems to be a signature of vertical evolution over very short evolutionary

timescales, the results overall suggest that most spacers are not maintained for very long but are continuously gained and lost. The only exception concerns a subset of divergent *L. casei* genomes (ANI ~ 98%, Fig. 2A), which contained plasmids carrying a CRISPR array sharing > 25% of the spacers.

To obtain an estimate of the CRISPR spacer turnover rate, we calculated how many novel CRISPR spacers would be acquired in each new generation in a community of defined size. We divided the number of unique spacers per genome by the number of nucleotide differences for each genome pair with an ANI > 99%. We found that one unique CRISPR spacer corresponds on average to 1355 core genome single nucleotide variants (SNVs) (sd = 2492). When accounting for a core genome mutation rate of  $8.9 \times 10^{-11}$  mutation per base-pair per generation [85] and a cell density of  $10^7$  cells per community, we calculated that the CRISPR turnover rate corresponds to a median of 2.8 CRISPR spacers per generation (Fig. 2B). This suggests that the acquisition of novel CRISPR spacers is extremely rapid and that at every bacterial generation several novel spacers can be incorporated. However, as we only considered fixed mutations, we may underestimate the time of divergence between these genomes which would result in a lower turnover rate. Indeed, previous estimates of CRISPR acquisition rates based on experimental data were ~ 1 magnitude lower (< 0.1 spacer/generation in [74] or ~ 0.5 spacer/generation in [55]).

Interestingly, we observed marked differences in spacer turnover rates between different CRISPR-Cas subtypes but not between different species. More specifically, the





spacer turnover rates of arrays belonging to CRISPR-Cas subtypes I-E, I-G, and III-A were generally lower than the median turnover rate (2.8 spacers per generation). On the contrary, spacers turnover rate of the CRISPR-Cas subtype II-C was generally higher than the median turnover rate. Finally, the turnover rates of arrays belonging to CRISPR-Cas subtypes I-C and II-A showed a bimodal distribution with some arrays having high and others low rates of spacer turnover (Fig. 2B). Variation in spacer turnover rate has been previously observed and was suggested to reflect differences in phage pressure acting on the different strains [2]. Our data suggest that it also depends, at least partially, on intrinsic properties of the CRISPR-Cas subtype within the species (Suppl. Fig. 10).

#### Extensive CRISPR spacer diversity in metagenomic datasets of cheese-associated communities

The high turnover rates of CRISPR spacers estimated from the isolate genomes suggests the presence of high levels of CRISPR spacer diversity within and across cheese-associated communities. Based on the identification of flanking CRISPR repeats we extracted 8226 non-redundant full-length spacer sequences from the Illumina reads of the 158 shotgun metagenomic samples presented in Fig. 1B. On average 5.24 (sd = 6.23) spacers per million reads were identified per sample (Fig. 3A). There was no difference in CRISPR spacer diversity between mesophilic (cheese that is made at  $\sim 30$  °C) and thermophilic (cheese that is made at  $> 45$  °C) communities. This was surprising, as mesophilic cheese communities are dominated by the non-CRISPR containing

species *L. lactis* and *Leuc. mesenteroides*, and suggests that subdominant community members harbor a high number of CRISPR spacers.

We compared our dataset to a previously published analysis of CRISPR spacer diversity in human microbiomes [53] and found that the diversity of CRISPR spacers in cheese-associated communities and the human microbiomes is not significantly different from each other (Fig. 3A). However, when accounting for the higher species diversity in the human microbiome (i.e., by normalizing to the richness of each community), we found that thermophilic communities harbor more CRISPR spacers per species than human microbiomes (Fig. 3B). This is in line with previous studies, which had shown that high CRISPR-Cas diversity is associated with anaerobic growth, high temperatures and non-host environments [10, 50, 82], all of which are characteristics of cheese environments.

Surprisingly, only a small fraction of the spacers identified across the metagenomic datasets (17.6%) matched to spacers detected in the 2778 isolate genomes of the predominant cheese-associated species (Fig. 3C). When also considering genomes of subdominant species, this number increased only little to 1'584 (19.3%) matching CRISPR spacers. As no other species were detected in the analyzed metagenomes (Fig. 1A, B), we conclude that the large majority (80.7%) of the metagenomic CRISPR spacers corresponds to within-species diversity not covered by the currently available isolate genomes. Further, we found little overlap in CRISPR spacer diversity between metagenomes. Only very few spacers (8%) were shared among more than two metagenomes (Suppl.

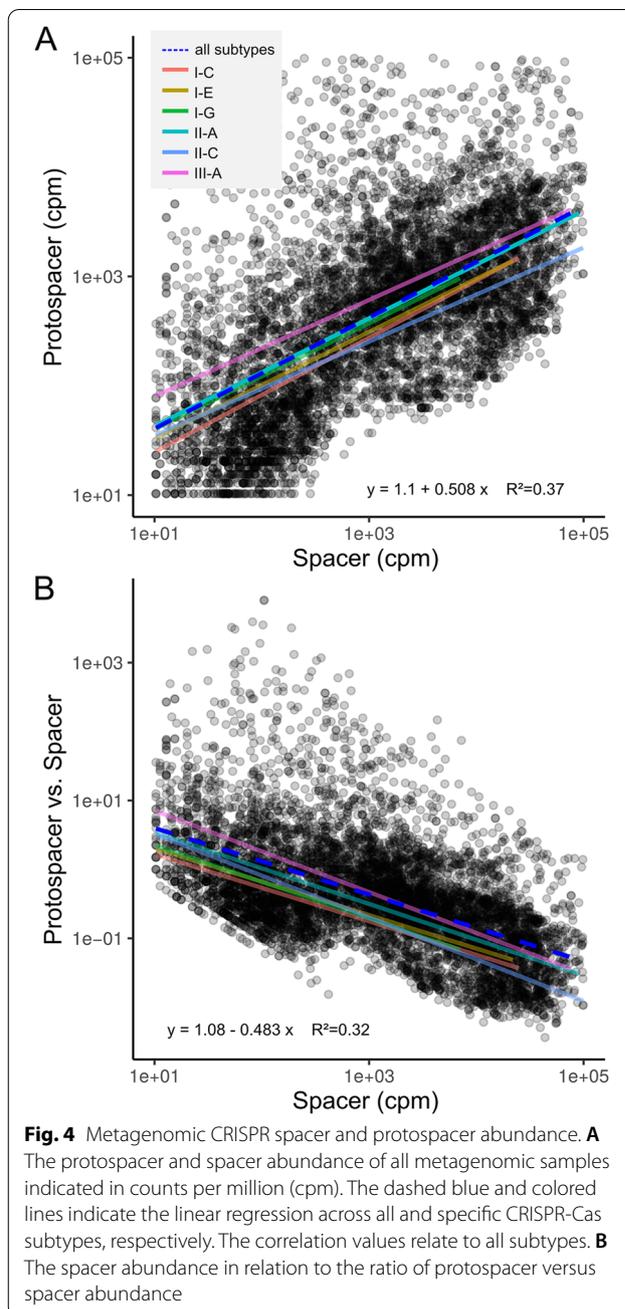
Fig. 11). Moreover, a rarefaction curve analysis showed that with the addition of each metagenomic sample, new spacers are being discovered (Fig. 3D). Together, this indicates that the CRISPR spacer diversity within and between cheese-associated communities is extensive and that we have only detected a fraction of this diversity in our study.

#### CRISPR spacer abundances correlates with target abundance

If the spacers have any ecological relevance [12, 29], one would expect to find a positive correlation between the abundances of phages and their matching spacer sequences. To quantify both spacer and target (i.e., phage) abundance directly from the metagenomic samples we mapped the metagenomic reads of the 158 datasets to all dereplicated repeat-spacer-repeat sequences identified in the 2778 isolate genomes. As spacers are usually shorter than Illumina reads, reads containing spacer and repeat sequence were considered to come from a CRISPR array (hereafter referred to as spacer reads). In contrast, reads mapping to only spacer sequences were considered to come from a target (e.g., a phage, hereafter referred to as protospacer reads) (Suppl. Fig. 12).

In each metagenome, we identified between 41 and 1961 repeat-spacer-repeat sequences, which recruited at least one spacer or protospacer read. In many cases, only protospacers (41.3%) or spacer (33.6%) reads were identified. For the remaining 25.1% of cases, we found a positive correlation between spacer and matching protospacer abundance (slope = 0.51,  $R^2 = 0.37$ , Fig. 4A), independent of the CRISPR-Cas subtype or the metagenomic sample (Suppl. Figs. 12 and 13). This is in line with previous results obtained for the Earth Microbiome Project [50] and supports the idea that spacers targeting highly abundant phages are under positive selection and thus dominant in the community [30]. Notably, in our previous study focusing on a single cheese starter culture, we had found the opposite pattern. This may be explained by the fact that a single phage dominated this community, causing chronic infections and thereby overcoming CRISPR-based immunity [74]. Interestingly, with increasing spacer abundance the ratio of protospacer to spacer abundance decreased (slope =  $-0.48$ ,  $R^2 = 0.32$ , Fig. 4B), suggesting that highly abundant spacers are effective in decreasing protospacer abundance. A similar correlation has previously been described for the viral fraction outside of the cells measured by virus-microbe-ratios (VMR) [36, 83].

Further, we wanted to see if the CRISPR spacers are genetically linked and verify if the entire strain or the



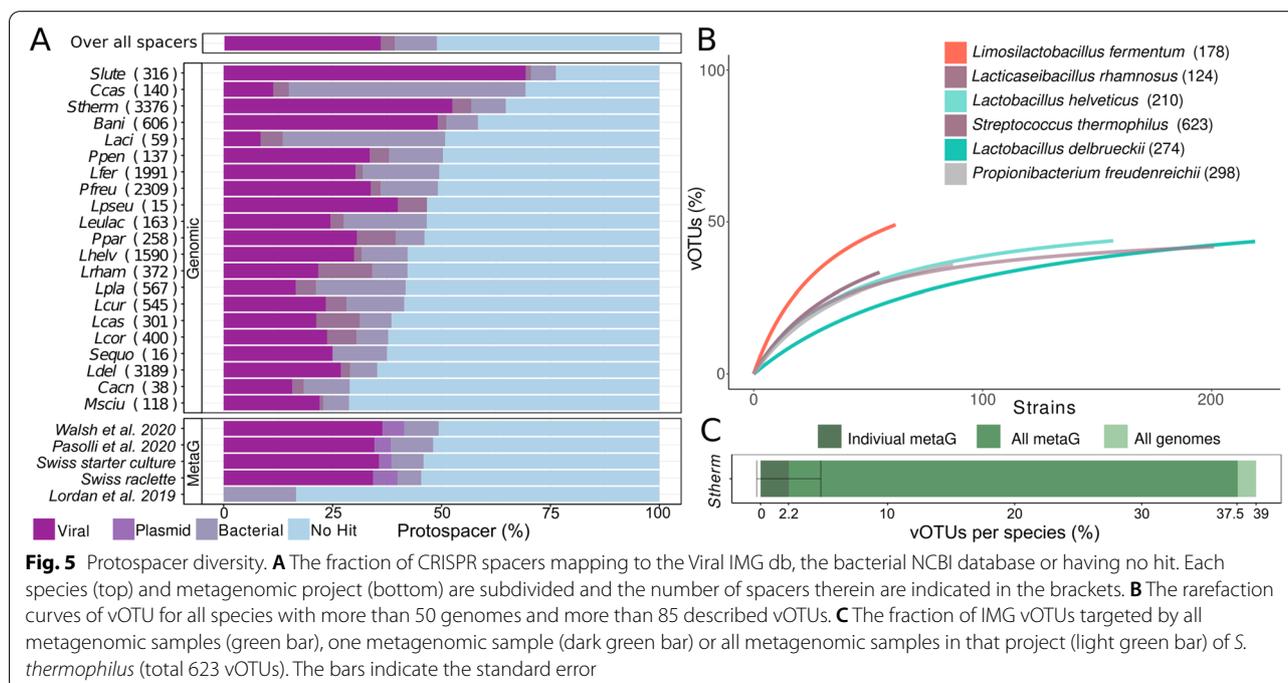
individual spacers are the level of selection. Therefore we looked at the spacer abundances within one metagenomic sample. The spacer abundance within a single metagenome clearly has a non-normal distribution, with few spacers being abundant, while the majority are of low frequency (Suppl. Fig. 14). This indicates that individual spacers can sweep through the population rather than the entire strain being selected for.

### No complete pan-immunity by CRISPR defense

For a large number of the spacers identified in the isolate genomes, we did not find a corresponding protospacer sequence in the metagenomes. However, in 41.3% of the cases (see above), we identified a protospacer, suggesting that not for all targets present in a given sample, CRISPR spacers have been acquired at a detectable level. To determine the identity of sequences targeted by the CRISPR-Cas system, we searched the 63,438 CRISPR spacers identified in the 2778 genomes and 158 metagenomes against the NCBI nucleotide database (i.e., bacteria), the IMG viral database (i.e., phages), and the plasmid PLSDDB database [27]. In total, 49% of the CRISPR spacers had a hit (maximal 2 mismatches) in at least one of these databases. The majority of spacers matched to phages (36%) followed by bacterial chromosomes (9%) and plasmids (3%) (Fig. 5A). These proportions are in a similar range to what has previously been reported for other bacteria [61]. Most of the sequences matching bacterial chromosomes and plasmid were found in genes belonging to the COG (Clusters of Orthologous Groups) category “Replication and repair”, which includes selfish elements such as transposons (Suppl. Fig. 15). The fraction of spacers targeting phages varied across species. In case of *S. thermophilus*, 61% of the spacers mapped to known phages, while for other species, e.g., *L. delbrueckii* or *L. helveticus*, only 30% and 36%, respectively, mapped to known phages, probably reflecting the extent of research conducted on the phage diversity of different bacterial species.

To assess the range of phages targeted by a given isolate, we categorize the phages into discrete viral operational taxonomic units (vOTU) inferred by the viral IMG database [65]. Further, we limited this analysis to the genomes of the six bacterial species with the largest diversity of phages represented in the database (i.e., > 85 described vOTU), namely phages of *S. thermophilus*, *L. delbrueckii*, *L. helveticus*, *L. fermentum*, *L. rhamnosus*, and *P. freudenreichii*. Within these species, we observed that each strain targets on average 9 vOTUs (sd = 8) (Suppl. Fig. 16), which corresponds to 2.5% (sd = 0.6%) of the known phage diversity of these species. This is in line with a recent study where they modeled that 1–10% of the total phage diversity is covered by the CRISPR-Cas system of a single strain [9]. Notably, most isolate genomes (83%) harbored only a single spacer against a given vOTU (Suppl. Fig. 17) with no major differences between species (Suppl. Fig. 18). This observation is in contrast to what has been observed in laboratory studies of phage-bacteria coevolution [13, 29] or in cases of chronic phage infections in *S. thermophilus* [74], where multiple spacers targeting the same vOTU were often found to be integrated into the CRISPR array.

Our results seem to suggest that the spacer repertoire of a given strain is aimed at targeting a broad range of different phages rather than being specialized towards a single vOTU. To assess whether the presence of several strains with diverse CRISPR spacers provides pan-immunity against a broad range of phages, we conducted



a rarefaction analysis of the CRISPR-vOTU matches identified in the isolate genomes. For all six analyzed species, the curve rapidly flattened with no more than 50% of the known vOTU having matching CRISPR spacers in the analyzed isolate genomes (Fig. 5B). For example for *S. thermophilus* only 39% of all phage vOTUs had matching spacers across the isolate genomes. This indicates that no combination of isolates results in complete CRISPR-based immunity against all known phages of a given species. Analysis of the metagenomic CRISPR spacers mapping to *S. thermophilus* phages confirmed these results: only 37.6% of the known *S. thermophilus* phages were targeted by CRISPR spacers identified across the 158 metagenomic datasets (Fig. 5C). Even fewer phages were targeted by spacers identified in individual metagenomic datasets (mean = 2.5%, sd = 2.2, Fig. 5C). Phages not targeted by any spacer did not seem to be rare as the vOTU clusters were not necessarily smaller than the clusters of targeted vOTU (Suppl. Fig. 19). Moreover, they did not contain more anti-CRISPR genes than phages that had matching CRISPR spacers in the communities (Suppl. Fig. 20). It is possible that these phages are integrated as prophages in the bacterial genomes and thereby avoid CRISPR-based immunity or that the bacteria and phages have not encountered each other due to spatial population structure or segregation into different communities that have not yet been sampled [79].

## Conclusions

Previous studies have used shotgun metagenomics to characterize CRISPR diversity in bacteria found on the human body, in the ocean or the soil [50, 53]. Cheese-associated communities are much simpler than these previously analyzed communities [20, 28]. They contain much fewer species and are propagated in relatively stable environments [57, 73]. A large amount of genomic data is available for cheese-associated communities as they are established experimental model systems to study bacteria-phage interactions. This allowed us to assess the intraspecific diversity and evolutionary dynamics of phage defense systems across a wide range of bacterial species and communities by analyzing publicly available genomes and metagenomic datasets. We found extensive diversity in innate and adaptive immune defense mechanisms across cheese-associated bacteria, despite the overall little genomic diversity present in these communities. Phages are known to be common in these environments and pose a risk for the cheese making process [40]. However, the extent of defense systems we have found seems to exceed the diversity present in previously analyzed ecosystems, which is surprising given that cheese

communities are closed systems with little opportunities for migration/invasion.

Our analysis revealed that innate immune systems were distributed in a strain-specific manner across the analyzed genomes of cheese-associated communities. They are part of the accessory genome and are rapidly gained and lost. Likewise, CRISPR spacer repertoires varied substantially across nearly clonal isolates and the amount of CRISPR spacers present in the metagenomic datasets seemed infinite. Accordingly, our estimation of the CRISPR spacer turnover rates suggested rapid gain and loss of CRISPR spacers with important differences found between different CRISPR-Cas subtypes but not necessarily species. The ecological relevance of the identified diversity is highlighted by the finding that metagenomic CRISPR spacers matching abundant target sequences (e.g., phages) were also abundant in the corresponding metagenomic sample which indicates specific bacteria-phage responses in terms of their ecological dynamics.

Our observations align with the pan-immunity model proposed for understanding the evolutionary ecology of innate immune systems [6] and suggest that this model can be extended to CRISPR-based adaptive immunity, as the three key points of this model seem to be fulfilled. First, we show that there is standing genetic diversity in CRISPR spacer diversity in cheese-associated bacteria and communities. Second, similar as for the horizontal mode of transfer proposed for the innate immune systems [33], the high turnover of spacers detected in our study reflects that novel defensive repertoires can be rapidly acquired (and lost) in the population. Third, CRISPR spacers seem to be selected for and hence functional as their abundance correlates with phage abundance.

On the contrary, for about 50% of the phages that were isolated from cheese-associated communities, we could not identify any matching CRISPR spacers across the genomes/metagenomic datasets. This may hint at the combined importance of both innate and adaptive immune systems and could suggest that CRISPR is not effective against all phages. It is possible that these phages avoid CRISPR targeting, making them interesting candidates to look for novel anti-CRISPRs mechanisms. Alternatively, the untargeted phages could be incompatible because of surface modifications, which are known to be crucial for phage resistance in the cheese-associated species *Streptococcus thermophilus* [49, 77]. Overall, our results allow us not only to understand the evolutionary ecology of phage-bacteria interactions but could also be instrumental in improving the protection of cheese starter culture by developing phage-based therapy [23] as a protection against pathogens [22] or invasive strains [63].

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-022-01328-6>.

**Additional file 1: Table S1.** Sample description of shotgun metagenomic samples.

**Additional file 2: Table S2.** Sample description of 16S community analysis.

**Additional file 3: Supplementary Figure 1.** A) Mean abundance and prevalence of the different species in the metagenomes illustrated in main Fig. 1A. B) The number of genomes for the different species downloaded from NCBI and in house cheese database. **Supplementary Figure 2.** Species environment assignment. The relative abundance of the dominant species in the mesophilic and thermophilic metagenomic samples (as indicated in Fig. 1A). Also the colors indicate the species assignment to either thermophilic, mesophilic or generalist species. **Supplementary Figure 3.** Number of ABI systems per strain across different species. **Supplementary Figure 4.** The number of different defense systems vs. average nucleotides between two genomes of the same species. Including only the most dominant species comparisons. This figure corresponds to Fig. 1G in the main text but includes the regression statistics. **Supplementary Figure 5.** The correlations between different phage defense systems. The heatmap illustrates the correlation coefficient (see legend to the right). **Supplementary Figure 6.** Number and variation of phage defense systems between A) CRISPR containing and no CRISPR containing species and B) within the CRISPR-containing species between CRISPR containing and no CRISPR containing strains. **Supplementary Figure 7.** Characteristics of CRISPR spacer and CRISPR repeats in the isolates of the different cheese-associated bacterial species illustrated by the following panels i) spacer GC content, ii) repeat GC content, iii) host GC content, iv) spacer length and v) number of spacers per strain. **Supplementary Figure 8.** The total number of CRISPR spacers per array is illustrated for Cas subtypes independent of species. Both the distribution as well as the actual number (points) are illustrated. **Supplementary Figure 9.** The ANI vs. shared CRISPR spacer plot shown for the predominant species separately. Only boxes with more than 30 comparisons are shown. **Supplementary Figure 10.** CRISPR spacer turnover rates divided by species and CRISPR-cas subtype. The colours represent the CRISPR-cas subtype. Only the most dominant species are illustrated. **Supplementary Figure 11.** The number of metagenomes that a spacer occurred in. If a spacer is unique for a single metagenome it is represented in the first column. Thereafter the columns illustrate the number of metagenomes a spacer is shared in. The large majority of spacers occur only in one or a few metagenomic samples. **Supplementary Figure 12.** Normalized spacer vs. protospacer abundances (counts per million) subdivided by the different metagenomic bioprojects. The blue line indicates the illustrated linear regression with the statistics represented at the top of the figures. **Supplementary Figure 13.** Normalized spacer vs. protospacer abundances (counts per million) subdivided by CRISPR-cas subtypes. The blue line indicates the illustrated linear regression with the statistics represented at the top of the figures. **Supplementary Figure 14.** The normalized spacer abundance (normalized by the most abundant spacer in the metagenome) of all metagenomic samples. Every individual metagenome contained between 41 and 1961 spacers. The large majority of spacers are of very low abundance. Apart from the accumulation of spacers at the low abundance spectrum there does not seem to be a large accumulation of spacers throughout the figure. This would have illustrated a dominant strain. **Supplementary Figure 15.** COG categories of the 859 proteins that were targeted by spacers (i.e. protospacers). **Supplementary Figure 16.** The number of vOTUs targeted by a given genome of the different species. Only depicted for the most abundant species and vOTUs. **Supplementary Figure 17.** The number of CRISPR spacers targeting the same vOTU in one genome. The percent is indicated above the bar. The large majority of spacers within a CRISPR array (~83%) target only a single vOTU. **Supplementary Figure 18.** Similar as the previous supplement figure, the number of CRISPR spacers targeting the same vOTU in one genome subdivided by the different species. The percent is indicated above the bar. Also here, the large majority of spacers within a CRISPR

array target only a single vOTU. **Supplementary Figure 19.** The cluster size of the vOTUS (indicating the number of known/sequence phages for this vOTU) for the vOTUs which have a CRISPR match in comparison to the vOTUS that remained unobserved (no target identified in the genomes). The Wilcoxon p-value is non-significant and indicated in the figure. **Supplementary Figure 20.** The number and fraction of vOTUs containing an anti-CRISPR sequence divided by vOTUs targeted by CRISPR or not targeted (no hit) that contain anti-CRISPR genes.

### Acknowledgements

We thank Aline Cuénod, Germán Bonilla-Rosso, and Malick Ndiaye for the feedback and discussions of the manuscript. Further we also thank Noam Shani, Petra Lüdin, Verena Schünemann, Abigail Bouwman, Meral Turgay, Marco Meola, Johann Bengtsson-Palme, and the Functional Genomics Center Zurich for the cheese samples sequencing and making their datasets available for this study.

### Authors' contributions

VS designed and executed the project and wrote the manuscript. TS executed the project and wrote the manuscript. HC, UvA RS, and RB advised the project. PE designed the project and wrote the manuscript. All authors gave feedback on the manuscript. The author(s) read and approved the final manuscript.

### Funding

The project was supported by the Master's program of the Universities of Fribourg and Bern, an ERC Starting Grant (MicroBeeOme, no. 714804), the NCCR microbiomes, a National Centre of Competence in Research and funded by the Swiss National Science Foundation (grant number 180575), a Swiss National Science Foundation project grant (31003A 160345) and Agroscope Switzerland.

### Availability of data and materials

All code for the bioinformatics is available here (<https://github.com/Thibault-Schowing/CRISPRscope>) and the figures here (<https://github.com/Freevini/CRISPRscope>). The data is deposited on zenodo (10.5281/zenodo.6444686).

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

All authors have approved the manuscript for submission and publication.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland. <sup>2</sup>Agroscope, Bern, Switzerland. <sup>3</sup>Interfaculty Bioinformatics Unit and Swiss Institute of Bioinformatics, University of Bern, Bern, Switzerland. <sup>4</sup>Institute for Integrative Biology, ETH Zürich, Zürich, Switzerland.

Received: 23 April 2022 Accepted: 17 July 2022

Published online: 27 August 2022

### References

- Alessandria V, Ferrocino I, De Filippis F, Fontana M, Rantsiou K, Ercolini D, et al. Microbiota of an Italian grana-like cheese during manufacture and ripening, Unraveled by 16S rRNA-based approaches. *Appl Environ Microbiol.* 2016;82:3988–95.
- Almeida GMF, Hoikkala V, Ravantti J, Rantanen N, Sundberg L-R. Mucin induces CRISPR-Cas defence in an opportunistic pathogen. *bioRxiv.* 2021:2021.08.10.455787 Available from: <https://www.biorxiv.org/content/10.1101/2021.08.10.455787v1.abstract>.
- Ankenbrand MJ, Keller A. bcgTree: automatized phylogenetic tree building from bacterial core genomes. *Genome.* 2016;59:783–91.

4. Bernheim A, Calvo-Villamañán A, Basier C, Cui L, Rocha EPC, Touchon M, et al. Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria. *Nat Commun.* 2017;8:1–9.
5. Bernheim A, Millman A, Ofir G, Meitav G, Avraham C, Shomar H, et al. Prokaryotic viperins produce diverse antiviral molecules. *Nature.* 2021;589 Available from: <https://pubmed.ncbi.nlm.nih.gov/32937646/>.
6. Bernheim A, Sorek R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat Rev Microbiol.* 2020;18:113–9.
7. Birkholz N, Jackson SA, Fagerlund RD, Fineran PC. A mobile restriction–modification system provides phage defence and resolves an epigenetic conflict with an antagonistic endonuclease. *Nucleic Acids Res.* 2022. <https://doi.org/10.1093/nar/gkac147>.
8. Bottari B, Levante A, Bancalari E, Sforza S, Bottesini C, Prandi B, et al. The inter-relationship between microbiota and peptides during ripening as a driver for parmigiano reggiano cheese quality. *Front Microbiol.* 2020;11:581658.
9. Bradde S, Nourmohammad A, Goyal S, Balasubramanian V. The size of the immune repertoire of bacteria. *Proc Natl Acad Sci U S A.* 2020;117:5144–51.
10. Burstein D, Sun CL, Brown CT, Sharon I, Anantharaman K, Probst AJ, et al. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat Commun.* 2016;7:10613.
11. Calasso M, Ercolini D, Mancini L, Stellato G. Relationships among house, rind and core microbiotas during manufacture of traditional Italian cheeses at the same dairy plant. *Food Microbiol.* 2016;54:115–26.
12. Childs LM, England WE, Young MJ, Weitz JS, Whitaker RJ. CRISPR-induced distributed immunity in microbial populations. *PLoS One.* 2014;9:e101710.
13. Common J, Morley D, Westra ER, van Houte S. CRISPR-Cas immunity leads to a coevolutionary arms race between *Streptococcus thermophilus* and lytic phage. *Philos Trans R Soc Lond B Biol Sci.* 2019;374:20180098.
14. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 2018;46:W246–51.
15. De Filippis F, La Stora A, Stellato G, Gatti M, Ercolini D. A selected core microbiome drives the early stages of three popular Italian cheese manufactures. *PLoS One.* 2014;9:e89680.
16. De Filippis F, Parente E, Ercolini D. Metagenomics insights into food fermentations. *Microb Biotechnol.* 2017;10:91–102.
17. De Pasquale I, Di Cagno R, Buchin S, De Angelis M, Gobbetti M. Spatial distribution of the metabolically active microbiota within Italian PDO Ewes' Milk Cheeses. *PLoS One.* 2016;11:e0153213.
18. Dimitriu T, Szczelkun MD, Westra ER. Evolutionary ecology and interplay of prokaryotic innate and adaptive immune systems. *Curr Biol.* 2020;30:R1189–202.
19. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery of anti-phage defense systems in the microbial pan-genome. *Science.* 2018;359 Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6387622/>.
20. Dugat-Bony E, Garnier L, Denonfoux J, Ferreira S, Sarthou A-S, Bonnarne P, et al. Highlighting the microbial diversity of 12 French cheese varieties. *Int J Food Microbiol.* 2016;238:265–73.
21. Duru IC, Laine P, Andreevskaya M, Paulin L, Kananen S, Tynkkynen S, et al. Metagenomic and metatranscriptomic analysis of the microbial community in Swiss-type Maasdam cheese during ripening. *Int J Food Microbiol.* 2018;281:10–22.
22. El Haddad L, Roy J-P, Khalil GE, St-Gelais D, Champagne CP, Labrie S, et al. Efficacy of two *Staphylococcus aureus* phage cocktails in cheese production. *Int J Food Microbiol.* 2016;217:7–13.
23. Endersen L, O'Mahony J, Hill C, Paul Ross R, McAuliffe O, Coffey A. Phage therapy in the food industry. *Annu Rev Food Sci Technol.* 2014;5:327–49. <https://doi.org/10.1146/annurev-food-030713-092415>.
24. Ercolini D, De Filippis F, La Stora A, Iacono M. "Remake" by high-throughput sequencing of the microbiota involved in the production of water buffalo mozzarella cheese. *Appl Environ Microbiol.* 2012;78:8142–5.
25. Erkus O, de Jager VCL, Spus M, van Alen-Boerrigter IJ, van Rijswijk IMH, Hazelwood L, et al. Multifactorial diversity sustains microbial community stability. *ISME J.* 2013;7:2126–36.
26. Escobar-Zepeda A, Sanchez-Flores A, Quirasco Baruch M. Metagenomic analysis of a Mexican ripened cheese reveals a unique complex microbiota. *Food Microbiol.* 2016;57:116–27.
27. Galata V, Fehlmann T, Backes C, Keller A. PLSDb: a resource of complete bacterial plasmids. *Nucleic Acids Res.* 2019;47:D195–202.
28. Gobbetti M, Di Cagno R, Calasso M, Neviani E, Fox PF, De Angelis M. Drivers that establish and assembly the lactic acid bacteria biota in cheeses. *Trends Food Sci Technol.* 2018;78:244–54. <https://doi.org/10.1016/j.tifs.2018.06.010>.
29. Guillemet M, Chabas H, Nicot A, Gatchich F, Ortega-Abboud E, Buus C, et al. Temperation and coevolution drive the evolution and the diversification of CRISPR immunity. *bioRxiv.* 2021. <https://doi.org/10.1101/2021.11.12.468349>.
30. Hille F, Richter H, Wong SP, Bratovič M, Ressel S, Charpentier E. The biology of CRISPR-Cas: backward and forward. *Cell.* 2018;172:1239–59.
31. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 2018;47:D309–14.
32. Hulsen T, de Vlieg J, Alkema W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics.* 2008;9:488.
33. Hussaini FA, Dubert J, Elsherbini J, Murphy M, VanInsberghe D, Arevalo P, et al. Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. *Science.* 2021;374:488–92.
34. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun.* 2018;9:5114.
35. Kamimura BA, De Filippis F, Sant'Ana AS, Ercolini D. Large-scale mapping of microbial diversity in artisanal Brazilian cheeses. *Food Microbiol.* 2019;80:40–9.
36. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobián-Güemes AG, et al. Lytic to temperate switching of viral communities. *Nature.* 2016;531:466–70.
37. Koonin EV, Makarova KS. Origins and evolution of CRISPR-Cas systems. *Philos Trans R Soc Lond B Biol Sci.* 2019;374:20180087.
38. Kronheim S, Daniel-Ivad M, Duan Z, Hwang S, Wong AI, Mantel I, et al. A chemical defence against phage infection. *Nature.* 2018;564:283–6.
39. Labrie SJ, Moineau S. Abortive infection mechanisms and prophage sequences significantly influence the genetic makeup of emerging lytic lactococcal phages. *J Bacteriol.* 2007;189:1482–7.
40. Lavelle K, Murphy J, Fitzgerald B, Lugli GA, Zomer A, Neve H, et al. A decade of *Streptococcus thermophilus* phage evolution in an Irish dairy plant. *Appl Environ Microbiol.* 2018;84. <https://doi.org/10.1128/aem.02855-17>.
41. LeGault KN, Hays SG, Angermeyer A, McKitterick AC, Johura F-T, Sultana M, et al. Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts. *Science.* 2021;373. <https://doi.org/10.1126/science.abg2166>.
42. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006;22:1658–9.
43. Lordan R, Walsh A, Crispie F, Finnegan L, Demuru M, Tsoupras A, et al. Caprine milk fermentation enhances the antithrombotic properties of cheese polar lipids. *J Funct Foods.* 2019;61:103507.
44. Mahony J. Cell surface polysaccharides represent a common strategy for adsorption among phages infecting lactic acid bacteria: lessons from dairy Lactococci and Streptococci. *mSystems.* 2021;6:e0064121.
45. Makarova KS, Wolf YI, Koonin EV. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* 2013;41:4360–77. <https://doi.org/10.1093/nar/gkt157>.
46. Makarova KS, Wolf YI, Snir S, Koonin EV. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J Bacteriol.* 2011;193:6039–56.
47. Marino M, Dubsky de Wittenau G, Saccà E, Cattonaro F, Spadotto A, Innocente N, et al. Metagenomic profiles of different types of Italian high-moisture Mozzarella cheese. *Food Microbiol.* 2019;79:123–31.
48. McDonald ND, Regmi A, Morreale DP, Borowski JD, Boyd EF. CRISPR-Cas systems are present predominantly on mobile genetic elements in *Vibrio* species. *BMC Genomics.* 2019;20:105.
49. McDonnell B, Hanemaaijer L, Bottacini F, Kelleher P, Lavelle K, Sadovskaya I, et al. A cell wall-associated polysaccharide is required for bacteriophage adsorption to the *Streptococcus thermophilus* cell surface. *Mol Microbiol.* 2020;114(1):31–45.
50. Meaden S, Biswas A, Arkhipova K, Morales SE, Dutilh BE, Westra ER, et al. High viral abundance and low diversity are associated with increased CRISPR-Cas prevalence across microbial ecosystems. *bioRxiv.* 2021:2021.06.24.449667 Available from: <https://www.biorxiv.org/content/10.1101/2021.06.24.449667v2>.

51. Milani C, Fontana F, Alessandri G, Mancabelli L, Lugli GA, Longhi G, et al. Ecology of lactobacilli present in Italian cheeses produced from raw milk. *Appl Environ Microbiol.* 2020;86. <https://doi.org/10.1128/AEM.00139-20>.
52. Millman A, Bernheim A, Stokar-Avihail A, Fedorenko T, Voichek M, Leavitt A, et al. Bacterial retrons function in anti-phage defense. *Cell.* 2020;183 Available from: <https://pubmed.ncbi.nlm.nih.gov/33157039/>.
53. Münch PC, Franzosa EA, Stecher B, McHardy AC, Huttenhower C. Identification of natural CRISPR systems and targets in the human microbiome. *Cell Host Microbe.* 2021;29:94–106.e4.
54. Oliveira PH, Touchon M, Rocha EPC. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* 2014;42:10618–31.
55. Paez-Espino D, Sharon I, Morovic W, Stahl B, Thomas BC, Barrangou R, et al. CRISPR immunity drives rapid phage genome evolution in *Streptococcus thermophilus*. *MBio.* 2015;6. <https://doi.org/10.1128/mBio.00262-15>.
56. Parente E, Coccolin L, De Filippis F, Zotta T, Ferrocino I, O'Sullivan O, et al. Food-Microbiome: a database for the visualisation and exploration of food bacterial communities based on network analysis. *Int J Food Microbiol.* 2016a;219:28–37.
57. Parente E, Guidone A, Matera A, De Filippis F, Mauriello G, Ricciardi A. Microbial community dynamics in thermophilic undefined milk starter cultures. *Int J Food Microbiol.* 2016b;217:59–67.
58. Parente E, Zotta T, Ricciardi A. Microbial association networks in cheese: a meta-analysis; 2021. <https://doi.org/10.1101/2021.07.21.453196>.
59. Pasolli E, De Filippis F, Mauriello IE, Cumbo F, Walsh AM, Leech J, et al. Large-scale genome-wide analysis links lactic acid bacteria from food with the gut microbiome. *Nat Commun.* 2020;11:2610.
60. Payne LJ, Todeschini TC, Wu Y, Perry BJ, Ronson CW, Fineran PC, et al. Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. *Nucleic Acids Res.* 2021;49:10868–78.
61. Pinilla-Redondo R, Russel J, Mayo-Muñoz D, Shah SA, Garrett RA, Nesme J, et al. CRISPR-Cas systems are widespread accessory elements across bacterial and archaeal plasmids. *Nucleic Acids Res.* 2021; Available from: <https://academic.oup.com/nar/advance-article-pdf/doi/10.1093/nar/gkab859/40506127/gkab859.pdf>.
62. Prakash A, Jeffries M, Bateman A, Finn RD. The HMMER web server for protein sequence similarity search. *Curr Protoc Bioinformatics.* 2017;60 Available from: <https://pubmed.ncbi.nlm.nih.gov/29220076/>.
63. del Río B, del Río B, Sánchez-Llana E, Redruello B, Magadan AH, Fernández M, et al. Enterococcus faecalis Bacteriophage 156 is an effective biotechnological tool for reducing the presence of tyramine and putrescine in an experimental cheese model. *Front Microbiol.* 2019;10. <https://doi.org/10.3389/fmicb.2019.00566>.
64. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
65. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, et al. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat Biotechnol.* 2019;37:29–37.
66. Roux S, Páez-Espino D, Chen I-MA, Palaniappan K, Ratner A, Chu K, et al. IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. *Nucleic Acids Res.* 2020;49:D764–75.
67. Ruggirello M, Dolci P, Coccolin L. Detection and viability of *Lactococcus lactis* throughout cheese ripening. *PLoS One.* 2014;9:e114280.
68. Russel J, Pinilla-Redondo R, Mayo-Muñoz D, Shah SA, Sørensen SJ. CRISPR-CasTyper: automated identification, annotation, and classification of CRISPR-Cas Loci. *CRISPR J.* 2020; Available from: <https://www.liebertpub.com/doi/abs/10.1089/crispr.2020.0059>.
69. Sant DG, Woods LC, Barr JJ, McDonald MJ. Host diversity slows bacteriophage adaptation by selecting generalists over specialists. *Nat Ecol Evol.* 2021;5:350–9.
70. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods.* 2012;9:811–4.
71. Shmakov SA, Sitnik V, Makarova KS, Wolf YI, Severinov KV, Koonin EV. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *MBio.* 2017;8. <https://doi.org/10.1128/mBio.01397-17>.
72. Skennerton CT, Imelfort M, Tyson GW. Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. *Nucleic Acids Res.* 2013;41:e105.
73. Smid EJ, Erkus O, Spus M, Wolkers-Rooijackers JCM, Alexeeva S, Kleerebezem M. Functional implications of the microbial community structure of undefined mesophilic starter cultures. *Microb Cell Factories.* 2014;13:S2. <https://doi.org/10.1186/1475-2859-13-s1-s2>.
74. Somerville V, Berthoud H, Schmidt RS, Bachmann H-P, Meng YH, Fuchsmann P, et al. Functional strain redundancy and persistent phage infection in Swiss hard cheese starter cultures. *bioRxiv.* 2021:2021.01.14.426499 Available from: <https://www.biorxiv.org/content/10.1101/2021.01.14.426499v2>.
75. Somerville V, Lutz S, Schmid M, Frei D, Moser A, Irmiler S, et al. Long-read based de novo assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *BMC Microbiol.* 2019;19:143.
76. Stellato G, De Filippis F, La Stora A, Ercolini D. Coexistence of lactic acid bacteria and potential spoilage microbiota in a dairy processing environment. *Appl Environ Microbiol.* 2015;81:7893–904. <https://doi.org/10.1128/aem.02294-15>.
77. Szymczak P, Filipe SR, Covas G, Vogensen FK, Neves AR, Janzen T. Cell wall glycans mediate recognition of the dairy bacterium *Streptococcus thermophilus* by bacteriophages. *Appl Environ Microbiol.* 2018;84(23):e01847–18.
78. Tesson F, Hervé A, Touchon M, D'humieres C, Cury J, Bernheim A. 2021. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Available from: <https://hal.archives-ouvertes.fr/hal-03383196/document>
79. Testa S, Berger S, Piccardi P, Oechslin F, Resch G, Mitri S. Spatial structure affects phage efficacy in infecting dual-strain biofilms of *Pseudomonas aeruginosa*. *Commun Biol.* 2019;2. <https://doi.org/10.1038/s42003-019-0633-x>.
80. Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev.* 2013;77:53–72.
81. Walsh AM, Macori G, Kilcawley KN, Cotter PD. Meta-analysis of cheese microbiomes highlights contributions to multiple aspects of quality. *Nat Food.* 2020;1:500–10.
82. Weissman JL, Laljani RMR, Fagan WF, Johnson PLF. Visualization and prediction of CRISPR incidence in microbial trait-space to identify drivers of antiviral immune strategy. *ISME J.* 2019;13:2589–602.
83. Weitz JS, Beckett SJ, Brum JR, Cael BB, Dushoff J. Lysis, lysogeny and virus-microbe ratios. *Nature.* 2017;549:E1–3.
84. Wickham H. Programming with ggplot2. *Use R!* 2016:241–53. [https://doi.org/10.1007/978-3-319-24277-4\\_12](https://doi.org/10.1007/978-3-319-24277-4_12).
85. Wielgoss S, Barrick JE, Tenaillon O, Cruveiller S, Chane-Woon-Ming B, Médigue C, et al. Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. *G3: Genes Genomes Genet.* 2011;1:183.
86. Wolfe BE, Button JE, Santarelli M, Dutton RJ. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. *Cell.* 2014;158:422–33.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

