GBE

Acquisition of Spacers from Foreign Prokaryotic Genomes by CRISPR-Cas Systems in Natural Environments

Katharina Sures[1], Sarah P. Esser[1], Till L.V. Bornemann[1,2], Carrie J. Moore[1], André R. Soares[1,2], Julia Plewka [1,3], Perla Abigail Figueroa-Gonzalez[1], S. Emil Ruff[4,5], Cristina Moraru[1], Alexander J. Probst [1,2,3,6,*]

[1]Environmental Metagenomics, Research Centre One Health of the University Alliance Ruhr, Faculty of Chemistry, University of Duisburg-%26%23132%3BEssen, Essen 45141, Germany

[2]Centre of Water and Environmental Research (ZWU), University of Duisburg-Essen, Essen 45141, Germany

[3]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[4]The Ecosystems Center, Marine Biological Laboratory, Woods Hole, MA 02543, USA

[5]Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA 02543, USA

[6]Centre of Medical Biotechnology (ZMB), University of Duisburg-Essen, Essen 45141, Germany

*Corresponding author: E-mail: alexander.probst@uni-due.de.

Accepted: October 17, 2025

Abstract

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) systems of bacteria and archaea provide immunities against mobile genetic elements, like viruses. In addition, protospacer analyses revealed a very specific acquisition of CRISPR spacers derived from genomes of related species or from closely interacting episymbiont genomes as recently shown for subsurface archaea. However, the origin of most of the spacers that can be found in CRISPR-Cas systems from natural environments has not been deciphered. Here, by analyzing CRISPR-Cas systems of metagenome-assembled genomes (MAGs) from two subsurface environments spanning more than 1 Tb of sequencing data, we show that a substantial proportion of CRISPR spacers are acquired from DNA of other prokaryotes inhabiting the same environment. As such, we found that the number of respective spacers can be up to three times higher than the number of self-targeting spacers. Statistical analyses demonstrated that the acquisition of CRISPR spacers from other prokaryotic genomes is partly explained by the relative abundance of the MAG containing the protospacer, as well as by other factors, such as the total number of CRISPR arrays present in a MAG with the respective spacers. Further, we found that spacer acquisition from foreign prokaryotic DNA occurs in almost all types of CRISPR-Cas systems, but shows preferences for subtypes of CRISPR-Cas systems that differ across the investigated ecosystems. Taken together, our results shed new light on the diversity of CRISPR spacers in natural microbial communities and provide an explanation for some of the many unmatched spacers in public databases.

Key words: CRISPR spacer, metagenomics, population genomics, protospacer.

Significance Statement

Spacers of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) systems can be used to infer infection histories of prokaryotes with mobile genetic elements like viruses. However, many of the spacers in recovered systems do not show any match to viral databases. Here, we show that CRISPR-Cas systems of metagenome-assembled genomes have acquired spacers originating from other prokaryotes inhabiting the same environment, thereby contributing to the overall pool of spacers found in CRISPR-Cas systems from natural environments.

Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (*cas*) genes constitute an adaptive immune system in prokaryotes providing sequence-specific immunity to invading mobile genetic elements (MGEs) like viruses (Barrangou et al. 2007) and plasmids (Marraffini and Sontheimer 2008). It is estimated that CRISPR-Cas systems are present in most archaea (85%) and about half (40%) of bacteria (Makarova et al. 2020). CRISPR-*cas* loci typically comprise an array of direct repeats (DR) separated by highly variable spacer sequences, and flanking *cas* genes (Barrangou et al. 2007). Based on the architecture of different *cas* genes and the presence of type-specific signature genes, CRISPR-Cas systems can be assigned to the main types I–VI and corresponding subtypes (Makarova et al. 2015, 2020). Following the exposure to an invading MGE, a specific sequence fragment called a protospacer is selected from the exogenous nucleic acid and incorporated into the CRISPR array as a spacer (Barrangou et al. 2007). Transcription of the CRISPR array generates a precursor CRISPR RNA (pre-crRNA) that is further processed into small crRNAs, each of which is composed of a spacer and parts of the adjacent repeats (Brouns et al. 2008). During the interference, the recognition of a MGE carrying the respective protospacer leads to the degradation of the intruding nucleic acid by Cas nucleases (Garneau et al. 2010; Jore et al. 2011).

A recent study on publicly available genomes showed that only an average of 7% of all CRISPR spacers can be matched to their respective protospacers, of which a large majority are located on MGEs (Shmakov et al. 2017). Only few spacers were found to match protospacers in the host genome (Shmakov et al. 2017; Zhang et al. 2018), generally referred to as self-targeting spacers. The acquisition of self-targeting spacers can be either random (Dy et al. 2013) or potentially be a defense against heavily recombining strains, as discussed for *Haloferax* (Stachler et al. 2017). In the former case, the spacer acquisition is sometimes followed by the mutation or deletion of the target site (Selle et al. 2015; Cañez et al. 2019) or the mutation of the CRISPR-Cas system, to counteract the activity of the system (Dy et al. 2013; Vercoe et al. 2013; Guan et al. 2017). Notably, self-targeting by RNA-directed CRISPR-Cas systems, such as type III, could degrade mRNA and therefore regulate gene expression (Jones et al. 2012; Sampson et al. 2013). Recently, Hwang et al. (2023) suggested that virus–host interactions across phylogenetic distances in densely populated ecosystems can also lead to the acquisition of spacers against foreign phages that randomly attached to a prokaryote other than the respective host, leading to an unsuccessful infection yet injection of its viral genome. The subsequent acquisition of spacers against viral genomes, whose phages do not successfully infect the respective prokaryote, has also been described for pure cultures (Hynes et al. 2014). Additionally, a recent study revealed that populations of *Ca.* Altiarchaeum can harbor hundreds of spacers matching a DPANN episymbiont's genome, possibly warding off the respective parasite (Esser et al. 2023).

However, these potential mechanisms and functions of spacer acquisition only account for about 7% of the overall spacer diversity (Shmakov et al. 2017) and thus cannot explain the tremendous diversity of spacers that is found in CRISPR-Cas systems from natural environments (England and Whitaker 2013; Childs et al. 2014). Consequently, we tested the hypothesis that CRISPR-Cas systems from subsurface environments carry spacers matching foreign DNA from other prokaryotes inhabiting the same environment and thus contribute to the overall sequence space of CRISPR spacers.

Results and Discussion

Focusing on the deep biosphere, we investigated oceanic and continental sites which are known to have high population density and/or high microbial activity, such as nutrient rich oceanic sediments (Dombrowski et al. 2018) and high-CO_2 aquifers (Figueroa-Gonzalez et al. 2024). Eleven public metagenomes from Guaymas Basin (GB; Gulf of California, Mexico) (Table S1) were previously acquired from three hydrothermal sampling sites and one nonhydrothermal background site (Dombrowski et al. 2018), while a time-series leveraging 28 metagenomes (size fractioned to 0.2-µm and 0.1-µm filters) came from the cold, CO_2-driven Geyser Wallender Born (WB; Wallenborn, Germany) (Figueroa-Gonzalez et al. 2024) (Table S3). The ~1 Tb of sequencing data resulted in 810 high- and medium-quality metagenome-assembled genomes (MAGs) for GB (Table S5) and 726 MAGs for WB (Table S6), respectively. Subsequent dereplication using dRep (95% average nucleotide identity; Olm et al. 2017) yielded 515 MAG clusters (i.e. sets of a representative MAG and its redundant MAGs) for GB and 90 MAG clusters for WB. Based on CRISPRCasFinder (evidence level 4; Couvin et al. 2018) followed by CD-HIT clustering of spacers (97% similarity; Fu et al. 2012), we identified 128 MAG clusters (out of 515 MAG clusters) with CRISPR-Cas systems from GB carrying a total of 4,071 spacer clusters, while the dataset from WB revealed 25 MAG clusters (out of 90 MAG clusters) with CRISPR-Cas systems and a spacer diversity of 965 clusters.

Identification of Spacers that Target Foreign Prokaryotic Genomes of Different Phyla

CRISPR-Cas based targeting of genomes of the same species (and, to a much smaller extent, also at higher taxonomic ranks) across the tree of life was previously revealed

based on publicly available genomes originating from a wide range of environments and other sample types (Shmakov et al. 2017). By contrast, we analyzed spacer-to-protospacer matches in prokaryotic genomes from the very same environments. To avoid the erroneous detection of matches arising from, e.g. mini-arrays (Shmakov et al. 2023), we masked any CRISPR array (evidence levels 1 to 4 based on CRISPRCasFinder (Couvin et al. 2018)) in the genomic data. In similar fashion, we excluded all viral scaffolds (detected by VirSorter2 (Guo et al. 2021), VIBRANT (Kieft et al. 2020) and assessed by CheckV (Nayfach et al. 2021)) from our MAGs (see Supplementary Information for details) to avoid confounding our results arising from matches with phages. We further excluded spacers that matched protospacers in prophage regions of genomes, as well as in regions adjacent to prophages, to avoid introducing biases arising from propagation of prophages in the natural populations. Based on these filtered results, we found that 42% of all MAG clusters with CRISPR-Cas systems (54 of 128 MAG clusters in GB, 10 of 25 MAG clusters in WB) had spacers that matched protospacers in foreign prokaryotic genomes from the same environment. MAG clusters with respective spacers were phylogenetically diverse, including eight archaeal and 15 bacterial phyla (Fig. 1a, b). The total proportion of the diversity of spacers with matches originating from foreign prokaryotic DNA was about 3% (123 of 4,071 spacer clusters in GB, 34 of 965 spacer clusters in WB) (for numbers see Fig. S1; Tables S7 and S8). While these results are based on 80% similarity cutoffs for spacer–protospacer matching (Rahlff et al. 2021), the patterns derived from these results were validated with more stringent similarity cutoffs for spacer–protospacer matches (90%) and are presented in Fig. S2. Interestingly, we found that the number of spacer clusters that matched foreign prokaryotic DNA in the same environment, henceforth referred to as prokaryote-targeting spacer clusters, varied greatly between the two ecosystems and was up to three times higher than the number of self-targeting spacer clusters (Fig. 1c). While Esser et al. (2023) described a similar finding based on the number of spacer clusters of an archaeal host targeting its DPANN symbiont (Esser et al. 2023), the data presented here is distributed across a diverse group of organisms rather than focused on a single symbiont–host pair. Thus, we conclude that prokaryote-targeting spacers make a substantial contribution (about 3%, see above) to the overall diversity of spacers in the analyzed ecosystems.

Expanding our analyses beyond the spacers detected within MAG clusters, we used predicted DR sequences (CRISPRCasFinder, Couvin et al. 2018) to extract spacers from the metagenomic reads via MetaCRAST (Moller and Liang 2017; see Supplementary Information for details). We detected a total of 141,557 spacer clusters (125,773 spacer clusters in GB, 15,784 spacer clusters in WB), of which ~4.65% (5,811 spacer clusters in GB, 778 spacer clusters in WB) matched protospacers originating from foreign prokaryotic genomes in the same environment. These results corroborate our findings from the analyses of CRISPR-Cas systems retrieved from MAGs.

Acquisition of CRISPR Spacers From Foreign Prokaryotic Genomes Cannot be Linked to the Diversity of DNA Uptake Mechanisms Encoded in Bacterial Genomes

While interphylum targeting, i.e. the existence of CRISPR-Cas spacers that match protospacers in the genomes from a different phylum, was recently described for archaea within the same ecosystem (Esser et al. 2023), we here found that also bacterial CRISPR-Cas systems can acquire spacers from other prokaryotic genomes that occur in the very same environment. To further elaborate on these findings, we selected only bacterial genomes for the following analyses. Approximately half of all MAG clusters in GB (249 of 515 MAG clusters) and the majority of those in WB (83 of 90 MAG clusters) were classified as bacterial (GTDB-Tk, Chaumeil et al. 2022). Of these, a total of 71 bacterial MAG clusters (49 in GB, 22 in WB) had CRISPR-Cas systems. The bacterial CRISPR arrays contained a total of 2,154 spacer clusters (1,282 spacer clusters in GB and 872 spacer clusters in WB), accounting for about 43% of the overall spacer diversity, with the rest being attributed to archaeal genomes. Spacer-to-protospacer matching revealed a total of 49 bacterial spacer clusters (21 spacer clusters in GB and 28 spacer clusters in WB) of prokaryotic origin, representing about 2% of the bacterial spacer diversity (Fig. S1, Tables S7 and S8). Targets of these spacers belonged to organisms of the same phylum as the targeting organism, to other bacterial phyla, and even genomes from archaea (Fig. 1a, b). The finding that protospacers originated from genomes of multiple different taxa suggests that the acquisition of spacers from foreign prokaryotic DNA may not be directed, as described for CRISPR-Cas-mediated immunity against viruses (Barrangou et al. 2007) and episymbionts (Esser et al. 2023).

The occurrence of spacers derived from DNA of other prokaryotic raised the question of why respective spacers were obtained and which molecular mechanisms drive the spacer acquisition. Bacteria retrieve foreign prokaryotic DNA by various mechanisms, including natural transformation (here defined as transfer of DNA from the environment), conjugation (*e.g.* transfer of DNA by direct of bacterial cells), transduction (e.g. transfer of bacterial DNA by phages), and other pathways (e.g. transfer of DNA by membrane vesicles) (Fig. 2). The uptake of DNA from the environment can be beneficial to the bacterial cell by promoting the acquisition of new genes via horizontal gene transfer (HGT) (reviewed in Arnold et al. 2022). At the same time, extracellular DNA provides the inhabiting communities with an estimated 2% to 4% of their carbon,

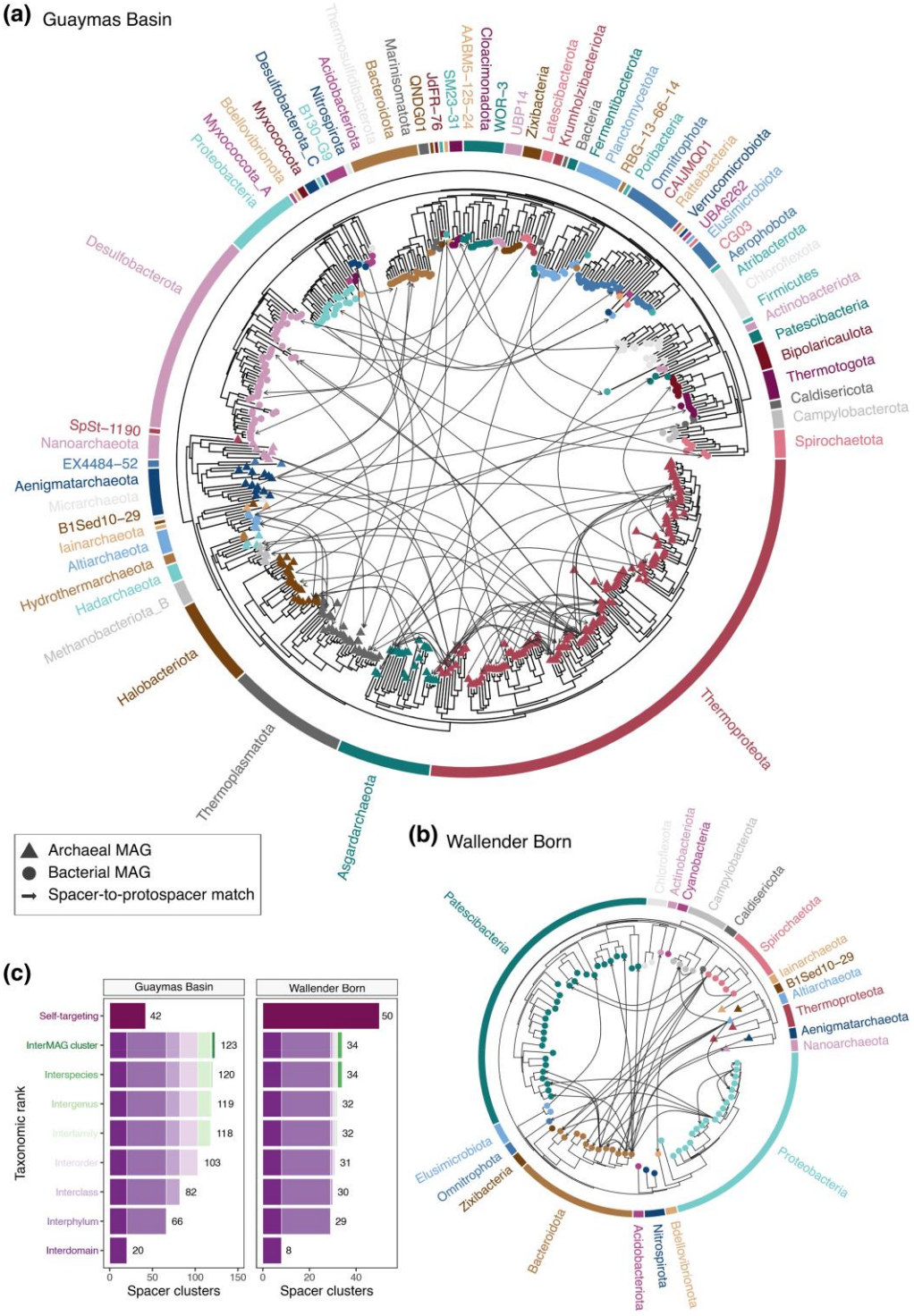

Fig. 1. Phylogenetic distribution of spacer-to-protospacer matches (80% sequence similarity) within foreign prokaryotic genomes based on trees of MAGs (GTDB-Tk, Chaumeil et al. 2022) in GB (a) and WB (b). For results based on 90% sequence similarity for spacer-to-protospacer matches, please see Fig. S2. (c) Distribution of spacer clusters that match at least one MAG cluster in the respective taxonomic rank. Self-targeting spacers refer to spacer clusters matching the same MAG cluster. InterMAG cluster refers to all spacer clusters that match prokaryotic MAG clusters in the respective dataset including those MAG clusters that fall into the same species. Interspecies to interphylum refer to spacer clusters that match MAG clusters of different taxonomic ranks, taking into account the respective hierarchy. For instance, Interclass refers to all spacer clusters that match across classes but also phyla and domain as these are included in that data and are thus indicated by different colors. Please note that the bars are the cumulative number of spacer clusters matching across taxonomic ranks and include all higher taxonomic ranks (also denoted by the numbers next to the bars).

4% to 7% of their nitrogen and 20% to 47% of their phosphorus requirements (Dell'Anno and Danovaro 2005; Corinaldesi et al. 2007; reviewed by Wasmund et al. 2021) in marine sediments, such as the GB hydrothermal sediments (Pérez Castro et al. 2021). It is thus intriguing to hypothesize that the incorporation of spacers from prokaryotic DNA (and potentially also from RNA) is a by-product of the acquisition of environmental nucleic acids by bacterial cells. Consequently, we searched the bacterial MAGs for genes encoding proteins that are known to mediate natural transformation in bacteria (reviewed in Dubnau and Blokesch 2019). We found that all bacterial MAG clusters with spacers matching foreign prokaryotic DNA encoded for respective proteins (Fig. S3, Tables S9 and S10), such as type IV pilus structural protein PilA or ATPase PilT (Chen and Dubnau 2004). However, bacterial MAG clusters containing prokaryote-targeting spacers were not significantly enriched in respective genes compared to bacterial MAG clusters with CRISPR arrays without such spacers (Wilcoxon rank-sum test, Fig. S4). In similar fashion, we found that 13 out of 25 bacterial genome clusters with prokaryote-targeting CRISPR-Cas systems had genes encoding extracellular nucleases. While these nucleases might play significant roles in the utilization of DNA as a nutrient and are less costly

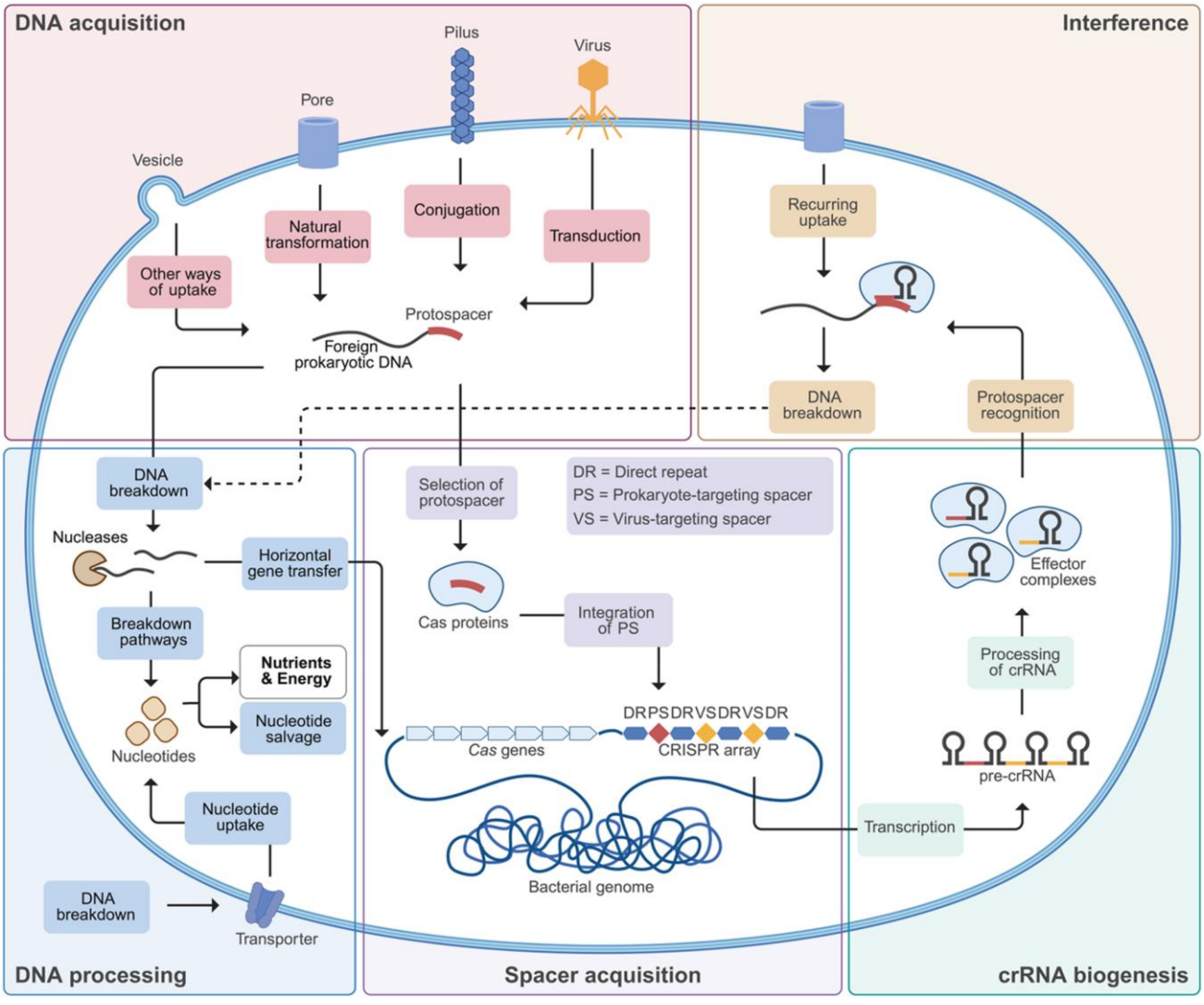

Fig. 2. Schematic representation of newly suggested CRISPR-Cas interferences with foreign prokaryotic nucleic acids. Bacteria can acquire foreign prokaryotic DNA via various uptake mechanisms. While (the uptake of) environmental DNA can promote HGT and/or provide a source of nutrients and energy, it is intriguing to consider that it may also enable the acquisition of spacers from foreign prokaryotic DNA. Transcription and processing of the CRISPR array generate small crRNAs and allow for the assembly of the effector complexes. Following the recognition of foreign prokaryotic DNA, Cas nucleases allow for the degradation of the target sequence, e.g. to potentially prevent genome spoilage (i.e. the impairment of cellular fitness caused by genomic rearrangements, deletions and insertions). Image created in BioRender (Sures 2025; https://BioRender.com/jzxgcl3).

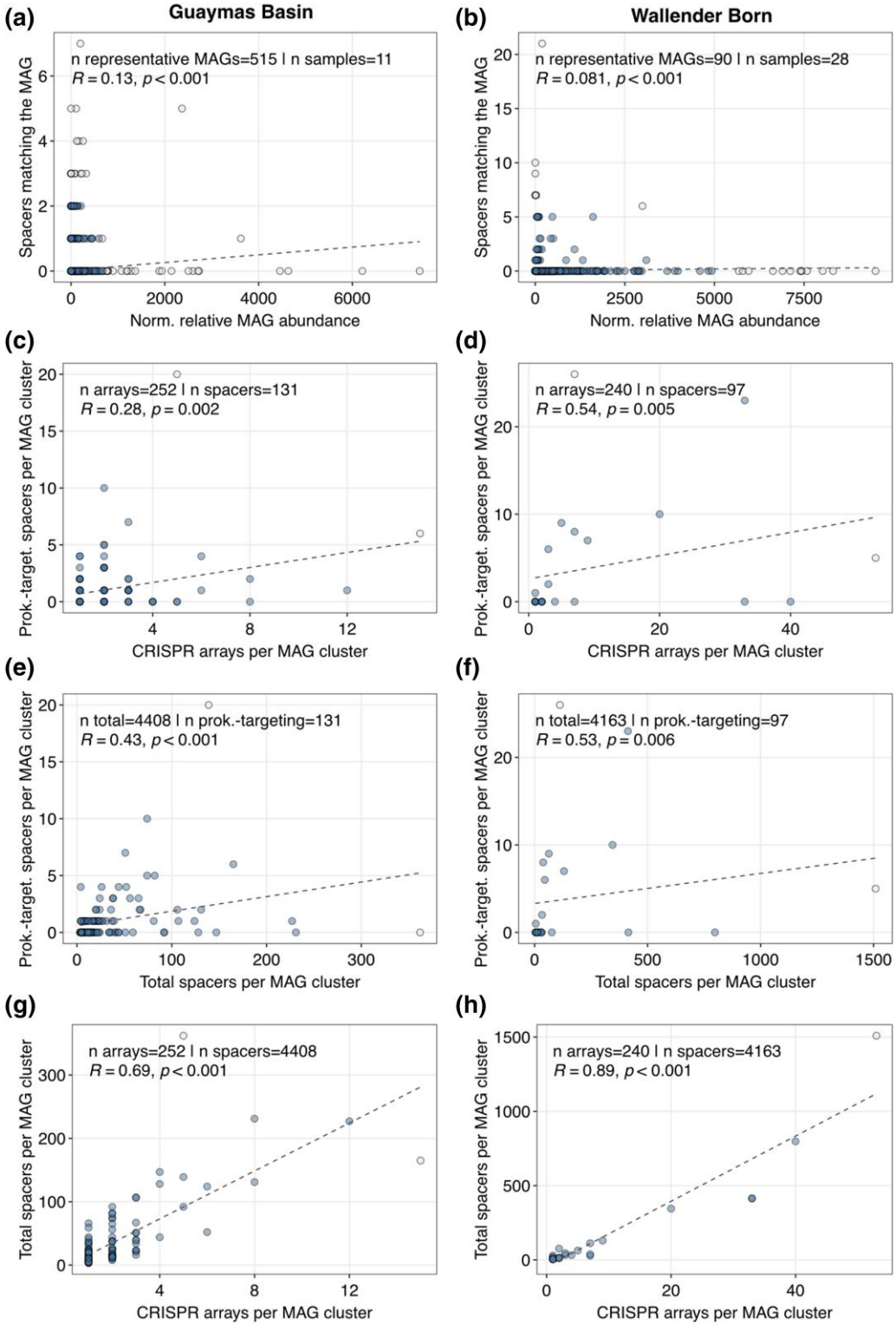

Fig. 3. Scatter plots showing the correlation of the number of spacers matching a MAG cluster and the normalized relative abundance of the same MAG cluster in all samples of GB (a) and WB (b), the correlation of the number of prokaryote-targeting spacers in a MAG cluster and the number of CRISPR arrays in a MAG cluster (c, d), the correlation of the number of prokaryote-targeting spacers in a MAG cluster and the number of total spacers in a MAG cluster (e, f) and the correlation of the total number of spacers in a MAG cluster and the number of CRISPR arrays in a MAG cluster (g, h). Spearman's rank correlation coefficients R and corresponding *P*-values were calculated for each correlation. Gray dashed lines show the linear regression fitted to all values of a panel. Extreme values smaller than the 0.5th and greater than the 95.5th percentile are highlighted in gray.

to maintain for DNA breakdown than Cas nucleases (Fig. S3, Tables S9 and S10), a Wilcoxon rank-sum test revealed no significant difference in the number of extracellular nucleases in bacterial MAG clusters with prokaryote-targeting spacers and clusters without these spacers in both environments (Fig. S5). This finding does not enable the conclusion that extracellular DNA breakdown outcompetes DNA breakdown via CRISPR-Cas systems for nutrient acquisition. Hence, we posit that metagenomic data is insufficient to prove or disprove if spacer acquisition from other prokaryotic genomes is driven by DNA uptake of the respective organism.

Relative Abundance of MAG Clusters Containing Protospacers and Total Number of CRISPR Arrays in a MAG Cluster are Related to the Acquisition of Spacers From Foreign Prokaryotic DNA

To corroborate the relationship between the total number of spacers that match a MAG cluster in a sample of one ecosystem and the normalized relative abundance of the representative MAG of the exact same MAG cluster in the respective sample, we performed correlation analyses of respective values, revealing a weak positive correlation in both environments (Spearman's rank correlations coefficient $R = 0.13$, P-value < 0.001 in GB and $R = 0.081$, P-value < 0.001 in WB; Fig. 3a, b). In other words, the greater the relative abundance of a MAG cluster in a sample, the greater the chance that a spacer matching that MAG cluster occurs in the CRISPR-Cas system of another MAG cluster. Further analyses revealed that the number of prokaryote-targeting spacers in a MAG cluster is positively correlated with the number of CRISPR arrays encoded in that MAG cluster (Spearman's rank correlation coefficient $R = 0.28$, $P = 0.002$ in GB and $R = 0.54$, $P = 0.005$ in WB; Fig. 3c, d), as well as with the total number of spacers found in a MAG cluster (Spearman's rank correlation coefficient $R = 0.43$, $P < 0.001$ in GB and $R = 0.53$, $P = 0.006$ in WB; Fig. 3e, f). However, we also detected a strong positive correlation between the overall spacer number and the number of CRISPR arrays in a MAG cluster (Spearman's rank correlation coefficient $R = 0.69$, $P < 0.001$ in GB and $R = 0.89$, $P < 0.001$ in WB; Fig. 3g, h), suggesting that MAGs with more CRISPR arrays generally have a greater chance of acquiring more spacers, including those that match other prokaryotes. Taken together, these results demonstrate that the relative abundance of the MAG clusters matched by the respective spacers, as well as the overall number of CRISPR arrays found in a MAG cluster containing prokaryote-targeting spacers at least in part determine the number of spacers matching other prokaryotic genomes in these ecosystems.

Prokaryote-targeting Spacers Exist for Multiple CRISPR-Cas Subtypes

Based on *cas* genes and the DR consensus sequences, we subtyped all CRISPR-Cas systems by CRISPRCasTyper (Russel et al. 2020). Overall, we classified 334 systems as five different types and 14 distinct subtypes, whereas 158 systems remained unclassified (Tables S11 and S12). Spacers that match foreign prokaryotic genomes came from arrays assigned to 12 different subtypes, however a substantial number of these spacers (54% in GB and 28% in WB) came from unsubtyped arrays. We observed a significant difference in the distribution of prokaryote-targeting spacers and non-targeting spacers across CRISPR subtypes in both environments (approximated Fisher's exact P-value = 1e−05 in GB and WB). Multiple subtypes showed significant enrichment in prokaryote-targeting spacers, such as subtype I-C in GB, as well as subtypes I-F, II-C, III-A, III-B, and IV-A in WB (Fisher's exact test; for adjusted P-values per subtype see Fig. S6). Interestingly, we observed a significant depletion of prokaryote-targeting spacers in subtype I-B in WB, unlike in the case of *Ca.* Altiarchaea which targets its episymbiont primarily via type I-B systems (Esser et al. 2023). Further, prokaryote-targeting spacers were significantly depleted in subtype I-E arrays in both environments (Fig. S6). Thus, our findings demonstrate that prokaryote-targeting spacers are associated with several CRISPR subtypes, with distinct patterns of preference for both environments investigated herein.

Conclusion

CRISPR-Cas systems in mixed populations confer adaptive immunity of their host against MGEs, regulate gene expression (Jones et al. 2012; Sampson et al. 2013) and potentially control strain diversification (Stachler et al. 2017). While analyses of publicly available genomes revealed targeting of other prokaryotic genomes (Shmakov et al. 2017), we here show that archaea and bacteria also acquire spacers matching other prokaryotic genomes from the same environment. Correlation-based analyses demonstrate that multiple factors can determine the acquisition of CRISPR spacers from foreign prokaryotic DNA, including the relative abundance of the matched genome in the metagenomic sample. While the true reasoning for the spacer acquisition has not been revealed, our analyses demonstrate an additional source of protospacers in metagenomic nucleic acids, which contribute to the overall spacer diversity. At least in our samples, these can be up to three times greater than self-targeting, suggesting an explanation for a fraction of the so far uncharacterized spacers in public databases (Shmakov et al. 2017).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

This research was supported by the Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen (Nachwuchsgruppe Dr. Alexander Probst) and the ERC Synergy grant "Archean Park" under 101118631 (Horizon Europe program). J.P. was supported by Aker BP as part of the GeneOil Project (given to A.J.P.). S.E.R. was supported by a grant from the Simons Foundation (824763). The authors thank Ivan A. Berg for insightful discussion, Ken Dreger for exemplary server administration, and Maximiliane Ackers and Ines Pothmann for administrative support.

Author Contributions

K.S. and J.P. performed the processing of the GB sediment metagenomes. K.S., S.P.E., T.L.V.B., C.J.M., and P.A.F.G. carried out the recovery and refining of MAGs. T.L.V.B. screened the genomes for proteins involved in taking up and degrading DNA. K.S., S.P.E., and T.L.V.B. performed the CRISPR-Cas system detection and spacer analysis. S.E.R. and C.M. assisted with discussion and data interpretation. K.S. and A.R.S. prepared the figures. K.S., A.R.S., and A.J.P. wrote the manuscript with input from all co-authors. A.J.P. designed the project and provided supervision.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

GB sediment metagenomes generated by Dombrowski et al. (2018) are deposited at the Microbial Genomes and Microbiomes (IMG/M) database (Chen et al. 2023). Respective IMG/M accessions are provided in Table S1. Metagenomes obtained from WB (Figueroa-Gonzalez et al. 2024) are available at the National Center for Biotechnology Information (NCBI) BioProject database (Sayers et al. 2025) under accession number PRJNA1001268. For individual BioSample (Sayers et al. 2025) accessions of metagenomes, see Table S3. MAGs from GB and WB analyzed as part of this study, along with the taxonomic assignments (based on GTDB-Tk; Chaumeil et al. 2022) of all representative MAGs, are publicly available on Figshare via this link: https://doi.org/10.6084/m9.figshare.c.7933379

Literature Cited

Arnold BJ, Huang I-T, Hanage WP. Horizontal gene transfer and adaptive evolution in bacteria. Nat Rev Microbiol. 2022:20:206–218. https://doi.org/10.1038/s41579-021-00650-4.

Barrangou R, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007:315:1709–1712. https://doi.org/10.1126/science.1138140.

Brouns SJJ, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008:321:960–964. https://doi.org/10.1126/science.1159689.

Cañez C, Selle K, Goh YJ, Barrangou R. Outcomes and characterization of chromosomal self-targeting by native CRISPR-Cas systems in *Streptococcus thermophilus*. FEMS Microbiol Lett. 2019:366:fnz105. https://doi.org/10.1093/femsle/fnz105.

Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics. 2022:38:5315–5316. https://doi.org/10.1093/bioinformatics/btac672.

Chen I, Dubnau D. DNA uptake during bacterial transformation. Nat Rev Microbiol. 2004:2:241–249. https://doi.org/10.1038/nrmicro844.

Chen I-MA, et al. The IMG/M data management and analysis system v.7: content updates and new features. Nucleic Acids Res. 2023:51:D723–D732. https://doi.org/10.1093/nar/gkac976.

Childs LM, England WE, Young MJ, Weitz JS, Whitaker RJ. CRISPR-induced distributed immunity in microbial populations. PLoS One. 2014:9:e101710. https://doi.org/10.1371/journal.pone.0101710.

Corinaldesi C, Dell'Anno A, Danovaro R. Early diagenesis and trophic role of extracellular DNA in different benthic ecosystems. Limnol Oceanogr. 2007:52:1710–1717. https://doi.org/10.4319/lo.2007.52.4.1710.

Couvin D, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018:46:W246–W251. https://doi.org/10.1093/nar/gky425.

Dell'Anno A, Danovaro R. Extracellular DNA plays a key role in deep-sea ecosystem functioning. Science. 2005:309:2179–2179. https://doi.org/10.1126/science.1117475.

Dombrowski N, Teske AP, Baker BJ. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. Nat Commun. 2018:9:4999. https://doi.org/10.1038/s41467-018-07418-0.

Dubnau D, Blokesch M. Mechanisms of DNA uptake by naturally competent bacteria. Annu Rev Genet. 2019:53:217–237. https://doi.org/10.1146/annurev-genet-112618-043641.

Dy RL, Pitman AR, Fineran PC. Chromosomal targeting by CRISPR-Cas systems can contribute to genome plasticity in bacteria. Mob Genet Elements. 2013:3:e26831. https://doi.org/10.4161/mge.26831.

England WE, Whitaker RJ. Evolutionary causes and consequences of diversified CRISPR immune profiles in natural populations. Biochem Soc Trans. 2013:41:1431–1436. https://doi.org/10.1042/BST20130243.

Esser SP, et al. A predicted CRISPR-mediated symbiosis between uncultivated archaea. Nat Microbiol. 2023:8:1619–1633. https://doi.org/10.1038/s41564-023-01439-2.

Figueroa-Gonzalez PA, et al. Metaproteogenomics resolution of a high-CO_2 aquifer community reveals a complex cellular adaptation of groundwater Gracilibacteria to a host-dependent lifestyle. Microbiome. 2024:12:194. https://doi.org/10.1186/s40168-024-01889-8.

Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012:28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

Garneau JE, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature. 2010:468:67–71. https://doi.org/10.1038/nature09523.

Guan J, Wang W, Sun B. Chromosomal targeting by the type III-A CRISPR-cas system can reshape genomes in *Staphylococcus aureus*. mSphere. 2017:2:e00403–e00417. https://doi.org/10.1128/mSphere.00403-17.

Guo J, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome. 2021:9:37. https://doi.org/10.1186/s40168-020-00990-y.

Hwang Y, Roux S, Coclet C, Krause SJE, Girguis PR. Viruses interact with hosts that span distantly related microbial domains in dense hydrothermal mats. Nat Microbiol. 2023:8:946–957. https://doi.org/10.1038/s41564-023-01347-5.

Hynes AP, Villion M, Moineau S. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. Nat Commun. 2014:5:4399. https://doi.org/10.1038/ncomms5399.

Jones CL, Sampson TR, Nakaya HI, Pulendran B, Weiss DS. Repression of bacterial lipoprotein production by *Francisella novicida* facilitates evasion of innate immune recognition: innate immune evasion via BLP repression. Cell Microbiol. 2012:14:1531–1543. https://doi.org/10.1111/j.1462-5822.2012.01816.x.

Jore MM, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol. 2011:18:529–536. https://doi.org/10.1038/nsmb.2019.

Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome. 2020:8:90. https://doi.org/10.1186/s40168-020-00867-0.

Makarova KS, et al. An updated evolutionary classification of CRISPR–Cas systems. Nat Rev Microbiol. 2015:13:722–736. https://doi.org/10.1038/nrmicro3569.

Makarova KS, et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol. 2020:18:67–83. https://doi.org/10.1038/s41579-019-0299-x.

Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science. 2008:322:1843–1845. https://doi.org/10.1126/science.1165771.

Moller AG, Liang C. MetaCRAST: reference-guided extraction of CRISPR spacers from unassembled metagenomes. PeerJ. 2017:5:e3788. https://doi.org/10.7717/peerj.3788.

Nayfach S, et al. Checkv assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021:39:578–585. https://doi.org/10.1038/s41587-020-00774-7.

Olm MR, Brown CT, Brooks B, Banfield JF. Drep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017:11:2864–2868. https://doi.org/10.1038/ismej.2017.126.

Pérez Castro S, et al. Degradation of biological macromolecules supports uncultured microbial populations in Guaymas Basin hydrothermal sediments. ISME J. 2021:15:3480–3497. https://doi.org/10.1038/s41396-021-01026-5.

Rahlff J, et al. Lytic archaeal viruses infect abundant primary producers in Earth's crust. Nat Commun. 2021:12:4642. https://doi.org/10.1038/s41467-021-24803-4.

Russel J, Pinilla-Redondo R, Mayo-Muñoz D, Shah SA, Sørensen SJ. CRISPRCasTyper: automated identification, annotation, and classification of CRISPR-Cas Loci. CRISPR J. 2020:3:462–469. https://doi.org/10.1089/crispr.2020.0059.

Sampson TR, Saroj SD, Llewellyn AC, Tzeng Y-L, Weiss DS. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature. 2013:497:254–257. https://doi.org/10.1038/nature12048.

Sayers EW, et al. Database resources of the National Center for Biotechnology Information in 2025. Nucleic Acids Res. 2025:53:D20–D29. https://doi.org/10.1093/nar/gkae979.

Selle K, Klaenhammer TR, Barrangou R. CRISPR-based screening of genomic island excision events in bacteria. Proc Natl Acad Sci U S A. 2015:112:8076–8081. https://doi.org/10.1073/pnas.1508525112.

Shmakov SA, et al. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. mBio. 2017:8:e01397-17. https://doi.org/10.1128/mBio.01397-17.

Shmakov SA, et al. Widespread CRISPR-derived RNA regulatory elements in CRISPR-Cas systems. Nucleic Acids Res. 2023:51:8150–8168. https://doi.org/10.1093/nar/gkad495.

Stachler A-E, et al. High tolerance to self-targeting of the genome by the endogenous CRISPR-Cas system in an archaeon. Nucleic Acids Res. 2017:45:5208–5216. https://doi.org/10.1093/nar/gkx150.

Sures K. Schematic representation of newly suggested CRISPR-Cas interferences with foreign prokaryotic nucleic acids. https://BioRender.com/jzxgcl3 2025.

Vercoe RB, et al. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. PLoS Genet. 2013:9:e1003454. https://doi.org/10.1371/journal.pgen.1003454.

Wasmund K, et al. Genomic insights into diverse bacterial taxa that degrade extracellular DNA in marine sediments. Nat Microbiol. 2021:6:885–898. https://doi.org/10.1038/s41564-021-00917-9.Zhang F, et al. CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions. Commun Biol. 2018:1:180. https://doi.org/10.1038/s42003-018-0184-6.

Associate editor: Tal Dagan