

High diversity and putative novel Class 2 CRISPR-Cas systems effectors from hot springs ¹

Alta diversidad y putativos nuevos efectores de sistemas CRISPR-Cas de clase 2 desde fuentes termales
Alta diversidade e possíveis novos efetores de sistemas CRISPR-Cas de classe 2 de fontes termais

Página | 1

Oscar Salgado*²

Pontificia Universidad Católica de Chile, Chile*

Fecha de Recepción: 5-3-2025. Fecha de Aceptación: 31-7-2025

Autor de correspondencia: Oscar Salgado, olsalgado@uc.cl

Cómo citar:

Salgado, O. (2025). High diversity and putative novel Class 2 CRISPR-Cas systems effectors from hot springs. *Revista Científica Cuadernos de Investigación*, 3, e51, 1-26. <https://doi.org/10.59758/rcci.2025.3.e51>

Abstract

CRISPR-Cas systems are present in ~42% of the bacterial and ~80% of archaeal genomes from all environments but enriched in thermophilic and hyperthermophilic members, reaching ~80% prevalence. Most descriptions of CRISPR-Cas in thermophiles consider hyperthermophilic archaeal species, where class 2 CRISPR-Cas systems are virtually absent. Despite their abundance in thermal environments, no study has described CRISPR-Cas diversity in hot springs communities worldwide in mesothermophilic (40°C-80°C) and circumneutral pH (6-8). **Objective:** To describe the types and subtypes of CRISPR-Cas systems in 37 hot springs metagenomes within this temperature and pH ranges, emphasizing the mining of novel class 2 variants. **Methodology:** Quantification of CRISPR-Cas systems through bioinformatic tools. **Results:** Searching all CRISPR-Cas systems revealed 2296 systems in hot springs, encompassing all types described to date. Using a phylogenetic and identity matrix approach, 57 class 2 effector proteins Cas9 were found, revealing potentially novel variants of Cas9. **Conclusions:** The large diversity observed could be the first glance toward further characterization of CRISPR-Cas loci and potential new variants with biotechnological applications. Furthermore, the results highlight hot springs for studies of the ecology and evolution of CRISPR-Cas by revealing new genetic sequences that can contribute to filling the gaps in the evolution of these systems.

Keywords: Cas9, Thermophiles, hot springs, CRISPR-Cas.

¹ Copyright: © 2025, Salgado. Este es un artículo de acceso abierto distribuido bajo la licencia **Creative Commons de Atribución No Comercial 4.0**. Permite su uso sin restricciones, su distribución y reproducción por cualquier medio, siempre que no se haga con fines comerciales y el trabajo original sea fielmente citado.

² <https://orcid.org/0000-0001-9297-968X>

Resumen

Los sistemas CRISPR-Cas están presentes en ~42% de los genomas bacterianos y ~80% de los genomas de arqueas de todos los ambientes, pero enriquecidos en microorganismos termófilos e hipertermófilos, alcanzando una prevalencia de ~80%. Las descripciones de CRISPR-Cas en termófilos consideran especies arqueales hipertermófilas, donde los sistemas CRISPR-Cas de clase 2 están virtualmente ausentes. A pesar de su abundancia en ambientes termales, ningún estudio ha descrito la diversidad de CRISPR-Cas en comunidades de termas del mundo en el rango mesotermofílico (40 °C-80 °C) y pH circunneutral (6-8). **Objetivo:** Describir los tipos y subtipos de sistemas CRISPR-Cas en 37 metagenomas de aguas termales, enfatizando en nuevas variantes de clase 2. **Metodología:** Cuantificación de sistemas CRISPR-Cas mediante herramientas bioinformáticas. **Resultados:** La búsqueda reveló 2296 sistemas en termas, que abarcan todos los tipos descritos hasta la fecha. Utilizando un enfoque filogenético y de matriz de identidad, se encontraron 57 proteínas efectoras de clase 2 Cas9, lo que reveló variantes potencialmente novedosas. **Conclusiones:** La gran diversidad observada podría permitir una mayor caracterización de los loci CRISPR-Cas y posibles nuevas aplicaciones biotecnológicas. Los resultados destacan las termas para estudios ecoevolutivos de CRISPR-Cas al revelar nuevas secuencias que contribuirían a dilucidar la evolución de estos sistemas.

Palabras clave: Cas9, termófilos, fuentes termales, CRISPR-Cas.

Resumo

Os sistemas CRISPR-Cas estão presentes em ~42% dos genomas bacterianos e ~80% dos genomas de arqueas de todos os ambientes, mas são enriquecidos em microrganismos termofílicos e hipertermofílicos, atingindo uma prevalência de ~80%. As descrições de CRISPR-Cas em termófilos consideram espécies de arqueas hipertermofílicas, onde os sistemas CRISPR-Cas de classe 2 estão virtualmente ausentes. Apesar de sua abundância em ambientes termais, nenhum estudo descreveu a diversidade de sistemas CRISPR-Cas em comunidades de fontes termais ao redor do mundo dentro da faixa mesotermofílica (40-80 °C) e pH circunneutro (6-8). **Objetivo:** Descrever os tipos e subtipos de sistemas CRISPR-Cas em 37 metagenomas de fontes termais, com ênfase em novas variantes de classe 2. **Metodologia:** Quantificação de sistemas CRISPR-Cas usando ferramentas de bioinformática. **Resultados:** A busca revelou 2296 sistemas CRISPR-Cas, abrangendo todos os tipos descritos até o momento. Utilizando uma abordagem filogenética e de matriz de identidade, 57 proteínas efectoras Cas9 de classe 2 foram encontradas, revelando variantes potencialmente novas. **Conclusões:** A alta diversidade observada pode permitir uma caracterização mais aprofundada dos loci CRISPR-Cas e potenciais novas aplicações biotecnológicas. Os resultados destacam o potencial para estudos ecoevolutivos de CRISPR-Cas, revelando novas sequências que contribuiriam para a elucidação da evolução desses sistemas.

Palavras Chave: Cas9, termófilos, fontes termais, CRISPR-Cas.

Introducción

The adaptive immune system CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated genes) is one of the most significant discoveries in the biological sciences in recent decades (Gostimskaya, 2022; Harsij et al., 2024; Lander, 2016; Mojica & Montoliu, 2016) due to its biotechnological applications but also for its fantastic biology. In general terms, all CRISPR-Cas systems described to date work through several Cas genes (Makarova et al., 2020, 2015, 2011; Mohanraju et al., 2016), which specifically recognize foreign genetic material (mostly from viruses, Shmakov et al., 2017; Shmakov et al., 2020) and incorporate it into the bacterial or archaeal host genome (Sternberg et al., 2016). When the invader's genetic material penetrates the cell, small fragments of that genetic material are integrated into a specific genome locus, thus extending the CRISPR array (Barrangou et al., 2007; Marraffini, 2015). This array is then transcribed and processed as small RNAs (crRNAs) to guide one or more Cas proteins (depending on the system type) that perform cleavage of the invader through base complementarity (Makarova et al., 2020; Marraffini, 2015), hence providing adaptive immunity to the host cell. The host-stored foreign genetic material allows immunological memory against the invader and can be inherited throughout generations (Koonin & Wolf, 2016) and transferred horizontally (Koonin & Makarova, 2009). The CRISPR-Cas revolution encouraged the search for new CRISPR-Cas systems and variants. Since the first CRISPR-Cas classification in 2011 with three types (Makarova et al., 2011), the systems have extended to two classes and six types to date (Makarova et al., 2020), evidencing the productive research in the field and suggesting that only the tip of the iceberg has been described so far.

The CRISPR-Cas system should be a robust adaptive advantage for the host cell against viruses. However, roughly 40% of the major bacterial groups harbor a CRISPR-Cas system, which raises an intriguing question about its importance in natural populations (Westra & Levin, 2020). Also, immunity against eventually beneficial genetic material (Jiang et al., 2013), e.g., plasmids, reinforces this question. Data suggest that at least four ecological factors may affect the prevalence of CRISPR-Cas (reviewed by Westra et al., 2016): (i) the force of infection referred to as the fitness cost imposed by the frequency of infection, (ii) the presence of defective viruses that are more frequently included as spacers, (iii) the observation that beneficial mobile genetic elements impact CRISPR-Cas presence and (iv) the genetic diversity of the virus. A handful of studies have already attempted to explain the dynamics between viral genetic diversity and CRISPR-Cas in nature through *in vitro* (Broniewski et al., 2020; Chevallereau et al., 2019; Childs et al., 2014; Weinberger et al., 2012a) and modeling (Iranzo et al., 2013; Weinberger et al., 2012a, 2012b) hypothesis. The results indicate that target viral population diversity determined the usefulness of CRISPR-Cas adaptive immunity. Analyzing natural environments, a comprehensive survey using 324 metagenomes corroborates the predicted importance of viral diversity on CRISPR-Cas prevalence, showing that the CRISPR-Cas systems abundance decreases when viral diversity increases (controlling for viral abundance) (Meaden et al., 2022). Furthermore, that study reveals that high viral abundance correlates with higher CRISPR repeat counts (Meaden et al., 2022). Interestingly, some of these studies (Iranzo et al., 2013; Weinberger et al., 2012b) mention the exceptionally high prevalence of CRISPR-Cas systems occurring in hot springs, which fit well with the proposed models and are consistent with *in vitro* and environmental evidence.

The high prevalence of CRISPR-Cas systems in the genomes of thermophilic bacteria and archaea compared to mesophilic and psychrophilic has been mentioned in several studies (Anderson et al., 2011; Lan et al., 2022; Makarova et al., 2006; Weinberger et al., 2012b; Weissman et al., 2019). Consistently with the aforementioned effect of viral diversity, a couple of studies (Iranzo et al., 2013; Weinberger et al., 2012b) indicate that the lower mutation rate of thermophilic prokaryotes (Drake, 2009; Friedman et al., 2004; Gault et al., 2021; Zeldovich et al., 2007) and their small population size (Iranzo et al., 2013), promotes the high prevalence of CRISPR-Cas systems in hot springs. On the other hand, the described lower diversity and small population size of viruses from hot springs (Parmar et al., 2018) agree with the ecosystem conditions promoting high CRISPR-Cas prevalence mentioned above. Therefore, hot springs environments are a convenient natural system for studying CRISPR-Cas systems, but comparatively fewer studies have been performed in these environments. In Yellowstone, Heidelberg et al. (2009) identified active coevolution through silent mutations in a viral gene evading CRISPR. A study with the hyperthermophilic archaea *Sulfolobus islandicus* revealed that CRISPR spacers tend to match the local viral community (Held & Whitaker, 2009). Also, a hyperthermophilic hot spring showed that previously unidentified viruses could be recovered using CRISPR spacers (Snyder et al., 2010). Between the few surveys performed in mesothermophilic (40-80°C) and circumneutral (pH 6-8) hot springs, Guajardo-Leiva et al. (2021, 2018) identified active CRISPR-Cas systems against a highly prevalent cyanophage in phototrophic microbial mats. Also, a comprehensive analysis of the Cas1 gene using hot spring metagenomes worldwide (Salgado et al., 2022) identifies new Cas1 clades, reinforcing the hypothesis of new CRISPR-Cas systems in hot springs worldwide.

About ten reports have been published describing new systems from hot springs, and all have considered Class 2 CRISPR-Cas systems (Adalsteinsson et al., 2021; Fan et al., 2021; Fuchs et al., 2022; Harrington et al., 2017; Le et al., 2020; Le & Sun, 2022; Mougiakos et al., 2017; Nguyen et al., 2022; Schmidt et al., 2019; Tian et al., 2020; Tsui et al., 2017). Compared to Class 1, Class 2 systems are biotechnologically interesting because they cleave the target with a straightforward single effector protein (Hille et al., 2018; Makarova et al., 2015). Class 2 CRISPR-Cas systems are classified into three types (types II, V, and VI) and 28 subtypes (Makarova et al., 2022), including subtype V-G described from a hot spring and with the smallest effector protein (Cas12g) (Yan et al., 2019). Type II is the most abundant Class 2 CRISPR-Cas system in nature (Makarova et al., 2022), characterized by the famous protein Cas9 (Shmakov et al., 2017). All type II subtypes described to date (II-A, II-B, II-C1, II-C2, II-D, Makarova et al., 2022) possess the HNH and RuvC domains. Type V also shows RuvC as the effector domain (Makarova et al., 2020), so a common ancestor for these systems has been proposed (Altae-Tran et al., 2021). Meanwhile, type VI has no common gene ancestors with type II and V (Makarova et al., 2022), but it is particularly relevant as a potential editing tool because it targets RNA (Makarova et al., 2020). In addition to the aforementioned thermal Cas12g, several Class 2 effectors that extend the DNA cleavage temperature limit of Cas9 (44°C for Cas9 from *S. pyogenes*) (Schmidt et al., 2019) have been searched in hot springs, and some have been probed as genome editing tools (Fan et al., 2021; Le et al., 2020; Le & Sun, 2022). In the genus *Geobacillus*, two Cas9 that cleavage DNA around 70°C were described (Harrington et al., 2017; Mougiakos et al., 2017). Furthermore, three other Cas9 have been discovered in *Acidothermus cellulolyticus* (Tsui et al., 2017), in an uncultured Ignavibacterium (Schmidt et al., 2019), and in *Thermus thermophilus* (Adalsteinsson et al., 2021). In addition to Cas9, thermostable effector proteins of types V and

VI (Cas12 and Cas13, respectively) have been discovered (Fuchs et al., 2022; Nguyen et al., 2022; Tian et al., 2020), showing a clear example of the relevance for mining these environments.

The mining and discovering of new class 2 effectors and/or CRISPR-Cas systems in hot spring metagenomes have at least two clear benefits. First, the extended thermostability of already known class 2 effector proteins (Cas9, Cas12, Cas14 or Cas13) expands the applications of already known and established genome editing protocols towards high-temperature species and systems, as well as promotes the development of new tools (e.g., Ali et al., 2020). Second, the description of unknown CRISPR-Cas systems could bring new Cas proteins potentially applicable for genome editing and regulation tools, besides its intrinsic value in understanding the ecology and evolution of CRISPR-Cas systems. Despite that class 2 CRISPR-Cas systems are virtually exclusively of the Bacteria domain (Makarova et al., 2020), and that metagenomic dataset from mesothermophilic (40-80°C) and circumneutral (pH 6-8) hot springs excludes most members of Archaea, no study has considered a detailed description of the types of CRISPR-Cas systems in these hot springs globally. The present work aimed to describe the genetic diversity of CRISPR-Cas systems in 37 mesothermophilic and circumneutral metagenomes from 14 hot springs (7 locations) worldwide, focusing on characterizing class 2 CRISPR-Cas systems. The results revealed a high diversity of CRISPR-Cas systems, including some of the uncommon types such as types IV (A, B, C and D) and VI (A and B), and also allowed the description of potential new variants of the most biotechnological applied class 2 protein, Cas9.

Materials and Methods

Metagenomic dataset selection, reads processing and de novo assembly

The 37 hot spring metagenomes used in this study were selected from the 48 metagenomic datasets used by Salgado et al. (2022). Eleven metagenomes were excluded here due to their low sequencing yield. The 37 selected are between 40-80°C and pH 6-8, encompassing 25 from the microbial mat, seven from sediment, and five from thermal water. These metagenomes are publicly available in NCBI SRA and JGI databases (Nordberg et al., 2014; Wheeler et al., 2007), and detailed information about location, database ID, geographic coordinates and bibliographic reference can be found in Supplementary Table S1. Metagenomic reads for each metagenome were processed as described in Guajardo-Leiva et al. (2018), while read assembling was performed according to Salgado et al. (2022). Assessment of *de novo* assembling quality was performed using metaQuast (Quast v5.1, Mikheenko et al., 2016) and is available in Supplementary Table S1.

CRISPR-Cas systems prediction and classification

CRISPR-Cas prediction was performed using the CRISPRCasTyper tool v1.8.0 (Russel et al., 2020) with default parameters except the --prodigal meta mode. High and medium-quality metagenome-assembled genomes (MAGs) were obtained, and the CRISPR-Cas systems present within the taxonomically assigned MAGs were assigned for accurate taxonomic assignment of the predicted systems. To obtain the MAGs, a refined binning protocol using the assembled contigs (Alcorta et al., 2020) but with some modifications was applied. Briefly, contigs larger than 1000 bp were binned using the metaWrap v1.3 (Uritskiy et al., 2018)

binning module. Consolidated MAGs with less than 10% contamination and more than 50% completeness were cleaned with RefineM tool v0.1.2 (Parks et al., 2017) and GUNC v1.0.5 (Orakov et al., 2021). The 1696 MAGs qualified with high or medium quality according to the Genomic Standards Consortium (Bowers et al., 2017) were taxonomically assigned using the “classify workflow” in the GTDB-tk v2.3.0 release 214 (Chaumeil et al., 2020). A detailed list of the taxonomy of MAGs used here can be found in Supplementary Table S2. Non-binned contigs were also used for CRISPR-Cas prediction with CRISPRCasTyper. These contigs were also taxonomically assigned using the MDMcleaner tool v0.8.7 (Vollmers et al., 2022), which works by alignment with GTDB (Chaumeil et al., 2020) release 214, Silva (Quast et al., 2013) release 138 and RefSeq (Pruitt et al., 2004) databases for contig classification. A CRISPR-Cas system was predicted if its interference module was at least 70% complete.

Definition of putative novel class 2 effectors

The predicted orphan CRISPRs (arrays without known Cas operon in the vicinity ± 10 Kbp, Russel et al., 2020) were used as input to the novel class 2 effector protein predictor module of the CRISPRimmunity suite (Zhou et al., 2023). Briefly, rare CRISPR arrays (below 90% identity and coverage of known arrays) are used to detect proteins over 499 amino acids near five upstream and downstream ORFs. Vicinity is annotated against CDD (Marchler-Bauer et al., 2017) with an e-value of <0.01 and the candidate is retained if a nuclease-associated domain is present, if this protein is located in the three ORFs near CRISPR, and if it has $<90\%$ similarity to known Cas. The putative novel class 2 effectors thus identified through CRISPRimmunity were used in subsequent analyses.

The CRISPRCasTyper tool predicts the CRISPR-Cas systems using a gene-scoring algorithm in which class 2 effector proteins are assigned under a specific e-value, sequence coverage, and HMM coverage (Russel et al., 2020). If a class 2 system has the same score for more than one subtype, the system is classified by CRISPRCasTyper as "Ambiguous". So, the effectors defined as ambiguous for class 2 in the 37 metagenomes were used under the rationale that CRISPRCasTyper recognizes the protein as Cas, but cannot assign it to a specific subtype due to its sequence variation concerning the canonical class 2 effector proteins in the databases. A non-redundant (90% identity and coverage, Shmakov et al. 2017) set of ambiguous Cas9 was constructed using CD-HIT v4.8.1 (Fu et al., 2012) and used for subsequent data analyses.

All the putative novel class 2 effectors identified, either using CRISPRimmunity or CRISPRCasTyper, were analyzed for protein domain architecture using CD-search (Marchler-Bauer & Bryant, 2004) in CDD v3.20 with an e-value of 0.01. In addition, these proteins were compared with the non-redundant (nr) protein sequence database using Blastp (Altschul et al., 1990) (-evalue $1e-5$ -qcov_hsp_perc 80).

Phylogenetic analysis of novel putative class 2 effector proteins

With the sequences of the new putative class 2 effector proteins obtained in this study, a phylogenetic tree was constructed, following the methodology of Shmakov et al. (2017) for class 2 CRISPR-Cas systems, and including reference sequences of the class 2 known subtype and the *IscB* gene (WP004892446) as an outgroup.

Briefly, sequences were aligned using MUSCLE v5.1 (Edgar, 2004), and positions with more than 0.5 gap fraction were removed using trimAL v1.4.rev22 (Capella-Gutiérrez et al., 2009). A maximum likelihood phylogenetic tree was built using IQ-TREE v1.6.12 (Nguyen et al., 2015) with the WAG evolutionary model and the discrete gamma model with 20 rate categories (-m WAG+G20, 10000 ultrafast bootstrap) (Shmakov et al., 2017). In order to deepen analyses of the genetic diversity in the candidate dataset used in the phylogenetic tree, a percentage identity matrix was constructed with the alignment results obtained using the Clustal Omega tool (Madeira et al., 2019).

Results

In the present study, 2296 CRISPR-Cas systems were predicted from 37 metagenomic datasets from three continents, representing 14 hot springs (Supplementary Table S3). This study represents the most detailed survey of mesothermophilic and circumneutral hot spring CRISPR-Cas systems regarding the number of samples, type of thermal habitats (water, sediment, microbial mat), and geographical distribution encompassed.

High diversity and uncommon CRISPR-as systems in hot springs

The CRISPR-Cas systems obtained were distributed in MAGs (1096) and non-binned contigs (1200) and showed a clear predominance of class 1 systems (Fig. 1). In fact, almost all class 1 described to date are present in the hot springs analyzed here with only variants of the subtypes III-E and IV-E being missing. Most of the class 1 CRISPR-Cas systems belong to subtype III-A, followed by I-B and III-B, while the least represented were subtypes I-F, IV-D and IV-A2. Interestingly, 45 systems belonging to the uncommonly reported subtype IV systems were observed here, with subtype IV-B being the most represented (Fig. 1). As for the taxonomy associated with all these class 1 systems, most of the prokaryotes from hot springs here show a large diversity of CRISPR-Cas types and subtypes. The phylum *Chloroflexota* presents the most remarkable diversity of subtypes, followed by a handful of commonly reported hot spring taxa such as *Bacteroidota*, *Armatimonadota* and *Pseudomonadota* (Fig. 1). However, it should be noted that some phyla tend to have specific subtypes. For example, 39% of the archaea phylum *Thermoproteota* have systems belonging to subtype I-A, while *Acidobacteriota* shows a higher frequency of subtype III-A and *Cyanobacteriota* of subtype I-D. For systems showing Cas operons with equally high scores and signature genes of the subtype, the systems were annotated as hybrid (Russel et al., 2020). Several combinations of types I and III were hybrid, most belonging to subtypes I-B and III-D.

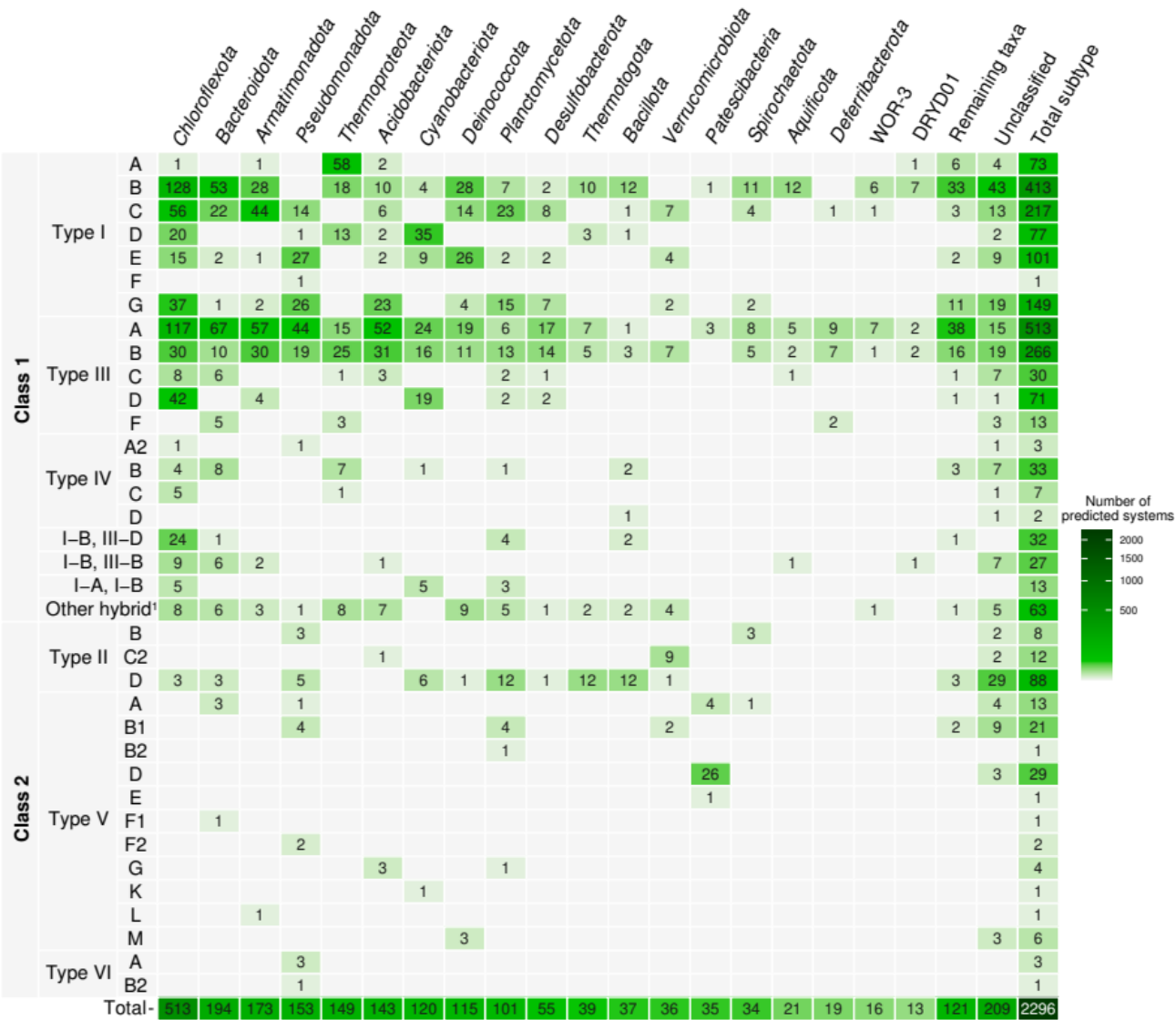


Figure 1. Absolute abundance of CRISPR-Cas systems obtained from the 37 mesothermophilic (40°C-80°C) and circumneutral (pH 6-8) metagenomes from hot springs worldwide. Classification by class, type and subtype is indicated on the left, and the taxonomy of the main microbial phyla is at the top. ¹Other hybrid systems encompass I-C,III-B; I-G,III-B; I-A,III-D; I-B,III-A; III-A,III-B; I-A,III-B; I-A,III-F; I-B,III-C,III-D; III-A,III-D; I-A,III-C; I-B,III-F; I-B,III-C; I-C,III-C; I-D,IV-C; I-D,III-B; I-D,III-C; I-G,III-A; I-G,III-C and III-B,III-D.

On the other hand, Class 2 CRISPR-Cas systems are present with members of all three types described to date (II, V and VI, Makarova et al., 2022) (Fig. 1). However, the prevalence of class 2 CRISPR-Cas systems in hot springs is low compared to class 1 systems. In fact, only 186 systems (8%) belong to class 2, with subtype II-D being significantly the most represented, followed by V-D and V-B1. Nevertheless, class 2 CRISPR-Cas systems tend to present subtypes in specific taxonomic groups. The phyla *Planctomycetota*, *Thermotogota* and *Bacillota* harbor 41% of the predicted II-D subtype, whereas the subtype V-D appears as exclusive of *Patescibacteria* phylum (Fig. 1). Several other class 2 subtypes are present only in one classified bacterial groups, e.g. V-B2 (*Planctomycetota*), V-E (*Patescibacteria*), V-F1 (*Bacteroidota*), V-F2 (*Pseudomonadota*), V-

K (*Cyanobacteriota*), V-L (*Armatimonadota*), V-M (*Deinococcota*), VI-A and VI-B2 (*Pseudomonadota*). Also, some subtypes are present only in two groups, as II-B (*Pseudomonadota* and *Spirochaetota*), II-C2 (*Acidobacteriota* and *Verrucomicrobiota*), V-F2 (*Pseudomonadota*), and V-G (*Acidobacteriota* and *Planctomycetota*). Finally, class 2 systems described to date were absent in the phyla *Thermoproteota*, *Aquificota*, *Deferribacterota*, and the groups WOR-3 and DRYD01. In addition, it is noteworthy that 52 class 2 CRISPR-Cas systems (28% of class 2 systems described in this survey) were predicted on contigs that were not taxonomically assigned beyond the bacterial domain and were therefore considered unclassified (Fig. 1). Most of these contigs are short and thus were not classified, so several are likely present in the groups mentioned here. Despite the lower representation of class 2 CRISPR-Cas systems compared to class 1, the genetic diversity of the 16 subtypes described in hot springs indicates that these systems are helpful for the microorganisms that possess them, suggesting the possibility of identifying new variants in hot springs environments.

Putative novel class 2 CRISPR-Cas systems from hot springs

The CRISPRimmunity suite (Zhou et al., 2023) predicted eight putative novel class 2 effector proteins, but checking these candidates in the CDD database under the threshold used by the tool (e-value 0.01) or even relaxing the threshold to e-value 0.1, do not reveal any recognized domain. Also, no known domains were detected using HHPred (<https://toolkit.tuebingen.mpg.de/tools/hhpred>) (Zimmermann et al., 2018). With exploratory intentions, a screening phylogenetic tree was constructed with the eight candidate sequences and class 2 reference sequences available to date (Makarova et al., 2022) (Supp. Fig. S1). Six sequences group with type V reference sequences and one with type VI, having one singleton. No sequences group with type II reference sequences.

Were identified 538 ambiguous Cas9 from 37 metagenomes using CRISPRCasTyper. No ambiguous types V or VI were identified in the dataset used here. The 493 cluster-representative Cas9 sequences were filtered by length considering the size of the smaller Cas9 homolog IscB (Shmakov et al., 2017), so only 200 Cas9 sequences over 400 aa were maintained. Then, to avoid possible errors of truncated Cas9 genes at the contig edges, the 57 only sequences surrounded by other ORFs were maintained and used in subsequent analyses. The phylogenetic tree was constructed with 57 putative novel Cas9 proteins and the reference sequences of types II-A, II-B, II-C (variants 1 and 2) and II-D (Fig. 2). The tree follows the phylogeny of Cas9 already published by previous studies (Fonfara et al., 2014; Ipoutcha et al., 2019; Schmidt et al., 2019) in terms of the topology of the reference sequence II-B at the root of the subtype II-A and variant 1 of subtype II-C reference sequences, but the most recently described Cas9 sequences of subtype II-D and II-C2 are the closest to the IscB ancestor (Fig. 2). Of the 57 putative novels Cas9 used here, Clade 1 lies between ancestral sequences and the Cas9 proteins II-B, and five of the six sequences of clade 1 belong to the phylum *Verrucomicrobiota*, specifically to the orders *Pedosphaerales* and *Chthoniobacterales*. In addition, clades 2 and 3 located in the tree near reference Cas9 sequences were also defined with operational intentions. In general terms, a taxonomic trend was observed in the different clades. Clade 2 has 43 members; among those classified, there are 15 belonging to the phylum *Bacteroidota*, 12 to *Planctomycetota*, and 10 to *Pseudomonadota*. Clade 3 harbors only Cas9 belonging to the phylum *Pseudomonadota* and the class *Alphaproteobacteria*. The Cas9 sequences

that group near the reference sequence of subtype II-C1 show only members of *Pseudomonadota* belonging to the class *Gammaproteobacteria* (Fig. 2). These results suggest that hot spring metagenomes seem to harbor new variants of Cas9 orthologs, some of which may be related to sequences closer to the proposed ancestor as subtypes II-D and II-C2.

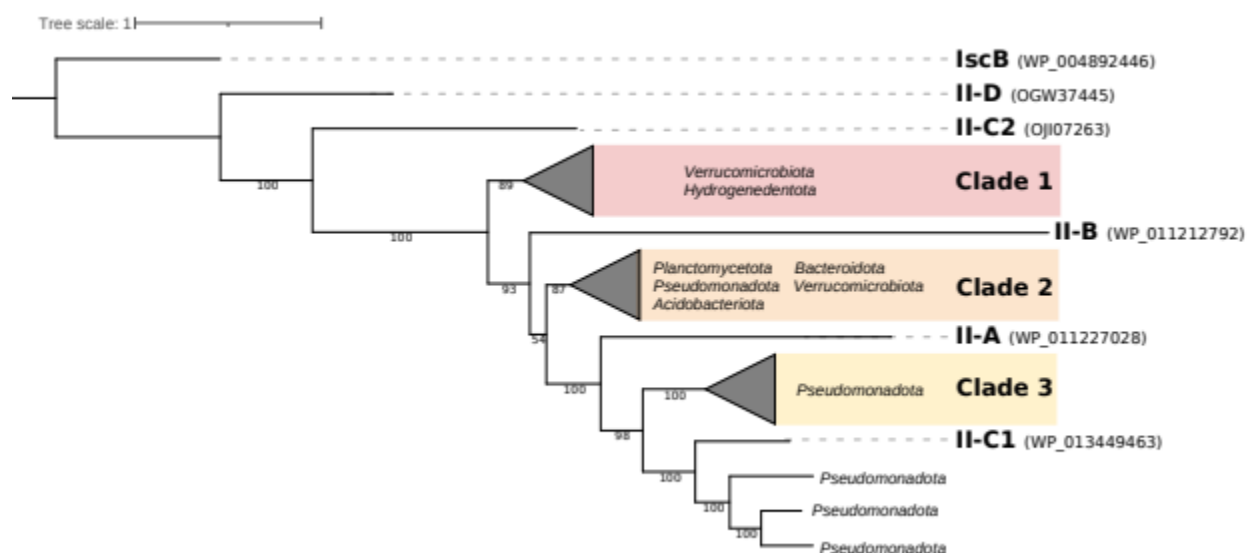


Figure 2. Maximun-Likelihood phylogenetic tree of 57 Cas9 proteins alongside reference sequences (with GenBank IDs) and IscB as an outgroup. Clades composed exclusively of hot spring sequences are colored, and related phyla inside the clade are indicated.

To further analyze the similarity between the 57 putative novel Cas9 recovered from hot springs, a similarity matrix was constructed using the tree sequences (Fig. 3). Most sequences show similarity values close to or below 25%. The Cas9 members of clade 1 showed higher similarity values among some of them but low with the rest of the sequences (similarity close to or below 25%, Fig. 3). Cas9 clade 2 shows low similarity values with the rest of the proteins in the matrix build, especially with reference and clade 1 sequences. The outgroup IscB and the type II reference sequences show low similarity values with all the selected hot spring sequences. Additionally, the similarity of these Cas9 to the nr database by BLASTp was analyzed. Of the 57 proteins, 38 show percentage identity values below 70% (detailed percent of identity with the best hit of nr alignment can be retrieved in Supplementary Table S4). Domain analyses performed with CD-search revealed that the 57 Cas9 obtained present the domains described for Cas9 (RuvC and HNH). To confirm the specific loci structure of these proteins, due to low similarity values obtained by comparison to public databases, the four upstream and downstream four ORFs were annotated in the Cas9 sequences that presented similarity values lower than 70% and harbored a CRISPR array in their vicinity (using CD-search with the above-mentioned parameters and database). In total, 20 Cas9 loci were considered, and the annotation of ORFs was analyzed. The most common genes in the vicinity of Cas9 were Cas1, Cas2, and a protein with the pfam18962 domain (present six times), a domain related to secretion systems in *Bacteria* (Glew et al., 2012). The remaining ORFs were present only once or twice, and no Cas4 protein was identified. Phylogenetic and similarity data indicate high genetic variability in these Cas9 with potentially new variants phylogenetically close to the mentioned ancestral homologs of the gene.

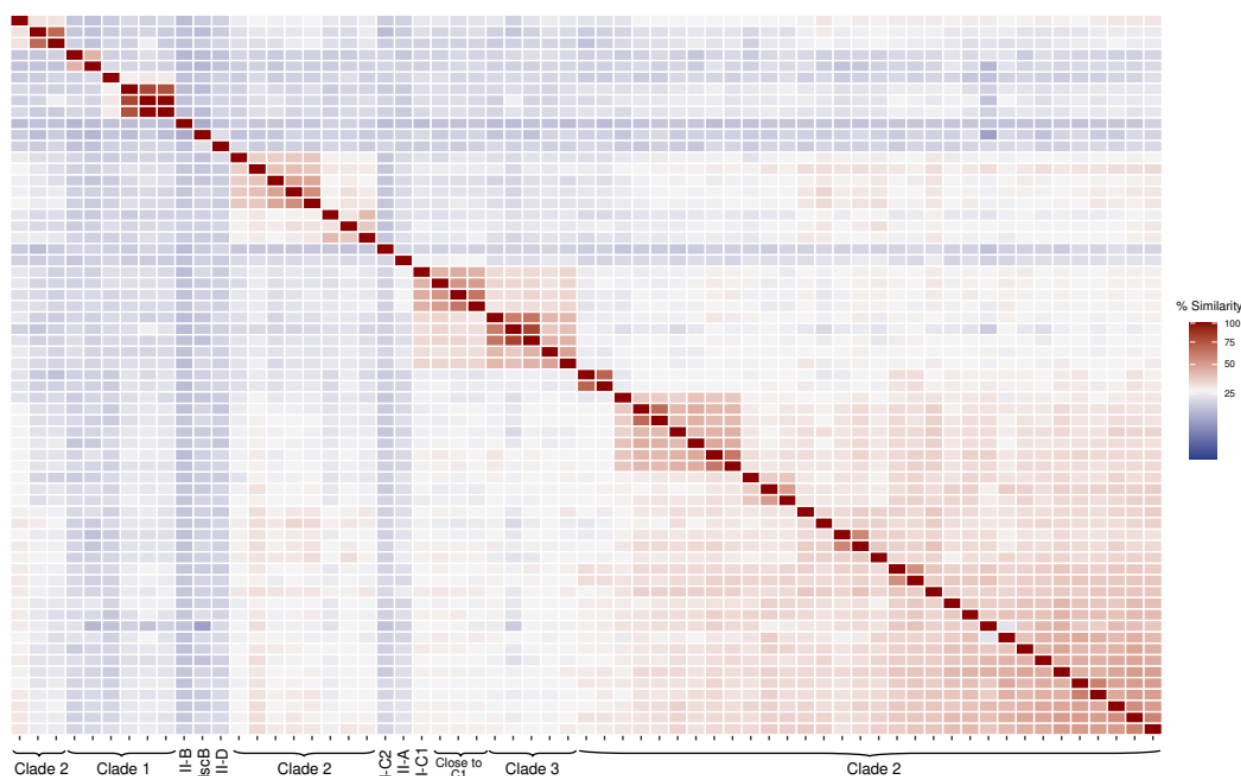


Figure 3. Percent identity matrix of 57 Cas9 proteins alongside reference and IscB outgroup sequences (sequences used in the phylogenetic tree of **Figure 2**). Phylogenetic clades are indicated at the bottom.

Discussion

The discovery and application of CRISPR-Cas systems have provided new insights into bacterial and archaeal immunity systems and a new paradigm for gene editing tools (Burstein et al., 2016; Doron et al., 2018). Expanding the descriptions by searching for new CRISPR-Cas systems in natural environments could contribute to a better understanding of the physiology of defense and regulatory mechanisms in *Archaea* and *Bacteria* and to developing better and, eventually, new biotechnological tools. In this study, the description of known and novel (particularly class 2) CRISPR-Cas systems in their most thriving environment known to date, such as thermal environments (Anderson et al., 2011; Lan et al., 2022; Weissman et al., 2019), was addressed. For this task, a delimited temperature range was considered, which excluded most archaeal members in which class 2 CRISPR-Cas systems are practically absent (Makarova et al., 2020). In fact, the archaeal phylum *Thermoproteota* does not reveal class 2 systems here (Fig. 1). Furthermore, the temperature range used here is of interest as it is less explored regarding viruses (Zablocki et al., 2018) and virus-host relationships, which could contribute to elucidating the molecular mechanisms involved in the CRISPR-Cas immune response.

Several uncommon class 1 types in hot springs worldwide

The abundance of class 1 CRISPR-Cas systems observed in hot springs is consistent with previous database descriptions and other environments (Crawley et al., 2018; Makarova et al., 2020; Russel et al., 2020). In fact, since the first classification of CRISPR-Cas systems to date, class 1 has been the most frequent and abundant in nature (Makarova et al., 2015, 2011). Data further suggest that class 1 CRISPR-Cas systems can generate a wider spacer diversity than class 2, and that multiprotein effectors are more versatile for expression and regulation, allowing a better fitness of class 1 systems (Watson et al., 2021). As hot spring CRISPR-Cas prevalence follows the tendency observed in databases (Makarova et al., 2020), it seems reasonable that this would also apply to hot spring environments. However, it might also be speculated that high temperatures affect class 2 systems because the big and single protein effector could be more affected by mutations. It would be interesting to probe if hot springs' highly deleterious non-silent mutation rate (Drake, 2009) is more important for the multidomain architecture of class 2 effectors proteins than for class 1. Interestingly, the second most represented class 1 subtype in hot springs, I-B after III-A, has been described to be involved in non-defensive functions (Faure et al., 2019). Specifically, some I-B and I-F systems have been observed in transposons guided by RNA-effector complex (CAST, CRISPR-associated transposase) (Peters et al., 2017; Strecker et al., 2019). Considering the discovery of transposable elements families exclusive from hot springs and related to CRISPR-Cas (Salgado et al., 2022), the research on subtypes I-B of hot springs eventually could shed light on CAST evolution.

Página | 12

The description of four type IV subtypes in the hot springs analyzed also reinforces the suggested role of CRISPR-Cas in non-defensive functions. For instance, all these systems have been identified in MGE from several environments (acid, thermal, alkaline, halophilic) (Moya-Beltrán et al., 2021; Makarova et al., 2020) and are assumed to be involved in plasmid exclusion (Makarova et al., 2022). Furthermore, type IV systems are closely related to the CAST mentioned above. Considering the high cell density of some hot spring habitats like microbial mats, it is possible that type IV systems could participate in the regulatory process where conjugative elements should be finely regulated. In all, Cas related to systems mentioned non-defensive task also suggests hot springs as a model for the study of alternative functions of CRISPR-Cas. For this reason, a detailed description of the uncommon loci of systems I-B, I-F and subtypes IV is relevant, as these could shed light on the intriguing interface between immunity and the regulation of self and foreign genetic material in natural microbial communities.

Compact Cas9 and taxa-specific Cas12 and Cas13 are present in hot springs

Regarding class 2 CRISPR-Cas systems, the high prevalence of subtype II-D in hot springs worldwide is interesting because this subtype presents small Cas9 effectors (Altae-Tran et al., 2021). This locus was described without Cas1 and Cas2, suggesting that adaptation occurs in trans (Altae-Tran et al., 2021), and the small size of class 2 effectors has proven helpful in genomic editing tools (Yang & Patel, 2019). Therefore, the hot springs Cas9 described here could eventually be helpful for already established editing protocols. As for the remaining class 2 types (V and VI) reported in this study, the prevalence of V-D (Cas12d) subtype in the phylum *Patescibacteria* is striking (Fig. 1). This CRISPR-Cas subtype that was described in the same

bacterial group (Burstein et al., 2016), is characterized by the epibiont lifestyle on *Actinomycetota* of its few cultured members (Wang et al., 2023). The high cost of maintaining and expressing the class 2 CRISPR-Cas system (Zaayman & Wheatley, 2022) is probably offset by its mutualistic relationship. However, it could also be speculated to be associated with defensive cooperation between both microorganisms. The epibiont could defend against viruses by reducing and regulating the virus's contact with the host. This could be particularly important in a cell-dense space, such as hot spring sediments or microbial mats.

As for type VI CRISPR-Cas systems (Cas13), those observed in the phylum *Pseudomonadota* (Fig. 1) draw considerable attention because these systems are recognized as RNA degraders (Barrangou & Gersbach, 2017; Zhao et al., 2022). RNA viruses have been described in hot springs (Bolduc et al., 2012; Le Lay et al., 2023), thus making it plausible that some strains use defensive mechanisms against those viruses. On the other hand, as the subtype VI-A effector (Cas13a) has been one of the most applied tools beyond gene editing (Zhao et al., 2022), it is suggested that Cas13 proteins found in hot springs could be good candidates for biotechnological applications. The class 2 diversity described here, encompassing 16 subtypes, encourages further characterization of these proteins from these environments to elucidate their potential as molecular tools. In all cases, the new variants from hot springs found here also contribute to a better understanding of the evolution of these systems.

As rare the taxa, as rare the Cas9

The strategy defined to search for putative novel class 2 systems included the use of two independent tools (see methods). For the CRISPRimmunity, the results were ambiguous and inconclusive, whereas 57 putative novel genes were recognized through CRISPRCasTyper tool. Regarding the search for novel class 2 systems with the CRISPRimmunity suite, the results are inconclusive because no recognized domains were identified in the candidates. Trying to replicate the published pipeline does not allow for the obtaining of putative novel variants. The authors of CRISPRimmunity were contacted two times by e-mail to ask about this result, but there was no answer. Since the publication (Zhou et al., 2023) indicates that putative novel candidates presenting nuclease domains are searched using the same database under the same parameters used here (CDD), additional information about the CRISPRimmunity suite is needed to understand these results. As the strategy defined here does not rely only on CRISPRimmunity suite, the work progressed using only the results of CRISPRCasTyper.

The Cas9 tree with 57 variants from hot spring (Fig. 2) is consistent with the published phylogenetic analyses using Cas9 proteins in revealing a high diversity of the gene (Al-Shayeb et al., 2020; Fonfara et al., 2014; Gasiunas et al., 2020; Ipoutcha et al., 2019; Pearson et al., 2015; Schmidt et al., 2019). However, most phylogenetic reconstructions of Cas9 do not include an outgroup despite the ancestral relationship depicted for the IscB (Altae-Tran et al., 2021), making evolutionary tracking difficult. Additionally, comprehensive analyses of Cas9 orthologs (Esquerra-Ruvira et al., 2023; Gasiunas et al., 2020) do not consider reference sequences of subtypes II-D and II-C2 due to their recent description (Aliaga Goltsman et al., 2022). Here, a rooted tree where clade 1 is located close to the likely ancestor of Cas9 (Fig. 2) was used. Clade 1 is composed of six Cas9 sequences, and only one does not belong to the phylum *Verrucomicrobiota*, a group that is among

the less abundant and less characterized in hot springs (Hou et al., 2008; Nixon et al., 2019), so may be expected to find uncommon Cas9 variants as a result of quite specific virus-host coevolution in the phylum. Some of the more abundant microbes in hot springs globally, such as members of *Bacteroidota* and *Pseudomonadota*, show high representation in clades 2, and 3, with sequences close to the reference Cas9 of type II-C2. Also, it is plausible that the particular virus-host relationship in hot springs (Guajardo-Leiva et al., 2021; Jarett et al., 2020; Sharma et al., 2018) is influencing the prevalence of some type II system variants present in specific bacterial groups. The specificity of viruses infecting bacterial species could maintain the utility of these Cas9 proteins and promote their evolution. Further studies that consider virus-host relationships in specific taxa in detail would allow a better understanding of the importance of subtype specificity for these taxa. Similarly, the sequence similarity among tree members (Fig. 3), which indicates small clusters of similarity, supports the hypothesis of a local evolution in hot springs. However, it should be noted that the continuous increase in sequenced metagenome availability and the astonishing diversity of Cas9 orthologs described in the database (Gasiunas et al., 2020; Ipoutcha et al., 2019) urge a cautious approach when proposing the identification of rare Cas9. Undoubtedly, similarity inside Cas9, with low similarity values among hot spring orthologs, strongly suggests additional experimental descriptions of Cas9 loci to demonstrate their putative novelty.

The potential function of Cas9 in enhanced cell security

The description of pfam18962 as the third most common domain in the vicinity of selected Cas9 agrees with recent findings showing a strong relationship between CRISPR-Cas systems and the membranome (Makarova et al., 2019; Medina-Aparicio et al., 2021; Rubio et al., 2023; VanderWal et al., 2023). That domain is described as targeting outer membrane secretion systems (Glew et al., 2012; Veith et al., 2013), and a link to iron uptake has also been mentioned (Conrad et al., 2022). An ancillary function enhancing the immune response could be involved, as described for subtypes I-F (Medina-Aparicio et al., 2021) and VI-B (VanderWal et al., 2023). However, no descriptions in this line have been made for Cas9 despite being the most numerous class 2 protein (Makarova et al., 2022). An intriguing hypothesis suggests that the CRISPR-Cas system might become necessary for the cell when a beneficial membrane protein is expressed but used by viruses to enter the cell (Rubio et al., 2023). Therefore, it might be speculated that the coexpression of pfam18962 and Cas9 assures an immune level control on the secretion system's function to target incoming viruses eventually. Additional molecular characterizations of the Cas9 loci vicinity members and their response to viral challenges could shed light on their function.

Hot springs as a CRISPR-Cas museum

The diversity of CRISPR-Cas systems, including type II systems, can be explained through striking events of exaptation, accretion, recombination and rearrangements (Koonin & Makarova, 2022; Makarova et al., 2022; Shams et al., 2021). Therefore, attempting to track the evolution of such diverse and fast-evolving systems (Makarova et al., 2018) seems impossible, especially for several known subtypes. However, explorations in specific environments like hot springs could greatly help. Despite being the least abundant in nature, the single effector-protein of class 2 CRISPR-Cas systems is a good candidate for tracking the evolution of these bacterial defense systems because they are monophyletic for some types (Koonin & Krupovic, 2023). The evolution of

type II but also type V CRISPR-Cas systems has been proposed from the *IscB* and *TnpB* transposon family IS200/IS605 genes, respectively (Makarova et al., 2020), and even the RNA-guided endonuclease activity of these proteins has been demonstrated (Altae-Tran et al., 2021). The mention performed here, including already-known class 2 systems from hot springs, is the first step towards a more profound analysis that could provide new information on their evolution. On the other hand, the evolution of class 1 systems presents a more complex picture than class 2 (Koonin & Makarova, 2019). In this regard, the numerous class 1 Cas proteins revealed here could fill the hypothetical gaps about their evolution and relationship with the MGEs and the membranome. For example, the diversity of type IV systems observed in hot springs globally clearly favors its study along with the systems already described. Further explorations considering locus analyses, ancillary gene type, CRISPR array repeats and spacers could better comprehend CRISPR-Cas systems in hot springs and its relationship with the systems described for other environments.

Conclusions

Hot springs environments have been reported to harbor the highest prevalence of CRISPR-Cas systems in nature. However, no survey has examined a detailed description on a global scale in the mesothermophilic and circumneutral pH ranges. Moreover, the scarce description of these CRISPR-Cas systems in thermophiles has been mainly concentrated in a few species of hyperthermophilic organisms. This study described the types and subtypes of CRISPR-Cas systems in mesothermophilic (40°C-80°C) and circumneutral pH (6-8) ranges globally in hot springs. All types already described to date were identified in hot springs, spreading in 32 of the 37 known subtypes, and revealing that class 1 types are the most abundant. Notably, all type IV systems described to date were identified in hot springs, pointing to these environments as relevant sites to understand the evolution of class 1 systems better. Novel putative variants of class 2 CRISPR-Cas systems, including the extensively used Cas9 protein, were also identified. Phylogenetic analyses of these Cas9 alongside reference sequences reveal distinct clades between already-known proteins and more recently described proteins located close to the root of the tree. Interestingly, the new sequences exhibit a very low similarity value between them and the Cas9 deposited in public databases, encouraging further studies for *in vitro* validation. The uncommon class 1 types observed and the potential new Cas9 variants, including new potential associations with membrane proteins, corroborate that CRISPR-Cas systems are poorly explored in natural communities, among which hot springs are particularly left behind. All these results indicate that hot springs are suitable environments to study the ecology and evolution of CRISPR-Cas systems and to mine for new effectors with biotechnological potential.

Conflict of interest

The author declare absence of any conflict of interest.

Supplementary material

Supplementary material (Supplementary Tables S1-S4) is available under request.



- Probst, A. J.; Power, M. E.; Tringe, S. G.; Li, W.-J.; Wrighton, K.; Harrison, S.; Morowitz, M.; Relman, D. A.; Doudna, J. A.; Lehours, A.-C.; Warren, L.; Cate, J. H. D.; Santini, J. M., & Banfield, J. F. (2020). Clades of huge phages from across Earth's ecosystems. *Nature*, 578, 425-431. <https://doi.org/10.1038/s41586-020-2007-4>
- Altae-Tran, H.; Kannan, S.; Demircioglu, F. E.; Oshiro, R.; Nety, S. P.; McKay, L. J.; Dlakić, M.; Inskip, W. P.; Makarova, K. S.; Macrae, R. K.; Koonin, E. V., & Zhang, F. (2021). The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science*, 374, 57-65. <https://doi.org/10.1126/science.abj6856>
- Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Anderson, R. E.; Brazelton, W. J. & Baross, J. A. (2011). Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiology Ecology*, 77(1), 120-133. <https://doi.org/10.1111/j.1574-6941.2011.01090.x>
- Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D. A. & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819), 1709-1712. <https://doi.org/10.1126/science.1138140>
- Barrangou, R. & Gersbach, C. A. (2017). Expanding the CRISPR Toolbox: Targeting RNA with Cas13b. *Molecular Cell*, 65(5), 582-584. <https://doi.org/10.1016/j.molcel.2017.02.002>
- Bolduc, B.; Shaughnessy, D. P.; Wolf, Y. I.; Koonin, E. V.; Roberto, F. F. & Young, M. (2012). Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone hot springs. *Journal of Virology*, 86(10), 5562-5573. <https://doi.org/10.1128/JVI.07196-11>
- Bowers, R. M.; Kyrpides, N. C.; Stepanauskas, R.; Harmon-Smith, M.; Doud, D.; Reddy, T. B. K.; Schulz, F.; Jarett, J.; Rivers, A. R.; Elie-Fadrosh, E. A.; Tringe, S. G.; Ivanova, N. N.; Copeland, A.; Clum, A.; Becraft, E. D.; Malmstrom, R. R.; Birren, B.; Podar, M.; Bork, P. . . . & Woyke, T. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, 35(7), 725-731. <https://doi.org/10.1038/nbt.3893>
- Broniewski, J. M.; Meaden, S.; Paterson, S.; Buckling, A. & Westra, E. R. (2020). The effect of phage genetic diversity on bacterial resistance evolution. *ISME Journal*, 14(4), 828-836. <https://doi.org/10.1038/s41396-019-0577-7>
- Burstein, D.; Harrington, L. B.; Strutt, S. C.; Probst, A. J.; Anantharaman, K.; Thomas, B. C.; Doudna, J. A. & Banfield, J. F. (2016). New CRISPR-Cas systems from uncultivated microbes. *Nature*, 542(7641), 237-241. <https://doi.org/10.1038/nature21059>
- Capella-Gutiérrez, S.; Silla-Martínez, J. M. & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972-1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Chaumeil, P. A.; Mussig, A. J.; Hugenholtz, P. & Parks, D. H. (2020). GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics*, 36(6), 1925-1927. <https://doi.org/10.1093/bioinformatics/btz848>

- Chevallereau, A.; Meaden, S.; van Houte, S.; Westra, E. R. & Rollie, C. (2019). The effect of bacterial mutation rate on the evolution of CRISPR-Cas adaptive immunity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1772), 20180094. <https://doi.org/10.1098/rstb.2018.0094>
- Childs, L. M.; England, W. E.; Young, M. J.; Weitz, J. S. & Whitaker, R. J. (2014). CRISPR-induced distributed immunity in microbial populations. *PLoS ONE*, 9(4), e101710. <https://doi.org/10.1371/journal.pone.0101710>
- Conrad, R. A.; Evenhuis, J. P.; Lipscomb, R. S.; Pérez-Pascual, D.; Stevick, R. J.; Birkett, C.; Ghigo, J.-M. & McBride, M. J. (2022). *Flavobacterium columnare* ferric iron uptake systems are required for virulence. *Frontiers in Cellular and Infection Microbiology*, 12. <https://doi.org/10.3389/fcimb.2022.1029833>
- Crawley, A. B.; Henriksen, J. R. & Barrangou, R. (2018). CRISPRdisco: An automated pipeline for the discovery and analysis of CRISPR-Cas systems. *CRISPR Journal*, 1(3), 171-181. <https://doi.org/10.1089/crispr.2017.0022>
- Doron, S.; Melamed, S.; Ofir, G.; Leavitt, A.; Lopatina, A.; Keren, M.; Amitai, G. & Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*, 359(6379), eaar4120. <https://doi.org/10.1126/science.aar4120>
- Drake, J. W. (2009). Avoiding dangerous missense: Thermophiles display especially low mutation rates. *PLoS Genetics*, 5(1), e1000520. <https://doi.org/10.1371/journal.pgen.1000520>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792-1797. <https://doi.org/10.1093/nar/gkh340>
- Esquerra-Ruvira, B.; Baquedano, I.; Ruiz, R.; Fernández, A.; Montoliu, L. & Mojica, F. J. M. (2023). Identification of the EH CRISPR-Cas9 system on a metagenome and its application to genome engineering. *Microbial Biotechnology*, 16(6), 1505-1523. <https://doi.org/10.1111/1751-7915.14266>
- Fan, C.; Zhang, W.; Su, X.; Ji, W.; Luo, H.; Zhang, Y.; Liu, B.; Yao, B.; Huang, H. & Xu, X. (2021). CRISPR/Cas9-mediated genome editing directed by a 5S rRNA-tRNA^{Gly} hybrid promoter in the thermophilic filamentous fungus *Humicola insolens*. *Biotechnology for Biofuels*, 14(1), 206. <https://doi.org/10.1186/s13068-021-02057-y>
- Faure, G.; Shmakov, S. A.; Yan, W. X.; Cheng, D. R.; Scott, D. A.; Peters, J. E.; Makarova, K. S. & Koonin, E. V. (2019). CRISPR-Cas in mobile genetic elements: Counter-defense and beyond. *Nature Reviews Microbiology*, 17(8), 513-525. <https://doi.org/10.1038/s41579-019-0204-7>
- Fonfara, I.; Le Rhun, A.; Chylinski, K.; Makarova, K. S.; Lécrivain, A.-L.; Bzdrenga, J.; Koonin, E. V. & Charpentier, E. (2014). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Research*, 42(4), 2577-2590. <https://doi.org/10.1093/nar/gkt1074>
- Friedman, R.; Drake, J. W. & Hughes, A. L. (2004). Genome-wide patterns of nucleotide substitution reveal stringent functional constraints on the protein sequences of thermophiles. *Genetics*, 167(4), 1507-1512. <https://doi.org/10.1534/genetics.104.026344>
- Fu, L.; Niu, B.; Zhu, Z.; Wu, S. & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150-3152. <https://doi.org/10.1093/bioinformatics/bts565>

- Fuchs, R. T.; Curcuru, J. L.; Mabuchi, M.; Noireterre, A.; Weigele, P. R.; Sun, Z. & Robb, G. B. (2022). Characterization of Cme and Yme thermostable Cas12a orthologs. *Communications Biology*, 5, 1-16. <https://doi.org/10.1038/s42003-022-03275-2>
- Gasiunas, G.; Young, J. K.; Karvelis, T.; Kazlauskas, D.; Urbaitis, T.; Jasnauskaite, M.; Grusyte, M. M.; Paulraj, S.; Wang, P.-H.; Hou, Z.; Dooley, S. K.; Cigan, M.; Alarcon, C.; Chilcoat, N. D.; Bigelyte, G.; Curcuru, J. L.; Mabuchi, M.; Sun, Z.; Fuchs, R. T.; Schildkraut, E.; Weigele, P. R.; Jack, W. E.; Robb, G. B.; Venclovas, Č. & Siksnys, V. (2020). A catalogue of biochemically diverse CRISPR-Cas9 orthologs. *Nature Communications*, 11(1), 5512. <https://doi.org/10.1038/s41467-020-19344-1>
- Gault, S.; Higgins, P. M.; Cockell, C. S. & Gillies, K. (2021). A meta-analysis of the activity, stability, and mutational characteristics of temperature-adapted enzymes. *Bioscience Reports*, 41(6), BSR20210336. <https://doi.org/10.1042/BSR20210336>
- Glew, M. D.; Veith, P. D.; Peng, B.; Chen, Y.-Y.; Gorasia, D. G.; Yang, Q.; Slakeski, N.; Chen, D.; Moore, C.; Crawford, S. & Reynolds, E. C. (2012). PG0026 is the C-terminal signal peptidase of a novel secretion system of *Porphyromonas gingivalis*. *Journal of Biological Chemistry*, 287(30), 24605-24617. <https://doi.org/10.1074/jbc.M112.369223>
- Gostimskaya, I. (2022). CRISPR-Cas9: A history of its discovery and ethical considerations of its use in genome editing. *Biochemistry (Moscow)*, 87(7), 777-788. <https://doi.org/10.1134/S0006297922080090>
- Guajardo-Leiva, S.; Pedrós-Alió, C.; Salgado, O.; Pinto, F. & Díez, B. (2018). Active crossfire between cyanobacteria and cyanophages in phototrophic mat communities within hot springs. *Frontiers in Microbiology*, 9, 2039. <https://doi.org/10.3389/fmicb.2018.02039>
- Guajardo-Leiva, S.; Santos, F.; Salgado, O.; Regeard, C.; Quillet, L. & Díez, B. (2021). Unveiling ecological and genetic novelty within lytic and lysogenic viral communities of hot spring phototrophic microbial mats. *Microbiology Spectrum*, 9(6), e00694-21. <https://doi.org/10.1128/spectrum.00694-21>
- Harrington, L. B.; Paez-Espino, D.; Staahl, B. T.; Chen, J. S.; Ma, E.; Kyrpides, N. C. & Doudna, J. A. (2017). A thermostable Cas9 with increased lifetime in human plasma. *Nature Communications*, 8, 1-8. <https://doi.org/10.1038/s41467-017-01408-4>
- Harsij, Z.; Ghafoorzadeh, Z. & Goharian, E. (2024). The CRISPR revolution: Unraveling the mysteries of Life's genetic code. *Gene*, 892, 147870. <https://doi.org/10.1016/j.gene.2023.147870>
- Heidelberg, J. F.; Nelson, W. C.; Schoenfeld, T. & Bhaya, D. (2009). Germ warfare in a microbial mat community: CRISPRs provide insights into the coevolution of host and viral genomes. *PLoS ONE*, 4(3), e4169. <https://doi.org/10.1371/journal.pone.0004169>
- Held, N. L. & Whitaker, R. J. (2009). Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. *Environmental Microbiology*, 11(2), 457-466. <https://doi.org/10.1111/j.1462-2920.2008.01784.x>
- Hille, F.; Richter, H.; Wong, S. P.; Bratovič, M.; Ressel, S. & Charpentier, E. (2018). The biology of CRISPR-Cas: Backward and forward. *Cell*, 172(6), 1239-1259. <https://doi.org/10.1016/j.cell.2017.11.032>
- Hou, S.; Makarova, K. S.; Saw, J. H.; Senin, P.; Ly, B. V.; Zhou, Z.; Ren, Y.; Wang, J.; Galperin, M. Y.; Omelchenko, M. V.; Wolf, Y. I.; Yutin, N.; Koonin, E. V.; Stott, M. B.; Mountain, B. W.; Crowe, M. A.; Smirnova, A. V.; Dunfield, P. F.; Feng, L.; Wang, L. & Alam, M. (2008). Complete genome

- sequence of the extremely acidophilic methanotroph isolate V4, *Methylacidiphilum infernorum*, a representative of the bacterial phylum Verrucomicrobia. *Biology Direct*, 3, 26. <https://doi.org/10.1186/1745-6150-3-26>
- Ipoutcha, T.; Tsarnopoulos, I.; Talenton, V.; Gaspin, C.; Moisan, A.; Walker, C. A.; Brownlie, J.; Blanchard, A.; Thebault, P. & Sirand-Pugnet, P. (2019). Multiple origins and specific evolution of CRISPR/Cas9 systems in minimal bacteria (*Mollicutes*). *Frontiers in Microbiology*, 10, 2701. <https://doi.org/10.3389/fmicb.2019.02701>
- Iranzo, J.; Lobkovsky, A. E.; Wolf, Y. I. & Koonin, E. V. (2013). Evolutionary dynamics of the prokaryotic adaptive immunity system CRISPR-Cas in an explicit ecological context. *Journal of Bacteriology*, 195(17), 3834-3844. <https://doi.org/10.1128/JB.00412-13>
- Jarett, J. K.; Džunková, M.; Schulz, F.; Roux, S.; Paez-Espino, D.; Eloë-Fadrosh, E.; Jungbluth, S. P.; Ivanova, N.; Spear, J. R.; Carr, S. A.; Trivedi, C. B.; Corsetti, F. A.; Johnson, H. A.; Becraft, E.; Kyrpides, N.; Stepanauskas, R. & Woyke, T. (2020). Insights into the dynamics between viruses and their hosts in a hot spring microbial mat. *ISME Journal*, 14(10), 2527-2541. <https://doi.org/10.1038/s41396-020-0705-4>
- Jiang, W.; Maniv, I.; Arain, F.; Wang, Y.; Levin, B. R. & Marraffini, L. A. (2013). Dealing with the evolutionary downside of CRISPR immunity: Bacteria and beneficial plasmids. *PLoS Genetics*, 9(6), e1003844. <https://doi.org/10.1371/journal.pgen.1003844>
- Koonin, E. V. & Krupovic, M. (2023). New faces of prokaryotic mobile genetic elements: Guide RNAs link transposition with host defense mechanisms. *Current Opinion in Systems Biology*, 36, 100473. <https://doi.org/10.1016/j.coisb.2023.100473>
- Koonin, E. V. & Makarova, K. S. (2022). Evolutionary plasticity and functional versatility of CRISPR systems. *PLoS Biology*, 20(7), e3001481. <https://doi.org/10.1371/journal.pbio.3001481>
- Koonin, E. V. & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1772), 20180087. <https://doi.org/10.1098/rstb.2018.0087>
- Koonin, E. V. & Makarova, K. S. (2009). CRISPR-Cas: An adaptive immunity system in prokaryotes. *F1000 Biology Reports*, 6, 1-6. <https://doi.org/10.3410/B1-95>
- Koonin, E. V. & Wolf, Y. I. (2016). Just how Lamarckian is CRISPR-Cas immunity: The continuum of evolvability mechanisms. *Biology Direct*, 11, 9. <https://doi.org/10.1186/s13062-016-0111-z>
- Lan, X. R.; Liu, Z. L. & Niu, D. K. (2022). Precipitous increase of bacterial CRISPR-Cas abundance at around 45°C. *Frontiers in Microbiology*, 13, 542. <https://doi.org/10.3389/fmicb.2022.773114>
- Lander, E. S. (2016). The heroes of CRISPR. *Cell*, 164(1), 18-28. <https://doi.org/10.1016/j.cell.2015.12.041>
- Le Lay, C.; Stott, M. B.; Shi, M.; Sadiq, S. & Holmes, E. C. (2023). A metatranscriptomic analysis of geothermal hot springs reveals diverse RNA viruses including the phylum Lenarviricota. *Virology*, 587, 109873. <https://doi.org/10.1016/j.virol.2023.109873>
- Le, Y.; Fu, Y. & Sun, J. (2020). Genome editing of the anaerobic thermophile *Thermoanaerobacter ethanolicus* using thermostable Cas9. *Applied and Environmental Microbiology*, 87(18), e01773-20. <https://doi.org/10.1128/AEM.01773-20>

- Le, Y. & Sun, J. (2022). Chapter one - CRISPR/Cas genome editing systems in thermophiles: Current status, associated challenges, and future perspectives. In G. M. Gadd & S. Sariaslani (Eds.), *Advances in Applied Microbiology* (pp. 1-30). Academic Press. <https://doi.org/10.1016/bs.aambs.2022.02.001>
- Madeira, F.; Park, Y. M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A. R. N.; Potter, S. C.; Finn, R. D. & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, 47(W1), W636-W641. <https://doi.org/10.1093/nar/gkz268>
- Makarova, K. S.; Gao, L.; Zhang, F. & Koonin, E. V. (2019). Unexpected connections between type VI-B CRISPR-Cas systems, bacterial natural competence, ubiquitin signaling network and DNA modification through a distinct family of membrane proteins. *FEMS Microbiology Letters*, 366(20), fnz088. <https://doi.org/10.1093/femsle/fnz088>
- Makarova, K. S.; Grishin, N. V.; Shabalina, S. A.; Wolf, Y. I. & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*, 1, 7. <https://doi.org/10.1186/1745-6150-1-7>
- Makarova, K. S.; Haft, D. H.; Barrangou, R.; Brouns, S. J. J.; Charpentier, E.; Horvath, P.; Moineau, S.; Mojica, F. J. M.; Wolf, Y. I.; Yakunin, A. F.; van der Oost, J. & Koonin, E. V. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology*, 9(6), 467-477. <https://doi.org/10.1038/nrmicro2577>
- Makarova, K. S.; Wolf, Y. I.; Alkhnbashi, O. S.; Costa, F.; Shah, S. A.; Saunders, S. J.; Barrangou, R.; Brouns, S. J. J.; Charpentier, E.; Haft, D. H.; Horvath, P.; Moineau, S.; Mojica, F. J. M.; Terns, R. M.; Terns, M. P.; White, M. F.; Yakunin, A. F.; Garrett, R. A.; van der Oost, J. ... & Koonin, E. V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*, 13(11), 722-736. <https://doi.org/10.1038/nrmicro3569>
- Makarova, K. S.; Wolf, Y. I.; Iranzo, J.; Shmakov, S. A.; Alkhnbashi, O. S.; Brouns, S. J. J.; Charpentier, E.; Cheng, D.; Haft, D. H.; Horvath, P.; Moineau, S.; Mojica, F. J. M.; Scott, D.; Shah, S. A.; Siksnys, V.; Terns, M. P.; Venclovas, Č.; White, M. F.; Yakunin, A. F.; Yan, W. ... & Koonin, E. V. (2020). Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants. *Nature Reviews Microbiology*, 18(2), 67-83. <https://doi.org/10.1038/s41579-019-0299-x>
- Makarova, K. S.; Wolf, Y. I. & Koonin, E. V. (2022). Evolutionary classification of CRISPR-Cas systems. In *Crispr* (pp. 13-38). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781683673798.ch2>
- Makarova, K. S.; Wolf, Y. I. & Koonin, E. V. (2018). Classification and nomenclature of CRISPR-Cas systems: Where from here? *CRISPR Journal*, 1(6), 325-336. <https://doi.org/10.1089/crispr.2018.0033>
- Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C. J.; Lu, S.; Chitsaz, F.; Derbyshire, M. K.; Geer, R. C.; Gonzales, N. R.; Gwadz, M.; Hurwitz, D. I.; Lu, F.; Marchler, G. H.; Song, J. S.; Thanki, N.; Wang, Z.; Yamashita, R. A.; Zhang, D. ... & Bryant, S. H. (2017). CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, 45(D1), D200-D203. <https://doi.org/10.1093/nar/gkw1129>
- Marchler-Bauer, A. & Bryant, S. H. (2004). CD-Search: Protein domain annotations on the fly. *Nucleic Acids Research*, 32(W1), W327-W331. <https://doi.org/10.1093/nar/gkh454>

- Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature*, 526(7571), 55-61. <https://doi.org/10.1038/nature15386>
- Meaden, S.; Biswas, A.; Arkhipova, K.; Morales, S. E.; Dutilh, B. E.; Westra, E. R. & Fineran, P. C. (2022). High viral abundance and low diversity are associated with increased CRISPR-Cas prevalence across microbial ecosystems. *Current Biology*, 32(1), 220-227.e5. <https://doi.org/10.1016/J.CUB.2021.10.038>
- Medina-Aparicio, L.; Rodriguez-Gutierrez, S.; Rebollar-Flores, J. E.; Martínez-Batallar, Á. G.; Mendoza-Mejía, B. D.; Aguirre-Partida, E. D.; Vázquez, A.; Encarnación, S.; Calva, E. & Hernández-Lucas, I. (2021). The CRISPR-Cas system is involved in OmpR genetic regulation for outer membrane protein synthesis in *Salmonella Typhi*. *Frontiers in Microbiology*, 12. <https://doi.org/10.3389/fmicb.2021.657404>
- Mikheenko, A.; Saveliev, V. & Gurevich, A. (2016). MetaQUAST: Evaluation of metagenome assemblies. *Bioinformatics*, 32(7), 1088-1090. <https://doi.org/10.1093/bioinformatics/btv697>
- Mohanraju, P.; Makarova, K. S.; Zetsche, B.; Zhang, F.; Koonin, E. V. & Van Der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science*, 353(6299), aad5147. <https://doi.org/10.1126/science.aad5147>
- Mojica, F. J. M. & Montoliu, L. (2016). On the origin of CRISPR-Cas technology: From prokaryotes to mammals. *Trends in Microbiology*, 24(9), 811-820. <https://doi.org/10.1016/j.tim.2016.06.005>
- Mougiakos, I.; Mohanraju, P.; Bosma, E. F.; Vrouwe, V.; Finger Bou, M.; Naduthodi, M. I. S.; Gussak, A.; Brinkman, R. B. L.; van Kranenburg, R. & van der Oost, J. (2017). Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nature Communications*, 8(1), 1647. <https://doi.org/10.1038/s41467-017-01591-4>
- Moya-Beltrán, A.; Makarova, K. S.; Acuña, L. G.; Wolf, Y. I.; Covarrubias, P. C.; Shmakov, S. A.; Silva, C.; Tolstoy, I.; Johnson, D. B., Koonin, E. V. & Quatrini, R. (2021). Evolution of Type IV CRISPR-Cas systems: Insights from CRISPR loci in integrative conjugative elements of *Acidithiobacillia*. *CRISPR Journal*, 4(6), 656-672. <https://doi.org/10.1089/crispr.2021.0051>
- Nguyen, L. T.; Macaluso, N. C.; Pizzano, B. L. M.; Cash, M. N.; Spacek, J.; Karasek, J.; Miller, M. R.; Lednický, J. A.; Dinglasan, R. R.; Salemi, M. & Jain, P. K. (2022). A thermostable Cas12b from *Brevibacillus* leverages one-pot discrimination of SARS-CoV-2 variants of concern. *eBioMedicine*, 77, 103926. <https://doi.org/10.1016/j.ebiom.2022.103926>
- Nguyen, L.-T.; Schmidt, H. A.; von Haeseler, A. & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(3), 268-274. <https://doi.org/10.1093/molbev/msu300>
- Nixon, S. L.; Daly, R. A.; Borton, M. A.; Solden, L. M.; Welch, S. A.; Cole, D. R.; Mouser, P. J.; Wilkins, M. J. & Wrighton, K. C. (2019). Genome-resolved metagenomics extends the environmental distribution of the *Verrucomicrobia* phylum to the deep terrestrial subsurface. *mSphere*, 4(1). <https://doi.org/10.1128/mSphere.00613-19>
- Nordberg, H.; Cantor, M.; Dusheyko, S.; Hua, S.; Poliakov, A.; Shabalov, I.; Smirnova, T.; Grigoriev, I. V. & Dubchak, I. (2014). The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Research*, 42(D1), D26-D31. <https://doi.org/10.1093/nar/gkt1069>

- Orakov, A.; Fullam, A.; Coelho, L. P.; Khedkar, S.; Szklarczyk, D.; Mende, D. R.; Schmidt, T. S. B. & Bork, P. (2021). GUNC: Detection of chimerism and contamination in prokaryotic genomes. *Genome Biology*, 22(1), 178. <https://doi.org/10.1186/s13059-021-02393-0>
- Parks, D. H.; Rinke, C.; Chuvochina, M.; Chaumeil, P. A.; Woodcroft, B. J.; Evans, P. N.; Hugenholtz, P. & Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, 2(11), 1533-1542. <https://doi.org/10.1038/s41564-017-0012-7>
- Parmar, K.; Dafale, N.; Pal, R.; Tikariha, H. & Purohit, H. (2018). An insight into phage diversity at environmental habitats using comparative metagenomics approach. *Current Microbiology*, 75(1), 132-141. <https://doi.org/10.1007/s00284-017-1357-0>
- Pearson, B. M.; Louwen, R.; van Baarlen, P. & van Vliet, A. H. M. (2015). Differential distribution of type II CRISPR-Cas systems in agricultural and nonagricultural *Campylobacter coli* and *Campylobacter jejuni* isolates correlates with lack of shared environments. *Genome Biology and Evolution*, 7(9), 2663-2679. <https://doi.org/10.1093/gbe/evv174>
- Peters, J. E.; Makarova, K. S.; Shmakov, S. & Koonin, E. V. (2017). Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proceedings of the National Academy of Sciences of the United States of America*, 114(40), E7358-E7366. <https://doi.org/10.1073/pnas.1709035114>
- Podar, P. T.; Yang, Z.; Björnsdóttir, S. H. & Podar, M. (2020). Comparative analysis of microbial diversity across temperature gradients in hot springs from Yellowstone and Iceland. *Frontiers in Microbiology*, 11, 1625. <https://doi.org/10.3389/fmicb.2020.01625>
- Pruitt, K. D.; Tatusova, T. & Maglott, D. R. (2004). NCBI reference sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 32(D1), D501-D504. <https://doi.org/10.1093/nar/gki025>
- Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J. & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), D590-D596. <https://doi.org/10.1093/nar/gks1219>
- Rubio, A.; Sprang, M.; Garzón, A.; Moreno-Rodríguez, A.; Pachón-Ibáñez, M. E.; Pachón, J.; Andrade-Navarro, M. A. & Pérez-Pulido, A. J. (2023). Analysis of bacterial pangenomes reduces CRISPR dark matter and reveals strong association between membranome and CRISPR-Cas systems. *Science Advances*, 9(4), eadd8911. <https://doi.org/10.1126/sciadv.add8911>
- Russel, J.; Pinilla-Redondo, R.; Mayo-Muñoz, D.; Shah, S. A. & Sørensen, S. J. (2020). CRISPRCasTyper: Automated identification, annotation, and classification of CRISPR-Cas loci. *CRISPR Journal*, 3(5), 462-469. <https://doi.org/10.1089/crispr.2020.0059>
- Salgado, O.; Guajardo-Leiva, S.; Moya-Beltrán, A.; Barbosa, C.; Ridley, C.; Tamayo-Leiva, J.; Quatrini, R.; Mojica, F. J. M. & Díez, B. (2022). Global phylogenomic novelty of the Cas1 gene from hot spring microbial communities. *Frontiers in Microbiology*, 13, 1069452. <https://doi.org/10.3389/fmicb.2022.1069452>
- Schmidt, S. T.; Yu, F. B.; Blainey, P. C.; May, A. P. & Quake, S. R. (2019). Nucleic acid cleavage with a hyperthermophilic Cas9 from an uncultured *Ignavibacterium*. *Proceedings of the National Academy of Sciences of the United States of America*, 116(47), 23100-23105. <https://doi.org/10.1073/pnas.1904273116>

- Shams, A.; Higgins, S. A.; Fellmann, C.; Laughlin, T. G.; Oakes, B. L.; Lew, R.; Kim, S.; Lukarska, M.; Arnold, M.; Staahl, B. T.; Doudna, J. A. & Savage, D. F. (2021). Comprehensive deletion landscape of CRISPR-Cas9 identifies minimal RNA-guided DNA-binding modules. *Nature Communications*, 12, 5664. <https://doi.org/10.1038/s41467-021-25992-8>
- Sharma, A.; Schmidt, M.; Kiesel, B.; Mahato, N. K.; Cralle, L.; Singh, Y.; Richnow, H. H.; Gilbert, J. A.; Arnold, W. & Lal, R. (2018). Bacterial and archaeal viruses of Himalayan hot springs at Manikaran modulate host genomes. *Frontiers in Microbiology*, 9, 3095. <https://doi.org/10.3389/fmicb.2018.03095>
- Shmakov, S.; Smargon, A.; Scott, D.; Cox, D.; Pyzocha, N.; Yan, W.; Abudayyeh, O. O.; Gootenberg, J. S.; Makarova, K. S.; Wolf, Y. I.; Severinov, K.; Zhang, F. & Koonin, E. V. (2017). Diversity and evolution of class 2 CRISPR-Cas systems. *Nature Reviews Microbiology*, 15(3), 169-182. <https://doi.org/10.1038/nrmicro.2016.184>
- Shmakov, S. A.; Sitnik, V.; Makarova, K. S.; Wolf, Y. I.; Severinov, K. V. & Koonin, E. V. (2017). The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *mBio*, 8(6), e01397-17. <https://doi.org/10.1128/mBio.01397-17>
- Shmakov, S. A.; Wolf, Y. I.; Savitskaya, E.; Severinov, K. V. & Koonin, E. V. (2020). Mapping CRISPR spaceromes reveals vast host-specific viromes of prokaryotes. *Communications Biology*, 3, 1-9. <https://doi.org/10.1038/s42003-020-1014-1>
- Snyder, J. C.; Bateson, M. M.; Lavin, M. & Young, M. J. (2010). Use of cellular CRISPR (clusters of regularly interspaced short palindromic repeats) spacer-based microarrays for detection of viruses in environmental samples. *Applied and Environmental Microbiology*, 76(23), 7251-7258. <https://doi.org/10.1128/AEM.01109-10>
- Sternberg, S. H.; Richter, H.; Charpentier, E. & Qimron, U. (2016). Adaptation in CRISPR-Cas systems. *Molecular Cell*, 61(6), 797-808. <https://doi.org/10.1016/j.molcel.2016.01.030>
- Strazzulli, A.; Fusco, S.; Cobucci-Ponzano, B.; Moracci, M. & Contursi, P. (2017). Metagenomics of microbial and viral life in terrestrial geothermal environments. *Reviews in Environmental Science and Biotechnology*, 16(3), 425-454. <https://doi.org/10.1007/s11157-017-9435-0>
- Strecker, J.; Ladha, A.; Gardner, Z.; Schmid-Burgk, J. L.; Makarova, K. S.; Koonin, E. V. & Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. *Science*, 363(6428), eaax9181. <https://doi.org/10.1126/science.aax9181>
- Tian, Y.; Liu, R. R.; Xian, W. D.; Xiong, M.; Xiao, M. & Li, W. J. (2020). A novel thermal Cas12b from a hot spring bacterium with high target mismatch tolerance and robust DNA cleavage efficiency. *International Journal of Biological Macromolecules*, 147, 376-384. <https://doi.org/10.1016/j.ijbiomac.2020.01.079>
- Tsui, T. K. M.; Hand, T. H.; Duboy, E. C. & Li, H. (2017). The impact of DNA topology and guide length on target selection by a cytosine-specific Cas9. *ACS Synthetic Biology*, 6(7), 1103-1113. <https://doi.org/10.1021/acssynbio.7b00050>
- Uritskiy, G. V.; DiRuggiero, J. & Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*, 6(1), 158. <https://doi.org/10.1186/s40168-018-0541-1>

- VanderWal, A. R.; Park, J.-U.; Polevoda, B.; Nicosia, J. K.; Molina Vargas, A. M.; Kellogg, E. H. & O'Connell, M. R. (2023). Csx28 is a membrane pore that enhances CRISPR-Cas13b-dependent antiphage defense. *Science*, 380(6633), 410-415. <https://doi.org/10.1126/science.abm1184>
- Veith, P. D.; Nor Muhammad, N. A.; Dashper, S. G.; Likić, V. A.; Gorasia, D. G.; Chen, D.; Byrne, S. J.; Catmull, D. V. & Reynolds, E. C. (2013). Protein substrates of a novel secretion system are numerous in the Bacteroidetes phylum and have in common a cleavable C-terminal secretion signal, extensive post-translational modification, and cell-surface attachment. *Journal of Proteome Research*, 12(10), 4449-4461. <https://doi.org/10.1021/pr400487b>
- Vollmers, J.; Wiegand, S.; Lenk, F. & Kaster, A.-K. (2022). How clear is our current view on microbial dark matter? (Re-)assessing public MAG & SAG datasets with MDMcleaner. *Nucleic Acids Research*, 50(13), e76. <https://doi.org/10.1093/nar/gkac294>
- Wang, Y.; Gallagher, L. A.; Andrade, P. A.; Liu, A.; Humphreys, I. R.; Turkarslan, S.; Cutler, K. J.; Arrieta-Ortiz, M. L.; Li, Y.; Radey, M. C.; McLean, J. S.; Cong, Q.; Baker, D.; Baliga, N. S.; Peterson, S. B. & Mougous, J. D. (2023). Genetic manipulation of Patescibacteria provides mechanistic insights into microbial dark matter and the epibiotic lifestyle. *Cell*, 186(22), 4803-4817.e13. <https://doi.org/10.1016/j.cell.2023.08.017>
- Watson, B. N. J.; Steens, J. A.; Staals, R. H. J.; Westra, E. R. & van Houte, S. (2021). Coevolution between bacterial CRISPR-Cas systems and their bacteriophages. *Cell Host & Microbe*, 29(5), 715-725. <https://doi.org/10.1016/J.CHOM.2021.03.018>
- Weinberger, A. D.; Sun, C. L.; Pluciński, M. M.; Denef, V. J.; Thomas, B. C.; Horvath, P.; Barrangou, R.; Gilmore, M. S.; Getz, W. M. & Banfield, J. F. (2012a). Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Computational Biology*, 8(6), e1002475. <https://doi.org/10.1371/journal.pcbi.1002475>
- Weinberger, A. D.; Wolf, Y. I.; Lobkovsky, A. E.; Gilmore, M. S. & Koonin, E. V. (2012b). Viral diversity threshold for adaptive immunity in prokaryotes. *mBio*, 3(4), e00456-12. <https://doi.org/10.1128/mBio.00456-12>
- Weissman, J. L.; Laljani, R. M. R.; Fagan, W. F. & Johnson, P. L. F. (2019). Visualization and prediction of CRISPR incidence in microbial trait-space to identify drivers of antiviral immune strategy. *ISME Journal*, 13(11), 2589-2602. <https://doi.org/10.1038/s41396-019-0411-2>
- Westra, E. R.; Dowling, A. J.; Broniewski, J. M. & van Houte, S. (2016). Evolution and ecology of CRISPR. *Annual Review of Ecology, Evolution, and Systematics*, 47, 307-331. <https://doi.org/10.1146/annurev-ecolsys-121415-032428>
- Westra, E. R. & Levin, B. R. (2020). It is unclear how important CRISPR-Cas systems are for protecting natural populations of bacteria against infections by mobile genetic elements. *Proceedings of the National Academy of Sciences*, 117(49), 27777-27785. <https://doi.org/10.1073/pnas.1915966117>
- Wheeler, D. L.; Barrett, T.; Benson, D. A.; Bryant, S. H.; Canese, K.; Chetvernin, V.; Church, D. M.; DiCuccio, M.; Edgar, R.; Federhen, S.; Geer, L. Y.; Kapustin, Y.; Khovayko, O.; Landsman, D.; Lipman, D. J.; Madden, T. L.; Maglott, D. R.; Ostell, J.; Miller, V. ... & Yaschenko, E. (2007). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 35(Database issue), D5-D12. <https://doi.org/10.1093/nar/gkl1031>

- Yan, W. X.; Hunnewell, P.; Alfonse, L. E.; Carte, J. M.; Keston-Smith, E.; Sothiselvam, S.; Garrity, A. J.; Chong, S.; Makarova, K. S.; Koonin, E. V.; Cheng, D. R. & Scott, D. A. (2019). Functionally diverse type V CRISPR-Cas systems. *Science*, 363(6425), 88-91. <https://doi.org/10.1126/science.aav7271>
- Yang, H. & Patel, D. J. (2019). CasX: A new and small CRISPR gene-editing protein. *Cell Research*, 29(6), 345-346. <https://doi.org/10.1038/s41422-019-0165-4>
- Zaayman, M. & Wheatley, R. M. (2022). Fitness costs of CRISPR-Cas systems in bacteria. *Microbiology*, 168(6), 001209. <https://doi.org/10.1099/mic.0.001209>
- Zablocki, O.; van Zyl, L. & Trindade, M. (2018). Biogeography and taxonomic overview of terrestrial hot spring thermophilic phages. *Extremophiles*, 22(5), 827-837. <https://doi.org/10.1007/s00792-018-1052-5>
- Zeldovich, K. B.; Berezovsky, I. N. & Shakhnovich, E. I. (2007). Protein and DNA sequence determinants of thermophilic adaptation. *PLoS Computational Biology*, 3(6), 0062-0072. <https://doi.org/10.1371/journal.pcbi.0030005>
- Zhao, L.; Qiu, M.; Li, X.; Yang, J. & Li, J. (2022). CRISPR-Cas13a system: A novel tool for molecular diagnostics. *Frontiers in Microbiology*, 13, 1069452. <https://doi.org/10.3389/fmicb.2022.1069452>
- Zhou, F.; Yu, X.; Gan, R.; Ren, K.; Chen, C.; Ren, C.; Cui, M.; Liu, Y.; Gao, Y.; Wang, S.; Yin, M.; Huang, T.; Huang, Z. & Zhang, F. (2023). CRISPRimmunity: An interactive web server for CRISPR-associated important molecular events and modulators used in genome editing tool identification. *Nucleic Acids Research*, 51(W1), W93-W107. <https://doi.org/10.1093/nar/gkad425>
- Zimmermann, L.; Stephens, A.; Nam, S.-Z.; Rau, D.; Kübler, J.; Lozajic, M.; Gabler, F.; Söding, J.; Lupas, A. N. & Alva, V. (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *Journal of Molecular Biology*, 430(12), 2237-2243. <https://doi.org/10.1016/j.jmb.2017.12.007>