Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Trade-off between CRISPR-based autoimmunity and CRISPR-Cas immunity against phages in bacteria

Bin Li

libin0571@zju.edu.cn

State Key Lab of Rice Biology and Breeding, Ministry of Agriculture and Rural Affairs Key Lab of Molecular Biology of Crop Pathogens and Insect Pests, Institute of Biotechnology, Zhejiang University
https://orcid.org/0000-0002-4581-4775

Wenting Lei
Zhejiang University

Mengju Liu
Zhejiang University

Aohan Xu
Zhejiang University

Munazza Ijaz
Zhejiang University

Shiqi Hu
Zhejiang University

Article

Keywords: self-targeting; cas; phage; autoimmunity; Promoter-bound regulators

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11693236/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Trade-off between CRISPR-based autoimmunity and CRISPR-Cas immunity against phages in bacteria

Wenting Lei[1], Mengju Liu[1], Aohan Xu[1], Munazza Ijaz[1], Shiqi Hu[1], Bin Li[1,*]

[1]State Key Laboratory of Rice Biology and Breeding, Ministry of Agriculture Key Laboratory of Molecular Biology of Crop Pathogens and Insects, Zhejiang Key Laboratory of Biology and Ecological Regulation of Crop Pathogens and Insects, Institute of Biotechnology, Zhejiang University, Hangzhou 310058, China;

*Correspondence: libin0571@zju.edu.cn (B.L.)

Tel.: +86-0571-88982412 (B.L.)

Abstract

CRISPR-Cas systems provide bacteria with sequence-specific anti-phage immunity via spacer-based memory. However, spacers matching the host chromosome can cause autoimmunity, and how bacteria balance these opposing outcomes in native host-phage systems remains unclear. Here, we identify two routes that reduce interference in self-targeting spacer-rich bacteria: loss of *cas*-locus integrity and host-mediated transcriptional repression. Across 2,339 type I-C-associated genomes, self-targeting spacer-rich bacteria frequently harboured incomplete *cas* loci, particularly in the interference module. In two closely related strains, phage-resistant strain transcribed *cas* genes and executed spacer-directed interference, whereas phage-susceptible strain carried more prophage-associated self-targeting spacers despite an intact locus. Although phage-susceptible strain encodes a catalytically active Cas3, promoter-bound host regulators repress *cas* transcription, and Cas3 is not detectably accumulated *in vivo*. Deletion of regulators increased *cas* expression and phage resistance while reducing bacterial growth, revealing a growth-defence trade-off associated with regulator loss. Together, these results identify genetic impairment and host-mediated repression as routes that limit interference in self-targeting spacer-rich bacteria, linking reduced autoimmune risk with a shift in phage defence.

Keywords: self-targeting; *cas*; phage; autoimmunity; Promoter-bound regulators

Introduction

Bacteria live under continuous predation by phages, and this pressure has driven the evolution of a broad range of antiviral defence systems [1-4]. Among them, CRISPR-Cas systems are widespread in bacteria and provide sequence-specific adaptive immunity. During phage infection, the CRISPR-Cas adaptation machinery selects short fragments of phage DNA, called protospacers, and integrates them into CRISPR arrays as spacers. If this bacterium is subsequently exposed to previously encountered phage, spacer-derived CRISPR RNAs guide Cas effector complexes to complementary protospacers, enabling degradation or cleavage of matching invasive nucleic acids [5-11]. This allows bacteria to convert previous encounters with phages into heritable immune memory, while spacer diversity can shape population-level defence against rapidly evolving phages [12].

In some cases, spacers stored in CRISPR arrays can become self-targeting spacers, directing CRISPR-Cas systems against protospacers of phages integrated in bacterial chromosomes [13,14]. Inappropriate activation against bacterial chromosomal targets can create a direct threat to genome integrity. For example, experimental genome targeting by CRISPR-Cas systems can be cytotoxic, which will induce genome remodeling or selectively eliminate bacterial cells carrying matching chromosomal targets [15-17]. Obviously, there is a high risk that adaptive memory becomes autoimmune when an immune target becomes part of bacterial genomes. Thus, self-targeting is involved in both CRISPR-Cas immunity and CRISPR-based autoimmunity, which must be balanced in bacteria.

Several routes have been reported to be able to resolve or avoid this self-targeting, such as mutations in the spacer or protospacer, changes in the protospacer adjacent motif (PAM), loss of *cas* genes or inhibition of CRISPR-Cas activity [13-15]. Phages can restore infectivity by encoding

anti-CRISPR proteins that directly inhibit CRISPR-Cas activity. In *Pseudomonas aeruginosa* phages, deletion of an anti-CRISPR gene rendered the phage unable to infect bacteria carrying a functional type I-F CRISPR-Cas system, whereas introducing the gene into a CRISPR-targeted phage restored phage infectivity [18]. Subsequent biochemical analyses showed that AcrF1 and AcrF2 inhibit target-DNA binding by the type I-F Csy surveillance complex, whereas AcrF3 binds Cas3 and prevents its recruitment to the DNA-bound complex, thereby blocking target DNA degradation [19]. These studies established phage-encoded inhibition of CRISPR surveillance or interference as a direct route to overcoming bacterial immunity [20-24]. However, most of these studies focus either on CRISPR-Cas immunity or CRISPR-based autoimmunity, little information was available about the correlation between them although they are both related to self-targeting spacers in bacteria.

To address how bacteria mediate this trade-off between CRISPR-based autoimmunity and CRISPR-Cas immunity against phages, we firstly collected 2,339 bacterial genomes and analysed the association of genome level self-targeting spacer richness with the integrity of *cas* locus in type I-C CRISPR-Cas systems, which have been found across diverse bacterial lineages. Furthermore, we experimentally verified this involvement of self-targeting spacers in the CRISPR-Cas activity using an interaction model established by phage resistant and susceptible strains of the rice pathogen *Acidovorax oryzae*, as well as their phage. In addition, we identified two promoter-bound host regulators, which represses *cas* transcription in phage-susceptible strain with rich self-targeting spacers using transcriptomic analysis, gene knockout, and phage infection, while the deletion increases *cas* transcription, resulting in the increased activity against phage infection and the impaired growth. Together, the data in this study revealed that promoter-bound host regulators

mediate the balance between CRISPR-based autoimmunity and the CRISPR-Cas activity in bacteria.

Results

Self-targeting spacer abundance is associated with interference-module defects

Self-targeting spacers can impose autoimmune risk when CRISPR-Cas systems remain active, suggesting that their long-term retention in bacterial genomes may depend on mechanisms that mitigate this risk [13-15]. To ask whether the distribution of self-targeting spacers in bacterial genomes is linked to the functional state of CRISPR-Cas loci, we firstly analysed a discovery cohort of 2,339 genomes from bacterial lineages in which type I-C CRISPR-Cas systems had been detected. The cohort combined CRISPR-Casdb-annotated type I-C genomes [25] with newly deposited NCBI genomes from the same lineages, while limiting the contribution of overrepresented genera (Supplementary Fig. 1). Our results showed that self-targeting spacers were found across the cohort, and their abundance was highly heterogeneous among the analysed bacterial genomes. Furthermore, 39.5% of bacterial genomes contained no detectable self-targeting spacers, whereas 60.5% carried at least one self-targeting spacer (Fig. 1a). In total, most genomes carried no or low number of self-targeting spacers but a subset retained multiple self-targeting spacers. In addition, a right-skewed distribution was found for self-targeting spacers, which rose steeply at a low number and then increased more gradually across the threshold with higher number (Supplementary Fig. 2a).

We also examined the association of prophage carriage with self-targeting spacers due to previous studies have showed that integrated prophages can provide chromosomal protospacers for CRISPR-Cas self-targeting [14]. Interestingly, the abundance of self-targeting spacers also differed significantly between prophage-containing and prophage-free genomes (Fig. 1b). Indeed,

114 self-targeting spacers showed a prevalence of 64.8% in prophage-containing genomes and a
115 prevalence of 55.6% in prophage-free genomes (Fig. 1c). Moreover, there was a positive
116 correlation between prophage number and the abundance of self-targeting spacers across the
117 cohort (Fig. 1d), consistent with an association between prophage carriage and self-targeting
118 spacer abundance.

119 Furthermore, we examined whether the abundance of self-targeting spacers is associated with
120 the integrity of *cas* locus. Compared with *cas*-intact genomes, genomes carrying incomplete *cas*
121 loci contained more self-targeting spacers. We observed a pronounced rightward shift in self-
122 targeting spacer abundance in *cas*-defect genomes, with the median number of self-targeting
123 spacers increasing from 1 in *cas*-intact genomes to 5 in *cas*-defect genomes (Fig. 1e). In parallel,
124 the proportion of genomes containing detectable self-targeting spacers increased from 57.5% in
125 *cas*-intact genomes to 72.3% in *cas*-defect genomes. This association was further resolved based
126 on module-level comparisons of type I-C CRISPR-Cas systems. To summarize the empirical
127 cumulative distribution function comparisons (Supplementary Fig. 2b), we used the Kolmogorov-
128 Smirnov D statistic, which measures the distributional divergence in self-targeting spacer numbers
129 between each defective class and *cas*-intact genomes (Fig. 1f). Interference-defect (*cas3*, *cas5*,
130 *cas8* and *cas7*) genomes showed the largest divergence from *cas*-intact genomes, with the highest
131 D statistic (D = 0.433) and the highest median self-targeting spacer numbers among the defective
132 subclasses (median = 7). Adaptation-defect (*cas4*, *cas1* and *cas2*) genomes showed a smaller
133 divergence from *cas*-intact genomes (D = 0.223; median = 3), whereas both-module-defect
134 genomes showed an intermediate divergence (D = 0.299; median = 4). The fraction of genomes
135 with detectable self-targeting spacers was also highest in the interference-defect group, reaching
136 77.9%, compared with 71.4% in adaptation-defect genomes and 67.2% in both-module-defect

genomes. Thus, genomes defective in interference modules showed the greatest elevation in self-targeting spacer abundance, indicating that the cost of self-targeting is largely determined by interference activity.

Closely related strains retain conserved type I-C loci but differ in phage susceptibility

Type I-C CRISPR-Cas systems have been found to be active and play an important role in bacterial resistance to phage infection although bacteria suppress the risk of CRISPR-based autoimmunity at the cost of *cas* locus integrity. To understand how bacteria balance the autoimmunity risk and active CRISPR-Cas, we carried out an analysis of spacer-protospacer matches using a native bacteria-phage system, including phage AP1 and two closely related *A. oryzae* strains that differ in phage susceptibility (RS-1, phage-resistant strain; RS-2, phage-susceptible strain) as shown in Fig. 2a. The genomes of two strains were approximately 5 Mb in size, with 87.4% of homologous regions (4470 homologous genes), and 97.9% of average nucleotide identity, revealing high sequences similarity between them. Both strains contained the conserved type I-C CRISPR-Cas loci including two CRISPR arrays (Fig. 2b) in similar genomic arrangements with 94-99% amino acid identity of the seven Cas proteins. Cas1-based phylogenetic analysis clustered the two strains in the type I-C lineage (Fig. 2c). The CRISPR arrays showed that both strains contained 99 and 101 spacers in CRISPR1, 44 and 28 spacers in CRISPR2, respectively (Fig. 2d).

We next examined spacer-protospacer relationships between the bacterial CRISPR arrays and the phage genome, because that CRISPR target recognition depends on both spacer-protospacer complementarity and the presence of a compatible protospacer adjacent motif [26,27]. In phage-resistant strain, six phage protospacers showed high similarity to CRISPR spacers (Supplementary Table 1). Several mismatches were located outside the seed region, and one phage protospacer

perfectly matched spacer 28 in CRISPR2 with an intact protospacer adjacent motif. These results revealed spacer-guided targeting potential against phage in phage-resistant strain. Phage-susceptible strain carried one phage-matching spacer, whose protospacer contained two mismatches, including one in the seed region (Supplementary Table 1).

The two strains differed more sharply in their self-targeting spacer profiles (Fig. 2e, Supplementary Table 2). The phage-resistant strain contained only one chromosomal protospacer matching an endogenous CRISPR spacer, and this site carried two PAM mismatches, making productive self-targeting unlikely. This protospacer was located in a gene encoding an integrating conjugative element protein and was not within predicted prophage regions (Supplementary Table 3). By contrast, phage-susceptible strain harbored three putative prophages and 24 chromosomal protospacers matching endogenous CRISPR spacers. Ten of these self-targeting spacers-associated protospacers were located within intact prophage region 2 (Supplementary Table 3), and 11 overlapped phage-related genes. Notably, three retained features compatible with CRISPR-Cas recognition: one perfectly matched CRISPR2 spacer 41 with an intact TTC PAM, and two contained mismatches outside the seed region while retaining intact PAMs. Therefore, phage-susceptible strain combined reduced phage-targeting potential with a higher abundance of prophage-associated self-targeting. The strain pair provided a matched system in which to compare the activity of conserved type I-C loci in distinct spacer contexts. We next tested whether these genomic differences were accompanied by differences in *cas* transcription and interference activity.

The phage-resistant strain expresses *cas* genes and executes spacer-directed interference

The contrasting phage-targeting and self-targeting spacer profiles suggest that both closely related strains may differ in activity of the conserved type I-C loci. We first examined the transcriptional activity of the conserved type I-C loci in both strains. The expression of all seven

183 *cas* genes in phage-resistant strain was revealed by analysis of FPKM-normalized RNA-seq (Fig.
184 3a), and confirmed by RT-qPCR (Fig. 3b), while read-coverage tracks further showed transcription
185 across the *cas* locus and both CRISPR arrays (Fig. 3c). In contrast, there was a great reduction in
186 FPKM-normalized expression of all seven *cas* genes in phage-susceptible strain compared to
187 phage-resistant strain (Fig. 3a). Fold changes calculated as FPKM ratios in phage resistant and
188 susceptible strains were 13.14 for *cas3*, 22.06 for *cas5*, 12.04 for *cas8*, 35.21 for *cas7*, 45.28 for
189 *cas4*, 7.57 for *cas1* and 22.64 for *cas2*. The markedly reduction in *cas* locus transcription
190 abundance were confirmed by RT-qPCR (Fig. 3b) and RNA-seq coverage (Fig. 3c). These results
191 showed that transcription across the conserved type I-C loci was high in the phage resistant strain
192 but strongly reduced in the phage-susceptible strain.

193 We next tested whether this expressed locus contributes to phage resistance by constructing
194 individual *cas* gene deletion mutants in phage-resistant strain and then assaying their susceptibility
195 to phage infection (Fig. 3d). Results indicated that deletion of each *cas* gene except *cas4* increased
196 phage susceptibility relative to the wild type, and complementation restored phage resistance in
197 the susceptible deletion mutants. Therefore, the deletion and complementation assays show that
198 the endogenous type I-C *cas* locus is required for efficient phage defence in phage-resistant strain.

199 To determine whether spacer-directed DNA interference was executed in phage-resistant
200 strain, we performed plasmid interference assays by introducing plasmids carrying phage-derived
201 protospacers matched to bacterial spacers, together with cognate PAM sequences into phage-
202 resistant strain. Results showed that transformation efficiency was strongly reduced by plasmid
203 interference, with colony numbers reaching only 11.43% of the control (Fig. 3e). This reduction
204 transformation efficiency is consistent with depletion of plasmids carrying a spacer-matched
205 protospacer and compatible PAM [26,27]. We then tested whether self-derived protospacer sequences

could also trigger interference when sequence mismatches were removed. A plasmid carrying a corrected self-targeting protospacer (CRISPR2-Spacer8), engineered to restore perfect spacer-protospacer complementarity (Fig. 3f), was also depleted in phage-resistant strain. Thus, the phage-resistant strain type I-C system can recognize a self-derived protospacer when the target sequence and PAM meet interference requirements. Therefore, a chromosomal protospacer can become an interference-competent self-target when complementarity and PAM requirements are met.

Collectively, the conserved type I-C *cas* locus found in both closely related bacterial strains occupied distinct functional states. Phage-resistant strain transcribes its *cas* genes and CRISPR arrays and executes spacer-directed interference against compatible DNA targets. In contrast, phage-susceptible strain, despite carrying a similar *cas* locus, shows markedly reduced transcription across the *cas* locus.

Catalytically active Cas3 does not accumulate detectably in the phage-susceptible strain

To determine whether reduced *cas* transcription in phage-susceptible strain reflected a defect in the effector itself, we first examined the *cas3* coding sequence and predicted Cas3 protein due to that Cas3 is the nuclease-helicase protein that mediates target DNA degradation during type I interference [28-31]. Cas3 proteins of phage resistant and susceptible strains were both 834 amino acids long and shared 97.84% amino acid identity (Supplementary Fig. 3a and Supplementary Table 4), with similar hydropathy profiles (Supplementary Fig. 3b, c). Conserved domain analysis showed that both proteins retained the canonical HD nuclease and DEXHc helicase domains characteristic of Cas3 proteins (Fig. 4a). Secondary-structure prediction showed comparable distributions of α-helices, random coils, extended strands and β-turns (Supplementary Fig. 3d). Three-dimensional modelling revealed near-complete overlap of the HD and DEXHc domains

between the two proteins (Fig. 4b). Therefore, it can be inferred that Cas3 proteins are structurally conserved in both phage resistant and susceptible strains.

We next tested whether Cas3 retained catalytic activity in phage resistant and susceptible strains by heterologous expression, purification (Supplementary Fig. 4a), and incubation with ssDNA or dsDNA substrates. Results indicated that both proteins exhibited strong *in vitro* nuclease activity in a Mg^{2+}-dependent and concentration-dependent manner, while the DNA substrates such as ssDNA and dsDNA were nearly completely degraded by both proteins at 1000 nM of concentration (Fig. 4c). We then examined whether Cas3 accumulated in both bacterial strains using a polyclonal antibody raised against Cas3 of phage-susceptible strain (Supplementary Fig. 4b). Western blotting detection revealed a clear Cas3 band in total protein extracts from phage resistant strain, whereas no corresponding signal was detected in total protein extracts from phage-susceptible strain (Fig. 4d). Because this antibody recognized Cas3 from phage resistant strain, the absence of a signal in phage-susceptible strain was unlikely to result from poor antibody cross-reactivity. Furthermore, a difference was found in Cas3 abundance between both strains based on LC-MS/MS analysis (Fig. 4e). Indeed, 11 out of 16 Cas3-derived peptides were identified in phage resistant strain (Supplementary Table 5), while no Cas3-derived peptides were detected in phage-susceptible strain. These results suggest that phage-susceptible strain encodes a structurally conserved and enzymatically active Cas3 that does not accumulate detectably *in vivo*.

Promoter-bound regulators mediate the trade-off between the CRISPR-Cas activity and bacterial growth in phage-susceptible strain

The absence of detectable Cas3 protein despite a catalytically active Cas3 sequence prompted us to examine promoter-level regulation. We therefore compared the 500 bp upstream regions of *cas3* in phage resistant and susceptible strains. Results indicated that both strains retained predicted

σ factor promoter elements, but the promoter of phage-susceptible strain contained a strain-specific predicted regulatory motif, 5′-CTTGTAAG-3′, that was absent from the promoter of phage-resistant strain (Fig. 5a). Using a biotin-labelled probe corresponding to the *cas3* upstream region, a DNA pull-down assay was performed to screen proteins that bind the promoter of phage-susceptible strain. LC-MS/MS identified 45 specific proteins from bands enriched in the labelled-probe fraction by silver staining relative to the control (Supplementary Fig. 5a and Supplementary Table 6). From these candidates, 11 proteins were selected for yeast one-hybrid screening based on enrichment and annotation (Supplementary Table 7). The screening was performed using a reporter containing three tandem copies of the predicted regulatory motif as bait, because the native promoter fragment showed strong autoactivation in yeast (Supplementary Fig. 5b). This screen identified four candidate promoter-binding proteins: p21, p1981, p2762 and p4207 (Supplementary Fig. 5c).

We next tested direct binding of these purified candidate proteins to the *cas3* promoter of phage-susceptible strain by electrophoresis mobility shift assay (EMSA), a standard assay for protein-DNA interactions [32]. Results indicated that p1981 and p2762 produced clear mobility shifts with the *cas3* promoter probe (wild-type, WT) (Fig. 5b), whereas p21 and p4207 did not (Fig. 5c and Supplementary Fig. 5d, e). The mutant 1 (MT1) probe behaved like WT and did not reduce p1981 or p2762 binding. Mutation of the predicted regulatory motif in the redesigned mutant 2 (MT2) probe partially reduced the binding of p1981 and p2762, indicating that this motif contributed to the interaction but was not the only binding determinant. We then used the mutant 3 (MT3) probe, in which nucleotides immediately upstream of this motif and adjacent to the predicted σ factor binding region were altered. p1981 failed to bind the MT3 probe, identifying 5′-GTTGCTTGTAAG-3′ as a critical sequence for p1981 recognition (Fig. 5d). In contrast, p2762

retained binding across multiple mutated probes, suggesting a broader or less sequence-restricted DNA-binding pattern. Competition EMSA further supported the specificity of p1981 binding, as increasing concentrations of unlabeled competitor reduced the shifted band and a 100-fold excess competitor abolished the signal (Fig. 5e).

The p1981-binding sequence overlapped the predicted σ factor recognition region, providing a plausible basis for transcriptional repression through promoter occupancy. Annotation identified p1981 as an H-NS-like protein, whereas p2762 was annotated as a hypothetical protein and is therefore considered here as an uncharacterized promoter-binding regulator. We then tested whether p1981 and p2762 regulate *cas* transcription *in vivo*. Deletion of either p1981 or p2762 in phage-susceptible strain partially restored *cas* gene transcription, with the p2762 mutant showing a stronger increase than the p1981 mutant (Fig. 5f). Both proteins therefore contributed to *cas* transcriptional repression in phage-susceptible strain.

We next examined whether derepression of the type I-C locus altered phage susceptibility. Deletion of p1981 or p2762 reduced bacterial susceptibility to phage infection, whereas complementation restored phage susceptibility (Fig. 5g). These results revealed that the two regulators that repressed *cas* transcription contributed to phage susceptibility in phage-susceptible strain. We also asked whether the two regulators act as a complex to affect *cas* transcription and phage susceptibility. A bacterial two-hybrid assay revealed that there was no interaction between p1981 and p2762 (Supplementary Fig. 5f), suggesting that they act through independent regulatory mechanisms rather than as a single protein complex.

Finally, we assessed also whether relief of *cas* repression was accompanied by a growth cost in bacteria. Deletion of p1981 or p2762 impaired growth performance relative to wild type of the phage-susceptible strain (Fig. 5h). At 18 h, the OD600 values of Δp1981 and Δp2762 decreased

by 41.0% and 53.1%, respectively, compared with the wild type. Consistently, area-under-the-curve (AUC) analysis showed that overall growth was reduced by 38.2% in Δp1981 and 50.6% in Δp2762 (Fig. 5i). Complementation partially restored growth, reaching 72.8% and 62.9% of the wild-type AUC, respectively. Derepression of the type I-C locus was therefore accompanied by reduced phage susceptibility and impaired growth in phage-susceptible strain. These results suggest that promoter-bound regulators mediate the trade-off between the CRISPR-Cas activity and bacterial growth in phage-susceptible strain.

Discussion

Balancing CRISPR-Cas-mediated phage defence with protection from self-targeting is essential for maintaining adaptive immunity, particularly when phage-derived sequences persist in bacterial chromosomes. Here, genome-scale analyses revealed that self-targeting spacer-rich genomes were associated with reduced integrity of *cas* interference modules. In a pair of closely related strains with conserved CRISPR-Cas loci but contrasting phage susceptibility, host-mediated regulation repressed *cas* transcription in the susceptible strain, which carried more self-targeting spacers. These findings link genetic impairment and host-mediated repression to reduced CRISPR-Cas interference. Together, they show how limiting interference can enable self-targeting spacers to persist while compromising phage defence.

Our results not only revealed that bacteria suppress the risk of CRISPR-based autoimmunity at the cost of CRISPR-Cas locus integrity, but also showed that different strategies have been developed in bacteria to suppress the risk of autoimmunity in both closely related strains with conserved CRISPR-Cas loci but different susceptibility to phage. Bacteria can accommodate self-targeting spacers through genetic impairment or regulatory repression of CRISPR-Cas interference, potentially at the expense of antiviral defence.

These findings extend previous work on CRISPR-Cas autoimmunity. Self-targeting spacers have been recognized across diverse CRISPR-Cas systems, and prophages have been proposed as major sources of chromosomal protospacers that drive self-targeting risk [13,14]. Our genome-scale analysis suggests that, in genomes from type I-C-associated lineages, genome-level self-targeting spacers is not simply a passive record of past spacer acquisition. Instead, elevated self-targeting spacer abundance is associated with prophage carriage and with reduced integrity of *cas* loci, especially the interference module. Rollie et al. shown that targeting temperate phages can drive the loss of type I CRISPR-Cas systems [33]. In type I systems, target recognition by the Cascade surveillance complex recruits Cas3, which mediates target DNA degradation[11,28,29]. This pattern suggests that weakening or repressing the interference machinery lowers the autoimmune cost of carrying self-targeting spacers, allowing them to persist. The strongest association with interference-module defects is consistent with the idea that active DNA targeting may constrain the persistence of self-targeting spacer-containing genomes. Together, reduced interference capacity permits the persistence of self-targeting spacers.

Genetic impairment is not the only possible route by which self-targeting spacers can coexist with CRISPR-Cas systems. When the *cas* locus remains intact, phage-derived sequences embedded in the host chromosome may remain potential targets of CRISPR interference. This blurring of the boundary between foreign and self may create a liability when target recognition leads to Cas3-mediated DNA degradation. Previous work has shown that CRISPR-Cas loci can be transcriptionally regulated, including by H-NS-mediated silencing in *Escherichia coli* [34,35]. In *Pseudomonas aeruginosa*, host alginate regulators repress type I-F CRISPR-Cas expression and immunity [36]. In a type II-A system, tracr-L-guided Cas9 autorepression limits immune activity, with derepression increasing both antiviral immunization and autoimmune toxicity [37]. These

studies show that CRISPR-Cas activity can be regulated at the transcriptional or effector level, but they did not address how such regulation relates to self-targeting spacer abundance in a native host-phage system. Our findings extend this framework by linking host-mediated repression of an intact type I-C locus to self-targeting spacer persistence and phage susceptibility.

Type I CRISPR-Cas systems offer a tractable setting in which to examine this trade-off. Their molecular machinery and mechanisms of interference have been extensively characterized [38,39]. We addressed this question using a matched *A. oryzae* host-phage system with contrasting self-targeting spacer abundance and phage susceptibility. These two strains carry comparable type I-C loci, yet they present different immune states to the same phage. Notably, a higher genome-wide abundance of self-targeting spacers coexists with prophage carriage and reduced *cas3* expression in the phage-susceptible strain. This finding is consistent with the association between interference-module defects and elevated self-targeting spacer abundance observed in the genome-scale analysis. Promoter-bound host regulators bind the *cas3* promoter and contribute to repression of *cas* transcription. Deletion of these regulators partially restores *cas* transcription and reduces phage susceptibility. Together, the data link host-mediated *cas* repression in a self-targeting spacer-rich genome with reduced type I-C activity and increased susceptibility to phage. The comparison also illustrates a limitation of genotype-based defence prediction. The presence of a CRISPR-Cas locus did not by itself predict functional defence. Genome-based prediction of bacterial immunity should therefore consider the regulated activity of a defence system alongside its genotype.

The strain pair allowed us to distinguish locus conservation from functional immune activity and to connect the genome-scale signal from CRISPR interference-module defects to a native host-phage phenotype. Cas3 abundance provides the main link between promoter regulation and type

I-C activity in this system. Type I systems are particularly sensitive to self-targeting because target recognition can recruit Cas3, a nuclease-helicase that degrades DNA processively [11,28,29]. Cas3 from the phage-susceptible strain retains conserved domains, structural similarity to its counterpart in the phage-resistant strain and nuclease activity in vitro. The block occurs before Cas3 effector accumulation. The phage-susceptible strain therefore represents a regulatory route to an interference-limited state, rather than a Cas3-defective state at the protein level. In contrast to *cas* gene defects, transcriptional repression can reduce interference while retaining the genetic capacity for immunity. This route converges with the interference-module defects observed in the genome-scale analysis. Both mechanisms limit the conversion of spacer-protospacer matching into Cas3-mediated DNA degradation. By reducing this interference step, bacteria may tolerate high self-targeting spacer abundance while sacrificing CRISPR-Cas immunity against phages. This explains how an intact type I-C locus can coexist with a high self-targeting spacer abundance in phage-susceptible strain.

The promoter-bound regulators identified in phage-susceptible strain clarify how this repressed state is maintained. Annotation identified p1981 as an H-NS-like protein. This connects *cas* repression to H-NS-like silencing, a bacterial strategy known to control horizontally acquired or costly DNA, including CRISPR-Cas promoters [34,35,40-42]. The promoter-binding and deletion results support a role for p1981 in repressing *cas* transcription. By contrast, p2762 was annotated as a hypothetical protein. It nevertheless bound the *cas3* promoter and contributed to repression *in vivo*. Deletion of p2762 resulted in a greater increase in *cas* transcription than deletion of p1981. These data identify p2762 as a component of the type I-C regulatory state, although its precise binding site within the *cas3* promoter remains unresolved. CRISPR-Cas regulation may therefore involve both conserved global silencers and specific factors that are not evident from genome

annotation. The consequences of this regulation extended beyond *cas* transcript abundance. Removing p1981 or p2762 increased *cas* transcription and reduced phage susceptibility, indicating that promoter-mediated repression contributes to the phage-sensitive state. The same perturbations reduced growth performance, consistent with a physiological cost accompanying *cas* derepression in a self-targeting spacer-rich genome. This interpretation is supported by studies showing that genome-targeting CRISPR-Cas activity can be deleterious or selectively eliminate cells carrying matching chromosomal targets [15-17]. A physiological cost of CRISPR-Cas derepression has also been observed in a native bacterial host [43]. Regulator deletion couples *cas* derepression and increased phage resistance with reduced growth, demonstrating a physiological cost of this regulatory state; defining the contribution of self-targeting interference to that cost will require targeted follow-up.

A related example comes from Wimmer et al., who examined two extensively self-targeting type I CRISPR-Cas systems in *Xanthomonas albilineans*. They showed that these systems retained Cascade-mediated target recognition and Cas3-dependent DNA degradation, and identified distinct anti-CRISPR proteins that inhibited DNA degradation [44]. Similar to our study, their work showed that extensive self-targeting can coexist with a retained CRISPR-Cas locus. The mechanisms act at different levels: the anti-CRISPR proteins identified by Wimmer et al. inhibit DNA targeting or Cas3-mediated degradation, whereas p1981 and p2762 in our system repress *cas* transcription upstream of Cas3 accumulation. Both mechanisms converge on reduced interference capacity. Unlike direct anti-CRISPR inhibition, transcriptional repression reduces interference before effector proteins accumulate, while preserving the underlying *cas* locus and the capacity for defence to be restored. Their implications extend beyond the persistence of self-targeting spacers to the outcome of phage infection. A large body of work has shown that phages can

overcome CRISPR-Cas immunity through target escape [27,45] or anti-CRISPR-mediated inhibition of surveillance or interference [18-24]. In our system, many chromosomal protospacers in the phage-susceptible strain were located within prophage regions or overlapped phage-related genes. Several of these protospacers retained features compatible with CRISPR-Cas recognition. These observations highlight how phage-derived sequences can become potential chromosomal targets, creating a conflict between anti-phage defence and avoidance of self-targeting. Previous studies have shown that prophage integration and excision can remodel CRISPR loci and spacer content [46], whereas phage dormancy can facilitate spacer acquisition and subsequent anti-phage protection [47]. Modelling and comparative genomic analyses further suggest that the balance between anti-phage defence and autoimmunity constrain CRISPR arrays size [48]. These studies provide a broader context for interpreting the coexistence of prophage-associated self-targeting and reduced interference in our system. Self-targeting spacer accumulation, prophage carriage and reduced interference capacity may interact over evolutionary time. Phage-derived sequences integrated into the host chromosome can serve as targets for CRISPR spacers, while reduced interference may allow these spacers and their chromosomal targets to coexist. Host-mediated repression limits self-targeting interference while compromising defence against infecting phages. Phage susceptibility may therefore reflect host regulation as well as phage counter-defence.

In current study, we identify host-mediated transcriptional repression of an intact type I-C CRISPR-Cas locus in a self-targeting spacer-rich bacterium. Together with the genome-scale association between self-targeting spacer abundance and interference-module defects, our results support genetic impairment and transcriptional repression as two routes for accommodating self-targeting spacers by limiting interference. This trade-off between potential autoimmune costs and

anti-phage protection provides a framework for understanding how self-targeting spacers persist in genomes with either impaired or intact *cas* loci.

Figures and Figures captions

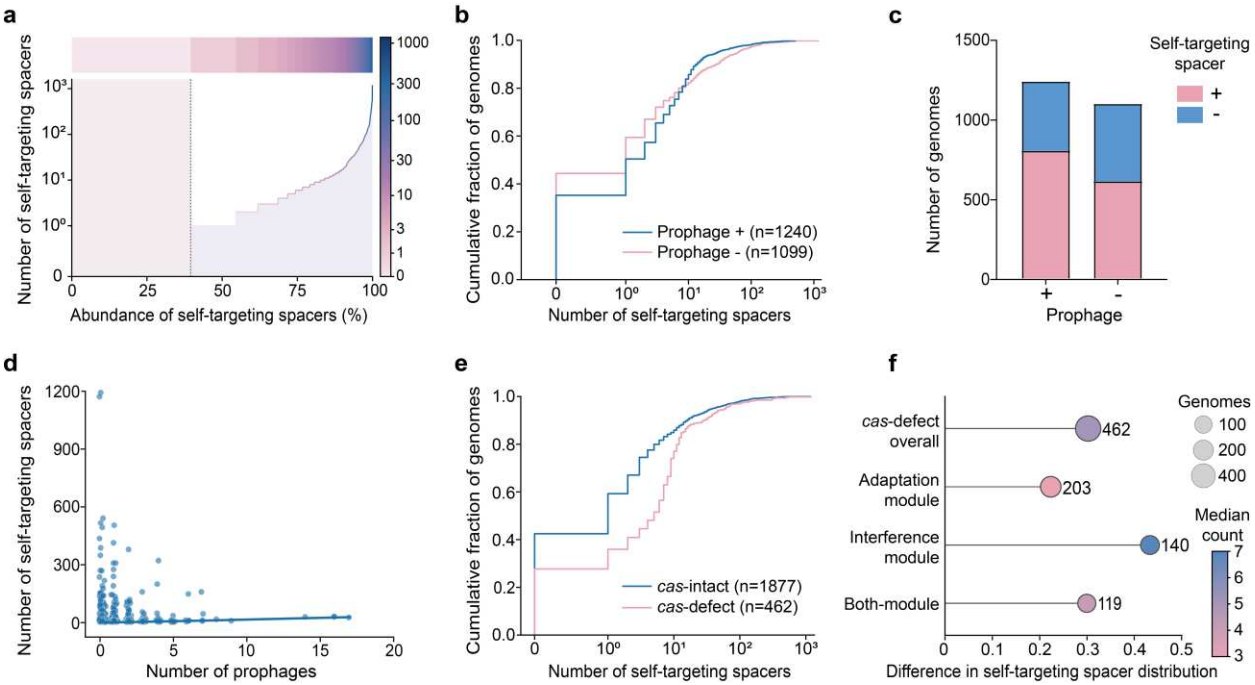


Fig. 1 Bacteria suppress the risk of CRISPR-based autoimmunity at the cost of CRISPR-Cas locus integrity.

a, Ranked distribution of genome-wide self-targeting spacers across 2,339 genomes from type I-C-associated lineages. **b**, Empirical cumulative distribution functions (ECDFs) of self-targeting spacer abundance in prophage-containing and prophage-free genomes. Distributions were compared using a two-sided Mann-Whitney U test ($P = 2.62 \times 10^{-4}$). **c**, Proportion of prophage-containing and prophage-free genomes with or without detectable self-targeting spacers. Association was assessed using Fisher's exact test ($P < 0.001$). **d**, Relationship between predicted prophage and self-targeting spacer number. Correlation was assessed using Spearman's rank correlation ($\rho = 0.137$, $P = 3.21 \times 10^{-11}$). **e**, ECDFs of self-targeting spacer numbers in *cas*-intact

and *cas*-defect genomes. Distributions were compared using a two-sided Kolmogorov-Smirnov test (D = 0.299, $P = 8.76 \times 10^{-30}$). **f**, Module-level effect-size summary of self-targeting spacer number distributions relative to *cas*-intact genomes. Horizontal lines indicate the Kolmogorov-Smirnov D statistic from two-sided comparisons with *cas*-intact genomes. Bubble size represents the number of genomes in each class, with genome numbers shown next to each bubble. Bubble colour indicates the median self-targeting spacer number. Adaptation-module defects were defined by disruption or absence of *cas1*, *cas2* or *cas4*, and interference-module defects were defined by disruption or absence of *cas3*, *cas5*, *cas8* or *cas7*, consistent with the modular organization of type I CRISPR-Cas systems [11,38]. The top row compares all *cas*-defect genomes with *cas*-intact genomes; lower rows show module-level subclasses. Two-sided Kolmogorov-Smirnov statistics were: *cas*-defect, D = 0.299, $P = 8.76 \times 10^{-30}$; adaptation- module, D = 0.223, $P = 1.67 \times 10^{-8}$; interference- module, D = 0.433, $P = 1.21 \times 10^{-22}$; both-module, D = 0.299, $P = 2.19 \times 10^{-9}$.

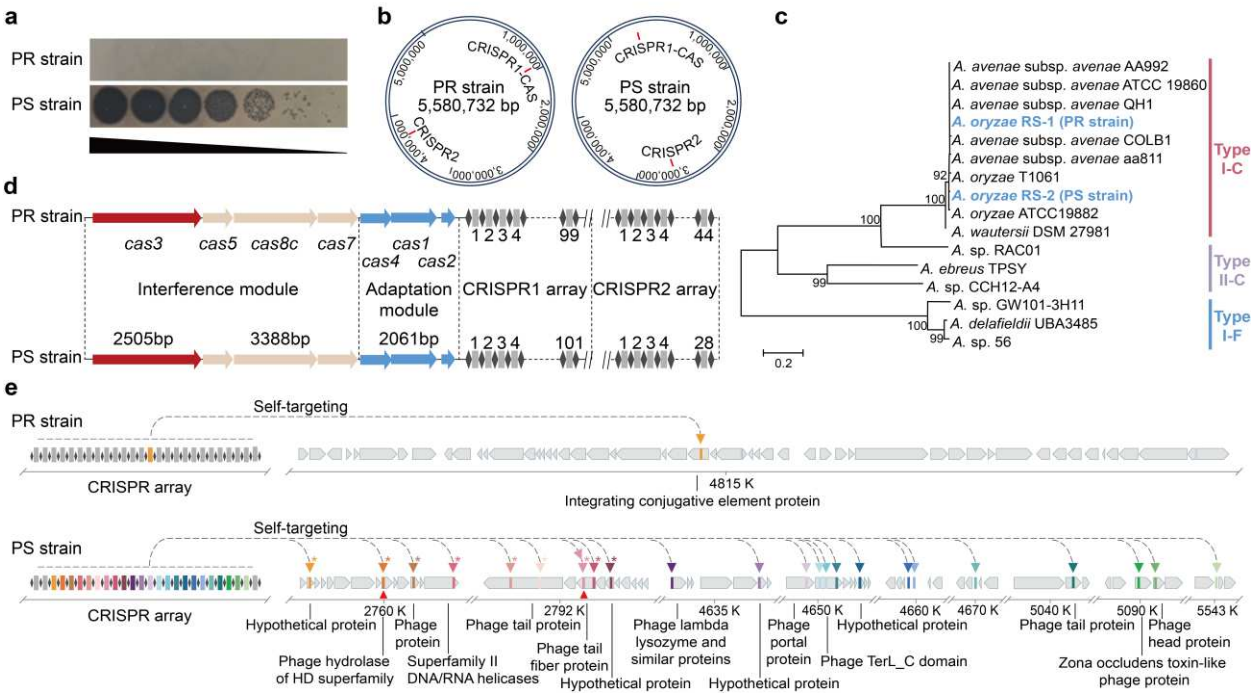


Fig. 2 CRISPR-based autoimmunity risk in both closely related strains with conserved CRISPR-Cas loci but different susceptibility to phage.

a, Susceptibility of *A. oryzae* strains RS-1 (phage-resistant strain, PR strain) and RS-2 (phage-susceptible strain, PS strain) to phage AP1, assessed by plaque formation. **b**, Chromosomal positions of CRISPR1 and CRISPR2 arrays in PR and PS strains. **c**, Phylogenetic placement of PR and PS strains based on Cas1 protein sequences from representative *Acidovorax* strains carrying complete CRISPR-Cas loci. Fourteen representative *Acidovorax* genomes were selected from NCBI. **d**, Organization of the type I-C CRISPR-Cas loci. Arrows indicate *cas* genes, black diamonds indicate repeat sequences and grey rectangles indicate spacer sequences. **e**, Genomic distribution of chromosomal protospacers matching endogenous CRISPR spacers. Coloured arrows indicate protospacers targeted by spacers from the type I-C CRISPR arrays, and gene names are shown where annotated. Double arrows denote protospacers targeted by two distinct spacers. Asterisks indicate protospacers located within predicted prophage regions. Red triangles mark

476 protospacers retaining recognition-compatible features, defined here by an intact 5′-TTC-3′PAM
477 and absence of mismatches in the seed region.

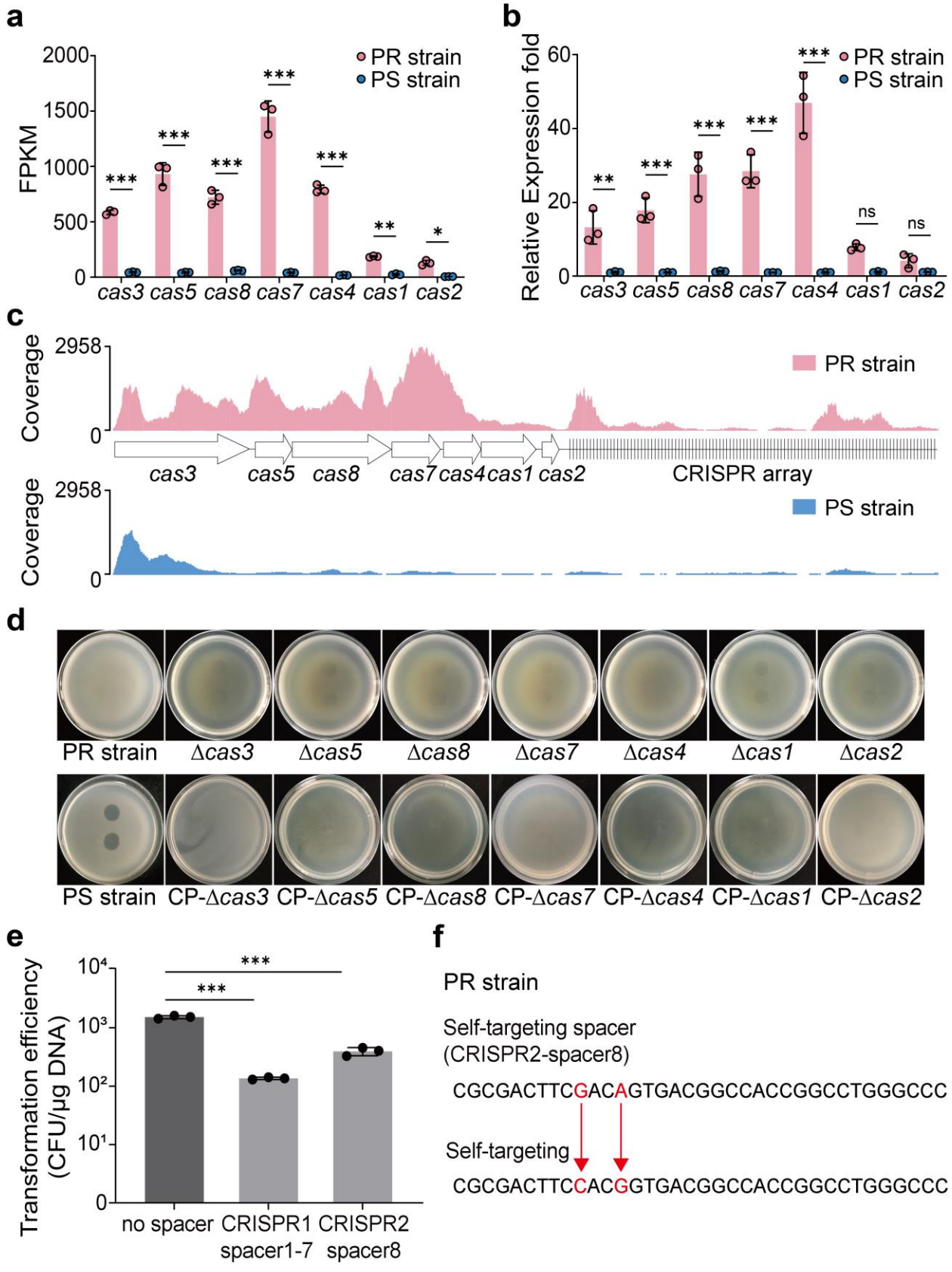
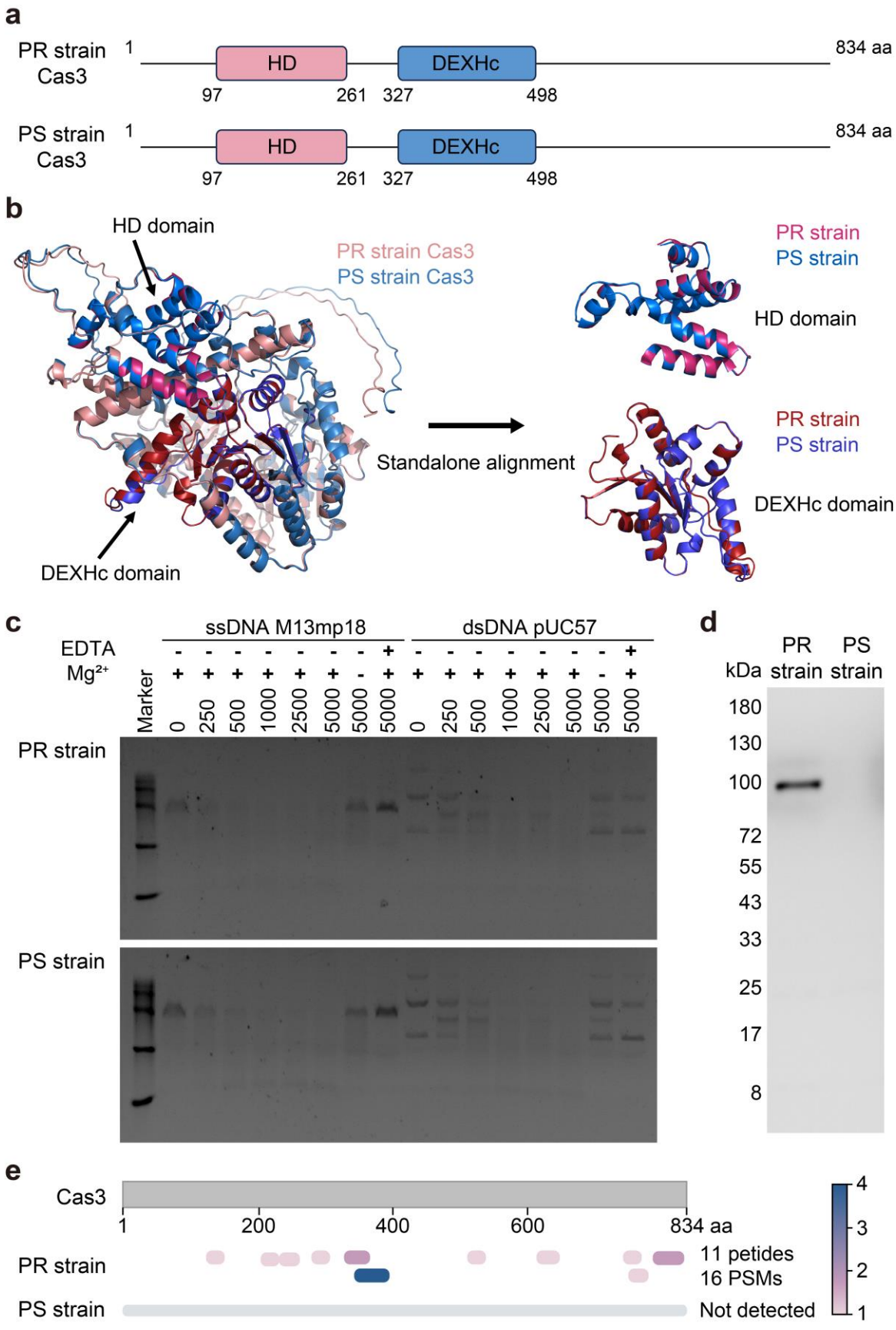

a
FPKM
2000
1500
1000
500
0
PR strain
PS strain
***
***
***
***
***
**
*
cas3 cas5 cas8 cas7 cas4 cas1 cas2
b
Relative Expression fold
60
40
20
0
PR strain
PS strain
**
***
***
***
***
ns
ns
cas3 cas5 cas8 cas7 cas4 cas1 cas2
c
Coverage
2958
0
PR strain
cas3
cas5
cas8
cas7
cas4
cas1
cas2
CRISPR array
Coverage
2958
0
PS strain
d
PR strain
Δcas3
Δcas5
Δcas8
Δcas7
Δcas4
Δcas1
Δcas2
PS strain
CP-Δcas3
CP-Δcas5
CP-Δcas8
CP-Δcas7
CP-Δcas4
CP-Δcas1
CP-Δcas2
e
Transformation efficiency (CFU/μg DNA)
10⁴
10³
10²
10¹
0
***
***
no spacer
CRISPR1 spacer1-7
CRISPR2 spacer8
f
PR strain
Self-targeting spacer (CRISPR2-spacer8)
CGCGACTTCGACAGTGACGGCCACCGGCCTGGGCCC
Self-targeting
CGCGACTTCCACGGTGACGGCCACCGGCCTGGGCCC

Fig. 3 Different strategy to suppress the risk of autoimmunity in both closely related strains with conserved CRISPR-Cas loci but different susceptibility to phage.

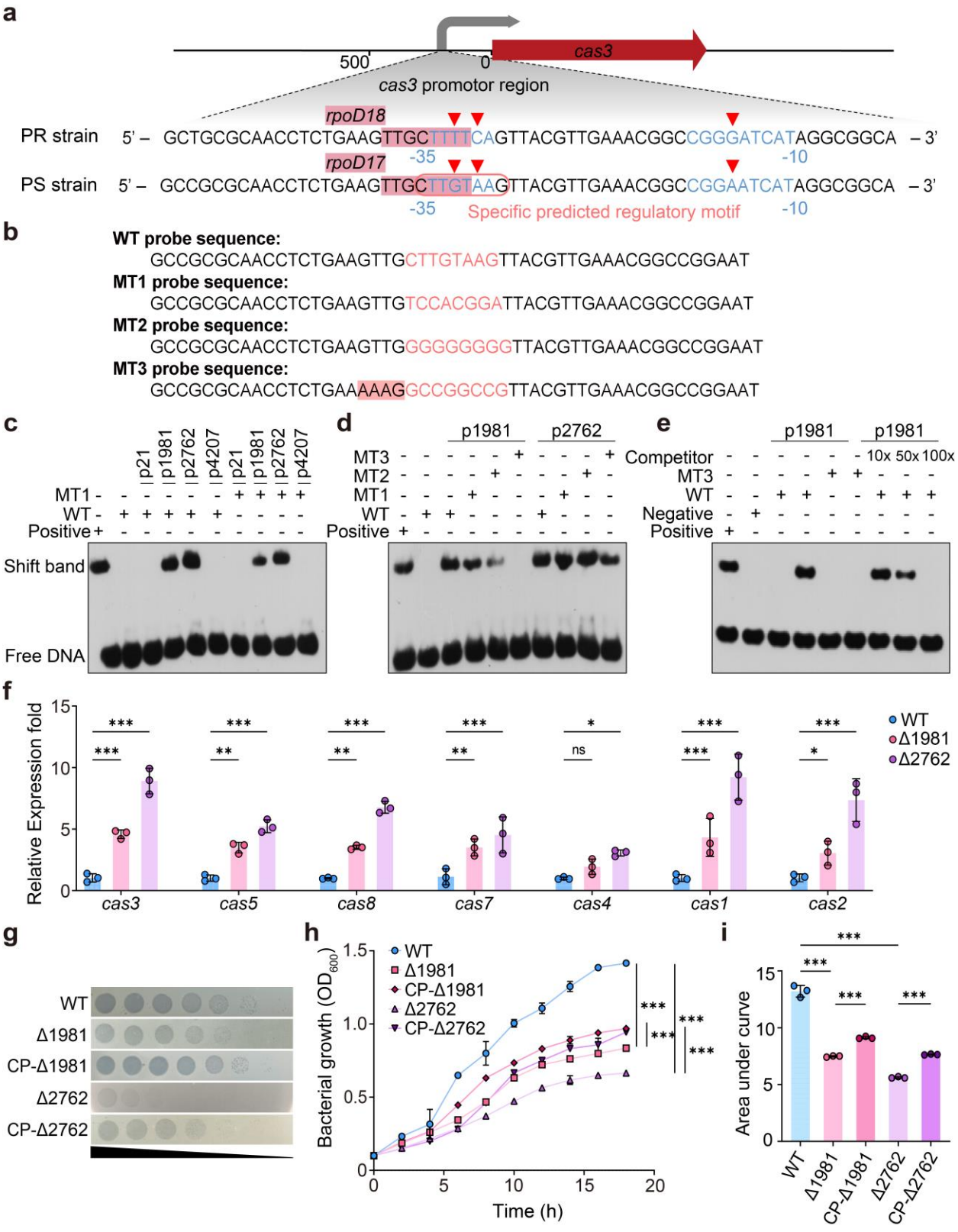
a, FPKM-normalized RNA-seq expression of type I-C *cas* genes in PR and PS strains. **b**, RT-qPCR analysis of *cas* transcript abundance. Expression was normalized to recA and calculated using the $2^{-\Delta\Delta Ct}$ method. **c**, RNA-seq read coverage across the type I-C *cas* loci and CRISPR arrays in PR and PS strains. Coverage indicates sequencing depth. **d**, Phage susceptibility of PR strain *cas* gene deletion mutants and complemented strains. Δ indicates the corresponding deletion mutant, and CP-Δ indicates the complemented strain. **e**, Plasmid interference assays in PR strain using pHERD30T constructs carrying spacer-matched protospacers with cognate PAM sequences. The CRISPR1 spacer 1-7 construct was used to assess spacer-directed interference against phage-derived protospacers, and the CRISPR2 spacer 8 construct was used to test a corrected self-derived protospacer. Transformation efficiency was normalized to the empty-vector control. **f**, Schematic of the corrected CRISPR2 spacer 8 protospacer used in the plasmid interference assay. Mismatches were corrected to restore perfect spacer-protospacer complementarity while preserving the type I-C-recognized 5′-TTC-3′ PAM. Statistical comparisons in **a**, **b** and **e** were performed using two-way ANOVA with multiple comparisons. Ns: not significant; *: $P \leq 0.05$; **: $P \leq 0.01$; ***: $P \leq 0.001$. Bars show the mean of three independent biological replicates, each with three technical replicates.


a
PR strain Cas3
1
HD
DEXHc
834 aa
97
261
327
498
PS strain Cas3
b
HD domain
PR strain Cas3
PS strain Cas3
DEXHc domain
Standalone alignment
PR strain
PS strain
HD domain
DEXHc domain
c
ssDNA M13mp18
dsDNA pUC57
EDTA
Mg2+
Marker
0
250
500
1000
2500
5000
PR strain
PS strain
d
PR strain
PS strain
kDa
180
130
100
72
55
43
33
25
17
8
e
Cas3
1
200
400
600
834 aa
PR strain
11 petides
16 PSMs
PS strain
Not detected
4
3
2
1

497

Fig. 4 Phage-susceptible strain reduced the risk of CRISPR-based autoimmunity at cost of the CRISPR-Cas activity

a, Domain organization of Cas3 from the PR and PS strains, showing the positions of the HD nuclease and DEXHc helicase domains. **b**, Structural comparison of Cas3. Left, structural alignment of full-length Cas3 from both strains predicted using AlphaFold 3 [73]. The predicted TM-score (pTM) was 0.87 for both models. Right, structural alignment of the extracted HD and DEXHc domains. **c**, Nuclease activity assay of recombinant Cas3 *in vitro*. Purified Cas3 proteins were incubated with 20 nM M13mp18 ssDNA or pUC57 dsDNA at the indicated protein concentrations for 2 h at 37 °C in the presence or absence of 20 mM MgCl2 or 67.5 mM EDTA. **d**, Western blot analysis of Cas3 in total protein extracts from PR and PS strains using a polyclonal antibody raised against PS-strain Cas3. **e**, LC-MS/MS peptide coverage of Cas3 in total protein extracts from PR and PS strains. The grey bar indicates the full-length Cas3 protein sequence. Coloured horizontal segments indicate Cas3-derived peptides detected in PR strain and mapped to their corresponding positions along Cas3; colour intensity denotes peptide-spectrum match (PSM) numbers. Peptide and PSM totals are shown on the right.


a
cas3
500
0
cas3 promotor region
rpoD18
PR strain 5' – GCTGCGCAACCTCTGAAGTTGCTTTTCAGTTACGTTGAAACGGCCGGGATCATAGGCGGCA – 3'
-35
-10
rpoD17
PS strain 5' – GCCGCGCAACCTCTGAAGTTGCTTGTAAGTTACGTTGAAACGGCCGGAATCATAGGCGGCA – 3'
-35
Specific predicted regulatory motif
-10
b
WT probe sequence:
GCCGCGCAACCTCTGAAGTTGCTTGTAAGTTACGTTGAAACGGCCGGAAT
MT1 probe sequence:
GCCGCGCAACCTCTGAAGTTGTCCACGGATTACGTTGAAACGGCCGGAAT
MT2 probe sequence:
GCCGCGCAACCTCTGAAGTTGGGGGGGGGTTACGTTGAAACGGCCGGAAT
MT3 probe sequence:
GCCGCGCAACCTCTGAAAAAGGCCGGCCGTTACGTTGAAACGGCCGGAAT
c
p21
p1981
p2762
p4207
MT1
WT
Positive
Shift band
Free DNA
d
p1981
p2762
MT3
MT2
MT1
WT
Positive
e
p1981
p1981
Competitor
10x 50x 100x
MT3
WT
Negative
Positive
f
Relative Expression fold
WT
Δ1981
Δ2762
cas3
cas5
cas8
cas7
cas4
cas1
cas2
g
WT
Δ1981
CP-Δ1981
Δ2762
CP-Δ2762
h
Bacterial growth (OD600)
WT
Δ1981
CP-Δ1981
Δ2762
CP-Δ2762
Time (h)
i
Area under curve
WT
Δ1981
CP-Δ1981
Δ2762
CP-Δ2762

513

Fig. 5 Promoter-bound regulators mediate the tradeoff between the CRISPR-Cas activity and bacterial growth in phage-susceptible strain.

a, Comparison of the 500 bp *cas3* upstream regions in PR and PS strains. Predicted -35 and -10 promoter elements, RNA polymerase σ-factor binding sites and the PS strain-specific predicted regulatory motif are indicated. Red triangles mark nucleotide differences within predicted regulatory regions. **b**, Schematic of the EMSA probes. The wild-type (WT) probe was a 50 bp sequence centred on the 8 bp PS strain-specific predicted regulatory motif. In the mutant 1 (MT1) probe, the eight nucleotides of the motif were mutated by C to T and G to A transitions. In the mutant 2 (MT2) probe, all eight nucleotides of the motif were replaced with G, the mutation that abolished the predicted binding site within the motif. In the mutant 3 (MT3) probe, the first three of the four nucleotides immediately upstream of the motif, adjacent to the RNA polymerase σ-factor binding site (*rpoD17*), were mutated to A. Mutated nucleotides are indicated in the probe schematics. **c**, Direct EMSA analysis of candidate regulator binding to the PS strain *cas3* promoter. Biotin-labelled WT and mutated MT probes were incubated with the indicated purified proteins. The *E. coli* cAMP receptor protein bound to the lac promoter-operator region was used as a positive control. **d**, EMSA-based mapping of promoter sequences required for regulator binding. **e**, Competition EMSA using labelled WT promoter probe and unlabelled competitor DNA. The lac promoter-operator sequence was used as a nonspecific competitor control. Shifted and free probes indicate protein-bound and unbound DNA, respectively. **f**, RT-qPCR analysis of *cas* genes expression in PS strain regulator deletion mutants. Expression was normalized to recA and calculated using the $2^{-\Delta\Delta Ct}$ method. Bars show the mean of three biological replicates, each measured with three technical replicates. **g**, Phage susceptibility of wild-type PS strain, regulator deletion mutants and complemented strains. Δ indicates the corresponding deletion mutant, and

CP-Δ indicates the complemented strain. **h**, Growth curves of wild-type PS strain, regulator deletion mutants and complemented strains. OD600 was measured every 2 h for 18 h. **i**, Area-under-the-curve analysis of the growth curves shown in **h**. Statistical comparisons in **f** and h were performed using two-way ANOVA with multiple comparisons. Area-under-the-curve values in **i** were analysed using one-way ANOVA with multiple comparisons. ns: not significant; *: $P \leq 0.05$; **: $P \leq 0.01$; ***: $P \leq 0.001$.

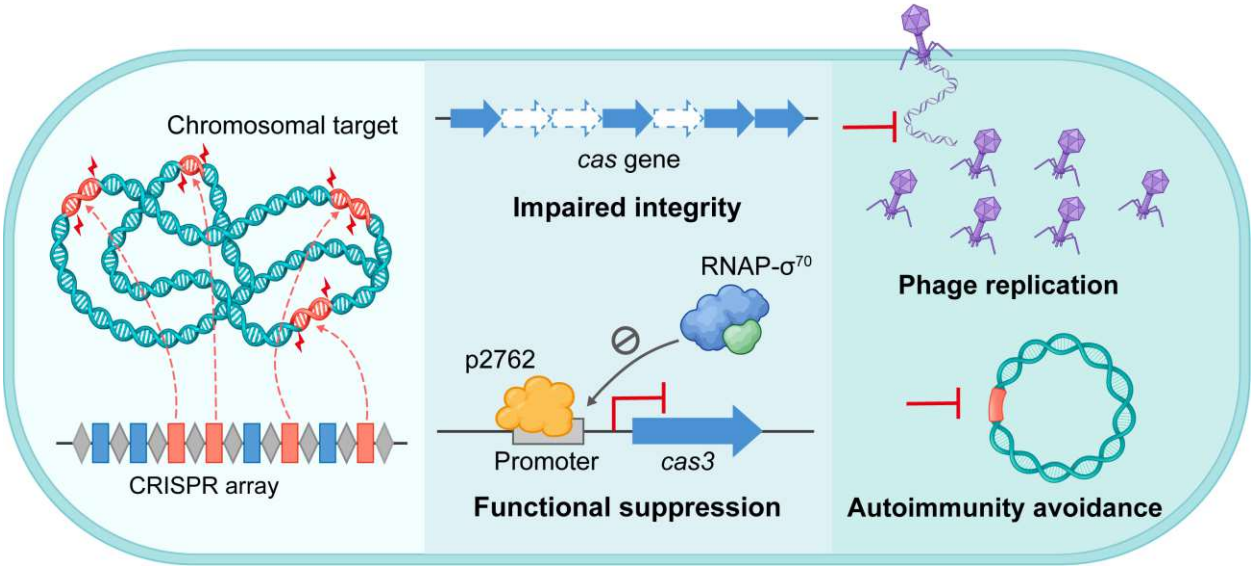
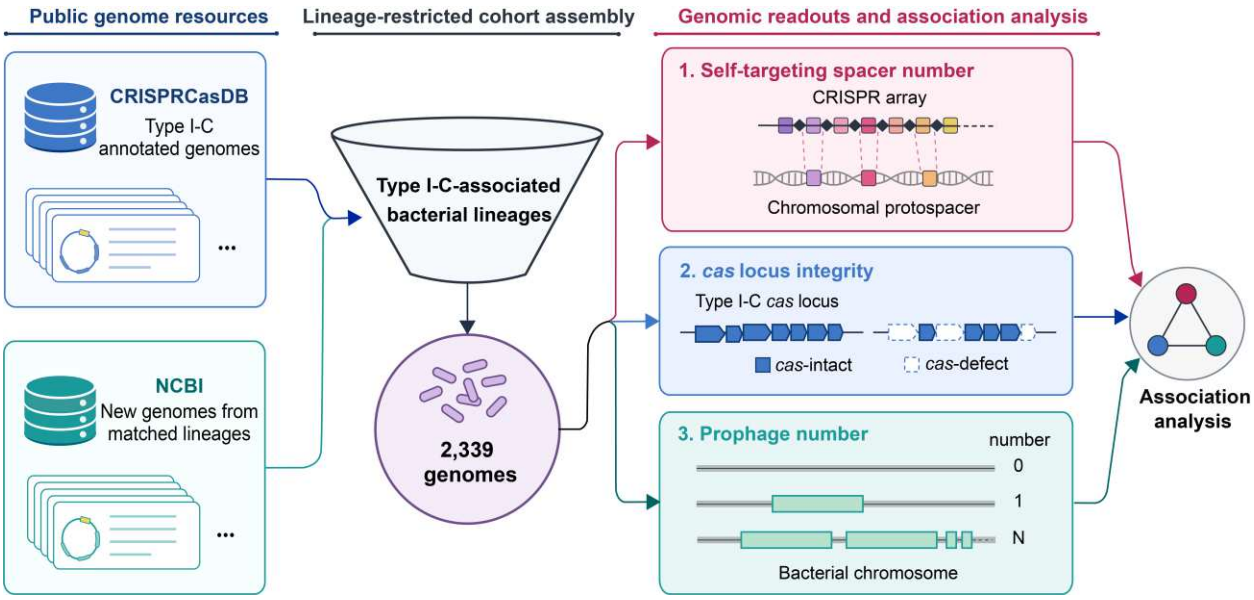


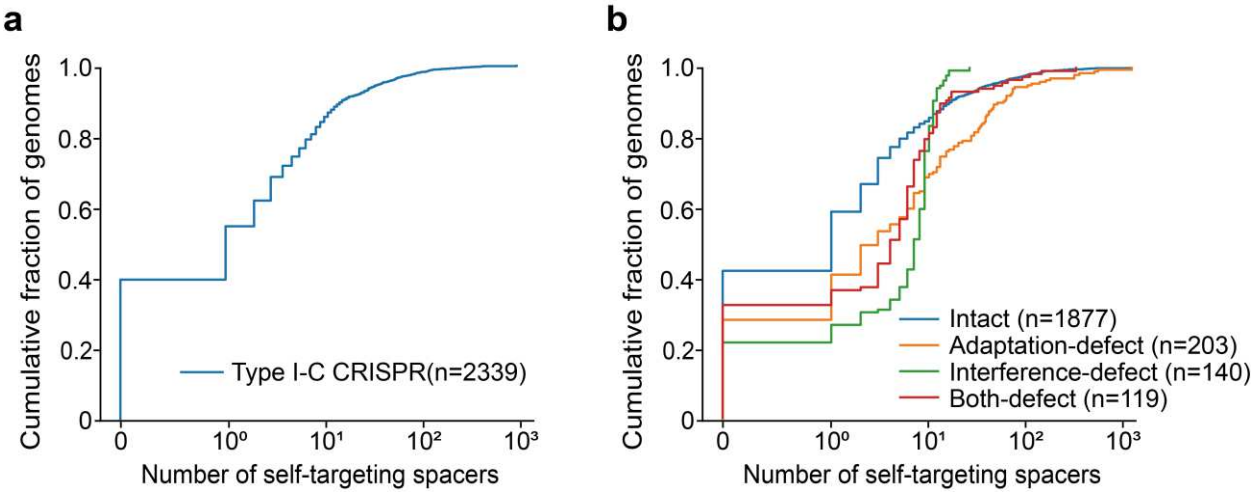
Fig. 6 The trade-off between avoidance of CRISPR-mediated autoimmunity and phage defence.

In bacteria enriched in self-targeting spacers, matches between CRISPR spacers and chromosomal sequences create a potential risk of autoimmunity. Two ways may mitigate this risk in type I-C CRISPR-Cas systems: loss of system integrity through deletion of *cas* genes, and functional suppression through transcriptional repression. In the latter route, the promoter-binding regulator p2762 represses *cas3* expression, potentially by interfering with promoter binding of the RNA polymerase (RNAP) σ^{70} holoenzyme. Limiting self-targeting interference protects host DNA from autoimmune damage. However, the accompanying reduction in CRISPR-Cas-mediated anti-phage immunity facilitates phage replication. Red segments denote self-targeting spacers and their

matching chromosomal targets. Dashed gene outlines indicate deletion *cas* genes. Red inhibitory bars represent suppression of transcription or inhibition of CRISPR-mediated DNA targeting, and the grey prohibition symbol indicates the inhibition of RNAP-σ^{70} binding to the promoter.

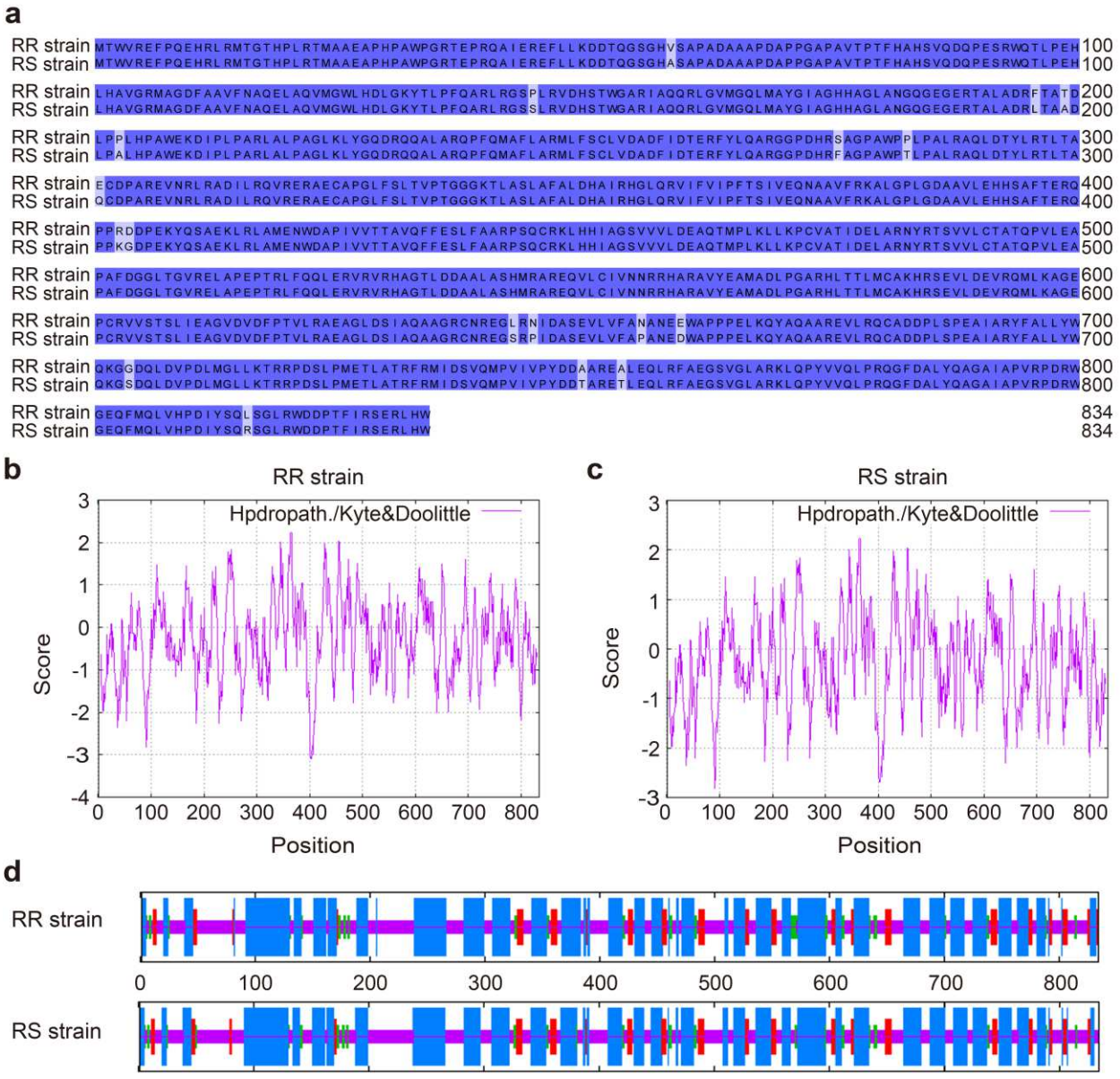


Supplementary Fig. 1 Workflow used to assemble a genome-scale discovery cohort from CRISPRCasdb-annotated type I-C genomes and newly deposited NCBI genomes from type I-C-associated bacterial lineages.



Supplementary Fig. 2 Distributional analyses of genome level self-targeting spacer abundance across genomes from type I-C-associated lineages.

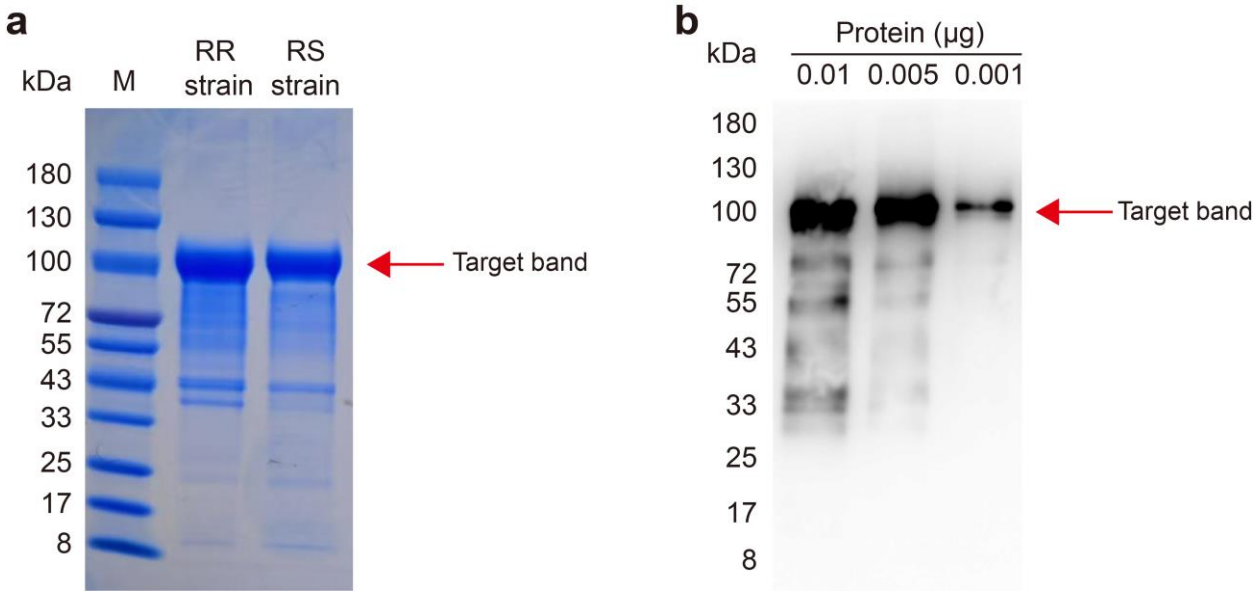
a, Empirical cumulative distribution functions (ECDFs) of self-targeting spacer numbers across the 2,339-genome discovery cohort. **b**, ECDFs of self-targeting spacer numbers stratified by type I-C module integrity. Adaptation-defect genomes carried disruption or absence of *cas1*, *cas2* or *cas4*, interference-defect genomes carried disruption or absence of *cas3*, *cas5*, *cas8* or *cas7*, and both-defect genomes carried defects in both modules.



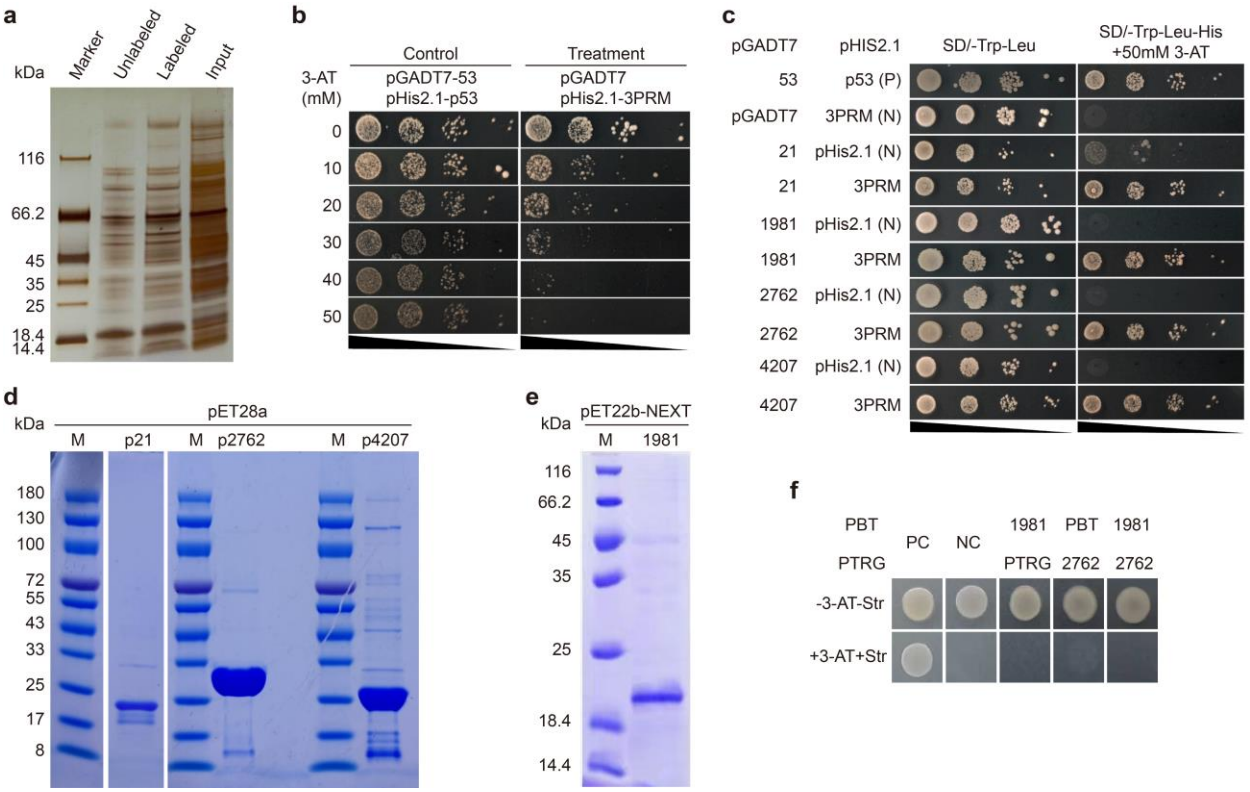
Supplementary Fig. 3 Comparative bioinformatics analysis of Cas3 from both the phage-resistant (PR) and phage-susceptible (PS) strains.

a, Sequence alignment. Cas3 from the two strains shared 97.84% amino acid identity. **b, c**, Kyte&Doolittle hydropathy profiles (ProtScale). ProtParam analysis predicted aliphatic indices of 86.1 and 85.41 and GRAVY scores of −0.262 and −0.263 for PR-strain Cas3 and PS-strain Cas3, respectively (GRAVY, grand average of hydropathicity; range, −2 to 2; positive values, hydrophobic; negative values, hydrophilic). **d**, Secondary-structure prediction using SOPMA.

Predicted α-helices accounted for 48.32% and 47.12% of PR-strain Cas3 and PS-strain Cas3, respectively; random coils accounted for 37.65% and 39.21%, extended strands for 8.99% and 8.87%, and β-turns for 5.04% and 4.80%.



Supplementary Fig. 4 Validation of recombinant Cas3 protein preparation and antibody detection. **a**, SDS-PAGE analysis of refolded and concentrated recombinant Cas3 proteins of the PR and PS strains expressed from pET28a. Prominent bands at approximately 100 kDa correspond to the expected Cas3 fusion proteins. M, molecular weight marker. **b**, Western blot validation of the PS strain Cas3-specific polyclonal antibody using recombinant PS-strain Cas3 expressed from pET28a, loaded at 0.01 μg, 0.005 μg and 0.001 μg per lane.



Supplementary Fig. 5 Identification of phage-susceptible strain *cas3* promoter-binding proteins.

a, Silver-stained SDS-PAGE of proteins recovered by DNA pull-down using the PS strain *cas3* promoter probe. The unlabeled lane shows pull-down products obtained with unlabeled DNA probe, beads and PS strain total protein; the labelled lane shows pull-down products obtained with biotin-labelled DNA probe under the same conditions; input indicates total PS strain protein extract. Selected differential bands were excised for mass-spectrometry analysis. Protein concentrations were 1.24 mg ml^{-1} for the unlabeled sample and 1.47 mg ml^{-1} for the labelled sample. Loading volumes were 12 µl for the unlabeled sample, 10 µl for the labelled sample and 10 µl for input. M, molecular weight marker. **b**, Autoactivation control of the 3PRM yeast one-hybrid bait. Three tandem copies of the predicted regulatory motif (PRM) from the PS strain *cas3* promoter were cloned into pHis2.1 and co-transformed with empty pGADT7 into yeast strain Y187. Growth was

assessed on SD/-Leu-Trp-His plates containing 0-50 mM 3-AT. The positive control was p53-pHis2.1 co-transformed with pGADT7-53. **c**, Yeast one-hybrid screening of candidate PS strain *cas3* promoter-binding proteins. P, positive control; N, negative control; pGADT7 and pHis2.1, empty vectors; 3PRM, three tandem PRM bait. **d**, Heterologous expression of p21, p2762 and p4207 using pET28a. **e**, Heterologous expression of p1981 using pET22b-NEXT. M, molecular weight marker. **f**, Interaction between p1981 and p2762 by bacterial two-hybrid analysis. PC indicates the positive control containing pBT-LGF2 and pTRG-Gal11p. NC indicates the negative control containing empty pBT and pTRG vectors. Data represent three independent biological replicates, each with three technical replicates.

Methods

Bacterial strains, phage and growth conditions

All bacterial strains, plasmids and phage used in this study are listed in Supplementary Table 8. *Acidovorax oryzae* strains RS-1 and RS-2 were cultured at 28 °C in LB medium. *Escherichia coli* DH5*α*, BL21(DE3) and XL1-Blue strains were grown at 37 °C in LB medium, and *Saccharomyces cerevisiae* strain Y187 was cultured at 30 °C in YPDA medium. Solid and semi-solid LB media contained 1.5% and 0.8% agar, respectively. When required, antibiotics were used at the following final concentrations: kanamycin, 50 μg ml^{-1} for pJP5603, pET28a and pHis2.1 derivatives and 30 μg mL^{-1} for *E. coli* XL1-Blue; ampicillin, 100 μg mL^{-1} for pRADK, pET22b-NEXT and pGADT7 derivatives; spectinomycin, 50 μg mL^{-1} for pHM1 derivatives; gentamicin, 30 μg mL for pHERD30T derivatives; chloramphenicol, 34 μg mL^{-1} for pBT derivatives; and tetracycline, 12.5 μg mL^{-1} for pTRG derivatives. Phage AP1 was propagated on wild-type RS-2 and stored in Sodium Chloride-Magnesium sulfate (SM) buffer at 4 °C. Phage AP1 was unable to infect strain RS-1 (phage-resistant strain), but can infect strain RS-2 (phage-susceptible strain) [49].

Genome cohort assembly

To examine whether self-targeting spacers abundance is associated with type I-C CRISPR-Cas at genome level, we assembled a lineage-bounded discovery cohort from public databases. Genomes encoding type I-C systems were first retrieved from CRISPRCasdb [25] as of 6 August 2022 to define type I-C-associated taxa. To capture newly sequenced diversity within these lineages, we then collected genomes belonging to the same taxa, which were deposited in NCBI between August 2022 and January 2026. Newly retrieved NCBI genomes were screened against CRISPRCasdb-derived records to identify assembly-version overlaps. For overlapping assemblies, the most recent version was retained and counted once in the final cohort. Because public genome collections are taxonomically imbalanced, genomes were grouped by the lowest reliably assigned taxonomic rank, preferentially genus; when genus-level assignment was unavailable, the lowest available higher-rank lineage label was used. All genomes were retained for groups with fewer than 100 genomes, whereas 100 genomes were randomly sampled without replacement from groups with 100 or more genomes. This strategy was designed to identify genome-wide associations of self-targeting spacers with *cas* locus integrity or prophage carriage within type I-C-associated lineages.

Detection and classification of CRISPR-Cas systems

CRISPR arrays, *cas* loci and *cas* genes were initially annotated using CRISPRCasFinder with default settings [50], including the *cas* gene detection module. To verify Cas protein assignment and reduce false absence calls, predicted cas proteins were re-analysed using CasFinder v3.1.0. The combined CRISPRCasFinder and CasFinder v3.1.0 annotations were used as the basis for defining *cas* locus completeness and module integrity in downstream analyses. The final classification of each genome can be found in Supplementary Table 9-10.

Prophage detection

Prophage regions were predicted from the genome assemblies used for the genome-wide self-targeting spacer analysis. Candidate viral regions were first identified with VirSorter2, which was used to recover putative viral nucleotide sequences from each assembly [51]. Candidate regions were then evaluated with CheckV to assess viral-genome quality and refine provirus boundaries [52]. Predictions were retained when estimated completeness was at least 50%, CheckV classified the region as a provirus, and low-confidence, low-quality or undetermined predictions were excluded. For retained proviruses, the host-trimmed viral intervals inferred by CheckV were mapped back to the corresponding VirSorter2 candidate regions to obtain corrected genomic coordinates. The predicted prophages are presented in Supplementary Table 11.

Detection of self-targeting spacers

Self-targeting spacers were detected by comparing CRISPR spacers with their source genomes using BLASTN-based spacer-genome searches [53], following the general spacer-to-source-genome strategy used in previous self-targeting spacer analyses [14]. Self-targeting spacers were identified from all CRISPR arrays in genomes harboring at least one type I-C locus. BLASTN hits were retained as candidate self-targeting events when the alignment started at the first nucleotide of the spacer, sequence identity was at least 90%, the number of mismatches was no more than three, the absolute difference between alignment length and spacer length was no more than six, and the sum of the mismatch number and this length difference was no more than nine. Matches corresponding to CRISPR-array spacer positions were not considered self-targets. Chromosomal protospacers targeted by self-targeting spacers were annotated using Bakta [54]. The results of self-targeting spacer detection are shown in Supplementary Table 12.

Prophage carriage and genome level self-targeting spacer analysis

For each genome in the cohort, curated VirSorter2-CheckV prophage coordinates were used to determine prophage carriage and prophage number. Self-targeting spacer calls were summarized as self-targeting spacer presence or absence and as the number of self-targeting spacers per genome. The prevalence and abundance of self-targeting spacers were then compared between prophage-containing and prophage-free genomes, and the relationship between prophage number and self-targeting spacer number was analysed as described in Statistics and reproducibility. To determine whether selected self-targeting protospacers were located within prophage regions or phage-related genes, self-targeting spacers protospacer coordinates were intersected with curated prophage coordinates and local gene annotations in two closely related strains with conserved CRISPR-Cas loci but different susceptibility to phage.

Phylogenetic analyses

Whole-genome phylogeny of *Acidovorax* strains was inferred from core protein sequences. Genome sequences were annotated using the RAST prokaryotic genome annotation server [55], and PGAP was run using the GeneFamily method for 108 whole-genome sequences [56]. After the first clustering step, core protein sequences were extracted from the Orthologs_Cluster output, concatenated and aligned with MAFFT [57]. A genome phylogenetic tree was then constructed in MEGA7 [58] using the neighbor-joining method with the Poisson model and 1,000 bootstrap replicates. For Cas1 phylogeny, Cas1 protein sequences were analysed in MEGA7.0, and a maximum-likelihood tree was constructed using the Jones-Taylor-Thornton evolutionary model [59,60].

Analysis of phage protospacers

Phage protospacers matching endogenous spacers in bacteria were first identified using NCBI BLASTN with the 'somewhat similar sequences' option and otherwise default parameters. Retained self-targeting spacer hits were then annotated at the protospacer level using the BLASTN alignment and genome coordinates. Mismatch number and mismatch position were recorded from each spacer-protospacer alignment, with particular attention to mismatches in the seed-proximal region. Protospacer-flanking sequences were extracted from the corresponding genome assemblies to evaluate the type I-C Protospacer Adjacent Motif (PAM) used in this study, including intact TTC PAMs.

RNA sequencing

Genomes of *A. oryzae* strains RS-1 and RS-2 have been sequenced in our previous study with NCBI accession numbers CP169238 and CP169239, respectively [49]. For transcriptome sequencing, each strain was first grown overnight in LB medium to an OD600 of approximately 1.0. A 500 μl aliquot of each culture was then transferred into 50 ml fresh LB medium and grown to an OD600 of approximately 0.6, with three biological replicates per strain. Cells were collected by centrifugation at 11,000 *g* for 10 min at 4 °C, the supernatant was discarded, and pellets were stored at -80 °C before shipment on dry ice. Library construction and transcriptome sequencing were performed by Benagen. Raw image data were converted into sequence reads, and reads containing adapters, reads containing ambiguous N bases and low-quality reads were removed. Low-quality reads were defined as reads in which bases with Qphred scores of 20 or lower accounted for more than 50% of the read. Clean reads were aligned to the corresponding RS-1 or RS-2 genome and gene annotation files using Bowtie2 [61]. HTSeq v0.6.1 was used to count reads

mapped to each gene [62], and FPKM values were calculated according to gene length and mapped read number. Read alignments were visualized using Integrative Genomics Viewer [63].

RT-qPCR

Total RNA was extracted as described for RNA sequencing. Genomic DNA was removed, and cDNA was synthesized using HiScript II Q RT SuperMix for qPCR with gDNA Wiper (Vazyme, Nanjing, China). Quantitative PCR was performed using ChamQ SYBR qPCR Master Mix (Vazyme) in 20-μl reactions containing 10 μl of 2x SYBR master mix, 0.4 μl of each primer at 10 μM, 1 microlitre of cDNA and nuclease-free water. Cycling conditions were 95 °C for 30 s followed by 40 cycles of 95 °C for 10 s and 60 °C for 30 s. Melting-curve analysis was used to assess amplification specificity. Expression levels were normalized to recA and calculated using the $2^{-\Delta\Delta Ct}$ method. Each assay included three biological replicates, each with three technical replicates.

Gene deletion and complementation

Gene deletion mutants of *A. oryzae* were generated by homologous recombination using the suicide plasmid pJP5603, as described previously [64]. For each target gene, an internal fragment of approximately two-thirds of the coding sequence was amplified from bacterial genomic DNA, digested and ligated into pJP5603. Recombinant plasmids were transformed into *E. coli* S17-1 λpir, selected on kanamycin-containing LB agar and verified by Sanger sequencing. Confirmed plasmids were introduced into recipient strains by electroporation, and gene disruption mutants were selected on kanamycin-containing LB agar. For complementation, full-length target genes together with approximately 500 bp of upstream native promoter sequence were amplified with

high-fidelity polymerase (Vazyme), cloned into pRADK or pHM1, verified by sequencing and introduced into the corresponding mutant strains.

Phage infection assays

Single bacterial colonies were inoculated into LB medium and cultured at 30 °C with shaking until the cultures reached an OD600 of 1.0. For each assay, 500 µl of bacterial culture was mixed with 5 ml of molten semi-solid LB medium and immediately poured onto an LB agar plate to generate a double-layer plate. The plates were left in a clean bench for 30 min to allow the upper agar layer to solidify. Concentrated phage lysates were serially diluted tenfold, and 2.5 µl of each dilution was spotted onto the surface of the double-layer plate. After the droplets were allowed to absorb for 30 min, the plates were incubated at 30 °C for 24 h. Each treatment was performed in triplicate and repeated three times.

Plasmid interference assays

Spacer-directed interference was tested by transformation of protospacer-bearing plasmids. For single-spacer assays, oligonucleotides encoding 32-nt protospacer sequences with 5-nt flanking regions, including the cognate PAM, were annealed and cloned into pHERD30T. For self-targeting tests, selected protospacer sequences were synthesized after correcting mismatches to generate perfect spacer-protospacer complementarity. For multiplex assays, a 312 bp synthetic fragment containing the first seven CRISPR1 spacers, each flanked by 5 nt on both sides including PAM sequences, was cloned into pHERD30T. Recombinant plasmids and the empty pHERD30T control were introduced into PR strain by electroporation. Transformants were recovered on gentamicin-containing LB agar at 28 °C for 48 h. Colony-forming units were counted, and transformation efficiency was normalized to the empty-vector control from the same experiment.

Protein expression and purification

Coding sequences of target genes were amplified with high-fidelity DNA polymerase (Vazyme) and cloned into pET28a or pET22b-NEXT. Constructs were verified by Sanger sequencing and transformed into *E. coli* BL21(DE3). Small-scale expression tests were used to determine soluble expression. p21, p2762 and p4207 were expressed as His-tagged proteins from pET28a. Induction conditions were 1 mM IPTG (Solarbio, Beijing, China) at 37 °C for pET28a-p21, 0.5 mM IPTG at 37 °C for pET28a-p2762 and 0.5 mM IPTG at 16 °C for pET28a-p4207. p1981 showed improved soluble expression after subcloning into pET22b-NEXT after induction with 0.5 mM IPTG at 30 °C. Cas3 from both the phage-resistant and phage-susceptible strains remained largely insoluble under the tested pET28a expression conditions and were purified under denaturing conditions. Cells were harvested, resuspended in 1 × phosphate-buffered saline (PBS), frozen at -80 °C, lysed by sonication and centrifuged to separate soluble and insoluble fractions. His-tagged proteins were purified using Ni-NTA affinity purification (Beyotime, Shanghai, China). Protein concentration was measured by BCA assay, and purity was assessed by SDS-PAGE and Coomassie staining.

Cas3 antibody preparation and western blotting

Cas3 purified from the phage-susceptible strain was refolded by stepwise dialysis against buffers containing 6, 4, 3, 2, 1 and 0 M urea. Dialysis buffer contained PBS supplemented with 5% glycerol, 1% L-arginine and 2% glycine. Refolded protein was concentrated by ultrafiltration, analysed by SDS-PAGE and quantified. A total of 2-3 mg purified protein was used for rabbit polyclonal antibody production (Huaan, Hangzhou, China). For western blotting, total cellular proteins from phage resistant and susceptible strains were separated by SDS-PAGE and transferred to PVDF membranes (Merck, Darmstadt, Germany). Membranes were blocked, incubated

overnight at 4 °C with rabbit anti-Cas3 antibody diluted 1:5,000, washed and incubated with goat anti-rabbit IgG secondary antibody. Signals were developed using an ECL chemiluminescence kit (Beyotime) and imaged with an ImageQuant LAS 4000 mini system.

LC-MS/MS analysis

Bacteria were cultured in LB medium at 30 °C with shaking at 220 rpm to an OD600 of approximately 0.8. Cells were harvested, washed and lysed by sonication in PBS. Total protein extracts were mixed with SDS loading buffer, denatured at 95 °C for 5 min and separated by SDS-PAGE. Gels were stained with Coomassie Brilliant Blue, and visible bands were excised for in-gel digestion followed by LC-MS/MS analysis. Mass spectrometry was performed on an LTQ Orbitrap Elite instrument at the Mass Spectrometry Platform of Zhejiang University.

Cas3 nuclease activity assays

Nuclease activity of purified Cas3 was tested using single-stranded M13mp18 DNA and supercoiled double-stranded pUC57 plasmid substrates. Reactions contained 20 nM DNA substrate in NB1 buffer consisting of 20 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$ and 0.2 mg ml^{-1} BSA, together with the indicated concentration of Cas3 diluted in storage buffer containing 10 mM Tris-HCl pH 7.5, 300 mM KCl, 50% glycerol and 1 mM DTT. Substrate and protein solutions were pre-incubated separately at 37 °C, mixed and incubated for 120 min. Reactions were terminated with 3x stop solution containing 67.5 mM EDTA, 27% glycerol and 0.3% SDS. Products were resolved by 0.8% agarose gel electrophoresis and visualized after staining.

Cas3 sequence and structural analyses

Cas3 amino acid composition and physicochemical properties were analysed using BioEdit [65] and ProtParam [66]. Sequence alignment was performed using Clustal Omega through the EMBL-EBI

Job Dispatcher framework [67] and visualized with Jalview v2.2 [68]. Hydrophobicity profiles were calculated using the Kyte-Doolittle method implemented in ProtScale [69]. Conserved domains were identified using NCBI Conserved Domain Search [70]. Secondary structure was predicted using SOPMA [71]. Three-dimensional models were generated using AlphaFold 3 [73] and visualized and compared in PyMOL v3.0.3. Putative transcription factor binding sites within the 500 bp upstream regions of Cas3 coding sequences were predicted using Softberry.

DNA pull-down

A DNA fragment of approximately 500 bp upstream of Cas3 was amplified in phage-susceptible strain by PCR to generate promoter probes. Biotin-labelled and unlabelled probes were prepared (GeneCreate, Wuhan, China) and purified after agarose gel electrophoresis. Biotin-labelled DNA was immobilized on streptavidin-coated magnetic beads. Total protein was extracted using RIPA buffer (Beyotime), sonication and centrifugation at 12,000 g for 10 min at 4 °C, and protein concentration was determined by BCA assay. Equal amounts of total protein were incubated overnight at 4 °C with DNA-bound magnetic beads in binding buffer containing 50 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 0.05% Tween 20, 50 mM KCl, 1 mM $MgCl_2$, 10% glycerol and 1 mM PMSF. Beads were washed four times with binding buffer, and bound proteins were eluted by boiling at 95 °C. Input, labelled-probe pull-down and unlabelled-probe control samples were separated by SDS-PAGE and analysed by silver staining. Differential bands were excised and identified by LC-MS/MS. MS/MS spectra were searched against the RS-2 genome annotation database using MaxQuant [72], and proteins were retained at a false discovery rate (FDR) threshold of 0.05.

Yeast one-hybrid assay

Yeast one-hybrid assays were used to identify proteins capable of *in vivo* binding the *cas3* promoter region in phage-susceptible strain. 11 candidates were selected from the DNA pull-down LC-MS/MS results and genome annotations: three proteins annotated as transcriptional regulators (p21, p815 and p3595), one H-NS-like protein previously implicated in repression of CRISPR-Cas systems in *E. coli* (p1981), and seven hypothetical proteins of unknown function (p649, p692, p973, p1013, p1536, p2762 and p4207). Bait fragments were inserted into pHis2.1 to generate pBait-His constructs, and candidate coding sequences were cloned into pGADT7 using homologous recombination-based cloning. The pBait-His construct and each pGADT7 prey construct were co-transformed into yeast strain Y187. Transformants were selected on SD/-Leu-Trp plates at 30 °C for 3 d. Because the native promoter fragment showed autoactivation, 3-amino-1,2,4-triazole (3-AT) was titrated from 0 to 50 mM to determine inhibitory conditions, and a reporter containing three tandem copies of the phage-susceptible strain-specific regulatory motif was used for validation. Colonies were serially diluted and spotted onto SD/-Leu-Trp-His plates containing the selected 3-AT concentration to assess protein-DNA interactions.

Electrophoretic mobility shift assays

DNA probes corresponding to the predicted promoter-binding regions were synthesized as biotin-labelled probes, unlabeled competitors and mutant probes (GeneCreate). Complementary oligonucleotides were annealed to generate double-stranded probes. EMSA reactions were performed in 10 μl containing 1x EMSA/Gel-shift binding buffer, labelled probe, 2 μg purified protein and glycerol. For competition assays, unlabeled competitor probe was added before the labelled probe. Reaction components except the labelled probe were pre-incubated at room temperature for 10 min, after which labelled probe was added and incubation continued for 50 min.

Samples were resolved on 6% native polyacrylamide gels in 0.5x TBE at 80 V under cooling conditions. DNA-protein complexes were transferred to PVDF membranes, UV-crosslinked, blocked with 1% casein in PBS, incubated with HRP-conjugated streptavidin and detected by chemiluminescence using an ImageQuant LAS 4000 mini system.

Bacterial two-hybrid assays

Protein-protein interactions among candidate regulators were tested using bacterial two-hybrid assays. Target genes were cloned into pBT and pTRG vectors, and plasmid pairs were co-transformed into *E. coli* XL1-Blue. Co-transformants were plated on non-selective medium containing kanamycin, chloramphenicol and tetracycline and on dual-selective M9 His-dropout medium containing 3-AT. Plates were incubated at 30 °C for 12-24 h. pBT-LGF2 and pTRG-Gal11p were used as positive controls, and empty-vector combinations were used as negative controls.

Bacterial growth assays

Growth curves were used to assess the fitness effects of deleting and complementing candidate regulatory genes. Bacteria were cultured overnight at 28 °C in 5 ml NB liquid medium, collected by centrifugation at 8,000 *g* for 3 min and washed twice with sterile double-distilled water. Cells were resuspended in NB liquid medium and adjusted to an OD600 of 0.1. Cultures were incubated at 28 °C with shaking at 220 rpm. OD600 was measured every 2 h for 18 h using a DeTie HBS-1096A V4.4 instrument. Growth curves were compared by two-way analysis of variance, and overall growth performance was quantified by area-under-the-curve analysis followed by one-way analysis of variance with multiple comparisons.

Statistics and reproducibility

Statistical tests were selected according to the data type and comparison described in the Results. Distributional differences in genome level self-targeting spacer abundance between *cas*-intact and disrupted or module-defect groups were analysed using the two-sided Kolmogorov-Smirnov test. Categorical associations between prophage carriage and self-targeting spacers detection were analysed using Fisher's exact test. Differences in self-targeting spacer abundance between prophage-containing and prophage-free genomes were analysed using the Mann-Whitney U test. Correlations between prophage number and self-targeting spacer number were analysed using Spearman's rank correlation. RNA-seq, RT-qPCR and growth-curve data were analysed using two-way analysis of variance. Area-under-the-curve values were analysed by one-way analysis of variance with multiple comparisons. Empirical cumulative distribution functions, symlog-scaled plots and other genome-scale visualizations were calculated and generated using Python scripts. Exact *P* values are reported in the Results where available.

Data availability

The genome sequences analysed in this study were retrieved from the NCBI databases and CRISPRCasdb, with accession numbers listed in Supplementary Table 9. The complete genome sequences of *A. oryzae* strains RS-1 and RS-2 are available under GenBank accessions CP169238 and CP169239, respectively. All other necessary data supporting this study are included in the Supplementary Files. Additional data will be made available from the corresponding author on request.

Acknowledgements

The work is partially supported by the National Natural Science Foundation of China (32372614 and 32572912), the Zhejiang Provincial Natural Science Foundation of China (LZ24C140004), and Zhejiang Province "San Nong Jiu Fang" Science and Technology Plan Project (2025SNJF027).

Author contributions

W.T.L., M.J.L. and B. L. conceived and designed the study. W.T.L. and M.J.L. performed the experiments. A.H.X. conducted the bioinformatic analyses. W.T.L. and M.J.L. analysed the data and wrote the first draft of the manuscript. W.T.L., M. I. and B. L. revised and edited the manuscript. B. L. acquired funding and supervised the project. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT in order to improve the language and readability of some sentences. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

1. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317–327 (2010).

915 2. Doron, S. et al. Systematic discovery of antiphage defence systems in the microbial pangenome.
916 *Science* **359**, eaar4120 (2018).
917 3. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their
918 phage foes. *Nature* **577**, 327–336 (2020).
919 4. Bernheim, A. & Sorek, R. The pan-immune system of bacteria: antiviral defence as a community
920 resource. *Nat. Rev. Microbiol.* **18**, 113–119 (2020).
921 5. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes.
922 *Science* **315**, 1709–1712 (2007).
923 6. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in
924 staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).
925 7. Brouns, S. J. J. et al. Small CRISPR RNAs guide antiviral defence in prokaryotes. *Science* **321**,
926 960–964 (2008).
927 8. Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR—a widespread system that provides acquired
928 resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* **6**, 181–186 (2008).
929 9. Levy, A. et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA.
930 *Nature* **520**, 505–510 (2015).
931 10. Jackson, S. A. et al. CRISPR-Cas: adapting to change. *Science* **356**, eaal5056 (2017).
932 11. Hille, F. et al. The biology of CRISPR-Cas: backward and forward. *Cell* **172**, 1239–1259
933 (2018).
934 12. van Houte, S. et al. The diversity-generating benefits of a prokaryotic adaptive immune system.
935 *Nature* **532**, 385–388 (2016).

936 13. Stern, A., Keren, L., Wurtzel, O., Amitai, G. & Sorek, R. Self-targeting by CRISPR: gene
937 regulation or autoimmunity? *Trends Genet.* **26**, 335–340 (2010).
938 14. Nóbrega, F. L., Walinga, H., Dutilh, B. E. & Brouns, S. J. J. Prophages are associated with
939 extensive CRISPR-Cas auto-immunity. *Nucleic Acids Res.* **48**, 12074–12084 (2020).
940 15. Vercoe, R. B. et al. Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape
941 bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet.* **9**, e1003454
942 (2013).
943 16. Gomaa, A. A. et al. Programmable removal of bacterial strains by use of genome-targeting
944 CRISPR-Cas systems. *mBio* **5**, e00928-13 (2014).
945 17. Bikard, D. et al. Exploiting CRISPR-Cas nucleases to produce sequence-specific
946 antimicrobials. *Nat. Biotechnol.* **32**, 1146–1150 (2014).
947 18. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that
948 inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013).
949 19. Bondy-Denomy, J. et al. Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR
950 proteins. *Nature* **526**, 136–139 (2015).
951 20. Pawluk, A. et al. Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse
952 bacterial species. *Nat. Microbiol.* **1**, 16085 (2016).
953 21. Borges, A. L., Davidson, A. R. & Bondy-Denomy, J. The discovery, mechanisms, and
954 evolutionary impact of anti-CRISPRs. *Annu. Rev. Virol.* **4**, 37–59 (2017).
955 22. Rauch, B. J. et al. Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell* **168**, 150–
956 158.e10 (2017).

957 23. Hynes, A. P. et al. Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range
958 of Cas9 proteins. *Nat. Commun.* **9**, 2919 (2018).
959 24. Landsberger, M. et al. Anti-CRISPR phages cooperate to overcome CRISPR-Cas immunity.
960 *Cell* **174**, 908–916.e12 (2018).
961 25. Pourcel, C. et al. CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas
962 genes from complete genome sequences, and tools to download and query lists of repeats and
963 spacers. *Nucleic Acids Res.* **48**, D535–D544 (2020).
964 26. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif
965 sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**,
966 733–740 (2009).
967 27. Semenova, E. et al. Interference by clustered regularly interspaced short palindromic repeat
968 (CRISPR) RNA is governed by a seed sequence. *Proc. Natl Acad. Sci. USA* **108**, 10098–10103
969 (2011).
970 28. Gong, B. et al. Molecular insights into DNA interference by CRISPR-associated nuclease-
971 helicase Cas3. *Proc. Natl Acad. Sci. USA* **111**, 16359–16364 (2014).
972 29. Huo, Y. et al. Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated
973 DNA unwinding and degradation. *Nat. Struct. Mol. Biol.* **21**, 771–777 (2014).
974 30. He, L., St. John James, M., Radovcic, M., Ivancic-Bace, I. & Bolt, E. L. Cas3 protein—a
975 review of a multi-tasking machine. *Genes* **11**, 208 (2020).
976 31. Nimkar, S. & Anand, B. Cas3/I-C mediated target DNA recognition and cleavage during
977 CRISPR interference are independent of the composition and architecture of Cascade
978 surveillance complex. *Nucleic Acids Res.* **48**, 2486–2501 (2020).

32. Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2**, 1849–1861 (2007).

33. Rollie, C. et al. Targeting of temperate phages drives loss of type I CRISPR–Cas systems. *Nature* **578**, 149–153 (2020).

34. Pul, U. et al. Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* **75**, 1495–1512 (2010).

35. Westra, E. R. et al. H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* **77**, 1380–1393 (2010).

36. Borges, A. L. et al. Bacterial alginate regulators and phage homologs repress CRISPR–Cas immunity. *Nat. Microbiol.* **5**, 679–687 (2020).

37. Workman, R. E. et al. A natural single-guide RNA repurposes Cas9 to autoregulate CRISPR-Cas expression. *Cell* **184**, 675–688.e19 (2021).

38. Makarova, K. S. et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).

39. Mohanraju, P. et al. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147 (2016).

40. Kahramanoglou, C. et al. Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res.* **39**, 2073–2091 (2011).

41. Grainger, D. C. Structure and function of bacterial H-NS protein. *Biochem. Soc. Trans.* **44**, 1561–1569 (2016).

42. Singh, K., Milstein, J. N. & Navarre, W. W. Xenogeneic silencing and its impact on bacterial genomes. *Annu. Rev. Microbiol.* **70**, 199–213 (2016).

43. Stoltzfus, M. J., Workman, R. E., Keith, N. C. & Modell, J. W. A dynamic subpopulation of CRISPR–Cas overexpressers allows Streptococcus pyogenes to rapidly respond to phage. *Nat. Microbiol.* **9**, 2410–2421 (2024).

44. Wimmer, F. et al. Interrogating two extensively self-targeting Type I CRISPR-Cas systems in *Xanthomonas albilineans* reveals distinct anti-CRISPR proteins that block DNA degradation. *Nucleic Acids Res.* **52**, 769–783 (2024).

45. Deveau, H. et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).

46. Varble, A. et al. Prophage integration into CRISPR loci enables evasion of antiviral immunity in *Streptococcus pyogenes*. *Nat. Microbiol.* **6**, 1516–1525 (2021).

47. Keith, N. C., Snyder, R. A., Euler, C. W. & Modell, J. W. Bacteria exploit viral dormancy to establish CRISPR-Cas immunity. *Cell Host Microbe* **33**, 330–340.e6 (2025).

48. Chen, H., Mayer, A. & Balasubramanian, V. A scaling law in CRISPR repertoire sizes arises from the avoidance of autoimmunity. *Curr. Biol.* **32**, 2897–2907.e5 (2022).

49. Liu, M. et al. Strain-specific infection of phage AP1 to rice bacterial brown stripe pathogen *Acidovorax oryzae*. *Plants* **13**, 3182 (2024).

50. Couvin, D. et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251 (2018).

51. Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).

52. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).

53. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

54. Schwengers, O. et al. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb. Genom.* **7**, 000685 (2021).

55. Aziz, R. K. et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).

56. Zhao, Y. et al. PGAP: pan-genomes analysis pipeline. *Bioinformatics* **28**, 416–418 (2012).

57. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

58. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

59. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282 (1992).

60. da Silva Xavier, A. et al. Characterization of CRISPR-Cas systems in the *Ralstonia solanacearum* species complex. *Mol. Plant Pathol.* **20**, 223–239 (2019).

61. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

62. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

63. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

64. Zhang, Y. et al. Characterization and functional analysis of *clpB* gene from *Acidovorax avenae* subsp. *avenae* RS-1. *Plant Pathol.* **66**, 1369–1379 (2017).

65. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).

66. Gasteiger, E. et al. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–607 (Humana Press, 2005).

67. Madeira, F. et al. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res.* **52**, W521–W525 (2024).

68. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

69. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

70. Marchler-Bauer, A. et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226 (2015).

71. Geourjon, C. & Deléage, G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* **11**, 681–684 (1995).

72. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

73. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable112.xlsx
- SupplementaryFigure15.pdf