Non-cognate immunity proteins provide broader defenses against interbacterial effectors in microbial communities

Abigail Knecht*[1,4], Denise Sirias*[1], Daniel R. Utter[2,3], and Karine A. Gibbs‡[1,4]

Departments of [1]Molecular and Cellular Biology and [2]Organismic and Evolutionary Biology, Harvard University; [3]Division of Geological and Planetary Sciences, California Institute of Technology; [4] Department of Plant & Microbial Biology, University of California, Berkeley

*Co-first authors

‡For correspondence: kagibbs [at] berkeley.edu

Key Words:

Effector-immunity protein pairs; DNA nuclease; PD-(D/E)XK-containing protein; bacterial communities; self/non-self recognition; *Proteus mirabilis*; *Rothia*; bacterial immunity proteins; multi-domain proteins; human microbiome; human metagenome; Type VI Secretion (T6SS); cell-cell signaling

Abstract

Dense microbial communities, like the gut and soil microbiomes, are dynamic societies. Bacteria can navigate these environments by deploying proteins (effectors) that alter foreign cells' behavior. Immunity proteins preferentially protect neighboring sibling cells, in contrast to canonical toxin-antitoxin systems. A prevailing hypothesis is that when immunity proteins are bound to specific (cognate) protein partners, it is sufficient to disrupt their function; further, there is little-to-no crosstalk with other (non-cognate) effectors. Here, we build on sporadic reports challenging these hypotheses. We show that immunity proteins from a newly defined protein family can bind and protect against non-cognate PD-(D/E)XK-containing effectors from diverse phyla. We describe the domains essential for binding and function and show that binding alone is insufficient for protective activity. Moreover, we found that these effector and immunity genes co-occur in individual human microbiomes. These results expand the growing repertoire of bacterial protection mechanisms and force us to reconsider how non-cognate interactions impact community structure within complex ecosystems.

Introduction

Specificity between protein-protein interactions is key for several biological processes, such as metabolism, development, and intercellular signaling. Binding to an incorrect partner or disrupted binding of the specific (cognate) protein can cause a diseased state or cell death[1]. For bacterial social behaviors, cognate protein-protein interactions between cells impact organism fitness and community structure, such as excluding foreign cells[2]. The importance of specificity between cognate partners in social behaviors remains largely unchallenged. In other contexts, flexible ("promiscuous") binding allows protein partners to retain their interactions when undergoing rapid mutational changes, such as during immune recognition of viral particles[3,4]. Unknown is whether flexible binding between noncognate proteins can occur during bacterial social behaviors. An expanded protective function would reveal new bacterial behaviors that influence individual fitness and community structure in dynamic ecosystems.

Microbes often exist within dense, multi-phyla communities, like the human gut microbiome, where they communicate and compete with neighbors. Bacteria can use effector-immunity protein (EI) pairs in these environments to gain advantages[5,6]. Unlike toxin-antitoxin systems in which a single cell produces both toxic and neutralizing proteins, bacteria inject cell-modifying proteins (effectors) directly into neighboring cells via several contact-dependent transport mechanisms, including the type VI secretion system (T6SS), type IV secretion system (T4SS), and contact-dependent inhibition (CDI)[7,8]. Clonal siblings produce the matching immunity protein that modifies the effector's activity. For lethal effectors, both clonal and non-clonal cells are negatively

impacted when binding is disrupted or absent[9]. These interactions between EI pairs can shape community composition by changing bacterial fitness.

The binding specificity between matching EI pairs has historically defined immunity protein protection, but recent studies raise doubts. Antitoxins have multiple mechanisms for neutralizing their cognate toxin, usually within a single cell. By contrast, EI pairs interact within a neighboring cell which puts unique restrictions on both their protection mechanisms and their evolution[10]. The predominant model is that T6SS-associated EI pairs act like a tumbler lock-and-key, where each effector protein has a single cognate partner[11]. Most immunity proteins bind their cognate effectors, often at the active site, to neutralize effector activity[12]. As such, EI pairs typically have increased protein-protein specificity between matching pairs to protect sibling cells[13]. However, experiments with engineered proteins reveal that changes to an immunity protein could allow it to bind effectors other than its cognate partner[14,15]. Also, the T6SS-associated effector and immunity proteins from *Salmonella enterica* subsp. *enterica* serovar Typhimurium and *Enterobacter cloacae*, which are phylogenetically close, bind each other *in vitro* and protect against the other *in vivo*[16]. These studies suggest that the predominant model needs to be revised to explain the breadth of immunity protein protection.

We studied an EI pair in *Proteus mirabilis* to examine the prevailing model. This opportunistic pathogen resides in human and animal guts and can cause recurrent and persistent urinary tract infections[17]. *P. mirabilis* encodes a lethal T6SS-dependent EI pair that impacts competition and collective motility[18]. The molecular functions of the putative effector and immunity proteins remained unknown. Here, we characterized this EI pair and determined the critical residues for activity, leading to the discovery of two protein

families. We showed that these immunity proteins bind non-cognate effectors produced by bacteria from different phyla. Our results indicated that binding was insufficient to neutralize effector proteins within this immunity protein family. Further, we found that these flexible EI pairs from various phyla naturally co-occur in individual human microbiomes. This work provides compelling evidence for cross-protection and supports a fundamental reinterpretation of EI pairs and their ecological significance.

Results

RdnE is a DNA nuclease and seeds a PD-(D/E)XK subfamily.

To compete against other strains, *P. mirabilis* strain BB2000 needs both the *idrD* gene and the T6SS, suggesting that the *idrD*-encoded protein functions as a T6SS-associated effector[18]. Similar effectors often contain an enzymatic domain in the C-terminus[19,20]. As a result, we investigated the function of the final 138 amino acids at IdrD's C-terminus, now named RdnE for recognition DNA nuclease effector. We measured bacterial growth using a strain derived from BB2000 that has disruptions in its native *idrD* and downstream genes[18]. These *P. mirabilis* cells had 1000 fewer cells per mL when engineered to overproduce RdnE *in trans* than the negative control containing no RdnE protein (Figure 1A). An equivalent growth pattern occurred in *Escherichia coli* cells under the same conditions (Supplemental Figure 1). Thus, RdnE was lethal *in vivo*.

RdnE's initial 86 amino acids contain a PD-(D/E)XK motif, which is suggestive of nucleotide degradation. The PD-(D/E)XK superfamily includes proteins with broad functions, including effectors that degrade DNA or RNA[21–23]. Three residues in the catalytic site—D, D/E, and K—are required for activity[24]. Therefore, we changed the

corresponding residues in RdnE (D39, E53, and K55) to alanine, separately and together. *P. mirabilis* producing these mutant proteins showed growth equivalent to the negative control lacking RdnE (Figure 1A). We also noticed that *E. coli* cells that were producing RdnE had morphologies that were indicative of DNA damage or stress. These cells were elongated, and the DAPI-stained DNA was distributed irregularly within the cells (Supplemental Figure 1). Cells producing a defunct D39A mutant ($RdnE_{D39A}$) did not have this appearance (Supplemental Figure 1). Therefore, the PD-(D/E)XK motif was essential for growth arrest.

The importance of the PD-(D/E)XK motif for activity suggested that RdnE was a nuclease, but defining its molecular target required direct analysis. We synthesized RdnE with a C-terminus FLAG epitope tag using *in vitro* translation because of its lethality (Supplemental Figure 1). We measured the protein concentrations, added phage lambda DNA (methylated or unmethylated) to progressively higher RdnE protein concentrations, and then performed agarose gel electrophoresis analysis. Degradation of lambda DNA occurred in the presence of RdnE, regardless of the DNA methylation state (Figure 1B). The $RdnE_{D39A}$ construct caused a slight reduction in lambda DNA, while the negative control showed no DNA loss (Figure 1B). RdnE also caused a reduction in plasmid DNA, indicating its endonuclease activity (Supplemental Figure 1). These results revealed that RdnE caused DNA degradation *in vitro* in a PD-(D/E)XK-dependent manner.

RdnE likely has two different domains, as there is a region directly following the PD-(D/E)XK motif. A two-domain architecture is similar to that described for DNases[25,26] and other T6SS effectors like some PoNe-containing DNases[21]. We made independent deletions of each potential RdnE domain to assess whether both were essential for

DNase activity. One construct deleted the first alpha helix without disturbing the catalytic residues; the other deleted the region after the PD-(D/E)XK motif, which we termed "region 2." The resulting proteins, produced via *in vitro* translation, were assayed for DNase activity as described above. The truncated proteins resulted in no loss of lambda DNA (Figure 1C), indicating that both domains are necessary for degradation activity.

We next asked whether RdnE homologs also act as DNA nucleases. A bioinformatics search revealed the closest RdnE homolog (outside of *Proteus*) was found in the Actinobacteria, *Rothia dentocariosa* C6B. *Rothia* species are inhabitants of the normal oral flora, dwelling in biofilms within the human oral cavity and pharynx[27]. The two RdnE proteins (^{Proteus}RdnE and ^{Rothia}RdnE) share approximately 55% amino acid sequence identity, mostly within the PD-(D/E)XK domain, and have similar predicted secondary structures (Figure 1D). Given this, we hypothesized that ^{Rothia}RdnE also acted as a DNA nuclease.

We analyzed ^{Rothia}RdnE for PD-(D/E)XK-dependent DNA nuclease activity. We produced it and a predicted null mutant, ^{Rothia}RdnE$_{D39A}$, using *in vitro* translation. Samples containing the ^{Rothia}RdnE$_{D39A}$ protein or a negative control had similar DNA levels (Figure 1E). By contrast, samples with the wild-type ^{Rothia}RdnE protein showed a loss of lambda DNA regardless of methylation state, indicating that ^{Rothia}RdnE also had DNA nuclease activity (Figure 1E). To further analyze RdnE's architecture, we exchanged region 2 between the ^{Proteus}RdnE and ^{Rothia}RdnE sequences and assayed for nuclease activity. The hybrid proteins degraded lambda DNA, unlike the negative control (Figure 1F). While DNA degradation by these trans-phyla hybrid proteins required more protein than the wild-type enzymes, the cross-phyla protein fragments could complement one another.

These data indicated that these RdnE proteins form a novel PD-(D/E)XK-containing DNA nuclease subfamily.

RdnI binds and neutralizes RdnE.

Effectors have cognate immunity proteins, often adjacently located on the chromosome. We hypothesized that the gene adjacent to *rdnE* in BB2000*, rdnI* (previously named "*idrE"*), encodes the cognate immunity protein (Figure 2A). RdnI did not have defined domains, and its function was unknown. We assessed RdnI's activity using microscopic and cell growth analysis. Swarming *P. mirabilis* cells are normally elongated with DAPI-stained DNA found along the cell body (Figure 2B). By contrast, swarming cells producing RdnE *in trans* did not elongate, had a reduced DAPI signal, and had an accumulation of misshapen cells (Figure 2B). Cell shape and DNA-associated fluorescence levels returned to normal when cells concurrently produced the RdnE and RdnI proteins (Figure 2B). RdnI production also rescued cell growth in *E. coli* cells producing RdnE (Supplemental Figure 2). These data suggest that RdnI inhibits RdnE's lethality.

To determine whether RdnI provided protection against injected RdnE within mixed communities, we used swarm competition assays, which combine one-to-one mixtures of *P. mirabilis* strains and then measure dominance[18]. The control strain was wild-type strain BB2000 (herein called "BB2000"), which naturally produced RdnE and RdnI. The other was strain ATCC29906, which did not naturally produce RdnE and RdnI. These two strains formed a visible boundary between swarming monoculture colonies (Figure 2C). The mixed-strain colony merged with BB2000 in one-to-one competitions, demonstrating

BB2000's dominance (Figure 2C). A similar outcome was seen when ATCC29906 was producing Green Fluorescent Protein (GFPmut2). However, BB2000 did not outcompete ATCC29906 engineered to produce RdnI with a C-terminal StrepII epitope tag ("RdnI-StrepII"), seen by a merging of the mixed-strain colony with ATCC29906 (Figure 2C). Thus, RdnI protected cells against injected RdnE within mixed communities.

Based on the prevailing model, we predicted that a cognate EI pair should bind to one another, which we evaluated *in vivo* and *in vitro.* We used the defunct mutant ($RdnE_{D39A}$-FLAG) for these assays because producing the wild-type RdnE protein kills cells. For *in vivo* analysis, we used bacterial two-hybrid assays (BACTH) in which the reconstitution of the T18 and T25 fragments of adenylate cyclase results in the colorimetric change to blue in the presence of the substrate, X-gal[28,29]. Constructed vectors contained genes for $RdnE_{D39A}$-FLAG, RdnI-StrepII, or GFPmut2 on the C-termini of the T18 or the T25 fragments. When the reporter strain produced $RdnE_{D39A}$-FLAG or RdnI-StrepII with GFPmut2, the resultant yellow color was equivalent to when X-gal was absent (Figure 2D). There was also minimal color change when an individual protein was produced on both fragments (Figure 2D). However, the reporter strains made blue colonies when X-gal was present and the cells concurrently produced $RdnE_{D39A}$-FLAG and RdnI-StrepII. These results indicated that RdnE and RdnI bind to each other *in vivo*.

We used batch *in vitro* co-immunoprecipitation assays to confirm the *in vivo* binding result. Separate *E. coli strains* produced either $RdnE_{D39A}$-FLAG or had a negative control, exogenous FLAG-BAP (*E. coli* bacterial alkaline phosphatase with a FLAG epitope tag) added to cell lysate. An anti-FLAG western blot showed both FLAG-BAP (~ 50 kDa) and $RdnE_{D39A}$-FLAG (~ 17 kDa) in the soluble and elution fractions. RdnI-StrepII

eluted with $RdnE_{D39A}$-FLAG but not the negative control (Figure 2E). The western blot results corresponded with the Coomassie blue-stained gels (Supplemental Figure 2). Overall, our data showed that *Proteus* RdnE and RdnI form a cognate EI pair. However, questions about their prevalence among bacteria remained, because RdnE and RdnI matched no known effector or immunity families.

Expansion of the RdnE and RdnI protein families revealed similar gene architecture and secondary structures.

Gene neighborhood analysis can guide homology inference and protein comparisons. We conducted consecutive searches with BLAST[30] and HMMER[31] to identify sequences that encoded proteins with high similarity to RdnE and RdnI (Supplemental Figure 3). The final list contained 21 EI pairs from a variety of phyla (Table 3). Although the genes surrounding these putative EI pairs differed, many shared mobile and transport-associated elements, such as recombinant hotspot (Rhs*)* sequences or other similar peptide-repeat sequences (Figure 3A). Several gene neighborhoods had transport-associated genes, such as the T6SS-associated *vgrG/tssC* gene and the CDI-associated *cdiB* gene. A few also included putative immunity proteins from other reported families, like immunity protein 44 (Pfam15571) in *Taylorella asinigenitalis* MCE3 and immunity protein 51 (Pfam15595) in *Chryseobacterium populi* CF314. Notably, these organisms varied widely in origin and residence. Some were from the soil rhizosphere (*Pseudomonas ogarae* and *C. populi*) and others from the human microbiome (*P. mirabilis*, *R. dentocariosa,* and *Prevotella jejuni*) (Figure 3A). The phylogenetic tree

showed distribution across bacterial phyla (Figure 3B), perhaps reflecting these proteins' role in community interactions.

We next constructed maximum likelihood trees to examine the relationship between the identified proteins. Both the RdnE and RdnI trees diverged from the species tree (Figure 3B) and each other (Figure 3C). Proteins from evolutionarily distant bacteria appeared to share more similarities compared to those from more closely related genera. These results are consistent with potential horizontal gene transfer seen for other EI pairs[7].

Yet, despite the differences in phyla and gene neighborhoods, the RdnE and RdnI proteins displayed the defining characteristics of protein families. The predicted secondary structures of the RdnE proteins were similar even though the amino acid sequences differed (Figure 3D, Supplemental Figure 3). The RdnE proteins showed two distinct domains, as with the *Proteus* and *Rothia* results (Figure 1): the PD-(D/E)XK region followed by a sequence variable region (region 2). Further, the predicted tertiary structures were consistent with PD-(D/E)XK folds (three β-sheets flanked by two α-helices, α/β/α) found in other proteins (Figure 3E, Supplemental Figure 4)[24,32]. These findings suggested that domains in RdnE are conserved across diverse phyla and the sequences seed a distinct PD-(D/E)XK subfamily.

While immunity proteins within a family have diverse overall amino acid sequences, conserved secondary structures and some conserved residues are not uncommon among immunity protein families. Indeed, they are often used to characterize these families[33]. We found that, while the RdnI proteins shared minimal primary amino acid sequence identity, they were predicted to contain several α-helices (Figure 3D,

Supplemental Figure 3) and had similar AlphaFold2-predicted tertiary structures (Figure 3E, Supplemental Figure 4). We also discovered a region with three alpha-helices and several conserved residues, which we named the "conserved motif" (Figure 3D). Interestingly, the addition of this conserved motif indicates that the RdnI architecture mirrors that of the two-domain RdnE proteins (Figure 3D). Therefore, the RdnI conserved motif may also indicate a novel immunity domain and seed a protein family.

Binding flexibility in RdnI allows for cross-species protection.

We deployed a structure-function approach to determine the conserved motif's role in the *Proteus*RdnI's activity. Bioinformatics using AlphaFold2[32] and Consurf[34] revealed seven highly conserved residues within this region; four of these (Y197, H221, P244, E246) clustered together within the AlphaFold2 structure and are identical between sequences (Figures 4A). In a sequence-optimized (SO) RdnI, we individually changed each of these four to alanine and discovered that each alanine-substituted variant behaved like the wildtype and inhibited RdnE activity (Supplemental Figure 5). We then replaced all seven residues (Y197, S235, K258, and the original four) with alanine (*Proteus*RdnI$_{7mut}$-StrepII) and found that, unlike wild-type *Proteus*RdnI-StrepII, this construct was not protective in swarm competition assays (Figure 4B). However, the *Proteus*RdnI$_{7mut}$-StrepII mutant still bound RdnE$_{D39A}$-FLAG in bacterial two-hybrid assays (Figure 4C). Therefore, the seven residues in the conserved motif are critical for RdnI's neutralizing function but dispensable for binding RdnE.

Given that the conserved motif and nearby regions are likely involved in function, we queried for potential functions in the remainder of the RdnI protein. We engineered

variants that were either (1) the first 85 amino acids, (2) amino acids 150 to 305, which contained an intact conserved motif, or (3) amino acids 235 to 305, which contained the last alpha helix of the conserved motif (Supplemental Figure 3). None of these constructs protected against RdnE's lethality *in vivo* (Figure 4D), demonstrating that the entire protein is likely essential for function. The variant containing the first 85 amino acids of ^{Proteus}RdnI was the only construct to bind ^{Proteus}RdnE, indicating that the N-terminal region is sufficient for binding between this *P. mirabilis* EI pair (Figure 4E). Thus, binding is necessary but insufficient for neutralization, and the neutralization activity likely resides within the second half of RdnI. These results, as well as the functional domain swaps between the *Proteus* and *Rothia* RdnE proteins (Figure 1F), contradicted the prevalent model's assertion that one-to-one binding is both necessary and sufficient for immunity protein function against a cognate effector.

Therefore, we explored the relationship between non-cognate RdnE and RdnI proteins from various phyla. We first asked whether non-cognate RdnI immunity proteins could protect against injected ^{Proteus}RdnE. Using the swarm competition assays, we competed BB2000 against ATCC29906 engineered to produce RdnI homologs from *R. dentocariosa*, *P. jejuni*, or *P. ogarae* (^{Rothia}RdnI-StrepII, ^{Prevotella}RdnI-StrepII, and ^{Pseudomonas}RdnI-StrepII, respectively). BB2000 dominated the swarm when ATCC29906 produced GFPmut2, ^{Prevotella}RdnI-StrepII, or ^{Pseudomonas}RdnI-StrepII (Figure 4F). However, ATCC29906 outcompeted BB2000 when making ^{Proteus}RdnI-StrepII or ^{Rothia}RdnI-StrepII (Figure 4F). Consistent with these findings, the RdnI immunity proteins from *Proteus* and *Rothia* consistently bound ^{Proteus}RdnE$_{D39A}$ in *in vivo* and *in vitro* assays, but the *Prevotella* and *Pseudomonas* variants did not (Figure 4G, Supplemental Figure 7). Thus, ^{Rothia}RdnI

bound and protected against ^{Proteus}RdnE, demonstrating that cross-protection between non-cognate EI pairs from different phyla is possible, provides a fitness benefit during competition, and influences community structure.

Given that the entire immunity protein is necessary for protection, we next evaluated which region(s) of RdnI allowed for cross-protection. We first moved the conserved motif of the three foreign RdnI homologs into ^{Proteus}RdnI-StrepII and measured neutralizing activity using swarm competition assays and binding activity using BACTH. The conserved motifs from *Rothia* and *Prevotella* were sufficient to preserve ^{Proteus}RdnI's neutralizing (Figure 4H) and binding functions (Figure 4I). However, the conserved motif from *Pseudomonas* was insufficient to neutralize ^{Proteus}RdnE (Figure 4H), even though the construct could still bind ^{Proteus}RdnE$_{D39A}$ (Figure 4I). We then moved the *Proteus* conserved motif into the RdnI variants from *Prevotella* and *Pseudomonas*, but the motif was insufficient to confer neutralizing (Figure 4H) or binding function (Figure 4I). These results showed that binding RdnE is necessary but not sufficient for RdnI protection. Further, within RdnI, the conserved motif is necessary for neutralization.

<u>RdnE and RdnI proteins from diverse phyla are present in individual human microbiomes.</u>

Our findings indicate that immunity proteins such as RdnI could provide a broader protective umbrella for a cell beyond inhibiting the effector proteins of their siblings. If so, one would expect to find evidence of RdnE and RdnI homologs from different phyla in the same environment or microbial community. We tested this hypothesis by analyzing about 500,000 publicly available microbiomes (metagenomes) for the specific *rdnE* and *rdnI* genes examined in Figure 4 (Figure 5A). 2,296 human and terrestrial metagenomes

contained reads matching with over 90% identity to these *rdnE* sequences (Figure 5B). The reads mapped to the expected niche for each organism, underscoring the presence of these specific effector proteins in naturally occurring human-associated microbiomes.

Further, the *rdnE* and *rdnI* genes from various human-associated bacteria occurred concurrently in individual human oral and, to a lesser extent, gut metagenomes. The *rdnE* and *rdnI* genes from *Rothia* and *Prevotella* co-occurred in approximately 5% of the metagenomes analyzed (Figure 5C). Stringent detection parameters were utilized, so the true number could be higher. We then compared the abundance of *rdnI to rdnE*, since metagenomic coverage (i.e., the number of short reads that map to a gene) approximates the underlying gene's abundance in the sampled community. In most gut samples, *rdnI* recruited more reads than *rdnE*, although there was substantial variance (Figure 5D). These data could indicate the presence of orphan *rdnI* genes, which is consistent with published T6SS orphan immunity alleles[35,36]. These metagenomic patterns suggest that a single community can produce multiple RdnE and RdnI proteins from different phyla to defend against each other, further highlighting the intricate nature of bacterial defense in crowded environments.

Discussion

Using these results as a foundation, we propose an alternate model of "EI skeleton keys," in which flexible ("promiscuous") binding between immunity and effector proteins enables protection in mixed-species communities (Figure 5E). This model expands the current paradigm of (selective) cognate EI partners[37] and has far-reaching implications for immunity protein evolution and molecular function. Unlike most known immunity

proteins that fit within the prevailing model, we demonstrated that flexible EI binding is necessary but not sufficient to neutralize RdnE. We also showed that both RdnE and RdnI proteins have multiple domains, one conserved with functional activity and the other variable across phyla. The combined findings point to a two-step mechanism for RdnI protection: the N-terminal variable-sequence domain mediates binding to an effector, while the C-terminal conserved domain is required for neutralization in a function that has yet to be determined (Figure 5E).

The RdnE and RdnI protein families have parallel structural organizations. RdnE's conserved region contains the enzymatic PD-(D/E)XK motif, which is necessary for DNA nuclease activity. Its variable region comprises the second half of the protein, which is also necessary for degradation activity. Despite amino acid variability among variants, these domains can be swapped between phyla and retain function. On the other hand, RdnI's variable and conserved regions likely determine binding and function, respectively. *Proteus*RdnI's first 85 amino acids are sufficient and necessary for binding *Proteus*RdnE, while the conserved motif, found in the last two-thirds of RdnI, is necessary for function. RdnI's protective ability can be preserved when the conserved motif is swapped between species, even among those that cannot otherwise bind. The interchangeability between domains in RdnE and RdnI variants could reflect constraints on their evolutionary paths.

Compared to the prevailing model, our proposed skeleton-key model better explains RdnI's conserved motif, orphan immunity genes, and other non-cognate protein interactions. Sequence diversity among EI pairs is often attributed to the coevolution between effector and immunity proteins, which is thought to maintain specificity and binding[33]. However, this coevolution model does not fully explain immunity proteins'

conserved regions or the presence and protective activity of orphan immunity proteins[38,39]. A recent study with the Tri1 immunity protein shows that it has conserved enzymatic activity, allowing for protection against foreign effectors, but there is no cross-binding between non-cognate EI pairs[40]. This data suggests that conserved enzymatic activity can be maintained alongside the coevolution necessary for strict cognate pair binding. Therefore, we propose that EI protein domains may evolve independently instead of coevolution across the full protein. For example, RdnE's PD-(D/E)XK motif and RdnI's conserved motif might need to be maintained for activity, while the variable domains may diversify in sequence more freely. Depending on the selective pressures, the variable regions could reinforce specificity between cognate EI pairs as they coevolve, like the Tri1 family. Additional evolutionary analysis would reveal how the balance between specificity and flexibility evolves in EI pairs, both within domains and across the entire protein.

Flexible protection likely applies beyond contact-dependent EI pairs. Indeed, this model is conceptually similar to the cross-binding seen in type II antitoxins and bacteriocins. Within the ParD antitoxin family, 45% of the twenty antitoxins tested show cross-protective activity with at least one non-cognate ParE toxin[41]. For the E group of colicins, non-cognate immunity proteins show a lower binding affinity than for their cognate immunity proteins, but they still offer cross-protection[13,42]. Similarly, immunity proteins from *Carnobacterium piscicola* and *Enterococcus faecium* offer cross-protection against the other's cognate bacteriocin[43]. In these systems, as with RdnE-RdnI, flexible binding between a harmful protein and a non-cognate neutralizing protein can result in a potential fitness advantage for the bacteria.

When considering the impacts on community structure, the broadened activity of RdnI proteins against RdnE effectors from multiple phyla likely increases bacterial fitness, which is advantageous in dense environments. RdnI production increased cells' fitness and modified the community in experimental setups like the swarm competition assays as it enabled vulnerable bacteria to inhabit previously restricted spaces. Supporting this experimental data, both gut and oral metagenomes showed evidence of multiple *rdnE-rdnI* pairs within individual samples, particularly between *Rothia* and *Prevotella*. Interestingly, the oral microbiome had roughly equivalent abundance between the effector and immunity genes, which might reflect that bacteria occupy distinct spatiotemporal niches within oral microbiomes, e.g. *R. dentocariosa* is predominantly on tooth surfaces[44]. By contrast, *rdnI* genes had greater abundance compared to *rdnE* in the gut microbiomes, which may reflect the greater diversity in member species and community structures found in the gut[45]. Orphan immunity genes are indeed a known phenomenon in T6SS EI literature but are usually documented through single isolate sequencing. This community level assessment affirms the presence of *rdnI* orphan genes on a population scale and points to relatively widespread immunity genes in hundreds or thousands of samples.

Given the ability of immunity genes to protect against non-cognate effectors, the presence of diverse orphan *rdnI* genes hints at the ecological complexity surrounding RdnE and RdnI. This community of immunity proteins is reminiscent of the model for shared immunity proteins within an ecosystem, also called a "hyper-immunity state," which was seen among colicins in wild field mice[46]. In this hyper-immunity state, a set of immunity proteins shared among a community could offer an advantage against pathogens. Invading bacteria would be unable to protect themselves from certain

effectors while the community would be protected as they share the immunity proteins necessary for defense. Flexible binding like RdnI could contribute to such a “hyper-immunity state” to help a bacterial community maintain its niche.

Indeed, bacteria have a diverse set of protective measures to ward off foreign effectors in addition to flexible immunity proteins. Recent work has identified non-specific mechanisms of protection including stress-response, physical barriers, and a stronger offense[11]. Orphan immunity genes also exist throughout many bacterial genomes and may be a part of this system[38], for example, these genes offer a fitness advantage in mouse microbiomes[39]. Our data extends the current repertoire of protection mechanisms by adding another tool: a flexible immunity protein collection, where each immunity protein acts as a skeleton key against a wider class of effectors. This collection could be useful in dense, diverse communities where contact-dependent competition using EI pairs is critical to maintain one’s population. Taken together, the physical interactions between, and evolution of, effector and immunity proteins remain a rich area for new explorations.

Acknowledgements

We thank Caroline Boyd, Neils Bradshaw, Emma Keteku, Alecia Septer, Nora Sullivan, Adnan Syed, and Larissa Wenren for contributing experimental materials to this project. Rachelle Gaudet, Colleen Cavanaugh, and members of the Gibbs Lab provided valued advice on the manuscript. The David and Lucile Packard Foundation, the George W. Merck Fund, Harvard University, and the University of California, Berkeley and its

Rausser College of Natural Resources funded our research. A.K., D.S., D.U., and K.A.G designed and performed research as well as analyzed data. A.K., D.S., D.U., and K.A.G wrote the paper. We have no competing interests to declare.

Materials and Methods

Bacterial strains and media

All strains are described in Table 1. Strains for bacterial two-hybrid assays were transformed fresh the day before. Overnight cultures were grown aerobically at 37°C in LB (Lennox) broth[47]. All *E. coli* strains were plated on LB (Lennox) agar surfaces (1.5% Bacto agar) and all *P. mirabilis* strains were plated on LSW^{-} agar[47] or 25 mL CM55 media (blood agar base agar [Oxoid, Basingstoke, England]). When necessary, antibiotics were used in the media at the following concentrations: 35 µg/mL kanamycin, 100 µg/mL carbenicillin.

Plasmid construction

Plasmids were constructed according to Table 2. Primers and gBlocks were ordered from Integrated DNA Technologies (IDT), Coralville, IA. P_{idrA}-RdnE was constructed by Polymerase Chain Reaction (PCR) amplifying the last 416 bp of the *idrD* gene from BB2000 and cloning it into the SacI and AgeI sites of the pBBR1-NheI vector, resulting in plasmid pAS1054. RdnE is the final 138 amino acids of IdrD (out of its total of 1581). P_{idrA}-*rdnE-rdnI* was constructed by PCR amplifying the last 416 bp of the *idrD* gene through the end of the *rdnI* gene from BB2000, resulting in the plasmid pAS1059. The gBlock and primer sequences will be archived on an OSF website (https://osf.io/scb7z/) and made publicly available upon publication.

We used several standard protocols for vector construction. Seamless ligation cloning extract (SLiCE) was adapted from Zhang et al[48]. Restriction-digest reactions were based on manufacturer's protocols. Overlap extension (SOE) PCR Amplification was

adapted from Heckman 2007[49]. Plasmids were transformed into OmniMax *E. coli* and confirmed using Sanger Sequencing (UC Berkeley DNA Sequencing Facility and Genewiz, South Plainfield NJ).

In vitro DNase assay

RdnE proteins were produced using the New England Biolabs PURExpress *In Vitro* Protein Synthesis Kit (New England BioLabs Inc., Ipswich MA). Template DNA contained the *rdnE* gene and required elements specified by the PURExpress kit. We adapted this protocol from prior *in vitro* DNA-degradation assays[50]. Reactions were performed with 250 ng of template DNA (no template DNA added to negative control reaction) and incubated at 37°C for two hours. Protein amount was determined using an anti-FLAG western blot with a known gradient of FLAG-BAP (2.5, 5, 10, and 20 ng). Synthesized protein (2.5, 5, and 10 ng) was added to 0.5 µg of lambda DNA (methylated and unmethylated), 5 µL of New England Biolabs Buffer 3.1, and up to a final volume of 25 µL. For plasmid DNase assays, 10 ng of synthesized protein was added to 250 ng of circular or linear plasmid DNA ($pids_{BB}$[51]). This reaction was incubated for one hour at 37°C, then Proteinase K (New England Biolabs Inc., Ipswich MA) was added and incubated for an additional 15 minutes at 37°C. Reaction was then run on a 1% agarose gel for analysis.

E. coli liquid growth and viability assays

Overnight cultures were grown at 37°C in a shaking incubator in LB broth with appropriate antibiotics. Cultures were normalized to an optical density at 595 nm (OD_{595})

of 1 and diluted 1:100 into LB broth with 35 µg/mL kanamycin, with and without 200 nM anhydrotetracycline (aTc). Some samples were analyzed for OD_{595} every thirty minutes for 16 hours in a 96-well plate using a TECAN. Other samples were incubated at 37°C for 6 hours while rocking. At indicated time points, 100 µL of sample was removed, diluted, and then plated on fresh LB agar plates to measure colony forming units per mL (CFU) after overnight growth at 37°C using standard protocols.

Microscopy

We performed microscopy on *P. mirabilis* strain *idrD*::Tn5 (Cm^R) (also called, *idrD**), which has a transposon insertion to disrupt *rdnE* and *rdnI* expression[18], carrying either vector pBBR1-NheI or pDS0002 (producing RdnE) and on *E. coli* carrying either pBBR1-NheI, pDS0002, or pDS0048 (producing $RdnE_{D39A}$). *P. mirabilis* cells were normalized to OD_{595} of 0.1 after overnight growth in LB broth supplemented with kanamycin. Cells were inoculated onto CM55 swarm pads containing 10 µg/mL DAPI and 10 nM aTc and grown in humidified chambers at 37°C. Images were taken at five and six hours after growth. From overnight cultures, *E. coli* cells were grown in LB broth plus kanamycin until mid-logarithmic phase and then mounted directly onto glass slides. Glass coverslips were sealed with nail polish. For all microscopy, we captured phase contrast and DAPI (150 ms exposure) images using a Leica DM5500B microscope (Leica Microsystems, Buffalo Grove IL) and CoolSnap HQ CCD camera (Photometrics, Tucson AZ) cooled to -20°C. MetaMorph version 7.8.0.0 (Molecular Devices, Sunnyvale CA) was used for image acquisition.

Sequence Optimized RdnI

The *P. mirabilis rdnI-StrepII* sequence was difficult to genetically engineer into due to its low GC% content (23%). As such, we engineered the sequence to have a higher GC%, called "Sequence optimized (SO) ^{Proteus}RdnI-StrepII" but no change to the amino acid sequence. The change to the nucleotide sequence did not affect the construct's ability to offer protection to a vulnerable strain (Figure 3G).

Swarm competition assay

The swarm competition (territoriality) assay was adapted from Wenren et. al. 2013[18]. 5 mL cultures were grown in LB broth with appropriate antibiotics overnight in a 37°C rocking incubator. Overnight cultures were normalized to an OD_{595} of 1. For the competition samples, the strains were mixed 1:1. 2 µL of each sample were inoculated onto CM55 agar with the appropriate antibiotic. Plates were incubated at 37°C for 22 hours and then photographed and assessed for boundary formation.

BACTH assay

The vectors are described in Battesti and Bouveret 2012[28] with an added linker region between the T25 or T18 fragments and multiple cloning sites. BTH101 cultures were grown at 30°C overnight in LB broth with kanamycin and carbenicillin. 10 µL of the overnight culture were inoculated onto LB agar with kanamycin, carbenicillin, 1 mM IPTG, and 0 or 40 µg/mL of X-gal (Thermo Fisher, Waltham MA), and grown at 30°C for 24 hours. Color was amplified by an additional 24 hours at 4°C, and then samples were imaged.

FLAG co-immunoprecipitation assays

The protocol was adapted from Cardarelli et. al. 2015[2]. *E. coli* cells were harvested from LB broth, grown for either 3 hours after induction with 200 nM aTc at 37°C or 16 to 20 hours at 16°C after induction with 1 mM IPTG. Cells were then pelleted by centrifugation and flash frozen in liquid nitrogen. RdnE-containing samples were lysed in 50 mM Tris pH 7.4, 150 mM NaCl, and 1x Protease Inhibitor Cocktail (Selleck Chemicals LLC, Houston TX), via bead bashing for 20 minutes at 4°C. RdnI-containing samples were lysed in 100 mM Tris-HCl pH 8, 180 mM NaCl, and 1x Protease Inhibitor Cocktail via 10x 10 second sonication pulses. The soluble fraction for both samples was obtained after centrifugation at 15,000 rpm for 15 minutes. FLAG epitope containing samples were incubated with prepared resin for two hours at 4°C. The resin was then washed twice (50 mM Tris pH 7.4, 150 mM NaCl, and 1% Tween-20), incubated with approximately 1 mL of the soluble fraction of the RdnI-StrepII-containing samples for another 2 hours, and washed thrice more. The protein was finally eluted with 50 µL of 300 ng/µL 3x FLAG peptide (Sigma-Aldrich, St. Louis, MO) for 45 minutes at 4°C. Sample buffer (63 mM Tris pH 6.8, 2% Sodium Dodecyl Sulfate, 10% glycerol, 5% 2-Mercaptoethanol) was added to samples, boiled at 95°C for 10 minutes, and frozen at -80°C.

SDS-Page and Western blotting

The protocol was adapted from Cardarelli et. al. 2015[2]. Protein samples were separated by gel electrophoresis using 13% Tris-Tricine polyacrylamide gels and either transferred to a 0.45-µm nitrocellulose membrane (Bio-Rad Laboratories, Hercules CA) or stained with Coomassie blue (Bio-Rad Laboratories, Hercules CA). Western blot

membranes were probed with primary antibody (either 1:4000 rabbit anti-FLAG [Sigma Aldrich, St. Louis MO] or 1:4000 mouse anti-StrepII [Genscript, Piscataway NJ]) for 1 hour at room temperature or overnight at 4°C and with secondary antibody (either 1:5000 goat anti-rabbit or anti-mouse respectively conjugated to horseradish peroxidase (HRP) [KPL, Inc., Gaithersburg MD]) for 30 minutes at room temperature. Samples were finally visualized using Immun-Star HRP substrate kit (Bio-Rad Laboratories, Hercules CA) and the Chemidoc XRS system (Bio-Rad Laboratories, Hercules, CA). TIFF files were analyzed on Fiji (ImageJ, Madison, WI).

Bioinformatics search for RdnE and RdnI homologs

A BLAST[30] search of the *P. mirabilis* RdnE protein sequence revealed seven RdnE homologs from a variety of species. The downstream genes of these RdnE homologs were identified using the DOE Joint Genome Institute (JGI) Integrated Microbial Genomes and Microbiomes (IMG/M). The seven RdnE and RdnI amino acid sequences were separately aligned with MUSCLE using Jalview[52]. These alignments were then used as seeds for a second homology search using HMMERsearch (HmmerWeb version 2.41.2) and the Ensembl Database[53]. The two data sets were then compared for genomes that contained both *rdnE* and *rdnI* genes next to one another within their respective genomes. Any EI pairs that contained disrupted PD-(D/E)XK motifs within their RdnE sequence were removed.

Gene neighborhood and primary conservation analyses

Gene neighborhoods were obtained using JGI's IMG/M Neighborhood viewer and then redrawn using Adobe Illustrator (Adobe Inc., 2022). Locations of predicted functions are approximate primarily based on the Pfam domain calling by IMG/M.

The final 21 RdnE and RdnI sequences were aligned with MUSCLE using Jalview[52]. The conserved residues were identified using Jalview and the cartoons were created using Adobe Illustrator (Adobe Inc., 2022). The sequence logo for the RdnI conserved motif was generated with WebLogo[54] and constrained to only visualize the conserved motif.

Secondary and tertiary structure predictions

Secondary structure predictions of the MUSCLE aligned sequences were determined with Ali2D from the MPI Bioinformatics toolkit[55,56]. The resulting predictions were made into cartoons manually using Adobe Illustrator (Adobe Inc., 2022). Tertiary structure predictions were done with AlphaFold2 using MMseqs2 on Google Colab[32]. Query protein sequences were inputted into the program and then run producing 5 models ranked 1-5. Rank 1 models are shown. pIDDT scores indicate confidence levels for each amino acid position. Structures were analyzed in PyMOL (The PyMOL Molecular Graphics System, Version 2.2.3 Schrödinger, LLC.). The pIDDT graphs are in the supplemental data.

Metagenomic analyses

A sourmash-based approach was used to screen approximately 500,000 public metagenomes stored on NCBI's SRA (*https://github.com/sourmash-bio/2022-search-sra-with-mastiff*) for the presence of the ten genomes shown in Figure 3A. Hits with a containment score greater than 0.2 were downloaded for further analysis, representing 9,137 metagenomes. Each metagenome was then mapped with bbmap[57] against the 10 *rdnE* and *rdnI* sequences with a stringency of 90% (minid = 0.9), along with quality filtering (trim1 = 20, minaveragequality = 10). After mapping, metagenomes were retained if they had (1) a mean coverage greater than 2X, (2) at least one base covered greater than 5X, and (3) more than half of the bases on reference *rdnE-rdnI* sequence receiving coverage. 2,857 metagenomes met these criteria, of which 2,296 contained *P. mirabilis*, *R. dentocariosa*, *P. jejuni*, or *P. ogarae* sequences and could be confidently assigned to samples obtained from the human gut or oral microbiome or from terrestrial sources. Gene-level coverage in a sample was then summarized as each gene's average nucleotide's coverage. The ratio of *rdnI* to *rdnE* coverage was then calculated for each sample and log10-transformed, and the distribution of ratios was summarized with Python's seaborn kdeplot using a bandwidth of 0.4[58].

Open Science statement

The sequence files and associated data will be archived on an OSF website (https://osf.io/scb7z/) and made publicly available upon publication.

Tables

Table 1: List of strains used in this study.

Strain	Strain Name	Description	Reference or Source
P. mirabilis BB2000 *idrD** + pDS0062	DS349	BB2000 *idrD*::Tn5 (Cm^R) producing GFPmut2 under the aTc-inducible promoter	This study
P. mirabilis BB2000 *idrD** + pDS0002	DS104	BB2000 *idrD*::Tn5 (Cm^R) producing RdnE under the aTc-inducible promoter	This study
P. mirabilis BB2000 *idrD** + pDS0058	DS344	BB2000 *idrD*::Tn5 (Cm^R) producing $RdnE_{D39A}$ under the aTc-inducible promoter	This study
P. mirabilis BB2000 *idrD** + pDS0059	DS345	BB2000 *idrD*::Tn5 (Cm^R) producing $RdnE_{E53A}$ under the aTc-inducible promoter	This study
P. mirabilis BB2000 *idrD** + pDS0060	DS346	BB2000 *idrD*::Tn5 (Cm^R) producing $RdnE_{K55A}$ under the aTc-inducible promoter	This study
P. mirabilis BB2000 *idrD** + pDS0061	DS347	BB2000 *idrD*::Tn5 (Cm^R) producing $RdnE_{D39A}$ E53A K55A under the aTc-inducible promoter	This study
E. coli MG1655 + pBBR1-NheI	DS068	MG1655 carrying empty vector	This study
E. coli MG1655 + pDS0002	DS151	MG1655 producing RdnE under the aTc-inducible promoter	This study
E. coli MG1655 + pDS0058	DS336	MG1655 producing $RdnE_{D39A}$ under the aTc-inducible promoter	This study
E. coli MG1655 + pDS0059	DS337	MG1655 producing $RdnE_{E53A}$ under the aTc-inducible promoter	This study
E. coli MG1655 + pDS0060	DS338	MG1655 producing $RdnE_{K55A}$ under the aTc-inducible promoter	This study
E. coli MG1655 + pDS0061	DS339	MG1655 producing $RdnE_{D39A\ E53A\ K55A}$ under the aTc-inducible promoter	This study
P. mirabilis BB2000 *idrD** + pDS0003	DS092	BB2000 *idrD*::Tn5 (Cm^R) co-producing RdnE followed by RdnI under the aTc-inducible promoter	This study
E. coli MG1655 + pDS0003	DS170	MG1655 co-producing RdnE followed by RdnI under the aTc-inducible promoter	This study

P. mirabilis BB2000 + pBBR1-NheI	ANS1127	BB2000 carrying empty vector	[18]
P. mirabilis ATCC29906 + pBBR1-NheI	ANS1280	ATCC29906 carrying empty vector	This study
P. mirabilis ATCC29906 + pAK043	AK0132	ATCC29906 producing RdnI with a C-terminal Strep-II tag with the *fla* promoter	This study
P. mirabilis ATCC29906 + pLMW04-gfp	AK387	ATCC29906 producing GFPmut2 with the *fla* promoter	This study
E. coli MG1655 + pDS0048	DS248	MG1655 producing $RdnE_{D39A}$ with a C-terminal FLAG tag under an aTc-inducible promoter	This study
E. coli BL21(plysS)DE3 + pAK023	AK024	BL21(plysS)DE3 producing RdnI with a C-terminal Strep-II tag under the T7 promoter	This study
P. mirabilis ATCC29906 + pAK044	AK0135	ATCC29906 producing the *Rothia*RdnI with a C-terminal Strep-II tag under the *fla* promoter	This study
P. mirabilis ATCC29906 + pAK045	AK0138	ATCC29906 producing the *Prevotella*RdnI with a C-terminal Strep-II tag under the *fla* promoter	This study
P. mirabilis ATCC29906 + pAK046	AK0141	ATCC29906 producing the *Pseudomonas*RdnI with a C-terminal Strep-II tag under the *fla* promoter	This study
E. coli BL21(plysS)DE3 + pAK058	AK261	BL21(plysS)DE3 producing *Rothia*RdnI with a C-terminal Strep-II tag under the T7 promoter	This study
E. coli BL21(plysS)DE3 + pAK059	AK262	BL21(plysS)DE3 producing *Prevotella*RdnI with a C-terminal Strep-II tag under the T7 promoter	This study
E. coli BL21(plysS)DE3 + pAK060	AK263	BL21(plysS)DE3 producing *Pseudomonas*RdnI with a C-terminal Strep-II tag under the T7 promoter	This study
P. mirabilis ATCC29906 + pAK063	AK318	ATCC29906 producing the recoded *Proteus*RdnI sequence with a C-terminal Strep-II tag under the *fla* promoter	This study
P. mirabilis ATCC29906 + pAK065	AK320	ATCC29906 producing the *Rothia*RdnI sequence (aa195-271) inserted between aa192-266 in the recoded *Proteus*RdnI with a C-terminal Strep-II tag under the *fla* promoter	This study

P. mirabilis ATCC29906 + pAK066	AK321	ATCC29906 producing the *Prevotella*RdnI sequence (aa170-245) inserted between aa192-266 in the recoded *Proteus*RdnI with a C-terminal Strep-II tag under the *fla* promoter	This study
P. mirabilis ATCC29906 + pAK067	AK322	ATCC29906 producing the *Pseudomonas*RdnI sequence (aa181-255) inserted between aa192-266 in the recoded *Proteus*RdnI with a C-terminal Strep-II tag under the *fla* promoter	This study
P. mirabilis ATCC29906 + pAK086	AK381	ATCC29906 producing the recoded *Proteus*RdnI sequence (aa192-266) inserted between aa170-245 in the *Prevotella*RdnI with a C-terminal Strep-II tag under the *fla* promoter	This study
P. mirabilis ATCC29906 + pAK087	AK382	ATCC29906 producing the recoded *Proteus*RdnI sequence (aa192-266) inserted between aa181-255 in the *Pseudomonas*RdnI with a C-terminal Strep-II tag under the *fla* promoter	This study
P. mirabilis ATCC29906 + pAK064	AK319	ATCC producing the recoded *Proteus*RdnI sequence with seven alanine mutations (Y197A, H221A, S235A, P244A, E246A, R254A, K258A) and a C-terminal Strep-II tag under the *fla* promoter	This study
OneShot OmniMax 2 T1R Competent Cells		*E. coli* strain for cloning	Thermo Fisher Scientific, Waltham, MA
MFDpir		Mu-free *E. coli* mating strain to introduce plasmids into *P. mirabilis*	[59]

Table 2: Plasmids used in this study.

Plasmid Name	Description	Cloning Method or Source
pBBR1-NheI	empty vector with pBBR1 origin.	[51]
pLMW04-gfp	GFPmut2 with a constitutive *fla* promoter, (pBBR1 origin, Kan resistance).	[18]
pDS0002	*rdnE* with the anhydrotetracycline (aTc)-inducible promoter, P_{tet}, (pBBR1 origin, Kan resistance)	gDS0005 was recombined into amplified pAS1054 by SLiCE
pDS0062	*gfpmut2* with the aTc-inducible promoter, (pBBR1 origin, Kan resistance)	restriction digest using amplified *gfpmut2* and pDS0002
pDS0048	$rdnE_{D39A}$-FLAG with the aTc-inducible promoter, (pBBR1 origin, Kan resistance)	gDS0025 was recombined into pDS0002 using restriction digest
pDS0058	$rdnE_{D39A}$ with the aTc-inducible promoter, (pBBR1 origin, Kan resistance)	pDS0048 was recombined into pDS0002 using restriction digest
pDS0059	$rdnE_{E53A}$ with the aTc-inducible promoter, (pBBR1 origin, Kan resistance)	gDS0026 was recombined into pDS0002 using restriction digest
pDS0060	$rdnE_{K55A}$ with the aTc-inducible promoter, (pBBR1 origin, Kan resistance)	gDS0027 was recombined into pDS0002 using restriction digest
pDS0061	$rdnE_{D39A,\ E53A,\ K55A}$ with the aTc-inducible promoter, (pBBR1 origin, Kan resistance)	gDS0028 was recombined into pDS0002 using restriction digest
pDS0034	*rdnE*-FLAG with the aTc-inducible promoter, (pBBR1 origin, Kan resistance)	gDS0023 was recombined into amplified pDS0002 using SLiCE
pDS0003	*rdnE-rdnI* with the aTc-inducible promoter, (pBBR1 origin, Kan resistance)	Amplified *rdnE-rdnI* from pAS1059 was recombined into pDS0002 using SOE PCR
pAK023	*rdnI*-StrepII with the T7 promoter, (pUC, Amp resistance)	gAK001 and pDS0003 were recombined into pET17b vector using SOE PCR
pAK043	*rdnI*-StrepII with the *fla* promoter, (pBBR1 origin, Kan resistance)	*rdnI*-StrepII amplified from pAK023 was recombined into pLMW04 using restriction digest
pAK070	T25-linker-$rdnE_{D39A}$-FLAG with the IPTG-inducible *lac* promoter, (p15A, Kan resistance)	$rdnE_{D39A}$-FLAG tag amplified from pDS0048 was recombined into pKT25 using restriction digest

pAK071	T25-linker-*rdnI* with the *lac* promoter, (p15A, Kan resistance)	*rdnI*-Strep-II amplified from pAK043 was recombined into pKT25 using restriction digest
pAK074	T18-linker-*rdnE*$_{D39A}$-FLAG with the *lac* promoter, (Col E1 origin, Amp resistance)	*rdnE*$_{D39A}$-FLAG tag amplified from pDS0048 was recombined into pUT18C using restriction digest
pAK075	T18-linker-*rdnI*-StrepII with the *lac* promoter, (Col E1 origin, Amp resistance)	*rdnI*-Strep-II tag amplified from pAK043 was recombined into pUT18C using restriction digest
pAK076	T25-*gfpmut2* with the *lac* promoter, (p15A origin, Kan resistance)	Amplified *gfpmut2* was recombined into pKT25 using restriction digest
pAK077	T18-*gfpmut2* with the *lac* promoter, (Col E1 origin, Amp resistance)	Amplified *gfpmut2* was recombined into pUT18C using restriction digest
pAK044	^{Rothia}*rdnI*-StrepII with the *fla* promoter, (pBBR1 origin, Kan resistance)	gAK003 was recombined into pAK043 using restriction digest
pAK045	^{Prevotella}*rdnI*-StrepII with the *fla* promoter, (pBBR1 origin, Kan resistance)	gAK004 was recombined into pAK043 using restriction digest
pAK046	^{Pseduomonas}*rdnI*-StrepII with the *fla* promoter, (pBBR1 origin, Kan resistance)	gAK005 was recombined into pAK043 using restriction digest
pAK079	T18-^{Rothia}*rdnI*-StrepII with the *lac* promoter, (pUC, Amp resistance)	gAK003 was recombined into pUT18C using restriction digest
pAK081	T18-^{Prevotella}*rdnI*-StrepII with the *lac* promoter, (pUC, Amp resistance)	gAK004 was recombined into pUT18C using restriction digest
pAK083	T18-^{Pseudomonas}*rdnI*-StrepII with the *lac* promoter, (pUC, Amp resistance)	gAK005 was recombined into pUT18C using restriction digest
pAK058	^{Rothia}*rdnI*-StrepII with the T7 promoter, (pUC, Amp resistance)	gAK003 was recombined into pAK023 using restriction digest
pAK059	^{Prevotella}*rdnI*-StrepII with the T7 promoter, (pUC, Amp resistance)	gAK004 was recombined into pAK023 using restriction digest
pAK060	^{Pseudomonas}*rdnI*-StrepII with the T7 promoter, (pUC, Amp resistance)	gAK005 was recombined into pAK023 using restriction digest
pAK063	Sequence optimized *rdnI*-StrepII with the *fla* promoter, (pBBR1 origin, Kan resistance)	gAK024 was recombined into pAK043 using restriction digest
pAK065	$^{Pm/Rd}$*rdnI*-StrepII (*Proteus rdnI* with *Rothia* conserved motif insert) with the *fla* promoter, (pBBR1 origin, Kan resistance)	gAK026 was recombined into pAK043 using restriction digest
pAK066	$^{Pm/Pj}$*rdnI*-StrepII (*Proteus rdnI* with *Prevotella* conserved motif insert) with the	gAK028 was recombined into pAK043 using restriction digest

	fla promoter, (pBBR1 origin, Kan resistance)	
pAK067	$^{Pm/Po}$*rdnI*-StrepII (*Proteus rdnI* with *Pseudomonas* conserved motif insert) with the *fla* promoter, (pBBR1 origin, Kan resistance)	gAK027 was recombined into pAK043 using restriction digest
pAK086	$^{Pj/Pm}$*rdnI*-StrepII (*Prevotella rdnI* with *Proteus* conserved motif insert) with the *fla* promoter, (pBBR1 origin, Kan resistance)	gAK029 was recombined into pAK043 using restriction digest
pAK087	$^{Pf/Pm}$*rdnI*-StrepII (*Pseudomonas rdnI* with *Proteus* conserved motif insert) with the *fla* promoter, (pBBR1 origin, Kan resistance)	gAK030 was recombined into pAK043 using restriction digest
pAK092	T18-linker-Sequence optimized ^{Proteus}*rdnI* with a C-terminal StrepII with the *lac* promoter (Col E1 origin, Amp resistance)	gAK024 was recombined into pAK075 using restriction digest
pAK093	T18-linker-Sequence optimized $^{Pm/Ra}$*rdnI* (*Proteus rdnI* with *Rothia* conserved motif insert) with a C-terminal StrepII with the *lac* promoter (Col E1 origin, Amp resistance)	gAK026 was recombined into pAK075 using restriction digest
pAK094	T18-linker-Sequence optimized $^{Pm/Pj}$*rdnI* (*Proteus rdnI* with *Prevotella* conserved motif insert) with the *lac* promoter (Col E1 origin, Amp resistance)	gAK028 was recombined into pAK075 using restriction digest
pAK095	T18-linker-Sequence optimized $^{Pm/Pf}$*rdnI* (*Proteus rdnI* with *Pseudomonas* conserved motif insert) with the *lac* promoter (Col E1 origin, Amp resistance)	gAK027 was recombined into pAK075 using restriction digest
pAK096	T18-linker- $^{Pj/Pm}$*rdnI* (*Prevotella rdnI* with *Proteus* conserved motif insert) with the *lac* promoter (Col E1 origin, Amp resistance)	gAK029 was recombined into pAK075 using restriction digest
pAK097	T18-linker-$^{Pf/Pm}$*rdnI* (*Pseudomonas rdnI* with *Proteus* conserved motif insert) with a C-terminal StrepII with the *lac* promoter (Col E1 origin, Amp resistance)	gAK030 was recombined into pAK075 using restriction digest
pAK064	Sequence optimized *rdnI*$_{7mut}$ -StrepII with the *fla* promoter, (pBBR1 origin, Kan resistance)	gAK025 was recombined into pAK043 using restriction digest
pAK085	T18-linker-Sequence optimized *rdnI*$_{7mut}$-StrepII with the *lac* promoter (Col E1 origin, Amp resistance)	gAK025 was recombined into pAK075 using restriction digest

Table 3: RdnE and RdnI homolog species.

Genus Species Strain	Phylum	Isolation Location	RdnE JGI unique ID	RdnI JGI unique ID
Acinetobacter baumannii BJAB0715	Proteobacteria	Fresh water	2562302616	2562302617
Acinetobacter baumannii XH858	Proteobacteria	Human sputum	2686809281	2686809282
Burkholderia sp. TSV86	Proteobacteria	water	2766166119	2766166118
Cellulophaga baltica 18	Bacteroidota	water	2815949879	2815949878
Chryseobacterium indologenes NBRC 14944	Bacteroidota	Human trachea	2565567985	2565567984
Chryseobacterium populi CF314	Bacteroidota	Soil rhizosphere	2511231970	2511231971
Chryseobacterium sp. IHB B 17019	Bacteroidota	Soil undefined subtype	2686963654	2686963655
Cronobacter turicensis 564	Proteobacteria	Human undefined subtype	2532469359	2532469360
Cronobacter turicensis z3032	Proteobacteria	Human blood culture	646327905	646327906
Endozoicomonas numazuensis DSM 25634	Proteobacteria	Marine sponge	2574519540	2574519541
Paenibacillus elgii M63	Firmicutes	Hot spring	2744846532	2744846531
Paenibacillus sp. Aloe-11	Firmicutes	Soil rhizosphere	2549870597	2549870598
Prevotella jejuni CD3:33	Bacteroidota	human intestine biopsy	2804797915	2804797916
Prevotella sp. C561	Bacteroidota	Human respiratory tract	2514485316	2514485315
Prevotella sp. F0108	Bacteroidota	Human oral cavity	647936965	647936966
Proteus mirabilis BB2000	Proteobacteria	Human intestine biopsy	2546214711	2546214712
Pseudomonas ogarae F113	Proteobacteria	Soil rhizosphere	2511826458	2511826457

Pseudomonas syringae pv. coriandricola ICMP 12471	Proteobacteria	undefined	2714543877	2714543878
Rothia dentocariosa C6B	Actinobacteriota	Human oral cavity	2611822673	2611822674
Tannerella forsythia ATCC 43037	Bacteroidota	Human oral cavity	2512371376	2512371377
Taylorella asinigenitalis MCE3	Proteobacteria	Mammal reproductive system	2511725471	2511725470

References

1. Kuzmanov, U., and Emili, A. (2013). Protein-protein interaction networks: Probing disease mechanisms using model systems. Genome Medicine *5*, 37. 10.1186/gm441.

2. Cardarelli, L., Saak, C., and Gibbs, K.A. (2015). Two proteins form a heteromeric bacterial self-recognition complex in which variable subdomains determine allele-restricted binding. mBio *6*, e00251-15. 10.1128/mBio.00251-15.

3. Burton, D.R., Stanfield, R.L., and Wilson, I.A. (2005). Antibody vs. HIV in a clash of evolutionary titans. Proc. Natl. Acad. Sci. U.S.A. *102*, 14943–14948. 10.1073/pnas.0505126102.

4. Schreiber, G., and Keating, A.E. (2011). Protein binding specificity versus promiscuity. Curr Opin Struct Biol *21*, 50–61. 10.1016/j.sbi.2010.10.002.

5. Speare, L., Cecere, A.G., Guckes, K.R., Smith, S., Wollenberg, M.S., Mandel, M.J., Miyashiro, T., and Septer, A.N. (2018). Bacterial symbionts use a type VI secretion system to eliminate competitors in their natural host. PNAS *115*, E8528–E8537. 10.1073/pnas.1808302115.

6. Xiong, L., Cooper, R., and Tsimring, L.S. (2018). Coexistence and pattern formation in bacterial mixtures with contact-dependent killing. Biophysical Journal *114*, 1741–1750. 10.1016/j.bpj.2018.02.012.

7. Ruhe, Z.C., Low, D.A., and Hayes, C.S. (2020). Polymorphic toxins and their immunity proteins: diversity, evolution and mechanisms of delivery. Annu Rev Microbiol *74*, 497–520. 10.1146/annurev-micro-020518-115638.

8. Sgro, G.G., Oka, G.U., Souza, D.P., Cenens, W., Bayer-Santos, E., Matsuyama, B.Y., Bueno, N.F., dos Santos, T.R., Alvarez-Martinez, C.E., Salinas, R.K., et al. (2019). Bacteria-killing type IV secretion systems. Frontiers in Microbiology *10*.

9. Russell, A.B., Peterson, S.B., and Mougous, J.D. (2014). Type VI secretion system effectors: poisons with a purpose. Nature Reviews Microbiology *12*, 137–148. 10.1038/nrmicro3185.

10. Jurėnas, D., Fraikin, N., Goormaghtigh, F., and Van Melderen, L. (2022). Biology and evolution of bacterial toxin–antitoxin systems. Nat Rev Microbiol, 1–16. 10.1038/s41579-021-00661-1.

11. Hersch, S.J., Manera, K., and Dong, T.G. (2020). Defending against the type six secretion system: beyond immunity genes. Cell Reports *33*, 108259. 10.1016/j.celrep.2020.108259.

12. Benz, J., Sendlmeier, C., Barends, T.R.M., and Meinhart, A. (2012). Structural insights into the effector – immunity system Tse1/Tsi1 from *Pseudomonas aeruginosa*. PLOS ONE *7*, e40453. 10.1371/journal.pone.0040453.

13. Wallis, R., Leung, K.-Y., Pommer, A.J., Videler, H., Moore, G.R., James, R., and Kleanthous, C. (1995). Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 2. Cognate and noncognate interactions that span the millilmolar to femptomolar affinity range. Biochemistry *34*, 13751–13759. 10.1021/bi00042a005.

14. Alteri, C.J., Himpsl, S.D., Zhu, K., Hershey, H.L., Musili, N., Miller, J.E., and Mobley, H.L.T. (2017). Subtle variation within conserved effector operon gene products contributes to T6SS-mediated killing and immunity. PLoS Pathog *13*. 10.1371/journal.ppat.1006729.

15. Levin, K.B., Dym, O., Albeck, S., Magdassi, S., Keeble, A.H., Kleanthous, C., and Tawfik, D.S. (2009). Following evolutionary paths to protein-protein interactions with high affinity and selectivity. Nat Struct Mol Biol *16*, 1049–1055. 10.1038/nsmb.1670.

16. Zhang, H., Gao, Z.-Q., Wei, Y., Xu, J.-H., and Dong, Y.-H. (2013). Insights into the cross-immunity mechanism within effector families of bacteria type VI secretion system from the structure of StTae4-EcTai4 complex. PLOS ONE *8*, e73782. 10.1371/journal.pone.0073782.

17. Schaffer, J.N., and Pearson, M.M. (2015). *Proteus mirabilis* and urinary tract infections. Microbiology Spectrum *3*, 10.1128/microbiolspec.uti-0017–2013. 10.1128/microbiolspec.uti-0017-2013.

18. Wenren, L.M., Sullivan, N.L., Cardarelli, L., Septer, A.N., and Gibbs, K.A. (2013). Two independent pathways for self-recognition in *Proteus mirabilis* are linked by type VI-dependent export. mBio *4*, e00374-13. 10.1128/mBio.00374-13.

19. Koskiniemi, S., Lamoureux, J.G., Nikolakakis, K.C., t'Kint de Roodenbeke, C., Kaplan, M.D., Low, D.A., and Hayes, C.S. (2013). Rhs proteins from diverse bacteria mediate intercellular competition. Proc Natl Acad Sci U S A *110*, 7032–7037. 10.1073/pnas.1300627110.

20. Ma, J., Sun, M., Dong, W., Pan, Z., Lu, C., and Yao, H. (2017). PAAR-Rhs proteins harbor various C-terminal toxins to diversify the antibacterial pathways of type VI secretion systems. Environmental Microbiology *19*, 345–360. 10.1111/1462-2920.13621.

21. Jana, B., Fridman, C.M., Bosis, E., and Salomon, D. (2019). A modular effector with a DNase domain and a marker for T6SS substrates. Nat Commun *10*, 1–12. 10.1038/s41467-019-11546-6.

22. Kosinski, J., Feder, M., and Bujnicki, J.M. (2005). The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism

and functional predictions for domains of (hitherto) unknown function. BMC Bioinformatics *6*, 172. 10.1186/1471-2105-6-172.

23. Yadav, S.K., Magotra, A., Ghosh, S., Krishnan, A., Pradhan, A., Kumar, R., Das, J., Sharma, M., and Jha, G. (2021). Immunity proteins of dual nuclease T6SS effectors function as transcriptional repressors. EMBO reports *22*, e51857. 10.15252/embr.202051857.

24. Steczkiewicz, K., Muszewska, A., Knizewski, L., Rychlewski, L., and Ginalski, K. (2012). Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. Nucleic Acids Res *40*, 7016–7045. 10.1093/nar/gks382.

25. Lowey, B., Whiteley, A.T., Keszei, A.F.A., Morehouse, B.R., Mathews, I.T., Antine, S.P., Cabrera, V.J., Kashin, D., Niemann, P., Jain, M., et al. (2020). CBASS immunity uses CARF-related effectors to sense 3′–5′- and 2′–5′-linked cyclic oligonucleotide signals and protect bacteria from phage infection. Cell *182*, 38-49.e17. 10.1016/j.cell.2020.05.019.

26. Schiltz, C.J., Lee, A., Partlow, E.A., Hosford, C.J., and Chappie, J.S. (2019). Structural characterization of Class 2 OLD family nucleases supports a two-metal catalysis mechanism for cleavage. Nucleic Acids Research *47*, 9448–9463. 10.1093/nar/gkz703.

27. Wilbert, S.A., Mark Welch, J.L., and Borisy, G.G. (2020). Spatial ecology of the human tongue dorsum microbiome. Cell Reports *30*, 4003-4015.e3. 10.1016/j.celrep.2020.02.097.

28. Battesti, A., and Bouveret, E. (2012). The bacterial two-hybrid system based on adenylate cyclase reconstitution in *Escherichia coli*. Methods *58*, 325–334. 10.1016/j.ymeth.2012.07.018.

29. Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proceedings of the National Academy of Sciences *95*, 5752–5756. 10.1073/pnas.95.10.5752.

30. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of Molecular Biology *215*, 403–410. 10.1016/S0022-2836(05)80360-2.

31. Eddy, S.R. (2009). A new generation of homology search tools based on probabilistic inference. Genome Inform *23*, 205–211.

32. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat Methods *19*, 679–682. 10.1038/s41592-022-01488-1.

33. Zhang, D., de Souza, R.F., Anantharaman, V., Iyer, L.M., and Aravind, L. (2012). Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. Biology Direct *7*, 18. 10.1186/1745-6150-7-18.

34. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Research *44*, W344–W350. 10.1093/nar/gkw408.

35. Kirchberger, P.C., Unterweger, D., Provenzano, D., Pukatzki, S., and Boucher, Y. (2017). Sequential displacement of Type VI Secretion System effector genes leads to evolution of diverse immunity gene arrays in *Vibrio cholerae*. Scientific Reports *7*, 45133. 10.1038/srep45133.

36. Koskiniemi, S., Garza-Sánchez, F., Sandegren, L., Webb, J.S., Braaten, B.A., Poole, S.J., Andersson, D.I., Hayes, C.S., and Low, D.A. (2014). Selection of orphan Rhs toxin expression in evolved *Salmonella enterica* serovar Typhimurium. PLoS Genet *10*. 10.1371/journal.pgen.1004255.

37. Jurėnas, D., and Journet, L. (2021). Activity, delivery, and diversity of Type VI secretion effectors. Molecular Microbiology *115*, 383–394. 10.1111/mmi.14648.

38. Barretto, L.A.F., and Fowler, C.C. (2020). Identification of a putative T6SS immunity islet in *Salmonella* Typhi. Pathogens *9*, 559. 10.3390/pathogens9070559.

39. Ross, B.D., Verster, A.J., Radey, M.C., Schmidtke, D.T., Pope, C.E., Hoffman, L.R., Hajjar, A.M., Peterson, S.B., Borenstein, E., and Mougous, J.D. (2019). Human gut bacteria contain acquired interbacterial defense systems. Nature *575*, 224–228. 10.1038/s41586-019-1708-z.

40. Ting, S.-Y., Bosch, D.E., Mangiameli, S.M., Radey, M.C., Huang, S., Park, Y.-J., Kelly, K.A., Filip, S.K., Goo, Y.A., Eng, J.K., et al. (2018). Bifunctional immunity proteins protect bacteria against FtsZ-targeting ADP-ribosylating toxins. Cell *175*, 1380-1392.e14. 10.1016/j.cell.2018.09.037.

41. Aakre, C.D., Herrou, J., Phung, T.N., Perchuk, B.S., Crosson, S., and Laub, M.T. (2015). Evolving new protein-protein interaction specificity through promiscuous intermediates. Cell *163*, 594–606. 10.1016/j.cell.2015.09.055.

42. Li, W., Keeble, A.H., Giffard, C., James, R., Moore, G.R., and Kleanthous, C. (2004). Highly discriminating protein–protein interaction specificities in the context of a conserved binding energy hotspot. Journal of Molecular Biology *337*, 743–759. 10.1016/j.jmb.2004.02.005.

43. Franz, C.M.A.P., van Belkum, M.J., Worobo, R.W., Vederas, J.C., and Stiles, M.E.Y. 2000 (2000). Characterization of the genetic locus responsible for production and immunity of carnobacteriocin A: the immunity gene confers cross-protection to

enterocin BThe GenBank/EMBL/DDBJ accession number for the sequence reported in this paper is AF207838. Microbiology *146*, 621–631. 10.1099/00221287-146-3-621.

44. Welch, J.L.M., Dewhirst, F.E., and Borisy, G.G. (2019). Biogeography of the oral microbiome: the site-specialist hypothesis. Annual Review of Microbiology *73*, null. 10.1146/annurev-micro-090817-062503.

45. Donaldson, G.P., Lee, S.M., and Mazmanian, S.K. (2016). Gut biogeography of the bacterial microbiota. Nat Rev Microbiol *14*, 20–32. 10.1038/nrmicro3552.

46. Riley, M.A., and Wertz, J.E. (2002). Bacteriocins: evolution, ecology, and application. Annual Review of Microbiology *56*, 117–137. 10.1146/annurev.micro.56.012302.161024.

47. Belas, R., Erskine, D., and Flaherty, D. (1991). Transposon mutagenesis in *Proteus mirabilis*. Journal of Bacteriology *173*, 6289–6293. 10.1128/jb.173.19.6289-6293.1991.

48. Zhang, Y., Werling, U., and Edelmann, W. (2014). Seamless ligation cloning extract (SLiCE) cloning method. Methods Mol Biol *1116*, 235–244. 10.1007/978-1-62703-764-8_16.

49. Heckman, K.L., and Pease, L.R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. Nat Protoc *2*, 924–932. 10.1038/nprot.2007.132.

50. Hughes, F.M.Jr., and Cidlowski, J.A. (1997). Utilization of an *in vitro* assay to evaluate chromatin degradation by candidate apoptotic nucleases. Cell Death Differ *4*, 200–208. 10.1038/sj.cdd.4400221.

51. Gibbs, K.A., Urbanowski, M.L., and Greenberg, E.P. (2008). Genetic determinants of self identity and social recognition in bacteria. Science *321*, 256–259. 10.1126/science.1160033.

52. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191. 10.1093/bioinformatics/btp033.

53. Cunningham, F., Allen, J.E., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Bennett, R., et al. (2022). Ensembl 2022. Nucleic Acids Research *50*, D988–D995. 10.1093/nar/gkab1049.

54. Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A Sequence Logo Generator. 3.

55. Gabler, F., Nam, S.-Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., Lupas, A.N., and Alva, V. (2020). Protein sequence analysis using the MPI bioinformatics toolkit. Current Protocols in Bioinformatics *72*, e108. 10.1002/cpbi.108.

56. Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N., and Alva, V. (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol *430*, 2237–2243. 10.1016/j.jmb.2017.12.007.

57. Bushnell, B. (2014). BBMap: A fast, accurate, splice-aware aligner (Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States)).

58. Waskom, M.L. (2021). seaborn: statistical data visualization. Journal of Open Source Software *6*, 3021. 10.21105/joss.03021.

59. Ferrières, L., Hémery, G., Nham, T., Guérout, A.-M., Mazel, D., Beloin, C., and Ghigo, J.-M. (2010). Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. Journal of Bacteriology *192*, 6418–6427. 10.1128/jb.00621-10.

60. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics *35*, 4453–4455. 10.1093/bioinformatics/btz305.

Figures

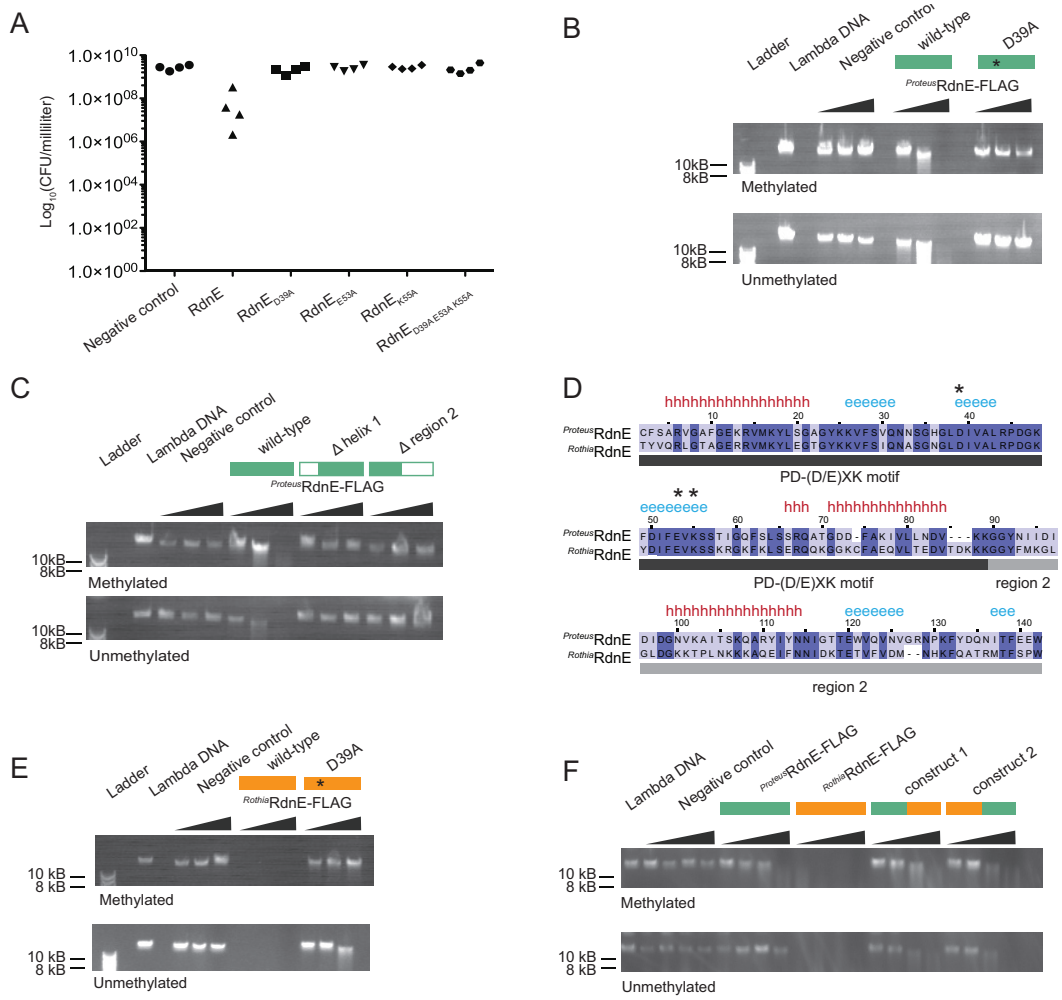

Figure 1. RdnE homologs act as DNA endonucleases and contain interchangeable domains. A) Cell viability (colony forming units per mL) after protein production in swarms of *P. mirabilis* strain *idrD**, which does not produce RdnE and RdnI. Cells produced GFPmut2, RdnE, or mutant variants in the predicted PD-(D/E)XK motif: D39A, E53A, K55A, or all. **B)** *In vitro* DNA degradation assay for ^{Proteus}RdnE. Increasing concentrations of a negative control, ^{Proteus}RdnE-FLAG, or ^{Proteus}RdnE$_{D39A}$-FLAG were incubated with methylated or unmethylated lambda DNA (48,502 bp) and analyzed by gel electrophoresis. Plasmid DNA degradation is in Supplemental Figure 1. **C)** *In vitro* DNA degradation assay for domain deletions of ^{Proteus}RdnE. The first construct removed the first alpha helix without disturbing the catalytic residues, and the second construct contained the PD-(D/E)XK motif and removed region 2. Increasing concentrations were analyzed as in **(B)**. **D)** Multiple sequence alignment between *P. mirabilis* and *R. dentocariosa* RdnE sequences. Black bar show the PD-(D/E)XK motif and the grey bar marks the variable region 2 domain. Conserved residues are highlighted in dark blue. Secondary structure predictions identified using Ali2D[55,56] (h for alpha helix, e for beta sheet); the catalytic residues (stars) are noted above the alignment. **(E,F)** *In vitro* DNA degradation assay and analysis as in **(B)**. **(E)** Increasing concentrations of either a negative control, ^{Rothia}RdnE-FLAG, or ^{Rothia}RdnE$_{D39A}$-FLAG. **(F)** The PD-(D/E)XK motifs were swapped between the ^{Rothia}RdnE (orange) and the ^{Proteus}RdnE (green) sequences and compared to the wild-type RdnE proteins.

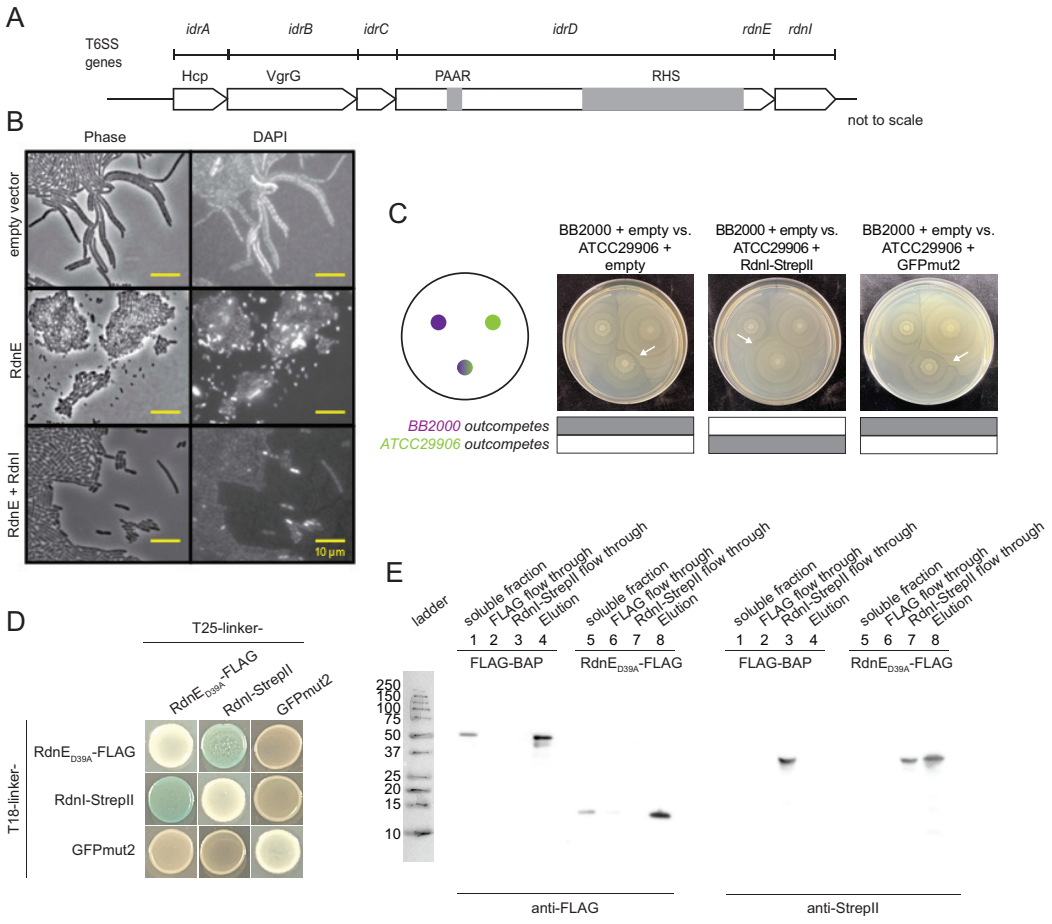

Figure 2. RdnI binds to and protects against RdnE *in vivo* and *in vitro*. A) Domain architecture for the *idr* locus in *P. mirabilis* strain BB2000[18]. At the top are genes with Pfam domains listed below them. Gray boxes denote PAAR and Rhs domains in the N-terminal region of the full-length IdrD protein. **B)** Micrographs of *P. mirabilis* strain *idrD** cells carrying an empty vector, a vector for producing RdnE, or a vector for producing RdnE and RdnI. DNA was visualized by DAPI stain. Phase, left; fluorescence, right. **C)** Swarm competition assay[18] of wild-type *P. mirabilis* strain BB2000 (donor) competed against the vulnerable target, which is *P. mirabilis* strain ATCC29906 carrying an empty vector, a vector for producing RdnI-StrepII, or a vector for producing GFP. Left, schematic of swarm competition assay where top left colony is BB2000, top right colony is ATCC29906 with its vector cargo, and bottom colony is a 1:1 mixture of BB2000 and ATCC29906 with its vector cargo. Gray boxes underneath indicate whether BB2000 (top) or ATCC29906 (bottom) dominated in the 1:1 mixture and white arrows point to the boundary line that forms between different strains. **D)** Bacterial two-hybrid (BACTH) assay with RdnE$_{D39A}$-FLAG, RdnI-StrepII, and GFPmut2. The colorimetric change was discerned in the presence of the substrate Xgal and inducer IPTG. **E)** An anti-FLAG batch co-immunoprecipitation of RdnE$_{D39A}$-FLAG and RdnI-StrepII. RdnE$_{D39A}$-FLAG or exogenous FLAG-BAP (soluble fraction) was incubated with anti-FLAG resin (FLAG flow through). RdnI-StrepII was then added to the resin (RdnI-StrepII flow through). Any proteins bound to resin were eluted with FLAG-peptide (Elution) and analyzed by anti-FLAG and anti-StrepII western blots.

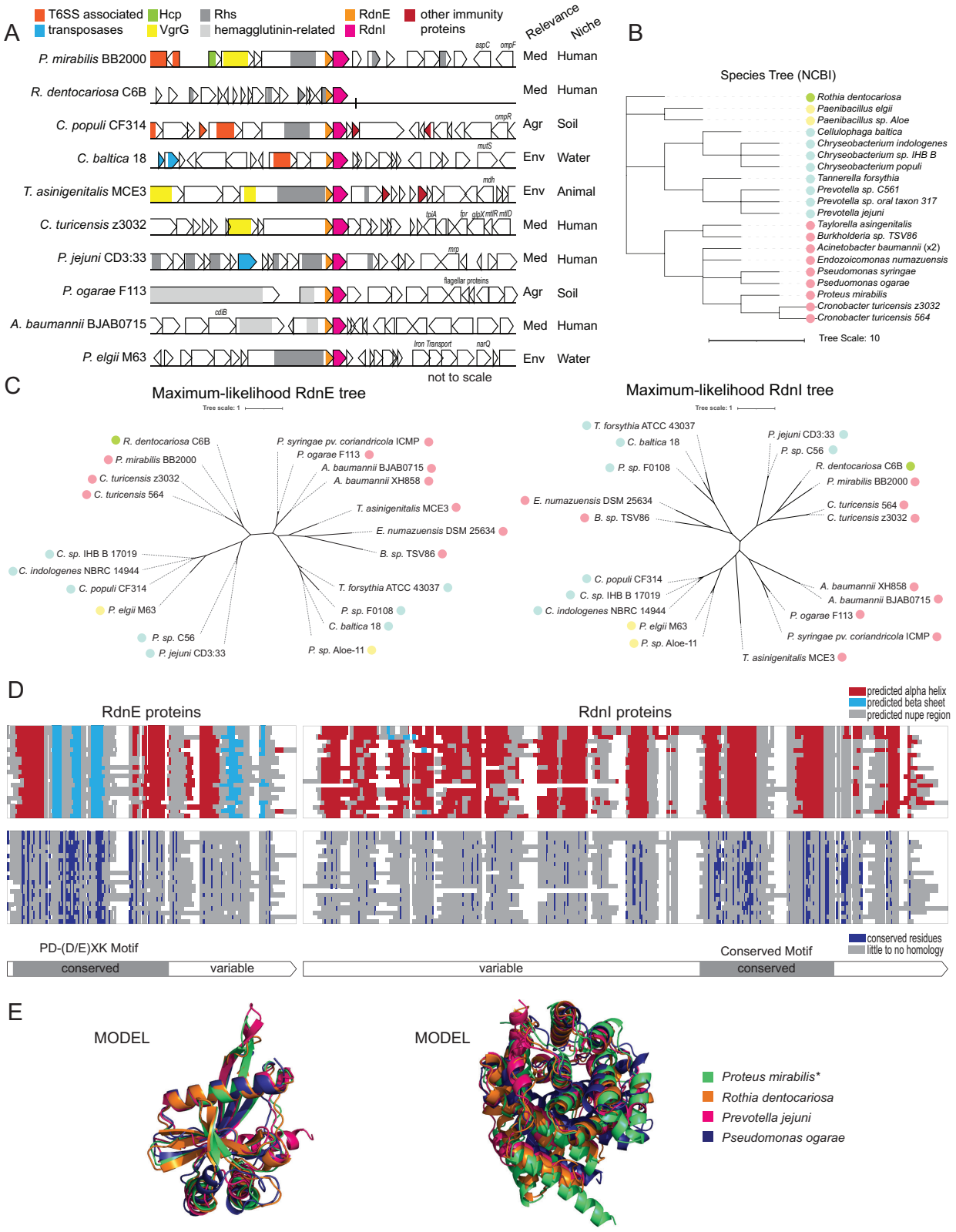

Figure 3. RdnE and RdnI protein families share conserved residues and predicted structures. A) Gene neighborhoods for RdnE and RdnI homologs. Listed are gene neighborhoods, relevance, and niche,

which we identified using IMG/M from the Joint Genomics Institute. Colors highlight conserved function/genes (not to scale). (Agr: Agriculture, Med: Medical, Env: Environmental), and the site of isolation. **B)** Phylogenetic tree based on NCBI taxonomy. Scale is located below the graph. The colored circles represent phyla (green: Actinobacteriota; yellow: Firmicutes; blue: Bacteroidota; pink: Proteobacteria). **C)** Unrooted maximum likelihood trees of the RdnE (left) and RdnI (right) homologs. Trees were created with RaxML[60], and the scale is annotated below. The colored circles represent phyla (same as in **B**). **D)** Protein alignments overlaid with either predicted secondary structures (top) or conserved residues (bottom) of the RdnE and RdnI homologs. MUSCLE alignments[52] are highlighted by secondary structures (red: alpha helices, light blue: beta sheets), or conserved residues (dark blue). White represents gaps in the protein alignment. The bars below mark the predicted conserved and variable domains. **E)** Alignments of AlphaFold2 predictions for RdnE and RdnI sequences from *P. mirabilis* (green), *R. dentocariosa* (orange), *P. jejuni* (magenta), and *P. ogarae* (dark blue). Structures were generated using ColabFold[32] and aligned using PyMol. The *P. mirabilis* RdnI sequence is a natural variant with two residue substitutions (V101E and I222M) as compared to BB2000 (supplemental figure 6).

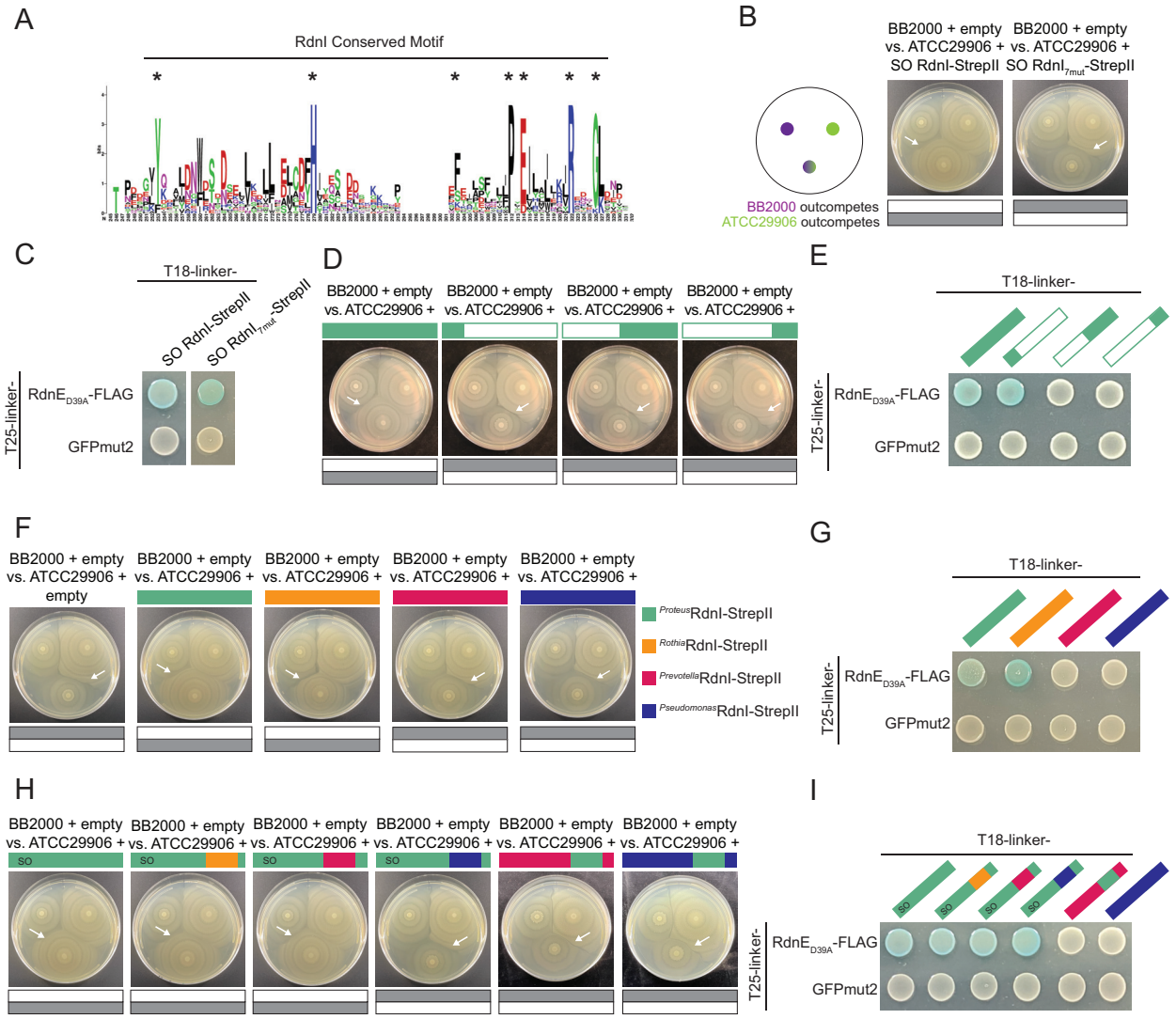

Figure 4. The RdnI protein family can offer cross-protection due to an interchangeable conserved domain that is critical for function. A) Sequence logo of the RdnI's conserved motif. Stars indicate the seven analyzed residues. **B)** Swarm competition assay with ATCC29906 producing either RdnI-StrepII or RdnI$_{7mut}$-StrepII, which contains mutations in all seven conserved residues. We used a sequence-optimized (SO) RdnI protein that had a higher GC% content and an identical amino acid sequence for ease of cloning. Left: schematic of swarm competition assay as in Figure 2. Gray boxes indicate which strain dominated over the other. White arrows point to the boundary formed between different strains. **C)** BACTH assay of RdnE$_{D39A}$-FLAG with SO RdnI-StrepII or RdnI$_{7mut}$-StrepII. GFPmut2 was used as a negative control. **D)** Swarm competition assay with ATCC29906 expressing either the wild-type RdnI or a RdnI truncation. The three truncations were in the first alpha helix (amino acids 1-85), the second half of RdnI (amino acids 150-305), and the end of the protein (amino acids 235-305). **E)** BACTH assay of RdnE$_{D39A}$-FLAG with wild-type RdnI and the three RdnI truncations. **F)** Swarm competition assay with ATCC29906 expressing foreign RdnI proteins. **G)** BACTH assay of RdnE$_{D39A}$-FLAG with each of the foreign RdnI proteins. GFPmut2 was used as a negative control. **H)** Swarm competition assay with ATCC29906 producing SO RdnI with swapped conserved motifs. **I)** BACTH assay of RdnE$_{D39A}$-FLAG with SO RdnI with swapped conserved motifs. Colored bars denote RdnI-StrepII proteins from *P. mirabilis* (green), *R. dentocariosa* (orange), *P. jejuni* (magenta), or *P. ogarae* (dark blue).

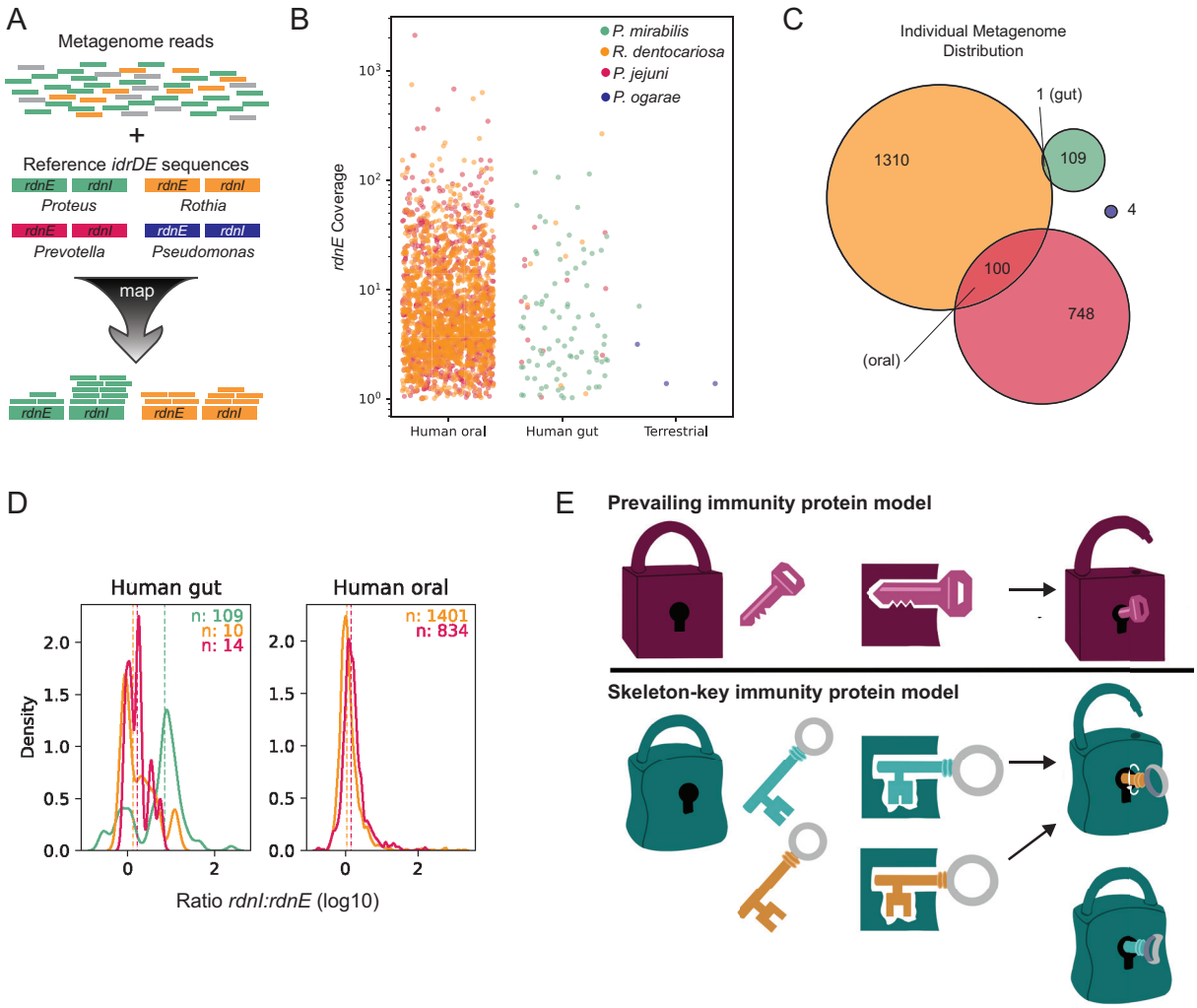

Figure 5. The RdnI protein family has the potential for broader protection within oral and gut microbiomes. A) Methodology used to identify *rdnE and rdnI* genes in publicly available metagenomic data. Metagenomes were mapped against sequences with a stringency of 90%. "Coverage" denotes the average depth of short reads mapping to a gene in a single sample. Colors represent *rdnE* and *rdnI* from *P. mirabilis* (green), *R. dentocariosa* (orange), *P. jejuni* (magenta), or *P. ogarae* (dark blue). **B)** The experimentally tested *rdnE* gene sequences from different organisms (colors) are found in thousands of human-associated metagenomes. Each dot represents a single sample's coverage of an individual *rdnE* gene, note $\log_{10}$-transormed y-axis. Only samples with >1x coverage are shown. **C)** Euler diagram showing the number of samples with co-occurring *rdnE* genes from different taxa (colors). **D)** Kernel density plot of the ratio of *rdnI* to *rdnE* coverage. The ratio of *rdnI* to *rdnE* was defined as $\log_{10}(I/E)$ where I and E are the mean nucleotide's coverage for *rdnI* and *rdnE*, respectively. The distribution of ratios was summarized as a probability density function (PDF) for each taxon (color) in each environment (subpanel). Here, the y-axis (unitless) reflects the probability of observing a given ratio (x-axis) in that dataset. The colored numbers in the top right of each panel show the number of metagenomes above the detection limit for both *rdnE* and *rdnI* for each taxon. Dashed vertical lines represent the median ratio. **E)** Skeleton-key model for immunity protein protection. Top, the current prevailing model for immunity proteins is that protection is defined by necessary and sufficient binding between cognate effectors (locks) and immunity proteins (keys). Bottom, our proposed skeleton-key model for protection is that multiple immunity proteins (skeleton-keys) can bind a single effector due to a flexible (promiscuous) binding site. Protection is a two-step process of binding and then neutralization.