nature communications

Article

https://doi.org/10.1038/s41467-024-51738-3

The role of TIR domain-containing proteins in bacterial defense against phages

Received: 2 May 2024

Accepted: 15 August 2024

Published online: 27 August 2024

Check for updates

Shuangshuang Wang[1,2,3], Sirong Kuang[1,2,3], Haiguang Song[1,2,3], Erchao Sun[1,2,3], Mengling Li[1,2,3], Yuepeng Liu[1,2,3], Ziwei Xia[1,2,3], Xueqi Zhang[1,2,3], Xialin Wang[1,2,3], Jiumin Han[1,2,3], Venigalla B. Rao[4], Tingting Zou[1], Chen Tan[1,2,3] & Pan Tao[1,2,3]

Toll/interleukin-1 receptor (TIR) domain-containing proteins play a critical role in immune responses in diverse organisms, but their function in bacterial systems remains to be fully elucidated. This study, focusing on *Escherichia coli*, addresses how TIR domain-containing proteins contribute to bacterial immunity against phage attack. Through an exhaustive survey of all *E. coli* genomes available in the NCBI database and testing of 32 representatives of the 90% of the identified TIR domain-containing proteins, we found that a significant proportion (37.5%) exhibit antiphage activities. These defense systems recognize a variety of phage components, thus providing a sophisticated mechanism for pathogen detection and defense. This study not only highlights the robustness of TIR systems in bacterial immunity, but also draws an intriguing parallel to the diversity seen in mammalian Toll-like receptors (TLRs), enriching our understanding of innate immune mechanisms across life forms and underscoring the evolutionary significance of these defense strategies in prokaryotes.

Toll/interleukin-1 receptor (TIR) domain-containing proteins play a vital role in the immune response of various organisms, including mammals, plants, and bacteria[1]. To date, at least 25 genes with TIR domains have been identified in the human genome, the vast majority of which are involved in pathogen recognition and the initiation of signaling pathways that activate the immune responses[2]. Similarly, in plants, proteins containing the TIR domain also play an important role in the immune system[3,4]. Specifically, TIR domains are nicotinamide adenine dinucleotide (NAD^+)-hydrolyzing enzymes that participate in signal transduction pathways essential for plant defense mechanisms against various pathogens[5,6]. Recent research has also elucidated the defensive role of TIR-containing proteins in bacteria, particularly in their protection against phage infection[7–11]. These findings underscore a universally conserved defense mechanism that spans various life forms.

In contrast to the numerous TIR domain-containing proteins involved in pathogen recognition in mammals[2], only six TIR-containing antiphage defense systems have been identified in bacteria (Supplementary Table 1)[8,9,12–15]. The Thoeris defense system employs TIR-domain proteins (ThsB) and a second protein (ThsA) with NADase activity. Upon phage infection, ThsB generates a signal molecule that activates ThsA, leading to NAD^+ depletion and abortion of the infection, thus preventing phage propagation[13,16,17]. In a type of Short prokaryotic Argonaute defense system, the TIR domain-containing protein and Short Argonaute form heterodimeric complexes (SPARTA). The NADase activity of the TIR domain of SPARTA is activated in the presence of highly transcribed multicopy plasmid DNA, leading to cell death through $NAD(P)^+$ depletion[9]. In a type of the CBASS defense system, the protein containing the TIR and STING domains forms a filament structure in response to phage infection[14,18]. The filament structure coordinates the activation of the NADase activity of the TIR domains, resulting in NAD^+ depletion and abortive infection to inhibit phage propagation[14]. Similarly, some Pycsar systems also use the TIR domain to deplete cellular

[1]State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan, Hubei 430070, China. [2]Hubei Hongshan Lab, Wuhan, Hubei 430070, China. [3]Cooperative Innovation Center for Sustainable Pig Production, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan, Hubei 430070, China. [4]Bacteriophage Medical Research Center, Department of Biology, The Catholic University of America, Washington, DC 20064, USA. e-mail: taopan@mail.hzau.edu.cn

NAD^+ to induce abortive infection. In these systems, the NADase activity of the TIR domain was activated by cyclic pyrimidines, which were produced by uridylate cyclases in response to phage infection[8]. TIR domains have also been identified in the Retron[15,19] and Avs (antiviral STAND)[12,20] defense systems, but their mechanisms for preventing phage replication are unknown.

While a large number of TIR domain-containing proteins have been identified in bacteria[8,9,12–15], the specific functions of many of these proteins, particularly in the context of antiphage activity, are not yet fully understood. This gap in knowledge represents a significant area for exploration, similar to the extensive roles that have been identified for TIR domain-containing proteins in the human immune response. In addition, phage components that specifically trigger the activation of bacterial TIR systems are unknown. Deciphering these triggers is critical to understanding the mechanisms by which TIR systems sense and respond to phage attack, which could have profound implications for our understanding of bacterial defense strategies. This study aims to systematically explore the landscape of the TIR defense system in *E. coli* by utilizing all available *E. coli* genome sequences from NCBI, together with a comprehensive collection of *E. coli* phages, thereby aiming to bridge these identified knowledge gaps.

Results

Extensive antiphage activity of TIR domain-containing proteins in *E. coli*: unveiling a robust bacterial defense system

To explore the landscape of TIR domain-encoding potential defense systems in *E. coli*, we first identified all TIR domain-containing proteins (781) out of 11,688,950 annotated *E. coli* proteins using all 2289 *E. coli* genomes available in NCBI by October 2020 (Fig. 1a). To ensure that all TIR-containing proteins are encoded by the *E. coli* genome, only the genomic DNAs were used for the analysis. These TIR-containing proteins were clustered into 64 clusters based on their sequences. We analyzed the genomic structure spanning 5 genes upstream and downstream of TIR genes in each cluster to identify the operon configuration, which were considered as candidate defense systems (Fig. 1a). For the 21 clusters with more than 10 members (Fig. 1b), we cloned one representative candidate system from each cluster from our *E. coli* collections. For the remaining clusters with 1–8 members, we were able to clone the representative candidate system from 11 clusters using our *E. coli* collections (Fig. 1b). In total, we cloned 32 candidate systems representing 90% (703 of the 781 cluster members) of *E. coli* TIR domain-containing proteins. The candidate systems with their native promoter were cloned into pSEC1 vector and transformed into *E. coli* MG1655 strain to evaluate their defense activities against a collection of 111 different phage strains[21].

E. coli MG1655 with or without the candidate TIR defense system was infected with 10-fold serially diluted phages to observe plaque formation. The defense activity of the TIR systems against each phage was determined based on their ability to inhibit plaque formation compared to the pSEC1 empty vector control, and expressed as the fold reduction in Efficiency of plating (EOP). Twelve systems have defense activities against a variable number of phages, and 9 of them are new defense systems and named TIR-I to IX (Fig. 2a). The antiphage activity of each TIR system depends on the specific phage, with strong inhibitory effects for some phages and little or no inhibitory effects for others. However, all TIR systems can inhibit at least one

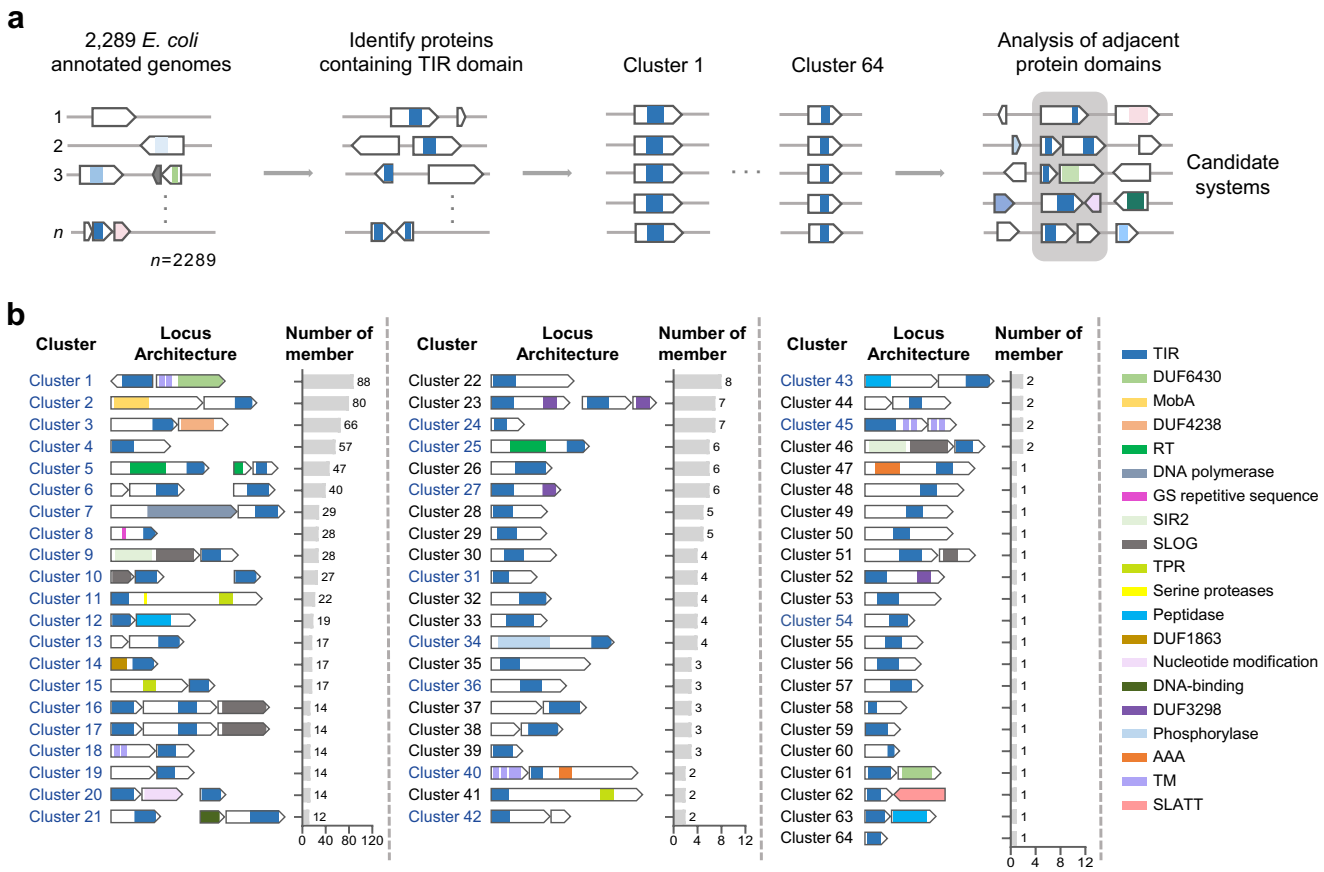

Fig. 1 | Screening of candidate antiphage defense systems containing TIR domains. a Flowchart to identify candidate systems containing TIR domains from all *E. coli* genomes available in the NCBI database. **b** Locus architectures of 64 TIR candidate defense systems. The bar graph shows the members of each cluster in the sequenced *E. coli* genomes. The cluster highlighted in blue indicates the 32 TIR systems cloned in this study. Domains: RT reverse transcriptase, DUF domain of unknown function, TPR tetratricopeptide repeat, TM transmembrane.

https://doi.org/10.1038/s41467-024-51738-3

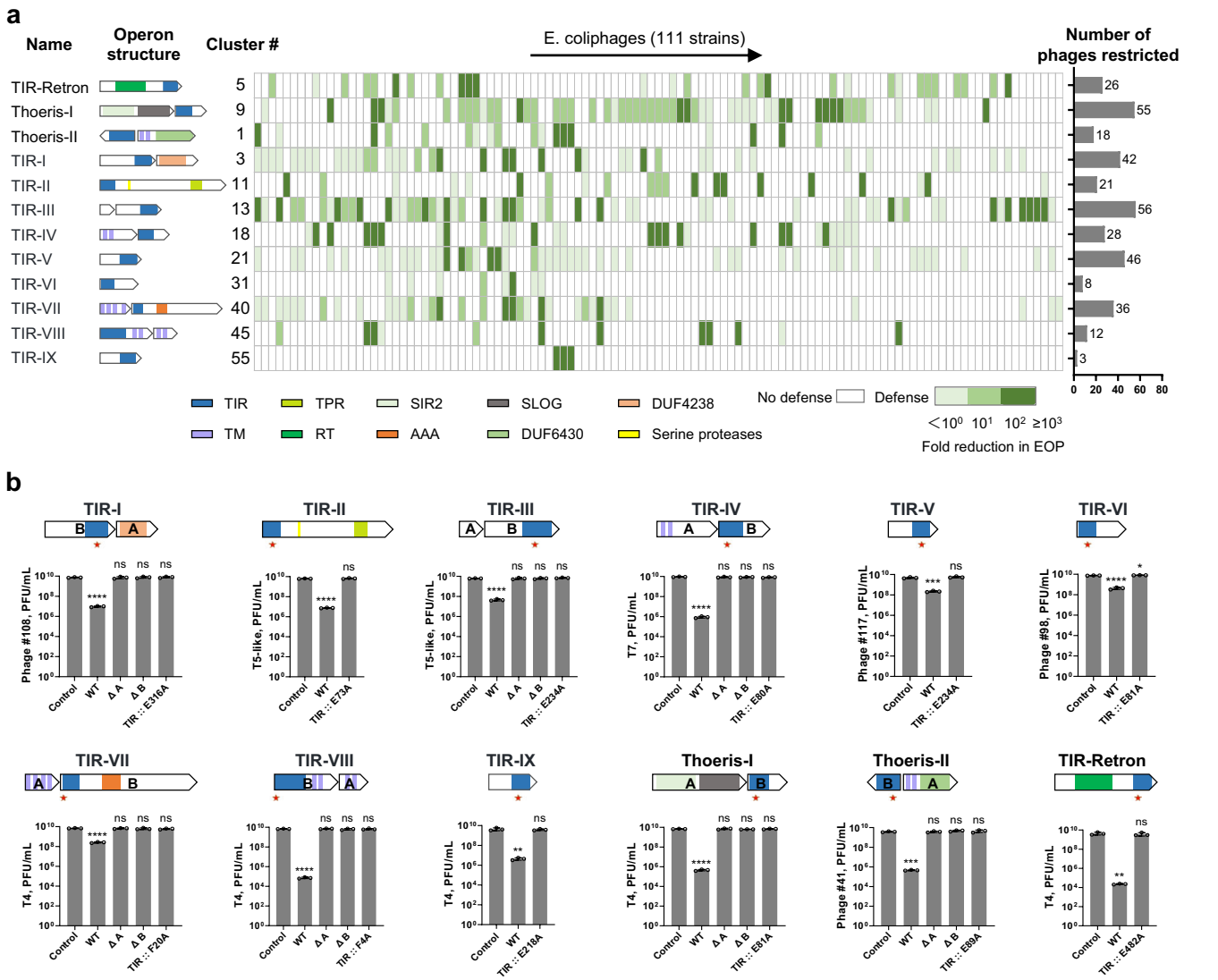

Fig. 2 | Identification of 12 antiphage systems with TIR domains in *E. coli*. a The domain organization of each system was shown, and their defense activities were determined in triplicate using *E. coli* MG1655 against 111 different phages as described in Methods. The color indicates the fold reduction in EOP of the representative results of each TIR system against each phage in EOP experiments. The bar graph shows the number of phages restricted in each system. **b** Essential domain architectures and mutational analysis of the defense systems. Conserved amino acids or the putative active sites were indicated with red stars. Assays were done in triplicate and data were presented as the mean ± S.D. *E. coli* MG1655 cells containing the empty vector were used as controls. Asterisks indicate statistical significance relative to the control. *, **, ***, and **** indicate $p < 0.05$, $p < 0.01$, $p < 0.001$, and $p < 0.0001$, respectively (ANOVA). The ns indicates not significant. Source data are provided as a Source data file.

phage with a 10^3-fold reduction in EOP (highlighted in dark green in Fig. 2a), and 95.5% of the phages (106/111) are targeted by at least one of these TIR defense systems.

Five of these new systems (I, III, IV, VII, and VIII) contain two proteins, and the single gene deletion assays showed that both proteins of each system are required for their antiphage activities (Fig. 2b). All other new defense systems contain only one protein. In addition to the TIR domains, four new systems also contain either transmembrane (TM), AAA, tetratricopeptide repeat (TPR), or DUF4238 domains, which have also been found in other defense systems[7,12,19,20]. For the remaining new defense systems, only the TIR domain was identified (Fig. 2). Mutation of the conserved amino acid residues within the active sites of TIR NADase[22] revealed that TIR domains are required for the defense activities of all the 12 systems (Fig. 2b). Thoeris systems have previously been cloned from *Bacillus subtilis*[7,13,23,24], but their homologous systems in *E. coli* have not yet been confirmed. We have shown that both type I (cluster 9) and type II Thoeris (cluster 1) are present in *E. coli*, with type I having broader antiphage activities (Fig. 2a).

For all TIR systems, phage infection at a multiplicity of infection (MOI) ≥ 2 resulted in more rapid death of *E. coli* cells expressing an individual defense system compared to cells without the defense system (Supplementary Fig. 1). The one-step growth curve experiment assays showed that no or very few progeny phages were produced in *E. coli* cells expressing TIR-VII, TIR-VIII, or TIR-Retron. For the remaining TIR systems, the progeny phages were significantly reduced in *E. coli* cells expressing an individual defense system compared to cells without the defense system (Supplementary Fig. 2).

Evolutionary analysis revealed that the TIR-Retron systems have a wide genetic diversity, while the remaining 11 TIR systems are divided into two phylogenetic branches (Supplementary Fig. 3). Prevalence analysis of TIR systems in prokaryotes revealed that TIR-II is the most abundant and widespread system (Supplementary Fig. 4a). Interestingly, this system was also found in Caudoviricetes

of the Duplodnaviria phylum (Supplementary Fig. 4a), suggesting that it may be a mechanism for interspecies competition among phages. TIR-IV systems are mainly found in *Proteobacteria* and *Bacillota* (Supplementary Fig. 4b). Most of the TIR-III and TIR-VIII systems were found in *Proteobacteria*, whereas TIR-I and TIR-VII systems were found only in *Proteobacteria* (Supplementary Fig. 4c–f). Since TIR-V, TIR-VI, and TIR-IX are single-protein systems with only a short TIR domain, they were excluded from our phylogenetic analysis. Nearly 30% of TIR-II systems contain an ATPase domain in addition to a TPR domain, which has also been found in the Avs system[12,25]. The TIR and TPR domains can be present in the same protein or can be located separately on two different proteins (Supplementary Fig. 4g). In some cases, TIR-II contains one or two additional proteins. TIR-I, TIR-III, and TIR-IV systems appear in diverse domains contexts, whereas the TIR-VII and TIR-VIII systems were found in a single form (Supplementary Fig. 4g).

Taken together, our results indicate that TIR domain-containing proteins, like restriction-modification (RM) and other prevalent defense systems[26], play an important role in defense against phage infection in *E. coli*. While these results suggested that a significant proportion of TIR proteins (12 of the 32 tested TIR proteins) have defensive roles, it is important to note that these results are based on the subset of proteins we tested and may not represent the activity of all TIR domain-containing proteins in *E. coli*.

Diverse recognition spectrum of *E. coli* TIR systems: covering all stages of phage replication

To elucidate the phage components that trigger the activation of the TIR systems, we first selected the phage mutants that have evolved the ability to escape the constraints imposed by TIR systems over multiple generations (Fig. 3a). In total, we screened 62 phage-host pairs and isolated 24 escaped phages, which partially or completely overcame 10 of the 12 TIR defense systems (Fig. 3b). We isolated multiple resistant phages for the TIR-III, TIR-IV, TIR-V, TIR-VII, TIR-VIII, TIR-Retron, and type II Thoeris systems, but only a single escaped phage was isolated for the TIR-I, TIR-VI, and TIR-IX systems. Escapes from the type I Thoeris and TIR-II systems were not achieved.

Comparative sequence analysis between the escaped phages and their wild-type (WT) counterparts identified 25 phage genes that could trigger TIR systems (Fig. 3b). Of these, 20 encode proteins with diverse known functions, including protein kinase, DNA helicase, ssDNA binding protein, capsid protein, and tail protein. These phage proteins are involved in different stages of phage replication, covering the early, middle, and late phage periods (Fig. 3c). These results suggest that the TIR defense systems are capable of recognizing different phage components, similar to mammalian Toll-like receptors (TLRs). TLRs are groups of proteins that contain the TIR domain and are a crucial part of the innate immune system. To date, 10 TLRs have been identified in humans, and each TLR recognizes specific patterns found in various pathogens, known as pathogen-associated molecular patterns (PAMPs), and plays a critical role in initiating immune responses[27]. Similarly, our results showed that *E. coli* employs multiple TIR systems to recognize different phage-associated molecular patterns (PhAMPs) and trigger the antiphage responses. It is important to note that the TIR domains of mammalian TLRs do not directly recognize pathogens[27]. Similarly, the TIR domains of *E. coli* do not necessarily recognize phages directly.

Since all 12 TIR systems protect against phages by inducing cell death or growth arrest, we determined the activity of the identified proteins to induce cell death by co-expressing individual proteins and their corresponding TIR system. We found that the expression of tail tube protein of phage #90 was toxic in cells containing the TIR-III system but not in cells lacking the defense system, whereas the mutant tail tube protein was not toxic even in the presence of the TIR-III system (Fig. 3d). Similarly, the expression of the WT, but not the mutant, tail fiber protein of phage #75 (Fig. 3e), the hypothetical protein of phage #114 (Fig. 3f), and the ssDNA-binding protein of phage #45 (Fig. 3g) were toxic only in cells containing the TIR-VI, TIR-VIII, and TIR-Retron systems, respectively. These results suggest that these proteins can trigger the activation of corresponding TIR systems and lead to cell death or growth arrest. We then analyzed in detail all mutations that conferred phage resistance to the defense system, which were discussed below.

Individual TIR defense system can be sensitive to the same components of different phages

In humans, each of the 10 TLRs can recognize similar PAMPs from different pathogens. For instance, TLR4 is capable of recognizing lipopolysaccharide (LPS) from a wide range of Gram-negative bacteria[28]. Similarly, in *E. coli*, we found that some TIR systems can be sensitive to the same components of different phages. For instance, all four escaped phages (#45, #41, #75, and #106) from TIR-IV have mutations only in their major capsid proteins (Figs. 3b and 4a). The major capsid proteins of phages #41, #75, and #106 have 97–99.2% sequence similarity, but the mutation sites of each escaped phage are different (V89A, V106D, and A176E for phages #41, #75, and #106, respectively) (Fig. 4a). Interestingly, these three phages showed no sequence similarity to phage #45, indicating that the activation of TIR-IV is not sequence specific. Similarly, two of the three escaped phages (#95 and #97) from TIR-Retron have mutations in their DNA primase-helicase proteins, and two of the four escaped phages (#75 and #117) from type II Thoeris have mutations in their tail proteins (Figs. 3b, 4b, and 4c). Amino acid sequence analysis showed that there is only 35.5% and 46.5% sequence similarity for the DNA primase-helicase and tail proteins, respectively, indicating that these TIR systems may recognize the conserved structure or function rather than the sequence.

We then focused on the TIR-IV system to determine how it recognizes the major capsid proteins. First, we modeled the structures of the major capsid proteins of phages #45 and #41 using AlphaFold2[29]. Since phages #75 and #106 share 97% and 99.2% DNA sequence identity with phage #41 respectively, they were not included in the structural analysis. Interestingly, the major capsid proteins of phages #45 and #41 share a similar structure although without sequence similarity (RMSD = 4.73), indicating that TIR-IV recognizes the specific structure (Fig. 4d). However, none of these major capsid proteins can directly activate TIR-IV to induce cell death (Supplementary Fig. 5a), indicating that the major capsid protein is not efficient in activating TIR-IV. In addition, the transmission electron microscopy analysis of the purified phages showed that all the escaped phages have similar morphology to WT phages, indicating that the mutations did not alter the capsid structures (Fig. 4e).

To determine whether mutations on the major capsid protein affect the infection kinetics, *E. coli* cells with WT or inactivated mutant TIR-IV system were infected with the WT or the escaped phages. As expected, *E. coli* culture containing WT TIR-IV system is resistant to WT phages, whereas the escaped phages can induce culture collapse after 1–2 h of infection (Supplementary Fig. 5b). Interestingly, for the phages #41, #45, and #75, the escaped phages lysed cultures containing the inactive TIR-IV system about 20–70 min slower than the WT phages (Fig. 4f). One-step growth curves also showed that these escaped phages had an increased latency and a decreased number of the released virions compared to the WT phages (Fig. 4g), which is consistent with the fact that the escaped phages generated smaller plaques compared to the WT phages (Supplementary Fig. 5c). These results indicate that the mutations on the major capsid protein can dramatically alter culture lysis kinetics, probably by affecting virion assembly to enable immune evasion. However, other mechanisms may also be involved, as escaped mutant of phage #106 showed similar infection kinetics to the WT phage (Fig. 4f, g and Supplementary Fig. 5b, c, last panels).

https://doi.org/10.1038/s41467-024-51738-3

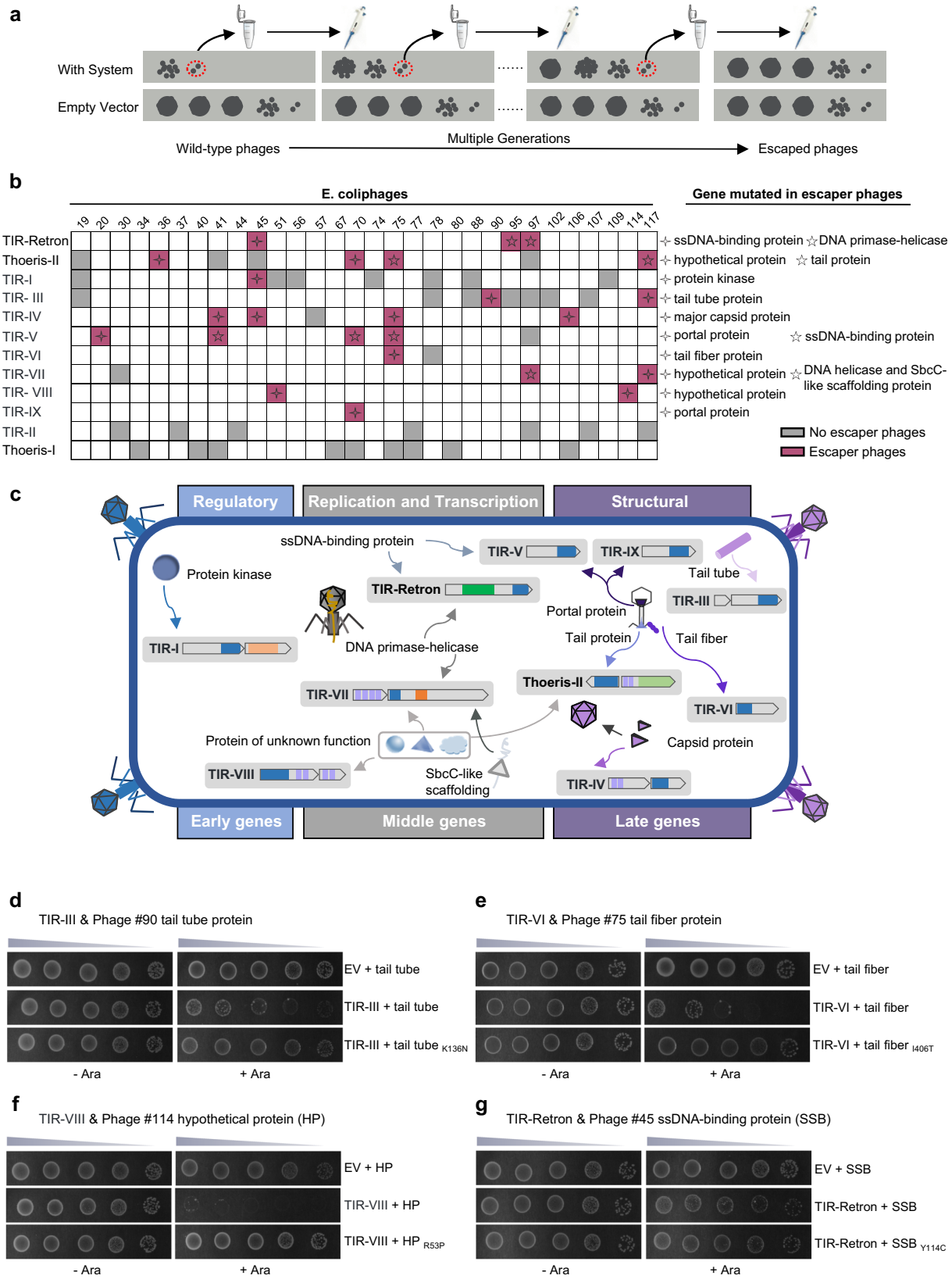

Fig. 3 | Broad recognition spectrum of *E. coli* TIR systems. a Pipeline for isolating mutant phages that overcome TIR defense systems. *E. coli* MG1655 cells containing a single TIR defense system were infected with phages. The plaque formed by high titer phages was harvested and subjected to a second round of selection. *E. coli* MG1655 cells containing empty vector were used as controls. **b** Identification of mutated genes in escaped phages. Gray blocks indicate that no escaped phage was isolated. Four- and five-pointed stars in each maroon block indicate the identity of the mutated gene as indicated on the right. **c** Schematic representation of genes sensitive to TIR systems, covering different stages of the phage replication. **d**–**g** Toxicity of TIR system sensitive genes in the presence of the corresponding TIR system. Protein expression was induced by the addition of arabinose. The empty vector lacking the TIR system was used as a negative control. The mutated genes were used to determine whether mutant proteins could confer phages resistant to the TIR system. Source data are provided as a Source data file.

Different TIR defense systems can be sensitive to the same components of different phages

In mammals, there are cases where different TLRs can recognize the same PAMPs, for example, both TLR7 and TLR8 can recognize single-stranded RNA from viruses[30,31]. Similarly, we have found that different TIR systems can recognize the same components of different phages. For instance, TIR-V recognizes the portal protein of phage #20, and TIR-IX recognizes the portal protein of phage #70 (Fig. 5a, b). In determining the toxicity of these portal proteins, we found that it is very difficult to transform the portal genes of phage #20 into *E. coli* cells containing the TIR-V system, and the transformation efficiency decreased 10-fold to 100-fold (Fig. 5c). In addition, the colonies generated are tiny. These results indicated the toxicity of the portal protein. We then randomly picked up two colonies from each transformation and sequenced the TIR-V and portal protein expression plasmids. Surprisingly, the sequencing results showed that all TIR-V systems were disrupted at the same position by the transposase S1 (IS1) or S4 (IS4) families when the WT or mutant portal genes were present (Fig. 5c and Supplementary Fig. 6a), whereas no mutations were found in the portal genes. The IS4 family of transposable elements, which includes a variety of insertion sequences characterized by specific transposase motifs, is involved in genomic rearrangements and is important for bacterial adaptability and evolution[32]. Our results showed that the disrupted TIR-V systems lost antiphage activity when transformed into *E. coli* MG1655 (Fig. 5d). These results indicated that co-expression of portal proteins, even the mutant, and the TIR-V system was lethal to *E. coli* cells.

Expression of the WT, but not the mutant, portal protein of phage #70 was toxic to *E. coli* cells containing the TIR-IX system (Fig. 5e), indicating that the portal protein can activate the TIR-IX system. To determine whether phage portal proteins are common triggers, we determined the toxicity of portal proteins of T7 or T4 phages in *E. coli* cells expressing these systems and found that T7 portal protein can activate TIR-V but not the TIR-IX systems, whereas T4 portal protein can activate both systems (Fig. 5f, g). Interestingly, both TIR-V and TIR-IX are unable to protect *E. coli* cells from T7 phage, whereas TIR-V is unable to protect *E. coli* cells from T4 phage (Supplementary Fig. 6b), suggesting that these phages may encode proteins for anti-TIR-V and anti-TIR-IX systems. Similar phenomenon has been reported in the CBASS system, where the T4 phage encodes the gp57B protein to hydrolyze 3′, 3′-cGAMP, a signal molecule to activate CBASS effectors[33]. Taken together, our results indicated that the portal proteins may be the common triggers of TIR-V and TIR-IX systems, and TIR-V and TIR-IX systems may share the same recognition pattern, while each has its own specific recognition pattern.

Amino acid sequence alignment showed that these portal proteins have no sequence similarity, indicating that TIR-V and TIR-IX systems recognize the similar structural pattern. Therefore, we modeled the structure of these portal proteins using AlphaFold2 and found that they are partially superimposed meanwhile with some differences (RMSD = 5.39–7.36) (Fig. 5h, i), which may partially explain the recognition pattern of TIR-V and TIR-IX systems. Both TIR-V and TIR-IX are single-protein systems, and structural analysis using AlphaFold2 showed that their C-termini are almost completely superimposed (RMSD = 0.76) although with very low sequence similarity, while their N-termini are not superimposed (Fig. 5j). This suggests that although both the TIR-V and TIR-IX systems recognize the portal proteins, the dissimilarity of the N-termini may make the portal proteins they recognize different. However, further studies are needed to confirm this speculation.

Different TIR defense systems can recognize different components of the same phage

In mammals, each TIR protein recognizes a different component of the same pathogen, forming a powerful defense network against infection by that pathogen[34,35]. To determine whether this is also the case in bacteria, we analyzed three phages, #45, #70, and #117, which can escape three different TIR systems during the directed evolution (Fig. 6a–c). The phage #117 escape mutants from Type II Thoeris (phage #117$_{\text{Thoeris-II}}$), TIR-III (phage #117$_{\text{TIR-III}}$), and TIR-VII (phage #117$_{\text{TIR-VII}}$) systems contain a point mutation on the tail protein, the tail tube protein, and a hypothetical protein, respectively (Fig. 3b). We found that the phage #117$_{\text{Thoeris-II}}$ can only escape the Type II Thoeris system, but not the TIR-III and TIR-VII systems (Fig. 6a). These results are further supported by the fact that the tail tube protein is toxic only in the presence of TIR-III, whereas the tail tube protein cannot induce cell death in the presence of either Thoeris-II or TIR-VII systems, indicating that the tail tube protein cannot activate these two TIR systems (Fig. 6d).

The phage #45 escape mutants from TIR-Retron (phage #45$_{\text{TIR-Retron}}$), TIR-I (phage #45$_{\text{TIR-I}}$), and TIR-IV (phage #45$_{\text{TIR-IV}}$) systems contain a point mutation on the ssDNA-binding protein, protein kinase, and major capsid protein, respectively (Fig. 3b). Similarly, we found that the phage #45$_{\text{TIR-Retron}}$ can only escape the TIR-Retron system, but not the TIR-I and TIR-IV systems (Fig. 6b). The ssDNA-binding protein of phage #45 has weak toxicity in the presence of TIR-Retron, but not in the presence of TIR-I and TIR-IV (Fig. 6e). Similar results were also observed for phage #70 that the escaped phages form Type II Thoeris can still be restricted by TIR-V and TIR-IX systems (Fig. 6c).

Taken together, these results indicate the strong defense network of the TIR systems, such that a phage escaped from one TIR system by mutation can still be recognized by the other type of TIR system. This is probably one of the reasons for the great diversity of bacterial TIR systems in nature. Although only 28.8% of *E. coli* strains contain more than one TIR system (Supplementary Fig. 7), the high frequency of horizontal gene transfer (HGT) in *E. coli* makes it easy to obtain new defense systems from the population.

Discussion

The comprehensive study presented here has elucidated the sophisticated nature of bacterial defense mechanisms against phage attack, with a particular focus on the role of TIR domain-containing proteins in *E. coli*. A key finding of our investigation is that a substantial proportion, 37.5% to be precise, of the tested *E. coli* TIR proteins exhibit antiphage activities. This not only underscores the robustness of TIR systems in bacterial immunity, but also points to a remarkable parallel with the diversity observed in mammalian TLRs. Such a parallel enhances our understanding of innate immune mechanisms across different biological domains and underscores the evolutionary importance of these defense systems in prokaryotes. TLRs are typically membrane-associated, share highly conserved structural features such as transmembrane domains and leucine-rich repeats, and often function as homo- or heterodimers with other TLRs[36]. However, it seems to us that some of these features are absent in bacterial TIR systems, and more studies are needed to further compare the mechanisms of bacterial TIR systems with mammalian TLRs.

Our study of the specific interactions between *E. coli* TIR proteins and various phage components underscores the versatility and specificity of these bacterial defense systems. The data showed that different TIR systems in *E. coli* can recognize and respond to a wide range of phage-associated molecular patterns, mirroring the adaptability seen in more complex immune systems such as those of mammals. This broad recognition capability underscores the ongoing evolutionary arms race between bacteria and phages, highlighting the critical role of diverse recognition strategies in bacterial survival. The ability of *E. coli* TIR systems to recognize a broad spectrum of phage components indicates a sophisticated evolutionary adaptation that provides these bacteria with a versatile toolkit to counter diverse phage threats. This versatility not only demonstrates the complexity of bacterial defense, but also sheds light on host-pathogen interactions at the molecular level and provides insights into their co-evolutionary dynamics.

https://doi.org/10.1038/s41467-024-51738-3

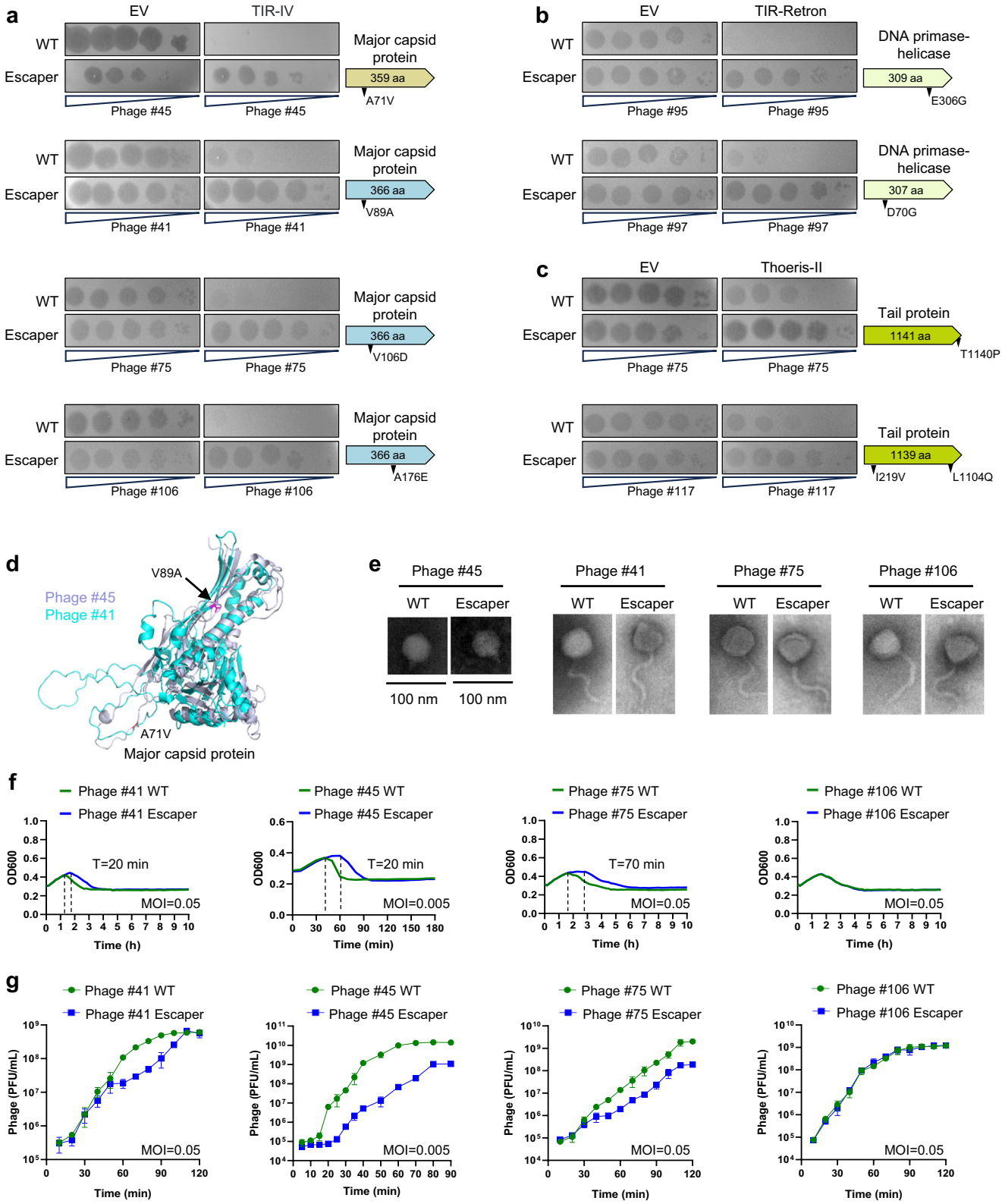

Fig. 4 | Individual TIR defense systems can be sensitive to the same components of different phages. a–c Plaque assays showed that the TIR-IV (**a**), TIR-Retron (**b**), and Thoeris-II (**c**) systems are sensitive to major capsid proteins, DNA primase-helicases, and tail proteins of different phages, respectively. The empty vector lacking the TIR system was used as a negative control. Tenfold serial dilutions of phage were used for infection. Mutation sites in the proteins were indicated by black arrows. **d** Structural analysis of the major capsid proteins of phages #41 and #45. Structures were modeled using AlphaFold2 and analyzed using PyMOL. The mutation sites were highlighted in red. **e** Electron micrographs of purified WT and escaped phage virions. Assays were performed in triplicate and the representative results were shown. **f** Infection kinetics of WT (green lines) and escaped phages (blue lines) #45, #41, #75, and #106 on *E. coli* MG1655 expressing the inactivated TIR-IV mutant. *E. coli* MG1655 cells expressing the WT TIR-IV system were used as controls, and the data were shown in Supplementary Fig. 5. The dashed lines indicated that the escaped phages lysed cultures about 20–70 min slower than the WT phages. **g** One-step growth curve of WT and escaped phages #45, #41, #75, and #106 on *E. coli* MG1655. Data were presented as the mean of three biological replicates ± S.D. Source data are provided as a Source data file.

https://doi.org/10.1038/s41467-024-51738-3

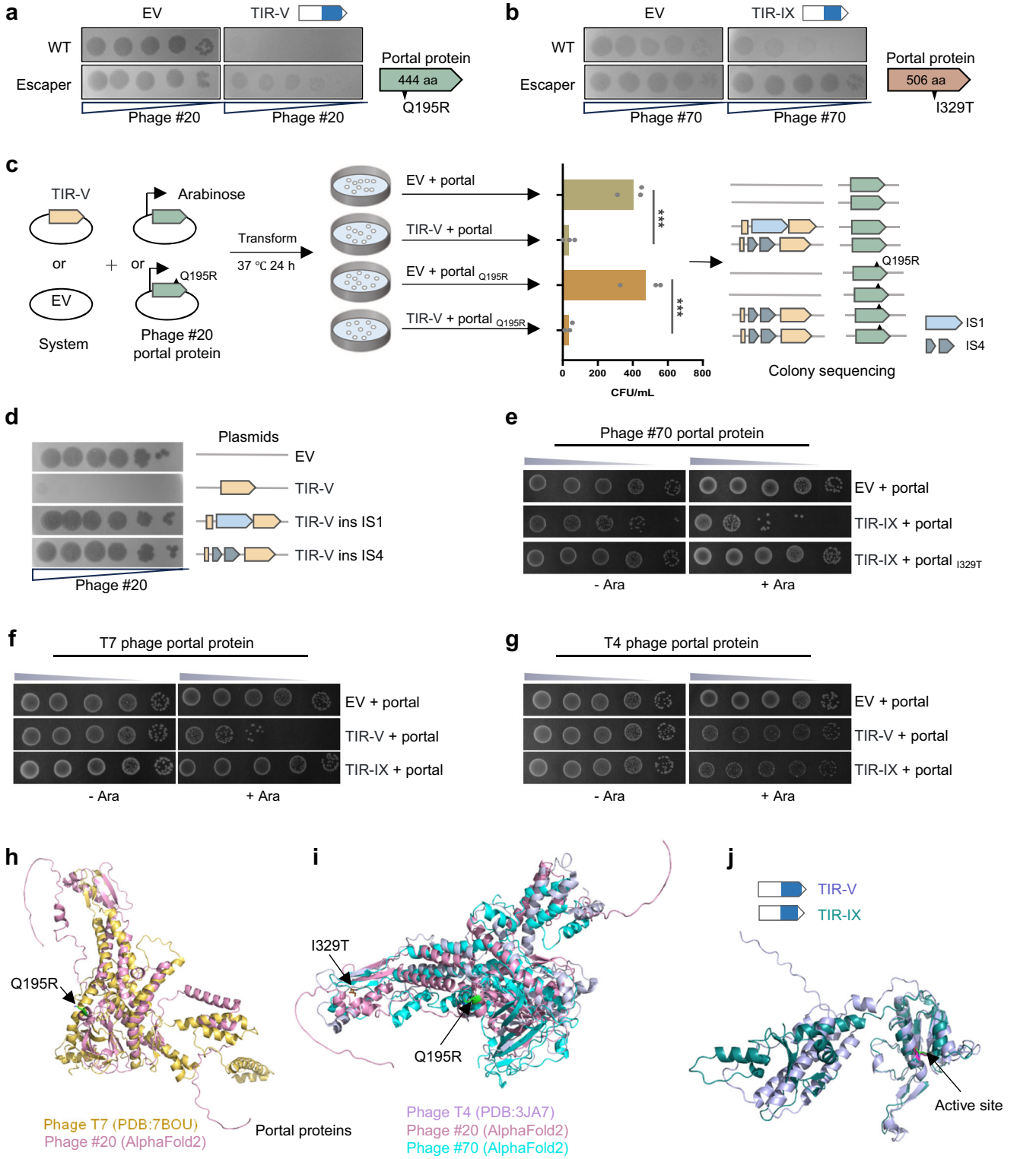

Fig. 5 | Different TIR defense systems can be sensitive to the same components of different phages. a, **b** Plaque assays showed that the TIR-V (**a**) and TIR-IX (**b**) systems are sensitive to portal proteins of phages #20 and #70, respectively. The empty vector lacking the TIR system was used as a negative control. Tenfold serial dilutions of phage were used for infection. Mutation sites in the proteins were indicated with black arrows. **c** Diagram showing that co-transformation of TIR-V and portal protein expression plasmids resulted in disruption of the TIR-V system by transposase S1/S4 insertions. The histogram showed that the transformation efficiency of the portal protein expression plasmids was significantly decreased in the presence of the TIR-V system. Data were presented as the mean of three biological replicates. *** indicates $p < 0.001$ (ANOVA). The right panel showed that the TIR-V system was disrupted by transposase S1/S4 insertions as determined by colony sequencing. **d** Plaque assays showed that the antiphage activity of TIR-V was abolished by transposase S1/S4 insertions. **e** Toxicity of the WT portal protein of phage #70 in the presence of the TIR-IX system. **f**, **g** The T7 phage portal protein can activate TIR-V but not the TIR-IX system (**f**), whereas the T4 phage portal protein can activate both (**g**). **h**, **i** Structural analysis of the portal proteins of phage T7 (dark yellow), #20 (violet), T4 (lilac), and #70 (cyan). Mutation sites of escaped phages were indicated. **j** Structural and conservation analysis of TIR-V and TIR-IX proteins. Structures were modeled using AlphaFold2 and analyzed using PyMOL. Source data are provided as a Source data file.

https://doi.org/10.1038/s41467-024-51738-3

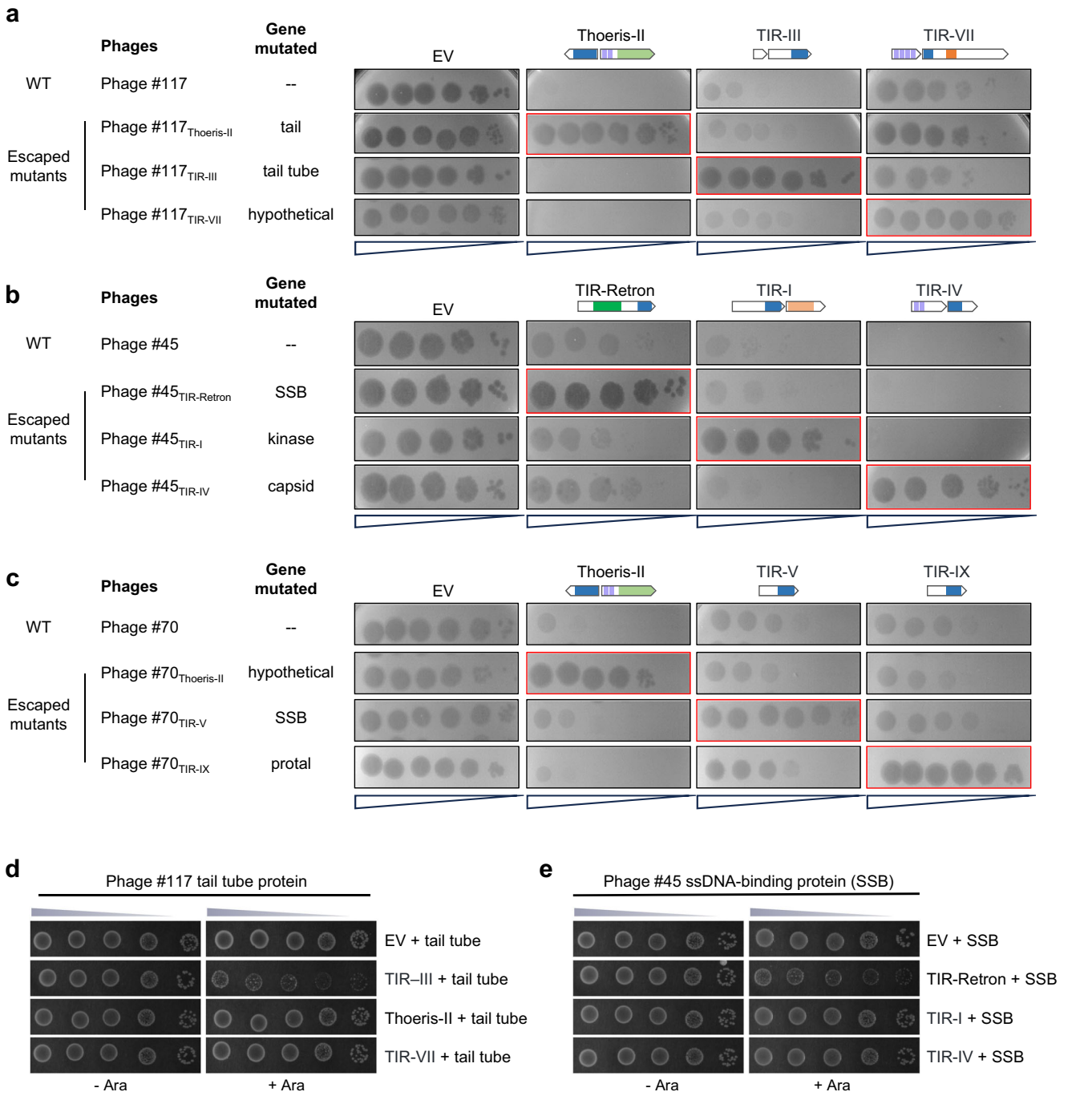

Fig. 6 | Different TIR defense systems can recognize different components of the same phage. a–**c** Plaque assays showed the resistance of different escaped phages of #117 (**a**), #45 (**b**), and #70 (**c**) from different systems. The escaped phages can only escape the TIR system, from which they were selected. The empty vector lacking the TIR system was used as a negative control. **d** Toxicity assays of the tail tube protein of phage #117 in the presence of the Thoeris-II, TIR-III, or TIR-VII systems. **e** Toxicity assays of the ssDNA-binding protein (SSB) of phage #45 in the presence of the TIR-Retron, TIR-I, or TIR-IV systems. Protein expression was induced by the addition of arabinose. The empty vector lacking the TIR system was used as a negative control. Source data are provided as a Source data file.

While our study provides valuable insights, the detailed mechanisms of action for each TIR system in *E. coli* have not been fully elucidated. Given the large number of TIR systems identified, dissecting the specific mechanisms of each system was beyond the scope of this study. This limitation points to the need for further research focusing on individual TIR systems to fully explore their mechanisms and roles in bacterial immunity. Future research should delve into the molecular intricacies that govern the specificity of each TIR system in phage recognition and response. Although all of these systems contain the TIR domain, the function of this domain in each system may be different. For instance, the TIR domain of Thoeris generates a signal molecule to activate the effector ThsA[13,16,17], whereas the TIR domain of the CBASS and Short Argonaute defense systems depletes cellular NAD^+ to induce abortive infection[9,14]. It is also important to note that while most of the genes identified from escape mutants are likely triggers for the TIR systems, some of the identified proteins were not toxic when co-expressed with the corresponding TIR system (Supplementary Fig. 8). This suggests that it's also possible that some of these protein mutants function as anti-TIR systems, a phenomenon observed

in other bacterial defense systems[37]. This gap highlights a potential area for further investigation.

Although our study used 111 diverse phages to assess the potential TIR systems, there remains a small possibility that some of these phages may have escaped detection by certain TIR systems, leading to potential false negatives. Furthermore, we were able to test the antiphage activity of 90% of the identified TIR proteins, leaving a gap where the remaining 10% could harbor unobserved antiphage activity. It's also important to recognize that although we analyzed all *E. coli* genomes available in the NCBI database, this comprehensive approach does not guarantee the identification of all TIR systems within *E. coli*. The extensive genetic diversity of the species and ongoing evolutionary processes suggest that additional, as yet undiscovered, TIR systems may exist.

In conclusion, this study enriches our understanding of the sophisticated nature of bacterial defense mechanisms, particularly in *E. coli*. While our findings have laid a foundation for understanding the role of TIR systems in bacterial immunity, they also highlight the need for more detailed studies to fully unravel the intricacies of these systems fully. The insights gained from this research not only advance our knowledge in microbiology, but also pave the way for innovative applications in medicine and biotechnology.

Methods

Bacteria and phages

E. coli strain DH5α (hsdR17(rK-mK+) sup2) was used for plasmid construction. *E. coli* strain MG1655 (F-λ-rph-1) was used for determination of the antiphage activities of the TIR-containing defense systems. A collection of 816 *E. coli* strains from our laboratory stocks was used to clone the TIR-containing candidate defense systems. Phages T4 and T7, along with 111 phages isolated in our previous study using *E. coli* MG1655 as a host were used individually to determine the antiphage activity of each candidate system.

Identification of the candidate defense systems containing TIR domains in *E. coli*

In this study, we systematically analyzed all 2289 *E. coli* genomes available in the NCBI database as of October 2020, as described previously[21]. Briefly, the protein-coding sequences were analyzed using Prodigal (release 2.6.3)[38], and protein domains were annotated using HMMER version 33.0[39]. We identified 781 proteins containing TIR domains and subsequently clustered them using the 'cd-hit' tool, with a sequence identity threshold of '-c 0.9' and a word size parameter of '-n 5'. To explore potential operonic structures, we examined the genomic context surrounding each TIR domain-containing protein, analyzing the five genes located both upstream and downstream. This effort aimed to identify genes in close proximity to the TIR proteins, indicating their potential inclusion within the same operons, thereby nominating entire operons as candidate defense systems. Additionally, the domain structures of these genes were analyzed for functional insights using the protein structural domain prediction tool at GenomeNet Motif Search (https://www.genome.jp/tools/motif/). The domain organization of the 12 TIR defense systems was analyzed using HHPRED with a PFAM and PDB database[40,41]. Structure of TIR domain-containing proteins and some phage proteins were modeled using AlphaFold2 with default setting[29] and analyzed using PyMOL (Supplementary Data 3). For proteins with little or no sequence similarity, the cealign command is used for structure alignment. For proteins with >30% sequence similarity, the align command is used for structure alignment.

Cloning of candidate systems into *E. coli* MG1655

In total, 64 clusters of candidate systems were identified. For 21 clusters with more than 10 members, we were able to clone one representative candidate system from each cluster from a collection of 816 *E. coli* strains stored in the laboratory. For the remaining clusters, we were able to clone the representative candidate system from 11 clusters using our *E. coli* collections. The candidate systems including their native promoters were amplified by PCR from our *E. coli* collections (Supplementary Data 1). The PCR products were cloned into the pSEC1 vector and confirmed by sequencing. In total, we cloned 32 candidate systems, representing 90% of *E. coli* TIR domain-containing proteins. The candidate systems were individually transformed into *E. coli* MG1655 strain to evaluate their defense activities against a collection of 111 different phage strains.

Validation of the candidate TIR defense systems

For each phage, *E. coli* MG1655 cells with or without the candidate TIR defense systems were infected with 10-fold serially diluted phages to observe plaque formation. The anti-phage activity of the TIR systems was determined based on their ability to inhibit phage plaque formation compared to the empty vector control. A total of 111 different phages isolated in the previous study were used to determine the defense spectrum of each TIR system. *E. coli* MG1655 cells transformed with the empty pSEC1 vector were used as controls. Phage plaque assays were performed as previously described[42]. Briefly, 400 μL of fresh *E. coli* culture was mixed with 5 mL top agar (30 g/L TSB, 7.5 g/L agar, and 50 μg/mL kanamycin) and poured onto a TSA-agar plate. Then, 100 μL of 10-fold serially diluted phages were dropped onto the plate, which were incubated at 37 °C overnight to generate phage plaques. Efficiency of plating (EOP) was calculated by dividing the input pfu by the number of plaques produced by infection of *E. coli* expressing the defense system.

Phylogenetic analysis

The homologous proteins of TIR systems from bacterial and archaeal genomes were downloaded from the NCBI database, and detail information of these proteins were included in the Supplementary Data 2. Protein domains were annotated using HMMER version 33.0[39]. The 'clusthash' option of MMseqs2 (release 6-f5a1c) was used to remove protein redundancies (using the '–min-seq-id 0.9' parameter). Sequences were aligned using MUSCLE (v3.8.1551)[43] with default parameters. Phylogenetic trees were generated using FastTree (v2.1.10)[44] with default parameters and visualized using the online tool iTOL (v.5)[45]. Phylum and genus annotations were based on information provided by the NCBI database (Supplementary Data 2).

Phage infection dynamics in liquid medium

An overnight culture of *E. coli* cells expressing a specific defense system was inoculated into fresh LB medium supplemented with 50 μg/mL kanamycin and incubated at 37 °C with a shaking speed of 200 rpm. *E. coli* cells containing the empty plasmid were used as controls. When the cell density reached OD600 of 0.3, 180 μL of cultures were transferred to each microwell of the 100-well plate, which contained 20 μL of 10-fold serially diluted phages. The plates were immediately placed on an automated growth curve analysis system (Oy Growth Curves Ab Ltd Bioscreen C). Cultures were kept at 37 °C with continuous shaking, and the OD600 was measured every 5 or 10 min for a total duration of 3 h.

Selection of TIR system escape phage mutants through directed evolution

Four hundred μL of fresh cultured *E. coli* MG1655 cells expressing the TIR defense system were mixed with 5 mL top agar (30 g/L TSB, 7.5 g/L agar, and 50 μg/mL kanamycin) and poured onto a 9-cm agar plate. Then, 4 μL of 10-fold serially diluted phages were dropped onto the plate. The plaque formed by high titer phages was picked and transferred to a 1.5 mL Eppendorf tube containing 500 μL of Pi-Mg buffer (26 mM Na_2HPO_4, 68 mM NaCl, 22 mM KH_2PO_4, 1 mM $MgSO_4$, pH 7.5). After 1 h incubation at room temperature with mixing every few minutes, the phages in the supernatant were serially diluted tenfold and dropped onto the plate plated with *E. coli* MG1655 cells as described above. The plate containing *E. coli* MG1655 cells without TIR defense

system was used in parallel as a control. The plaque formed by high titer phages was picked to repeat this cycle until single plaques were formed in the plate containing *E. coli* MG1655 cells expressing the TIR defense system. The single plaque was picked and subjected to several rounds of purification as described previously[42].

To identify the mutations in the escape mutants, genomic DNAs of wild-type or escape mutant phages were isolated by the phenol-chloroform-isoamyl alcohol extraction and sequenced. Briefly, the phage pellet was first digested with 100 μL 10% SDS and 100 μL 0.5 mol/L EDTA in a 65 °C water bath for 15–20 min. A mixture of phenol: chloroform: isoamyl alcohol (volume ratio of 25:24:1, 10 mL) was added, and the tube was inverted several times until the emulsion formed. After centrifugation at 10,000 × *g* for 10 min at 4 °C, the supernatant was collected and subjected to two more rounds of extraction. The genomic DNAs were precipitated by adding isopropanol. After centrifugation at 12,000 × *g* for 10 min at 4 °C, the genomic DNA pellet was washed twice with 800 μL precooled 70% ethanol and dissolved in 50 μL deionized water. Genomic DNA was sequenced on Illumina, and reads were mapped to the parental phage genome using bwa[46] and per-position coverage was obtained using SAMtools[47] depth. Pilon version 1.24 was used to correct errors in the assembled whole genome sequence to improve accuracy. Mutations were identified using Geneious Prime with the WT phage genome sequence as a reference.

Toxicity assays of phage genes in the presence of the TIR systems

The mutant genes of the escape phage (Supplementary Table 2) were cloned into the pBAD expression vector under the control of an arabinose-inducible promoter. The corresponding wild-type genes were also cloned and used as controls. The expression plasmid was co-transformed with the TIR system or empty vector (EV) into *E. coli* MG1655. To determine the toxicity of phage genes using solid media, 400 μL of fresh *E. coli* MG1655 culture was pelleted by centrifugation at 5000 g for 5 min, washed with 400 μL phosphate-buffered saline (PBS), and resuspended with 400 μL PBS. Four μL of 10-fold serially diluted cultures were then spotted on a TSA-agar plate with or without 0.2% arabinose and incubated overnight at 37 °C for bacterial plaque formation.

Electron microscopy analysis of phages

Phages were propagated and purified as described previously[42]. Briefly, log-phase *E. coli* MG1655 cells (~2×10^8 cells/mL) grown on LB medium were infected with phages at an MOI of 0.05. After 3–5 h incubation at 200 rpm at 37 °C, the culture was centrifuged at 8000 × *g* for 10 min, and the phages in the supernatant were collected by centrifugation at 30,000 × *g* for 50 min. The phages were resuspended with 2 mL of Pi-Mg buffer and further purified by CsCl step density gradient centrifugation at 148,000 × *g* for 1 h at 4 °C. Finally, the phages were dialyzed against dialysis solution I (10 mM Tris pH 8.0, 200 mM NaCl, and 5 mM $MgCl_2$) for 5 h followed by dialysis solution II (10 mM Tris pH 8.0, 50 mM NaCl, and 5 mM $MgCl_2$) for overnight at 4 °C. The purified phages were observed by transmission electron microscopy (TEM).

One-step phage growth

One-step growth experiments were carried out as described previously[42]. Briefly, the log phase *E. coli* was infected with phages at various MOIs as indicated in the Fig. 4g and Supplementary Fig. 2. After 7 min incubation at 37 °C, the cultures were then centrifuged for 2 min at 10,000 × *g* and washed twice with fresh LB medium to remove the unabsorbed phages. The infection mixtures were cultured at 37 °C. For T7 phage in the Supplementary Fig. 2, the culture was withdrawn every 5 min for the first 30 min and every 10 min thereafter until 90 min post infection. For remaining phages in the Supplementary Fig. 2, the culture was withdrawn every 10 min until 90 min post infection. For phage #45 in the Fig. 4g, the culture was withdrawn every 5 min for the first 30 min and every 10 min thereafter until 90 min post infection. For phages #41, #75, and #106 in the Fig. 4g, samples were withdrawn every 10 min until 2 h post infection. The *E. coli* cell pellet was resuspended and incubated at 37 °C with a shaking speed of 200 rpm. The samples were centrifuged at 10,000 × *g* for 2 min. The phages in the supernatant were titrated by plaque assay following serial dilutions.

Statistics and reproducibility

Statistical analyses were performed using GraphPad Prism 8 (GraphPad Software Inc., USA). Sample sizes were chosen according to similar experiments published in previous literatures[20,33]. Multi-group comparisons were assessed using one-way analysis of variance (ANOVA). All data are presented as the mean ± SD of three replicates, and $p < 0.05$ was considered statistically significant. All plaque assays were performed in triplicate and the representative images were shown.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data generated in this study are presented in the article or the supplementary information. The GenBank accession numbers for the new TIR systems were listed in the Supplementary Data 2. The pdb files generated using AlphaFold2 in Supplementary Data 3 were deposited in Figshare (https://figshare.com/s/5f8ece9a7268e5ffb254). Source data are provided with this paper.

References

1. Essuman, K., Milbrandt, J., Dangl, J. L. & Nishimura, M. T. Shared TIR enzymatic functions regulate cell death and immunity across the tree of life. *Science* **377**, eabo0001 (2022).
2. Verma, S. & Sowdhamini, R. A genome-wide search of Toll/Interleukin-1 receptor (TIR) domain-containing adapter molecule (TICAM) and their evolutionary divergence from other TIR domain containing proteins. *Biol. Direct* **17**, 24 (2022).
3. Nishimura, M. T. et al. TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **114**, E2053–e2062 (2017).
4. Nimma, S. et al. Structural evolution of TIR-domain signalosomes. *Front. Immunol.* **12**, 784484 (2021).
5. Huang, S. et al. Identification and receptor mechanism of TIR-catalyzed small molecules in plant immunity. *Science* **377**, eabq3297 (2022).
6. Jia, A. et al. TIR-catalyzed ADP-ribosylation reactions produce signaling molecules for plant immunity. *Science* **377**, eabq8180 (2022).
7. Doron, S. et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science (New York, N.Y.)* **359**, https://doi.org/10.1126/science.aar4120 (2018).
8. Tal, N. et al. Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell* **184**, 5728–5739.e5716 (2021).
9. Koopal, B. et al. Short prokaryotic Argonaute systems trigger cell death upon detection of invading DNA. *Cell* **185**, 1471–1486.e1419 (2022).
10. Hogrel, G. et al. Cyclic nucleotide-induced helical structure activates a TIR immune effector. *Nature* **608**, 808–812 (2022).
11. Bayless, A. M. et al. Plant and prokaryotic TIR domains generate distinct cyclic ADPR NADase products. *Sci. Adv.* **9**, eade8487 (2023).
12. Gao, L. A. et al. Prokaryotic innate immunity through pattern recognition of conserved viral proteins. *Science* **377**, eabm4096 (2022).
13. Ofir, G. et al. Antiviral activity of bacterial TIR domains via immune signalling molecules. *Nature* **600**, 116–120 (2021).

14. Morehouse, B. R. et al. Cryo-EM structure of an active bacterial TIR-STING filament complex. *Nature* **608**, 803–807 (2022).
15. Millman, A. et al. Bacterial retrons function in anti-phage defense. *Cell* **183**, 1551–1561 e1512 (2020).
16. Manik, M. K. et al. Cyclic ADP ribose isomers: production, chemical structures, and immune signaling. *Science* **377**, eadc8969 (2022).
17. Tamulaitiene, G. et al. Activation of Thoeris antiviral system via SIR2 effector filament assembly. *Nature* https://doi.org/10.1038/s41586-024-07092-x (2024).
18. Morehouse, B. R. et al. STING cyclic dinucleotide sensing originated in bacteria. *Nature* **586**, 429–433 (2020).
19. Gao, L. et al. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020).
20. Rousset, F. et al. Phages and their satellites encode hotspots of antiviral systems. *Cell Host Microbe* **30**, 740–753.e745 (2022).
21. Wang, S. et al. Landscape of new nuclease-containing antiphage systems in Escherichia coli and the counterdefense roles of bacteriophage T4 genome modifications. *J. Virol.* **97**, e0059923 (2023).
22. Essuman, K. et al. TIR domain proteins are an ancient family of NAD(+)-consuming enzymes. *Curr. Biol.* **28**, 421–430.e424 (2018).
23. Ka, D., Oh, H., Park, E., Kim, J. H. & Bae, E. Structural and functional evidence of bacterial antiphage protection by Thoeris defense system via NAD(+) degradation. *Nat. Commun.* **11**, 2816 (2020).
24. Sabonis, D. et al. TIR domains produce histidine-ADPR conjugates as immune signaling molecules in bacteria. 2024.2001.2003.573942, https://doi.org/10.1101/2024.01.03.573942 %J bioRxiv (2024).
25. Leipe, D. D., Koonin, E. V. & Aravind, L. STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J. Mol. Biol.* **343**, 1–28 (2004).
26. Georjon, H. & Bernheim, A. The highly diverse antiphage defence systems of bacteria. *Nat. Rev. Microbiol.* **21**, 686–700 (2023).
27. Weiss, H. J. & O'Neill, L. A. J. Of flies and men-the discovery of TLRs. *Cells* **11** https://doi.org/10.3390/cells11193127 (2022).
28. Nijland, R., Hofland, T. & van Strijp, J. A. Recognition of LPS by TLR4: potential for anti-inflammatory therapies. *Mar. Drugs* **12**, 4260–4273 (2014).
29. Cramer, P. AlphaFold2 and the future of structural biology. *Nat. Struct. Mol. Biol.* **28**, 704–705 (2021).
30. Lind, N. A., Rael, V. E., Pestal, K., Liu, B. & Barton, G. M. Regulation of the nucleic acid-sensing Toll-like receptors. *Nat. Rev. Immunol.* **22**, 224–235 (2022).
31. Fitzgerald, K. A. & Kagan, J. C. Toll-like receptors and the control of immunity. *Cell* **180**, 1044–1066 (2020).
32. De Palmenaer, D., Siguier, P. & Mahillon, J. IS4 family goes genomic. *BMC Evolut. Biol.* **8**, 18 (2008).
33. Hobbs, S. J. et al. Phage anti-CBASS and anti-Pycsar nucleases subvert bacterial immunity. *Nature* **605**, 522–526 (2022).
34. Akira, S., Takeda, K. & Kaisho, T. Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat. Immunol.* **2**, 675–680 (2001).
35. Leulier, F. & Lemaitre, B. Toll-like receptors–taking an evolutionary approach. *Nat. Rev. Genet.* **9**, 165–178 (2008).
36. Lee, C. C., Avalos, A. M. & Ploegh, H. L. Accessory molecules for Toll-like receptors and their function. *Nat. Rev. Immunol.* **12**, 168–179 (2012).
37. Huiting, E. et al. Bacteriophages inhibit and evade cGAS-like immune function in bacteria. *Cell* **186**, 864–876.e821 (2023).
38. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
39. Potter, S. C. et al. HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–w204 (2018).
40. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).
41. Hildebrand, A., Remmert, M., Biegert, A. & Söding, J. Fast and accurate automatic structure prediction with HHpred. *Proteins* **77**, 128–132 (2009).
42. Tao, P., Wu, X. & Rao, V. Unexpected evolutionary benefit to phages imparted by bacterial CRISPR-Cas9. *Sci. Adv.* **4**, eaar4134 (2018).
43. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinforma.* **5**, 113 (2004).
44. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
45. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–w296 (2021).
46. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
47. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, https://doi.org/10.1093/gigascience/giab008 (2021).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant no. 32170094 to P.T.), the Major Project of Hubei Hongshan Laboratory (grant no. 2022hszd023 to P.T.), Fundamental Research Funds for the Central Universities (grant no. 2662022DKYJ003 to P.T.), Natural Science Foundation of Hubei Province (grant no. 2021CFA016 to P.T.), and the earmarked fund for CARS-41.

Author contributions

P.T. and S.W. designed the experiments. S.W., S.K., H.S., E.S., M.L., Y.L., Z.X., X.Z., X.W. and J.H. conducted the experiments. S.W. analyzed the date. P.T. and C.T. supervised the study. P.T. and S.W. wrote the manuscript. P.T., V.B.R., T.Z. and C.T. revised and polished the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-51738-3.

Correspondence and requests for materials should be addressed to Pan Tao.

Peer review information *Nature Communications* thanks Giuseppina Mariano and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.