

<https://doi.org/10.1038/s42003-025-07459-4>

# In silico characterization of defense system hotspots in *Acinetobacter* spp.

Check for updates

Wenjing Yi<sup>1</sup>, Ning Zhu<sup>1</sup>, Zhihan Peng<sup>2</sup>, Xiao Chu<sup>2</sup>, Haotian Sun<sup>2</sup>, Lei Song<sup>2</sup>, Zhimin Guo<sup>3</sup>, Arnab Pain<sup>4</sup>,  
Zhaoqing Luo<sup>5</sup> & Qingtian Guan<sup>1</sup> ✉

The bacteria-phage arm race drives the evolution of diverse bacterial defenses. This study identifies and characterizes the defense hotspots in *Acinetobacter baumannii* using a reference-free approach. Among 4383 high-quality genomes, we found a total of 17,430 phage defense systems and with 54.54% concentrated in 21 hotspots. These hotspots exhibit distinct preferences for different defense systems, and co-occurrence patterns suggest synergistic interactions. Additionally, the mobile genetic elements are abundant around these hotspots, likely facilitating horizontal transfer and evolution of defense systems. The number of hotspots increases in species phylogenetically closer to *Acinetobacter baumannii*, but the number of defense systems per hotspot varies due to particular selective pressures. These findings provide critical insights into the genetic organization of phage defense systems, contributing to a broader understanding of bacterial immunity and the evolutionary dynamics that shape *Acinetobacter* genomes. This knowledge lays the foundation for developing targeted interventions to combat antibiotic resistance *Acinetobacter baumannii*.

Bacteria-phage interactions are critical to the microbial ecology and evolution of microbial communities. As viruses that infect bacteria, bacteriophages impose strong selective pressure on bacterial communities, which leads to the evolution of complex bacterial defense mechanisms. These defense systems include restriction-modification (RM) systems<sup>1</sup>, CRISPR-Cas systems<sup>2,3</sup>, abortive infection (Abi) systems<sup>4</sup>, toxin-antitoxin (TA) systems<sup>5</sup> and systems with unknown mechanisms<sup>6</sup>. New bacterial anti-phage defense systems are continued to be discovered and the diversity and complexity of the systems are being revealed. For example, the CBASS system<sup>7</sup>, is able to detect phage infection by using cyclic oligonucleotides<sup>8</sup> which will signal the presence of phage DNA and activate a series of defensive actions within the bacterial cell. The SspBCDE system operates through a phosphorothioation (PT) process to distinguish phage non-PT-modified DNA and prevent the phage from replicating and protecting the bacterial host<sup>9</sup>.

It is well-known that defense systems are non-uniformly distributed within bacterial genomes, and they often cluster in specific regions<sup>10</sup>, which are referred to as defense islands, are characterized by the co-localization of defense systems and mobile genetic elements (MGEs)<sup>11–13</sup>. These islands encompass various defense strategies, including those targeting phages.

Within defense islands, certain regions show significant enrichment specifically in phage defense systems, which are referred to as “phage defense hotspots”. Unlike defense islands, hotspots as defined in our study may or may not be associated with MGEs, reflecting their broader genomic context. Several studies have leveraged the concept of “guilt by association” to uncover novel defense systems by analyzing the non-uniform distribution of defense systems<sup>6,14,15</sup>. Moreover, the nature and significance of phage defense hotspots have been identified and characterized in different bacteria<sup>16</sup>. For example, a recent study on the hotspots of *Escherichia coli* identified 41 integration sites for phage defense systems. Most of these regions are highly variable and contain a mosaic of different defense mechanisms. It was found that a large portion of the *E. coli* pan-immune system is carried on the MGEs<sup>17</sup>. In addition, research by Vassallo et al. revealed that many of the defense systems in these hotspots are not yet characterized, with most genes annotated as hypothetical protein<sup>18</sup>. Rousset et al. discovered that P2-like phage and P4-like satellite families serve as large reservoirs of anti-phage systems within *E. coli*<sup>19</sup>, and they described an abortive infection system discovered in these hotspots. Studies of defense systems in *Vibrio* genomes have revealed that phage defense systems are often located within genomic regions associated with MGEs<sup>20</sup>, and exhibit rapid evolutionary

<sup>1</sup>Bioinformatics Laboratory, Infectious Diseases and Pathogen Biology Center, The First Hospital of Jilin University, Changchun, China. <sup>2</sup>Department of Respiratory Medicine, Infectious Diseases and Pathogen Biology Center, State Key Laboratory for Diagnosis and Treatment of Severe Zoonotic Infectious Diseases, Key Laboratory for Zoonosis Research of the Ministry of Education, The First Hospital of Jilin University, Changchun, China. <sup>3</sup>Department of Laboratory Medicine, Infectious Diseases and Pathogen Biology Center, The First Hospital of Jilin University, Changchun, China. <sup>4</sup>Pathogen Genomics Laboratory, Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, 23955-6900, Jeddah, Makkah, Saudi Arabia. <sup>5</sup>Department of Biological Sciences, Purdue University, West Lafayette, IN, 47907, USA. ✉e-mail: [qingtian\\_guan@jlu.edu.cn](mailto:qingtian_guan@jlu.edu.cn)

turnover, causing differential phage susceptibility among clonal bacterial strains despite invariant phage receptors. In contrast to MGEs transferred hotspots seen in *Vibrio* and *E. coli*, *Pseudomonas aeruginosa* features core defense hotspots (cDHS) with conserved genomic markers flanking immunity regions, which the cDHS regions are flanked by conserved genomic markers, which can reliably predict the presence of immune loci without an obvious single mobilizing mechanism<sup>21</sup>.

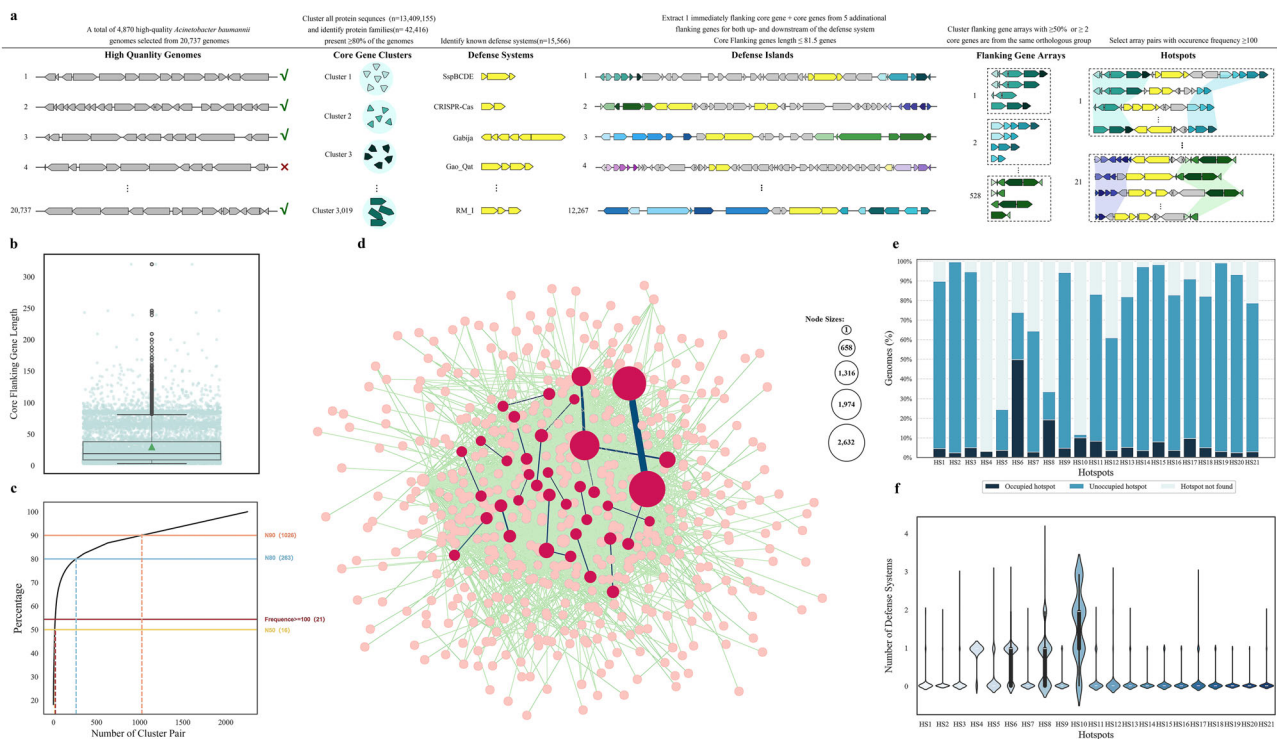
*A. baumannii* caused significant morbidity and mortality<sup>22</sup> in clinical settings and its persistence could partially be explained by the genomic plasticity. The growing interest in phage therapy as a potential treatment for pathogen infections underscores the importance of understanding the interactions between *A. baumannii* and phages<sup>23</sup>. In this study, we obtained 4870 high-quality genomes from a collection of 23,737 raw genomes, and identified 17,430 phage defense systems within the genomes. We found 21 regions that frequently co-occur with multiple defense systems, defining them as hotspots. These hotspots were further analyzed to understand the distribution and abundance of various defense systems, as well as their co-occurrence or mutually-exclusive patterns. Mapping defense system hotspots in *A. baumannii* provides insights into its genetic organization and adaptive responses. The knowledge will be helpful in developing targeted interventions against antibiotic resistance and phage susceptibility in *A. baumannii*. Furthermore, our findings contribute to the broader understanding of bacterial immunity and the evolutionary pressures shaping bacterial genomes in response to phage threats.

**Results**

**Identification of the defense system hotspots within *A. baumannii* genomes**

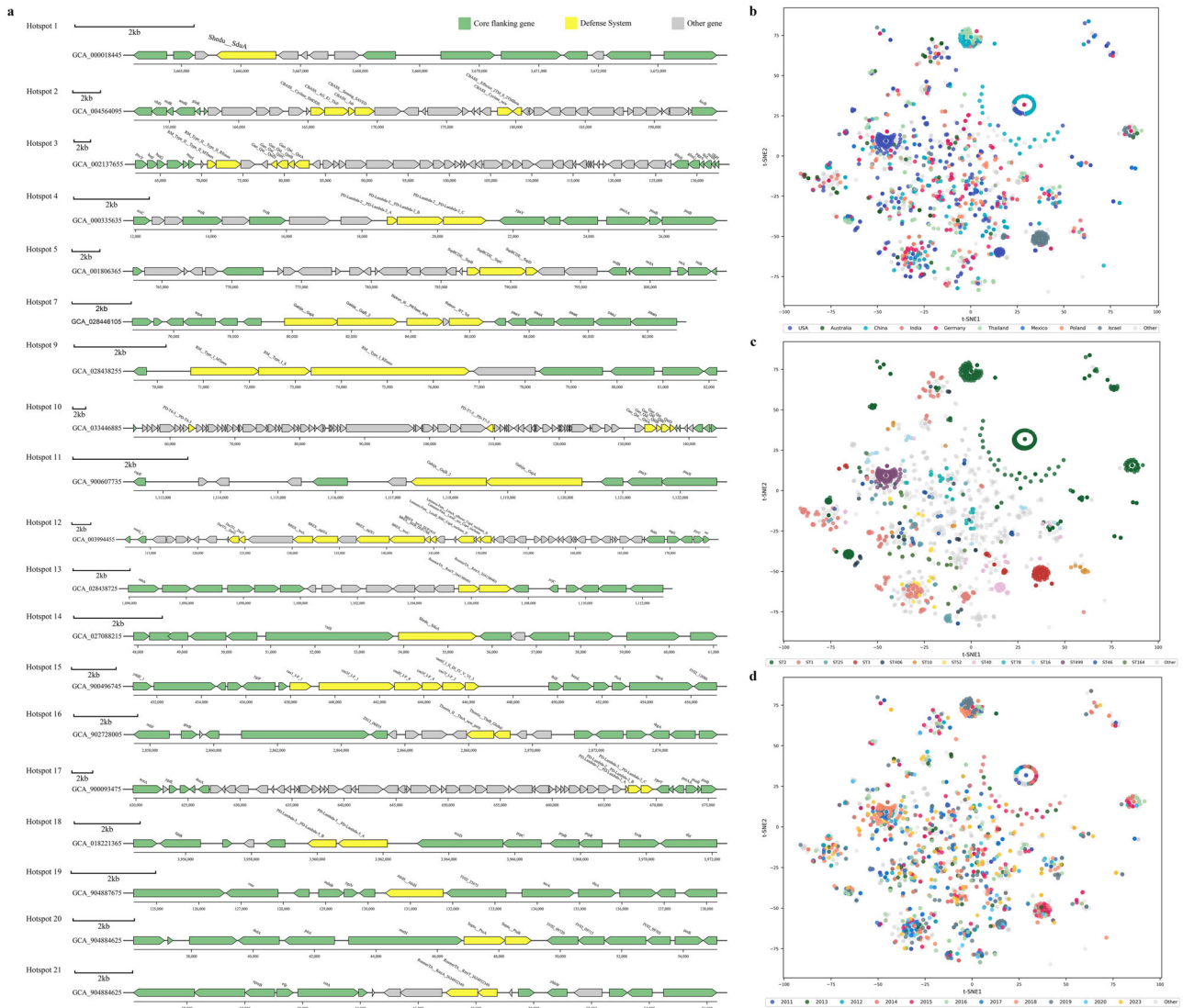
We have developed a “reference-free” pipeline (Fig. 1a), which incorporates all genome data and avoids bias associated with relying on a single reference genome. To identify and select hotspots within *A. baumannii* genomes, the cutoff value for core gene distance was determined by a boxplot analysis and 81.50 genes (Q3 + 1.50×IQR) were selected to provide a robust method to identify meaningful hotspots without the influence of anomalous data points (Fig. 1b). A total of 12,267 core flanking genes array pairs (the gene arrays surrounding the upstream and downstream of the core flanking genes) and 15,566 defense systems were identified from 4387 genomes (Supplementary Data 1, Supplementary Data 2). On average, each genome contains ~3.54 known defense systems, while each array pair carries an average of 1.27 defense systems.

The collective distribution of the defense systems reveals a steep initial rise (Fig. 1c) which indicates a significant proportion of the identified defense systems is captured by the first few array pairs. The curve approaches a plateau as more array pairs are being considered, suggesting the number of flanking genes surrounding the defense systems is becoming more diverse. Hence, we focused on the arrays that co-occurred more than 100 times, which are defined as defense system-carrying hotspots. A total of 54.43% of all detected defense systems were concentrated within these 21 hotspots (HS). This substantial concentration underscores the significance



**Fig. 1 | Identification and characterization of defense islands in *A. baumannii*.** **a** Schematic of the defense island identification method used in this study. **b** Distribution of distances between core genes across *A. baumannii* genomes. The boxplot represents the distribution of distances for all core genes between each pair of genomes in this study. The red diamonds indicate outliers, and the mean distance is represented by the green triangle. **c** Cumulative defense system distribution across array pairs. Number of defense systems in array pairs (cumulative) with respect to the number of arrays on the X-axis and Y-axis, respectively. Horizontal lines denote the cumulative number of defense systems at 50%, 54.43%, 80% or 90% of the total defense systems are covered, and vertical dashed line depicts minimum cluster pairs needed to cover at least 50%, 54.43%, 80% or 90% defense system. **d** Network of core gene arrays around defense systems in *A. baumannii*. This network diagram illustrates the connectivity of different core gene arrays around phage defense systems in

*A. baumannii*. Each circle represents a core gene array, with the size of the circle indicating the frequency of occurrence of that array in the genome. The thickness of the connecting lines represents the number of times two core gene arrays are connected. Core gene arrays that are connected more than 100 times are highlighted in cerise. **e** The bar chart illustrates the distribution of 21 defense system hotspots across 4,870 *A. baumannii* genomes. Each bar represents a hotspot, showing the proportion of genomes where the hotspot is occupied by defense systems (dark blue), unoccupied (blue), or not found (light blue). **f** The violin plot displays the number of defense systems present in each of the 21 hotspots. Each violin represents a hotspot, showing the distribution and frequency of defense systems within that hotspot. The width of each violin at different values indicates the density of genomes with the corresponding number of defense systems in that hotspot.



**Fig. 2 | Genomic loci and correlation analysis of defense systems in *A. baumannii*.** **a** Genomic loci of the 21 identified hotspots in *A. baumannii*. Each panel represents one hotspot, exemplified by a specific genome. The scale bar above each genome indicates the size of the genomic region. Different colors represent various types of genes within the loci: green denotes core flanking genes, yellow represents defense systems within the hotspots, and gray indicates other genes. t-SNE plots showing the

correlation between defense systems and **b** geographic location, **c** sequence type (ST), and **d** isolation year. Each point represents a genome, color-coded based on its metadata category. Categories with fewer data points (Sequence Type <100 occurrences, Country <200 occurrences, Year <250 occurrences) have been grouped into “Other” to improve readability.

of specific genomic regions in harboring the defense mechanisms (Fig. 1d, Supplementary Data 3).

According to our analysis, 17 out of the 21 hotspots were found in more than 60% of the *A. baumannii* genomes (Fig. 1e, Supplementary Data 4). Notably, HS19 is widespread in more than 99% of the genomes. Some hotspots, like HS4, HS8, and HS10, constantly contain phage defense systems even though the frequency is relatively low. For example, HS4 was found in only 3% of the genomes, but over 99% of these occurrences included defense systems. A majority of hotspots have a low density of genomes with defense systems (Fig. 1f), which reflects these hotspots experience less selective pressure from phage attacks. Conversely, HS10 stands out as an exception who frequently contains multiple defense systems with the hotspots.

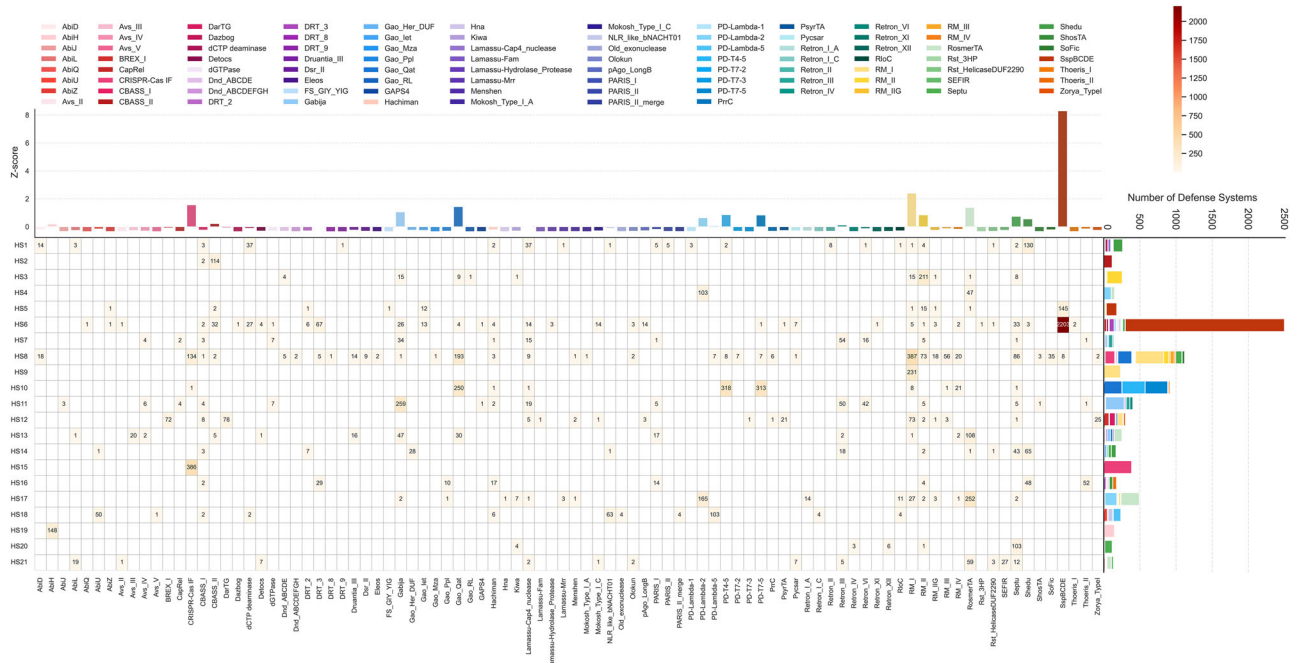
### Characterization and functional analysis of defense systems in *A. baumannii* hotspots

The distribution of core flanking genes and the associated defense systems (Fig. 2a) varies across different hotspots. The defense systems are conserved

at the flanking genes (marked in green) but differ in terms of the defense systems they carry (marked in yellow), contributing to the genetic diversity in *A. baumannii*. The diversity and distribution of the defense systems across different hotspots reflect the adaptive strategy of *A. baumannii* to survive under particular environmental conditions. To investigate the relationship between defense system composition/distribution and various bacterial characteristics, t-SNE analysis was performed based on geographic location, sequence type (ST), or year of isolation (Fig. 2b–d, Supplementary Data 5). The t-SNE plots show that the most distinct clustering is observed with sequence types, which suggests the distribution and composition of defense systems in *A. baumannii* are more closely associated with the bacterial genomic background than with the geographical origin or time of isolation.

Our analysis demonstrates that specific COG categories are more prevalent in the phage defense system hotspot regions of *A. baumannii* compared to the background (Supplementary Fig. 2a, Supplementary Data 6). These categories include genes for amino acid and nucleotide transport and metabolism, replication, recombination, repair, and defense





**Fig. 3 | Distribution and abundance of phage defense systems in *A. baumannii* across 21 hotspots.** The distribution and abundance of phage defense systems across 21 identified hotspots in *A. baumannii* are shown as a heatmap. The columns are different phage defense systems, and the rows refer to specific hotspots (HS1-HS21). The numbers in cells show how many defense systems are located at the given hotspot. This is achieved by intensity encoding the color from light red (low count)

to dark red (high count), where darker shades indicate a higher density of defense systems. A plot above the heatmap shows Z-scores of each defense system as a representation of how abundant this particular system is across all hotspots. On the right of the heatmap is a bar plot, which tells us the cumulative count for defense system within each hotspot.

mechanisms, indicating a strong presence of essential metabolic and regulatory genes. The KEGG pathway enrichment analysis reveals significant enrichment of pathways related to amino acid metabolism (e.g., seleno-compound metabolism, taurine and hypotaurine metabolism), nucleotide metabolism (e.g., pyrimidine metabolism, DNA replication, mismatch repair), and protein metabolism and signal transduction (e.g., MAPK signaling pathway, two-component system). Pathways involved in defense mechanisms (e.g., beta-lactam resistance, biosynthesis of ansamycins) were also enriched (Supplementary Fig. 2b). GO analysis (Supplementary Fig. 2c) provided further insights by identifying enriched biological processes such as propionate catabolic processes and amino sugar biosynthesis. Molecular functions like tRNA dihydrouridine synthase activity, although not directly linked to phage resistance, play a critical role in maintaining protein synthesis and cellular function under stress conditions. In addition, the enrichment of cellular components associated with the cytoplasmic side of the plasma membrane suggests that these regions are integral to maintaining membrane integrity.

**Distribution and abundance of phage defense systems in hotspots**

The distribution and prevalence of various phage defense systems in *A. baumannii* were analyzed across 21 identified hotspots (Supplementary Data 7). The quantitative data (Fig. 3 and Supplementary Fig. 3) show the frequency of each system within these hotspots. The SspBCDE system demonstrated significant prevalence, particularly in HS6, where it was identified 2203 times. The SspBCDE system, an anti-phage defense mechanism, is prevalent across various bacterial genera (Supplementary Fig. 4), including *Pseudomonas*, *Enterobacter*, *Escherichia*, etc. Within the phylogenetic tree, the genus *Acinetobacter* occupies bacterial branches near the central node, indicating complex and intertwined evolutionary relationships with multiple other genera. The *Acinetobacter* branches are not monophyletic, suggesting independent divergence events. High bootstrap values support the proximity of

*Acinetobacter* branches to those of other genera, such as *Duganella* and *Noviherbaspirillum*.

The system also exhibited high counts in HS5, with 145 instances, respectively, indicating its essential role in phage resistance across different environments within *A. baumannii*. Similarly, RM\_I and RM\_II systems were widely distributed in different hotspots, with RM\_I notably abundant in HS8 and HS9, showing 387 and 231 instances, respectively. This widespread occurrence suggests that RM systems are fundamental components of the bacterial defense strategy, providing a primary line of defense by recognizing and cutting foreign DNA.

Certain systems exhibit a high specificity to particular hotspot. For example, the AbiH system was exclusively found in HS19 (148 instances), suggesting a specialized defense role in this specific hotspot. Similarly, PD-T4-5 and PD-T7-5 are dominant in HS10, as well as Gabijia is dominant in HS11. While some hotspots displayed diverse defense systems, for example, in HS6, several systems, including SspBCDE, DRT\_3, and Retron\_IV, showed high frequencies, suggesting potential synergistic interactions that enhance bacterial defense. Similarly, HS8 and HS17 had high counts of multiple systems, indicating robust and multifaceted defense strategies within the same hotspots.

The alignment of these loci demonstrates the considerable variability in gene content and organization within the hotspots across different strains. Notably, the defense systems in HS6 (Fig. 4a, Supplementary Data 8) and HS8 show diverse arrangements (Fig. 4b, Supplementary Data 9), suggesting that these regions are subject to significant evolutionary pressure and adaptation. This diversity may contribute to the overall genomic plasticity and adaptability of *A. baumannii*, allowing it to survive in various environmental conditions and evade threats, including bacteriophages. The observed variability in the defense systems within these hotspots could be attributed to several factors, including horizontal gene transfer (HGT), selective pressure from bacteriophage exposure, genomic plasticity, recombination events, and the dynamic nature of the hotspots themselves. These factors collectively enable *A. baumannii* to rapidly adapt to different



**Fig. 4 | Representative genomic loci of HS6 and HS8 in *A. baumannii*.** **a** HS6 loci from various *A. baumannii* strains. **b** HS8 loci from various *A. baumannii* strains. Each of the HS6 and HS8 are linked with various defense systems. Green indicates

the flanking core genes, yellow indicates defense-related genes, and gray indicates other regional genes. The alignment of these loci elucidates the adaptability and consistency of the loci.

threats, leading to the retention and diversification of effective defense systems.

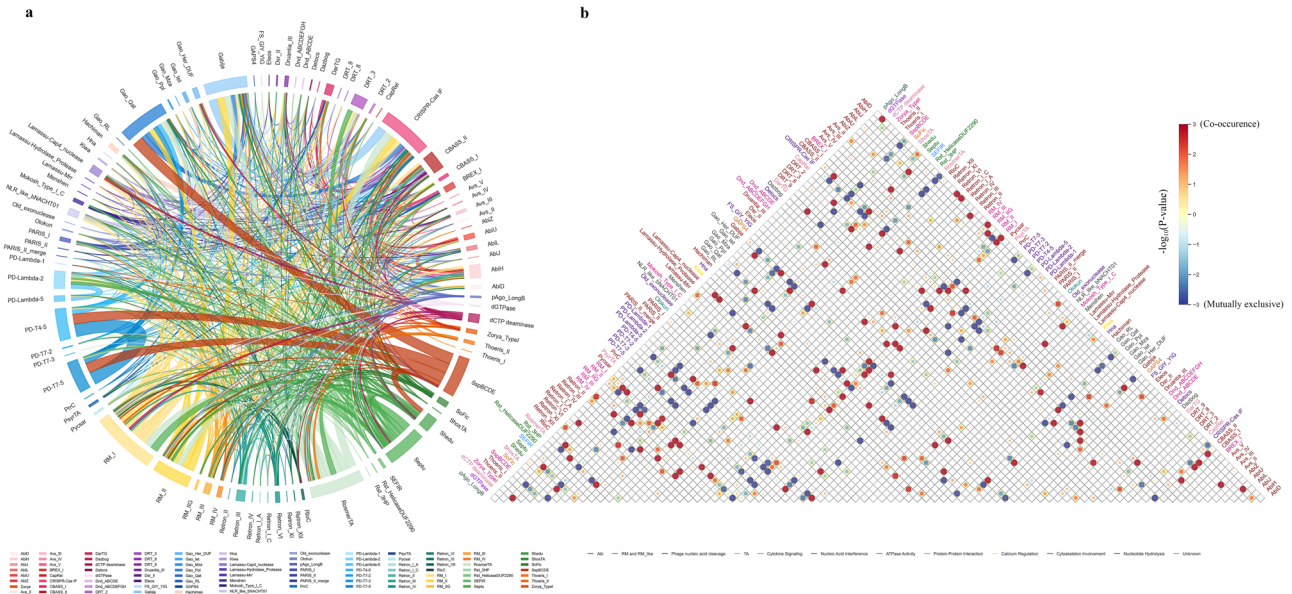
**Co-occurrence patterns of phage defense systems in *A. baumannii***

We further applied a phage defense system co-occurrence analysis in multiple bacterial genomes, which provides different co-occurrence patterns for different systems. Remarkably, many defense systems demonstrate strong co-occurrence with multiple other defense systems (Supplementary Data 10). Such defense systems include SspBCDE, RM\_I, CRISPR-Cas type I-F<sup>24</sup>, Gabijia<sup>25</sup> and Gao\_Qat<sup>6</sup> exponentially appear alongside a large number of other defense system (Fig. 5a, b). The frequent presence of both together suggests an important role for the two in supplying versatile and reliable defense traits. Interestingly, some defense systems like PD-Lambda-5 and PD-Lambda-2<sup>18</sup> show a clear tendency of coexistence with each other. Such a result might hint toward dedicated functional linkage or co-regulation used to target specific phage threats, and/or environmental conditions. On the other hand, three defense systems (Lamassu-Fam<sup>25</sup>, AbiQ<sup>26</sup> and Mokosh type I-A<sup>27</sup>) were never found to co-occur with any of those in this study. This is probably due to the fact that these defense systems were found in very few instances, being only one of each defense system present in our dataset—as such, it seems more reasonable to consider their isolation an issue related to sampling bias than a true scarcity of co-occurrence.

Beyond examining the co-occurrence events within each genome, we also investigated co-occurrence events within individual hotspots. Among all possible associations (3916) between the 89 defense systems detected in *Acinetobacter*, different phage defense systems are positively correlated in 203 cases (5.2%) and negatively correlated in 200 cases (5.1%) (Supplementary Fig. 5b, Supplementary Data 11). For instance, PD-T4-5 and PD-T7-5<sup>18</sup> exhibited a significant co-occurrence pattern within HS10 (Supplementary Fig. 5a). PD systems, often involved in phage defense through mechanisms such as TA systems, may collaborate to enhance bacterial survival against phages in specific contexts. Systems like DarTG<sup>28</sup> and BREX\_I<sup>14</sup>, dCTP deaminase<sup>29</sup> and Lamassu\_Cap4\_nuclease<sup>30</sup>, and Septu<sup>31</sup> and RM\_I frequently co-occur within the same hotspots, suggesting potential cooperative or layered defense strategies, indicating that the clustering of these defense genes is not random but likely a result of evolutionary pressures and genomic adaptations.

**Distribution and abundance of mobile genetic elements in hotspots and surrounding regions**

We investigated MGEs inside and outside of 21 genomic hotspots identified in *A. baumannii* genomes, which highlight a remarkable plasticity. Distribution of MGEs was assessed both within hotspots (Fig. 6a, Supplementary Data 12) and in the upstream and downstream



**Fig. 5 | Co-occurrence and correlation patterns of phage defense systems in *A. baumannii*.** **a** The co-occurrence patterns of identified phage defense systems within each genome of *A. baumannii* are shown by circos plot. Each segment represents a different phage defense system, and they are colored for ease of identification. The connecting lines (chords) express how often these systems occurred as paired at the hotspots between all the segments of this combination. The weight corresponds to the likelihoods, with heavier lines representing stronger associations. **b** Correlation

matrix of different phage defense systems. The yellow background indicates significant correlations ( $P < 0.05$ ). The size and color of the circles represent the  $p$ -value of the correlations, with the color scale ranging from blue (mutually exclusive) to red (co-occurrence). Larger circles indicate more significant  $p$ -values. The colors of the defense systems represent different categories of functional mechanisms.

20 kb regions encompassing each hotspot (Fig. 6b, Supplementary Data 13), revealing differing levels of contribution by MGEs. The majority of the hotspots (17/21) exhibit relatively low levels of MGEs content within the hotspots. In contrast, the surrounding regions of these hotspots show a higher percentage of genomes containing MGEs. For example, within HS2, less than 1.2% of the genomes contain MGEs, but the surrounding regions show a 75.0% probability of containing MGEs, with a 37.4% increase in ISs and a 30.0% increase in multiple types of MGEs. Similarly, within HS9, less than 3.7% of the genomes contain MGEs, but the surrounding regions have a 68.3% probability of containing MGEs, predominantly prophages. In particular, certain hotspots, such as HS10, exhibit a high percentage of prophages, accounting for over 90.8% of the genomes analyzed within the hotspot, highlighting its role as a potential MGE integration site.

We found prophages, plasmids and ISs to be the three most common types of MGEs that were located inside or near these regions (69.8% overall). While other types of MGEs, such as transposons and phage satellites are less common in these hotspots but more abundant in their surrounding regions. Around the hotspots such as HS2 or HS12, a large part of genes is related to putative transposons (2.4% and 1.1%), which are more prevalent in the surrounding regions than inside those islands per se. HS4 does not have any MGEs confirmed of, in or near the hotspot. No MGEs were detected in HS4 other than PD-Lambda-2 and RosmerTA<sup>27</sup>, implying localized activity or integration. This analysis further supports that the hotspots are denser regions of MGEs compared to their flanking genomic elements with a high turnover and gene mobility, revealing more dynamics for both within *A. baumannii*.

We also assessed tRNA genes in the hotspots (Fig. 6c, Supplementary Data 14), and their flanking regions displayed significant alterations (Fig. 6d, Supplementary Data 15). The tRNA genes are present within the hotspots only in the HS12 and HS21, but increased in other hotspots surrounding regions of 20 kb, as shown by the data. Enrichment analysis of AMR genes in these hotspots (Fig. 6e, f) highlights the co-localization of pertinent resistance determinants, specifically those for the AdeABC efflux pump such as *adeA*, *adeB* and its regulators protein-coding gene pair (*adeR/adeS*)

(Supplementary Data 16), thus further stressing their importance to promote bacterial survival under not only phage but also antibiotic selection pressure.

### Hotspot distribution and defense system variability in *Acinetobacter* species

The distribution and analysis of hotspots within *Acinetobacter* species reveal several critical insights into the genomic landscape and defense mechanisms (Fig. 7, Supplementary Data 17). The analysis indicates that the identified hotspots are exclusive to the *Acinetobacter* genus among 251 genera across the bacteria domain, suggesting the unique genomic characteristic of *Acinetobacter* species. The number of defense systems per genome increases in species phylogenetically closer to *A. baumannii*, which indicates a strong evolutionary adaptation within closely related species, potentially driven by similar environmental pressures and selective forces. Although the number of defense systems per genome is higher in species closer to *A. baumannii*, the number of defense systems per hotspot does not follow the same trend. The highest numbers of defense systems per hotspot are found in species such as *Acinetobacter thutiu* (1.00), *Acinetobacter bouvetii* (0.44), *Acinetobacter vivianii* (0.30), *Acinetobacter beijerinckii* (0.20) and *Acinetobacter venetianus* (0.18) while this number in *A. baumannii* is 0.12. These species do not share the monophyletic group with *A. baumannii*, indicating that these defense systems might be horizontally transferred within the genus, and the number of defense systems in each hotspot likely reflects the biological pressures each species faces, as these species are predominantly isolated from environmental samples rather than clinical settings, which may expose them to higher levels of viral predation<sup>32–36</sup>.

Although HS6 is the most prevalent hotspot in *A. baumannii* (Fig. 3), HS2 is the most common among other species among the genus. Hotspots such as HS9, HS15, and HS19, which carry only one defense system in *A. baumannii*, are consistent across different species, indicating potential conserved functions or selective advantages conferred by these single defense systems. This comprehensive analysis underscores the conservation of hotspots within the *Acinetobacter* genus, as they are absent in other genera, and also highlights their dynamic nature within *Acinetobacter*



**Fig. 6 | Distribution of MGEs, tRNA, and AMR genes in identified hotspots within/around *A. baumannii* genomes.** **a** Bar charts displaying the frequency of MGEs in each individual hotspot and **b** their surrounding upstream or downstream 20 kb regions. **c** Bar chart showing the distribution of tRNA genes within or **d** surrounding hotspots (upstream or downstream 20 kb regions of particular

hotspot) regions. **e** Bubble plot showing the enrichment of AMR genes within the identified hotspots. The size of each bubble indicates the percentage of genomes with this AMR gene in the set, and color represents enrichment significance ( $\log_{10}(1/FDR)$ ). **f** The top most significantly enriched AMR genes within the hotspots, as visualized by a bar plot of the  $-\log_{10}$  (adjusted *p*-value).

species, showing significant variation in hotspot distribution and defense system content.

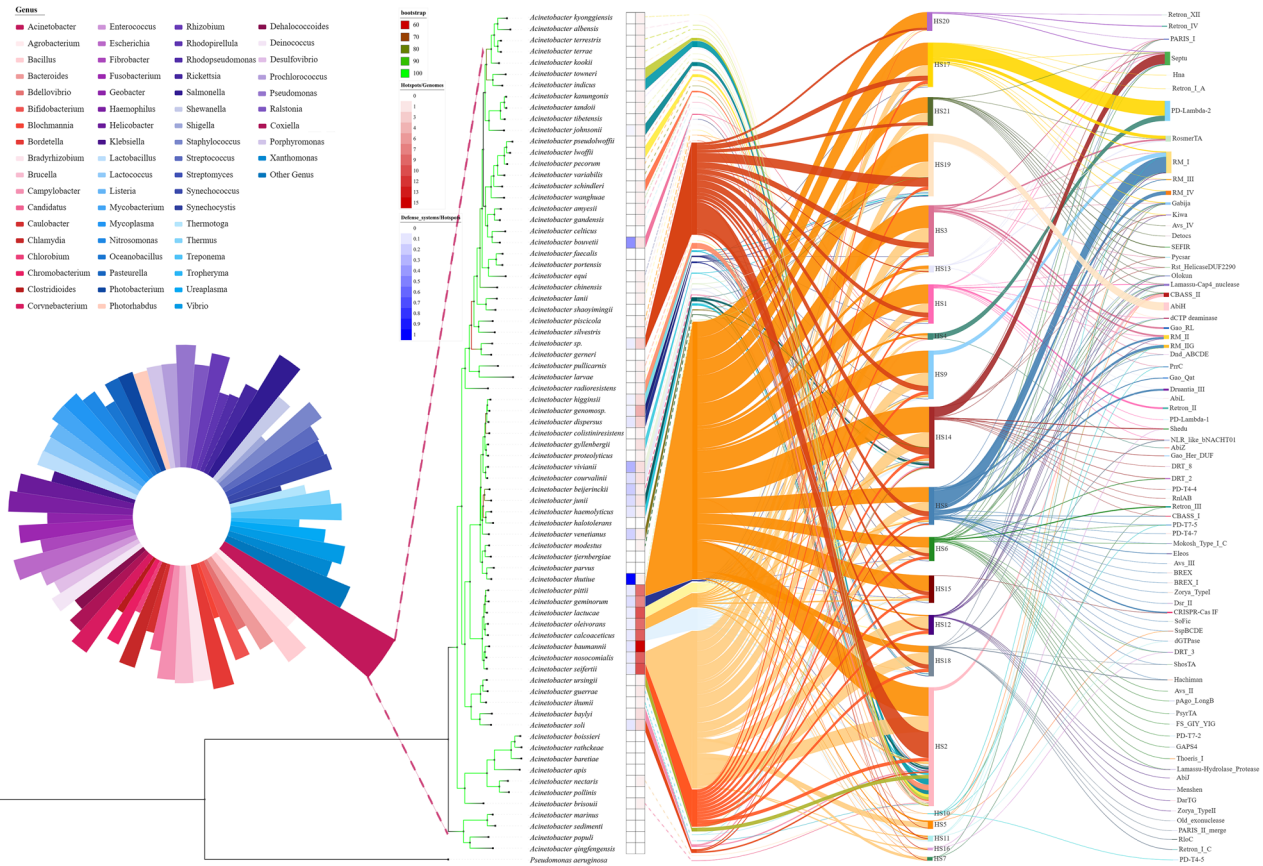
### Discussion

Phage–bacteria interactions are fundamental to microbial ecology and generate the evolutionary landscape of bacterial genomes. The application of phage therapy for *Acinetobacter* infections is an increasingly explored area with significant potential to resolve the problem of multidrug-resistant bacterial infection. Several studies highlighted the efficacy and versatility of phage therapy in both preclinical and clinical settings<sup>23,37–39</sup>. The

identification of genomic defense system hotspots provides a deeper understanding of *A. baumannii*'s adaptive strategies against phage attacks, which will enhance the development of phage therapy and ensure effective management of the challenging infections.

Our methodology (Fig. 1a and Supplementary Fig. 1) for identifying defense islands in the *A. baumannii* genome offers several advantages compared to previous studies<sup>17,21</sup>. By not mapping the hotspots to a reference genome (reference-free method), we avoided biases associated with reference genomes, allowing for a more accurate representation of the genetic diversity present in the pangenome. This approach ensures that





**Fig. 7 | Distribution of hotspots across different genera and within *Acinetobacter* species.** This figure shows the distribution of hotspots across *Acinetobacter* genus and different genera across bacteria domain. The Nightingale Rose Chart displays the different genera used in this study, and genera names are located at the top left corner of each chart. Each segment represents a genus (scales are in log10), and the length of each one corresponds to the number of genomes. The phylogenetic tree indicates the evolutionary relationships among the *Acinetobacter* species, and *Pseudomonas aeruginosa* was used as an outgroup. The first column of the

annotation shows the number of hotspots per genome for each *Acinetobacter* species, and the second column of the annotation shows the number of defense systems per hotspot. For simplicity, *A. baumannii* hotspots' information are excluded from the figure. The Sankey diagram shows the hotspots present in different *Acinetobacter* species and the specific defense systems contained within each hotspot. The width of each segment represents the number of genomes. The connections in the Sankey diagram illustrate the distribution and frequency of defense systems across different hotspots within various *Acinetobacter* species.

novel or unique defense systems are not overlooked. We acknowledge the high genomic plasticity of *A. baumannii* by applying a flexible cutoff of 50% for grouping the flanking gene arrays. This flexibility enables the detection of defense systems that may be missed by more rigid cutoffs (Fig. 1). The interesting observation that most hotspots are empty in a majority of genomes raises questions about the dynamics of phage defense system distribution (Fig. 1e, f). The predominance of empty hotspots likely reflects a complex interplay of ecological, evolutionary, and genomic factors.

The presence of conserved core flanking genes around these defense systems likely serve as anchors, ensuring the retention and functionality of defense systems within the genome (Fig. 2). In contrast, the variability observed in other genes within these loci contributes to the genetic diversity of *A. baumannii*, enabling it to adapt to different threats (Fig. 4). This dynamic balance between conservation and variability is a hallmark of genomic regions under strong selective pressure. The specificity of certain defense systems to particular hotspots suggests that these regions are tailored to local environmental conditions or specific phage threats (Fig. 3 and Supplementary Fig. 3). For example, the AbiH<sup>40</sup> system's predominance in a specific hotspot implies a specialized adaptation to phages encountered in that niche. Some systems, like SspBCDE, are widespread and highly prevalent, suggesting the fundamental role in providing robust defense against phages (Supplementary Fig. 4) in *A. baumannii*.

Our integrated analysis of COG, KEGG, and GO data underscores the multifaceted role of defense system hotspots in *A. baumannii* (Supplementary Fig. 2). The enrichment of specific COG categories and KEGG

pathways within these hotspots indicates that they are heavily involved in critical metabolic and genetic regulatory functions. The high proportions of genes related to amino acid and nucleotide transport and metabolism suggest a strong need for rapid synthesis of enzymes and nucleotides, which is essential for both robust cellular defense and the maintenance of genomic stability under phage attack. The presence of enriched transcription and signal transduction pathways within these hotspots further highlights their importance as central hubs for coordinating complex defense responses. These regions likely facilitate rapid cellular adaptations to environmental changes, including phage infections, by modulating gene expression and signaling pathways. This aligns with previous studies that have demonstrated the cooperative and regulatory interactions among various defense systems<sup>41</sup>, supporting the idea that these hotspots are crucial for the integrated management of bacterial immune responses. Moreover, while primarily associated with phage defense, these hotspots also exhibit functional enrichment in pathways related to antibiotic resistance, as indicated by KEGG analysis. Although our study focuses on phage resistance, the overlap with antibiotic resistance pathways, such as those involved in beta-lactam resistance, suggests that these genomic regions may play dual roles in defending against multiple threats, including antimicrobial agents.

The observed co-occurrence patterns of phage defense systems provide intriguing insights into bacterial immunity of *A. baumannii* (Fig. 5). Contrary to the potential expectation that bacteria might avoid using similar defense strategies simultaneously to enhance their overall defense capabilities, our findings reveal significant co-occurrence of certain defense



systems. For instance, AbiU and AbiH, both belonging to the Abi category<sup>4</sup>, significantly co-occur within genomes ( $p$ -value =  $1.56E-47$ ). Similarly, PD-Lambda-5 and PD-Lambda-2, involved in phage nucleic acid cleavage<sup>18</sup>, frequently co-occur ( $p$ -value =  $4.57E-140$ ), as do RM\_II and RM\_III systems<sup>42</sup> ( $p$ -value = 0.04), and Shed<sup>25</sup> and Septu, both cytoskeleton-related defense systems ( $p$ -value = 0.03). These findings suggest that despite belonging to the same functional categories, these systems may provide complementary functions, targeting various stages or mechanisms of phage attacks. Their co-occurrence might also be regulated by similar or complementary controls, allowing for coordinated expression under specific environmental pressures.

While existing studies indicate similar co-occurrence patterns across different bacteria (e.g., Druantia\_III and RM\_I systems significantly co-occurring in *A. baumannii*, *E. coli*, *Enterobacterales*, and *Pseudomonales* genomes, and Gabija and RM\_II significantly co-occurring in *A. baumannii*, *E. coli*, and *Enterobacterales* genomes<sup>43</sup>), our research demonstrates variability in these systems' co-occurrence or mutual exclusivity among different bacterial species. For example, Druantia\_III and RM\_III are mutually exclusive in *E. coli* and *Enterobacterales* but co-occur significantly in *A. baumannii* ( $p$ -value = 0.04). Similar patterns are observed with Hachiman and RM\_I systems ( $p$ -value =  $1.01E-4$ ). These differences suggest that different bacteria might face varied phage pressures in different ecological niches, leading to selective retention or loss of specific defense systems. The role of HGT is also evident, as defense systems can be transferred between bacteria, with different species acquiring these systems through diverse mechanisms and timings. In summary, the co-occurrence and mutual exclusion of phage defense systems within bacterial genomes highlight the sophisticated ecological adaptations and evolutionary mechanisms bacteria employ.

Overall, our analysis indicates the abundance of MGEs in certain hotspots and their flanking sequences. The vast majority of hotspots showed a low level of MGE content with the exception that these were significantly higher in a 20 kb upstream and downstream region surrounding these (Fig. 6a, c). Such a pattern has been known in previous studies and suggests that the regions surrounding genomic hotspots are more susceptible to MGE insertion, presumably due to favorable conditions for integration, such as specific nucleotide sequences or chromosome structures<sup>44</sup>. For example, similar observations have been made in studies on *E. coli*<sup>17</sup> and *V. cholerae*<sup>45</sup>, where the regions flanking genomic hotspots are densely populated with MGEs, thus contributing to genome fluidity, likely facilitating the survival of these bacteria in changing environments. These observations illustrate the crucial role of MGEs in promoting rapid genetic innovation to help secure bacterial populations under selective pressures, such as bacteriophage predation or antibiotic selection<sup>46,47</sup>.

The absence of these hotspots in other genera underscores the conservation of the discovered hotspots within *Acinetobacter* (Fig. 7), indicating a unique genomic characteristic that may be crucial for the genus's adaptability and survival that these hotspots have been maintained through evolutionary pressures. The trend where species closer to *A. baumannii* exhibit a higher number of defense systems per genome may be driven by similar environmental pressures faced by these closely related species, necessitating the acquisition of diverse defense mechanisms to thrive in respective niches<sup>48,49</sup>. In the meantime, the observation that species not phylogenetically close to *A. baumannii* possess the highest number of defense systems per hotspot indicates that HGT and MGEs play a significant role in shaping the genomic landscape across the genus. The presence of these defense systems suggests that whether or not the "cargo" (defense system) is loaded depends on the particular environmental pressures faced by the bacteria. This pattern supports the notion that defense systems can be horizontally transferred within the genus<sup>50</sup>, contributing to the observed genomic plasticity<sup>46,47,51,52</sup>.

In conclusion, the identification of defense system hotspots within *A. baumannii* genomes illustrates the complex and dynamic nature of bacterial adaptation to phage predation. The unique presence of these hotspots in *Acinetobacter* species and their variability influenced by selective pressures

provide a foundation for future research into bacterial immunity. Furthermore, the identification of these hotspots as focal points for diverse defense systems highlights the potential for discovering novel bacterial defense mechanisms.

## Methods

### Genomes collection and filtering

A total of 20,737 *A. baumannii* genomes were initially collected for analysis. These genomes were subjected to stringent quality control using CheckM (v1.2.2)<sup>53</sup>, ensuring completeness >99% and contamination <1%, with genomes containing fewer than 100 contigs being selected. This filtering process resulted in 4870 high-quality genomes for subsequent analysis. This *A. baumannii* dataset represents genomes coming from 54 countries and 261 institutions, covering the years 2007, 2008, and 2010 to 2023. These genomes were further annotated using Prokka (v1.14.6)<sup>54</sup> with default parameters.

We initially used 90% of the total filtered genomes ( $n = 4383$ ) to identify hotspots, and the remaining 10% ( $n = 487$ ) for validation to verify if the same hotspots were identified. The selected 4387 genomes were then analyzed using Cd-hit (v4.8.1)<sup>55</sup> with default parameters to cluster orthologous genes. This process identified 42,416 ortholog groups, which were used to determine the core genome. From the grouped orthologs, core genes were identified based on their presence across the majority of the genomes (80%). This step resulted in the identification of 3019 core gene groups, encompassing 13,409,155 genes. The validation dataset showed that the results were consistent with those obtained from the 90% dataset. Subsequent analyses were conducted using all 4870 genomes.

### Identification of defense system hotspots

A schematic figure of the defense island identification method used in this study is shown (Fig. 1a). Amino acid sequences of genes in all 4870 genomes were submitted to DefenseFinder (v1.2.0)<sup>30,56,57</sup>. This analysis detected a total of 17,430 systems associated with various phage defense systems. For each identified defense system, core flanking genes both upstream and downstream were analyzed within a genomic region of less than 81.5 genes, excluding outliers based on core gene distance. This analysis included one core gene immediately flanking the defense system and core genes from five additional flanking genes. Only genes located on a single contig were considered to ensure genomic continuity, and the core gene arrays were recorded and included in the analysis.

Considering the high genomic plasticity of *A. baumannii*, a relatively flexible criterion was applied to ensure that potential defense system hotspots were not missed due to rigid categorization. Any new gene arrays were compared with existing arrays, and were grouped with the array that fit the criteria: if an array shared more than 50% core gene similarity or contained more than two genes with different arrays, it was grouped with the array having the higher number of core gene matches. Specific technical parameters and procedures were recorded in Supplementary Fig. 1).

The occurrence of each array pair was counted across all genomes analyzed. Array pairs that appeared together more than 100 times were identified, and these frequently co-occurring array pairs were defined as hotspots. Cytoscape (v3.10.1)<sup>58</sup> was used to visualize the resulting network. These hotspots represent regions rich in defense systems within *A. baumannii* genomes.

### Multidimensional analysis of defense system distribution characteristics

To gain insight into the association between the defense system types and time, geographic location or ST classification of the *A. baumannii* genome, we performed a MLST (v2.23.0)<sup>59</sup> analysis on the sequences of the genomes. The individual ST profile matrix and isolation time, location data were then dimensionality reduced using the t-SNE, which would allow us to visualize the distribution of the defense systems.

### Enrichment analysis of antibiotic resistance genes in hotspots

Using RGI (Resistance Gene Identifier) (v6.0.3)<sup>60</sup>, we screened the protein sequences of the genomes for antibiotic resistance genes, filtering out genes in the hotspots that are not part of the defense systems or core genes. We then utilized the GSEAPy package to construct an antibiotic resistance gene set and performed enrichment analysis on the genes in the hotspots, visualizing the results with the dotplot function.

### Functional enrichment analysis

All proteins from hotspots containing defense systems and the remaining genes, along with their COG (Clusters of Orthologous Groups) functional annotations using eggNOG-mapper (v2.1.12)<sup>61</sup>. The number of each COG category (A-Z) in hotspots with defense systems and other genes without defense systems was tallied. For proteins with multiple functions, the quantity of each function was separately counted. The proportion of proteins within defense system hotspots performing a specific function relative to all proteins in these hotspots was calculated. Similarly, the proportion of proteins without defense systems performing the same function, relative to all other proteins, was determined. The data from the statistical analysis were visualized using the seaborn<sup>62</sup> package in Python (v3.11.5).

Protein sequences from 4870 genomes were annotated using eggNOG-mapper. Proteins annotated with KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways and GO (Gene Ontology) were filtered as a dataset, and all proteins within hotspots were selected. The hotspot proteins were compared with the KEGG and GO pathways from 4870 genomes by performing an enrichment analysis in a Python package GSEAPy<sup>63</sup>. In the enrichment analysis, those enriched results with a  $p$ -value  $< 0.01$  were selected by filtering feature in enrichr module of GSEAPy package. The results were plotted using the GSEAPy dotplot and bar plot functions.

### Phylogenetic analysis of the SspBCDE system

Due to the widespread distribution of SspBCDE in *A. baumannii*, we conducted a detailed phylogenetic analysis of the SspBCDE system. We selected the amino acid sequence of the essential protein SspB from the SspBCDE system and performed three iterations of Position-Specific Iterated BLAST (PSI-BLAST)<sup>64</sup> in the NCBI database to identify different matching species' protein sequences for phylogenetic tree construction. First, multiple sequence alignment was performed using MAFFT (v7.525)<sup>65</sup>. The aligned sequences were then trimmed using TrimAl (v1.4.1)<sup>66</sup> to remove poorly aligned regions. Following this, a maximum likelihood phylogenetic tree was constructed using IQ-TREE (v2.1.4\_beta)<sup>67</sup> with 1000 bootstrap replicates, employing the Q.plant+R5 evolutionary model. Finally, the phylogenetic tree was visualized with iTOL (v6.9.1)<sup>68</sup>.

### Identification of mobile genetic elements and tRNAs

Integrative conjugative elements (ICEs), insertion sequences (ISs), and transposons were identified through MobileElementFinder (v1.0.6)<sup>69</sup> by analyzing the nucleotide sequences of the genomes. Only ISs, ICEs, and transposons located within hotspots were considered for further analysis. Prophages and plasmids were identified using geNomad (v1.7)<sup>70</sup> analysis of the nucleotide sequences. A region was classified as a prophage only if geNomad identified it as a provirus with more than one gene matching a phage. The plasmid-associated protein sequences were obtained using geNomad, and the hotspots on the plasmids were identified using the method described in the article. Phage satellites within hotspots were detected using SatelliteFinder (v0.9.1), which analyzed the amino acid sequences of proteins in each hotspot. In cases where two or more types of MGEs were clearly integrated within the same region, the region was recorded as containing multiple MGE types. The tRNA sequences were annotated by Aragorn<sup>71</sup> and the distribution of tRNAs within the hotspots and in the 20 kb region upstream and downstream of the hotspots was analyzed.

### Statistical analysis of defense systems co-occurrence

Defense systems within hotspots were extracted and analyzed for co-occurrence. If two defense systems appeared in the same genome, the number of their co-occurrences was counted. A chord diagram illustrating the frequency of co-occurrence between pairs of defense systems was generated using the chord\_diagram function from the mpl\_chord\_diagram package in Python. To assess the correlation and significance of co-occurrence, Fisher's exact test was employed to calculate the odds ratio (OR) and  $p$ -value. An odds ratio greater than 1 indicates a positive correlation between two defense systems, with the significance of co-occurrence evaluated using the  $\log_{10}(p\text{-value})$ . Conversely, an odds ratio less than 1 indicates a negative correlation, with the significance of mutual exclusivity also assessed using  $\log_{10}(p\text{-value})$ .

### Mapping hotspots to different genera of prokaryotic organisms

We filtered all prokaryotic species from the tree of life provided by iTOL using genomes downloaded from the NCBI Assembly Database, selecting 100 genomes for each species. If a species had fewer than 100 genomes, the remaining genomes were randomly selected from the corresponding genus, resulting in 10,089 genomes from 251 genera in total across the bacteria domain. All genomes from *Acinetobacter* were downloaded, and the same CheckM quality control pipeline was applied. A Python script was used to align hotspots to the genomes of different strains, and DefenseFinder was employed to identify defense systems in each genome. Finally, we performed a statistical analysis of the number of hotspots across different species and the variety of defense systems within these hotspots.

### Statistics and reproducibility

We used the fisher\_exact function from the scipy package in Python to perform Fisher's exact test, specifically employing a two-sided approach, to determine the co-occurrence of different defense systems within the same genome or the same hotspots. The  $p$ -values from the co-occurrence and enrichment analyses were corrected using the FDR function in Python.

No statistical method was used to predetermine the sample size. The genome sequences were subjected to stringent quality control using CheckM (v1.2.2)<sup>53</sup>, ensuring completeness  $>99\%$  and contamination  $<1\%$ , with genomes containing fewer than 100 contigs being selected. The experiments were not randomized.

### Data availability

All data generated in this study, including raw genome sequence files, is available on Github<sup>72</sup> at <https://github.com/Guan-biolab/Phage-Hotspots>. Source data are provided in this paper. Source data can be obtained in Supplementary Data 3–17.

### Code availability

We share Python scripts used in this study via GitHub<sup>72</sup> (<https://github.com/Guan-biolab/Phage-Hotspots>).

Received: 5 October 2024; Accepted: 3 January 2025;

Published online: 10 January 2025

### References

1. Oliveira, P. H., Touchon, M. & Rocha, E. P. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* **42**, 10618–10631 (2014).
2. Ofir, G. & Sorek, R. Contemporary phage biology: from classic models to new insights. *Cell* **172**, 1260–1270 (2018).
3. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020).
4. Durmaz, E. & Klaenhammer, T. R. Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected *Lactococcus lactis*. *J. Bacteriol.* **189**, 1417–1425 (2007).

5. Pecota, D. C. & Wood, T. K. Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. *J. Bacteriol.* **178**, 2044–2050 (1996).
6. Gao, L. et al. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020).
7. Ledvina, H. E. et al. An E1-E2 fusion protein primes antiviral immune signalling in bacteria. *Nature* **616**, 319–325 (2023).
8. Millman, A., Melamed, S., Amitai, G. & Sorek, R. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. *Nat. Microbiol.* **5**, 1608–1615 (2020).
9. Wang, S. et al. SspABCD-SspFGH constitutes a new type of DNA phosphorothioate-based bacterial defense system. *mBio* **12**, e00613-21 (2021).
10. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* **41**, 4360–4377 (2013).
11. Mayo-Muñoz, D., Pinilla-Redondo, R., Birkholz, N. & Fineran, P. C. A host of armor: prokaryotic immune strategies against mobile genetic elements. *Cell Rep.* **42**, 112672 (2023).
12. Frost, L. S., Leplae, R., Summers, A. O. & Toussaint, A. Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* **3**, 722–732 (2005).
13. Rankin, D. J., Rocha, E. P. C. & Brown, S. P. What traits are carried on mobile genetic elements, and why? *Heredity* **106**, 1–10 (2011).
14. Goldfarb, T. et al. BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* **34**, 169–183 (2015).
15. Ofir, G. et al. DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nat. Microbiol.* **3**, 90–98 (2018).
16. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* **193**, 6039–6056 (2011).
17. Hochhauser, D., Millman, A. & Sorek, R. The defense island repertoire of the *Escherichia coli* pan-genome. *PLoS Genet.* **19**, e1010694 (2023).
18. Vassallo, C. N., Doering, C. R., Littlehale, M. L., Teodoro, G. I. C. & Laub, M. T. A functional selection reveals previously undetected anti-phage defence systems in the *E. coli* pangenome. *Nat. Microbiol.* **7**, 1568–1579 (2022).
19. Rousset, F. et al. Phages and their satellites encode hotspots of antiviral systems. *Cell Host Microbe* **30**, 740–753.e745 (2022).
20. Hussain, F. A. et al. Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. *Science* **374**, 488–492 (2021).
21. Johnson, M. C. et al. Core defense hotspots within *Pseudomonas aeruginosa* are a consistent and rich source of anti-phage defense systems. *Nucleic Acids Res.* **51**, 4995–5005 (2023).
22. Tacconelli, E. et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018).
23. LaVergne, S. et al. Phage therapy for a multidrug-resistant *Acinetobacter baumannii* craniectomy site infection. *Open forum Infect. Dis.* **5**, ofy064 (2018).
24. Makarova, K. S. et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
25. Doron, S. et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, aar4120 (2018).
26. Emond, E. et al. AbiQ, an abortive infection mechanism from *Lactococcus lactis*. *Appl. Environ. Microbiol.* **64**, 4748–4756 (1998).
27. Millman, A. et al. An expanded arsenal of immune systems that protect bacteria from phages. *Cell Host Microbe* **30**, 1556–1569.e1555 (2022).
28. Schuller, M. et al. Molecular basis for DarT ADP-ribosylation of a DNA base. *Nature* **596**, 597–602 (2021).
29. Tal, N. et al. Bacteria deplete deoxynucleotides to defend against bacteriophage infection. *Nat. Microbiol.* **7**, 1200–1209 (2022).
30. Payne, L. J. et al. Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. *Nucleic Acids Res.* **49**, 10868–10878 (2021).
31. Millman, A. et al. Bacterial retrons function in anti-phage defense. *Cell* **183**, 1551–1561.e1512 (2020).
32. Zhang, Y. et al. Complete genome sequence analysis of a novel alkane-degrading bacterial strain, *Acinetobacter vivianii* KJ-1, and its diesel degradation ability. *Front. Environ. Sci.* **10**, 1044754 (2022).
33. Fondi, M. et al. The genome sequence of the hydrocarbon-degrading *Acinetobacter venetianus* VE-C3. *Res. Microbiol.* **164**, 439–449 (2013).
34. Radisic, V., Nimje, P. S., Bienfait, A. M. & Marathe, N. P. Marine plastics from Norwegian west coast carry potentially virulent fish pathogens and opportunistic human pathogens harboring new variants of antibiotic resistance genes. *Microorganisms* **8**, 1200 (2020).
35. Carr, E. L., Kämpfer, P., Patel, B. K. C., Gürtler, V. & Seviour, R. J. Seven novel species of *Acinetobacter* isolated from activated sludge. *Int. J. Syst. Evolut. Microbiol.* **53**, 953–963 (2003).
36. Tuyet, N. L. T. & Kim, J. *Acinetobacter thutiae* sp. nov. isolated from oil-contaminated soil in motorbike repair workshops. *Diversity* **15**, 1179 (2023).
37. Liu, M. et al. Comparative genomics of *Acinetobacter baumannii* and therapeutic bacteriophages from a patient undergoing phage therapy. *Nat. Commun.* **13**, 3776 (2022).
38. Gordillo Altamirano, F. et al. Bacteriophage-resistant *Acinetobacter baumannii* are resensitized to antimicrobials. *Nat. Microbiol.* **6**, 157–161 (2021).
39. Gordillo Altamirano, F. L. et al. Phage-antibiotic combination is a superior treatment against *Acinetobacter baumannii* in a preclinical study. *EBioMedicine* **80**, 104045 (2022).
40. Forde, A. & Fitzgerald, G. F. Bacteriophage defence systems in lactic acid bacteria. *Antonie Van Leeuwenhoek* **76**, 89–113 (1999).
41. Tesson, F. & Bernheim, A. Synergy and regulation of antiphage systems: toward the existence of a bacterial immune system? *Curr. Opin. Microbiol.* **71**, 102238 (2023).
42. Pingoud, A. & Jeltsch, A. Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **29**, 3705–3727 (2001).
43. Wu, Y. et al. Bacterial defense systems exhibit synergistic anti-phage activity. *Cell Host Microbe* **32**, 557–572.e556 (2024).
44. Pfeifer, E., Sousa, J. M., Touchon, M. & Rocha, E. P. When bacteria are phage playgrounds: interactions between viruses, cells, and mobile genetic elements. *Curr. Opin. Microbiol.* **70**, 102230 (2022).
45. Hazen, T. H., Pan, L., Gu, J. D. & Sobecky, P. A. The contribution of mobile genetic elements to the evolution and ecology of *Vibrios*. *FEMS Microbiol. Ecol.* **74**, 485–499 (2010).
46. Koonin, E. V., Makarova, K. S., Wolf, Y. I. & Krupovic, M. Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. *Nat. Rev. Genet.* **21**, 119–131 (2020).
47. Johnson, C. M., Harden, M. M. & Grossman, A. D. Interactions between mobile genetic elements: an anti-phage gene in an integrative and conjugative element protects host cells from predation by a temperate bacteriophage. *PLoS Genet.* **18**, e1010065 (2022).
48. Zhao, Y., Wei, H. M., Yuan, J. L., Xu, L. & Sun, J. Q. A comprehensive genomic analysis provides insights on the high environmental adaptability of *Acinetobacter* strains. *Front. Microbiol.* **14**, 1177951 (2023).
49. Jia, J., Liu, M., Feng, L. & Wang, Z. Comparative genomic analysis reveals the evolution and environmental adaptation of *Acinetobacter johnsonii*. *Gene* **808**, 145985 (2022).



50. Godeux, A. S. et al. Interbacterial transfer of carbapenem resistance and large antibiotic resistance islands by natural transformation in pathogenic acinetobacter. *mBio* **13**, e0263121 (2022).
51. Pinilla-Redondo, R. et al. Discovery of multiple anti-CRISPRs highlights anti-defense gene clustering in mobile genetic elements. *Nat. Commun.* **11**, 5652 (2020).
52. Getz, L. J. & Maxwell, K. L. Diverse antiphage defenses are widespread among prophages and mobile genetic elements. *Ann. Rev. Virol.* **11**, 343–362 (2024).
53. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
54. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
55. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
56. Tesson, F. et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* **13**, 2561 (2022).
57. Abby, S. S., Néron, B., Ménager, H., Touchon, M. & Rocha, E. P. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS ONE* **9**, e110726 (2014).
58. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
59. Maiden, M. C. Multilocus sequence typing of bacteria. *Annu. Rev. Microbiol.* **60**, 561–588 (2006).
60. Alcock, B. P. et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **51**, D690–d699 (2023).
61. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
62. Okura, Y. [Fundamentals] 4. Visualization in Python Programming: how to use Matplotlib and seaborn]. *Nihon Hoshasen Gijutsu Gakkai Zasshi* **79**, 723–731 (2023).
63. Fang, Z., Liu, X. & Peltz, G. GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023).
64. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
65. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
66. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
67. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
68. 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).
69. Johansson, M. H. K. et al. Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder. *J. Antimicrob. Chemother.* **76**, 101–109 (2021).
70. Camargo, A. P. et al. Identification of mobile genetic elements with geneNomad. *Nat. Biotechnol.* **42**, 1303–1312 (2024).
71. Laslett, D. Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic acids Res.* **32**, 11–16 (2004).
72. Yi, W. et al. In silico characterization of defense system hotspots in *Acinetobacter* spp. Github <https://github.com/Guan-biolab/Phage-Hotspots/tree/main> (2024).

## Acknowledgements

We thank the colleagues from the Bioinformatics Laboratory of The First Hospital of Jilin University for their valuable discussions and suggestions, which have greatly improved the quality of the manuscript. The project was supported by the research fund of Prof. Qingtian Guan, grant number 04045970001.

## Author contributions

Q.G. conceptualized the study, obtained the funding, and supervised the work. W.Y., N.Z., and Z.P. performed the data analysis, Q.G. and W.Y. prepared the initial draft of the manuscript, followed by edits from X.C., H.S., L.S., Z.G., A.P., and Z.L. All authors have commented on various sections of the manuscript, which were finally curated and incorporated in the final version by Q.G.

## 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/s42003-025-07459-4>.

**Correspondence** and requests for materials should be addressed to Qingtian Guan.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Professor Haichun Gao and Aylin Bircan, Tobias Goris. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025