Analyses of *Xenorhabdus griffiniae* genomes reveal two distinct sub-species that display intra-species variation due to prophages

Jennifer K. Heppert[1*], Ryan Musumba Awori[2*], Mengyi Cao[3], Grischa Chen[3], Jemma McLeish[1], Heidi Goodrich-Blair[1#]

1. Department of Microbiology, University of Tennessee at Knoxville, Knoxville, Tennessee, USA.
2. Elakistos Biosciences, Nairobi, Kenya.
3. Division of Biosphere Sciences Engineering, Carnegie Institute for Science, Pasadena, California, USA.

[#] corresponding author
*denotes equal contribution

Abstract

Background:

Nematodes of the genus *Steinernema* and their *Xenorhabdus* bacterial symbionts are lethal entomopathogens that are useful in the biocontrol of insect pests, as sources of diverse natural products, and as research models for mutualism and parasitism. *Xenorhabdus* play a central role in all aspects of the *Steinernema* lifecycle, and a deeper understanding of their genomes therefore has the potential to spur advances in each of these applications.

Results:

Here, we report a comparative genomics analysis of *Xenorhabdus griffiniae*, including the symbiont of *Steinernema hermaphroditum* nematodes, for which genetic and genomic tools are being developed. We sequenced and assembled circularized genomes for three *Xenorhabdus* strains: HGB2511, ID10 and TH1. We then determined their relationships to other *Xenorhabdus* and delineated their species via phylogenomic analyses, concluding that HGB2511 and ID10 are *Xenorhabdus griffiniae* while TH1 is a novel species. These additions to the existing *X. griffiniae* landscape further allowed for

the identification of two subspecies within the clade. Consistent with other *Xenorhabdus*, the analysed *X. griffiniae* genomes each encode a wide array of antimicrobials and virulence-related proteins. Comparative genomic analyses, including the creation of a pangenome, revealed that a large amount of the intraspecies variation in *X. griffiniae* is contained within the mobilome and attributable to prophage loci. In addition, CRISPR arrays, secondary metabolite potential and toxin genes all varied among strains within the *X. griffiniae* species.

Conclusions:

Our findings suggest that phage-related genes drive the genomic diversity in closely related *Xenorhabdus* symbionts, and that these may underlie some of the traits most associated with the lifestyle and survival of entomopathogenic nematodes and their bacteria: virulence and competition. This study establishes a broad knowledge base for further exploration of not only the relationships between *X. griffiniae* species and their nematode hosts but also the molecular mechanisms that underlie their entomopathogenic lifestyle.

Keywords: *Xenorhabdus griffiniae*, nematode-bacterium symbiosis, prophage, CRISPR loci, pangenome, bacterial subspeciation, insect toxins, entomopathogenic bacteria

Background

Buried in soils across the world is living white gold, a rich, but as yet under-utilized bioresource: *Steinernema* nematodes. These insect-killing roundworms have been found in 51 countries to date [1–6] and are profitable commercial products for the control of insect crop pests. In addition, they are colonized by microbes, including obligate symbiotic

bacteria from the genus *Xenorhabdus*, that produce a battery of useful biomolecules [7]. To date, 35 *Xenorhabdus* species have been described [7–9, https://lpsn.dsmz.de/genus/xenorhabdus] found in association with *Steinernema* nematodes, and the two species work in tandem to infect and kill insects and exploit the nutrient rich cadaver for the reproductive stage of their shared lifecycle.

Mechanistically, the nematode's *Xenorhabdus* bacterium gut symbionts potentiate their insect-killing trait and serve as the primary food source for the nematode. In the non-feeding, host-seeking infective juvenile (IJ) stage of the nematode's lifecycle, *Steinernema* nematodes house their *Xenorhabdus* symbionts in a specialized tissue of the anterior intestine known as the receptacle. After an IJ successfully enters an insect *via* natural openings such as spiracles, it defecates into the haemolymph, its *Xenorhabdus* bacteria, which by their secretion of insect toxins, immunosuppression and growth in the haemolymph, kill the insect [7]. The resultant insect cadaver is an enclosed nutrient-rich niche that both nematode and bacterium leverage to reproduce proliferatively. Nematode fecundity is enhanced by the consumption of *Xenorhabdus* [10], and *Xenorhabdus* defends the niche by secreting bacteriocins, antimicrobials and scavenger deterrents, which antagonize both microbial and invertebrate competitors [11–13]. Before their exit from the insect cadaver and entry into the surrounding soil, nascent IJ nematodes are specifically colonized by *Xenorhabdus* in the anterior intestine and ultimately, the receptacle [14].

Each *Steinernema* nematode is colonized in the receptacle by a specific *Xenorhabdus* species. However, some *Xenorhabdus* species associate with multiple *Steinernema* nematode hosts [7] suggesting a relatively fluid partnership landscape in which the molecular determinants of host-symbiont specificity across the two genera are still being defined. Key to this understanding is a robust comparative analysis of the phylogenetic relatedness and genomic content of *Xenorhabdus* isolates. Genome assemblies that are lowly contaminated, with high levels of completeness and greater than 50x coverage, are sufficient for bacterial species delineation [15], comparative genomics and identification and analyses of genes-of-interest. Indeed, insights into the *Xenorhabdus-Steinernema* symbiosis, bacterial speciation, and entomopathogenicity have been gained through such analyses of *Xenorhabdus* genomes. For example, comparing the degree of genetic similarity across whole genomes using digital DNA-DNA hybridization (dDDH) has led to the delineation of five novel *Xenorhabdus* species to date [8, 9, 16, 17]. Also, a comparative analysis of *Xenorhabdus bovienii* CS03 and SS-2004 genomes revealed that CS03 is more adapted to destroying microbial competitors than is SS-2004 but encodes fewer genes associated with entomopathogenicity [18]. Pangenome analyses of *X. bovienii* strains revealed intra-species content variation including of prophage origin, suggestive of strain adaptation to specific host environments [19]. The biosynthetic gene cluster that encodes the production of GameXpeptides was found in 72% of a sample of 29 *Xenorhabdus* genomes [20], which suggests that many strains use this peptide to suppress insect immune pathways [21].

Of the >100 *Steinernema* nematode species described to date, only *Steinernema hermaphroditum* is a self-fertilizing hermaphrodite [22, 23]. This makes *S.*

hermaphroditum particularly well-suited for rigorous nematode genetic studies, including those on the dynamics of transmission of *Xenorhabdus* bacteria from one generation of their host nematode to another [24, 25]. To lay the groundwork for such studies [26], we aimed to comprehensively analyse genomes of *Xenorhabdus griffiniae*, gut symbionts of *S. hermaphroditum* [23, 27]. We hypothesize that insights into the *X. griffiniae-S. hermaphroditum* symbiosis are attainable through detailed comparative genome analyses of *X. griffiniae* strains and their close phylogenetic relatives. Here, we report our studies in which we delineated two species of *Xenorhabdus* among the analysed strains, reconstructed their phylogenetic relationships with the rest of the *Xenorhabdus* genus, and analysed their pangenome and unique loci including prophages, CRISPR, and those encoding phage tail-like structures, secondary metabolites and insect toxins, the last of which was substantiated through insect mortality assays.

Methods

Bacterial genome sequencing

X. griffiniae HGB2511 and *Xenorhabdus* sp. TH1 (respective Genbank accession numbers for *16s rDNA*: MZ913116-MZ913125 and OR047834) were isolated from infective juvenile nematodes of *Steinernema hermaphroditum* CS34 and *Steinernema adamsi*, respectively, as previously described [16]. Briefly, *Galleria mellonella* larvae were exposed to infective juvenile nematodes, died, and emergent infective juveniles were isolated by White trapping in sterile water. These nematodes were then surface sterilized, ground, plated on LB agar plates supplemented with pyruvate (0.1%), and incubated at 30°C. *X. griffiniae* ID10 was purchased from BacDive (DSM 17911). Cultures were

inoculated in lysogeny broth (LB) stored in the dark (dark LB:0.5% yeast extract, 1% tryptone, 0.5% NaCl) and grown overnight under agitation at 30°C. Genomic DNA was prepared for sequencing by phenol extraction and spooling (*X. griffiniae* HGB2511), using the Qiagen DNeasy Kit per the manufacturer's instructions with minor modification, (*X. griffiniae* ID10), or by both methods (*Xenorhabdus* sp. TH1). The Qiagen DNeasy Kit protocol was modified to prevent the viscous *Xenorhabdus* cell lysate from clogging the DNA-binding column by diluting the lysate with 10 mL of Qiagen QBT buffer before running it through the DNA-binding column by gravity. Purified DNA was sequenced using both short-read and long-read sequencing at the Millard and Muriel Jacobs Genetics and Genomics Laboratory, at the California Institute of Technology (for *X. griffiniae* HGB2511 and *Xenorhabdus* sp. TH1) or Novogene (for *X. griffiniae* ID10). For the *X. griffiniae* ID10 genome, short-read sequencing was performed using Illumina NovaSeq 150 base pairs (bp) paired-end short-read sequencing with library construction consisting of genomic DNA fragmentation, Illumina adapter ligation and PCR amplification, followed by size selection and purification, resulting in 33 gigabases (GB), of which 842 megabases (MB) were used for assembly. For primer-free long-read sequencing, library preparation consisted of size selection, adapter ligation and purification using Beckman Coulter AMPure XP beads. Sequencing was performed on an Oxford Nanopore (ONP) PromethION platform with base calling performed using Guppy software [28] with standard parameters. Prior to assembly, the ONP long-reads were filtered using Filtlong v0.2.1 resulting in 1.13 GB that was used in the assembly. Short- and long-reads were assembled using Unicycler v0.5.0 [29], resulting in a coverage of 432X. Sequencing of the TH1 and HGB2511 was done via a similar workflow. For each, 1.13GB and ~0.74GB

of bases were obtained from the Illumina and ONP runs, respectively. Likewise, a similar hybrid assembly method was also used, resulting in assemblies that were 406x and 496x for HGB2511 and TH1, respectively. Genome characteristics were determined via the PATRIC platform [30] (Table 1). Coverage was calculated by taking the total number of base pairs used in the assembly and dividing that by the genome size [31]. EvalG was used to determine the quality and completion of the assemblies [32]. GenBank accessions for ID10, HGB2511 and TH1 genomes are CP147737.1, CP147738.1 and CP147734.1, respectively. Names and GenBank accession numbers of other genomes used in this study are found in Supplementary sheet S17 in Additional file 2.

Tree generated using Bayesian inference

Phylogenetic analyses were performed as described previously [33]. Briefly, select *Xenorhabdus* (30) and *Photorhabdus* (1) species for which genomic sequences were publicly available were analysed using MicroScope MaGe's Gene Phyloprofile tool [34] to identify homologous open reading frames (ORF) sets (homologs with at least 50% identity with synteny) which were conserved across all assayed genomes. Putative paralogs were excluded from the downstream analysis to ensure homolog relatedness, resulting in 1235 homologous sets (one-to-one orthologs). Homolog sets were retrieved via locus tag indexing using Python v3.8.0, and nucleotide sequences were individually aligned using Muscle v3.8.31 [35], concatenated using Sequence Matrix v1.9 [36], and trimmed of nucleotide gaps using TrimAL v1.4 [37]. A General Time Reversible + γ variation substitution model was used for maximum likelihood and Bayesian analysis. Maximum likelihood analyses were performed via RAxML v8.2.10 [38] using rapid bootstrapping and 1,000 replicates and were visualized via Dendroscope v3.6.2 [39]. Nodes with less than 40% bootstrap support were collapsed. Bayesian analyses were

performed via MrBayes v3.2.7a in BEAGLE [40] on the CIPRES Science Gateway platform [41]. A total of 100,000 Markov chain Monte Carlo (MCMC) replicates were performed. Twenty-five per cent were discarded as burn-in, and posterior probabilities were sampled for every 500 replicates. Two runs were performed with three heated chains and one cold chain. The final average standard deviation of split frequencies was 0.052352. Bayesian trees were visualized with FigTree v1.4.4 [42]. Posterior probabilities are 100% except where otherwise indicated.

Digital DNA-DNA hybridisation and pangenome analyses

To determine pairwise digital DNA-DNA hybridisation (dDDH) values among 31 strains (ID10, VH1, XN45, TH1, HGB2511, BMMCB, BG5, Kalro, 97 and 22/26 validly published *Xenorhabdus* type strains), their fasta formatted genomes were uploaded to the TYGS server [43] and analysed as previously described [16]. Genomes of six *X. griffiniae* strains plus that of *Xenorhabdus* sp. TH1 were comparatively analysed using a pangenome approach in Anvio 7.1 [44]. Briefly, fasta formats of the seven genomes were reformatted to simplify the definition lines, then converted to anvio contig databases. On these, hidden Markov models (HMMs) and genes were identified using HMMER [45] and Prodigal [46], respectively. The functions of these genes were then predicted, based on orthology, using the Cluster of Orthologous Genes (COG) database [47] as a reference. These annotated contig databases were then used to construct a pangenome with the anvi-pan-genome program under the following parameters --use NCBI-BLAST, MCL inflation 10, minbit 0.5, --exclude-partial-gene-calls. The anvi-display-pan program was used to both display the pangenome as a sunburst chart and subsequently create selected bins. To obtain sub-pangenomes of only type VI secretion system-associated orthologous gene clusters (GC), all GCs annotated with “type VI” were identified using the search functions feature

and binned. The anvi-split program was then used to obtain a pangenome of only Type VI secretion system-associated GCs (Supplementary Sheet S3 in Additional File 2). For each analysis, pangenomes were represented both as sunburst charts and tabulated gene lists. To calculate average nucleotide identities (Supplementary Sheet S13 in Additional file 2), the fastANI [48] program was used within an Anvio environment. The average alignment fraction and fragments were 0.85 and 1425 respectively.

To identify the mobilome, all genes annotated with the COG category “X” were extracted from the main pangenome and used to create a sub-pangenome of the mobilome only. This was then used to calculate the total number of genes annotated as phage, transposase, and plasmid-related in each genome. For each genome, these totals were correlated with proteome size using Pearson’s adjusted (due to the small sample size) r-square at a 0.05 alpha level.

Identification and analysis of unique genes within the accessory genome

Using the tabulated output of the Anvio pangenome analysis and Microsoft Excel, core genes and genes unique to a given strain were identified using the number of genomes in which the gene cluster has hits identifier, where the value equalled 7 for core genes (19,196) and 1 for unique genes for the HGB2511 (340), ID10 (411) and TH1 (454) strains. Because the genomes of the sub-species clade which includes *X. griffiniae* Kalro are so similar, there were less than ten unique genes per strain [XN45(8), VH1(7), xg97(4), Kalro (0)]. Thus, to gain a better understanding of what genes might be unique within the sub-species clade, we compared the Kalro genome alone to the HGB2511, ID10 and TH1 genomes and found 454 unique genes for further analysis. The COG annotations assigned as part of the pangenome analysis were used to elucidate the functional categories that might be enriched among those genes we found to be unique

to a given species. The number of unique genes per COG category was plotted as a percentage of the total number of unique genes with a COG designation for each strain and the core genome using GraphPad PRISM v10.1.1. As expected, a much larger fraction of genes with no COG designation was found in each of the unique gene categories (approximately 50% in each case) compared to the core genome (6%).

Elucidation and analysis of prophages in *X. griffiniae* genomes

To identify prophages in bacterium genomes, we used VIBRANT 1.2.1 [49] under default parameters. Each prophage sequence was then separately reannotated with Pharokka [50] under "meta mode" and Bakta [51]. Genomad [52] was used to both taxonomically classify prophages and assess their quality (through CheckV [53]) and completion (Supplementary Sheet S14 in Additional File 2). Resultant gene lists of the annotated prophages are available in (Supplementary Sheets S8-S11 in Additional File 2).

To identify similar prophages across strains, we used progressive Mauve [54] to identify collinearity blocks between prophage sequences. Considerably similar pairs were selected, pairwise aligned, and visualised as dot plots using Geneious 8.1.9 [55].

To determine the effect of prophages on the dDDH values among *X. griffiniae* strains, all prophage sequences were deleted from their corresponding bacterium genomes in Geneious 8.1.9. Then, dDDH analyses were rerun using both original and "phageless" *X. griffiniae* bacterium genomes (Supplementary Sheet S12 in Additional File 2).

To identify strain-specific and subspecies-specific genes that are from prophages, a pangenome approach was used. Briefly, their entire prophage sequence from a strain — ID10, Kalro, TH1, xg97 and HGB2511— were merged into a fasta file. For example, all prophages from the ID10 genome were merged into a single fasta file "ID10_prophages.fasta". These five prophage fasta files, plus those of genomes of ID10,

HGB2511, VH1, XN45, xg97 and Kalro, were used to create a pangenome as aforementioned. Using the anvio-display-pan program, strain-specific or subspecies-specific GCs that were of prophage origin were visually identified and binned. Using the anvio-summary program, the number, annotation, and aa sequences of these strain-specific or subspecies-specific genes of prophage origin were obtained from resultant gene lists (Supplementary Sheet S14 in Additional File 2). This was then used to calculate what proportion of strain-specific and subspecies-specific genes were from prophages.

To identify genomic loci encoding complete type VI secretion systems (T6SS) within genomes, we used Secret6 [56] under default parameters. Geneious was used to create multiple sequence alignments of T6SS-encoding loci and to calculate pairwise nucleotide sequence identities.

To infer gene gain and loss events within the *X. griffiniae* clade, we used the above described anvi'o workflow to create aa pangenome of the seven strains plus *Xenorhabdus* sp. BG5; BG5 was included to specifically infer the evolution of gene content for strain TH1, as it is the strain that just diverges prior to the emergence of TH1(Fig. 1). We then used the tabulated version of the pangenome to manually construct a phyletic pattern (Supplementary Sheet S6 in Additional File 2). A dDDH-based phylogenomic tree of the eight strains only, was reconstructed via the TYGS pipeline. Both this tree and the phyletic pattern were used as input data in COUNT [57]. The evolution of gene content was inferred using Wagner parsimony with a gain penalty of two.

Identification and analysis of defence systems and CRISPR-Cas loci in *X. griffiniae* genomes

Anti-phage or anti-plasmid defence systems present within genomes were identified using the DefenseFinder bioinformatics search tool and results were interpreted using the

associated knowledge base [58–60] (Table S4 in Additional File 3, Supplementary Sheet S22 in Additional File 2). To identify CRISPR-Cas systems in *X. griffiniae* genomes we used BlastN on the Magnifying Genome platform to search for CRISPR repeats with the previously identified *X. nematophila* CRISPR repeats, XnCRISPR-E and XnCRISPR-G as queries [61]. Regions with similar or identical sequences were extracted and manually curated for repeat-spacer content. All results were then verified using CRISPRDetect [62]. The only distinction between the two approaches was for ID10, for which the manual approach had suggested five potential repeat sites of one spacer only, whereas CRISPRDetect did not identify any. We therefore conducted a second manual annotation of the repeats in this genome (Additional File 3). Based on conservation with consensus repeats called by CRISPRDetect in both upstream and downstream repeats, we considered ID10 1ai and ID10 1bii regions to be bona fide single spacer CRISPR loci. Cas protein-encoding genes were identified based on their annotation in the Magnifying Genomes platform and confirmed using CasFinder-3.1.0 [63]. To identify protospacers we searched for full-length identical sequences using each identified spacer as a BlastN query against the *X. griffiniae*, TH1, and BMMCB genomes in the Magnifying Genome platform. Protospacers were identified in both other strains and within the same genome (self-targeting protospacers). To gain a broader view of the non-self-protospacers, we used CRISPRTarget (http://crispr.otago.ac.nz/CRISPRTarget/crispr_analysis.html) to search the spacer sequences found by CRISPRDetect for TH1, ID10, HGB2511, and BMMCB genomes, against subset of the available databases (ALCAME genes, Genbank-Phage, RefSeq-Archaea, RefSeq-Plasmid, and RefSeq-Viral) for high confidence sequence matches to the potential protospacers [62, 64]. All putative protospacers for a

given spacer were listed (Table S6 in Additional File 3) and the top, annotated hit for each is listed in Table 2.

Elucidation of biosynthetic gene clusters

To predict which secondary metabolites can be produced by *X. griffiniae,* a fasta format of the ID10 genome was uploaded to the AntiSMASH [65] webserver (https://antismash.secondarymetabolites.org) as data input. Default parameters (i.e., relaxed detection strictness) and the use of all extra analyses were selected. Output data comprised detailed bioinformatic analyses of 23 biosynthetic gene clusters (BGCs) in the ID10 genome. Each of the 23 BGCs was then manually inspected in two main ways. First, the clusterBLAST feature was used to identify known homologous BGCs by comparing the gene synteny and sequence similarity of an ID10 BGC to those of known BGCs found in the MiBIG database [66]. An ID10 BGC was considered homologous to a known BGC if it contained >80% genes in perfect synteny, with each gene having BLASTp sequence similarity >47%, sequence coverage >40%, and E_{value} <2.78E-19 with the corresponding gene in the known BGC. Second, for ID10 non-ribosomal peptide synthetase (NRPS) BGCs, their Stachelhaus codes [67] for adenylation domains, as well as their epimerisation/dual condensation domains [68] were analysed to determine the amino acid sequence of the linearised non-ribosomal peptide (NRP) the NRPS was predicted to biosynthesize. Predicted NRPs were then compared to known NRP to identify NRPS BGCs that encode the biosynthesis of novel derivatives/peptides. This workflow was similarly applied for the analysis of BGCs in the HGB2511 genome. Chemical structures of the compounds whose production was predicted to be encoded by the known ID10

BGCs were obtained from the Natural Product Atlas [69] and edited in Chemdraw and Inkscape [70].

Putative toxin, secretion system, Cas protein, and Restriction modification system identification

Putative toxins predicted to impact insect or nematode virulence were identified using multiple approaches. The first approach used the loci of previously characterized known or suspected toxins [71]. BlastP was performed on the Magnifying Genomes platform, and the BLAST query accession proteins are listed in the putative toxins table (Table 3), this procedure was repeated for other novel putative toxins identified as well. The PathoFact software package [72] was used to identify novel putative toxins in the ID10, HGB2511, Kalro, TH1 and *X. nematophila* 19061 genomes with the standard settings and chromosomal genomes downloaded from the Magnifying Genomes platform. The toxin library outputs from PathoFact were compared and toxins potentially unique to each genome were further examined (Supplementary sheets S2-S3 in Additional File 4). This resulted in the identification of a potential hydrogen cyanide synthetase locus (XTH1_v2_1430-1432) which appears to be unique to the TH1 genome among those analysed. Further, this search revealed two zonula occludens toxin proteins (JASDYB01_14222 and JASDYB01_14237 and _14239) which appear to be unique to the Kalro genome (homologs of were also found in VH1, XN45, and xg97). To both confirm and expand the list resulting from the combination of the above analyses, the search term “toxin” was used on the Magnifying Genomes platform to further query the gene annotations for the HGB2511, ID10, TH1 and Kalro genomes. This search resulted in the identification of a protein annotated as an insecticidal toxin (XGHID_v1_0629) in the ID10 genome and a homolog (XGHIN1_v1_3228) was subsequently found in the

HGB2511 genome via BlastP. The products of toxin-antitoxin systems were generally not considered in our analysis. The highest confidence and most interesting results are summarized in the Putative Toxins Table (Table 3).

To systematically predict the types and numbers of secretion systems encoded by the *X. griffiniae* genomes (HGB2511, ID10, Kalro, VH1, XN45, xg97), TXSScan application from the Macsyfinder 2.0 [73] program was run on the Galaxy [74] platform specifying an ordered, circular replicon, diderm bacteria and with default HMMER options. A combination of the summary output for each genome is found in Table S1 in Additional File 1. Alignment of MARTX regions was performed using MUSCLE in MegAlignPro (DNA Star).

To confirm the type and numbers of Cas proteins encoded by the *X. griffiniae* genomes (HGB2511, ID10, Kalro, VH1, XN45, xg97), CasFinder application from the Macsyfinder 2.0 [73] program was run on the Galaxy Pasteur platform specifying an ordered, circular replicon, diderm bacteria and with default HMMER options. The summary table is compiled from the software summary output and the additional sheets contain the best solution predictions from the Cas Finder output and include the gene names and locus tags from the different Cas systems found (Supplementary Sheet S20 in Additional File 2).

Restriction modification systems were identified by searching the BMMCB, TH1, ID10, HGB2551 and Kalro genomes for restriction enzymes, anti-restriction proteins, and

restriction enzyme-associated methylases-based gene annotations on the MAGE Microscope platform. These genomes were further searched using BLAST 2.15.0+ to compare a list of “Gold Standard” methyltransferase and endonuclease protein sequences from the New England Biolabs’ REBASE with the protein sequences contained within each genome [75]. An E value cut-off of $\leq 1 \times 10^{-5}$ and 75% coverage were used to generate a list of high confidence candidates. This candidate list was compared with the initial list of restriction system proteins generated using MAGE, and redundant sequences were removed. Methyltransferase homologs that were not found to be near predicted restriction endonucleases were excluded from further analysis, though the presence of many ‘orphan’ methyltransferases may indicate a need to protect the bacterial chromosome from restriction modification systems [76]. Putative restriction enzyme loci were examined to identify or confirm neighbouring methylases in the case of Type I, II and III restriction enzymes, or lack thereof, in the case of the Type IV and the HNH restriction endonucleases identified [77]. Of the putative restriction modification genes examined, only three loci from the BMMCB genome (LDNM01_v1_10020, LDNM01_v1_400040, LDNM01_v1_1980001) were predicted to encode restriction endonucleases from Type I or III but were not observed to encode a proximal methyltransferase. Because of their incomplete nature, these loci were excluded from the final table and count of restriction systems was identified. However, each locus was near a contig break, so a less fragmented BMMCB genome assembly may reveal these to be complete predicted restriction endonuclease loci.

Insect larvae rearing and preparation

Eggs of *Manduca sexta*, the tobacco hornworm, were purchased from Carolina Biological Supply (North Carolina, USA) and were reared to the fifth instar according to a previously

described protocol [78]. Briefly, eggs were sterilized with 0.6 % (v/v) bleach solution on arrival and then transferred to a four-ounce plastic container with a Gypsy moth diet (MP Biomedicals, Ohio, USA). The eggs were then incubated at 26 °C in a humidified insect incubator with a 16-h light: 8-h dark photoperiod. Once hatching was complete, the larvae were transferred to a new four-ounce container for two days and then transferred to a new two-ounce container. Feeding and cleaning were performed every two to three days until the larvae reached the fifth instar stage. On the day of the experiment, the larvae were examined and sorted by weight ranging from 0.67 g to 3.5 g, and the larvae were randomly distributed across the conditions. Larvae (n=10, with n=5 for phosphate-buffered saline (PBS) control) were placed into individual two-ounce plastic containers. Groups of 10 larvae were carefully injected with 10 µL of various doses (10^{-5}, 10^{-4}, and 10^{-3}) of bacteria between the first set of abdominal prolegs using a Hamilton syringe. Following injections, the larvae were incubated at 26 °C in a humidified insect incubator with a 16-h light: 8-h dark photoperiod and monitored for survival over 72 h.

Preparation of bacteria for infection of *M. sexta*

Bacterial strains were streaked from –80 °C freezer stocks onto dark LB agar and cultivated at room temperature in the dark for two days. Broad streaks including multiple colonies were used to inoculate LB medium (5 mL) and incubated on a rotating wheel at 30°C for approximately 8 h. At 8 h the optical density (OD) at 600nm was measured (OD600), and cultures were all near an OD600 of 1.0. Cultures were normalized to an OD600 of 1.0 by adjusting the volume of culture taken (e.g., 500 µL OD600 of 1.0). Cells were spun down for 1 min at max speed in Eppendorf tubes and washed twice with 1 mL of sterile PBS. After the final wash, cells were resuspended in 500 µL of PBS. The washed cells were then diluted ten-fold, six times in sterile PBS in a 96-well plate. For each

dilution, 10 µL was inoculated onto LB agar plates (LBP) to quantify the number of colony-forming units (CFU) in each dilution. To test for sterility, PBS was also inoculated onto LBP. Based on a previous study [71], we estimated that 100, 1,000, and 10,000 CFU/10 µL was suitable for an observable virulence (insect mortality) and dose-response, and we further calculated that 10^{-5}, 10^{-4}, and 10^{-3} dilutions of OD600 of 1.0 culture would yield approximately that number of cells per 10 µL injected. *M. sexta* larvae were raised, injected as described above, and then observed over 72 h.

Results

Strains ID10 and HGB2511 belong to one of two subspecies of *Xenorhabdus griffiniae* while *Xenorhabdus* sp. TH1 is a novel species

An *X. griffiniae* bacterial isolate (HGB2511) and its nematode host, a strain of *Steinernema hermaphroditum* [24] are being developed as a genetically tractable model for interrogating bacteria-host interactions [24, 25, 79]. To clarify the relationship between HGB2511 and other related *Xenorhabdus* isolates, we sought to comparatively analyse their genomes. We first sequenced the genome of three strains: HGB2511 [24], ID10, the *X. griffiniae* type strain isolated from an Indonesian strain of *S. hermaphroditum* [27], and TH1 isolated from *Steinernema adamsi* from Thailand [80]. For all three strains, we obtained circularised genome assemblies of the bacterial chromosome that had <0.7% contamination, >99.6% completion and >50x depth (Table 1).

Table 1. Characteristics of genomes assembled in this study. Characteristics were identified using analysis tools on the Bacterial and Viral Bioinformatics Resource Center platform.

Strain	Total length* (bp)	Completeness (contamination)	GC content (%)	Total CDS (with functional assignments)	tRNA genes	rRNA genes	Depth
ID10	4767870	100% (0.4%)	43.91	4688 (70%)	80	22	432x
HGB2511	4594889	100% (0.2%)	43.78	4369 (72%)	81	22	406x
TH1	3770596	99.6% (0.7%)	43.48	3469 (78%)	81	22	496x

*** The total length was equal to the Contig N_{50} (bp) since all three chromosome genome assemblies circularised.**

We leveraged these high-quality genomes to conduct a wide range of comparative genome analyses, the first of which was taxonomic species delineation via both dDDH and phylogenomics. These analyses both showed the same relationships among the *X. griffiniae* and closely related species (Fig. 1 & Fig. S1 in Additional file 1). The type strain designation of strain ID10 was corroborated by its lack of >70% pairwise dDDH values, the threshold for conspecific strains [81], with any of the other type strains of *Xenorhabdus* (Fig. 1A). Five additional strains were delineated as members of the *X. griffiniae* species, as they each had pairwise dDDH values with strain ID10 that were above the 70% threshold. Among the six *X. griffiniae* strains, two subspecies were evident due to intragroup pairwise dDDH values that were all above the 80% threshold [82]. Strains from Kenya, XN45, VH1, xg97, and Kalro, belonged to one subspecies while strains ID10 and HGB2511 from Indonesia and India, respectively, belonged to another subspecies (Fig. 1A and 1B). Although *Xenorhabdus* sp. TH1 was the most closely related strain to *X. griffiniae* HGB2511 and *Xenorhabdus* sp. BG5, it is an undescribed species of the genus as it lacked pairwise dDDH values, with any of the type strains that were above 70% (Fig.

1A). Indeed, in a pangenome of the seven strains, TH1 had the highest number of strain-specific genes, and none of the ANI pairwise values between strain TH1 and its six closest relatives met the 95% threshold [48] for conspecific strains (Fig. 2A, Supplementary Sheet S13 in Additional file 2). Hence, *Xenorhabdus* sp. TH1 is a novel species of the *Xenorhabdus* genus and not a strain of *X. griffiniae* as was stated in the description of its nematode host [80].

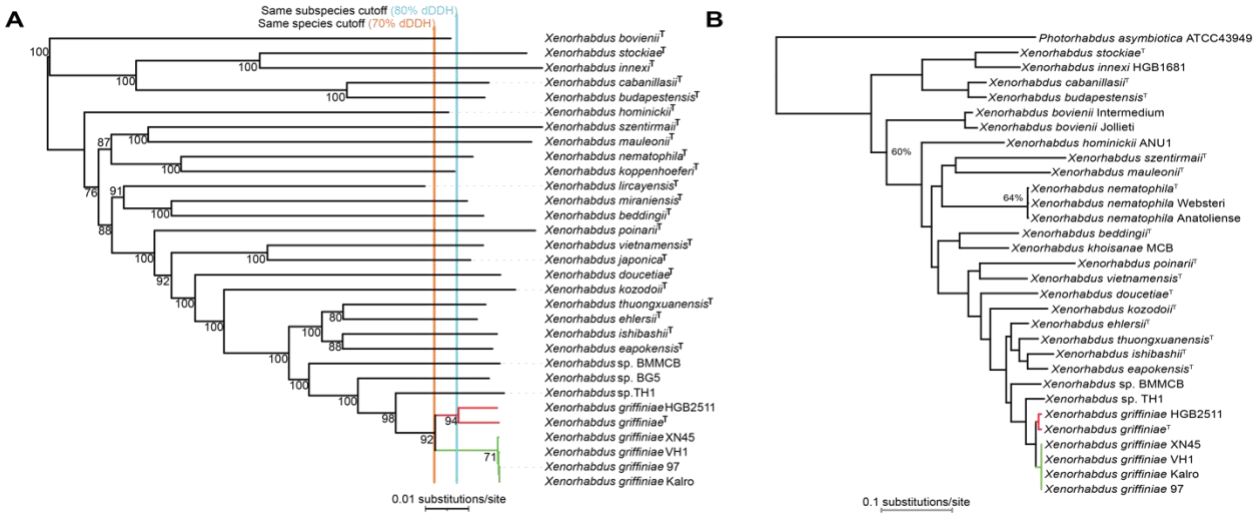


Fig. 1: Phylogenomic reconstruction of type strains of *Xenorhabdus* and other strains closely related to *Xenorhabdus griffiniae*. **A)** This neighbour joining tree was reconstructed using genomic distances calculated with the same formula (Genome Distance BLAST Phylogeny distance formula *d5)* used for species delineation by digital DNA-DNA hybridisation (dDDH) analyses. Orange and aqua lines correlate with dDDH boundaries for species (>70%) and subspecies (>80%), respectively. Strains of *X. griffiniae* formed two subspecies, those from India-Indonesia (red) and those from Kenya (green). **B)** Bayesian phylogenetic tree created using one-to-one orthologs from *Xenorhabdus* type strains, *X. griffiniae* strains, and *Photorhabdus asymbiotica* as an outgroup. Posterior probabilities are equal to 1 (100%) unless otherwise indicated at a given node. Strains of *X. griffiniae* formed two separate clades, those isolated from *S. hermaphroditum* nematodes found in Indonesia (ID10) and India (HGB2511) (red) and those isolated from nematodes found in Kenya (green).

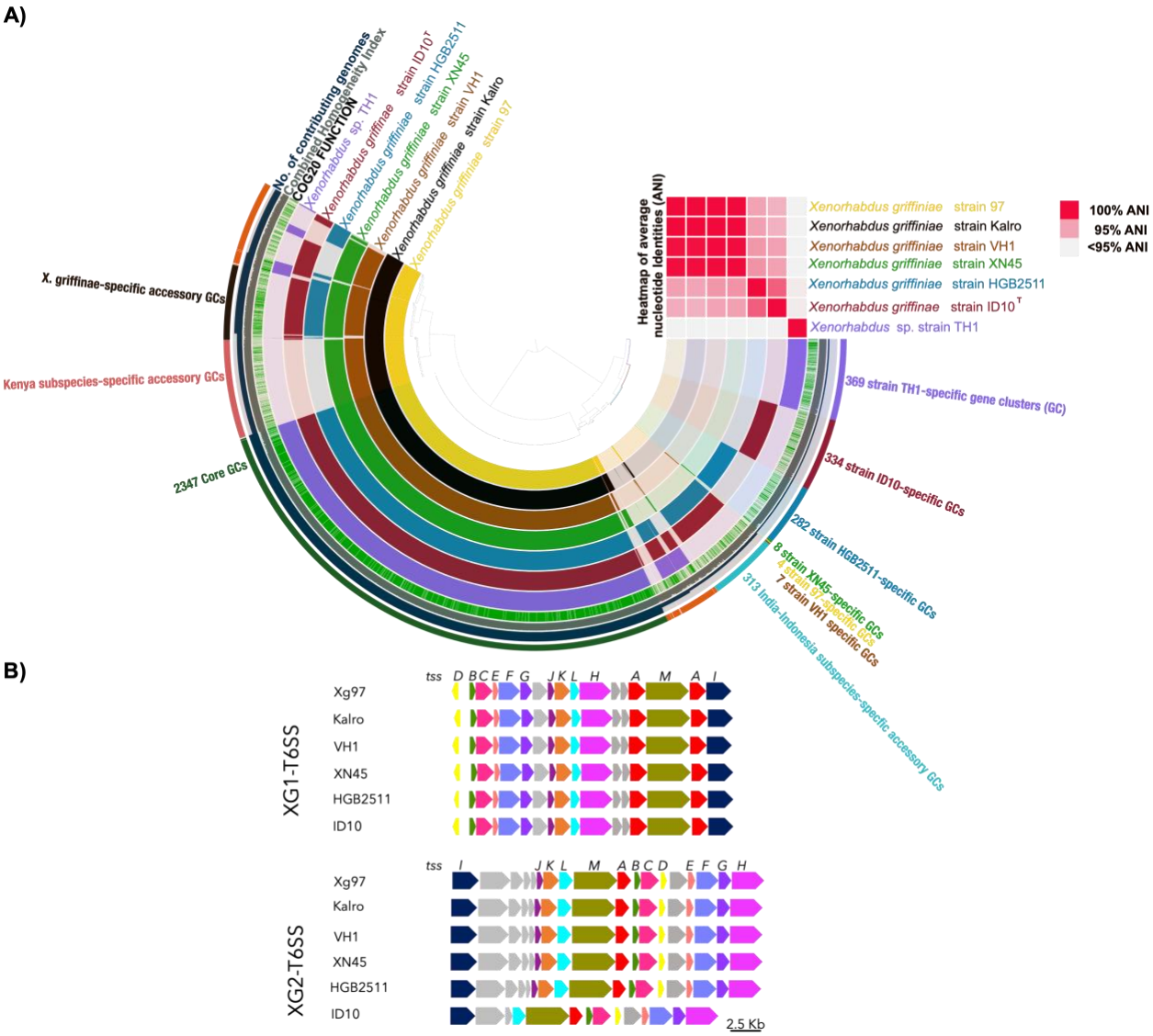


Fig. 2: Graphical representations of a pangenome of the *Xenorhabdus griffiniae* clade**. A**) Sunburst chart of a pangenome of the seven closest known relatives of the *X. griffiniae* type strain coupled to a heatmap of their average nucleotide identities (ANI). The pangenome contained 27,337 genes that were clustered into 5,113 groups of orthologs known as gene clusters (GC). The genomes from which each of the GCs was constituted are depicted in the sunburst chart as follows. Each concentric ring represents a genome and each radius represents a GC. For each radius, a dark shade across a concentric ring denotes that the GC is composed of genes from that genome. For the ANI heatmap, shades of red represent pairwise ANI values between 95% (blush) and 100% (rose), the threshold values for conspecific strains. **B)** Loci encoding type six secretions systems (T6SS) found only in *X. griffiniae* genomes, of those analysed here. Each genome encoded two different T6SS, XG1-T6SS and XG2-T6SS. The core T6SS-encoding genes are indicated (*tssA-M*). Other genes are in grey. Pairwise percentage nucleotide identities for shown genomic loci that encode XG1-T6SS and XG2-T6SS ranged between 98-100% and 80-100%, respectively.

X. griffiniae species encode type six secretion systems with subspecies-specific effectors

The pangenome of the six *X. griffiniae* strains plus *Xenorhabdus* sp. TH1 (Fig. 2), the most closely related known species to the *X. griffiniae* clade (Fig. 1), had a total of 27,337 genes. These were grouped into 5113 groups of orthologs, which we termed gene clusters (GC). Out of these, 2347 GCs were core in that they contained orthologs from every genome within the pangenome. On the other hand, a total of 369, 334, 282, 8, 4 and 7 GCs were unique to genomes of TH1, ID10, HGB2511, XN45, xg97 and VH1 respectively. Strain Kalro lacked unique GCs as all its GCs were present in the xg97 genome, even though their respective nematode hosts, *Steinernema* sp. Kalro and *Steinernema* sp. 97 are likely two different undescribed species [83]. Accessory GCs, each of which was composed of orthologs from between two and *n*-1 genomes, were 1739. Among these were 313 and 448 GCs, which were unique to genomes from the India-Indonesia and Kenyan subspecies, respectively. GCs that encode traits that define an *X. griffiniae* strain likely fall among the 319 GCs that were unique to *X. griffiniae* genomes, of which, only 202 (63%) had known functions. Among these, T6SS function was most enriched as it represented 12% of all *X. griffiniae*-specific genes (Sheet S2 in Additional file 2). For the T6SS GCs, those specifically encoding core components *tssA-M* were highly conserved as they had 95-100% combined homogeneity indexes—this is an anvi'o pangenome metric for estimating the similarity of orthologs within a GC calculated from sequence similarities and gap penalties derived from a multiple sequence alignment (MSA) of amino acid sequences [44]. The higher the value the more the positions with identical residues and no gaps within the MSA (Sheet S2 in Additional File 2). Upon deeper investigation, we found that all *X. griffiniae* genomes encode two complete T6SS that we designated XG1-T6SS and XG2-T6SS (Fig. 2B). XG1-T6SS loci

were almost identical across the six genomes, as the pairwise nucleotide percentage identities for this locus were between 98-100% (Sheet S16 in Additional File 2). Moreover, they were found in roughly the same chromosomal location (Fig. 3) in the four circularised genomes (xg97, Kalro, HGB2511 and ID10). For XG2-T6SS, none of the six corresponding genomic loci had pairwise nucleotide sequence identities that were less than 80% (Sheet S16 in Additional File 2). However, the ID10 XG2-T6SS encoding locus uniquely lacked *tssK, tssJ* and two other genes that were directly downstream of *tssJ* (Fig. 2). Like XG1-T6SS, the XG2-T6SS loci were also found in a similar chromosomal region across the four circularised genomes (Fig. 3). Based on the high pairwise nucleotide identities, we identified homologs of XG1-T6SS and XG2-T6SS-encoding loci in *X. szentirmaii* US123, *X. doucetiae*^{T}, *X. cabanillasi*^{T}, *X. hominickii* ANU, *X. nematophila*^{T}, *X. poinarii*^{T} and *X. bovienii* SS-2004 (Sheet S16 in Additional File 2). The strain SS-2004 homologs were those identified by Chaston *et al.* [84] and designated T6SS-1 and T6SS-2, respectively, by Kochanowsky *et al.* [85].

We further identified T6SS-associated GCs that are subspecies-specific. The Kenyan and India-Indonesia subspecies have eight and nine subspecies-specific GCs, respectively. The majority of these are predicted to encode spike proteins annotated as VgrG or PAAR-domain-containing Rhs proteins (Sheet S3 in Additional File 2). One Kenyan subspecies GC encodes a *tssF* that was not part of the two complete T6SS-encoding loci. We analysed the genes in the neighbourhood of the subspecies-specific PAAR-encoding loci for genes that encode T6SS effector proteins and their cognate immunity proteins. We found four such loci that are specific to the India-Indonesia subspecies

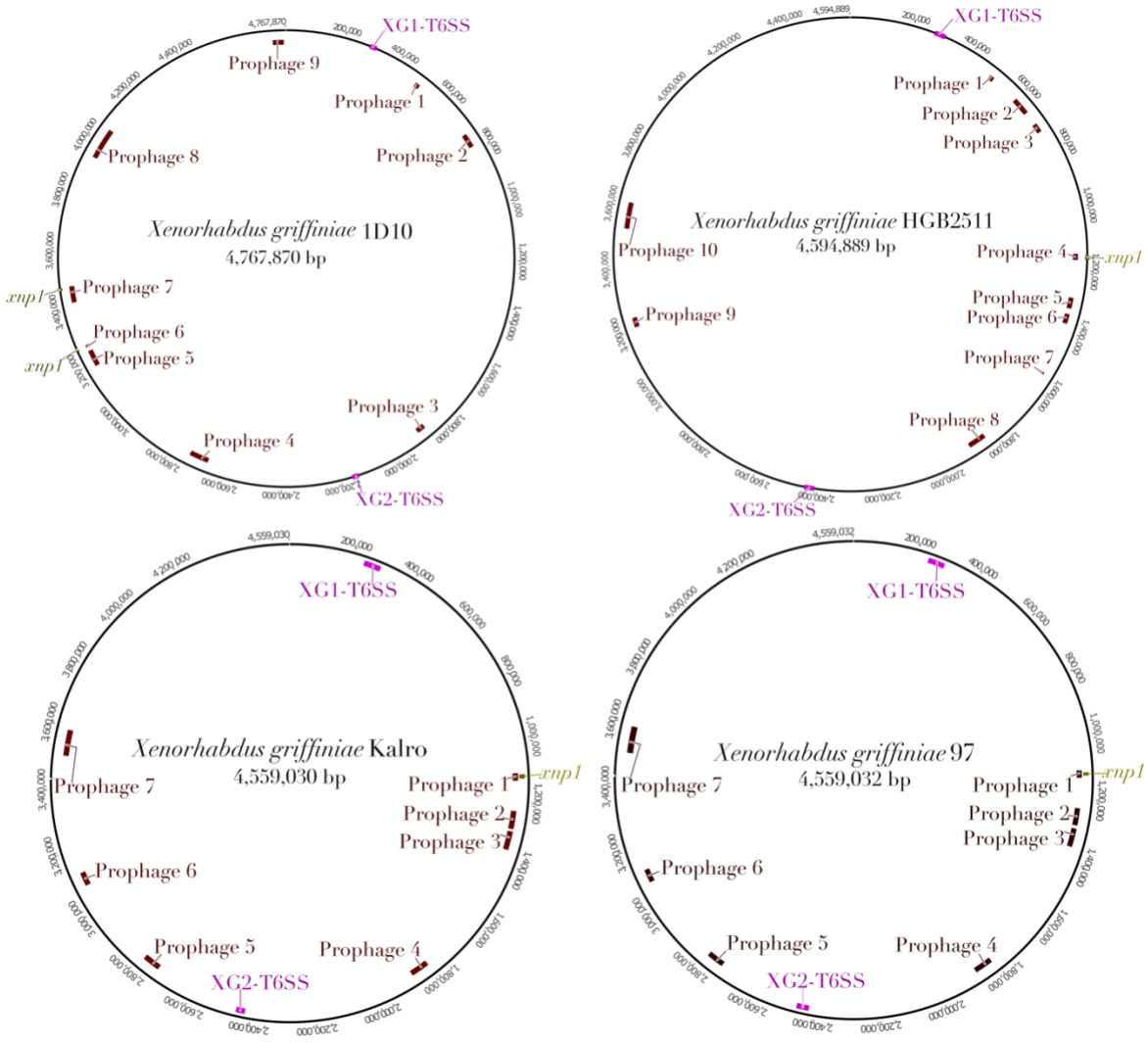


Fig. 3: Loci of prophages and gene clusters encoding complete type six secretion systems (T6SS) and xenorhabdicin (*xnp1*) in complete genomes of four *Xenorhabdus griffiniae* strains. Genomes *of X. griffiniae* and XN45 and VH1 were not included in this analysis as they were not circularised, which often results in a prophage locus being split over multiple contigs.

and that share similar gene content and synteny (Fig. S4 in Additional File 1). For each of these four loci, their encoded PAAR proteins are highly similar since their amino acid sequences had combined homogeneity indexes between 94-100% (Sheet S3 in Additional File 2). These findings indicate that in *X. griffiniae,* the T6SS spike and its cognate effector proteins may contribute to intraspecific traits. We extended these

findings by predicting other putative secretion systems encoded across the sequenced genomes using TXSScan (Table S1 in Additional File 1). The number of type I systems varied across the genomes analysed, and at least one copy of flagellum, Type 4a pilus (T4aP), Type 5a Secretion system (T5aSS), and Type 5b secretion system (T5bSS) were identified in all the genomes analysed.

Prophages mediated the acquisition of both subspecies-specific and strain-specific genes

We hypothesized that horizontal gene transfer was a major driver of subspeciation in *X. griffiniae*, since the mobilome constituted the largest fraction of functionally annotated, strain-specific genes (Fig. 4, sky blue=mobilome: prophages, transposons, plasmids), and strain-specific genes result in speciation when they confer ecologically useful traits [86].

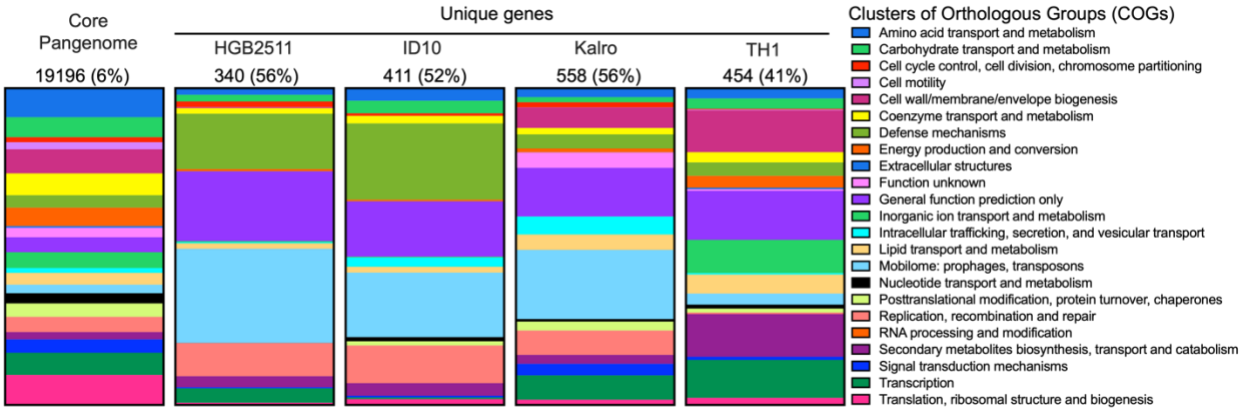


Fig. 4: Stacked bar charts depicting the Clusters of Orthologs Groups of proteins (COGs) of the core and unique genes among four closely related *Xenorhabdus* strains. The core pangenome bar shows the COGs of genes that are common to all genomes in the pangenome. The subsequent bars show the COGs of the unique genes from the HGB2511, ID10, Kalro, and TH1 genomes derived from the previously described pangenome analysis of *X. griffiniae* strains and *Xenorhabdus* sp. TH1 (Fig. 2A). The circularised Kalro genome was used to represent the XN45, VH1 and Xg97 genomes as these four genomes are highly similar

(>99.83% pairwise ANI values). The numbers on top of each bar are the total number of genes in each category, followed by the percentage of those genes without COG designation in parentheses.

We first investigated this by inferring the evolution of gene content among the six strains of *X. griffiniae* and *Xenorhabdus* sp. TH1.Strain BG5 was included in this analysis since it is most closely related known species that diverged just before TH1 in our phylogenomic reconstructions (Fig 1). We inferred that a net gene loss likely drove the speciation of *Xenorhabdus* sp. TH1 (Fig. 5), consistent with its smaller genome size when compared to those of *X. griffiniae* strains (Table 1).

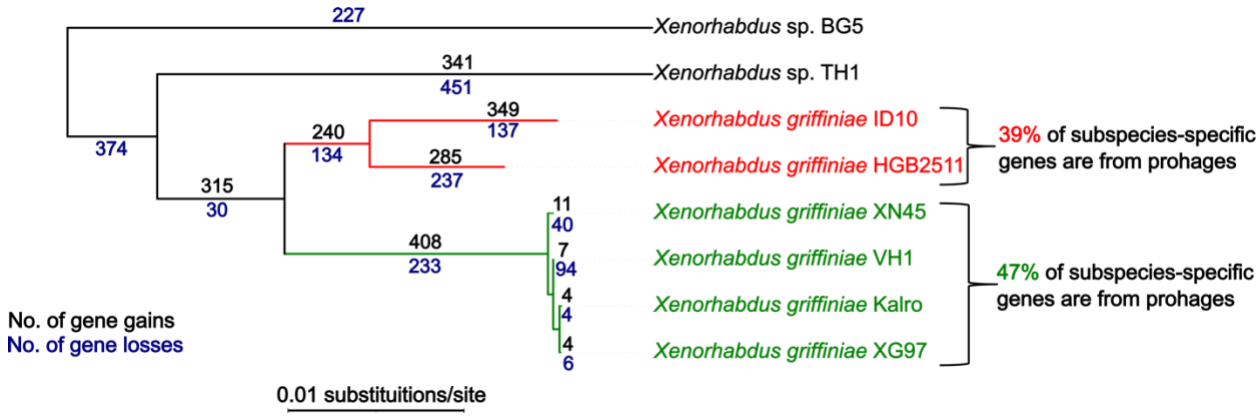


Fig. 5. Neighbour-joining phylogenomic tree depicting the evolution of gene content among strains of *Xenorhabdus griffiniae*. ThreeThree species are depicted in this tree, *Xenorhabdus* sp. BG5, *Xenorhabdus* sp. TH1 and *X. griffiniae*. For *X. griffiniae,* its India-Indonesia and Kenya subspecies are in red and green, respectively. The emergence of both subspecies was likely associated with net gene gains. Gene content analysis was conducted in COUNT applying Wagner parsimony.

Conversely, net gene gains possibly resulted in the formation of the two *X. griffiniae* subspecies. We addressed the question of whether horizontal gene transfer (HGT) may have mediated these gene gains by conducting a preliminary pangenome analysis of 49 *Xenorhabdus* strains. We found that the total number of phage-related genes accounted for 55% of the variation in the proteome sizes among *Xenorhabdus* genomes. In this analysis, the total number of phage-related genes accounted for 48.69% of the variation in proteome sizes among the seven strains (adjusted r^2=0.48691, p=0.04899983). Similar

correlations for transposable elements (adjusted r^2=0.23241, *p*=0.15) and plasmid-related genes(adjusted r^2= 0.1762, *p*=0.19) were insignificant. Based on this, we focused on the identification of subspecies-specific and strain-specific genes that were linked with prophages.

We first identified prophages in TH1, ID10, HGB2511, Kalro and xg97 (Fig. 3 and Fig. S2 in Additional File 1, Sheets S8-S11 and S15 in Additional File 2). Genomes of VH1 and XN45 were excluded as they were too fragmented to yield robust results. Taxonomically, all identified prophages in the five genomes belonged to the family *Caudoviricetes* (Sheet S15 in Additional File 2). The genomes of HGB2511, ID10, Kalro, 97 and TH1, had ten, nine, seven, seven, and three prophages, respectively.

To determine how these prophage numbers compared to those found in other strains, we similarly identified prophages in ten other *Xenorhabdus* strains whose chromosomal genomes were also assembled into one contig (Sheet S7 in Additional File 2). We found an average of seven prophages per genome, ranging from 3 to 13, indicating that HGB2511 and ID10 harbour higher-than-average numbers of prophages in their genomes. Three prophage loci were similar across genomes: 1) ID10 prophage 3 and HGB2511 prophage 3 which had 73% pairwise nucleotide percentage identities; 2) ID10 prophage 9 and HGB2511 prophage 5 with 61% pairwise nucleotide percentage identities (Fig. S3 in Additional File 1); and 3) the locus (*xnp1*) [87], whose conserved and variable regions (Fig. 6) had 95-96% and 51% average pairwise nucleotide percentage identities, respectively. These regions were previously elucidated in strains of *X. nematophila* and *X. bovienii* and shown to encode xenorhabdicin, an antimicrobial R-type pyocin, or tailocin

structure [88]. The *xnp1* locus, including genes essential for xenorhabdicin production and release, were detected in all seven genomes (Fig. 3; Fig. 6; Fig. S2 in Additional File 1).

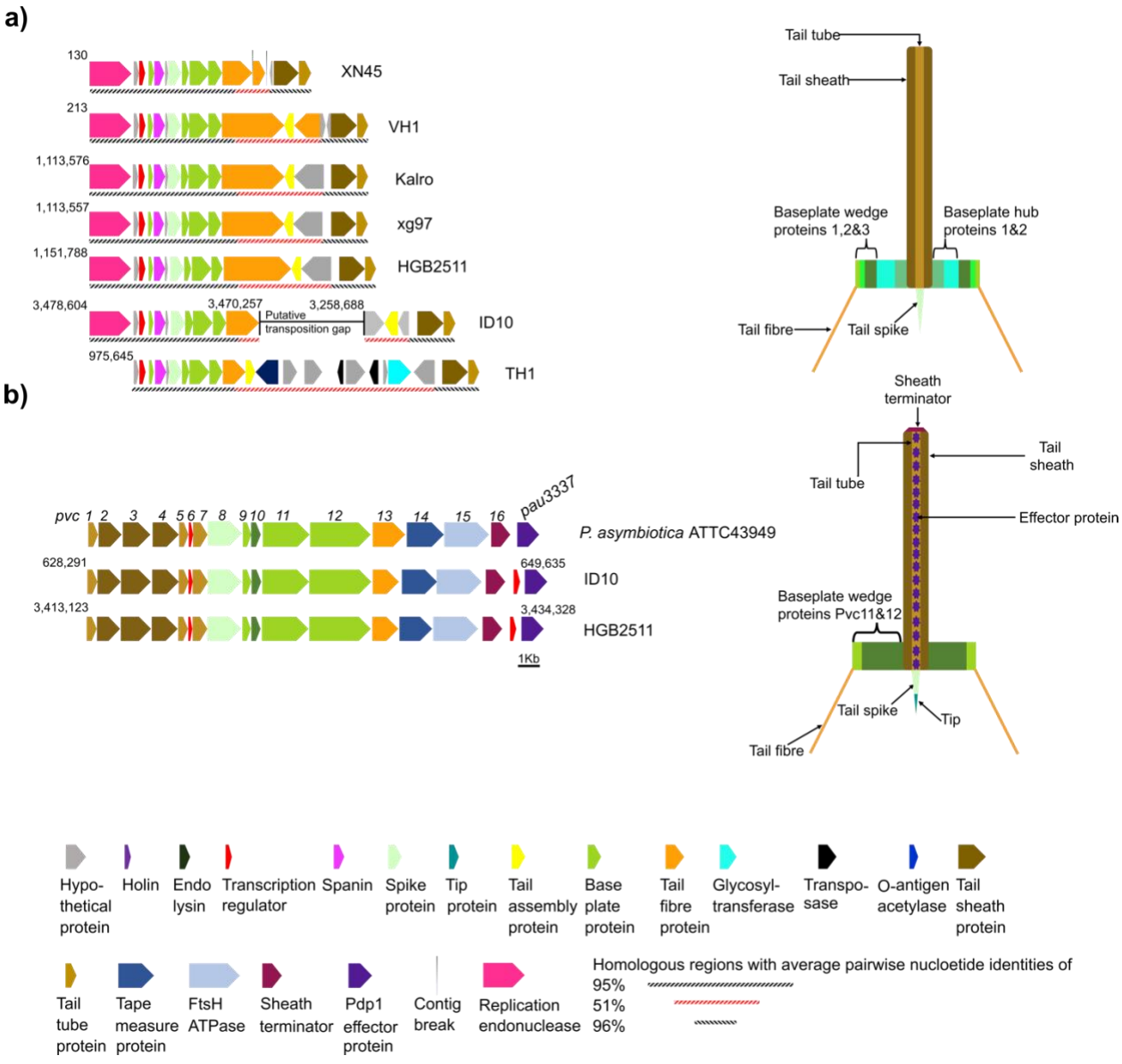


Fig. 6. Genomic loci encoding phage tail-like particles from *Xenorhabdus griffiniae*. A) The xenorhabdicin encoding genomic loci (*xnp1*) in six strains of *Xenorhabdus griffiniae* and *Xenorhabdus* sp. TH1 and a cartoon of the corresponding xenorhabdicin particle. B) Genomic loci (*pvc*) encoding an extracellular contractile injection system, the *Photorhabdus* Virulence Cassette (PVC) in *P. asymbiotica* ATCC43949, *X. griffiniae* ID10 and HGB2511. The genome coordinates of the loci are shown for the circularised genome assemblies. For XN45 and VH1, the shown coordinates are for contigs JACWFC010000129 and JADEUF010000065, respectively.

628 Comparative analysis of *xnp1* loci revealed that in ID10 it has been split into two loci that
629 were over in different parts of the genome(Fig. 3 & Fig. 6), probably due to transposition
630 events. The TH1 *xnp1* locus included genes that encode O-antigen acetylase and
631 glycosyltransferases (Fig. 6) which may be involved in conferring immunity to
632 xenorhabdicins through modification of its likely receptor, lipopolysaccharide O antigen
633 [89].
634
635 To identify the specific genes contained in the identified prophage regions, we took ID10
636 prophages as an example (Fig. 7). Genes predicted to encode viral replication and
637 hypothetical proteins constituted 35 and 15 percent, respectively (Sheet S19 in Additional
638 File 2) whereas 'cargo' genes with non-virus, functional annotations constituted the
639 remaining half. These annotated cargo genes encoded diverse products, including toxin-
640 antitoxin systems that have wide-ranging effects on bacterial physiology and mobile
641 genetic elements within genomes [90], Importin-11, predicted to encode a nuclear
642 transport receptor that presumably would be delivered for modulation of animal host cell
643 physiology [91], and diguanylate cyclase, predicted to be part of a signal transduction
644 cascade mediated through the second messenger cyclic-di-GMP [92] (Fig. 7).
645

646

647 **Fig. 7:** Graphical representation of predicted products of genes contained within prophage loci of
648 *Xenorhabdus griffiniae* ID10. The genes could be categorized into three broad categories; annotated cargo
649 genes, viral replication genes and those whose products are unknown.

650

651 The identified prophages contained 45 and 38% of strain-specific genes in the HGB2511
652 and ID10 genomes, respectively. Likewise, 39 and 47% of India-Indonesia and Kenyan
653 subspecies-specific genes were from prophages (Fig. 5). Indeed, removal of prophage
654 regions from the four genomes elevated their pairwise dDDH values: Pairwise dDDH

values between ID10 and strains HGB2511, Kalro, TH1 and xg97 rose by 3-3.2 percentage points when identified prophages were removed from all genomes (Sheet S12 in Additional File 2). These findings demonstrate that, in *X. griffiniae*, a considerable proportion of both subspecies and strain-specific genes were gained through prophages.

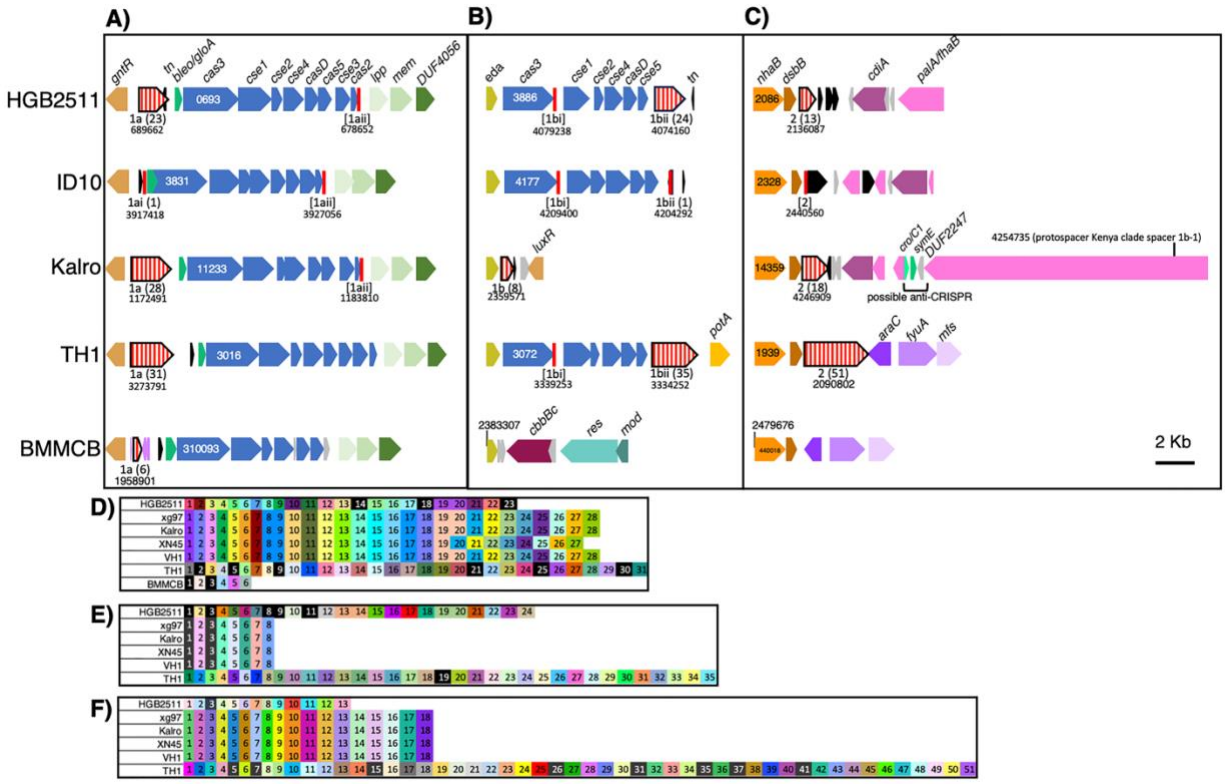


Fig. 8: CRISPR-Cas regions of *Xenorhabdus* strains. A-C) Comparison of three CRISPR-Cas related genomic regions in selected strains, anchored for synteny in the diagram using *gntR* **(A)**, *eda* **(B)**, or *nhaB* **(C)**. ORFs are indicated by solid block arrows (blue for *cas* genes, with the locus tag number of *cas3* provided) with annotated gene names indicated above, and identical colours indicating homology. CRISPR repeat arrays (red vertical stripe block arrows) for each region (1a, 1b, 2) had variable number of repeats (noted in parentheses). Start coordinates for each are shown underneath. Degenerate repeat arrays are indicated by brackets. A conserved sequence [1aii; vertical red line] containing a repeat, a spacer, and a degenerate repeat) was identified at the end of *cas2,* in all strains except BMMCB, which lacks this gene **(A)**. Another [1bi] was apparent at the end of *cas3* in the *eda* locus of HGB2511, ID10, and TH1 **(B)**. **D-F)** The spacer sequences of each CRISPR array found at *gntR* **(D)**, *eda* **(E),** or *nhaB* **(F)** were compared for identity to each other or other loci among the analysed strains. Each box represents a spacer, and different colours indicate different sequences. Spacers represented by black boxes and white

lettering have 100% identity to “target” loci outside of the array, either within the same genome or within one of the other genomes analysed here. *gntR:* DNA-binding transcriptional repressor; *bleo/gloA*: bleomycin resistance/glyoxalase; *tn*: transposase; *lpp:* lipoprotein; *mem*: membrane protein: DUF4056: domain of unknown function 4056 gene; *eda:* 4-hydroxy-2-oxoglutarate aldolase; *nhaB*: Na(+):H(+) antiporter NhaB; *dsbB:* protein thiolquinone oxidoreductase; *cdiA*: Deoxyribonuclease CdiA; *palA/fhbA*: filamentous hemagglutinin; *luxR*: LuxR family transcriptional regulator; *cro/CI*: HTH cro/C1-type domain-containing protein; *symE*: Type I addiction module toxin, SymE family; *potA*: spermidine preferential ABC transporter ATP binding subunit; *araC:* AraC family transcriptional regulator; *fyuA:* Putative TonB-dependent siderophore receptor; *mfs*: putative MFS transporter, signal transducer; *cbbBC*: Molybdopterin-binding oxidoreductase; *res*: Type III restriction endonuclease subunit R; *mod*: site-specific DNA-methyltransferase (adenine-specific). The circularised Kalro genome was used to represent the XN45, VH1 and Xg97 genomes as these four genomes are highly similar (>99.83% pairwise ANI values).

Comparative analysis of CRISPR-Cas, protospacer, and anti-CRISPR content

The high prevalence of prophages and prophage-mediated gene gains in *X. griffiniae* genomes suggests that these bacterial symbionts relatively frequently encounter phage-related foreign DNA. This prompted us to investigate the presence or absence of defence systems in three *X. griffiniae* strains HGB2511, ID10, and Kalro (representative of the XN45, VH1, and Xg97 genomes), as well as TH1 and BMMCB. Consistent with the diversity of defence systems observed among bacterial genomes, including between strains, each of the five strains analysed here encoded a unique repertoire of defence systems relative to the others [93] (See Table S4 in Additional file 3). Each genome had at least one type of restriction-modification system, which is described in more detail below, as well as at least one type of toxin/antitoxin system, predicted to encode a stable toxin and an unstable anti-toxin. All of the examined strains contained a copy of the Type II toxin-antitoxin system, MazEF that can elicit cell death in response to phage infection in a process known as abortive infection [94]. Abortive infection, which can be triggered by diverse defence pathways, kills the infected cell before phage reproduction can occur,

thereby protecting the rest of the population [95]. All of the strains also encoded a Class 1-Subtype-I-E CRISPR-Cas immunity system [90]. Because CRISPR-Cas systems can yield insights into strain diversity and prior history of phage exposure [96], we further explored the content of these loci.

The genomes of each of the analysed *X. griffiniae* genomes and the close relative TH1 encode a syntenic locus containing a full set of Class1-Subtype-I-E *cas* genes encoded adjacent to a *gntR* homolog (Fig. 8A; Table S5 in Additional file 3). The former closest known relative of ID10, BMMCB, has an incomplete set of *cas* genes on a single contig, with *casD* and *cas2* lacking. However, since the BMMCB genome is fragmented, we cannot rule out the possibility that these genes are encoded elsewhere. HGB2511, ID10, and TH1 also have a second set of Class1-Subtype-I-E genes (but lacking *cas1* and *cas2*) encoded adjacent to an *eda* homolog (Fig. 8B). All strains except ID10 and BMMCB also have three CRISPR arrays, comprising conserved repeats and variable targeting spacers that are predicted to be transcribed and cleaved into non-coding, small (61 nt), targeting CRISPR RNAs (crRNAs): array 1a, adjacent to the full set of *cas* genes at the *gntR* locus (Fig. 8A,D), array 1b, adjacent to the *eda* homolog (Fig. 8B,E)**,** and array 2, adjacent to a *dsbB* gene (Fig. 8C,F). BMMCB lacks the CRISPR arrays adjacent to *eda* and *dsbB*. Instead, BMMCB has a second CRISPR array with just two spacers in another region of the genome (at coordinate 3720455, not shown in figure) that encodes phage-related genes. This array falls at the edge of a contig break, so may not be an accurate reflection of the repeats that might be present at this locus.

Since spacer sequences are acquired in a directional manner in response to active infection by foreign nucleic acid material (e.g., phages or plasmids), comparisons of spacer content across related strains can be used to infer their shared life histories and prior exposure to such threats [97, 98]. The spacer contents of CRISPR arrays 1a, 1b and 2, is variable in number and sequence across the strains (Fig. 8D-F). Consistent with the close relatedness of the Kenyan subspecies strains xg97, Kalro, XN45, and VH1, their CRISPR spacer content is identical, except for the absence in XN45 of a duplicated spacer found in array 1a of the other strains (Fig. 8D). Otherwise, there is no overlap in spacer identity among CRISPR arrays of the different strains, indicating their divergence prior to the acquisition of existing spacer content. Compared to the other *X. griffiniae* and the close relative TH1, the CRISPR arrays in ID10 appear to have a limited number of targeting spacers. Based on the presence of conserved repeat sequences, ID10 encodes two (1ai and 1bii) bona fide, single repeat CRISPR loci at locations syntenic with regions 1a and 1b of the other strains. Three other loci (1aii, 1bi, and 2) appear to be remnants, with only a single clear left repeat, a spacer, and a degenerate right repeat (Fig. 8A-C; Additional File 3).

Table 2: Annotated spacer-protospacer content among *X. griffiniae* and related strains

Source Genome	Spacer ID	Spacer Sequence (5'-3')	HGB2511	ID10	Kalro	TH1	BMMCB
HGB2511	1a-14	CCAATGGTTCGAGAATTTTTATTGATACACCA					
HGB2511	1a-18	TACTCCGATCCGACCGTCTATTCAGAAGATGG					
HGB2511	1a-23	CCCAATTCAGGCATAGTAAGACCCCGATAAAC					
HGB2511	1b-1	CCTTCTCGCCACTGCATATTTACTACAAAGAC					
HGB2511	1b-3	CATCCCACTCAGAAATTCCGCGTTGGGATAAA					
HGB2511	1b-8	GGAATAATGCCCGTGATACGCAGCTCTTTCGGC					
HGB2511	1b-9	AGGTTATTCACATGGCAGATTGCCCGGTAGAA					
HGB2511	1b-11	AACGTTACAACCAGTGCCAGAATATCAATTGT					
HGB2511	2-3	CTGAACAGGGGGACGGATTCAGCCGCCGTGAT					
Kalro	1b-1	ACAATCGTGCCATCACTGACCGCTGCATGTGT			SELF		
Kalro	1b-3	GTCAGTAGAAATTTCTTTTTAGTACCTAAAGA			SELF		
TH1	1a-2	GCATTGCCGAAGACCAGATAATCCAGCACATA					
TH1	1a-5	CATTGATAATGTCAAAGTACTGGAACGCATTC				SELF	
TH1	1a-9	GCATTGATCCGCTGGAAACCCTCGTCTTATTG					
TH1	1a-21	CAATGTTCCGCGTTCACGGGGATTATCCCACC					

TH1	1a-25	GCATGGAATCACATCCCGCTTCTCTGTTGGGA					
TH1	1a-30	GAAAACGCTTTTTGTCTCAGGATGTAATACAG					
TH1	1b-19	CGACCCCTATCGCTTTTCCTGACTTACTGTCT					
TH1	2-5	CCATTCAGTTTGTCCGCTATACGGTGATGCCG					
TH1	2-7	AAACCAACCCAGAGTTACCATATTCGCCAGAC					
TH1	2-15	TTCATGTAATTTATGTTGGTGTTCGTCCAGTT					
TH1	2-26	CATAAAGGTTTGGAAAAATTCTTTAAAACAGG					
TH1	2-31	GCCCTATTTAACATAATGGCTCTTAGGCGA					
TH1	2-35	CGGCTATCCGATGTAATCACATATTTACCCAA					
TH1	2-37	CTGGAAGAGAAAAAACGCCGGCTGGGTCTCGG					
TH1	2-41	GCTGCTTGCCATCCTTGAGCAACACGAAACCG					
BMMCB	1a-1	TTCACGCAAATCAGCCAGATCGATGTTGCCGT					
BMMCB	1a-3	CATCGCAAAATATCATCAAACTGTTGCGGGAT					SELF

Protospacer annotations	
	Spacer-encoding strain
	phage region gene
	restriction modification system
	TIGR03750 conjugal family protein
	DUF3277 domain-containing protein
	DUF2829 domain-containing protein
	gene of unknown function
	filamentous hemagglutinin
	ABC transporter
	folD
	gcvP

To gain insights into the types of threats encountered by *X. griffiniae* and related strains, we searched for putative target sequences (known as protospacers) based on their identity with CRISPR array spacer sequences. We found some spacers have 100% identity to protospacers either within the same genome (self-targeting) or within one of the other genomes analysed here (Table 2; Table S6 in Additional File 3). In many cases, these protospacers were within phage-related, conjugation machinery, and restriction modification systems, in line with the role of CRISPR systems in defending against these types of mobile genetic elements [99, 100]. Consistent with their close relationship, the Kenya clade demonstrated identical protospacer content in genes throughout the genome, including several predicted to be targeted by CRISPR small RNAs from HGB2511 (Table 2; Table S6 in Additional File 3).

Additional functional genes with protospacer sequences that could be targeted by crRNA included those predicted to encode filamentous hemagglutinin (Kenya clade spacer 1b-

1), an ABC transporter (Kenya clade spacer 1b-3), and the enzymes FolD (TH1 spacer 1a-5) and GcvP (BMMCB spacer 1a-3). Curiously, the *palA/fhaB* filamentous hemagglutinin gene with self-identity to the Kenya clade spacer 1b-1 is encoded in the *dsb* locus, in proximity to the Kenya clade spacer region 2 (Fig. 8C). Since spacer self-identity would presumably result in self-intoxication, we hypothesize the genome also encodes an anti-CRISPR immunity mechanism such as anti-CRISPR proteins known as Acr. These proteins are difficult to predict with sequence similarity because they vary widely [100]. We manually searched for such loci in the selected genomes using a “guilt-by-association" approach of putative Acr by identifying small open reading frames in proximity to the protospacer-containing gene and a helix-turn-helix (HTH) domain-containing gene, which is predicted to be the anti-CRISPR regulator. Of the self-targeting protospacers we detected, only the one in *palA/fhaB* of the Kenyan subclade had a promising candidate based on these criteria (Fig. 8C; Additional File 3). The putative Acr is a DUF2247 domain-containing protein (e.g., JASDYB01_14371) which is predicted to encode a protein of 171 aa and is encoded near an HTH cro/C1-type domain-containing protein (e.g., JASDYB01_14369) that may be a putative Aca transcriptional regulator [100].

Restriction Modification Systems

In addition to CRISPR arrays, restriction-modification systems resist the introduction of foreign DNA, including phage infection, by detecting and cleaving non-chromosomal DNA. Restriction-modification systems can be classified based on their structure, cofactor requirements, DNA recognition site and relative cleavage locations [101, 102]. Type I, II, and III all encode both a restriction endonuclease and a methyltransferase, whereas Type

IV endonucleases cleave modified DNA (such as 5-hydroxymethylcytosine) and variably encode an adjoining methyltransferase [102]. There are also anti-restriction proteins which inhibit restriction modification systems by various mechanisms. We predicted the number and type of complete restriction-modification systems and anti-restriction proteins in the BMMCB, TH1, HGB2511, ID10 and Kalro genomes (Table S2 in Additional File 3). Type I and II systems were the most prevalent across the genomes. Only the HGB2511 genome encoded a complete Type III system, while ID10, Kalro, and TH1 genomes each encoded Type IV restriction endonucleases. The genomes of all *X. griffiniae* and *Xenorhabdus* sp. TH1 encoded at least one anti-restriction protein, with ID10 appearing to encode seven anti-restriction proteins, by far the largest complement (Table S2 in Additional File 3).

X. griffiniae **encode the biosynthesis of diverse natural products**

A *Xenorhabdus* bacterium occupies the uncanny ecological role of both a mutualist of soil-dwelling nematodes and a parasite of diverse insects. A means it uses to juggle this Jekyll and Hyde lifestyle is the production of potpourri of peptides, polyketides, siderophores, aminoglycosides, and toxin proteins and complexes. Hence, *Xenorhabdus* genomes often comprise loci, known generally as biosynthetic gene clusters (BGCs), which are responsible for the biosynthesis of some of these natural products. A BGC can include many genes, often under the control of one promoter, which collectively encode the production pathway of a single natural product and its derivatives. Commensurate with its genus, which ranks among those that produce the most diverse set of natural products [103], the *X. griffiniae* ID10 genome contained over 21 biosynthetic gene

clusters that were predicted to encode the production of over ten different types of natural products (Fig. 9; Table S3 in Additional File 1).

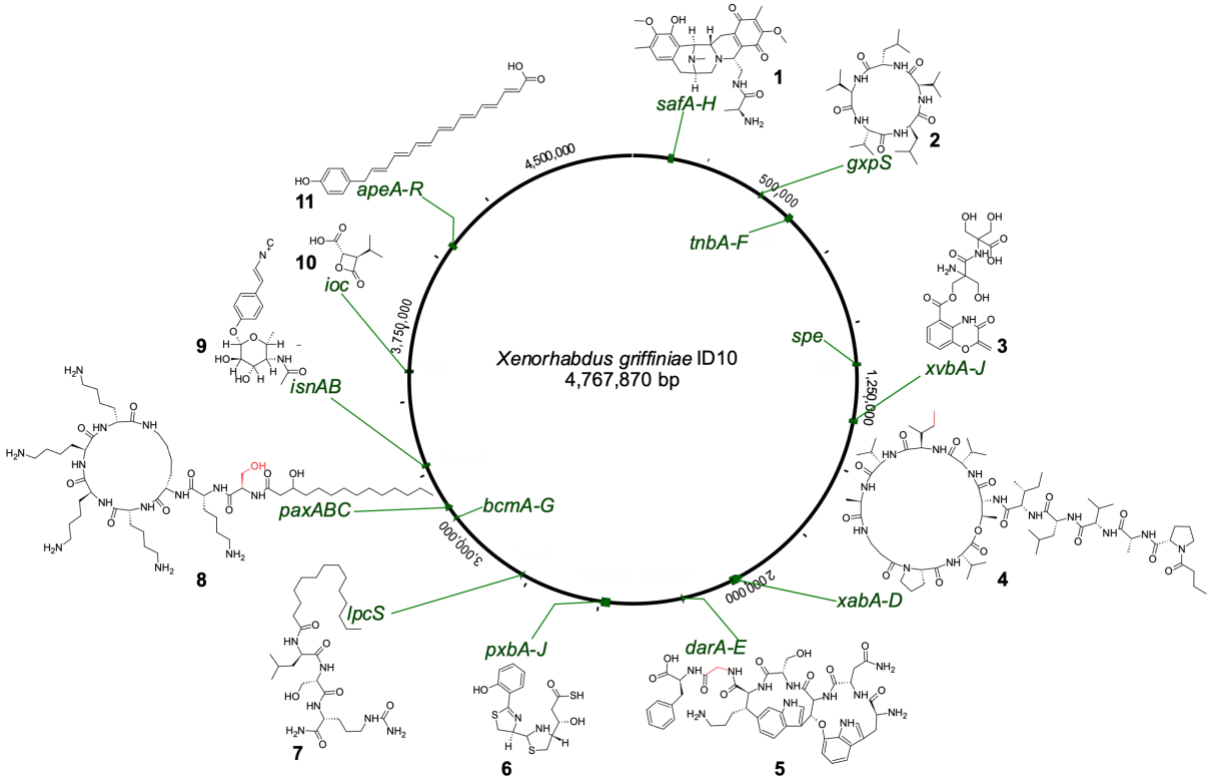


Fig. 9: Genomic loci of known biosynthetic gene clusters in the *X. griffiniae* ID10 genome and predicted chemical structures of the natural products whose biosynthesis they encode. The *paxABC*, *darABCDE* and *xabABCD* BGCs were predicted to encode the production of potentially novel derivatives of PAX peptides (8), darobactin (5) and xenoamicin (4), respectively, that differed from known structures in amino acid building blocks at positions highlighted in red. Safracin (1), gameXpeptide C (2), benzobactin (3), photoxenobactin (6), type 2 bovienimide (7), rhabduscin (9), 3-isopropyl-4-oxo-2-oxetanecarboxylic acid (10), arylpolyene (11).

Fourteen of these ID10 biosynthetic gene clusters (BGCs) are predicted to encode the biosynthesis of either known compounds or their derivatives (Fig. 9; Table S3 in Additional File 1). For example, the *lpcS* and *isnAB,GT* BGCs in the ID10 genome are predicted to encode the production of group IIA bovienimides [104] and rhabduscin [105], respectively, both of which are insect immunity suppressors, the *pxb* BGC which encodes

819 the production of the insecticidal photoxenobactins [20], and *safA-H, ioc/leu, xvbA-J,*
820 *bcmA-G,* and *ape* BGCs which respectively encode the production of safracin antibiotics,
821 3-isopropyl-4-oxo-2-oxetanecarboxylic acid (IOC), benzobactins, bicyclomycins, and aryl
822 polyenes [20, 106–108]. The *gxpS* BGC was predicted to encode the synthesis of
823 GameXPeptide C [109], as their predicted peptide sequence was ^{D}Val-^{L}Val-^{D}Leu-^{L}Val-
824 ^{L}Leu. In contrast, each member of the *X. griffiniae* India-Indonesia subspecies had a
825 unique set of BGCs. Specifically, the HGB2511 strain lacked BGCs that encoded the
826 production of rhabduscin, benzobactin, bicyclomycin and actinospectacin, all of which
827 were present in the ID10 genome (Table S3 in Additional File 1).
828
829 Notably, the ID10 genome contained known BGCs but the predicted biosynthetic products
830 are previously unknown derivatives. For example, the ID10 *paxABC* BGC, which encodes
831 the biosynthesis of PAX peptides, is predicted to encode an heptapeptide backbone of
832 ^{L}Ser-^{L}Lys-^{L}Lys-^{D}Lys-^{D}Lys-^{D}Lys-^{D}Lys, which differs from those of *X. nematophila* [110] and
833 *X. khoisanae* [111] by having ^{L}Ser at position one instead of Gly, since the respective
834 Stachelhaus code was DVWHLSLIDK and not DILQIGLIWK. The *xabABCD* BGC is
835 predicted to encode the biosynthesis novel xenoamicins that incorporate ^{D}Ile in lieu of
836 ^{D}Val [112] at position eight of the tetradodecapeptide backbone. However, since
837 predictions based on Stachelhaus codes are prone to inaccuracies, only the chemical
838 structure elucidation of these peptides can determine whether they are indeed novel
839 derivatives. The synthesis of novel derivatives is also predicted for BGCs that encode the
840 biosynthesis of the known ribosomally-synthesized and post-translationally modified
841 peptide (RiPP), darobactin [113], since the ID10 *darABCDE* BGC was predicted to

encode the biosynthesis a core peptide with the sequence Trp-Asn-Trp-Ser-Lys-Gly-Phe and not Trp-Asn-Trp-Ser-Lys-Ser-Phe.

X. griffiniae encode entomotoxins and are insecticidal to *Manduca sexta*

An essential part of the *Steinernema-Xenorhabdus* entomopathogenic lifecycle, is the ability of the host-symbiont pair to infect and kill insect hosts, and *Xenorhabdus* produce virulence factors that target other microorganisms competing for the nutritious insect cadaver [7]. To better understand the toxic potential of the *X. griffiniae* bacteria, we mined the HGB2511, ID10, Kalro and TH1 genomes for toxin-domain-containing loci. Using a list of known toxins found in other *Xenorhabdus* [19, 71] we identified homologs of genes encoding the known insecticidal toxins Mcf “makes caterpillars floppy” and PirAB in each of the strains [114–116], along with proteins homologous to the MARTX toxin family (Table 3). Similar to *X. innexi* HGB1681, the MARTX proteins in HGB2511, ID10, Kalro and TH1 each lack four of the A repeats at the *N*-terminus of the protein (A Δ3-7), leaving nine repeats compared to the 14 found in *X. nematophila* 19061, *X. bovienii* SS-2004 (Jollieti), and *Vibrio* species [71, 117] (Fig. S5 in Additional File 1). It remains unclear how these differences in repeat structure might impact MARTX protein function in *X. griffiniae*.

Insecticidal Tc toxins are three-part toxin complexes (TcA, TcB, and TcC type) commonly found in entomopathogenic bacteria, including in some *Xenorhabdus* species [118]. Tc toxin family proteins were notably absent from the *X. griffiniae* HGB2511, ID10 and Kalro genomes and the *Xenorhabdus* sp. TH1 genome (Table 3). No evidence was found of homologs of Shiga toxin (*stx1a*) related genes, such as those found in *X. bovienii* (Table

3) [19]. Within the ID10 genome we identified a putative insecticidal toxin (XGHID_v1_0629), a homolog of which was also found in the HGB2511 genome (XGHIN1_v1_3228). These genes are homologs of a *Photorhabdus asymbiotica* gene (PAU_03337) (Fig. 6B) that encodes *Photorhabdus* dNTP pyrophosphatase 1 (Pdp1), a cytotoxic protein that not only kills immune cells by reducing their intracellular deposits of deoxynucleotide triphosphates (dNTPs) but is also an effector protein of the extracellular contractile injection system (eCIS) known as *Photorhabdus* virulence cassette (PVC) [119]. Indeed, analysis of genes upstream *pdp1* revealed that both ID10 and HGB2511 encode PVCs (Fig 6B). Although PVCs are phage tail-like particles that are structurally similar to xenorhabdicin (Fig 6B), they differ by having within their tube, effector proteins that are translocated into the target cell, upon tail fibre-mediated binding and subsequent tail sheath contraction [120]. Notably, when loci that encode eCIS and T6SS were found on the same genome, core genes often lacked from T6SS encoding loci [121], probably explaining why *tssK, tssJ* are absent from the XG2-T6SS encoding locus of ID10.The *N*-terminus (50aa) of the *Photorhabdus asymbiotica* Pdp1 acts as a signal peptide for secretion through the PVC [122]. Amino acid alignment of the ID10 and HGB2511 Pdp1 proteins with the *Photorhabdus* Pdp1 and two non-PVC secreted homologs [122] revealed amino acids at the *N*-terminus of the *Xenorhabdus* proteins that could act as a signal peptide for PVC secretion (Fig. S6, Additional File 1).

A *de novo* search for other toxin homologs using PathoFact software on our genomes of interest revealed two strain-specific loci of particular interest [72]. The *Xenorhabdus* sp. TH1 genome contains a complete hydrogen cyanide synthase locus (*hcnABC)* (Table 3).

hcnABC is found in plant-associated and entomopathogenic bacteria [123] where it plays a role in insect killing. Notably, *hcnABC* was recently identified in the genome of a steinernematid-associated *Pseudomonas piscis* bacterium [83]. In the *X. griffiniae* Kalro genome, three proteins with zonular occludens toxin (zot) domains were identified (JASDYB01_14222, JASDYB01_14237, JASDYB01_14239) (Table 3). Zonula occludens toxin (Zot) domain-containing proteins target the eukaryotic cell cytoskeleton and compromise the structure of intercellular tight junctions, leading to a permeabilization of epithelia [124, 125]. Homologs of the three Zot domain-containing proteins found in Kalro were also identified in xg97, VH1 and XN45, and BLASTp revealed other Zot domain-containing proteins in other *Xenorhabdus* species (including *X. bovienii, X. khoisanae, X. eapokensis, X. ehlersii* and *X. innexi*). JASDYB01_14222 and JASDYB01_14237 are each predicted to encode a transmembrane helix and to be membrane embedded, whereas JASDYB01_14239 is a considerably shorter peptide lacking both transmembrane domains and secretory signals. These Zot domain-containing proteins may affect the insect midgut as part of an oral route of infection [126], or destroy insect epithelial tissues when the bacteria are released into the hemocoel.

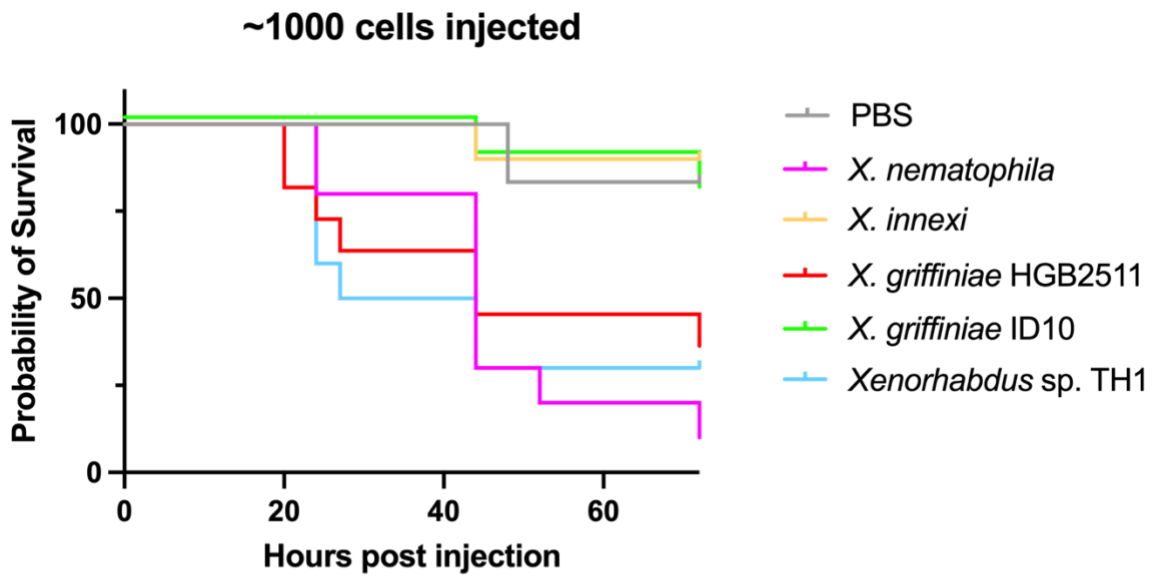
Table 3: Putative entomotoxin-encoding genes in *Xenorhabdus griffiniae* and TH1 genomes

Putative toxin gene name	blast sequence	2511	ID10	Kalro	TH1
mcf	XNC1_2265	XGHIN1_v1_3249	XGHID_v1_0647	JASDYB01_13066	XTH1_v2_2642
MARTX	XNC1_1381	XGHIN1_v1_2845	XGHID_v1_1587	JASDYB01_13503	XTH1_v2_2371
prtA	XNC1_4025	XGHIN1_v1_3782	XGHID_v1_0525	JASDYB01_12528	XTH1_v2_551
HIP57 (GroEL)	XNC1_3605	XGHIN1_v1_3711	XGHID_v1_1038	JASDYB01_12581	XTH1_v2_2959
xaxAB	XNC1_ 3766, XNC1_3767	none found	XGHID_v1_0927, XGHID_v1_0928	JASDYB01_12687, JASDYB01_12688	none found
xhlA* and *B*, *tpsA* and *B	XNC1_4556, XNC1_4555	XGHIN_v1_0252, XGHIN_v1_0253	XGHID_v1_0280, XGHID_v1_0281	JASDYB01_11642, JASDYB01_11641	XTH1_v2_0208, XTH1_v2_0209
***xhlA* hemolysin**	XNC1_3177	XGHIN1_v1_1428	XGHID_v1_3009	JASDYB01_10378	XTH1_v2_1082

stx1a	XBI1v2_2730004	none found	none found	none found	none found
xptA2	XNC1_2569	none found	none found	none found	none found
Tc toxins (A)	XNC1_2333 + 2334, XNC1_2560 + 2561, XNC1_2566, XNC1_2569, XNC1_3020 + 3021 + 3022 + 3023 + 3024, and XNC1_2187	none found	none found	none found	none found
Tc toxins (B)	XNC1_2186, XNC1_2335, XNC1_2568	none found	none found	none found	none found
Tc toxins (C)	XNC1_2188, XNC1_2336, XNC1_2567	none found	none found	none found	none found
Chitinase	XNC1_2562	none found	none found	none found	none found
putative chitinases	XGHIN1_v1_0490	XGHIN1_v1_0490	XGHID_v1_3568	JASDYB01_12511	none found
putative chitinases	XGHIN1_v1_3307	XGHIN1_v1_3307	XGHID_v1_0504	none found	none found
putative chitinases	XGHIN1_v1_4103	XGHIN1_v1_4103	none found	none found	none found
Pir toxins	XNC1_1142, and XNC1_1143	XGHIN1_v1_1121	XGHID_v1_3432	JASDYB01_10835	XTH1_v2_858
TPS-Fha (txp40)	XNC1_1129	none found	none found	none found	none found
Xenocin	XNC1_1221–1223	XGHIN1_v1_1178	XGHID_v1_3165	JASDYB01_10762	XTH1_v2_0927
insecticidal toxin (as annotated in ID10 genome)	XGHID_v1_0629	XGHIN1_v1_3228	XGHID_v1_0629	none found	none found
hcnABC	XTH1_v2_1430-1432	none found	none found	none found	XTH1_v2_1430-1432
zonula occludens toxin	JASDYB01_14222	none found	none found	JASDYB01_14222	none found
zonula occludens toxin	JASDYB01_14237	none found	none found	JASDYB01_14237	none found
zonula occludens toxin	JASDYB01_14239	none found	none found	JASDYB01_14239	none found

906

907



	PBS	*X. nematophila*	*X. innexi*	*X. griffiniae* HGB2511	*X. griffiniae* ID10	*Xenorhabdus* sp. TH1
% Mortality	16.6666667	90	10	60	20	70
% Mortality with Abbots Correction	0	88	-8	52	4	64

Fig. 10: Comparative entomopathogenicity of *Xenorhabdus* bacteria to *Manduca sexta* larvae. Lines represent survival curves for insect larvae injected with approximately 1000 cells of each strain. Values have been corrected with Abbot's formula within the table below.

The diversity of toxin coding potential within the analysed *Xenorhabdus* genomes suggest possible differences in their entomopathogenicity. To begin to interrogate this possibility, we assessed the survival of fifth instar *Manduca sexta* insect larvae over a 72 h period after injection with five *Xenorhabdus* strains at three concentrations. We aimed to compare the strains at an inoculum of ~1000 cells because at that dosage our controls, *X. nematophila* (19061) and *X. innexi* (HGB1681) were previously shown to induce near to 100% lethality and <10% lethality, respectively [71], and indeed, these trends were recapitulated in our study (Fig. 10). Each bacterial strain displayed a dose-dependent survival response, with the highest inoculum resulting in the greatest mortality (Fig. S7, Additional File 1). Insects injected with ~1000 cells of *X. griffiniae* ID10 displayed robust

survival, like *X. innexi,* whereas greater than 50% of the animals injected with approximately the same number of *X. griffiniae* HGB2511 and *Xenorhabdus* sp. TH1 cells succumbed within 72 h of injection, like the level observed for *X. nematophila* (Fig. 10). These results reveal that the isolates tested have different levels of virulence against Lepidopteran insects.

Discussion

As part of their symbiotic and entomopathogenic lifecycle, all *Xenorhabdus* must colonize and be transported by a nematode host, suppress insect immunity, establish a community within and consume the cadaver, and support reproduction of the nematode to ensure future transport [10]. Many will also compete or cooperate with other resident or transient microbial community members and respond to variations in abiotic factors and higher-order trophic interactions. Here we conducted a comparative genomics analysis to gain insights into the consequences of such variable selective pressures on the evolution of *Xenorhabdus* genome content. Our analysis centred around seven related *Xenorhabdus* strains, chosen based on their close phylogenetic proximity to *X. griffiniae*. Strains of this *Xenorhabdus* species are particularly relevant to the development of laboratory model systems used to study nematode-bacteria symbiosis because they are symbionts of *S. hermaphroditum,* currently the most genetically tractable steinernematid [24, 127, 25].

To allow detailed comparative genome analyses, we assembled new high-quality, circularised genomes for three strains: *X. griffiniae* ID10^{T}, HGB2511, and TH1 (Table 1) [27, 24, 80]. High-quality genomes were already available for four other strains previously

identified as *X. griffiniae*: xg97, Kalro, XN45 and VH1 [16, 83]. HGB2511 and the Kenyan isolates were verified as *X. griffiniae* species, as defined by a genome with pairwise values for dDDH and ANI that are >70% [128] and >95% [48] thresholds, respectively, with that of the type strain ID10 (Figure 1) [27]. Consistent with previous findings [16], our analysis confirmed that strain BMMCB is not an *X. griffiniae* strain as originally designated [129]. Instead, we found that strain BMMCB is likely conspecific with *Xenorhabdus* sp. SF857 (Sheet S1 in Additional File 2), the recently described type strain of a novel species *X. bakwenae* [8]. Strain TH1 is not conspecific to ID10 nor any other type strain, making it a novel species within the *Xenorhabdus* genus. These conclusions were further supported by phylogenetic reconstructions which revealed that *Xenorhabdus* sp. TH1 does not cluster with *X. griffiniae* but shares a last common ancestor with the progenitor of the *X. griffiniae* clade (Fig. 1). From our findings on the evolution of gene content, we speculate that TH1 diverged, primarily through gene losses, from the progenitor of the *X. griffiniae* clade to ultimately form its species (Fig. 5), similar to the speciation observed in *Bordetella pertussis* [130].

Phage-related genes are known drivers of genome variation between closely related strains [131] and previously were implicated as drivers of the differences in gene content among other close *Xenorhabdus* relatives of *X. griffiniae* [16]. The *X. griffiniae* genomes we analysed were found to be enriched in mobilome content, with phage-related genes specifically driving the diversification of subspecies and strains. We found that the India-Indonesia and Kenyan subspecies arose from the net gain of genes, of which 39-47% were from prophages (Fig. 4&5). Further, we found that the proportion of *X. griffiniae*

strain-specific genes of prophage origin is 35-48%. This is especially high, considering that for example, of 30 *Bifidobacterium* strains examined, the highest proportion of strain-specific genes of prophage origin observed was only between 0.03-35.4% [132]. Moreover, *X. griffiniae* prophage regions reduced pairwise dDDH values among them by three percentage points. In *Salmonella enterica* prophage sequences have been shown to be highly variable and differentially conserved among strains, making them key drivers of genome diversification and useful markers for serovar typing [133, 134], Similarly, we conclude that in *X. griffiniae* prophages possibly are major drivers of subspeciation and strain differentiation.

An essential component of the *Xenorhabdus* lifestyle is interaction with other organisms, including competing microbes, the mutualistic nematode host, and the prey insect host. Among the molecular machines that facilitate bacterial manipulation of other organisms is the T6SS, which delivers effectors into target (non-self) cells [135]. In this study, we found a high sequence similarity of corresponding *X. griffiniae* T6SS-encoding loci, but gene content variability in loci encoding concomitant effector proteins. This indicates that effector proteins possibly contribute to traits that vary within a species. One of these traits is nematode host specificity, which varies within both *X. griffiniae* and *X. bovienii* species [7]. In *X. bovienii*, the inactivation of *vgrG* in XG2-T6SS loci in strain SS-2004 resulted in the near loss (200-fold decrease) of the bacterium's capacity to colonise its nematode host [136]. Although we found that the presence and sequence of VgrG does not appear to vary across *X. griffiniae* (or other *Xenorhabdus*), the genes encoding the effector proteins transported by VgrG do vary among the *X. griffiniae* species. Therefore, we

speculate that one or more *X. griffiniae* VgrG-associated effector proteins may determine a strain's capacity to naturally colonise its specific nematode host, similar to the conclusion reached about the T6SS function in *X. bovienii* SS-2004 [136].

CRISPR content and spacer identities support the conclusion that the strains studied here are diversifying due to phage pressure and reflect the taxonomic relationships we observed. The Kenyan subclade has nearly identical CIRSPR-Cas loci, consistent with their very close relationship. The only difference is that the XN45 CRISPR array 1a appears to lack a repeated spacer that the others have. Considering the Kenyan subclade as a single group, all the genomes were distinct from each other with respect to CRISPR spacer content, indicative of their unique histories in exposure to, and successful defence from phages and mobile genetic elements. Consistent with this, non-self-targeting protospacers identified within the group could be found within prophage genes and conjugation machinery in the genomes of the other strains. Our analysis identified instances of potential self-targeting, which offered the opportunity to search for anti-CRISPR genes, which are a key aspect of the co-evolution of phage and defence systems and have potential utility in applications of CRISPR technologies [100]. In the Kenyan subclade, we identified one clear candidate for such an anti-CRISPR locus comprising an HTH-domain Aca regulatory candidate and a small DUF2247 domain-containing protein ORF of 171 amino acids (aa), adjacent to the protospacer-containing gene *palA/fhaB,* which is also in the same region as a CRISPR array. However, DUF2247 proteins are also known as "imm38" and their presence within polymorphic toxin loci, such as *palA/fhaB,* has implicated them as immunity proteins to these toxins [137]. Consistent

1015 with this possibility, members of the Kenyan clade are the only strains of those analysed
1016 here that appear to have both a DUF2247 ORF and full-length *palA/fhaB* genes in that
1017 locus.
1018
1019 The *X. griffiniae* and *Xenorhabdus* sp. TH1 genomes we compared all bear hallmarks of
1020 the entomopathogenicity characteristic of the *Xenorhabdus* genus. However, when
1021 cultures were injected into *M. sexta* insect larva, the ID10 strain displayed attenuated
1022 virulence when compared with HGB2511 and TH1. The magnitude of this difference was
1023 directly comparable to the difference in virulence observed between *X. innexi* and *X.*
1024 *nematophila* species [71]. Members of the *X. bovienii* species group have demonstrated
1025 a similar range of virulence phenotypes when injected in the absence of the vectoring
1026 nematodes [138]. These differences may be due to genomic variation between the closely
1027 related species. Notably, we identified strain specific toxin loci, such as the *hcnABC* locus
1028 in TH1 and the zot domain containing toxins in the Kenyan clade that may underly
1029 different mechanisms or levels of virulence in *X. griffiniae* and its closest relative (Fig. 10;
1030 Table 3). It is also possible that the lack of virulence observed for ID10 is due to gene
1031 expression programs which control phenotypic variation locking the isolate in a state of
1032 attenuated virulence [139, 140]. If so, we predict that other isolates of the ID10 strain may
1033 retain high levels of virulence, similar to what has been observed for other *Xenorhabdus*
1034 species [139, 141, 142]. Alternatively, the ecological insect host range of ID10 may be
1035 distinctive enough from other *X. griffiniae* that it has lost the ability to infect the
1036 Lepidopteran insect *Manduca sexta* that we tested here, or the nematode may carry most
1037 of the virulence potential between the host-symbiont pair [143].

Conclusion

This study yielded three complete genome assemblies, which were of *X. griffiniae* ID10, *X. griffiniae* HGB2511 and *Xenorhabdus* sp. TH1. *Xenorhabdus* sp. TH1 is a novel bacterial species and putative type strain with the temporary designation *Candidatus Xenorhabdus lamphunensis,* while *X. griffiniae* contained two subspecies. Both CRISPR loci and loci encoding T6SS effector proteins divided along these *X. griffiniae* subspecies lines. Intraspecific variation, including subspeciation, was largely driven by prophages. In terms of biosynthetic potential, *X. griffiniae* genomes encoded the production pathways of diverse and biotechnologically useful natural products such as antibacterials, antiprotozoals, and insecticidal toxins. Intraspecific variation in biosynthetic potential was observed, which we substantiated by the different levels of entomopathogenicity, among *X. griffiniae* strains, to *M. sexta*. Ultimately, these genome assemblies and genomic insights are foundational for continuing studies into the symbiosis between *X. griffiniae* and its self-fertilizing nematode host, *S. hermaphroditum*.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated during the current study are available in the GenBank repository BioProject ID: PRJNA1085699. Previously reported data is available from

public repositories (see Methods), and additional data is provided within the manuscript or supplementary information files.

Competing interests

Not applicable

Funding

This work was supported by a National Science Foundation EDGE grant (IOS-2128266).

Authors' contributions

JKH, RMA, and HGB collected, and analysed the data, and wrote the original draft of the manuscript. MC and GC performed DNA extraction, sequencing, and assembly for HGB2511 and TH1 genomes. JKH performed all other “bench” experiments, extractions and assemblies, with help from JM on the insect virulence assay. HGB supervised and provided resources for the study. All authors reviewed the manuscript.

Acknowledgements

The authors would like to thank Dr.Apichat Vitta, Department of Microbiology and Parasitology, Faculty of Medical Science, Naresuan University, Thailand, for isolating the *Steinernema adamsii* nematodes, the host of the TH1 bacterial symbiont, *Candidatus Xenorhabdus lamphunensis*. Zachary Burcham for assistance running the PathoFact software used in the toxin analysis and Sarah Kauffman for assistance with insect injections.

References

1. Bhat AH, Chaubey AK, Askary TH. Global distribution of entomopathogenic nematodes, *Steinernema* and *Heterorhabditis*. Egyptian Journal of Biological Pest Control. 2020;30:31.

2. Edmunds C, Wilding CS, Rae R. Pathogenicity and environmental tolerance of commercial and UK native entomopathogenic nematodes (*Steinernema* and

Heterorhabditis spp.) to the larvae of mosquitoes (*Aedes aegypti* and *Ochlerotatus detritus*). International Journal of Pest Management. 2021;67:232–40.

3. Kovtun A. New locality records for *Steinernema* and *Heterorhabditis* (Nematoda: Rhabditida: *Steinernematidae*, *Heterorhabditidae*) fauna of Ukraine. Quarantine and plant protection. 2023;:39–45.

4. Ali M, Allouf N, Ahmad M. First report of entomopathogenic nematode *Steinernema affine* (Nematoda: *Steinernematidae*) in Syria and its virulence against *Galleria mellonella* L. (Lepidoptera: *Pyralidae*). Egyptian Journal of Biological Pest Control. 2022;32:101.

5. Kizi, NKS. First report on local entomopathogenic nematode *Steinernema feltiae* in Uzbekistan. International Journal of Advance Scientific Research. 2023;3:225–35.

6. Mackiewicz JP, Kramarz PE, Rożen A. Thermal sensitivity of *Xenorhabdus bovienii* (Enterobacterales: *Morganellaceae*) isolated from *Steinernema feltiae* (Rhabditida: *Steinernematidae*) originating from different habitats. Applied Entomology and Zoology. 2022;57:347–55.

7. Awori RM. Nematophilic bacteria associated with entomopathogenic nematodes and drug development of their biomolecules. Frontiers in Microbiology. 2022;13.

8. Ritter CL, Malan AP, Dicks LMT. *Xenorhabdus bakwenae* sp. n., associated with the entomopathogenic nematode *Steinernema bakwenae*. Nematology. 2023;25:1169–79.

9. Machado RAR, Bhat AH, Castaneda-Alvarez C, Askary TH, Půža V, Pagès S, et al. *Xenorhabdus aichiensis* sp. nov., *Xenorhabdus anantnagensis* sp. nov., and *Xenorhabdus yunnanensis* sp. nov., Isolated from *Steinernema* Entomopathogenic Nematodes. Curr Microbiol. 2023;80:300.

10. Mucci NC, Jones KA, Cao M, Wyatt MR 2nd, Foye S, Kauffman SJ, et al. Apex Predator Nematodes and Meso-Predator Bacteria Consume Their Basal Insect Prey through Discrete Stages of Chemical Transformations. mSystems. 2022;7:e0031222.

11. Murfin KE, Ginete DR, Bashey F, Goodrich-Blair H. Symbiont-mediated competition: *Xenorhabdus bovienii* confer an advantage to their nematode host *Steinernema affine* by killing competitor *Steinernema feltiae*. Environ Microbiol. 2018. https://doi.org/10.1111/1462-2920.14278.

12. Bashey F, Young SK, Hawlena H, Lively CM. Spiteful interactions between sympatric natural isolates of *Xenorhabdus bovienii* benefit kin and reduce virulence. J Evol Biol. 2012;25:431–7.

13. Gulcu B, Hazir S, Kaya HK. Scavenger deterrent factor (SDF) from symbiotic bacteria of entomopathogenic nematodes. Journal of Invertebrate Pathology. 2012;110:326–33.

14. Chaston JM, Murfin KE, Heath-Heckman EA, Goodrich-Blair H. Previously unrecognized stages of species-specific colonization in the mutualism between *Xenorhabdus* bacteria and *Steinernema* nematodes. Cellular Microbiology. 2013;15:1545–59.

15. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. International Journal of Systematic and Evolutionary Microbiology,. 2018;68:461–6.

16. Awori RM, Waturu CN, Pidot SJ, Amugune NO, Bode HB. Draft genomes, phylogenomic reconstruction and comparative genome analysis of three *Xenorhabdus* strains isolated from soil-dwelling nematodes in Kenya. Access Microbiology. 2023;5.

17. Castaneda-Alvarez C, Prodan S, Zamorano A, San-Blas E, Aballay E. *Xenorhabdus lircayensis* sp. nov., the symbiotic bacterium associated with the entomopathogenic nematode *Steinernema unicornum*. International Journal of Systematic and Evolutionary Microbiology,. 2021;71.

18. Bisch G, Ogier J-C, Médigue C, Rouy Z, Vincent S, Tailliez P, et al. Comparative Genomics between Two *Xenorhabdus bovienii* Strains Highlights Differential Evolutionary Scenarios within an Entomopathogenic Bacterial Species. Genome Biol Evol. 2016;8:148–60.

19. Murfin KE, Whooley AC, Klassen JL, Goodrich-Blair H. Comparison of *Xenorhabdus bovienii* bacterial strain genomes reveals diversity in symbiotic functions. BMC Genomics. 2015;16:889.

20. Shi Y-M, Hirschmann M, Shi Y-N, Ahmed S, Abebew D, Tobias NJ, et al. Global analysis of biosynthetic gene clusters reveals conserved and unique natural products in entomopathogenic nematode-symbiotic bacteria. Nature Chemistry. 2022;14:701–12.

21. Jin G, Hrithik MTH, Lee D-H, Kim I-H, Jung J-S, Bode HB, et al. Manipulation of GameXPeptide synthetase gene expression by a promoter exchange alters the virulence of an entomopathogenic bacterium, *Photorhabdus temperata*, by modulating insect immune responses. Front Microbiol. 2023;14:1271764.

22. Stock SP, Griffin CT, Chaerani R. Morphological and molecular characterisation of *Steinernema hermaphroditum* n. sp.(Nematoda: *Steinernematidae*), an entomopathogenic nematode from Indonesia, and its phylogenetic relationships with other members of the genus. Nematology. 2004;6:401–12.

23. Bhat AH, Chaubey AK, Shokoohi E, William Mashela P. Study of *Steinernema hermaphroditum* (Nematoda, Rhabditida), from the West Uttar Pradesh, India. Acta Parasitol. 2019;64:720–37.

24. Cao M, Schwartz HT, Tan C-H, Sternberg PW. The entomopathogenic nematode *Steinernema hermaphroditum* is a self-fertilizing hermaphrodite and a genetically tractable system for the study of parasitic and mutualistic symbiosis. Genetics. 2022.

25. Schwartz HT, Tan C-H, Peraza J, Raymundo KLT, Sternberg PW. Molecular identification of a peroxidase gene controlling body size in the entomopathogenic nematode *Steinernema hermaphroditum*. Genetics. 2024;226:iyad209.

26. Alani OS, Cao M, Goodrich-Blair H, Heppert JK. Conjugation and transposon mutagenesis of *Xenorhabdus griffiniae* HGB2511, the bacterial symbiont of the nematode *Steinernema hermaphroditum* (India). 2023. https://doi.org/10.17912/micropub.biology.000772.

27. Tailliez P, Pagès S, Ginibre N, Boemare N. New insight into diversity in the genus *Xenorhabdus*, including the description of ten novel species. International Journal of Systematic and Evolutionary Microbiology,. 2006;56:2805–18.

28. Sherathiya VN, Schaid MD, Seiler JL, Lopez GC, Lerner TN. GuPPy, a Python toolbox for the analysis of fiber photometry data. Sci Rep. 2021;11:24212.

29. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Computational Biology. 2017;13:e1005595.

30. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, et al. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. Nucleic Acids Research. 2022;51:D678–89.

31. Lander ES, Waterman MS. Genomic mapping by fingerprinting random clones: A mathematical analysis. Genomics. 1988;2:231–9.

32. Parrello B, Butler R, Chlenski P, Olson R, Overbeek J, Pusch GD, et al. A machine learning-based service for estimating quality of genomes using PATRIC. BMC Bioinformatics. 2019;20:486.

33. Grossman AS, Mauer TJ, Forest KT, Goodrich-Blair H. A Widespread Bacterial Secretion System with Diverse Substrates. mBio. 2021;12:10.1128/mbio.01956-21.

34. Médigue C, Calteau A, Cruveiller S, Gachet M, Gautreau G, Josso A, et al. MicroScope-an integrated resource for community expertise of gene functions and comparative analysis of microbial genomic and metabolic data. Brief Bioinform. 2019;20:1071–84.

35. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004;32:1792–7.

36. Vaidya G, Lohman DJ, Meier R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 2011;27:171–80.

37. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.

38. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35:4453–5.

39. Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol. 2012;61:1061–7.

40. Ayres DL, Cummings MP, Baele G, Darling AE, Lewis PO, Swofford DL, et al. BEAGLE 3: Improved Performance, Scaling, and Usability for a High-Performance Computing Library for Statistical Phylogenetics. Systematic Biology. 2019;68:1052–61.

41. M. A. Miller, W. Pfeiffer, T. Schwartz. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE). 2010. p. 1–8.

42. Rambaut A. Figtree. 2019.

43. Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nature Communications. 2019;10:2182.

44. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol. 2021.

45. Eddy SR. Accelerated Profile HMM Searches. PLOS Computational Biology. 2011;7:1–16.

46. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.

47. Galperin MY, Kristensen DM, Makarova KS, Wolf YI, Koonin EV. Microbial genome analysis: the COG approach. Briefings in Bioinformatics. 2019;20:1063–70.

48. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications. 2018;9:5114.

49. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome. 2020;8:90.

50. Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald P-J, Vreugde S. Pharokka: a fast scalable bacteriophage annotation tool. Bioinformatics. 2023;39.

51. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom. 2021;7.

52. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. Identification of mobile genetic elements with geNomad. Nat Biotechnol. 2023. https://doi.org/10.1038/s41587-023-01953-y.

53. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nature Biotechnology. 2021;39:578–85.

54. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010;5:e11147.

55. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9.

56. Zhang J, Guan J, Wang M, Li G, Djordjevic M, Tai C, et al. SecReT6 update: a comprehensive resource of bacterial Type VI Secretion Systems. Sci China Life Sci. 2023;66:626–34.

57. Csűös M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics. 2010;26:1910–2.

58. Tesson F, Hervé A, Mordret E, Touchon M, d'Humières C, Cury J, et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun. 2022;13:2561.

59. Néron B, Denise R, Coluzzi C, Touchon M, Rocha EPC, Abby SS. MacSyFinder v2: Improved modelling and search engine to identify molecular systems in genomes. Peer Community Journal. 2023;3.

60. Tesson F, Planel R, Egorov A, Georjon H, Vaysset H, Brancotte B, et al. A Comprehensive Resource for Exploring Antiphage Defense: DefenseFinder Webservice, Wiki and Databases. 2024;:2024.01.25.577194.

61. Veesenmeyer JL, Andersen AW, Lu X, Hussa EA, Murfin KE, Chaston JM, et al. NilD CRISPR RNA contributes to *Xenorhabdus nematophila* colonization of symbiotic host nematodes. Mol Microbiol. 2014;93:1026–42.

62. Biswas A, Staals RHJ, Morales SE, Fineran PC, Brown CM. CRISPRDetect: A flexible algorithm to define CRISPR arrays. BMC Genomics. 2016;17:356.

63. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018;46:W246–51.

64. Biswas A, Gagnon JN, Brouns SJJ, Fineran PC, Brown CM. CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. RNA Biol. 2013;10:817–27.

65. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. 2023;51:W46–50.

66. Terlouw BR, Blin K, Navarro-Muñoz JC, Avalon NE, Chevrette MG, Egbert S, et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. Nucleic Acids Research. 2022;51:D603–10.

67. Stachelhaus T, Mootz HD, Marahiel MA. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem Biol. 1999;6:493–505.

68. Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC Evol Biol. 2007;7:78.

69. van Santen JA, Poynton EF, Iskakova D, McMann E, Alsup TA, Clark TN, et al. The Natural Products Atlas 2.0: a database of microbially-derived natural products. Nucleic Acids Research. 2021;50:D1317–23.

70. Bah T. Inkscape: guide to a vector drawing program. Prentice Hall Press; 2011.

71. Kim I-H, Aryal SK, Aghai DT, Casanova-Torres ÁM, Hillman K, Kozuch MP, et al. The insect pathogenic bacterium *Xenorhabdus innexi* has attenuated virulence in multiple insect model hosts yet encodes a potent mosquitocidal toxin. BMC Genomics. 2017;18:927.

72. de Nies L, Lopes S, Busi SB, Galata V, Heintz-Buschart A, Laczny CC, et al. PathoFact: a pipeline for the prediction of virulence factors and antimicrobial resistance genes in metagenomic data. Microbiome. 2021;9:49.

73. Abby SS, Denise R, Rocha EPC. Identification of Protein Secretion Systems in Bacterial Genomes Using MacSyFinder Version 2. Methods Mol Biol. 2024;2715:1–25.

74. Jalili V, Afgan E, Gu Q, Clements D, Blankenberg D, Goecks J, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. Nucleic Acids Research. 2020;48:W395–402.

75. Papoulis SE, Wilhelm SW, Talmy D, Zinser ER. Nutrient Loading and Viral Memory Drive Accumulation of Restriction Modification Systems in Bloom-Forming Cyanobacteria. mBio. 2021;12:10.1128/mbio.00873-21.

76. Kobayashi I. Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution. Nucleic Acids Res. 2001;29:3742–56.

77. Teklemariam AD, Al-Hindi RR, Qadri I, Alharbi MG, Ramadan WS, Ayubu J, et al. The Battle between Bacteria and Bacteriophages: A Conundrum to Their Immune System. Antibiotics. 2023;12:381.

78. Hussa E, Goodrich-Blair H. Rearing and Injection of *Manduca sexta* Larvae to Assess Bacterial Virulence. JoVE. 2012;:e4295.

79. Rodak NY, Tan C-H, Sternberg PW. An improved solid medium-based culturing method for Steinernema hermaphroditum. MicroPubl Biol. 2024;2024.

80. Baniya A, Subkrasae C, Ardpairin J, Anesko K, Vitta A, Dillman AR. *Steinernema adamsi* n. sp. (Rhabditida: *Steinernematidae*), A new entomopathogenic nematode from Thailand. Journal of Parasitology. 2024;110:22–39.

81. Meier-Kolthoff JP, Klenk H-P, Göker M. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol. 2014;64 Pt 2:352–6.

82. Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, et al. Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. Standards in Genomic Sciences. 2014;9:2.

83. Awori RM, Hendre P, Amugune NO. The genome of a steinernematid-associated *Pseudomonas piscis* bacterium encodes the biosynthesis of insect toxins. Access Microbiology. 2023;5.

84. Chaston JM, Suen G, Tucker SL, Andersen AW, Bhasin A, Bode E, et al. The Entomopathogenic Bacterial Endosymbionts *Xenorhabdus* and *Photorhabdus*: Convergent Lifestyles from Divergent Genomes. PLOS ONE. 2011;6:1–13.

85. Kochanowsky RM, Bradshaw C, Forlastro I, Stock SP. *Xenorhabdus bovienii* strain jolietti uses a type 6 secretion system to kill closely related *Xenorhabdus* strains. FEMS Microbiol Ecol. 2020;96.

86. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proceedings of the National Academy of Sciences. 2005;102:2567–72.

87. Morales-Soto N, Forst SA. The xnp1 P2-like tail synthesis gene cluster encodes xenorhabdicin and is required for interspecies competition. J Bacteriol. 2011;193:3624–32.

88. Morales-Soto N, Gaudriault S, Ogier J-C, Thappeta KRV, Forst S. Comparative analysis of P2-type remnant prophage loci in *Xenorhabdus bovienii* and *Xenorhabdus nematophila* required for xenorhabdicin production. FEMS Microbiology Letters. 2012;333:69–76.

89. Carim S, Azadeh AL, Kazakov AE, Price MN, Walian PJ, Lui LM, et al. Systematic discovery of pseudomonad genetic factors involved in sensitivity to tailocins. ISME J. 2021;15:2289–305.

90. Jurėnas D, Fraikin N, Goormaghtigh F, Van Melderen L. Biology and evolution of bacterial toxin–antitoxin systems. Nature Reviews Microbiology. 2022;20:335–50.

91. Lee J-Y, Rahman FU, Kim E-K, Cho S-M, Kim H-R, Lee K, et al. Importin-11 is Essential for Normal Embryonic Development in Mice. Int J Med Sci. 2020;17:815–23.

92. Yu Z, Zhang W, Yang H, Chou S-H, Galperin MY, He J. Gas and light: triggers of c-di-GMP-mediated regulation. FEMS Microbiol Rev. 2023;47.

93. Georjon H, Bernheim A. The highly diverse antiphage defence systems of bacteria. Nat Rev Microbiol. 2023;21:686–700.

94. Engelberg-Kulka H, Hazan R, Amitai S. mazEF: a chromosomal toxin-antitoxin module that triggers programmed cell death in bacteria. J Cell Sci. 2005;118 Pt 19:4327–32.

95. Lopatina A, Tal N, Sorek R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. Annu Rev Virol. 2020;7:371–84.

96. Lemay M-L, Horvath P, Moineau S. The CRISPR-Cas app goes viral. Current Opinion in Microbiology. 2017;37:103–9.

97. Hampton HG, Watson BNJ, Fineran PC. The arms race between bacteria and their phage foes. Nature. 2020;577:327–36.

98. Thompson CP, Doak AN, Amirani N, Schroeder EA, Wright J, Kariyawasam S, et al. High-Resolution Identification of Multiple *Salmonella* Serovars in a Single Sample by Using CRISPR-SeroSeq. Appl Environ Microbiol. 2018;84.

99. Kushwaha SK, Narasimhan LP, Chithananthan C, Marathe SA. Clustered regularly interspaced short palindromic repeats-Cas system: diversity and regulation in Enterobacteriaceae. Future Microbiology. 2022;17:1249–67.

100. Makarova KS, Wolf YI, Koonin EV. In Silico Approaches for Prediction of Anti-CRISPR Proteins. J Mol Biol. 2023;435:168036.

101. Bickle TA, Krüger DH. Biology of DNA restriction. Microbiol Rev. 1993;57:434–50.

102. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, et al. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Research. 2003;31:1805–12.

103. Gavriilidou A, Kautsar SA, Zaburannyi N, Krug D, Müller R, Medema MH, et al. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. Nature Microbiology. 2022;7:726–35.

104. Li J-H, Cho W, Hamchand R, Oh J, Crawford JM. A Conserved Nonribosomal Peptide Synthetase in *Xenorhabdus bovienii* Produces Citrulline-Functionalized Lipopeptides. J Nat Prod. 2021;84:2692–9.

105. Crawford JM, Portmann C, Zhang X, Roeffaers MBJ, Clardy J. Small molecule perimeter defense in entomopathogenic bacteria. Proceedings of the National Academy of Sciences. 2012;109:10821–6.

106. Grammbitter GLC, Schmalhofer M, Karimi K, Shi Y-M, Schöner TA, Tobias NJ, et al. An Uncommon Type II PKS Catalyzes Biosynthesis of Aryl Polyene Pigments. J Am Chem Soc. 2019;141:16615–23.

107. Vior NM, Lacret R, Chandra G, Dorai-Raj S, Trick M, Truman AW. Discovery and Biosynthesis of the Antibiotic Bicyclomycin in Distantly Related Bacterial Classes. Appl Environ Microbiol. 2018;84.

108. Velasco A, Acebo P, Gomez A, Schleissner C, Rodríguez P, Aparicio T, et al. Molecular characterization of the safracin biosynthetic pathway from *Pseudomonas fluorescens* A2-2: designing new cytotoxic compounds. Mol Microbiol. 2005;56:144–54.

109. Nollmann FI, Dauth C, Mulley G, Kegler C, Kaiser M, Waterfield NR, et al. Insect-specific production of new GameXPeptides in *Photorhabdus luminescens* TTO1, widespread natural products in entomopathogenic bacteria. Chembiochem. 2015;16:205–8.

110. Fuchs SW, Proschak A, Jaskolla TW, Karas M, Bode HB. Structure elucidation and biosynthesis of lysine-rich cyclic peptides in *Xenorhabdus nematophila*. Org Biomol Chem. 2011;9:3130–2.

111. Dreyer J, Rautenbach M, Booysen E, van Staden AD, Deane SM, Dicks LMT. *Xenorhabdus khoisanae* SB10 produces Lys-rich PAX lipopeptides and a Xenocoumacin in its antimicrobial complex. BMC Microbiology. 2019;19:132.

112. Zhou Q, Grundmann F, Kaiser M, Schiell M, Gaudriault S, Batzer A, et al. Structure and Biosynthesis of Xenoamicins from Entomopathogenic Xenorhabdus. Chemistry – A European Journal. 2013;19:16772–9.

113. Imai Y, Meyer KJ, Iinishi A, Favre-Godal Q, Green R, Manuse S, et al. A new antibiotic selectively kills Gram-negative pathogens. Nature. 2019;576:459–64.

114. Daborn PJ, Waterfield N, Silva CP, Au CPY, Sharma S, ffrench-Constant RH. A single *Photorhabdus* gene, *makes caterpillars floppy* (*mcf*), allows *Escherichia coli* to persist within and kill insects. Proceedings of the National Academy of Sciences. 2002;99:10742–7.

115. Waterfield NR, Daborn PJ, Dowling AJ, Yang G, Hares M, ffrench-Constant RH. The insecticidal toxin makes caterpillars floppy 2 (Mcf2) shows similarity to HrmA, an avirulence protein from a plant pathogen. FEMS Microbiol Lett. 2003;229:265–70.

116. Wang H-C, Lin S-J, Wang H-C, Kumar R, Le PT, Leu J-H. A bacterial binary toxin system that kills both insects and aquatic crustaceans: *Photorhabdus* insect-related toxins A and B. PLOS Pathogens. 2023;19:e1011330.

117. Satchell KJF. Multifunctional-autoprocessing repeats-in-toxin (MARTX) Toxins of Vibrios. Microbiol Spectr. 2015;3.

118. Roderer D, Raunser S. Tc Toxin Complexes: Assembly, Membrane Permeation, and Protein Translocation. Annual Review of Microbiology. 2019;73:247–65.

119. Wang X, Cheng J, Shen J, Liu L, Li N, Gao N, et al. Characterization of *Photorhabdus* Virulence Cassette as a causative agent in the emerging pathogen *Photorhabdus asymbiotica*. Sci China Life Sci. 2022;65:618–30.

120. Jiang F, Li N, Wang X, Cheng J, Huang Y, Yang Y, et al. Cryo-EM Structure and Assembly of an Extracellular Contractile Injection System. Cell. 2019;177:370-383.e15.

121. Chen L, Song N, Liu B, Zhang N, Alikhan N-F, Zhou Z, et al. Genome-wide Identification and Characterization of a Superfamily of Bacterial Extracellular Contractile Injection Systems. Cell Rep. 2019;29:511-521.e2.

122. Jiang F, Shen J, Cheng J, Wang X, Yang J, Li N, et al. N-terminal signal peptides facilitate the engineering of PVC complex as a potent protein delivery system. Sci Adv. 2022;8:eabm2343.

123. Flury P, Vesga P, Péchy-Tarr M, Aellen N, Dennert F, Hofer N, et al. Antimicrobial and Insecticidal: Cyclic Lipopeptides and Hydrogen Cyanide Produced by Plant-Beneficial *Pseudomonas* Strains CHA0, CMR12a, and PCL1391 Contribute to Insect Killing. Front Microbiol. 2017;8:100.

124. Fasano A, Fiorentini C, Donelli G, Uzzau S, Kaper JB, Margaretten K, et al. Zonula occludens toxin modulates tight junctions through protein kinase C-dependent actin reorganization, in vitro. J Clin Invest. 1995;96:710–20.

125. Pierro MD, Lu R, Uzzau S, Wang W, Margaretten K, Pazzani C, et al. Zonula Occludens Toxin Structure-Function Analysis: Identification of the fragment biologically active on tight junctions and of the zonulin receptor binding domain. Journal of Biological Chemistry. 2001;276:19160–5.

126. Gao M, Dong S, Hu X, Zhang X, Liu Y, Zhong J, et al. Roles of Midgut Cadherin from Two Moths in Different *Bacillus thuringiensis* Action Mechanisms: Correlation among Toxin Binding, Cellular Toxicity, and Synergism. J Agric Food Chem. 2019;67:13237–46.

127. Cao M. CRISPR-Cas9 genome editing in *Steinernema* entomopathogenic nematodes. 2023;:2023.11.24.568619.

128. Hugenholtz P, Chuvochina M, Oren A, Parks DH, Soo RM. Prokaryotic taxonomy and nomenclature in the age of big sequence data. The ISME Journal. 2021;15:1879–92.

129. Mothupi B, Featherston J, Gray V. Draft Whole-Genome Sequence and Annotation of *Xenorhabdus griffiniae* Strain BMMCB Associated with the South African Entomopathogenic Nematode *Steinernema khoisanae* Strain BMMCB. Genome Announcements. 2015;3:e00785-15.

130. Linz B, Ivanov YV, Preston A, Brinkac L, Parkhill J, Kim M, et al. Acquisition and loss of virulence-associated factors during genome evolution and speciation in three clades of *Bordetella* species. BMC Genomics. 2016;17:767.

131. Kirchberger PC, Schmidt ML, Ochman H. The Ingenuity of Bacterial Genomes. Annu Rev Microbiol. 2020;74:815–34.

132. Lugli GA, Milani C, Turroni F, Tremblay D, Ferrario C, Mancabelli L, et al. Prophages of the genus Bifidobacterium as modulating agents of the infant gut microbiota. Environ Microbiol. 2016;18:2196–213.

133. Mottawea W, Duceppe M-O, Dupras AA, Usongo V, Jeukens J, Freschi L, et al. *Salmonella enterica* Prophage Sequence Profiles Reflect Genome Diversity and Can Be Used for High Discrimination Subtyping. Frontiers in Microbiology. 2018;9.

134. Yu J, Xu X, Wang Y, Zhai X, Pan Z, Jiao X, et al. Prophage-mediated genome differentiation of the *Salmonella* Derby ST71 population. Microb Genom. 2022;8.

135. Singh RP, Kumari K. Bacterial type VI secretion system (T6SS): an evolved molecular weapon with diverse functionality. Biotechnol Lett. 2023;45:309–31.

136. Pothula R, Lee M-W, Patricia Stock S. Type 6 secretion system components hcp and vgrG support mutualistic partnership between *Xenorhabdus bovienii* symbiont and *Steinernema jollieti* host. J Invertebr Pathol. 2023;198:107925.

137. Zhang D, de Souza RF, Anantharaman V, Iyer LM, Aravind L. Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. Biology Direct. 2012;7:18.

138. McMullen JG, McQuade R, Ogier J-C, Pagès S, Gaudriault S, Patricia Stock S. Variable virulence phenotype of *Xenorhabdus bovienii* (γ-Proteobacteria: *Enterobacteriaceae*) in the absence of their vector hosts. Microbiology (Reading). 2017;163:510–22.

139. Park Y, Herbert EE, Cowles CE, Cowles KN, Menard ML, Orchard SS, et al. Clonal variation in *Xenorhabdus nematophila* virulence and suppression of *Manduca sexta* immunity. Cellular Microbiology. 2007;9:645–56.

140. Sugar DR, Murfin KE, Chaston JM, Andersen AW, Richards GR, deLéon L, et al. Phenotypic variation and host interactions of *Xenorhabdus bovienii* SS-2004, the entomopathogenic symbiont of *Steinernema jollieti* nematodes. Environ Microbiol. 2012;14:924–39.

141. Hussa EA, Casanova-Torres ÁM, Goodrich-Blair H. The Global Transcription Factor Lrp Controls Virulence Modulation in *Xenorhabdus nematophila*. J Bacteriol. 2015;197:3015–25.

142. Cao M, Goodrich-Blair H. *Xenorhabdus nematophila* bacteria shift from mutualistic to virulent Lrp-dependent phenotypes within the receptacles of *Steinernema carpocapsae* insect-infective stage nematodes. Environmental Microbiology. 2020;22:5433–49.

143. Bisch G, Pagès S, McMullen JG, Stock SP, Duvic B, Givaudan A, et al. *Xenorhabdus bovienii* CS03, the bacterial symbiont of the entomopathogenic nematode *Steinernema weiseri,* is a non-virulent strain against lepidopteran insects. Journal of Invertebrate Pathology. 2015;124:15–22.

Supplementary information

Additional file 1: Word document containing supporting data for the detection and analysis of secretion systems, maximum likelihood phylogenies, prophages and

type six effector protein loci identification, MARTX locus alignments, and dose response assays for strains used in insect virulence assays

Table S1: Comparison of the number of putative secretion systems that are present *Xenorhabdus griffinae* strains. **Table S2.** Loci encoding putative restriction modification systems identified by homology and genome annotation searches. **Table S3:** Biosynthetic gene clusters (BGCs) in the *Xenorhabdus griffiniae* ID10 genome. **Figure S1:** Maximum likelihood phylogenetic tree created using RAxML with one-to-one orthologs from *Xenorhabdus* type strains, closely related *X. griffiniae* strains, and *Photorhabdus asymbiotica* as an outgroup. **Figure S2**: Loci of prophages and *xnp1* in the complete *Xenorhabdus* sp. TH1 genome. **Figure S3**: Dotplots of prophage loci that were considerably similar between strains. **Figure S4**: Schematic of subspecies-specific type six secretion system (T6SS) effector-encoding loci in six strains of *Xenorhabdus griffiniae.* **Figure S5**: MARTX protein multiple sequence alignments **Figure S6:** Pdp1 protein multiple sequence alignment **Figure S7:** Percentage survival of *Manduca sexta* post injection with multiple concentrations of *Xenorhabdus* bacteria strains tested.

Additional file 2. **Excel workbook containing raw data and results of all genome analyses.**

Data tables which collectively contain raw data and values from pangenome, dDDH, ANI, biosynthetic gene cluster, prophage analyses; defence systems; phyletic patterns; genome accession numbers and names of strains used.

Additional file 3. Word document with additional details of the detection of *Xenorhabdus griffiniae* defense systems and analysis of defense systems CRISPR loci

Table S4. Summary of defense systems in *X. griffiniae* and related strains. **Table S5.** Locus tags and coordinates of CRISPR-Cas features. **Table S6**. CRISPR spacers, protospacers, concomitant annotation and genome wherein they are located. Methods and results for *X. griffiniae* CRISPR repeats, protospacers and self-targeting immunity.

Additional file 4. Excel workbook of putative insect toxins encoded by *Xenorhabdus* strains

These sheets contain raw data of the toxin domain-containing proteins that are unique among the *X. griffiniae* and *Xenorhabdus* sp. TH1 species. Also contained herein, is raw data from the comparison between the combined toxin library outputs for the *X. griffiniae* and *Xenorhabdus* sp. TH1 species with the output for *X nematophila* 19061. Also provided are the raw “toxin library” summary outputs from the PathoFact software used in the analyses.