

# Comparative Genomics and Pathogenicity Analysis of Two Bacterial Symbionts of Entomopathogenic Nematodes: The Role of The GroEL Protein in Virulence

**Abraham Rivera-Ramírez**

Universidad Autónoma del Estado de Morelos

**Rosalba Salgado-Morales**

Universidad Autónoma del Estado de Morelos

**Alfredo Jiménez-Pérez**

Instituto Politécnico Nacional

**Rebeca Pérez-Martínez**

Universidad Autónoma del Estado de Morelos

**Blanca Inés García-Gómez**

Universidad Nacional Autónoma de México

**Edgar Dantán-González** (✉ [edantan@uaem.mx](mailto:edantan@uaem.mx))

Universidad Autónoma del Estado de Morelos

---

## Research Article

**Keywords:** Comparative genomics, pathogenicity, entomopathogenic, GroE

**Posted Date:** November 15th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1032559/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Bacteria of the genera *Xenorhabdus* and *Photorhabdus* are symbionts of entomopathogenic nematodes. Despite their close phylogenetic relationship, they show differences in their pathogenicity and virulence mechanisms in target insects. These differences can be explored by the analysis of the pangenome, as it provides a framework for characterizing and defining the gene repertoire. Here, we report the genome of strain SC 0516. In addition, we performed the first pangenome analysis of 91 strains of *Xenorhabdus* and *Photorhabdus*, obtaining a total of 23,603 gene clusters and a core genome of 348 genes. Phylogenetic analysis performed with the core genome showed that our strain belonged to the *X. nematophila* group. Biological tests showed that whole cells of *X. nematophila* SC 0516 were more virulent than those of *P. luminescens* HIM3 when both were injected into *Galleria mellonella* larvae. In addition, we cloned and expressed the GroEL proteins of both bacteria, as this protein has been previously indicated to show insecticidal activity in the genus *Xenorhabdus*. Cpn60-Xn was found to be the most toxic at all concentrations tested, with an LC<sub>50</sub> value of 102.34 ng/larva. Sequence analysis suggested that the Cpn60-Xn toxin was homologous to Cpn60-PI; however, Cpn60-Xn contained thirty-five differentially substituted amino acid residues that could be responsible for its insecticidal activity.

## Introduction

*Xenorhabdus* and *Photorhabdus* are closely related phylogenetic groups belonging to the family Enterobacteriaceae; they are Gram-negative gammaproteobacteria that have evolved to form symbiotic associations with soil entomopathogenic nematodes of the families *Steinernematidae* and *Heterorhabditidae*, respectively (1). These bacteria have a complex life cycle involving a mutualistic symbiotic stage, in which the bacteria become established and colonize the nematode gut, and a pathogenic stage, in which susceptible insects become infected with the nematode via its entry through natural openings such as respiratory spiracles, the mouth or the anus (2). Once the bacteria are inside the hemocoel, they actively replicate and release compounds that have the potential to suppress the host insect's immune response as a protective strategy to maintain symbiosis with the nematode. The bacterium enters the stationary phase of its growth cycle, resulting in the production of a wide range of compounds, including toxins, phospholipases, proteases and antibiotics. Degradative enzymes bioconvert the macromolecules of the insect, thus providing a supply of essential nutrients to the nematode in its reproduction and development cycle, while antibiotic substances inhibit the growth of other microorganisms that might compete for the carcass (3, 4). During the final stages of development, the nematode and bacteria reassociate, and there is a corresponding decrease in nutrients. These processes lead to the differentiation of the nematode from the juvenile stages to the infective juvenile (IJ) stage, in which it no longer feeds on the insect. Infective juveniles carrying bacteria in their intestinal tract emerge from insect carcasses in search of new prey (5–7).

Although the life cycle of this nematode-bacteria symbiotic complex shows a high degree of similarity, there are differences in the life cycle itself and in the mechanism of pathogenicity, which involves evasion of the insect immune system and the expression of virulence factors (8). Some studies have suggested

the existence of functionally divergent mechanisms of these functions (9–11). One approach to the study of these differences is comparative genomics, which allows the comparison of complete genomes and provides abundant genetic information to elucidate genomic structural landmarks, novel gene repertoires and phylogenetic relationships between different organisms. The fundamental goal of comparative genomics is to achieve pangenome analysis from genomes (12, 13). The pangenome is composed of a core genome, which includes all the genes present in all the strains studied and probably encodes functions related to the basic biology and phenotypes of the species, and an "accessory or dispensable genome," which is composed of genes present in some but not all strains studied as well as strain-specific genes. The dispensable genome is generally associated with nonessential functions, in addition to conferring selective advantages such as adaptability to the ecological niche, the ability to colonize new environments and antibiotic resistance (13–16). To date, there have been no pangenome analyses of the entomopathogenic bacteria of the genera *Xenorhabdus* and *Photorhabdus*.

In this work, we sequenced and reported the genome of *Xenorhabdus nematophila* SC 0516. In addition, we performed a pangenome analysis with the complete genome sequence data available from 91 different strains of the genera *Xenorhabdus* and *Photorhabdus* and determined the phylogenetic position of our strain using the core genome. We also characterized the virulence capacity of *X. nematophila* SC 0516 in a strain previously assessed in our laboratory, identified as *Photorhabdus luminescens* HIM3, and we further characterized a protein component with unique properties, GroEL, which is a chaperonin shared by both strains with differential insecticidal activity.

## Results And Discussion

### Genomic Features and Assembly Metrics of the SC 0516 strain

In the present study, bacterial strain SC 0516 was isolated from the hemolymph of infected larvae as well as from a macerate of infective juvenile *Steinernema carpocapsae* larvae. In addition, the bacterial genome of SC 0516 was sequenced and assembled de novo. The draft genome of strain SC 0516 was 4,179,879 bp in length, with 142-fold genome coverage and an approximately 43.5% G+C content. A total of 103 contigs were generated, with an N50 contig length of 73 255 bp and an L50 value of 16 contigs. No plasmids were detected in the analysis.

To functionally analyze the genome of strain SC 0516, the contigs were subjected to subsystem annotation on the RAST server. A total of 4174 protein-coding genes, including 51 tRNAs, 5 rRNAs and 5 noncoding RNAs, were assigned to 336 annotated subsystems, which can be defined as biological processes that are components of metabolism or structural complexes supported by a set of functional roles. Most of these genes were predicted to be involved in the metabolism of amino acids and derivatives (309 genes), proteins (189 genes) and carbohydrates (150 genes).

This whole-genome shotgun project has been deposited in GenBank under accession number JACDOS000000000.1. The version described in this article is the first version.

## Pangenome and phylogenetic analysis of *Xenorhabdus nematophila* SC 0516

A pangenome consists of core genes (i.e., genes found in all strains of a genus or species), flexible genes (found in more than one strain but not all strains), and single genes (found in only one strain, also known as accessory genes). We included 77 drafts and 14 complete genomes of *Xenorhabdus* and *Photorhabdus* bacteria in the analysis (Supplementary Data 1). As a selection criterion, genomes that were highly fragmented (> 300 contigs) were excluded. The consensus pangenomic matrix of *Xenorhabdus* and *Photorhabdus* from our dataset consisted of 23,603 gene clusters (Figure 1A). More than half of the complete set of genes constituting the pangenome (15,662 genes, 66%) were found to be uniquely present (referred to as a cloud genome), meaning that each strain contributed approximately 172 new genes on average to the pangenome. The shell genome (genomes  $\geq$  50%) and the soft core (genomes = 95%) consisted of 6,311 and 1,630 gene clusters, respectively (Figure 1A). This analysis showed that the pangenome of both genera has an open form (Figure 1B). This occurs when the number of new gene families continues to increase in a taxonomic lineage and this increase does not appear to be asymptotic, regardless of how many new genomes are added to the pangenome; higher rates of gene gains by horizontal gene transfer (HGT) are characteristic of these lineages (16, 17). The core genome was characterized as the set of genes present in all the genomes analyzed. We established that the core genome contained approximately 348 genes that were present in all 91 genomes studied, representing 1.4% of the pangenome (Figure 1C). However, the core genome presented a negative trend due to the addition of new strains, as the probability of gene sharing between strains decreased as new strains were incorporated into the study sample (Figure 1D). This was consistent with studies revealing a general negative relationship between pangenome size and the proportion of core genes, where larger "open" pangenomes have a lower proportion of core genes (18). Overall, the *Xenorhabdus-Photorhabdus* pangenome showed a high level of genome variability, with only 1% of its genome being constant. Thus, the remaining 66% of the pangenome is presented as a variable piece of DNA composed of a wide repertoire of genes and molecular functions.

To our knowledge, this was the first pangenomic analysis of the *Xenorhabdus* and *Photorhabdus* genera in which the core genomes were examined using BDBH, COG and OMCL strategies. Based on most of the genes that are part of the core genome grouping according to the NCBI Clusters of Orthologous Groups (COGs) database, a high percentage of the genes are involved in the generation and conversion of energy processes (29.35%), translation, ribosome structure and biogenesis (19.18%), amino acid transport and metabolism (12.40%) and replication, recombination and repair (11.61%), as shown in Figure 2.

The taxonomic position of strain SC 0516 was assessed using a maximum likelihood core genome phylogeny calculated from the 281 highest-scoring markers selected by the GET\_PHYLOMARKERS pipeline from the 348 consensus groups calculated by GET\_HOMOLOGUES. Three complete genomes of *Yersinia pestis* were included in the phylogenetic analysis. The core genome phylogeny showed that our

strain, SC 0516, was clustered with the species *X. nematophila*, as shown in Figure 3. Strain SC 0516 is closely related to *X. nematophila* YL001 (GenBank accession number CP032329.1), which was isolated from nematodes from a soil sample collected in the locality of Shanxi, China. The strength of the phylogenetic tree obtained using this core genome-based approach resolved clades with maximum support.

In addition, this analysis allowed us to locate genus-specific genes of *Xenorhabdus* and *Photorhabdus*, highlighting those genes involved in pathogenicity and virulence. Overall, we found that the genus *Xenorhabdus* did not possess unique elements related to pathogenicity and virulence. However, the distinctive genes present in the genus included genes associated with tellurium resistance and polyamine transport. On the other hand, the genus *Photorhabdus* presented 29 unique elements associated with virulence and pathogenicity related to the type III secretion system, pilus structures and fimbriae (**Supplementary Data 2**). This analysis was consistent with previous observations that the genus *Photorhabdus* possesses genes related to the type III secretion system that are absent in *Xenorhabdus*; these genes play an important role in host insect invasion as well as the secretion of toxins and various virulence factors (10, 11, 19, 20). These results reveal that, although the members of these genera are phylogenetically related bacteria with similar lifestyles, they may differ drastically in their molecular mechanisms of pathogenicity in the same host, and *Xenorhabdus* probably uses different effectors and secretion systems, which could be reflected in the differences in virulence capacity between the two genera.

### Evaluation of pathogenicity in *Galleria mellonella*

To evaluate the pathogenicity of our strain, *X. nematophila* SC 0516, survival experiments were performed in *G. mellonella* by injecting different doses of colony forming units (CFUs). For a comparative analysis of pathogenicity, we used a bacterial strain previously characterized in our working group, identified as *Photorhabdus luminescens* HIM3 (21). Bioassays showed that both bacteria were pathogenic to *G. mellonella* larvae, as demonstrated by the mortality rate visualized in Kaplan–Meier survival curves, which differed significantly from the control (*E. coli* DH5 $\alpha$ ) ( $\chi^2= 457.636$ ,  $df= 6$ , and  $p < 0.001$ ), as shown in Figure 4. This analysis revealed that *X. nematophila* SC 0516 was more virulent than *P. luminescens* HIM3 at all doses used (Figure 4). This difference was also illustrated by the median survival times (the time for which 50% of the larvae survive) estimated from the Kaplan–Meier survival analysis. In the case of *X. nematophila* SC 0516, the median times were 26.4 and 24 hours for doses of 102 and 103 CFUs, respectively, and 30.4 hours for a dose of 101 CFUs. However, *P. luminescens* HIM3 showed median times of 37.2 and 36 hours for doses of 102 and 103 CFUs, respectively, and 41 hours for a dose of 101 CFUs (Table 1). We also observed differences in the phenotypes of the dead larvae. In the larvae treated with *X. nematophila* SC 0516, a slight darkening of the body was observed after 48 hours of treatment, while those treated with *P. luminescens* HIM3 developed a dark reddish color at the same time. No external symptoms were observed in the control larvae (treated with *E. coli* DH5 $\alpha$ ) (**Supplementary Figure S1**).

Table 1

Survival time (h) of *Galleria mellonella* larvae injected with different doses of colony forming units (CFUs) of *X. nematophila* SC 0516 and *P. luminescens* HIM3.

	10 <sup>1</sup> CFU	10 <sup>2</sup> CFU	10 <sup>3</sup> CFU
<i>E.coli</i> DH5α	-	-	48 A
<i>P.luminescens</i> HIM3	41 ± 0.771 B	37.2 ± 0.469 B	36 B
<i>X. nematophila</i> SC 0516	30.4 ± 0.779 C	26.4 ± 0.625 C	24 C
Log-Rank test:457.636, df: 6, p <0.001.			
Survival time on the same column followed by the same letter are not significantly different (Holm-Sidak test, p>0.05)			

To explain these differences, we looked for unique virulence- and pathogenicity-related elements present in the genomes of both bacteria. Our analysis showed that *P. luminescens* HIM3 presented a higher number and diversity of pathogenicity- and virulence-associated components than *X. nematophila* SC 0516 in almost all categories (Figure 5). The only unique elements presented by *X. nematophila* SC0516 were the mcf (makes caterpillars floppy) toxin (reported in the literature as a highly insecticidal toxin) and a higher number of genes associated with the type IV secretion system (T4SS) and the type I toxin-antitoxin system (Figure 5). An *et al.*, 2009 compared the gene expression of *Photorhabdus temperata* and *Xenorhabdus koppenhoeferi* *in vivo* in the insect *Rhizotrogus majalis* and found that more than 60% of the genes were uniquely induced in one of the two bacteria (22). However, in *Xenorhabdus* bacteria, the mechanisms of toxin delivery to the insect are not completely known, although some mechanisms of secretion through the flagellar apparatus or by vesicular systems of the outer membrane have been suggested (23, 24). The pathogenicity of some Gram-negative bacteria depends on their ability to secrete virulence factors into the mammalian host through the release of outer membrane vesicles (OMVs). Some virulence factors of OMVs include proteases, hemolysins, phospholipids and lipopolysaccharides (25, 26). The analysis of OMV proteins led to the identification of a 58 kDa GroEL homolog as a major component of a complex that has been characterized as a virulence factor with insecticidal activity in some *Xenorhabdus* species, including *X. nematophila* (27, 28). These factors, termed moonlighting proteins, escaped our pangenome analysis, as was the case for the GroEL chaperonin, and could explain the differences in pathogenicity and virulence found between the two bacteria.

## Evaluation of the insecticidal activity of GroEL proteins

In this case, we decided to clone and express the GroEL proteins from both bacteria, and we named them Cpn60-Xn and Cpn60-PI (**Supplementary Figure S2**). It should be noted that genomic data obtained from these bacteria show that they possess only a single copy of the groEL gene. The biological activity of the two purified proteins was evaluated by a direct hemolymph injection method. This method introduced the

protein directly into the insect, mimicking the release of toxins by bacteria, a phenomenon that occurs shortly after a nematode infects a target insect.

The bioassay results indicated that only the Cpn60-Xn protein killed a high percentage of *G. mellonella* larvae and that the mortality rate depended on the concentration of protein used. The relationship between larval death and the protein concentration was assessed by semilogarithmic linear regression analysis, and the slope values (m) representing toxicity were 42.73 and 16.87 for Cpn60-Xn and Cpn60-PI, respectively; thus, the difference between the slopes was highly significant ( $F= 31.50$ ,  $DFn= 1$ ,  $DFd= 26$ ,  $p<0.001$ ). (Figure 6). The 50% lethal concentration ( $LC_{50}$ ) of the purified Cpn60-Xn protein was found to be 102.34 ng/larvae. No external symptoms or mortality were observed in control larvae infiltrated with PBS or BSA (2000 ng), or simply punctured. GroEL toxicity has been reported in *Enterobacter aerogenes* (EnGroEL) and some *Xenorhabdus* species. EnGroEL was shown to be a paralytic toxin that ultimately killed cockroaches of the genus *Blattella* when injected into the hemolymph at a minimum dose of  $2.7\pm 1.6$  ng (29). Later, a GroEL protein from *X. nematophila* was purified and found to show insecticidal activity against larvae of the cotton bollworm, *Helicoverpa armigera*, after oral administration but showed no effect when injected into the hemolymph of this insect (30). Two other proteins from *X. budapestensis* (HIP57) and *X. ehlersii* (XeGroEL) were shown to be toxic in *G. mellonella* larvae when injected into the hemolymph, with  $LC_{50}$  values of 206.81 and  $0.76 \pm 0.08$  ng/larva, respectively (31, 32). In the genus *Photorhabdus*, there is an initial report that GroEL is a protein component of an 860 kDa complex found in the toxic fraction of the *P. luminescens* W-14 extract (33); however, no direct evidence supporting this finding has been reported. Our work provides the first evaluation of the activity of this protein.

A global alignment of these two protein sequences showed that they were very similar, sharing 89.86% identity of 548 residues. Cpn60-Xn showed 35 substitutions relative to Cpn60-PI, which were distributed throughout the protein. A three-dimensional homology-generated model of Cpn60-Xn showed that 13 of these substitutions were located in the apical domain, 2 in the intermediate domain and 20 in the equatorial domain (**Supplementary Figure S2**). These point substitutions could be responsible for conferring the greater toxicity of Cpn60-Xn relative to Cpn60-PI.

## Conclusion

In this study, we sequenced and reported the genome of *X. nematophila* strain SC0516, which allowed us to perform the first robust pangenomic analysis of 91 genomes of *Xenorhabdus* and *Photorhabdus*, using three different algorithms. The results revealed that the differences between these two genera mainly involved genes belonging to the type III secretion systems and related to the fimbria and pilus, which are absent in the genus *Xenorhabdus*. Despite this, data from biological experiments related to pathogenicity and virulence showed that both the *X. nematophila* SC0516 and *P. luminescens* HIM3 strains were pathogenic; however, the former was more virulent at all concentrations tested, indicating differences in the mechanism of pathogenicity. In addition, the GroEL protein (reported as a virulence factor in the genus *Xenorhabdus*) from *X. nematophila* SC0516 (Cpn60-Xn) showed significantly more insecticidal activity than that from *P. luminescens* HIM3 (Cpn60-PI) at all concentrations used in *Galleria*

*mellonella* larvae when injected. The analysis of the GroEL sequences of both bacteria showed that this protein was highly conserved in *X. nematophila* SC 0516 and *P. luminescens* HIM3, although the former presented 13 specific differences that could contribute to explaining the difference in insecticidal activity.

## Methods

**Maintenance of insects and nematodes** *Galleria mellonella* was maintained at a temperature of  $25 \pm 2^\circ\text{C}$  with a light/dark (LD) photoperiod of 12:12 h and relative humidity (RH) of  $70 \pm 10\%$  and was cultured on an artificial diet. The nematodes were reared on last-instar *G. mellonella* larvae, and infective juveniles (IJs) were recovered from white traps at 12-14 days after inoculation with IJs.

**Bacterial growth** Bacterial strains were grown on NBTA medium (nutrient agar supplemented with 0.025 g bromothymol blue and 0.04 g 2,3,5-triphenyltetrazolium chloride per liter). For experimental purposes, a 48-h colony was transferred to 50 mL nutrient broth medium (NB) in a 250 mL Erlenmeyer flask and incubated at  $25 \pm 1^\circ\text{C}$  at 120 rpm for 24 h.

**Genome collection** A total of 91 sequences were collected from the annotated genomes of *Xenorhabdus* and *Photorhabdus* available in the open access RefSeq: NCBI Reference Sequence Database, among which 14 were complete, and 77 were drafts in various stages of completion. As a selection criterion, genomes that were highly fragmented ( $> 300$  contigs) were excluded. Three complete *Yersinia pestis* genomes were included in the phylogenetic analysis.

**Genome sequencing and annotation** Genomic DNA from *X. nematophila* SC 0526 was extracted using the ZR Fungal/Bacterial MiniPrep™ Kit (Zymo Research, Irvine, CA, USA), following the manufacturer's instructions. Subsequently, 5  $\mu\text{g}$  of genomic DNA was sequenced on the Illumina MiSeq platform. Read quality was analyzed with FastQC (33). Illumina adapter sequences were removed using the ILLUMINACLIP trimming step of Trimmomatic v0.39 software (34). Low-quality bases were removed from Illumina paired-end reads using the DynamicTrim algorithm of the SolexaQA++ v3.1.7.1 software package with a Phred quality score of  $Q=13$  (35). Paired-end reads were assembled de novo using the SPAdes v3.14.1 program with the following options: i) only run the assembly module (`-only-assembler`); ii) reduce the number of mismatches (`-careful`); and iii) k-mer lengths between 21 and 71, which generated 248 contigs (36). Contigs of less than 500 bp in length were discarded, and the remaining contigs were used for a multidraft-based analysis using the genome of the *X. nematophila* YL001 strain via the MeDuSa v1.6 scaffolder (37). A final assembly polishing step was performed by reassigning the filtered high-quality sequence reads to the ordered scaffolds using BWA and passing the resulting ordered binary alignments to SAMtools for indexing (38). The indexed alignments were used for analysis by Pilon v1.23 (39). The draft bacterial genome was automatically annotated using the RAST server version 2.0 (40) (available at <https://rast.nmpdr.org/rast.cgi>), and 16S rRNA gene sequences were obtained using the RNAmmer server (41) (<http://www.cbs.dtu.dk/services/RNAmmer/>).

**Phylogenomic analyses** The core genome phylogeny was estimated under the maximum likelihood (ML) criterion using the GET\_HOMOLOGUES (42) and GET\_PHYLOMARKERS (43) software suites. The

get\_homologues.pl program was used in combination with compare\_clusters.pl to compute a consensus core genome resulting from clustering BLASTP results (with 90% query coverage) with the BDBH (Bi-directional Best Hits), COG triangles and OrthoMCL (Markov Clustering of orthologs) algorithms implemented in GET\_HOMOLOGUES (44). Domain search was allowed in PFAM for the last runs. A consensus pangenome was similarly computed from COG triangles and OMCL clusters. The pangenome consensus clusters were input into the GET\_PHYLOMARKERS pipeline to select alignments with optimal phylogenetic attributes (no significant evidence of recombination, producing tree topologies and branch lengths that did not deviate significantly from the expected distribution of these parameters and showing average branch support values > 0). 6). The alignments that passed these filters were concatenated, and an ML phylogeny was estimated with IQ-TREE 1.6.1 (45) using the best-fit model and choosing the phylogeny with the highest probability score of those found among independent searches. Finally, the phylogeny was rooted using the *Y. pestis* genomes. Phylogenetic trees were visualized and edited with FigTree v1.4.3 (46).

**Pathogenicity assays in *Galleria mellonella*** Pathogenicity bioassays in *G. mellonella* were performed by the direct injection of  $10^1$ ,  $10^2$  or  $10^3$  CFUs of *X. nematophila* SC 0516 into the hemolymph of fifth-instar larvae of *G. mellonella*. Larvae were 6/13 selected by weight (250 to 350 mg) in all trials. A final volume of 10  $\mu$ l of LB containing the bacterial suspension was injected into the hemolymph of individual insects in the last left pro-leg with a 31-gauge insulin needle (BD Medical-Diabetes Care, Holdrege, NE, USA). Equivalent doses of *P. luminescens* HIM3 and *E. coli* DH5 $\alpha$  (negative control) were used in all assays. Fifteen larvae were used per dose, and each larva was placed in a 9 cm Petri dish. Experiments were performed four times in independent bioassays (n=60), and survival was assessed at 12, 24, 36 and 48 hours post infection.

**Cloning, expression and purification of GroEL proteins** Based on the genomic information obtained from the bacteria, specific primers were designed for the PCR amplification of the groEL sequences of *X. nematophila* SC 0516 (Fw:CATATGGCAGCTAAAGACGTAAAATTTG and Rv:GAATTCACATCATGCCGCCATTCCAC) and *P. luminescens* HIM3 (Fw:CATATGGCAGCTAAAGACGTAAAATTTGG and Rv:GTCGACTTACATCATGCCGCCATACCG). The 1.7 kb coding sequences were cloned into the pJET1.2 blunt vector (Thermo Scientific), and the resultant constructs were transformed into *E. coli* DH5 $\alpha$  cells following the protocol described by the manufacturer. Subsequently, the 1.7 Kb fragments of both vectors were ligated into the expression vector pET28a, and these constructs were transformed into *E. coli* BL21DE3, thus producing the recombinant protein-producing strains XnGroEL and PIGroEL. These cells were grown at 37°C and 150 rpm in LB medium containing 50  $\mu$ g/ml kanamycin until an O.D. of 0.5 was reached, when gene expression was immediately induced by the addition of 1 mM IPTG for 4-5 hours. Cells were washed and lysed by sonication, and the cell debris-free supernatant was filtered through a cold agarose-acetic acid-nickel agarose affinity column (Ni-NTA Agarose QIAGEN) following the manufacturer's recommended protocol.

**Evaluation of the insecticidal activity of GroEL proteins** The insecticidal activity of the GroEL proteins from *X. nematophila* SC 0526 (named Cpn60-Xn) and *P. luminescens* HIM3 (named Cpn60-PI) was

evaluated in fifth-instar *G. mellonella* larvae using doses of 100, 200, 500, 1000 and 2000 ng of protein per larva. Injection was performed directly into the insect hemolymph with 0.3 ml 31G X 6 mm Ultra Fine U-100 insulin syringes (BD Medical-Diabetes Care, Holdrege, NE, USA). The final injection volume was 10 µl in all treatments, and BSA (2000 ng) was used as a negative control. The assays were repeated three times using 15 larvae per treatment. The 50% lethal dose of the purified proteins was assessed using semilogarithmic regression, applying the logarithm base 10 of the protein concentrations and using the following formula:  $y = 42.73 \cdot x - 35.79$ .

**Sequence and Structure Analysis** The global alignment of the GroEL sequence of *X. nematophila* SC 0516 with that of *P. luminescens* HIM3 was performed with the program Clustal Omega (47). I-TASSER (48) was used to generate the three-dimensional structure, which was visualized and analyzed with VMD (49, 50).

**Statistical analysis** Survival analysis, which is generally a set of methods for analyzing data in which the outcome variable is time to the occurrence of an event of interest (mortality), was performed. Data from the experiments (n=60) were plotted using the Kaplan–Meier estimator, a nonparametric statistic, and a log-rank test was performed to detect significant differences between treatments. To isolate one or more groups that differed from the others, a multiple comparison procedure was carried out using the Holm-Sidak method. In both cases,  $p > 0.005$  was the rejection probability.

In the case of semilogarithmic linear regressions, ANCOVA was performed to detect significant differences between the two lines, with a probability of rejection of  $p > 0.005$ .

Statistical analyses were performed with GraphPad Prism version 8.4.3 (GraphPad Software Inc., San Diego, CA, USA).

## References

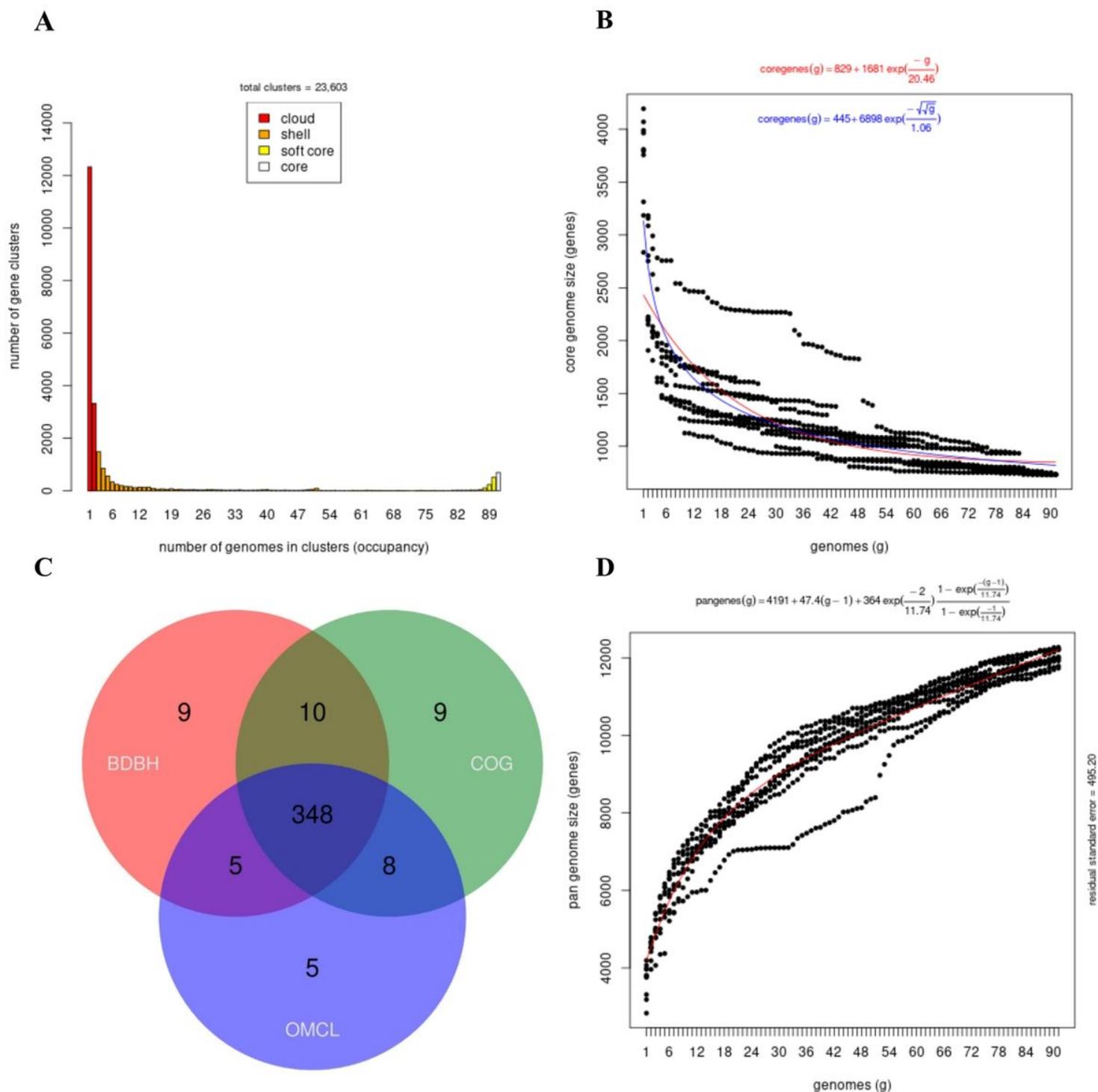
1. Akhurst, R. & Dunphy, G. Tripartite interactions between symbiotically associated entomopathogenic bacteria, nematodes, and their insect hosts. (1993).
2. Akhurst, R. Bacterial symbionts of entomopathogenic nematodes—the power behind the throne. *Nematodes biological control insect pests* 127–136 (1993).
3. Forst, S., Dowds, B., Boemare, N. & Stackebrandt, E. *Xenorhabdus* and *Photorhabdus* spp.: bugs that kill bugs. *Annu. review microbiology* 51, 47–72 (1997).
4. Owuama, C. Entomopathogenic symbiotic bacteria, *Xenorhabdus* and *Photorhabdus* of nematodes. *World J. Microbiol. Biotechnol.* 17, 505–515 (2001).
5. Adams, B. J. et al. Reprint of “biodiversity and systematics of nematode–bacterium entomopathogens” [biol. control 37 (2006) 32–49]. *Biol. control* 38, 4–21 (2006).
6. Gulcu, B., Cimen, H., Raja, R. K. & Hazir, S. Entomopathogenic nematodes and their mutualistic bacteria: their ecology and application as microbial control agents. *Biopestic* 13, 79–112 (2017).

7. Sajnaga, E. & Kazimierczak, W. Evolution and taxonomy of nematode-associated entomopathogenic bacteria of the genera *Xenorhabdus* and *Photorhabdus*: An overview. *Symbiosis* 80, 1–13 (2020).
8. Hinchliffe, S. J., Hares, M. C., Dowling, A. J. et al. Insecticidal toxins from the *Photorhabdus* and *Xenorhabdus* bacteria. *The Open Toxinology J.* 3 (2010).
9. Poinar, G. Origins and phylogenetic relationships of the entomophilic rhabditids, heterorhabditis and steinernema. *Fundamental Appl. Nematol.* 16, 333–338 (1993).
10. Goodrich-Blair, H. & Clarke, D. J. Mutualism and pathogenesis in *Xenorhabdus* and *Photorhabdus*: two roads to the same destination. *Mol. microbiology* 64, 260–268 (2007).
11. Chaston, J. M. et al. The entomopathogenic bacterial endosymbionts *Xenorhabdus* and *Photorhabdus*: convergent lifestyles from divergent genomes. *PloS one* 6, e27909 (2011).
12. Lu, Q.-F. et al. Genus-wide comparative genomics analysis of *Neisseria* to identify new genes associated with pathogenicity and niche adaptation of *neisseria* pathogens. *Int. journal genomics* (2019).
13. Kim, Y., Gu, C., Kim, H. U. & Lee, S. Y. Current status of pan-genome analysis for pathogenic bacteria. *Curr. opinion biotechnology* 63, 54–62 (2020).
14. Tettelin, H. et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci.* 102, 13950–13955 (2005).
15. Tettelin, H., Riley, D., Cattuto, C. & Medini, D. Comparative genomics: the bacterial pan-genome. *Curr. opinion microbiology* 11, 472–477 (2008).
16. McInerney, J. O., McNally, A. & O’connell, M. J. Why prokaryotes have pangenomes. *Nat. microbiology* 2, 1–5 (2017).
17. Treangen, T. J. & Rocha, E. P. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS genetics* 7, e1001284 (2011).
18. Brockhurst, M. A. et al. The ecology and evolution of pangenomes. *Curr. Biol.* 29, R1094–R1103 (2019).
19. Brugirard-Ricaud, K. et al. Variation in the effectors of the type iii secretion system among *Photorhabdus* species as revealed by genomic analysis. *J. bacteriology* 186, 4376–4381 (2004).
20. Brugirard-Ricaud, K. et al. Site-specific antiphagocytic function of the *Photorhabdus luminescens* type iii secretion system during insect colonization. *Cell. microbiology* 7, 363–371 (2005).
21. Salgado-Morales, R. et al. Assessing the pathogenicity of two bacteria isolated from the entomopathogenic nematode *Heterorhabditis indica* against *Galleria mellonella* and some pest insects. *Insects* 10, 83 (2019).
22. An, R., Sreevatsan, S. & Grewal, P. S. Comparative in vivo gene expression of the closely related bacteria *Photorhabdus temperata* and *Xenorhabdus koppenhoeferi* upon infection of the same insect host, *Rhizotrogus majalis*. *BMC genomics* 10, 1–16 (2009).
23. Cowles, K. N. & Goodrich-Blair, H. Expression and activity of a *Xenorhabdus nematophila* haemolysin required for full virulence towards *Manduca sexta* insects. *Cell. microbiology* 7, 209–219 (2005).

24. Darsouei, R., Karimi, J. & Dunphy, G. B. Functional characterization of outer membrane proteins (omps) in *Xenorhabdus nematophila* and *Photorhabdus luminescens* through insect immune defense reactions. *Insects* 10, 352 (2019).
25. Schwechheimer, C. & Kuehn, M. J. Outer-membrane vesicles from gram-negative bacteria: biogenesis and functions. *Nat. reviews microbiology* 13, 605–619 (2015).
26. Avila-Calderón, E. D. et al. Outer membrane vesicles of gram-negative bacteria: An outlook on biogenesis. *Front. Microbiol.* 12, 345 (2021).
27. Khandelwal, P. & Banerjee-Bhatnagar, N. Insecticidal activity associated with the outer membrane vesicles of *Xenorhabdus nematophilus*. *Appl. Environ. Microbiol.* 69, 2032–2037 (2003).
28. Khandelwal, P., Bhatnagar, R., Choudhury, D. & Banerjee, N. Characterization of a cytotoxic pilin subunit of *Xenorhabdus nematophila*. *Biochem. biophysical research communications* 314, 943–949 (2004).
29. Yoshida, N. et al. Chaperonin turned insect toxin. *Nature* 411, 44–44 (2001).
30. Joshi, M. C. et al. An insecticidal groel protein with chitin binding activity from *Xenorhabdus nematophila*. *J. Biol. Chem.* 283, 28287–28296 (2008).
31. Yang, J. et al. An insecticidal protein from *Xenorhabdus budapestensis* that results in prophenoloxidase activation in the wax moth, *Galleria mellonella*. *J. invertebrate pathology* 110, 60–67 (2012).
32. Shi, H. et al. An insecticidal protein from *xenorhabdus ehlersii* triggers prophenoloxidase activation and hemocyte decrease in *Galleria mellonella*. *Curr. microbiology* 64, 604–610 (2012).
33. Guo, L. et al. *Photorhabdus luminescens* w-14 insecticidal activity consists of at least two similar but distinct proteins: purification and characterization of toxin a and toxin b. *J. Biol. Chem.* 274, 9836–9842 (1999).
34. Wingett, S. W. & Andrews, S. Fastq screen: A tool for multi-genome mapping and quality control. *F1000Research* 7 (2018).
35. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120 (2014).
36. Cox, M. P., Peterson, D. A. & Biggs, P. J. Solexaqa: At-a-glance quality assessment of illumina second-generation sequencing data. *BMC bioinformatics* 11, 1–6 (2010).
37. Bankevich, A. et al. Spades: a new genome assembly algorithm and its applications to single-cell sequencing. *J. computational biology* 19, 455–477 (2012).
38. Bosi, E. et al. Medusa: a multi-draft based scaffold. *Bioinformatics* 31, 2443–2451 (2015).
39. Li, H. & Durbin, R. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* 25, 1754–1760 (2009).
40. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one* 9, e112963 (2014).

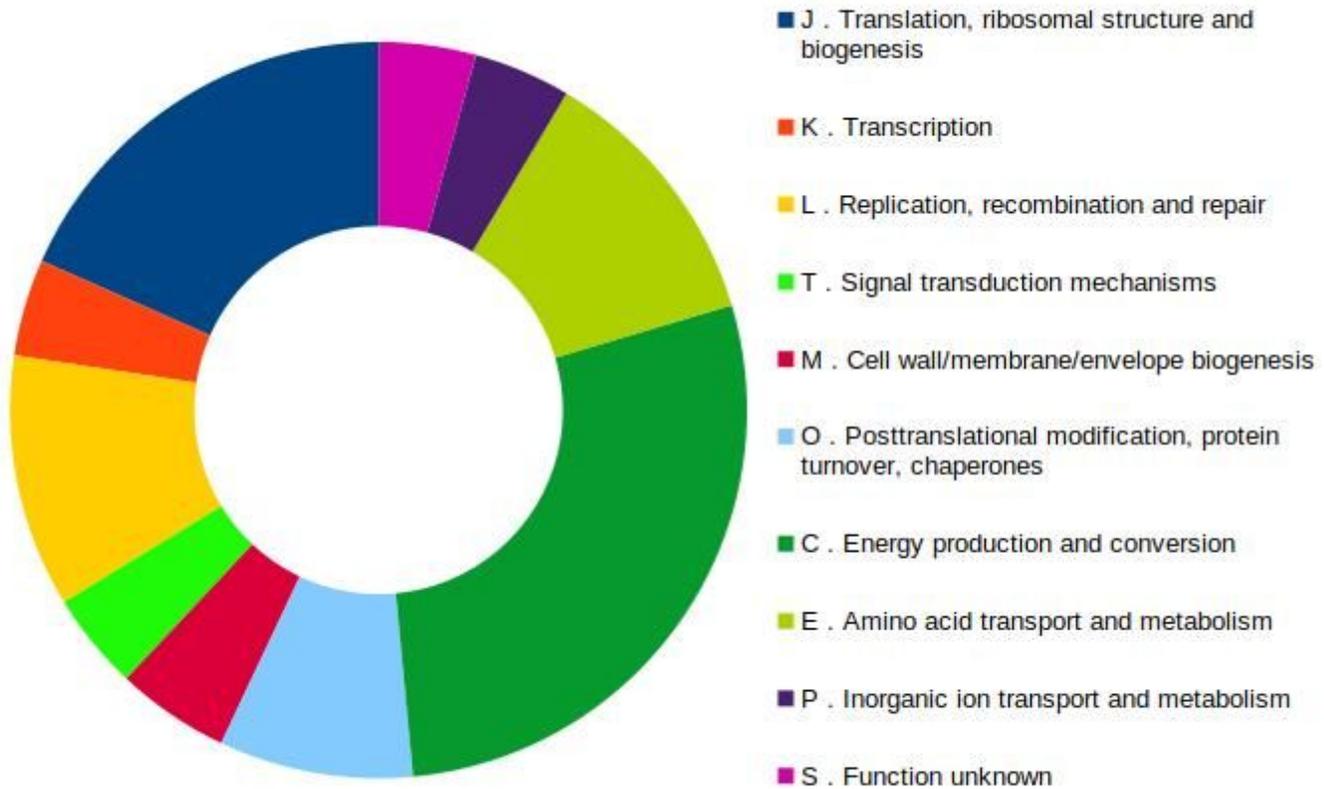
41. Aziz, R. K. *et al.* The rast server: rapid annotations using subsystems technology. *BMC genomics* 9, 1–15 (2008).
42. Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal rna genes. *Nucleic acids research* 35, 3100–3108 (2007).
43. Contreras-Moreira, B. & Vinuesa, P. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. environmental microbiology* 79, 7696–7701 (2013).
44. Vinuesa, P., Ochoa-Sánchez, L. E. & Contreras-Moreira, B. GET\_PHYLOMARKERS a software package to select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies, used for a critical geno-taxonomic revision of the genus *Stenotrophomonas*. *Front. microbiology* 9, 771 (2018).
45. Vinuesa, P. & Contreras-Moreira, B. Robust identification of orthologues and paralogues for microbial pan-genomics using get\_homologues: a case study of pinca/c plasmids. *In Bacterial Pangenomics*, 203–232 (Springer, 2015).
46. Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. biology evolution* 32, 268–274 (2015).
47. Rambaut, A. Figtree v1. 4.2, a graphical viewer of phylogenetic trees. Available <http://tree.bio.ed.ac.uk/software/figtree/> (2014).
48. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools apis in 2019. *Nucleic acids research* 47, W636–W641 (2019).
49. Yang, J. & Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic acids research* 43, W174–W181 (2015).
50. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. molecular graphics* 14, 33–38 (1996).
51. Roberts, E., Eargle, J., Wright, D. & Luthey-Schulten, Z. Multiseq: unifying sequence and structure data for evolutionary analysis. *BMC bioinformatics* 7, 1–11 (2006).

## Figures



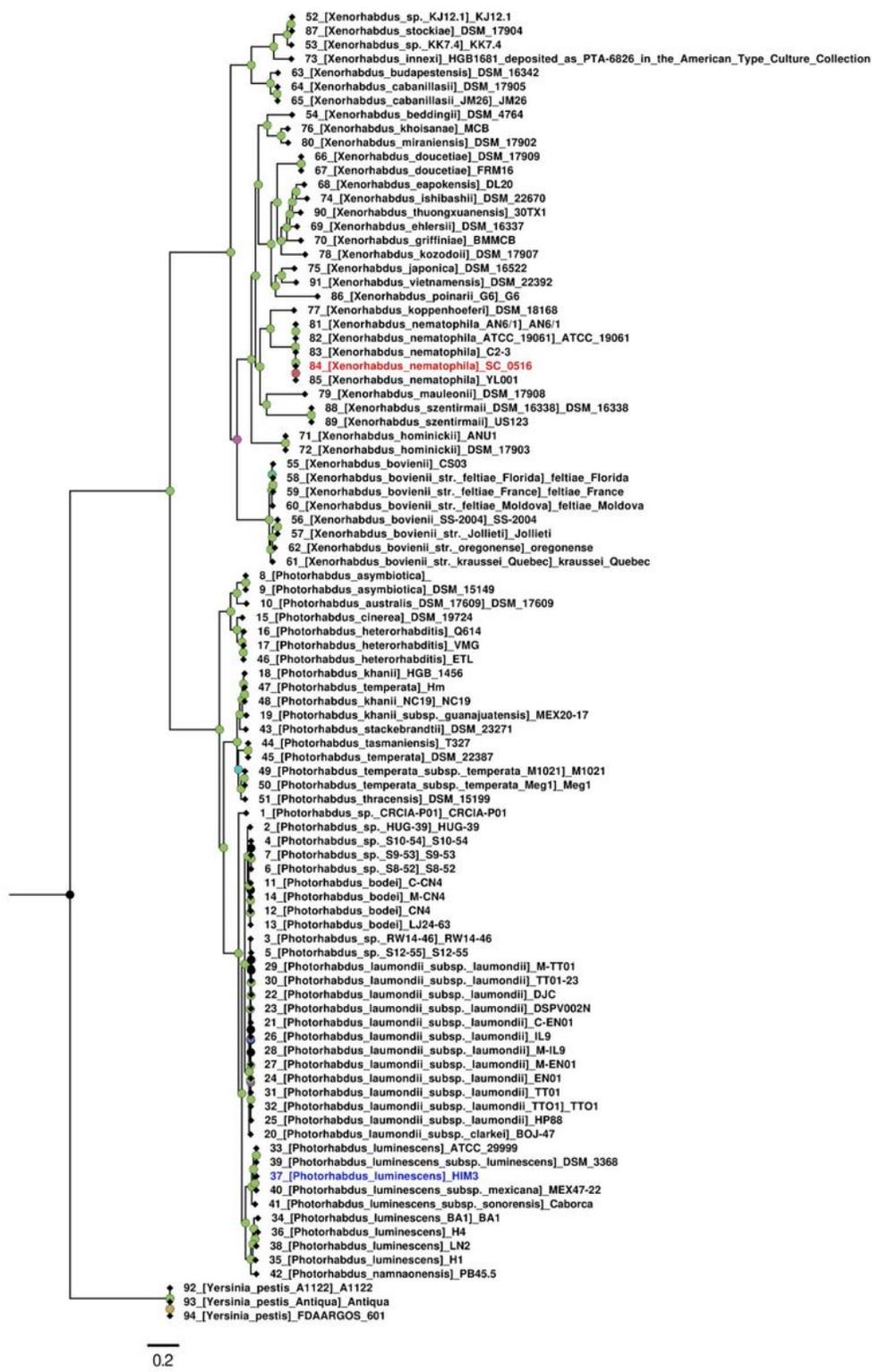
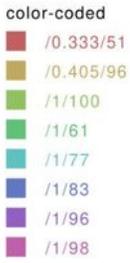
**Figure 1**

(A) Barplot of the frequency of genes within the 91 genomes included in this analysis. Genes present in a single genome represent lineage-specific genes, while at the opposite end of the scale, genes found in all 91 genomes represent the core genome. (B) Plot of the estimated pangenome sizes of *Xenorhabdus* - *Photorhabdus* fitted with the Tettelin function. (C) Venn diagram of the core genome generated by the BDBH, COG, and OMCL algorithms. (D) Plot of the estimated core genome sizes of *Xenorhabdus* - *Photorhabdus*.



**Figure 2**

COG functional classification of the genes belonging to the core genome. COG database annotation was used to classify the 348 members belonging to the core genome. The abundance of specific gene classes is shown.



**Figure 3**

Maximum likelihood core genome phylogeny. The ML tree is based on 281 consensus protein-coding genes defined by the GET\_HOMOLOGES package and filtered through GET\_PHYLOMARKERS selected for their optimal phylogenetic attributes. Nodal support values are color-coded as shown in the legend. The first value corresponds to approximate Bayesian support values, and the second corresponds to ultrafast bootstrap values, as implemented in IQTREE. The scale represents the number of expected substitutions

per site under the best-fit GTR2+F+R3 (binary) model. The phylogeny corresponds to the tree with the highest score (lnL= -3045265.731) found among 10 independent searches in IQTREE.

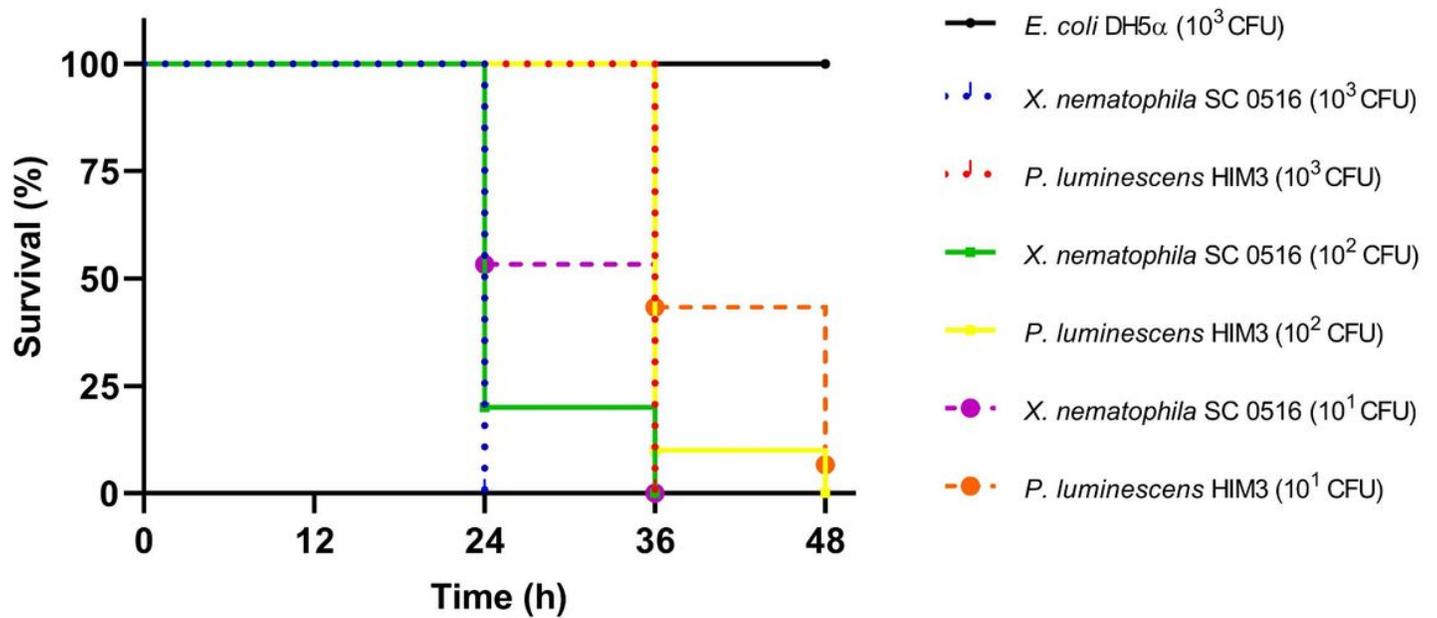
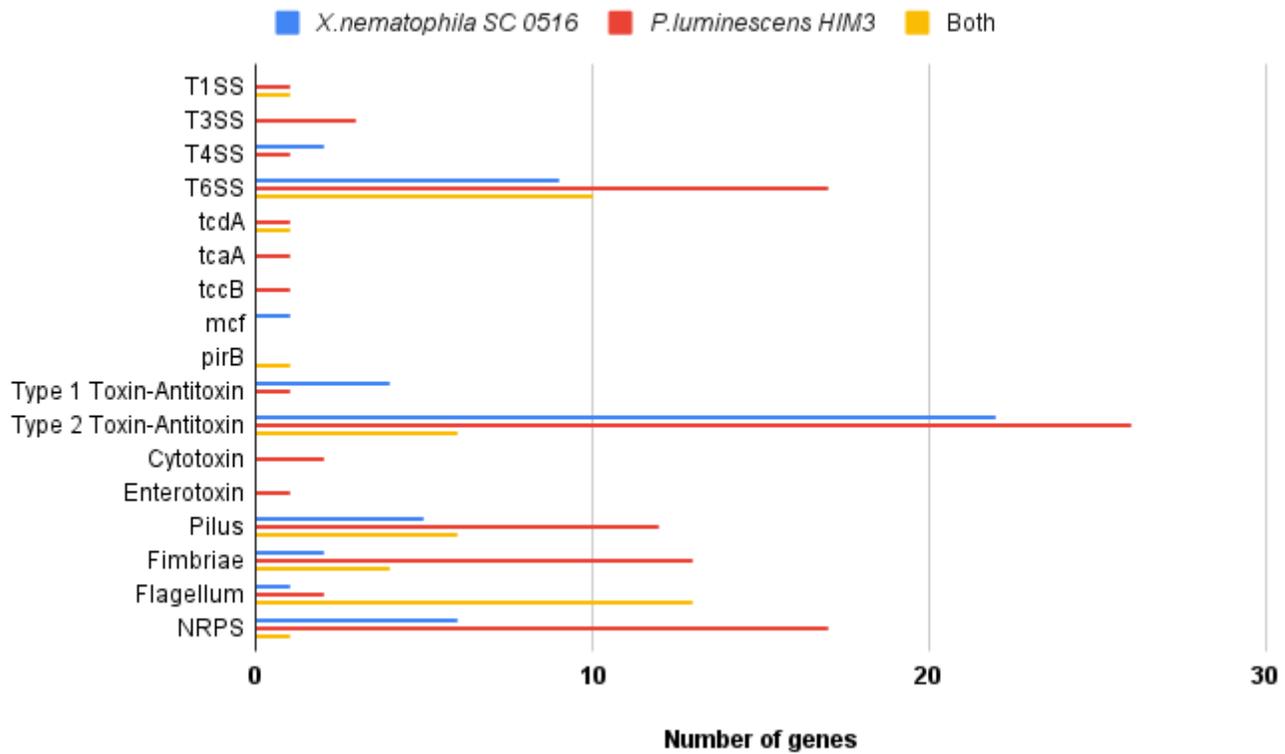


Figure 4

Kaplan–Meier survival curves of *Galleria mellonella* larvae after direct injection with *X. nematophila* SC 0516. Experiments were performed four times in independent bioassays, and survival was assessed at 12, 24, 36 and 48 hours after infection. Differences in survival (n=60) were calculated using the log-rank test  $\chi^2 = 457.636$ , df= 6,  $p < 0.001$ .



**Figure 5**

Unique and shared genes between *X. nematophila* SC 0516 and *P. luminescens* HIM3. The X axis indicates the total number of genes present in one or both bacteria, and the Y axis indicates the types of genes associated with virulence: Type Secretion System (TSS), Toxin Complex (tc), Makes Caterpillars Floppy (mcf), Photorhabdus Insect Related Toxins (Pir) and NRPS (Non-Ribosomal Peptide Synthetases).

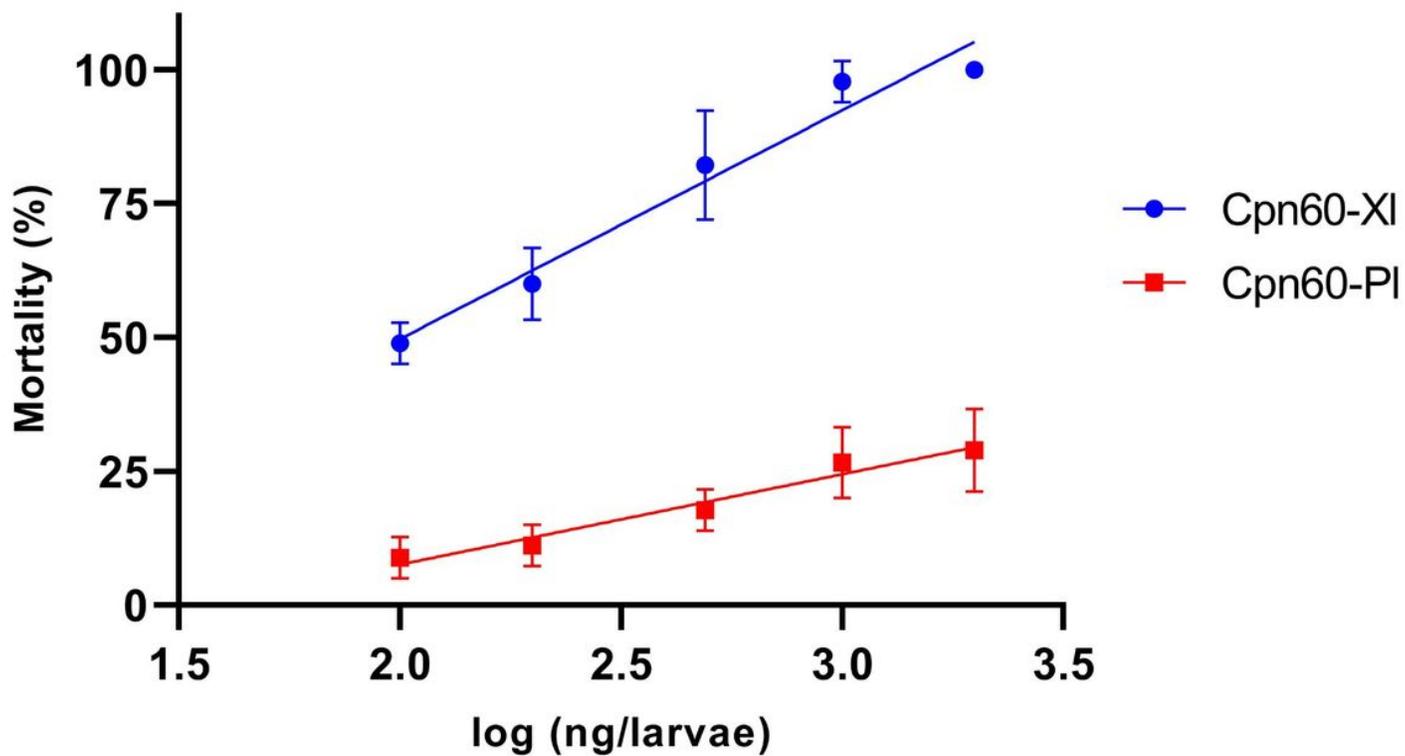


Figure 6

Insecticidal activity of GroEL from *X. nematophila* SC 0516 (Cpn60-Xn) and *P. luminescens* HIM3 (Cpn60-PI) against *G. mellonella* larvae. Plot of percent mortality against the protein concentration. Each dot in the plot represents the average of three individual experiments  $\pm$  standard error.