

Article

Not peer-reviewed version

Genome-Guided Insights into Xenobiotic Biodegradation and Secondary Metabolite Bioprocess Potential in *Enterobacter pseudoroggenkampii* G2.8

[Wellinton Cupozak-Pinheiro](#) , [Francine Piubeli](#) , Kauanny Plenz , [Maricy Bonfá](#) , [Rodrigo Pereira](#) *

Posted Date: 11 March 2026

doi: 10.20944/preprints202603.0890.v1

Keywords: functional genomics; fipronil; microbial genomics; One Health; bioremediation; bioprocesses; environmental biotechnology



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Genome-Guided Insights into Xenobiotic Biodegradation and Secondary Metabolite Bioprocess Potential in *Enterobacter pseudoroggenkampii* G2.8

Wellinton Cupozak-Pinheiro ¹, Francine Piubeli ², Kauanny Plenz ³, Maricy Bonfá ³
and Rodrigo Pereira ^{3,*}

¹ Faculty of Exact Sciences and Technology, Federal University of Grande Dourados (UFGD), Dourados, MS 79804-970, Brazil

² Department of Genetic, Faculty of Biology, University of Sevilla, 41012 Sevilla, Spain

³ Faculty of Biological and Environmental Sciences, Federal University of Grande Dourados (UFGD), Dourados 79804970, Brazil

* Correspondence: rodrigopereira@ufgd.edu.br

Abstract

Fipronil is a phenylpyrazole agrochemical widely used in agriculture and livestock production, posing persistent challenges of environmental contamination due to its toxicity and the formation of stable transformation products. Genome-based analyses provide a powerful framework for exploring the biotechnological potential of environmental microorganisms. The G2.8 isolate, obtained from fipronil-contaminated soil, was initially classified as *Enterobacter chengduensis*; however, taxonomic reassessment based on whole-genome sequencing combined with average nucleotide identity and digital DNA-DNA hybridization (ANI/dDDH $\approx 97\%$) reclassified this strain as *Enterobacter pseudoroggenkampii*. The occurrence of this species in a contaminated environmental niche highlights its relevance beyond previously reported clinical or plant-associated contexts and supports its potential role in bioremediation. The draft genome of *E. pseudoroggenkampii* G2.8 was assembled and subjected to rigorous quality assessment and functional annotation using genome-scale approaches. Functional analyses revealed 14 biosynthetic gene clusters, including non-ribosomal peptide synthetases, hybrid NRPS/polyketide synthases, and siderophore-related clusters, indicating potential for secondary metabolite production. In addition, genes encoding oxidoreductases, hydrolases, and esterases associated with xenobiotic transformation were identified, supporting the experimentally observed capacity of this strain to degrade fipronil and its toxic metabolites. Within a One Health framework, the genome exhibited only intrinsic antimicrobial resistance determinants, mainly related to efflux systems and chromosomal β -lactamases, with no evidence of mobile resistance elements, supporting an environmental safety profile. Overall, genome-guided functional and comparative analyses provide a robust foundation for identifying metabolic pathways involved in both biosynthesis and biodegradation, positioning *E. pseudoroggenkampii* G2.8 as a promising genome-guided candidate for metabolite-driven environmental biotechnology and reinforcing the value of microbial genomics in the development of sustainable bioprocesses.

Keywords: functional genomics; fipronil; microbial genomics; One Health; bioremediation; bioprocesses; environmental biotechnology

1. Introduction

The intensive use of synthetic pesticides in modern agriculture has significantly contributed to increased crop productivity. Fipronil is a highly relevant broad-spectrum agrochemical belonging to the phenylpyrazole class and is widely employed for the control of agricultural pests and urban vectors [1–4]. In Brazil, it is used in cotton, rice, potato, sugarcane, maize, soybean, and barley seed

crops, as well as in beans, pastures, wheat, and eucalyptus seedlings [5]. Its metabolites, fipronil-sulfone and fipronil-sulfide, are frequently detected in ecosystems and represent significant risks to non-target organisms, including soil biota, insects, and pollinators [4,6,12].

Its large-scale use combined with high environmental persistence may pose serious risks to human and animal health, causing hepatotoxicity, nephrotoxicity, neurotoxicity, and alterations in reproductive development and endocrine systems. In addition, it negatively affects native vegetation, soil microbiota, and natural resources such as water, highlighting the urgent need for sustainable mitigation strategies [7–9].

In response to this global concern, conventional strategies for pesticide removal such as excavation, incineration, and chemical treatments are limited by high costs, low selectivity, and the generation of secondary environmental impacts. In this context, microbial transformation of xenobiotics represents a sustainable alternative for the attenuation of contaminated environments, based on the ability of microorganisms to convert toxic compounds into less harmful forms or achieve mineralization [10–12]. This degradation may occur through specialized catabolic pathways expressed under xenobiotic exposure, involving enzymes such as oxygenases and hydrolases [25]. However, the effectiveness of these processes should be evaluated not only by the removal of the parent compound but also by the elimination of persistent metabolites such as fipronil-sulfone, which is frequently associated with chronic soil contamination [4,12].

Soil and rhizosphere bacteria play a central role in these processes, as they are constantly exposed to xenobiotic compounds and exhibit high metabolic plasticity. Recent studies demonstrate that certain bacteria can integrate pesticide degradation with typical plant growth promoting rhizobacteria (PGPR) traits, such as siderophore production, phosphate solubilization, and phytohormone synthesis, thereby simultaneously promoting the recovery of soil biological functionality [13,14]. This multifunctionality is particularly relevant in contaminated agricultural environments, where remediation must occur in synergy with productive sustainability.

Gram-negative bacteria of the *Enterobacter cloacae* complex are notable for their metabolic versatility and broad distribution in soils and the rhizosphere, where they can adapt to contaminated conditions [15]. Species within this group exhibit functions in xenobiotic metabolism, production of bioactive metabolites, and plant interaction, and may occur in environmental niches as well as in the human microbiota. However, some species, such as *Enterobacter cloacae*, are opportunistic pathogens of clinical relevance [16–19]. Comparative genomics studies demonstrate that environmental and plant-associated lineages display distinct genomic profiles characterized by the absence of major clinical virulence determinants and by the presence of genes related to rhizosphere adaptation and environmental metabolism [17,20,21]. These findings reinforce the need for genome-based approaches to differentiate environmentally adapted isolates with biotechnological potential from clinically relevant lineages.

With advances in bioinformatics, genome-scale functional and comparative analyses have become essential tools for uncovering metabolic potential and ecological adaptation in environmental microorganisms. High-quality genome assemblies enable the identification of biosynthetic gene clusters, xenobiotic-related enzymes, intrinsic resistance determinants, and regulatory networks that cannot be reliably inferred from phenotypic assays alone [11,24–26]. However, obtaining robust draft genomes from environmental isolates remains challenging due to genome plasticity, repetitive regions, and assembly fragmentation, which may affect gene recovery and functional annotation. Therefore, the integration of whole-genome sequencing (WGS), quality-controlled assembly strategies, and comparative genomics is critical for accurate taxonomic resolution, biosafety assessment, and functional inference in microorganisms with environmental relevance [22,23].

The isolate *Enterobacter pseudoroggenkampii* G2.8, previously described as *Enterobacter chengduensis*, has been experimentally shown to degrade fipronil and its most toxic metabolites in soil, representing a suitable model for genome-guided investigation of xenobiotic-associated metabolic traits [17]. The aim of this study was to perform a functional genomic characterization of the draft genome of *E. pseudoroggenkampii* G2.8, isolated from fipronil-contaminated soil, to identify

genes and metabolic pathways associated with xenobiotic transformation and secondary metabolite biosynthesis, and to evaluate its biotechnological potential and biosafety based on genome-scale analyses.

2. Materials and Methods

2.1. DNA Extraction and Genome Sequencing

The *Enterobacter pseudoroggenkampii* G2.8 strain was previously isolated and initially identified by Prado et al. [17] using 16S rRNA gene sequencing, being described at that time as *Enterobacter chengduensis* G2.8. The strain was obtained from soil samples collected at the experimental farm of the Federal University of Grande Dourados (UFGD), Brazil.

For the present study, genomic DNA was extracted at the Environmental Biotechnology Laboratory of UFGD using the BioSpin Omni Genomic DNA Extraction Kit (BioFlux), following the manufacturer's instructions. Whole-genome sequencing was performed using the Illumina MiSeq platform at the Central Multiuser Laboratory for Large-Scale DNA Sequencing and Gene Expression Analysis (LMSEQ), São Paulo State University (UNESP), Jaboticabal campus, according to the manufacturer's protocols. For a more comprehensive understanding, the critical steps used in this work are described in detail in Figure 1.

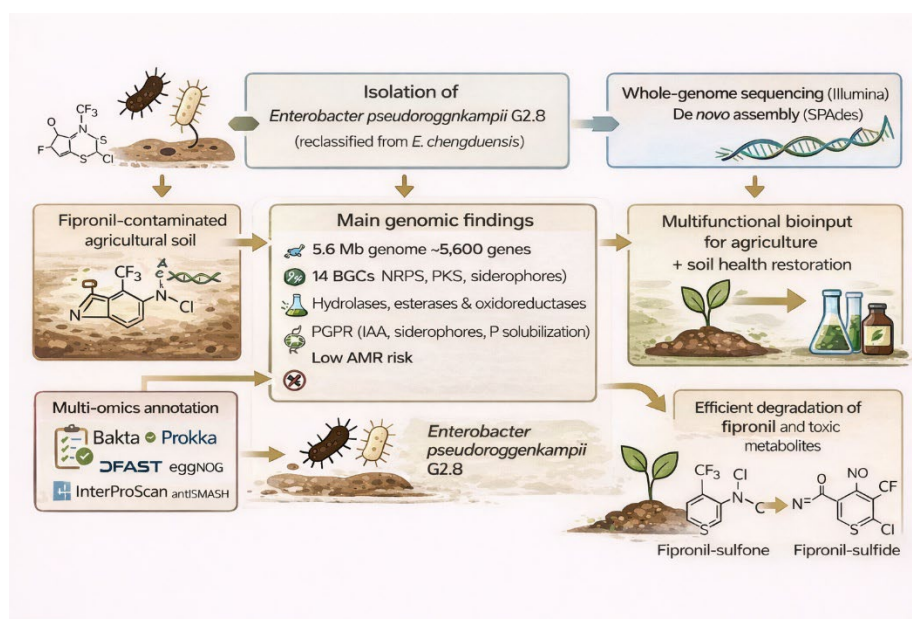


Figure 1. Schematic representation of the main steps carried out in this work.

2.2. Data Processing and Genome Assembly

2.2.1. Quality Control, Adapter Removal, Draft Genome Assembly, and Refinement

All bioinformatics analyses were performed using the Galaxy platform (<https://usegalaxy.org> and <https://usegalaxy.eu>), an open-source, web-based environment that ensures transparency and reproducibility of computational workflows. Tools were executed using default parameters unless otherwise specified, within the UseGalaxy pipeline (v24.1).

Sequence quality and adapter contamination were assessed using FastQC v0.74 [28], and adapter removal was performed with Trim Galore! v0.67 [29].

Genome assembly and quality assessment were re-performed in this study to ensure an optimized and high-confidence draft genome suitable for genome-scale functional, biosynthetic, and biosafety analyses. Although the isolate had been previously sequenced, reassembly using multiple

assemblers allowed a comparative evaluation of assembly performance, minimized assembler-specific biases, and ensured robustness for downstream analyses [22].

Draft genome assemblies were generated using multiple assembler tools implemented within the Galaxy environment, including SPAdes v3.15.5 [31], Velvet v1.2.10 [32], Shovill (SKESA) v1.1.0 [33], MEGAHIT v1.2.9 [34]. The use of different assemblers allowed a comparative evaluation of assembly performance and ensured the selection of the most suitable draft genome for downstream analyses.

Assembly quality was evaluated using QUAST v5.2.0 [36] based on standard metrics, including N50, N75, L50, L75, total assembly length (≥ 0 bp), and number of contigs. The SPAdes assembly was selected for subsequent analyses due to its superior balance between contiguity and fragmentation, as indicated by higher N50 values and a lower number of contigs (Table S1). Assembled contigs were subjected to post-assembly refinement by filtering sequences shorter than 500 bp using the “Filter sequences by length” tool (Galaxy version 1.2). Consensus polishing was performed using Pilon v1.20.1 [37], employing raw reads to correct base errors and small insertions/deletions. Taxonomic validation of the assembly was conducted using Kraken2 v2.1.3+galaxy1 [38] to ensure genome consistency and the absence of contamination. Assembly structure and integrity were further evaluated using QUAST v5.2.0 [36], generating metrics such as N50, L50, and GC content. Genome completeness was estimated using BUSCO v5.8.0 [39] with the Enterobacteriales odb10 dataset.

2.2.2. Gene Prediction and Functional Annotation, Resistance and Virulence Genes, Biosynthetic Clusters, and Plasmid Detection

Whole-genome average nucleotide identity (ANI) was rapidly calculated without alignment using FastANI (Galaxy version 1.3), considering values $\geq 95\%$ as indicative of species-level affiliation [43]. Genome-based taxonomic identification was also performed using the Type (Strain) Genome Server (TYGS) [44,45], with digital DNA–DNA hybridization (dDDH) estimates, for which values $\geq 70\%$ are recommended for species delimitation.

Protein-coding gene prediction for the *E. pseudoroggenkampii* G2.8 genome was performed using multiple independent pipelines: Prokka v1.11 [40] with RefSeq and UniProt databases; Bakta v1.9.4 [41] for standardized INSDC-compatible annotation, generating GFF3 and FASTA files for proteins and RNAs; and DFAST v1.6.0 [42] for CDS, rRNA, and tRNA annotation.

Functional assignment Kyoto Encyclopedia of Genes and Genomes (KO), Enzyme Commission number (EC), Clusters of Orthologous Groups (COG) and Gene Ontology (GO), was carried out using eggNOG-mapper v2.1.8 with default DIAMOND settings and the eggNOG v5.0.2 database [46,47]. InterProScan v5.59-91.0+galaxy3 [48] was used to identify protein domains and families (Pfam, TIGRFAM, SMART, PANTHER, Gene3D, SUPERFAMILY, PROSITE, MobiDBLite, Coils).

CRISPR arrays and CRISPR/Cas-associated genes were identified using the Proksee pipeline [49] and CRISPRCasFinder v1.1.0 [50] with default parameters. Biosynthetic gene clusters (BGCs) related to secondary metabolite biosynthesis were detected using antiSMASH v6.1.1+galaxy1 [26]. Antimicrobial resistance (AMR) gene screening was performed using ABRicate v1.0.1 [51] with the ResFinder database and CARD Resistance Gene Identifier (RGI) v1.3.1 [52], implemented through the Proksee pipeline [49]. Plasmid replicons were investigated using MOB-Recon v3.1.9+Galaxy0 [53], which did not detect complete plasmids; plasmidSPAdes v4.2.0+galaxy0 [62], which generated 406 additional contigs without robust plasmid evidence; MOB-Typer v3.1.9+Galaxy0 [53]; and PlasmidFinder v2.1.6+galaxy1 [54], which confirmed the chromosomal nature of the final 140 sequences. The genome was therefore considered essentially chromosomal, with no detectable plasmids.

3. Results and Discussion

3.1. Quality Control and Draft Genome Assembly

This study was guided by the hypothesis that strain G2.8, isolated from fipronil-contaminated soil and previously shown to efficiently degrade this insecticide and its toxic metabolites, harbors a genomic repertoire shaped by xenobiotic exposure, including metabolic, regulatory, and biosynthetic traits associated with environmental adaptation and biotechnological potential. Accurate genome reconstruction and taxonomic resolution are therefore essential prerequisites for interpreting functional traits and assessing environmental biosafety of candidate bioremediation organisms.

To ensure the selection of a robust and representative draft genome for downstream analyses, multiple genome assemblies were generated using different assembly algorithms. This comparative strategy allowed evaluation of assembly consistency and minimized assembler-specific biases, providing a reliable basis for selecting the most suitable genome for structural and functional analyses. When evaluating assemblies generated by SPAdes, Velvet, Shovill-SKESA, and MEGAHIT, SPAdes presented the highest N50 and lowest L50 values (Table S1). N50 represents the length of the shortest contig accounting for 50% of the assembly, whereas L50 indicates the number of scaffolds comprising more than 50% of the genome. Differences among assemblers reflect distinct algorithmic strategies and heuristics that influence contiguity, structural accuracy, and gene recovery depending on genome characteristics [31]. Therefore, considering its superior assembly continuity and performance metrics, SPAdes was selected for subsequent genome reconstruction and downstream analyses.

After polishing, the draft genome comprised approximately 5.1 Mb with ~55% GC content and 4,900–5,100 predicted Coding Sequences (CDSs), consistent with genomes of the *Enterobacter cloacae* complex. Genome completeness assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs) yielded 99.8% completeness (Complete: 99.8% [Single-copy: 83.6%, Duplicated: 16.1%]; Fragmented: 0.2%; Missing: 0.0%; n = 440), with no detectable contamination. These metrics indicate that the assembly is near-complete and structurally reliable, supporting its suitability for functional and comparative genomic analyses.

For environmental biotechnology applications such as bioremediation, accurate taxonomic identification is critical to ensure biosafety and functional reliability. Methods based solely on the 16S rRNA gene often lack resolution within metabolically diverse genera such as *Enterobacter*, whereas whole-genome sequencing (WGS) combined with average nucleotide identity (ANI) and digital DNA–DNA hybridization (dDDH) provides the current gold standard for species-level classification and biological risk assessment [22].

Average Nucleotide Identity (ANI) analysis revealed in Figure 2A a value of 97.90% between strain G2.8 and *Enterobacter pseudoroggenkampii* (reference genome ASM3040616v1), exceeding the ≥95% threshold recommended for species-level assignment. This result was supported by 1,418 orthologous matches distributed across 1,576 genomic fragments, indicating extensive genome-wide similarity and a strong phylogenomic relationship. Whole-genome taxonomic analysis using the Type (Strain) Genome Server (TYGS) further corroborated this classification, yielding dDDH = 80.5% relative to the type strain, well above the ≥70% species boundary (Table 1). Synteny analysis revealed high collinearity and conservation of genomic blocks between genomes, reinforcing taxonomic coherence.

Phylogenomic analysis using TYGS integrates genome similarity with genomic parameters such as GC content, genome size, protein count, and reference type strains, and all indicators consistently placed strain G2.8 within *E. pseudoroggenkampii*. In contrast, comparisons with *Enterobacter chengduensis* yielded ANI and dDDH values below species thresholds, supporting its exclusion. Together, ANI, dDDH, and synteny analyses provide robust genomic evidence for definitive reclassification of strain G2.8 as *Enterobacter pseudoroggenkampii* (Figure 2B).

This taxonomic reassignment provides the genomic framework necessary to interpret functional traits observed in strain G2.8, since metabolic capacity, biosynthetic potential, and biosafety profiles

are lineage-dependent within the *Enterobacter cloacae* complex. Notably, *E. pseudoroggenkampii* has been described as an environmentally associated species exhibiting genomic plasticity and adaptation to heterogeneous ecological niches, consistent with the contaminated-soil origin and biodegradation phenotype of strain G2.8 [20,21,55].

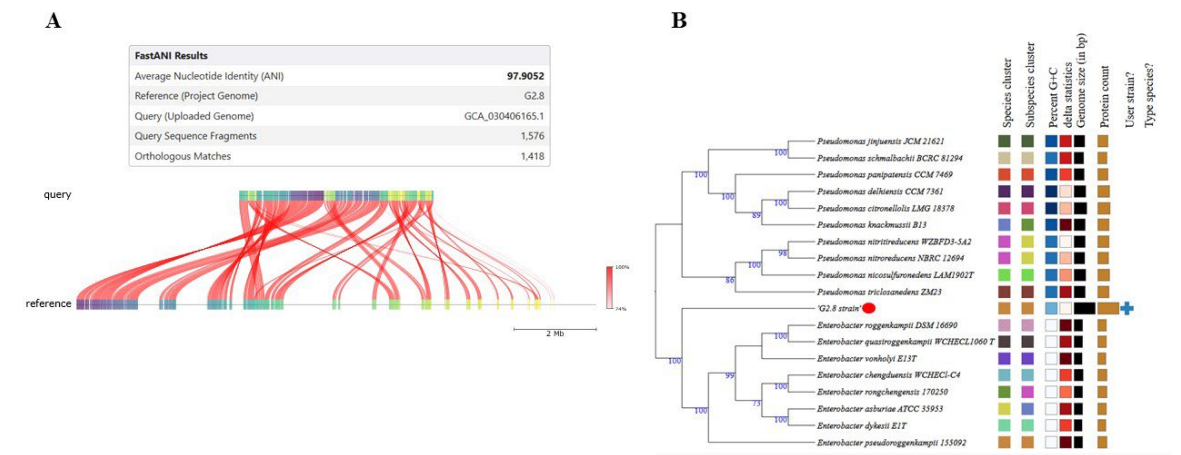


Figure 2. Genomic similarity and phylogenetic placement of strain G2.8. **A.** Global genome alignment between *E. pseudoroggenkampii* G2.8 and the reference genome based on average nucleotide identity (ANI). **B.** Tree inferred with FastME 2.1.6.1 from GBDP distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d5. The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 93.8 %. The tree was rooted at the midpoint.

Table 1. Taxonomic analyses used for the identification of the draft genome of <i>E. pseudoroggenkampii</i> G2.8.				
Method	Comparison	Results	Reference Thresholds	Interpretation
FastANI	G2.8 × <i>E. pseudoroggenkampii</i> (ASM3040616v1)	97,86%	≥95%	Same species
dDDH (d4, TYGS)	G2.8 × type strain	80,50%	≥70%	Same species
Syntenia	G2.8 × <i>E. pseudoroggenkampii</i>	High collinearity	—	Genomic conservation
Comparison with <i>E. chengduensis</i>	ANI / dDDH	< threshold	≥95 / ≥70	Distinct species

3.2. Prediction and Functional Annotation, Resistance and Virulence Genes, Biosynthetic Clusters, and Plasmid Detection

To assess the functional potential of the G2.8 strain, draft genome annotation was performed using independent and widely adopted prokaryotic annotation pipelines, such as Prokka, Bakta, and DFAST, allowing for a comparative and integrative evaluation of predicted coding sequences (Table 2). Prokka provides rapid gene prediction and annotation based on curated protein databases, Bakta applies a standardized ontology-driven framework with cross-referencing to multiple reference datasets, and DFAST offers a flexible annotation workflow optimized for draft genomes, improving functional consistency between annotations. The integration of multiple annotation platforms is

recommended for prokaryotic genome analysis because differences in gene prediction algorithms, reference databases, and functional ontologies can substantially influence CDS identification and functional assignment [41]. Such genome-driven functional inference approaches have been widely applied in environmental biotechnology to reveal metabolic capabilities and adaptive traits even when conventional annotations are incomplete [11]. In this context, the combined annotation strategy adopted here aimed to maximize functional resolution and reduce tool-specific bias in the interpretation of the genomic repertoire of strain G2.8.

In addition to core genome annotation, specialized in silico analyses were performed to screen for antimicrobial resistance determinants (ABRicate/CARD), virulence-associated genes, biosynthetic gene clusters (antiSMASH), and plasmid-associated sequences (MOB-recon and PlasmidFinder). The integration of these complementary tools is widely recommended in genome-based characterization of environmental bacteria, as it enables simultaneous assessment of metabolic capacity, biosynthetic potential, and biosafety-related traits [22,23]. This combined approach provided a comprehensive overview of the genomic architecture of strain G2.8, highlighting both its metabolic and ecological potential and the absence of complete plasmid replicons. The lack of plasmids and mobile resistance elements is consistent with environmental Enterobacter lineages, which typically exhibit intrinsic stress-tolerance determinants rather than clinically relevant resistance or virulence factors [20,21]. Such genome-wide screening is particularly important in candidate bioremediation organisms to distinguish environmentally specialized strains from clinically associated taxa and to ensure biosafety in applied contexts [22,23].

Between 75% and 80% of the predicted proteins were functionally annotated, while approximately 20%–25% remained classified as hypothetical [27,62], and no complete plasmids were identified (Figure 3A). Proportions of hypothetical proteins within this range are commonly reported in bacterial draft genomes and reflect the persistence of poorly characterized or lineage-specific genes in public databases, particularly in environmentally adapted taxa [27,62]. The circular genome map further highlights relatively homogeneous GC content across the chromosome, with localized variations and a clear GC skew pattern, consistent with a single chromosomal replicon and indicative of the origin and terminus of replication. These features reflect a structurally stable genome architecture, while subtle GC deviations may correspond to regions associated with genomic plasticity, potentially including horizontally acquired loci involved in environmental adaptation.

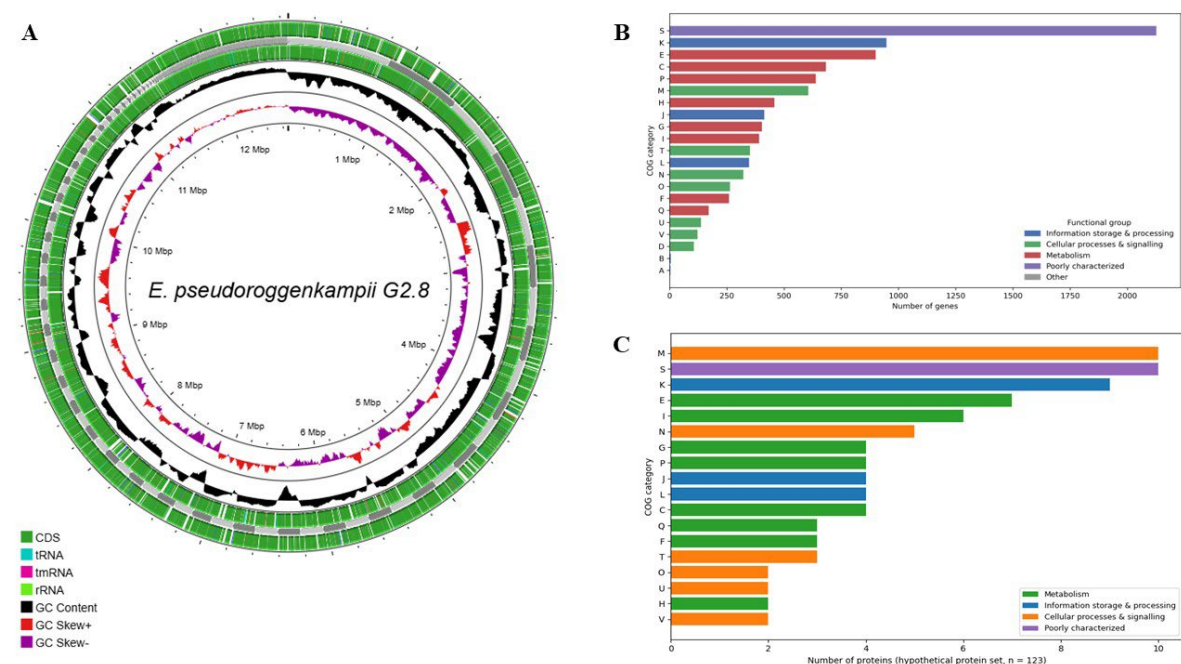


Figure 3. Genomic architecture and COG functional distribution of *Enterobacter pseudoroggenkampii* G2.8. **A.** Circular genome map showing CDSs on forward and reverse strands, RNA genes (tRNA, rRNA, tmRNA and ncRNA), GC content and GC skew. **B.** Genome-wide distribution of COG functional categories grouped into Information storage and processing, Cellular processes and signaling, Metabolism, and Poorly characterized. **C.** COG functional distribution of the 123 unique consensus hypothetical proteins inferred by eggNOG-mapper. Bars indicate the number of proteins per category; a single protein may be assigned to multiple COG classes. COG functional categories were defined as follows: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control, cell division and chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, posttranslational modification, protein turnover and chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; S, function unknown; T, signal transduction mechanisms; U, intracellular trafficking, secretion and vesicular transport; and V, defense mechanisms.

As shown in Table 2, Prokka annotated a total of 10,794 genomic elements, including 10,623 CDSs, 151 tRNAs, 18 rRNAs, and 2 tmRNAs. Bakta provided broader structural coverage, totaling 11,018 features, of which 10,618 correspond to CDSs, in addition to tRNAs, rRNAs, ncRNAs, small ORFs, and regulatory regions. In contrast, DFAST annotated 1,470 features, predominantly CDSs (1,450) (Table 2). These differences likely reflect the distinct prediction criteria and reference databases employed by each annotation tool. Such variation highlights intrinsic differences among algorithms, prediction strategies, and database resources, particularly regarding the identification of small ORFs, non-coding RNAs, and regulatory regions, which are more extensively explored by Bakta [41].

Comparison among the annotation pipelines revealed differences in the classification of hypothetical proteins. Prokka presented the highest absolute number of CDSs annotated as hypothetical (3,491), followed by Bakta (482) and DFAST (157) (Table 2). This discrepancy reflects the use of more up-to-date databases and alignment-independent annotation strategies in Bakta, resulting in improved functional assignment for CDSs previously classified as hypothetical by more conservative approaches [27,41]. Cross comparison among the pipelines did not reveal hypothetical proteins shared simultaneously by all three tools when genes were compared based on genomic coordinates (contig, start position, end position, and strand), demonstrating the strong dependence on annotation criteria and database selection (Table 2) [27]. Therefore, Bakta annotations were used as the primary reference for subsequent analyses, while Prokka and DFAST results were used in a complementary manner for validation and cross-comparison.

Consensus hypothetical proteins were defined as CDSs annotated as “hypothetical protein” by both pipelines and presenting exact matching genomic coordinates. A total of 133 consensus hypothetical proteins were initially identified. However, 10 of these corresponded to duplicated locus tags, in which the same gene was annotated in more than one entry. These duplicates were removed, resulting in a final set of 123 consensus hypothetical proteins, representing a highly conservative subset of CDSs with no functional assignment by the two main annotation pipelines employed (Tables S1 and S2). The use of consensus annotation across independent pipelines reduces tool-specific bias and increases confidence that these CDSs correspond to genuinely uncharacterized proteins rather than annotation artifacts. Such conservative subsets are commonly used to define high-confidence genomic “dark matter” in bacterial genomes and reflect proteins lacking detectable orthology or conserved domains despite multi-tool annotation [27,41,46–48].

Table 2.

1. Main statistics		2. Protein and enzyme prediction by tool		3. Secondary metabolites (antiSMASH)		4. Resistance genes (AMR)				5. Metabolic functionality	
Categories	N° of genes	Program	N° of genes / proteins	BGC type	N° of genes	CARD -RGI (36 genes)	N° of loci (occurrences)	ABRi cate (9 genes)	N° of loci (occurrences)	Functional category	N° of genes
Total de features (CDS + RNAs)	5.666	PROKKA – total CDS	10.794	CDS in BGC	195	<i>acrD</i>	4	<i>oqxB_1</i>	3	COG	5.542 genes with functional categories
Proteins annotated (Prokka)	5.666	Hypothetical proteins (Prokka)	3.491	Total clusters	14	<i>msbA</i>	3	<i>qnrE1_1</i>	2	KEGG KO	5.542 genes with annotated metabolic pathways
Proteins annotated (Bakta)	5.662	BAKTA – total CDS	11.018	NRPS	60	<i>leuO</i>	3	<i>fosA_1</i>	2	InterProScan	10.983 functional domains identified
Genes with eggNOG	5.541	Hypothetical proteins (Bakta)	482	NRPS + T1PKS	45	<i>adeF</i>	3	<i>oqxB_1</i> ; <i>oqxA_1</i>	1	-	-
Genes with COG	5.542	DFAST – total CDS	1.470	β-lactone	26	PBP3 (<i>H. influenzae</i>)	3	<i>blaACT-4_2</i>	1	-	-
Genes with KEGG KO	5.542	Hypothetical proteins (DFAST)	157	RiPP	21	<i>oqxB</i>	2	-	-	-	-
-	-	-	-	hserlactone	22	<i>soxR</i>	2	-	-	-	-
-	-	-	-	siderophore	13	<i>vanX</i>	2	-	-	-	-
-	-	-	-	butyrolactone	8	<i>vanH</i>	2	-	-	-	-
						<i>emrA / emrB</i>	2	-	-	-	-

CDS, coding DNA sequence; BGC, biosynthetic gene cluster; COG, Clusters of Orthologous Groups; KEGG KO, Kyoto Encyclopedia of Genes and Genomes Orthology; CARD-RGI, Comprehensive Antibiotic Resistance Database – Resistance Gene Identifier; NRPS, non-ribosomal

peptide synthetase; T1PKS, type I polyketide synthase; RiPP, ribosomally synthesized and post-translationally modified peptide.

EggNOG-mapper and InterProScan analyses applied to the complete set of predicted coding sequences (Table 2) consistently identified genes encoding AMP-binding, condensation, and oxidoreductase domains, which are characteristic of biosynthetic enzymes. Hydrolases and esterases associated with xenobiotic degradation, oxidative stress response, and metabolic adaptation were also detected across both annotation platforms. These enzyme classes are widely implicated in pesticide biodegradation and environmental stress tolerance, where oxidation–reduction and hydrolytic reactions mediate the transformation of complex xenobiotics into less toxic or metabolizable intermediates [8–11,15]. The predominance of these catalytic domains therefore supports a genome-wide metabolic architecture compatible with the experimentally observed degradation phenotype of strain G2.8.

At the genome-wide level, COG functional classification revealed a predominant distribution in categories related to unknown function (S), cell wall/membrane/envelope biogenesis (M), and transcription (K), in addition to categories associated with metabolism, transport, and cellular signaling (Figure 3B). Similar COG distributions have been reported in environmental and plant-associated *Enterobacter* lineages, in which regulatory and envelope-associated functions are enriched due to adaptation to fluctuating physicochemical and xenobiotic stresses in soil and rhizosphere environments [20,21].

To infer the functional role of proteins initially classified as hypothetical, a targeted analysis of the 123 unique consensus hypothetical proteins was performed using eggNOG-mapper. As a result, 84 proteins (68.3%) received functional annotations based on orthology, and 54 proteins (43.9%) were associated with at least one KEGG Orthology (KO) identifier (Table S2). Analysis of the COG functional categories assigned to this subset showed a similar distribution pattern, with enrichment in categories S, M, and K, as well as metabolism-related functions (Figure 2C), suggesting that a substantial fraction of these proteins may play regulatory and structural roles linked to environmental adaptation and xenobiotic stress responses. The recovery of functional signals from initially unannotated CDSs reinforces the concept that bacterial genomic “dark matter” often encodes niche-adaptive and stress-associated functions that remain poorly represented in reference databases [27,46–48].

To investigate the presence and organization of CRISPR–Cas defense systems and to assess potential interactions with mobile genetic elements, the genomic distribution of CRISPR arrays and isolated cas genes was analyzed (Figure 4A). Multiple CRISPR arrays were detected across distinct contigs of the draft genome and classified with high-confidence evidence levels, consistent with bona fide CRISPR loci (Table S3). The presence of multiple spacer-containing arrays indicates recurrent exposure to foreign genetic elements, such as bacteriophages and plasmids, as CRISPR spacers are typically acquired from invading DNA and retained as a molecular record of past infections. Such spacer diversity is commonly interpreted as evidence of historical phage–host interactions in environmental bacteria inhabiting dynamic microbial communities [61].

Although isolated cas genes were identified, including cas3 associated with Type I systems, these genes were not organized into complete operons nor directly associated with the detected CRISPR arrays (Figure 4B–C). Consequently, no complete CRISPR–Cas system could be classified in the analyzed genome. This genomic configuration is consistent with the presence of orphan CRISPR arrays and fragmented cas loci, which are commonly interpreted as evolutionary remnants of previously functional CRISPR–Cas systems that have undergone decay or partial loss during genome evolution [61].

In environmental bacteria, erosion or modular loss of CRISPR–Cas operons has been associated with genome plasticity and adaptation to habitats characterized by intense horizontal gene transfer and fluctuating selective pressures, such as contaminated soils and rhizosphere environments [27,61]. The occurrence of orphan CRISPR elements in strain G2.8 is therefore compatible with its environmental origin and suggests historical exposure to mobile genetic elements coupled with subsequent genomic restructuring. Such patterns have been reported in metabolically versatile environmental Enterobacterales, where residual CRISPR loci persist despite the absence of a fully functional adaptive immune system.

Alternatively, orphan CRISPR arrays may be retained as structural or regulatory genomic elements, potentially influencing genome organization or gene expression, although their functional significance in strain G2.8 remains to be elucidated.

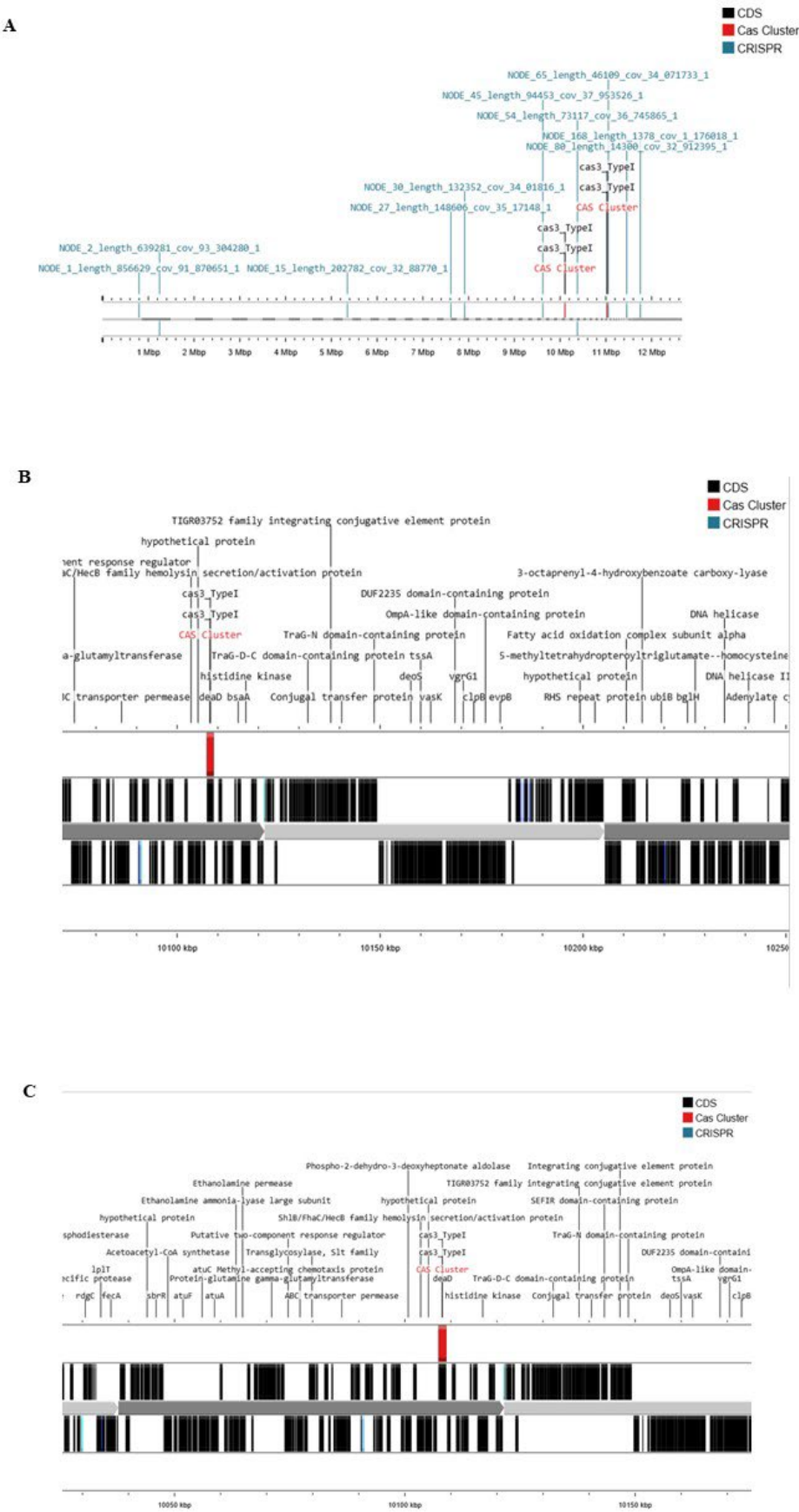


Figure 4. Distribution of CRISPR arrays and Cas gene clusters in the draft genome of *Enterobacter pseudoroggenkampii* G2.8. A. Genome-wide view showing coding sequences (CDS, green), CRISPR arrays (purple), and Cas gene clusters (red). B–C. Detailed views of the two isolated Cas loci, highlighting the presence of *cas3* (Type I) genes and surrounding genomic regions. Neither Cas cluster is colocalized with CRISPR arrays nor organized into a complete CRISPR–Cas operon, indicating fragmented and orphan Cas systems.

Using antiSMASH (antibiotics & Secondary Metabolite Analysis Shell), a specialized bioinformatic platform for the identification and characterization of biosynthetic gene clusters (BGCs) involved in secondary metabolite production, a total of 14 BGCs were detected in the draft genome of *Enterobacter pseudoroggenkampii* G2.8 (Table 2). These clusters included non-ribosomal peptide synthetases (NRPS), hybrid NRPS/type I polyketide synthases (NRPS/T1PKS), β -lactones, ribosomally synthesized and post-translationally modified peptides (RiPP-like), siderophores, arylpolyenes, butyrolactones, and homoserine lactone-related regions (Table S4).

NRPS and NRPS/T1PKS clusters are typically associated with the biosynthesis of bioactive peptides and polyketides with antimicrobial, siderophore, or signaling functions [24,26]. RiPP-like clusters encode ribosomally synthesized peptides that undergo post-translational modifications, often involved in microbial competition, ecological interactions, and stress tolerance [24]. Siderophore clusters are related to iron acquisition, arylpolyene clusters are associated with protection against oxidative stress, β -lactones are linked to antimicrobial activity, and butyrolactone- and homoserine lactone-related regions are commonly involved in quorum sensing and regulatory signaling processes [24–26].

Several BGCs exhibited similarity to known clusters such as aerobactin, turnerbactin, arylpolyenes, and crochelin A, although most showed low to moderate similarity values, suggesting the presence of potentially novel biosynthetic pathways [24,26]. These metabolite classes are commonly associated with iron acquisition, antimicrobial activity, oxidative stress protection, microbial competition, and ecological adaptation in soil and plant-associated bacteria [25,26]. The predominance of siderophore- and redox-related clusters is consistent with environmental *Enterobacter* lineages adapted to fluctuating nutrient availability and oxidative stress conditions typical of contaminated soils [20,21].

Consensus hypothetical proteins were also detected within BGC regions, particularly in NRPS and hybrid NRPS/T1PKS clusters, suggesting possible accessory, regulatory, or tailoring roles in secondary metabolite biosynthesis. The frequent occurrence of poorly characterized genes within BGC boundaries is widely reported and reflects the modular and rapidly evolving nature of specialized metabolism loci, which often harbor lineage-specific enzymatic components [24,27]. Together, these genomic features indicate that strain G2.8 harbors a diverse and potentially unique repertoire of biosynthetic gene clusters, supporting a genetic capacity for secondary metabolite production that remains to be functionally validated.

In addition to secondary metabolite biosynthetic clusters, genes associated with plant growth-promoting traits were identified in the genome of strain G2.8 (Table 3). These included genes involved in siderophore-mediated iron acquisition (*ent* and *fep* operons), indole-3-acetic acid (IAA) biosynthesis via the indole-3-pyruvate pathway (*ipdC*), phosphate solubilization through the pyrroloquinoline quinone (PQQ) system (*pqqABCDE* and *gcd*), as well as determinants related to rhizosphere colonization and environmental fitness, such as motility (*motAB*), chemotaxis (*cheAY*), and oxidative stress tolerance (*katG*, *sodA*, *sodB*).

The coexistence of nutrient acquisition, phytohormone production, and colonization-associated determinants indicates a genomic repertoire compatible with plant-associated ecological interactions. Similar combinations of siderophore production, IAA biosynthesis, and phosphate solubilization genes are widely reported in plant-associated *Enterobacter* spp. and are considered core genomic determinants of plant growth-promoting rhizobacteria (PGPR) functioning under abiotic and chemical stress conditions [13,14]. Comparative genomic analyses further demonstrate that environmental and plant-associated *Enterobacter* lineages, including *E. pseudoroggenkampii*, are

enriched in rhizosphere-adaptation and xenobiotic-response traits relative to clinical counterparts [20,21].

In strain G2.8, the presence of these PGPR-related genes together with biosynthetic clusters and oxidative-stress enzymes suggests a multifunctional ecological strategy combining plant interaction, environmental persistence, and metabolic adaptability. Such trait convergence is consistent with microorganisms adapted to contaminated agricultural soils, where nutrient limitation, oxidative stress, and xenobiotic exposure select for bacteria capable of both pollutant transformation and plant-associated fitness. Therefore, the PGPR genomic signature observed in strain G2.8 complements its biodegradation potential and supports its prospective application in integrated soil bioremediation and plant-assisted remediation systems, pending experimental validation.

Table 3. Genes detected in the draft genome of *E. pseudoroggenkampii* G2.8 associated with plant growth promotion (PGPR).

Functional category	Genes	Function	Evidence tool
Iron acquisition (siderophores)	<i>entA, entB, entC, entD, entE, entF</i>	Enterobactin biosynthesis	antiSMASH / eggNOG / KEGG
Iron acquisition (siderophores)	<i>fepA, fepB, fepC, fepD, fepG</i>	Siderophore transport	eggNOG / KEGG
Phytohormone production (IAA)	<i>ipdC</i>	Indole-3-pyruvate pathway (IAA)	eggNOG / KEGG
Phosphate solubilization	<i>pqqB, pqqC, pqqD, pqqE</i>	PQQ biosynthesis	eggNOG
Phosphate solubilization	<i>gcd</i>	Glucose dehydrogenase	KEGG
Rhizosphere colonization	<i>motA, motB</i>	Flagellar motilit	eggNOG
Rhizosphere colonization	<i>cheA, cheY</i>	Bacterial chemotaxis	eggNOG
Environmental stress tolerance	<i>katG, sodA, sodB</i>	Detoxification of reactive oxygen species	InterPro / KEGG
Microbial competition	NRPS-like BGC	Antimicrobial metabolite production	antiSMASH
Aromatic compound metabolism	<i>pcaH, pcaG</i>	Phenolic compound catabolism	KEGG

PGPR, plant growth-promoting rhizobacteria; IAA, indole-3-acetic acid; PQQ, pyrroloquinoline quinone; BGC, biosynthetic gene cluster; NRPS, non-ribosomal peptide synthetase; KEGG, Kyoto Encyclopedia of Genes and Genomes; eggNOG, evolutionary genealogy of genes: Non-supervised Orthologous Groups; InterPro, integrated protein signature database; antiSMASH, Antibiotics & Secondary Metabolite Analysis Shell (tool for prediction of secondary metabolite biosynthetic gene clusters).

Screening with ABRicate (ResFinder) and CARD-RGI identified genes associated with a restricted intrinsic antimicrobial resistance profile, mainly related to efflux systems and chromosomal β -lactamases, including *acrD*, *oqxA/oqxB*, *msbA*, *adeF*, and *blaACT*-like variants (Table 2). No acquired resistance determinants, virulence factors, or plasmid-associated resistance genes were detected.

This genomic configuration is consistent with environmental *Enterobacter* isolates, in which intrinsic efflux systems and chromosomal β -lactamases are commonly associated with tolerance to environmental toxicants rather than with clinically relevant multidrug resistance [18,20]. In contrast, clinical *Enterobacter* strains frequently harbor mobile resistance determinants and virulence genes located on plasmids and genomic islands, facilitating horizontal dissemination in healthcare environments [18,19,23].

The absence of high-risk virulence markers and mobile resistance elements in strain G2.8 therefore supports its environmental lineage and low biosafety concern, in agreement with genome-based frameworks used to distinguish environmental from clinically relevant *Enterobacter*

populations [20,21]. From an ecological perspective, the predominance of intrinsic resistance mechanisms together with stress-response and metabolic genes suggests adaptation to chemically impacted habitats rather than to host-associated niches.

Recent studies on pesticide biodegradation report that many bacterial degraders exhibit incomplete mineralization or accumulation of toxic intermediates, particularly in the case of fipronil, whose transformation products such as fipronil-sulfone often persist in the environment [8–10,15]. In contrast, strain G2.8 has been experimentally demonstrated to degrade both fipronil and its major toxic metabolites in contaminated soil systems [17], indicating a more advanced degradation performance than that reported for several previously described isolates [8,57]. Genome analysis of G2.8 additionally revealed the presence of diverse biosynthetic gene clusters, including NRPS and hybrid NRPS/TIPKS systems, together with plant growth-promoting-related genes and intrinsic resistance determinants associated with environmental stress tolerance. Notably, although G2.8 belongs to *E. pseudoroggenkampii*, it was isolated from fipronil-contaminated soil, whereas the reference genome (ASM3040616v1) originates from a plant-associated (rice) environment. This ecological distinction is reflected in the genomic repertoire of G2.8, which includes multiple features associated with xenobiotic exposure and environmental adaptation. While many reported pesticide-degrading bacteria are primarily characterized by catabolic potential alone, the coexistence in G2.8 of genomic features related to xenobiotic transformation, secondary metabolism, rhizosphere-associated traits, and adaptive resistance suggests a broader ecological potential in contaminated soil environments. This integrated genomic profile distinguishes G2.8 both from many described degraders and from plant-associated reference representatives of the same species, supporting its classification as an environmentally adapted strain rather than a strictly plant-associated lineage.

Future investigations should prioritize experimental validation of the predicted biosynthetic and metabolic capabilities of strain G2.8 through controlled bench-scale assays and functional analyses. In particular, metabolomic profiling, expression studies of biosynthetic gene clusters, and biodegradation experiments under defined environmental conditions will be essential to confirm the production and activity of secondary metabolites inferred from genomic data. Such approaches may elucidate the ecological roles and potential applications of these metabolites in xenobiotic transformation, microbial interaction, and plant-associated processes. Additionally, evaluating metabolite production dynamics and degradation efficiency in soil microcosms or bioreactor systems will be critical to assess the feasibility of deploying strain G2.8 or its metabolic products in environmental and agricultural bioprocesses. These future studies will enable translation of genome-based predictions into practical applications and clarify the biotechnological relevance of secondary metabolites encoded in the G2.8 genome.

4. Conclusions

In this study, we generated a high-quality draft genome of *Enterobacter pseudoroggenkampii* G2.8, an environmental strain isolated from fipronil-contaminated soil. Genome-scale analyses revealed a multifunctional genetic repertoire including biosynthetic gene clusters related to NRPS and siderophore production, enzymes associated with xenobiotic transformation, and regulatory features linked to environmental adaptation. Although no genes explicitly annotated as fipronil-degrading were detected, the presence of oxidoreductases, hydrolases, esterases, and secondary metabolite pathways is consistent with the previously demonstrated degradation phenotype of this strain. The absence of plasmid-borne resistance determinants and high-virulence factors further supports a favorable biosafety profile consistent with an environmental lineage.

Overall, the genome-guided characterization of strain G2.8 demonstrates the coexistence of genomic features associated with xenobiotic transformation, secondary metabolism, and plant-associated functions within a single organism. This integrated repertoire highlights its relevance for studies on microbial metabolite processes in contaminated environments. Collectively, these findings establish *E. pseudoroggenkampii* G2.8 as an environmentally adapted strain with combined

biodegradation and biosynthetic potential, providing a genomic foundation for future experimental validation and exploration of metabolite-oriented environmental biotechnology applications

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org., Table S1. Quality assessment of the draft genome assembly of the *E. pseudoroggenkampii* G2.8 isolate using QUAST.

Author Contributions: W.J.C.P., R.M.P. and M.R.L.B.; methodology: W.J.C.P and R.M.P.; software: W.J.C.P. and R.M.P.; validation: W.J.C.P., R.M.P. and F.P.; formal analysis: W.J.C.P., M.R.L.B. and R.M.P.; investigation: W.J.C.P., R.M.P., M.R.L.B. and K.W.P; resources: R.M.P.; date curation: W.J.C.P., R.M.P., M.R.L.B., and F.P.; writing-original draft preparation: W.J.C.P., R.M.P., and F.P.; writing, review and editing: W.J.C.P., R.M.P., M.R.L.B. and F.P.; supervision: W.J.C.P., R.M.P., and F.P.; project administration: R.M.P. and M.R.L.B.; funding acquisition: R.M.P. All authors have read and agreed to the published version of the manuscript.

Funding: Fundect—Support Foundation for the Development of Education, Science and Technology of the State of Mato Grosso do Sul. CNPq -National Council for Scientific and Technological Development. CAPES- National Council for Scientific and Technological Development and Coordination for the Improvement of Higher Education Personnel.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data of *Enterobacter pseudoroggenkampii* strain G2.8 were deposited in the Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI) and can be accessed by BioProject PRJNA1403648. The SRA database is public and can be accessed free of charge through the NCBI website.

Acknowledgments: The Centralized Multiuser Laboratory for Large Scale DNA Sequencing -LMSeq of UNESP Campus Jaboticabal, SP Brazil. Universidade Federal da Grande Dourados, MS Brazil -UFGD.

Conflicts of Interest: The authors declare no conflicts of interest.:

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

References

1. Tingle, C.C.D.; Rother, J.A.; Dewhurst, C.F.; Lauer, S.; King, W.J. Fipronil: Environmental fate, ecotoxicology, and human health concerns. *Rev. Environ. Contam. Toxicol.* 2007, 176, 1–66.
2. Silva, M.R.; Sartori, S.; Castro, A.P.; Krüger, R.H.; et al. Soil bacterial communities in the Brazilian Cerrado: Response to vegetation type and management. *Acta Oecol.* 2019, 100, 103463. <https://doi.org/10.1016/j.actao.2019.103463>.
3. Provase, M.; Boeing, G.A.N.S.; Tsukada, E.; Salla, R.F.; Abdalla, F.C. Impact of environmental concentrations of fipronil on DNA integrity and brain structure of *Bombus atratus* bumblebees. *Environ. Toxicol. Pharmacol.* 2024, 110, 104536. <https://doi.org/10.1016/j.etap.2024.104536>.
4. Goulart, L.M.; et al. Fate and toxicity of 2,4-D and fipronil in mesocosm systems. *Environ. Pollut.* 2024, 332, 121899. <https://doi.org/10.1016/j.envpol.2023.121899>.
5. ANVISA. F43 Fipronil. Monografias Autorizadas; Agência Nacional de Vigilância Sanitária: Brasília, Brazil, 2022. Available online: <https://encurtador.com.br/dlyF6> (accessed on 15 January 2026).

6. Lozano, V.L.; Defarge, N.; Mesnage, R.; Hennequin, D.; Cassier, R.; de Vendômois, J.S.; Panoff, J.M.; Séralini, G.E.; Amiel, C. Sex-dependent impacts of fipronil on non-target organisms: Oxidative stress and genotoxicity. *Ecotoxicol. Environ. Saf.* 2021, 208, 111491. <https://doi.org/10.1016/j.ecoenv.2020.111491>.
7. Essack, S.Y. Environment: The neglected component of the One Health triad. *Lancet Planet. Health* 2018, 2, e238–e239. [https://doi.org/10.1016/S2542-5196\(18\)30124-4](https://doi.org/10.1016/S2542-5196(18)30124-4).
8. Zhou, Z.; Wu, X.; Lin, Z.; Pang, S.; Mishra, S.; Chen, S. Biodegradation of fipronil: Current state of mechanisms of biodegradation and future perspectives. *Appl. Microbiol. Biotechnol.* 2021, 105, 7695–7708. <https://doi.org/10.1007/s00253-021-11605-3>.
9. Bhatt, P.; Gangola, S.; Ramola, S.; Bilal, M.; Bhatt, K.; Huang, Y.; Zhou, Z.; Chen, S. Insights into the toxicity and biodegradation of fipronil in contaminated environment. *Microbiol. Res.* 2023, 266, 127247. <https://doi.org/10.1016/j.micres.2022.127247>.
10. Gupta, A.; Kumar, V.; Chauhan, A.; Singh, N.; Chandra, R. Progress in bioremediation of pesticide residues in soil and water: Microbial consortia and enzyme-assisted degradation. *Environ. Eng. Res.* 2020, 25, 446–461. <https://doi.org/10.4491/eer.2020.446>.
11. Bonfá, M.R.L.; Durrant, L.R.; Piubeli, F.A.; Prado, C.C.A.; Guima, S.E.S.; Pereira, R.M. Bioprospecting of microbial enzymes with application in environmental biotechnology: An omics approach. In *Bioremediation and Bioeconomy*; Wiley: Hoboken, NJ, USA, 2024; Chapter 17. <https://doi.org/10.1002/9783527844340.ch17>.
12. Bonfá, M.R.L.; Prado, C.C.A.; Piubeli, F.A.; Durrant, L.R. Fipronil microbial degradation: An overview from bioremediation to metabolic pathways. In *Pesticides Bioremediation*; Siddiqui, S.; Meghvansi, M.K.; Chaudhary, K.K., Eds.; Springer: Cham, Switzerland, 2022. https://doi.org/10.1007/978-3-030-97000-0_3.
13. Haque, M.A.; et al. Unveiling chlorpyrifos mineralizing and tomato plant-growth activities of *Enterobacter* sp. *Front. Microbiol.* 2022, 13, 1060554. <https://doi.org/10.3389/fmicb.2022.1060554>.
14. Aswathi, A.; et al. Plant growth-promoting traits of rhizosphere bacteria under pesticide stress. *Appl. Soil Ecol.* 2024, 193, 104941. <https://doi.org/10.1016/j.apsoil.2023.104941>.
15. Ali, S.; Khan, I.; Ali, M.; Shah, N.S.; Shin, J.H. Biodegradation and detoxification of pesticides: Insight into microorganisms and enzymes. *Biodegradation* 2021, 32, 271–296. <https://doi.org/10.1007/s10532-021-09940-7>.
16. Wu, W.; Feng, Y.; Zong, Z. Characterization of a strain representing a new *Enterobacter* species, *Enterobacter chengduensis* sp. nov. *Antonie van Leeuwenhoek* 2019, 112, 491–500. <https://doi.org/10.1007/s10482-018-1180-z>.
17. Prado, C.; et al. Fipronil degradation in soil by *Enterobacter chengduensis* strain G2.8. *Life* 2023, 13, 1935. <https://doi.org/10.3390/life13091935>.
18. Zagui, G.S.; Moreira, N.C.; Santos, D.V.; Paschoalato, C.F.P.R.; Sierra, J.; Nadal, M.; Domingo, J.L.; Darini, A.L.C.; Andrade, L.N.; Segura-Muñoz, S.I. Multidrug-resistant *Enterobacter* spp. in wastewater and surface water: Molecular characterization of β -lactam resistance and metal tolerance genes. *Environ. Res.* 2023, 233, 116443. <https://doi.org/10.1016/j.envres.2023.116443>.
19. Gonçalves, D.L.D.R.; Chang, M.R.; Nobrega, G.D.; Venancio, F.A.; Higa Júnior, M.G.; Fava, W.S. Hospital sewage in Brazil: A reservoir of multidrug-resistant carbapenemase-producing *Enterobacteriaceae*. *Braz. J. Biol.* 2024, 84, e277750. <https://doi.org/10.1590/1519-6984.277750>.
20. Wu, W.; et al. Taxonogenomic insights into pathogenic and environmental *Enterobacter* lineages. *Microb. Genom.* 2023, 9, mgen000932. <https://doi.org/10.1099/mgen.0.000932>.
21. Taboadela-Hernanz, J.; et al. Genomic differentiation of plant-associated and clinical *Enterobacter* strains. *Microorganisms* 2025, 13, 114. <https://doi.org/10.3390/microorganisms13010114>.
22. Chun, J.; et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 2018, 68, 461–466. <https://doi.org/10.1099/ijsem.0.002516>.
23. Didelot, X.; et al. Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* 2016, 17, 601–612. <https://doi.org/10.1038/nrg.2016.78>.
24. Medema, M.H.; et al. Minimum information about a biosynthetic gene cluster (MIBiG). *Nat. Chem. Biol.* 2015, 11, 625–631. <https://doi.org/10.1038/nchembio.1890>.

25. Plaza, G.A.; Banat, I.M.; Mrozik, W.; Briguglio, N.; Kaczorek, E. Biodegradation of xenobiotics by bacterial and fungal isolates: A review of mechanisms and environmental applications. *Int. J. Environ. Res. Public Health* 2020, 17, 196. <https://doi.org/10.3390/ijerph17010196>.
26. Blin, K.; et al. antiSMASH 7.0: Improved detection and comparative analysis of biosynthetic gene clusters. *Nucleic Acids Res.* 2023, 51, W46–W50. <https://doi.org/10.1093/nar/gkad344>.
27. Galperin, M.Y.; et al. The dark matter of bacterial genomes. *Nat. Rev. Microbiol.* 2019, 17, 491–504. <https://doi.org/10.1038/s41579-019-0210-2>.
28. Andrews, S. FastQC: A quality control tool for high throughput sequence data; Babraham Bioinformatics: Cambridge, UK, 2010. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 15 January 2026).
29. Krueger, F. Trim Galore! 2021. Available online: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (accessed on 15 January 2026).
30. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; Pyshkin, A.V.; Sirotkin, A.V.; Vyahhi, N.; Tesler, G.; Alekseyev, M.A.; Pevzner, P.A. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 2012, 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>.
31. Zerbino, D.R.; Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008, 18, 821–829. <https://doi.org/10.1101/gr.074492.107>.
32. Seemann, T. Shovill: Faster SPAdes assembly of Illumina reads. 2017. Available online: <https://github.com/tseemann/shovill> (accessed on 15 January 2026).
33. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015, 31, 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
34. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 2013, 29, 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
35. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; Earl, A.M. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 2014, 9, e112963. <https://doi.org/10.1371/journal.pone.0112963>.
36. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015, 31, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
37. Wood, D.E.; Salzberg, S.L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 2014, 15, R46. <https://doi.org/10.1186/gb-2014-15-3-r46>.
38. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 2014, 30, 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
39. Schwengers, O.; et al. Bakta: Rapid and standardized annotation of bacterial genomes. *Microb. Genom.* 2021, 7, 000685. <https://doi.org/10.1099/mgen.0.000685>.
40. Tanizawa, Y.; Fujisawa, T.; Nakamura, Y. DFAST: A flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* 2018, 34, 1037–1039.
41. Jain, C.; Rodriguez-R, L.M.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 2018, 9, 5114. <https://doi.org/10.1038/s41467-018-07641-9>.
42. Meier-Kolthoff, J.P.; Göker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* 2019, 10, 2182. <https://doi.org/10.1038/s41467-019-10210-3>.
43. Meier-Kolthoff, J.P.; Sardà Carbasse, J.; Peinado-Olarte, R.L.; Göker, M. TYGS and LPSN: A database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* 2022, 50, D801–D807. <https://doi.org/10.1093/nar/gkab902>.
44. Huerta-Cepas, J.; Szklarczyk, D.; Forslund, K.; Cook, H.; Heller, D.; Walter, M.C.; Rattei, T.; Mende, D.R.; Sunagawa, S.; Kuhn, M.; Jensen, L.J.; von Mering, C.; Bork, P. eggNOG 4.5: A hierarchical orthology

- framework with improved functional annotations. *Nucleic Acids Res.* 2016, 44, D286–D293. <https://doi.org/10.1093/nar/gkv1248>.
45. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; von Mering, C.; Bork, P. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* 2017, 34, 2115–2122. <https://doi.org/10.1093/molbev/msx148>.
 46. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; Pesseat, S.; Quinn, A.F.; Sangrador-Vegas, A.; Scheremetjew, M.; Yong, S.-Y.; Lopez, R.; Hunter, S. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 2014, 30, 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
 47. Grant, J.R.; Enns, E.; Marinier, E.; Mandal, A.; Herman, E.K.; Chen, C.; Graham, M.; Van Domselaar, G.; Stothard, P. Proksee: In-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res.* 2023, 51, W484–W490. <https://doi.org/10.1093/nar/gkad326>.
 48. Couvin, D.; Bernheim, A.; Toffano-Nioche, C.; Touchon, M.; Michalik, J.; Néron, B.; Rocha, E.P.C.; Vergnaud, G.; Gautheret, D.; Pourcel, C. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 2018, 46, W246–W251. <https://doi.org/10.1093/nar/gky425>.
 49. Seemann, T. ABRicate: Mass screening of contigs for antibiotic resistance genes. 2016. Available online: <https://github.com/tseemann/abricate> (accessed on 15 January 2026).
 50. Alcock, B.P.; Huynh, W.; Chalil, R.; Van, V.; Hong, Y.G.; et al.; McArthur, A.G. CARD 2023: Expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* 2023, 51, D690–D699. <https://doi.org/10.1093/nar/gkac920>.
 51. Robertson, J.; Quick, J.; Taylor, D.; Woodall, C.; Akpan, I.; Kontoravdi, C.; Loman, N.J. MOB-suite: Software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb. Genom.* 2018, 4, e000206. <https://doi.org/10.1099/mgen.0.000206>.
 52. Carattoli, A.; Hasman, H. PlasmidFinder and in silico pMLST: Identification and typing of plasmid replicons in whole-genome sequencing (WGS). In *Horizontal Gene Transfer*; Springer: New York, NY, USA, 2019; pp. 285–294. https://doi.org/10.1007/978-1-4939-9877-7_20.
 53. Wu, W.; Wei, L.; Feng, Y.; Kang, M.; Zong, Z. *Enterobacter huaxiensis* sp. nov. and *Enterobacter chuandaensis* sp. nov., recovered from human blood. *Int. J. Syst. Evol. Microbiol.* 2019, 69, 708–714.
 54. Guima, S.E.S.; Piubeli, F.; Bonfá, M.R.L.; Pereira, R.M. New insights into the effect of fipronil on the soil bacterial community. *Microorganisms* 2023, 11, 52. <https://doi.org/10.3390/microorganisms11010052>.
 55. Prado, C.C.A.; et al. Fipronil biodegradation and metabolization by *Bacillus megaterium* strain E1. *J. Chem. Technol. Biotechnol.* 2021, 97, 474–481. <https://doi.org/10.1002/jctb.6758>.
 56. Destoumieux-Garzón, D.; et al. The One Health concept: 10 years old and a long road ahead. *Front. Vet. Sci.* 2018, 5, 14. <https://doi.org/10.3389/fvets.2018.00014>.
 57. Hernández, S.; et al. One Health approach to antimicrobial resistance: Environmental perspective. *Sci. Total Environ.* 2019, 681, 476–488. <https://doi.org/10.1016/j.scitotenv.2019.05.052>.
 58. FAO/WHO. One Health approach to food safety. Food and Agriculture Organization of the United Nations/World Health Organization, 2017.
 59. Makarova, K.S.; et al. Evolutionary classification of CRISPR–Cas systems: A burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 2020, 18, 67–83. <https://doi.org/10.1038/s41579-019-0299-y>.
 60. Antipov, D.; Hartwick, N.; Shen, M.; Raiko, M.; Lapidus, A.; Pevzner, P.A. plasmidSPAdes: Assembling plasmids from whole genome sequencing data. *Bioinformatics* 2016, 32, 3380–3387. <https://doi.org/10.1093/bioinformatics/btw493>.
 61. Unknown genes in bacterial genomes: distribution, diversity and association with environmental niches. *Microbial Genomics* 2023, 9(4), 000996. <https://doi.org/10.1099/mgen.0.000996>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.