



Article

Whole Genome Sequencing Based Taxonomic Classification, and Comparative Genomic Analysis of Potentially Human Pathogenic *Enterobacter* spp. Isolated from Chlorinated Wastewater in the North West Province, South Africa

Tawanda E. Maguvu ^{*} and Cornelius C. Bezuidenhout

Unit for Environmental Sciences and Management-Microbiology, North-West University, Private Bag X6001, Potchefstroom 2520, South Africa; carlos.bezuidenhout@nwu.ac.za

* Correspondence: temaguvu@gmail.com



Citation: Maguvu, T.E.; Bezuidenhout, C.C. Whole Genome Sequencing Based Taxonomic Classification, and Comparative Genomic Analysis of Potentially Human Pathogenic *Enterobacter* spp. Isolated from Chlorinated Wastewater in the North West Province, South Africa. *Microorganisms* **2021**, *9*, 1928. <https://doi.org/10.3390/microorganisms9091928>

Academic Editor: Antonio Ventosa

Received: 17 August 2021
Accepted: 5 September 2021
Published: 10 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Comparative genomics, in particular, pan-genome analysis, provides an in-depth understanding of the genetic variability and dynamics of a bacterial species. Coupled with whole-genome-based taxonomic analysis, these approaches can help to provide comprehensive, detailed insights into a bacterial species. Here, we report whole-genome-based taxonomic classification and comparative genomic analysis of potential human pathogenic *Enterobacter hormaechei* subsp. *hoffmannii* isolated from chlorinated wastewater. Genome Blast Distance Phylogeny (GBDP), digital DNA-DNA hybridization (dDDH), and average nucleotide identity (ANI) confirmed the identity of the isolates. The algorithm PathogenFinder predicted the isolates to be human pathogens with a probability of greater than 0.78. The potential pathogenic nature of the isolates was supported by the presence of biosynthetic gene clusters (BGCs), aerobactin, and aryl polyenes (APEs), which are known to be associated with pathogenic/virulent strains. Moreover, analysis of the genome sequences of the isolates reflected the presence of an arsenal of virulence factors and antibiotic resistance genes that augment the predictions of the algorithm PathogenFinder. The study comprehensively elucidated the genomic features of pathogenic *Enterobacter* isolates from wastewaters, highlighting the role of wastewaters in the dissemination of pathogenic microbes, and the need for monitoring the effectiveness of the wastewater treatment process.

Keywords: *Enterobacter*; comparative genomics; pan-genome analysis; taxonomic classification

1. Introduction

The genus *Enterobacter* comprises common human pathogens, with the *Enterobacter cloacae* complex (ECC) species (*Enterobacter cloacae* subsp. *cloacae*, *Enterobacter asburiae*, *Enterobacter kobei*, *Enterobacter hormaechei*, *Enterobacter ludwigii* among others) being notoriously found in hospitals, causing a wide range of infections such as lower respiratory tract infections, urinary tract infections, bacteremia, and meningitis [1–3]. The clinical significance of the ECC strains is aided by the fact that most of the strains carry multiple antibiotic resistance genes [4,5]. Although the ECC strains are of clinical significance, global surveillance and characterization of the virulence factors associated with these species are still limited [6].

Precise species and subspecies assignment of bacterial isolates lays a foundation for understanding the epidemiology, pathogenesis, and microbiological features of bacteria, allowing for better global surveillance and comparative genomics. Taxonomic classification of the genus *Enterobacter* is complicated; recently, Wu et al. (2020) updated the taxonomy of *Enterobacter* species and subspecies based on whole-genome taxonomic analysis [7]. The complexity resulted from the fact that classification in the genus *Enterobacter* had been defined based on low-resolution analytical methods, and is therefore in need of careful re-examination [8–10]. As shown with the current severe acute respiratory syndrome

coronavirus 2 (SARS-CoV-2), proper identification of strains and variants is critical, as they have important implications for diagnosis, treatment, prognosis, and prevention. Whole-genome sequences have greatly improved the identification of species, as they elucidate functional profiles of taxonomic groups and easily resolve ambiguities in the phylogeny of higher taxa, which would have been difficult through traditional approaches [11,12].

In the current study, we present whole-genome-based taxonomic classification and comparative genomic analysis of potential human pathogenic *E. hormaechei* subsp. *hoffmannii* (designated as isolates S4 and S5) isolated from chlorinated wastewater. This study aids with the global baseline surveillance of ESKAPE pathogens and provides useful resources (genome sequences), which further helps understanding of the global dynamics of an important bacterial species, *Enterobacter*. More importantly, the taxonomic classification analysis from this study supports a recent report by Wu et al. (2020) that *E. cloacae* subsp. *dissolvens* and *E. hormaechei* subsp. *hoffmannii* are species rather than subspecies; *Enterobacter xiangfangensis*, *E. hormaechei* subsp. *oharae*, and *E. hormaechei* subsp. *steigerwaltii* are the same species. Thus, it is important to apply whole-genome-based taxonomic classification, particularly when dealing with isolates of clinical relevance, since their proper identification is critical for proper diagnosis, treatment, prognosis, and prevention.

2. Materials and Methods

2.1. Sample Isolation, Whole-Genome Sequencing, and Analysis of Genomic Features

Samples were isolated from tertiary wastewater effluent, collected from a wastewater treatment plant that uses (5 mg/L) chlorine for treatment in the North West Province, South Africa, as part of our laboratory routine ESKAPE pathogens monitoring program. An API 20E micro-organism identification kit (bioMerieux, Midrand, South Africa), and whole-genome sequencing were used to ascertain the identity of the isolates. Bacterial DNA was extracted from overnight broth cultures using a Nucleospin[®] Tissue extraction kit (Macherey-Nagel, Düren, Germany) following the manufacturer's protocol. Agarose gel electrophoresis and NanoDrop spectrophotometry (ND-100, NanoDrop Technologies Inc, Wilmington, DE, USA) were used to determine the integrity and the purity of the resultant DNA, respectively. Paired-end libraries were prepared from 1 ng bacterial DNA using a Nextera XT DNA Sample Preparation Kit and a Nextera XT Index kit (Illumina Inc., San Diego, California, USA) following the manufacturer's protocol. Sequencing was performed using a MiSeq 2000 Illumina platform (250 bp paired-end reads). Raw sequence reads were quality filtered using FastQC; trimming was performed using Trimmomatic [13]. De novo assembling of the quality reads was performed using SPAdes version 3.13.0 [14]. CheckM [15] was used to assess the quality of the SPAdes assembled genomes. SPAdes assembled genomes were annotated by RAST [16] with default settings.

2.2. Whole Genome-Based Taxonomic Analysis

For whole-genome-based taxonomic analysis, the SPAdes-assembled genomes were uploaded to the Type Strain Genome Server (TYGS) [17]. Pairwise comparison of the user-uploaded genomes (isolates S4 and S5) and the phylogenetically related type strains were performed using Genome Blast Distance Phylogeny (GBDP). Inter-genomic distances were inferred using the trimming algorithm and distance formula d_5 , with 100 replicates [18]. Digital DNA to DNA hybridization (dDDH) values and confidence intervals were calculated using the recommended settings of the Genome-to-Genome Distance Calculator (GGDC) 2.1. The resulting inter-genomic distances were used to infer a balanced minimum evolution tree with branch support via FASTME 2.1.4, including SPR post-processing [19]. The trees were rooted at the midpoint and visualized with PhyD3 [20]. FastANI version 0.1.2 [21] was used to estimate the average nucleotide identity (ANI) of the isolates using closely related genomes of *Enterobacter* species as reference genomes.

2.3. Comparative Genomics

RAST annotated whole-genome sequences of species closely related to the isolates (S4 and S5) were uploaded onto the Kbase server for downstream analysis [22]. The used species were *E. asburiae* (CP011863), *Enterobacter bugandensis* (GCA_015137655), *Enterobacter cancerogenus* (GCA_000478345), *Enterobacter chengduensis* (CP043318), *Enterobacter chuandaensis* (GCA_003594915), *E. cloacae* subsp. *dissolvens* (NC_018079), *E. cloacae* subsp. *cloacae* (CP001918), *E. hormaechei* subsp. *hormaechei* (GCA_001875655), *E. hormaechei* subsp. *steigerwaltii* (CP017179), *E. hormaechei* subsp. *hoffmannii* (CP017186), *Enterobacter huaxiensis* (RWHU01000000), *E. kobei* (CP017181), *E. ludwigii* (CP017279.1), *Enterobacter mori* (GCA_000211415), *Enterobacter oligotrophica* (AP019007), *Enterobacter quasihormaechei* (GCA_004331385), *Enterobacter roggkampii* (CP017184), *Enterobacter sichuanensis* (GCA_002939185), *Enterobacter soli* (GCA_001654845), *Enterobacter timonensis* (GCA_90002-1175), *Enterobacter wuhouensis* (GCA_004331265) and *Enterobacter xiangfangensis* (CP017183). Domains in the genome sets were annotated using HMMER version 3.1 b [23], Gapped BLAST and PSI-BLAST [24]. The annotated domains were used to view the difference in functional roles among the genomes based on The SEED and COG (<https://www.ncbi.nlm.nih.gov/research/cog-project/>, accessed on 25 January 2021) categories/roles.

3. Results and Discussions

3.1. Genomic Features and Assembling Metrics of the Genomes of the Isolates

The genomes assembled with SPAdes resulted in high-quality genomes, as shown in Table 1. CheckM quality assessment of the assembled genomes showed that the genomes of isolates S4 and S5 had an estimated 97.43% and 94.67% completeness, respectively, and had 0.08% predicted contamination. Isolate S4 had a genome size of 4,658,088 bp, which corresponds to an N50 of 89,616 bp, having 71 contigs. Isolate S5 had a genome size of 4,442,534 bp, which corresponds to an N50 of 39,615 bp, having 132 contigs. The GC content was homogeneous with S4 and S5 having a GC content of 55.26% and 55.35%, respectively. RAST annotation of the genomes predicted that isolate S4 had 4545 coding genes while isolate S5 had 4341 coding genes. The genes for S4 were predicted to have 3997 distinct functions, with 1991 of the genes having SEED annotation ontology across 1806 distinct SEED functions. For S5, there were 3651 predicted distinct functions, with 1930 of the genes having SEED annotation ontology across 1751 distinct SEED functions.

Table 1. Quality metrics of the isolates assembled genomes.

Isolate	S4	S5
Total Contigs	71	132
Total Length (bp)	4,658,088	4,442,534
Largest Contig (bp)	329,202	135,452
N50 (bp)	89,616	39,615
GC%	55.26	55.35
Coding Genes	4545	4341
Non Coding Repeats	16	10
Non Coding RNA	16	56

3.2. Whole-Genome-Based Taxonomic Analysis

Taxonomic classification in the genus *Enterobacter* has been problematic, since some species and most subspecies have been assigned based on low-resolution methods [8,9]. In addition, this complexity has been made worse by the fact that phenotype-based methods cause misidentification of *Enterobacter* species and are unreliable for precise species identification [6]. To correctly classify our isolates (S4 and S5), we applied a more robust whole-genome-based taxonomic classification through comparison of the whole-genome sequences of our isolates with the whole-genome sequences of established type strains; for this purpose, we used the Type Strain Genome Server (TYGS) [17]. The TYGS results showed that isolates S4 and S5 form a highly supported clade corresponding to a

species cluster with *Enterobacter hormaechei* subsp. *hoffmannii* (Figure 1). The main criterion for species affiliation is the 70% DNA–DNA hybridization threshold [25,26]. Digital DNA–DNA hybridization (dDDH) computed by the Genome–Genome Distance Calculator (GGDC) between the whole genomes of our isolates (S4 and S5) and that of *E. hormaechei* subsp. *hoffmannii* was well above the species cut-off point of 70%: 92.8% and 92.9%, respectively (Table 2), indicating that isolates S4 and S5 are indeed *E. hormaechei* subsp. *hoffmannii*. The GGDC results showed low delta values coupled with high average branch support, which denotes a high phylogenetic accuracy [27,28]. Another overall genome relatedness index important for the delineation of species based upon genome similarity is average nucleotide identity (ANI). This is one of the most important indices, since sequence identity has clockwise-like properties and provides the likelihood of correlation with times of divergence [29]. ANI is widely acknowledged as a measure of genomic relatedness, with an ANI $\geq 96\%$ being the proposed cut-off point for species delineation [30]. In the present study, the ANI of the isolates S4 and S5 were well above the cut-off point compared to the closely phylogenetically related species *E. hormaechei* subsp. *hoffmannii* (Supplementary Table S1), indicating that the isolates used in this study are *E. hormaechei* subsp. *hoffmannii*.

A

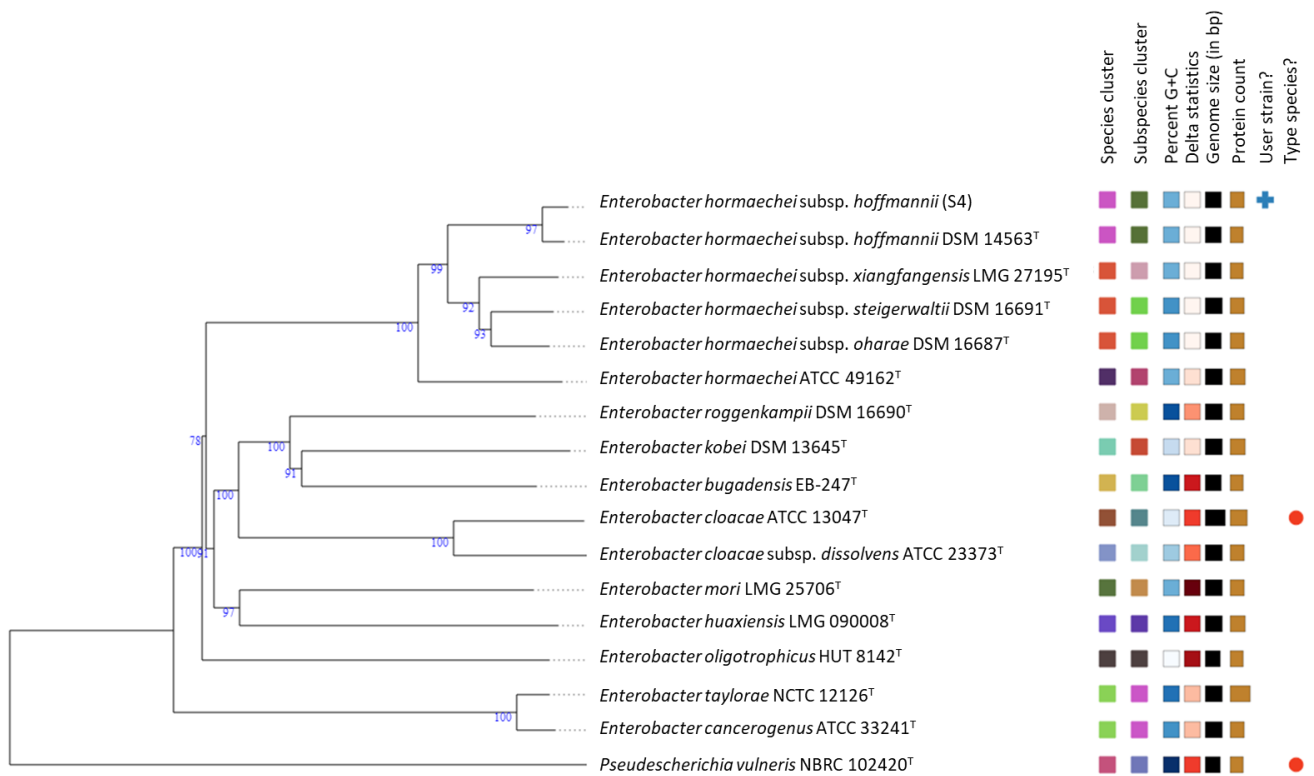


Figure 1. Cont.

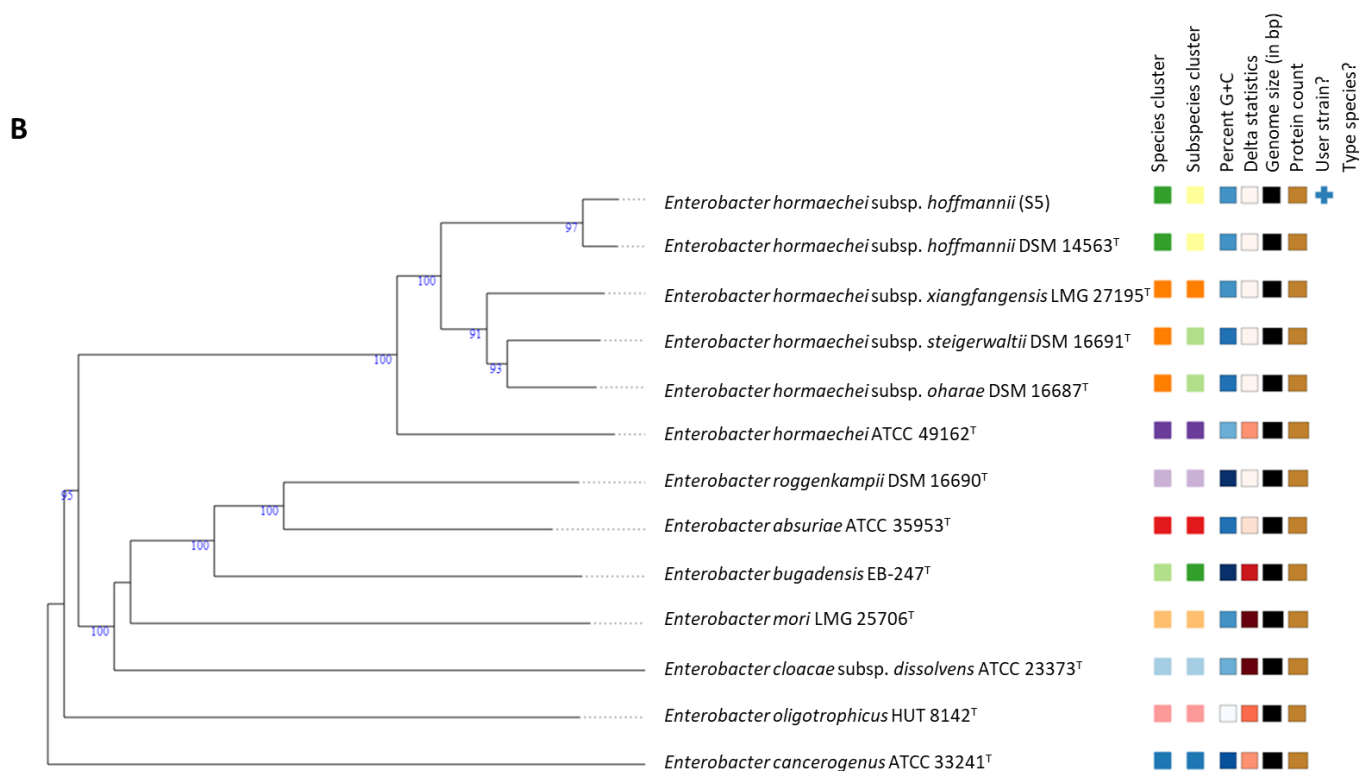


Figure 1. Trees inferred with FastME 2.1.6.1 from the Genome Blast Distance Phylogeny (GBDP) distances calculated from whole-genome sequences. Isolates S4 (A) and S5 (B) form a species cluster with *Enterobacter hormaechei* subsp. *hoffmannii*, having a dDDH of >70%, a clear indication that the isolates belong to the same species as *E. hormaechei* subsp. *hoffmannii*.

Table 2. Type Strain Genome Server calculation of digital DNA–DNA hybridizations of isolates (S4, S5), and some of their closely related type strains.

Query Strain	Subject Strain	dDDH (d4, in %)	C.I. (d4, in %)	G + C Content Difference (in %)
S4SPAdes.contigs.fa	<i>Enterobacter hormaechei</i> subsp. <i>hoffmannii</i> DSM 14563 ^T	92.9	[90.9–94.4]	0.07
S5SPAdes.contigs.fa	<i>Enterobacter hormaechei</i> subsp. <i>hoffmannii</i> DSM 14563 ^T	92.8	[90.8–94.4]	0.01
S4SPAdes.contigs.fa	<i>Enterobacter hormaechei</i> subsp. <i>oharae</i> DSM 16687 ^T	66.	[63.4–69.2]	0.32
S5SPAdes.contigs.fa	<i>Enterobacter hormaechei</i> subsp. <i>oharae</i> DSM 16687 ^T	66.4	[63.5–69.3]	0.24
S4SPAdes.contigs.fa	<i>Enterobacter hormaechei</i> subsp. <i>xiangfangensis</i> LMG 27195 ^T	66.1	[63.2–69.0]	0.02
S5SPAdes.contigs.fa	<i>Enterobacter hormaechei</i> subsp. <i>xiangfangensis</i> LMG 27195 ^T	66.2	[63.2–69.0]	0.07
S4SPAdes.contigs.fa	<i>Enterobacter hormaechei</i> subsp. <i>steigerwaltii</i> DSM 16691 ^T	66	[63.0–68.8]	0.29
S5SPAdes.contigs.fa	<i>Enterobacter hormaechei</i> subsp. <i>steigerwaltii</i> DSM 16691 ^T	65.9	[62.9–68.7]	0.21
S4SPAdes.contigs.fa	<i>Enterobacter hormaechei</i> ATCC 49162 ^T	57.3	[54.5–60.0]	0.02
S5SPAdes.contigs.fa	<i>Enterobacter hormaechei</i> ATCC 49162 ^T	57.3	[54.5–60.1]	0.1
S4SPAdes.contigs.fa	<i>Enterobacter bugandensis</i> EB-247 ^T	35.1	[32.7–37.6]	0.74
S5SPAdes.contigs.fa	<i>Enterobacter bugandensis</i> EB-247 ^T	35.3	[32.8–37.8]	0.65
S4SPAdes.contigs.fa	<i>Enterobacter mori</i> LMG 25706 ^T	34.2	[31.7–36.7]	0.04
S5SPAdes.contigs.fa	<i>Enterobacter mori</i> LMG 25706 ^T	34.3	[31.9–36.8]	0.05
S4SPAdes.contigs.fa	<i>Enterobacter roggkampii</i> DSM 16690 ^T	33.7	[31.3–36.3]	0.78
S5SPAdes.contigs.fa	<i>Enterobacter roggkampii</i> DSM 16690 ^T	34	[31.5–36.5]	0.7

Table 2. Cont.

Query Strain	Subject Strain	dDDH (d4, in %)	C.I. (d4, in %)	G + C Content Difference (in %)
S4SPAdes.contigs.fa	<i>Enterobacter oligotrophicus</i> HUT 8142 ^T	33.4	[31.0–35.9]	0.96
S5SPAdes.contigs.fa	<i>Enterobacter oligotrophicus</i> HUT 8142 ^T	33.5	[31.1–36.0]	1.04
S4SPAdes.contigs.fa	<i>Enterobacter cloacae</i> subsp. <i>dissolvens</i> ATCC 23373 ^T	32	[29.6–34.5]	0.1
S5SPAdes.contigs.fa	<i>Enterobacter cloacae</i> subsp. <i>dissolvens</i> ATCC 23373 ^T	32.1	[29.7–34.6]	0.18
S4SPAdes.contigs.fa	<i>Enterobacter cancerogenus</i> ATCC 33241 ^T	31.3	[28.9–33.8]	0.42
S5SPAdes.contigs.fa	<i>Enterobacter cancerogenus</i> ATCC 33241 ^T	31.4	[29.0–34.0]	0.34

Based on ANI and dDDH, the taxonomic classification analysis from this study supports the recent report that *E. cloacae* subsp. *dissolvens* and *E. hormaechei* subsp. *hoffmannii* are species rather than subspecies; and *E. xiangfangensis*, *E. hormaechei* subsp. *oharae*, and *E. hormaechei* subsp. *steigerwaltii* are the same species (Supplementary Table S1). Based on the recently updated taxonomic classification of the genus *Enterobacter* by Wu et al. (2020), isolates S4 and S5 are *E. hoffmannii* [7]. Whole-genome sequences have greatly improved the identification of species as they elucidate functional profiles of taxonomic groups, and easily resolve ambiguities in the phylogeny of higher taxa, which would have been difficult through traditional approaches [11,12]. Low-resolution methods were applied in the earlier classification of *Enterobacter* species [8–10], which caused the above-mentioned inconsistency; more robust methods should be applied when classifying *Enterobacter* species.

3.3. Predictions of Potential Human Pathogenicity, Resistome, Virulome, and Biosynthetic Gene Clusters

Identifying pathogenic bacterial strains and understanding the biological mechanisms of pathogenicity is important for timely intervention programs, designing control strategies, as well as the development of targeted vaccines. To predict the pathogenicity of our isolates, we used the PathogenFinder algorithm [31]; both S4 and S5 were predicted to be human pathogens with a probability of greater than 0.78. The genome sequence of isolate S4 matched a total of 63 pathogenic families from a broad range of pathogens, including *Yersinia pestis*, *Salmonella enterica* subsp. *arizonae*, several pathogenic *Escherichia coli* strains, *Klebsiella pneumoniae* strains, *Citrobacter koseri*, *Shigella boydii*, and *Enterobacter sakazakii*. Supplementary Materials Supporting Information 1a provides a detailed description of the matched families, which includes location (chromosomal or plasmid), completeness, and identity percentage for S4. Isolate S5 matched a total of 58 pathogenic families; Supplementary Materials Supporting Information 1b gives a detailed description of the matched families.

For effective colonization of the host, pathogens require an arsenal of strategies to enable adherence, persistence, aggression, and evasion of innate and adaptive immunity [32–34]. We used the Virulence Factors of Pathogenic bacteria database to determine the presence of genes encoding the above-described arsenal. Both isolates S4 and S5 have the necessary virulence factors for successful host colonization and perpetuation of the pathogenesis process, supporting their potential pathogenicity (Table 3). In addition to the virulence factors, the isolates have several antibiotic resistance genes, which included genes encoding for beta-lactamase resistance. The mechanisms for antibiotic resistance were predicted to be antibiotics inactivation, antibiotic target alteration, reduced permeability to antibiotic, and antibiotic efflux (Supplementary Table S2). The presence of all the arsenal required for pathogenesis and the ability of the isolates to survive the chlorination process emphasize the importance of continuously monitoring wastewaters and evaluating the wastewater treatment process. A number of biosynthetic gene clusters have been shown to offer competitive fitness advantages and aid virulence. For example, aryl polyenes (APes) were shown to function as fitness factors that increase protection from oxidative stress and contribute to biofilm formation [35]. In *E. coli*, APes are present in most pathogenic

strains, whereas they are typically absent from commensals and laboratory strains [36]. We used antiSMASH to identify the secondary metabolite biosynthetic gene clusters in our genome sequences [37]. Both S4 and S5 consist of aryl polyenes (Figure 2A,B); this could possibly explain their ability to survive the chlorination treatment process of the wastewater. We also identified aerobactin (Figure 2C,D), augmenting the predicted human pathogenic nature of the isolates. Lack of available iron is one of the first lines of defense to avoid bacterial infection [38–40]. Aerobactin, a hydroxymate-type siderophore, has been shown to offer a selective advantage under iron starvation and increases the virulence of *E. coli* strains [41,42].

Table 3. Virulence factors detected in the genome sequences of isolates S4 and S5.

Virulence Factor Class	Virulence Factors	S4	S5
Adherence	Type 3 fimbriae	1	1
	Type I fimbriae	8	6
	Curli fibers(<i>Escherichia</i>)	3	2
	Hemorrhagic <i>E. coli</i> pilus (HCP) (<i>Escherichia</i>)	2	2
	Type IV (pili) (<i>Yersinia</i>)	1	1
Antiphagocytosis	Capsule	15	15
Efflux pump	AcrAB	3	3
Iron uptake	Aerobactin	5	5
	Ent Siderophore	12	12
	Salmochelins	1	1
	Heme transport (<i>Shigella</i>)	1	1
	Heme uptake (<i>Escherichia</i>)	2	2
	Iron/Manganese transport (<i>Escherichia</i>)	1	1
Regulation	RcsAB	2	1
Secretion system	T6SS-I	10	9
	T6SS-II	4	4
	T6SS-III	0	2
	EPS type secretion system (<i>Vibrio</i>)	1	1
	Flagella (cluster I) (<i>Yersinia</i>)	32	32
	Hcp secretion island-1 encoded type VI secretion system (H-T6SS) (<i>Pseudomonas</i>)	3	3
	SCI-I T6SS (<i>Escherichia</i>)	1	0
	T2SS (Yst1) (<i>Yersinia</i>)	1	1
	T2SS (<i>Aeromonas</i>)	1	1
	TTSS (SPI-encode) (<i>Salmonella</i>)	1	0
	Serum resistance	LPS rfb locus	1
Toxin	Heat-stable cytotoxic toxin	1	1
Autotransporter	EhaB	1	
Endotoxin	LOS (<i>Haemophilus</i>)	1	1
Fimbrial adherence determinants	Agf/Csg (<i>Salmonella</i>)	4	4
	Fim (<i>Salmonella</i>)	5	5
	Sti (<i>Salmonella</i>)	1	1
	Stj (<i>Salmonella</i>)	1	0
	Stk (<i>Salmonella</i>)	4	4
Invasion	Flagella (<i>Burkholderia</i>)	6	6
Motility	Flagella (<i>Bordetella</i>)	1	1

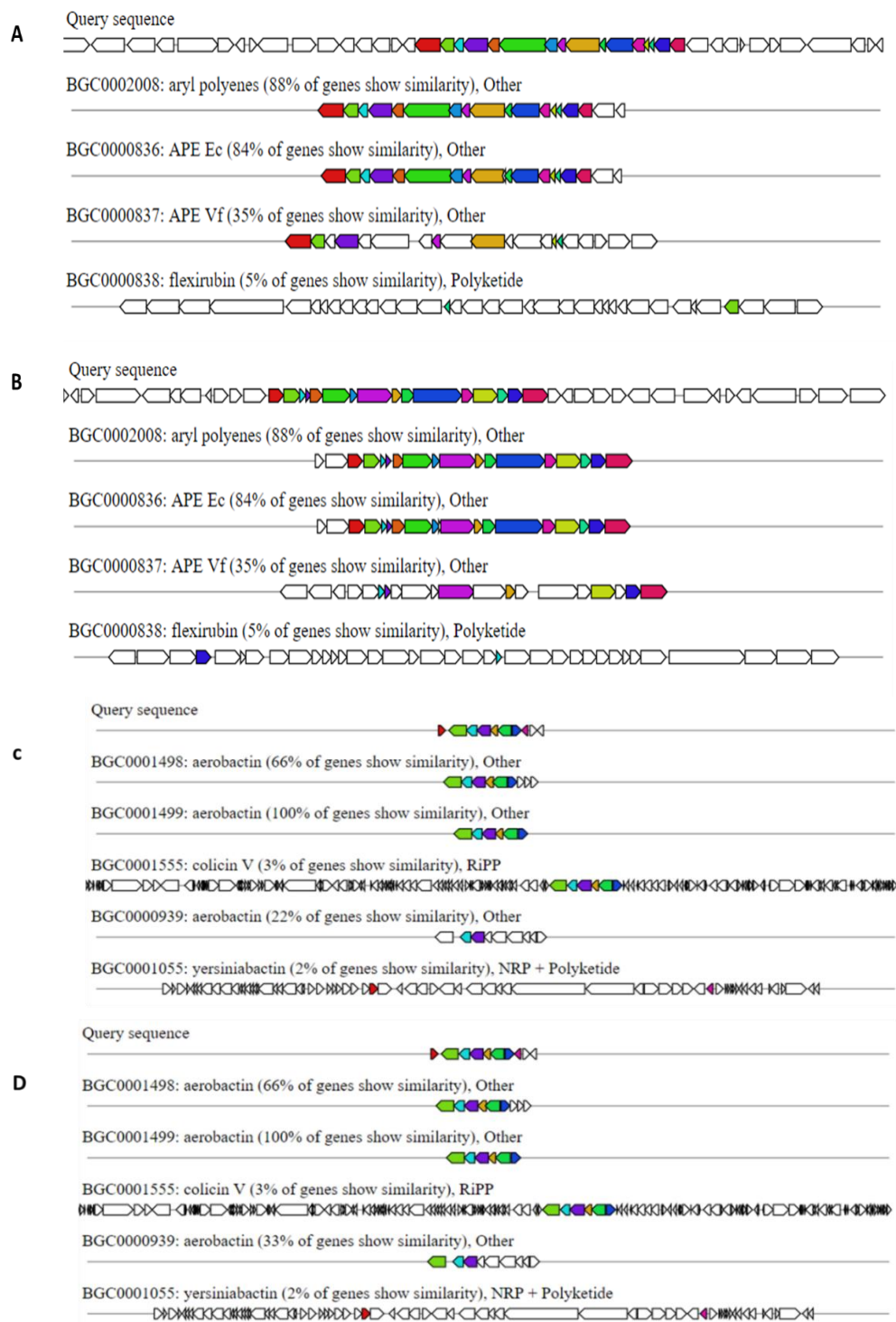


Figure 2. Biosynthetic gene clusters (BGCs) detected in the genome sequences of isolates S4 and S5. (A,B) Aryl polyenes (APEs) for genomes S4 and S5, respectively. (C,D) Aerobactin for genomes S4 and S5, respectively. Query sequences are the genome sequences of the isolates S4 and S5.

3.4. Comparative Genomics

The development of next-generation sequencing technologies (NGS) has allowed for comparative genomics of multiple genomes. Comparative genomics, in particular pan-genome analysis, more accurately reflects the notion of bacterial species [11,43]. Pan-genome refers to the whole gene repertoire of a study group [44,45]. The pan-genome is divided into three components: (1) the core genome, which is a set of all genes common to

all strains of the study; (2) non-core/accessory genome, which is a set of genes present in more than one, but not in all of the strains used in a study; and (3) singletons, which are genes unique to individual strains used in the study.

From the 24 *Enterobacter* genomes used in this study, there were a total of 6032 functions detected, of which 40.4% (2438) were core functions. The remaining 59.6% were part of the accessory genome. The core function genes covered the essential basics for survival; the clusters of orthologous genes (COGs) for the core functions are shown in Figure 3. They are homogeneously distributed, covering cellular processes and signaling, metabolism, and information storage and processing. Isolates S4 and S5 are more identical; thus, including all isolates in the pan-genome analysis masked how distinct the isolates are from other species (Figure 4A). However, removing one of the isolates (S4) revealed that the isolates are greatly distinct from other samples used in this study (Figure 4B), having a total of 234 singletons. Supplementary Materials Supporting Information 2 provides details of all the singletons for isolate S5 compared to all the other samples, excluding isolate S4. A large fraction of the accessory genomes is mobile genetic elements [45]. These accessory genomes place a host cell in an advantageous position to be viable under specific conditions [46,47]. The singletons comprise many virulence factors, antibiotic resistance genes, phage related proteins, with hypothetical proteins and genes of unknown function constituting 54% of the singletons. Of the singletons, some of the genes to note are (phage-encoded virulence determinant Bor, several phage related proteins/genes, type IV secretion complex proteins (VirB5, VirB6, VirB9, VirB1, VirD4, VirB2, VirB8, VirB10, VirB11), T6SS secretion lipoprotein TssJ (VasD), T1SS secreted agglutinin RTX, nitrilotriacetate monooxygenase component A (EC 1.14.13.-), multidrug resistance protein MdtH, SOS-response repressor and protease LexA (EC 3.4.21.88), inner-membrane proton/drug antiporter (MSF type) of tripartite multidrug efflux system, arsenite/antimonite pump-driving ATPase ArsA (EC 3.6.3.16), ferric hydroxamate outer membrane receptor FhuA, diguanylate cyclase/phosphodiesterase (GGDEF and EAL domains) with PAS/PAC sensor(s), and COG3121: P pilus assembly protein, chaperone PapD), a clear indication of the virulence nature of the isolates (Supplementary Materials Supporting Information 2). Expression of *bor* significantly increases the survival of the *Escherichia coli* host cell in animal serum. This property is a well-known bacterial virulence determinant [48,49]. Prophages are known to increase the virulence of pathogenic strains [50,51], and there is a positive correlation between the phage-related DNA content of a given Enterobacterium and its pathogenicity [52]. In this study, the singletons for S5 constitute a higher proportion of phage DNA, and the isolates were predicted to be human pathogens. Type IV secretion systems mediate the translocation of virulence factors (proteins and/or DNA) from Gram-negative bacteria into eukaryotic cells [53,54].

Comparison of protein domain content allows for the construction of bacteria phylogeny independent of gene sequence [55], which is a better indicator of shared physiology and ecology [56]. This also takes into account the impacts of mutations, gene loss, and horizontal gene transfer. The domain architecture is preserved at large phylogenetic distances [57,58]. Protein domains are distinct, functional units responsible for specific functions and interactions; thus, we also applied the SEED protein domain-class content to study the species diversity. Under the SEED category virulence, subcategory Type_4_secretion_and_conjugative_transfer, isolates S4 and S5 have a higher proportion of domains related to this function (Figure 5). These are also relatively higher in *E. coli*, *E. quasihormoachei*, *E. huaxiansis*, *E. chuandensis*, and *E. cancerogenus* compared to other genomes, albeit lower than in the isolates S4 and S5. Moreover, S4 and S5 have a considerably higher level of phage_capsid_protein, with S5 being the only genome harboring Tn552. In addition, S5 shows a relatively higher level of domains related to copper homeostasis (Figure 5).

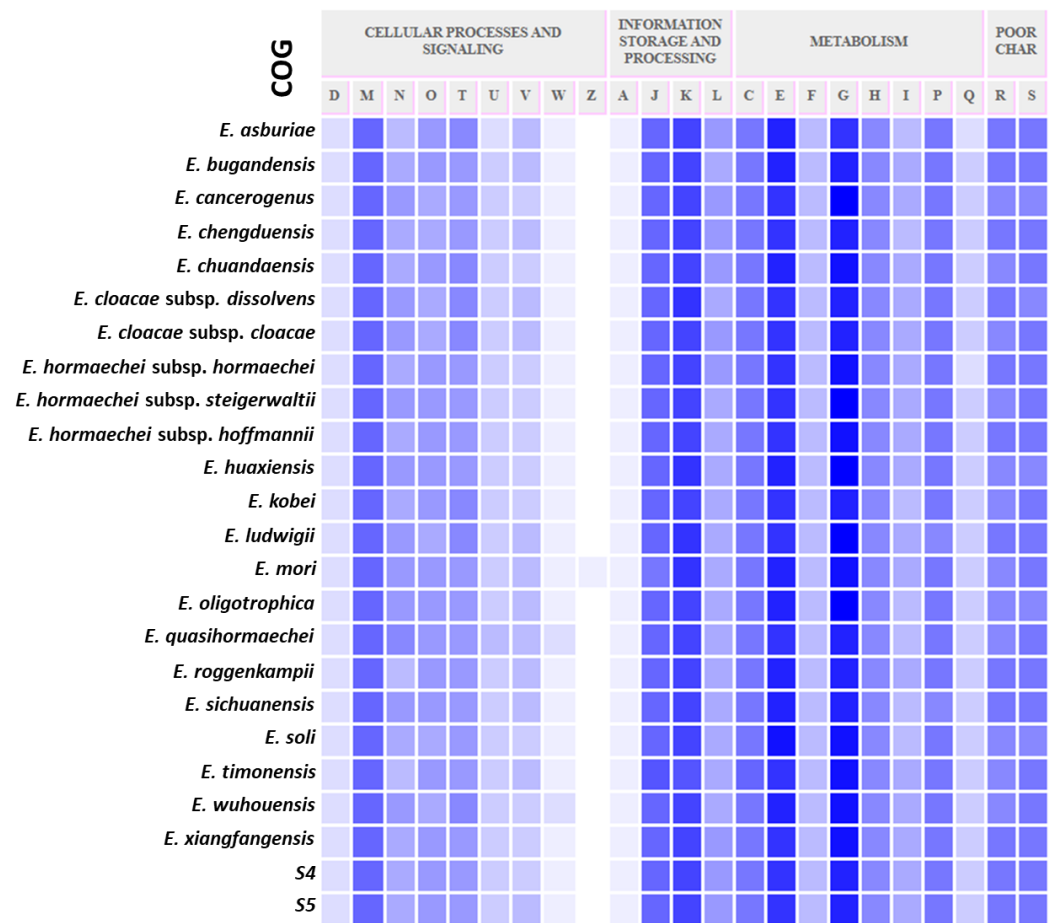


Figure 3. A heatmap showing clusters of orthologous genes (COGs) for isolates S4, S5, and a representative of all validly published *Enterobacter* species (see Materials and Methods for full species list and accession numbers). The COGs covered cellular processing and signaling (D: cell cycle control, cell division, chromosome partitioning; M: cell wall/membrane/envelope biogenesis; N: cell motility; O: post-translational modification, protein turnover, and chaperones; T: signal transduction mechanism; U: intracellular trafficking, secretion, and vesicular transportation; V: defense mechanisms; W: extracellular structures; Z: cytoskeleton), information storage and processing (A: RNA processing and modification; J: translation, ribosomal structure, and biogenesis; K: transportation; L: replication, recombination, and repair), metabolism (C: energy production and conversion; 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; P: inorganic ion transport and metabolism; Q: secondary metabolites biosynthesis, transport, and catabolism), as well as the poorly characterized (R and S).

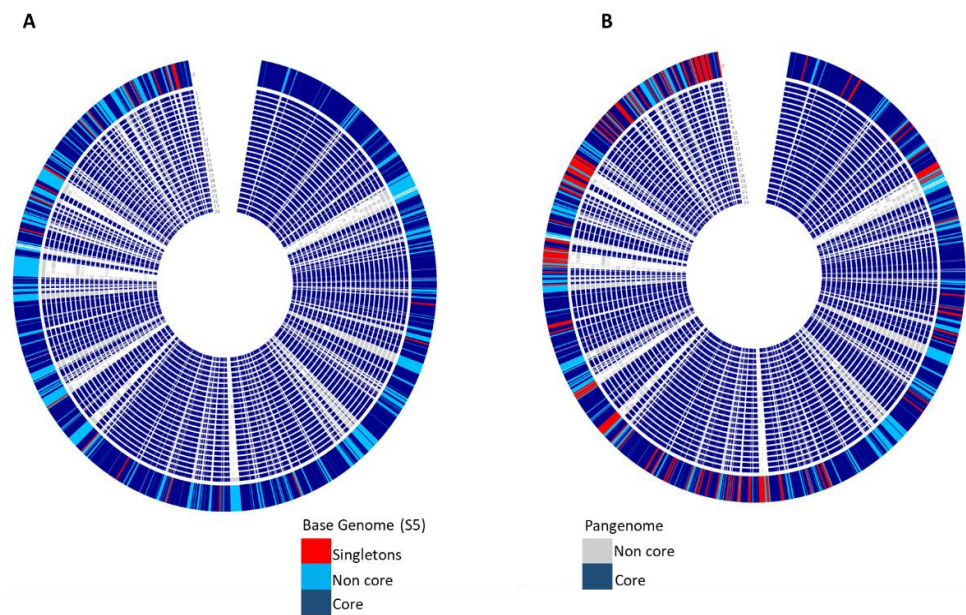


Figure 4. Pan-genome circular visualization output. **(A)** Isolates S4 and S5 compared to representatives of all validly published *Enterobacter* species (see Materials and Methods for full species list and accession numbers); each red strip indicates a gene that is unique to the isolate S5, which was used as a base genome. **(B)** Isolate S5 compared to validly published *Enterobacter* species; each red strip indicates a gene that is unique to the isolate S5. **(B)** Isolate S4 is not included in the analysis as it is highly identical to isolate S5 and would mask the true variation in the isolates compared to other *Enterobacter* species.

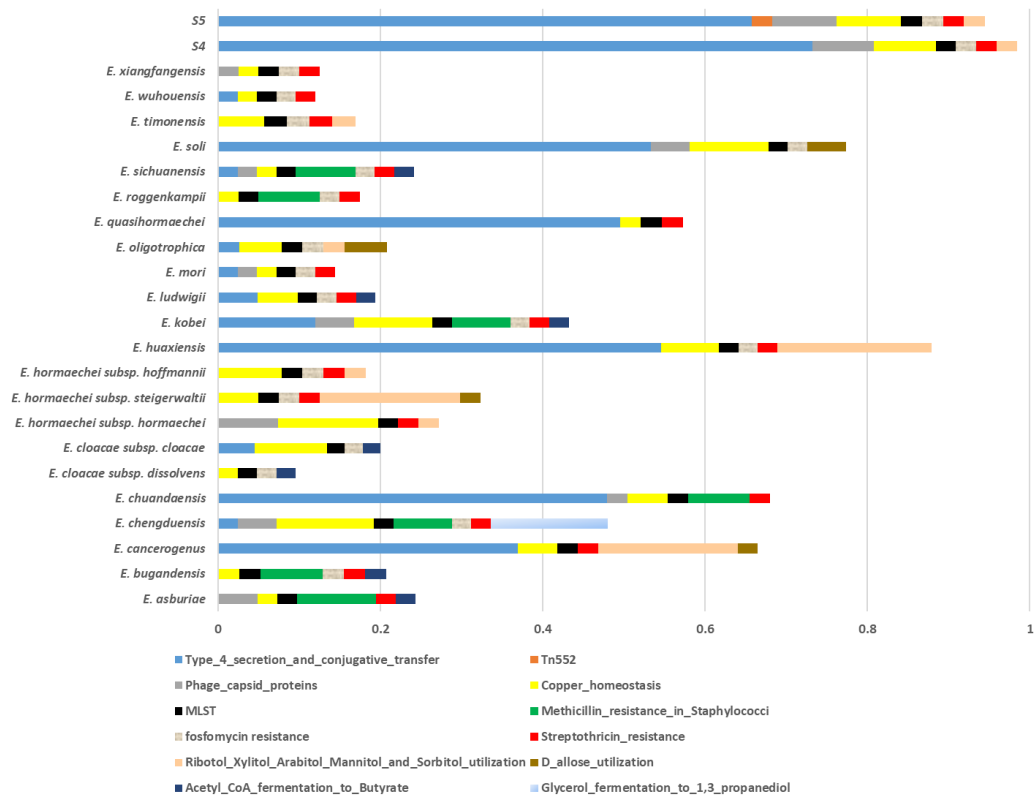


Figure 5. Distribution of a selected SEED domain categories between the isolates (S4, S5) and the representatives of all validly published *Enterobacter* species (see Materials and Methods for full species list and accession numbers). Interestingly, isolates S4 and S5 have the largest proportion of Type 4 Secretion System in addition to being the only genomes carrying Tn522.

4. Conclusions

This study provides comprehensive insights into the genomic structure of *E. hormaechi* subsp. *hoffmannii* from South Africa, and aids with the baseline of global surveillance and monitoring of *Enterobacter* species, a member of the clinically significant ESKAPE pathogens. From the genome sequences, several important features including virulence factors, antibiotic resistance genes, and biosynthetic gene clusters (BGCs) were explored. The genome sequences provide an important resource for future comparative genomics. Moreover, the study demonstrates the risks associated with wastewater and the need for continuous monitoring of the treated wastewaters to ensure efficacy in the removal of pathogens. It also helps local authorities to understand the potential dangers associated with the wastewater, and encourage them to reconsider their treatment processes. The results from whole-genome taxonomic classification support recent updates in the taxonomy of the genus *Enterobacter*, emphasizing the importance of using high-resolution analytic methods for species identification.

Supplementary Materials: The supplementary materials are available online at <https://www.mdpi.com/article/10.3390/microorganisms9091928/s1>.

Author Contributions: Conceptualization, T.E.M. and C.C.B.; data curation, T.E.M.; formal analysis, T.E.M.; funding acquisition, C.C.B.; investigation, C.C.B.; methodology, T.E.M.; project administration, C.C.B.; resources, C.C.B.; software, T.E.M.; supervision, C.C.B.; validation, T.E.M.; visualization, T.E.M.; writing—original draft, T.E.M.; writing—review and editing, T.E.M. and C.C.B. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge with thanks the financial support from the North West University, South Africa; and the Water Research Commission (WRC), South Africa (project 2019/2020-00224). The views expressed are those of the authors and not of the funders.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw sequences generated for this study were deposited into the National Centre for Biotechnology Information (NCBI) database under the BioProject ID: PRJNA735620, with BioSample Accessions SAMN19590289, SAMN19590290. All the other data supporting the findings of this study are available in this published article and its supplementary information files.

Conflicts of Interest: The authors declare that they have no competing interest.

References

1. Keller, R.; Pedroso, M.Z.; Ritchmann, R.; Silva, R.M. Occurrence of virulence-associated properties in *Enterobacter cloacae*. *Infect. Immun.* **1998**, *66*, 645–649. [[CrossRef](#)] [[PubMed](#)]
2. Streit, J.M.; Jones, R.N.; Sader, H.S.; Fritsche, T.R. Assessment of pathogen occurrences and resistance profiles among infected patients in the intensive care unit: Report from the sentry Antimicrobial Surveillance Program (North America, 2001). *Int. J. Antimicrob. Agents* **2004**, *24*, 111–118. [[CrossRef](#)] [[PubMed](#)]
3. Pauw, A.; Caspers, M.P.M.; Schuren, F.H.J.; Leverstein-van Hall, M.A.; Deletoile, A.; Montijn, R.; Verhoef, J.; Fluit, A.C. Genomic diversity within the *Enterobacter cloacae* complex. *PLoS ONE* **2008**, *3*, e3018. [[CrossRef](#)] [[PubMed](#)]
4. Leverstein-van Hall, M.A.; Blok, H.E.; Pauw, A.; Fluit, A.C.; Troelstra, A.; Mascini, E.M.; Bonten, M.J.M.; Verhoef, J. Extensive hospital-wide spread of a multidrug-resistant *Enterobacter cloacae* clone, with late detection due to a variable antibiogram and frequent patient transfer. *J. Clin. Microbiol.* **2006**, *44*, 518–524. [[CrossRef](#)]
5. Pauw, A.; Verhoef, J.; Fluit, A.C.; Blok, H.E.; Hopmans, T.E.; Troelstra, A.; Leverstein-van Hall, M.A. Failure to control an outbreak of qnrA1-positive multidrug-resistant *Enterobacter cloacae* infection despite adequate implementation of recommended infection control measures. *J. Clin. Microbiol.* **2007**, *45*, 1420–1425. [[CrossRef](#)]
6. Mezzatesta, M.L.; Gona, F.; Stefani, S. *Enterobacter cloacae* complex: Clinical impact and emerging antibiotic resistance. *Future Microbiol.* **2012**, *7*, 887–902. [[CrossRef](#)]
7. Wu, W.; Feng, Y.; Zong, Z. Precise species identification for *Enterobacter*: A genome sequence-based study with reporting of two novel species *Enterobacter quasiroggenkampii* sp. nov. and *Enterobacter quasimori* sp. nov. *mSystems* **2020**, *5*, e00527–20. [[CrossRef](#)]
8. Hoffmann, H.; Roggenkamp, A. Population genetics of the nomen-species *Enterobacter cloacae*. *Appl. Environ. Microbiol.* **2003**, *69*, 5306–5318. [[CrossRef](#)]

9. Hoffmann, H.; Stindl, S.; Ludwig, W.; Stumpf, A.; Mehlen, A.; Monget, D.; Piérard, D.; Ziesing, S.; Heesemann, J.; Roggenkamp, A.; et al. *Enterobacter hormaechei* subsp. *oharae* subsp. nov., *E. hormaechei* subsp. *hormaechei* comb. nov., and *E. hormaechei* subsp. *steigerwaltii* subsp. nov., three new subspecies of clinical importance. *J. Clin. Microbiol.* **2005**, *43*, 3297–3303. [[CrossRef](#)]
10. Sutton, G.G.; Brinkac, L.M.; Clarke, T.H.; Fouts, D.E. *Enterobacter hormaechei* subsp. *hoffmannii* subsp. nov., *Enterobacter hormaechei* subsp. *xiangfangensis* comb. nov., *Enterobacter roggenkampii* sp. nov., and *Enterobacter muelleri* is a later heterotypic synonym of *Enterobacter asburiae* based on computational analysis of sequenced *Enterobacter* genomes. *F1000Research* **2018**, *7*, 521.
11. Whitman, W.B. Genome sequences as the type material for taxonomic descriptions of prokaryotes. *Syst. Appl. Microbiol.* **2015**, *38*, 217–222. [[CrossRef](#)]
12. Ramasamy, D.; Mishra, A.K.; Lagier, J.C.; Padhmanabhan, R.; Rossi, M.; Sentausa, E.; Raoult, D.; Fournier, P.-E. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 384–391. [[CrossRef](#)] [[PubMed](#)]
13. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
14. Nurk, S.; Bankevich, A.; Antipov, D.; Gurevich, A.A.; Korobeynikov, A.; Lapidus, A.; Prjibelski, A.D.; Pyshkin, A.; Sirotkin, A.; Sirotkin, Y.; et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* **2013**, *20*, 714–737. [[CrossRef](#)] [[PubMed](#)]
15. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [[CrossRef](#)] [[PubMed](#)]
16. Aziz, R.K.; Bartels, D.; Best, A.; DeJongh, M.; Disz, T.; Edwards, R.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST Server: Rapid annotations using subsystems technology. *BMC Genom.* **2008**, *9*, 75. [[CrossRef](#)]
17. 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. [[CrossRef](#)]
18. Meier-Kolthoff, J.P.; Auch, A.F.; Klenk, H.-P.; Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* **2013**, *14*, 60. [[CrossRef](#)]
19. Lefort, V.; Desper, R.; Gascuel, O. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* **2015**, *32*, 2798–2800. [[CrossRef](#)]
20. Kreft, L.; Botzki, A.; Coppens, F.; Vandepoele, K.; Van Bel, M. PhyD3: A phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics* **2017**, *33*, 2946–2947. [[CrossRef](#)] [[PubMed](#)]
21. Jain, C.; Rodriguez, R.L.M.; Phillipy, A.M.; Konstantinidis, K.T.; Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **2017**, *9*, 5114. [[CrossRef](#)] [[PubMed](#)]
22. Arkin, A.P.; Cottingham, R.; Henry, C.S.; Harris, N.L.; Stevens, R.L.; Maslov, S.; Dehal, P.; Ware, D.; Perez, F.; Canon, S.; et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **2018**, *36*, 566–569. [[CrossRef](#)]
23. Eddy, S.R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **2011**, *7*, e1002195. [[CrossRef](#)]
24. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)] [[PubMed](#)]
25. Wayne, L.G.; Moore, W.E.C.; Stackebrandt, E.; Kandler, O.; Colwell, R.R.; Krichevsky, M.I.; Truper, H.G.; Murray, R.G.E.; Grimont, P.A.D.; Brenner, D.J.; et al. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* **1987**, *37*, 463–464. [[CrossRef](#)]
26. Tindall, B.J.; Rossello-Mora, R.; Busse, H.J.; Ludwig, W.; Kämpfer, P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* **2010**, *60*, 249–266. [[CrossRef](#)]
27. Auch, A.F.; Henz, S.R.; Holland, B.R.; Göker, M. Genome BLAST distance phylogenies inferred from whole plastid and whole mitochondrion genome sequences. *BMC Bioinform.* **2006**, *7*, 350. [[CrossRef](#)] [[PubMed](#)]
28. Holland, B.R.; Huber, K.T.; Dress, A.; Moulton, V. δ Plots: A Tool for Analyzing Phylogenetic Distance Data. *Mol. Biol. Evol.* **2002**, *19*, 2051–2059. [[CrossRef](#)] [[PubMed](#)]
29. Kuo, C.H.; Ochman, H. Inferring clocks when lacking rocks: The variable rates of molecular evolution in bacteria. *Biol. Direct* **2009**, *4*, 35. [[CrossRef](#)]
30. Konstantinidis, K.T.; Tiedje, J.M. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2567–2572. [[CrossRef](#)]
31. Cosentino, S.; Voldby Larsen, M.; Møller Aarestrup, F.; Lund, O. PathogenFinder—Distinguishing Friend from Foe Using Bacterial Whole Genome Sequence Data. *PLoS ONE* **2013**, *8*, e77302. [[CrossRef](#)]
32. Falkow, S. What is a pathogen. *ASM News* **1997**, *63*, 359–365.
33. Galán, J.E.; Collmer, A. Type III secretion machines: Bacterial devices for protein delivery into host cells. *Science* **1999**, *284*, 1322–1328. [[PubMed](#)]
34. Hacker, J.; Kaper, J.B. Pathogenicity islands and the evolution of microbes. *Ann. Rev. Microbiol.* **2000**, *54*, 641–679. [[CrossRef](#)]
35. Johnston, I.; Osborn, L.J.; Markley, R.L.; McManus, E.A.; Kadam, A.; Schultz, K.B.; Nagajothi, N.; Ahern, P.P.; Brown, J.M.; Claesen, J. Identification of essential genes for *Escherichia coli* aryl polyene biosynthesis and function in biofilm formation. *bioRxiv* **2020**. [[CrossRef](#)]

36. Cimermancic, P.; Medema, M.H.; Claesen, J.; Kurita, K.; Brown, L.C.W.; Mavrommatis, K.; Pati, A.; Godfrey, P.A.; Koehrsen, M.; Clardy, J.; et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **2014**, *158*, 412–421. [[CrossRef](#)] [[PubMed](#)]
37. Medema, M.H.; Blin, K.; Cimermancic, P.; De Jager, V.; Zakrzewski, P.; Fischbach, M.A.; Weber, T.; Takano, E.; Breitling, R. antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **2011**, *39*, W339–W346. [[CrossRef](#)]
38. Lawrence, R.M.; Hoeprieh, P.D. Resistance to infection. In *Infectious Diseases*; Hoeprieh, P.D., Ed.; Harper and Row Publishers: Philadelphia, PA, USA, 1983; pp. 76–84.
39. Bullen, J.J. The significance of iron in infection. *Rev. Infect. Dis.* **1981**, *3*, 1123–1138. [[CrossRef](#)] [[PubMed](#)]
40. Weinberg, E.D. Iron withholding: A defense against infection and neoplasia. *Physiol. Rev.* **1984**, *64*, 65–102. [[CrossRef](#)] [[PubMed](#)]
41. Williams, P.H. Novel iron uptake system specified by ColV plasmids; An important component in the virulence of invasive strains of *Escherichia coli*. *Infect. Immun.* **1979**, *26*, 925–932. [[CrossRef](#)]
42. Williams, P.H.; Warner, P.J. ColV plasmid-mediated colicin V-independent iron uptake system of invasive strains of *Escherichia coli*. *Infect. Immun.* **1980**, *29*, 411–416. [[CrossRef](#)]
43. Caputo, A.; Merhej, V.; Georgiades, K.; Fournier, P.-E.; Croce, O.; Robert, C.; Raoult, D. Pan-genomic analysis to redefine species and subspecies based on quantum discontinuous variation: The *Klebsiella* paradigm. *Biol. Direct* **2015**, *10*, 55. [[CrossRef](#)]
44. Medini, D.; Donati, C.; Tettelin, H.; Massignani, V.; Rappuoli, R. The microbial pangenome. *Curr. Opin. Genet. Dev.* **2005**, *15*, 589–594. [[CrossRef](#)]
45. Frost, L.S.; Leplae, R.; Summers, O.A.; Tossaint, A. Mobile genetic elements: The agents of open source evolution. *Nat. Rev. Microbiol.* **2005**, *3*, 722–732. [[CrossRef](#)]
46. Read, T.D.; Ussery, D.W. Opening the pan-genomics box. *Curr. Opin. Microbiol.* **2006**, *9*, 496–498. [[CrossRef](#)]
47. Mira, M.; Martin-Cuadrado, A.B.; Auria, G.D.; Rodriguez-Valera, F. The bacterial pan-genome: A new paradigm in microbiology. *Int. Microbiol.* **2010**, *13*, 45–57.
48. Taylor, P. Bactericidal and bacteriolytic activity of serum against gram-negative bacteria. *Microbiol. Rev.* **1983**, *47*, 46–83. [[CrossRef](#)] [[PubMed](#)]
49. Joiner, K.A.A. Complement evasion by bacteria and parasites. *Rev. Microbiol.* **1988**, *42*, 201–230. [[CrossRef](#)] [[PubMed](#)]
50. Boyd, E.F. Bacteriophage-encoded bacterial virulence factors and phage-pathogenicity islands interactions. *Adv. Virus Res.* **2012**, *82*, 91–118. [[PubMed](#)]
51. Varani, A.M.; Monteiro-Vitorello, C.B.; Nakaya, H.I.; Van Sluys, M.A. The role of prophage in plant pathogenic bacteria. *Annu. Rev. Phytopathol.* **2013**, *51*, 429–451. [[CrossRef](#)] [[PubMed](#)]
52. Bobay, L.M.; Rocha, E.P.C.; Touchon, M. The adaptation of temperate bacteriophages to their host genomes. *Mol. Biol. Evol.* **2013**, *4*, 737–751. [[CrossRef](#)] [[PubMed](#)]
53. Cascales, E.; Christie, P.J. The versatile bacteria type IV secretion systems. *Nat. Rev. Microbiol.* **2003**, *2*, 137–149. [[CrossRef](#)] [[PubMed](#)]
54. Llosa, M.; O’Callaghan, D. Euro conference on the Biology of Type IV Secretion Processes: Bacterial gates into the outer world. *Mol. Microbiol.* **2004**, *53*, 1–8. [[CrossRef](#)] [[PubMed](#)]
55. Yang, S.; Doolittle, R.F.; Bourne, P.E. Phylogeny determined by protein domain content. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 373–378. [[CrossRef](#)] [[PubMed](#)]
56. Koehorst, J.J.; Saccenti, E.; Schaap, P.J.; dos Santos, V.A.M.; Suarez-Diez, M. Protein domain architectures provide a fast, efficient and scalable alternative to sequence-based methods for comparative functional genomics. *F1000Research* **2016**, *5*, 1987. [[CrossRef](#)] [[PubMed](#)]
57. Basu, M.K.; Poliakov, E.; Rogozin, I.B. Domain mobility in proteins: Functional and evolutionary implications. *Brief Bioinform.* **2009**, *10*, 205–216. [[CrossRef](#)] [[PubMed](#)]
58. Zmasek, C.M.; Godzik, A. This Déjà Vu Feeling—Analysis of Multidomain Protein Evolution in Eukaryotic Genomes. *PLoS Comput. Biol.* **2012**, *8*, e1002701. [[CrossRef](#)] [[PubMed](#)]