



## Article

# Whole Genome Sequencing and Comparative Analysis of the First *Ehrlichia canis* Isolate in China

Jilei Zhang <sup>1,2,\*</sup> , Jiawei Wang <sup>1</sup> and Chengming Wang <sup>3,\*</sup> <sup>1</sup> College of Veterinary Medicine, Yangzhou University, Yangzhou 225009, China; wangjiawei@cwbio.cn<sup>2</sup> College of Medicine, University of Illinois Chicago, Chicago, IL 60612, USA<sup>3</sup> College of Veterinary Medicine, Auburn University, Auburn, AL 36849, USA

\* Correspondence: jileiz@uic.edu (J.Z.); wangche@auburn.edu (C.W.)

**Abstract:** *Ehrlichia canis*, a prominent tick-borne pathogen causing canine monocytic ehrlichiosis (CME), is one of the six recognized *Ehrlichia* species worldwide. Despite its widespread presence in ticks and host dogs in China, comprehensive genomic information about this pathogen remains limited. This study focuses on an in-depth analysis of *E. canis* YZ-1, isolated and cultured from an infected dog in China. The complete genome of *E. canis* YZ-1 was sequenced (1,314,789 bp, 1022 genes, 29% GC content, and 73% coding bases), systematically characterizing its genomic elements and functions. Comparative analysis with representative genomes of *Ehrlichia* species, including *E. canis* strain Jake, *E. chaffeensis*, *Ehrlichia* spp., *E. muris*, *E. ruminantium*, and *E. minasensis*, revealed conserved genes, indicating potential evolutionary connections with *E. ruminantium*. The observed reduction in virulence-associated genes, coupled with a type IV secretion system (T4SS), suggests an intricate balance between pathogenicity and host adaptation. The close relationship with *E. canis* Jake and *E. chaffeensis*, alongside nuanced genomic variations with *E. ruminantium* and *E. mineirensis*, underscores the need to explore emerging strains and advancements in sequencing technologies continuously. This genetic insight opens avenues for innovative medications, studies on probiotic resistance, development of new detection markers, and progress in vaccine development for ehrlichiosis. Further investigations into the functional significance of identified genes and their role in host–pathogen interactions will contribute to a more holistic comprehension of *Ehrlichia*'s biology and its implications for pathogenicity and transmission.

**Keywords:** *Ehrlichia canis*; whole genome sequencing; tick-borne pathogens; dogs



**Citation:** Zhang, J.; Wang, J.; Wang, C. Whole Genome Sequencing and Comparative Analysis of the First *Ehrlichia canis* Isolate in China. *Microorganisms* **2024**, *12*, 125. <https://doi.org/10.3390/microorganisms12010125>

Academic Editors: Galina N. Leonova and Maxim Khasnatinov

Received: 18 December 2023

Revised: 5 January 2024

Accepted: 6 January 2024

Published: 8 January 2024



**Copyright:** © 2024 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/>).

## 1. Introduction

Ehrlichiosis, attributed to a cluster of emerging rickettsial tick-borne pathogens, comprises Gram-negative obligate intracellular bacteria within the genus *Ehrlichia*. The six recognized species within this genus include *E. canis*, *E. chaffeensis*, *E. ewingii*, *E. muris*, *E. minasensis*, and *E. ruminantium* [1–3]. Of particular significance is *E. canis*, the causative agent of canine monocytic ehrlichiosis (CME), also known as canine rickettsiosis, canine typhus, canine hemorrhagic fever, tropical canine pancytopenia, and tracker dog disease. This disease targets platelets, monocytes, and granulocytes, gaining widespread recognition in the 1970s when significant mortalities were observed in German Shepherd dogs during the Vietnam War, despite its initial description in Algeria in 1938 [4–6].

While most dogs with CME recover through treatment with doxycycline or rifampicin, alongside appropriate supportive care, chronic cases leading to fatalities are not uncommon [4,7]. Notably, *E. canis* infections have also been reported in humans [8,9].

Despite several decades of study, genetic resources for *E. canis* remain sparse, with only one complete genome of *E. canis* strain Jake, initially isolated in North America and published in 2005 [10,11], and some from Australia [12]. The challenges of in vitro cultivation of *E. canis* and the complexities of obtaining genomes from their intracellular

locations contribute to this scarcity [4,13]. To address this, we include representative genomes of all six *Ehrlichia* species [14]. Typically, *Ehrlichia* genomes range from 1.2 to 1.5 Mbp, containing approximately 1000 genes, making them relatively small compared with most bacteria [14,15].

In light of these challenges, we embarked on a genomic analysis of *E. canis* isolated and cultured in the laboratory from an infected dog during an ehrlichiosis epidemiological investigation in China. Our primary objective was to acquire genomic data specific to the Chinese strain, enabling comparative analyses with other publicly available *Ehrlichia* genomes. This genomic dataset and analysis aim to enhance our understanding of the relatedness and evolution of *E. canis* in China, providing valuable insights into potential molecular markers for detection, treatment, and vaccination against CME.

## 2. Materials and Methods

### 2.1. Preparation of *E. canis* for Sequencing

The canine monocytic cell line DH82 (ATCC CRL-10389, ATCC, Manassas, VA, USA) was cultivated in Minimum Essential Medium Eagle (MEM) (Sigma-Aldrich, Saint Louis, MO, USA) medium, supplemented with 10% fetal bovine serum (Gibco, Billings, MT, USA) as described before [4]. The *E. canis* strain YZ-1 used in this study was isolated from an adult female beagle dog (*Canis lupus familiaris*) living in a commercial canine facility in Taizhou, Jiangsu, China. This strain was maintained in our laboratory by continuous passage for more than 20 passages in DH82 cells [15]. For propagation, the *E. canis* was inoculated onto a monolayer of DH82 cells and incubated for six days at 37 °C under 5% CO<sub>2</sub>. Subsequently, the culture medium and cells were harvested using a 23 cm cell scraper (Thermo Fisher Scientific, Waltham, MA, USA) and preserved at −80 °C until nucleic acid extraction.

### 2.2. Nucleic Acid Purification

The suspension of pathogens and cells was centrifuged at 500× rpm, 25 °C for 5 min to remove the cell debris. Then, the supernatant was transferred to a new tube, followed by centrifugation at 14,000× rpm, 25 °C for 15 min. The supernatant was removed, and the pellets were resuspended with a sucrose–phosphate–glutamate (1:1:1) solution. The DNAs were extracted from the suspension with a High Pure PCR Template Preparation Kit per the manufacturer's description [3,16,17].

### 2.3. Whole Genome Sequencing and Assembly

Total DNA obtained for WGS underwent quality control, involving the electrophoretic separation of 1 µL of DNA on an agarose gel and quantification using the Qubit system. The genome of *E. canis* YZ-1 was sequenced utilizing Single-Molecule, Real-Time (SMRT) technology on the 3rd generation PacBio platform, conducted at Beijing Novogene Bioinformatics Technology Co., Ltd. (Beijing, China). The SMRT Analysis 2.3.0 software was employed to filter low-quality reads and assemble them into a single, gap-free contig using the filtered reads. Subsequently, the genome underwent de novo assembly through the SMRT portal software, guided by the valid sequencing data. The automatically annotated *E. canis* YZ-1 genome draft was generated using the NCBI Prokaryotic Genomes Annotation Pipeline (NCBI\_PGAP). The complete genome sequence was deposited in the NCBI database with the GenBank accession number CP025749 [15].

### 2.4. *Ehrlichia canis* YZ-1 Genome Analysis

#### 2.4.1. Genomic Components Analysis

Following the acquisition of the complete genome of *E. canis* YZ-1, various genomic components, including coding genes, repeats, noncoding RNA (ncRNA), and genomic islands (GIs), were predicted utilizing specialized software. Coding genes were identified and analyzed using GeneMark, a widely employed tool for gene prediction in bacteria, archaea, and metagenomics [18]. The analysis of repeats, encompassing tandem repeats (TR)

and interspersed repeats (IR), was conducted through the Tandem Repeats Finder (TRF) and RepeatMasker software 4.1.6, respectively [19,20]. The prediction of ncRNA, crucial in life activities without transcription, involved rRNA database blasting or RNAmmer software for rRNA, tRNAscan software for tRNA location and structure, and Rfam software for sRNA [21–23]. Indicative of horizontal origins with potential involvement in symbiosis or pathogenesis, genomic islands were predicted using IslandPath-DIMOB software.

#### 2.4.2. Genomic Function Annotation

The annotation of genes in *E. canis* YZ-1 was carried out using multiple databases, including Gene Ontology (GO), the Kyoto Encyclopedia of Genes and Genomes (KEGG), the Cluster of Orthologous Groups of Proteins (COG), the Non-Redundant Protein Database (NR), and Swiss-Prot. The GO knowledgebase [24] encompasses three main categories—cellular component, molecular function, and biological process—providing comprehensive bioinformatics information on gene functions. KEGG [25,26], a comprehensive database, facilitated the systematic analysis of cellular metabolic pathways and gene expression functions in organisms. COG, NR, and Swiss-Prot [27] are protein-specific annotation databases focusing on different protein functions, such as nonredundant protein information in NR.

Additionally, pathogenesis and antibiotic resistance genes were analyzed using the *E. canis* YZ-1 genome. The PHI (Pathogen–Host Interactions) database [28] annotated interactions between *E. canis* and hosts by blasting target amino acid sequences with reference sequences in PHI. SignalP software 5.0 [29] was employed for the recognition of secretion proteins, while EffectiveT3 software was utilized for the type recognition of the secretion system [30]. Simultaneously, the VFDB (Virulence Factors of Pathogenic Bacteria) database [31] was used to identify virulence factors in *E. canis* YZ-1, and the ARDB (Antibiotic Resistance Genes Database) was consulted for antibiotic resistance analysis [32].

#### 2.5. Macroscopic Comparative Genomic and Phylogenetic Analysis

The genome content of *E. canis* YZ-1 (CP025749) underwent a thorough comparison analysis against the reference genome of *E. canis* strain Jake (NC\_007354) and other publicly available ehrlichial species, including *E. chaffeensis* (CP000236), *Ehrlichia* spp. (NZ\_CP007474), *E. muris* (CP006917), *E. ruminantium* (CR767821), and *E. minasensis* (NZ\_CDGH01000070). Pairwise genomic comparisons were conducted using the Artemis Comparison Tool (ACT) [33] and Geneious 9 [34] with alignments produced by progressive Mauve [35] and MAFFT [36]. Whole genome comparisons were performed with Blastn and tBlastx algorithms using BRIG [37] and Easyfig [38]. The graphical representation of the *E. canis* YZ-1 genome and its components was achieved using DNA-Plotter [39], providing a visual overview of the genomic landscape and facilitating a comprehensive understanding of genomic variations and relationships among the studied *Ehrlichia* species.

### 3. Results

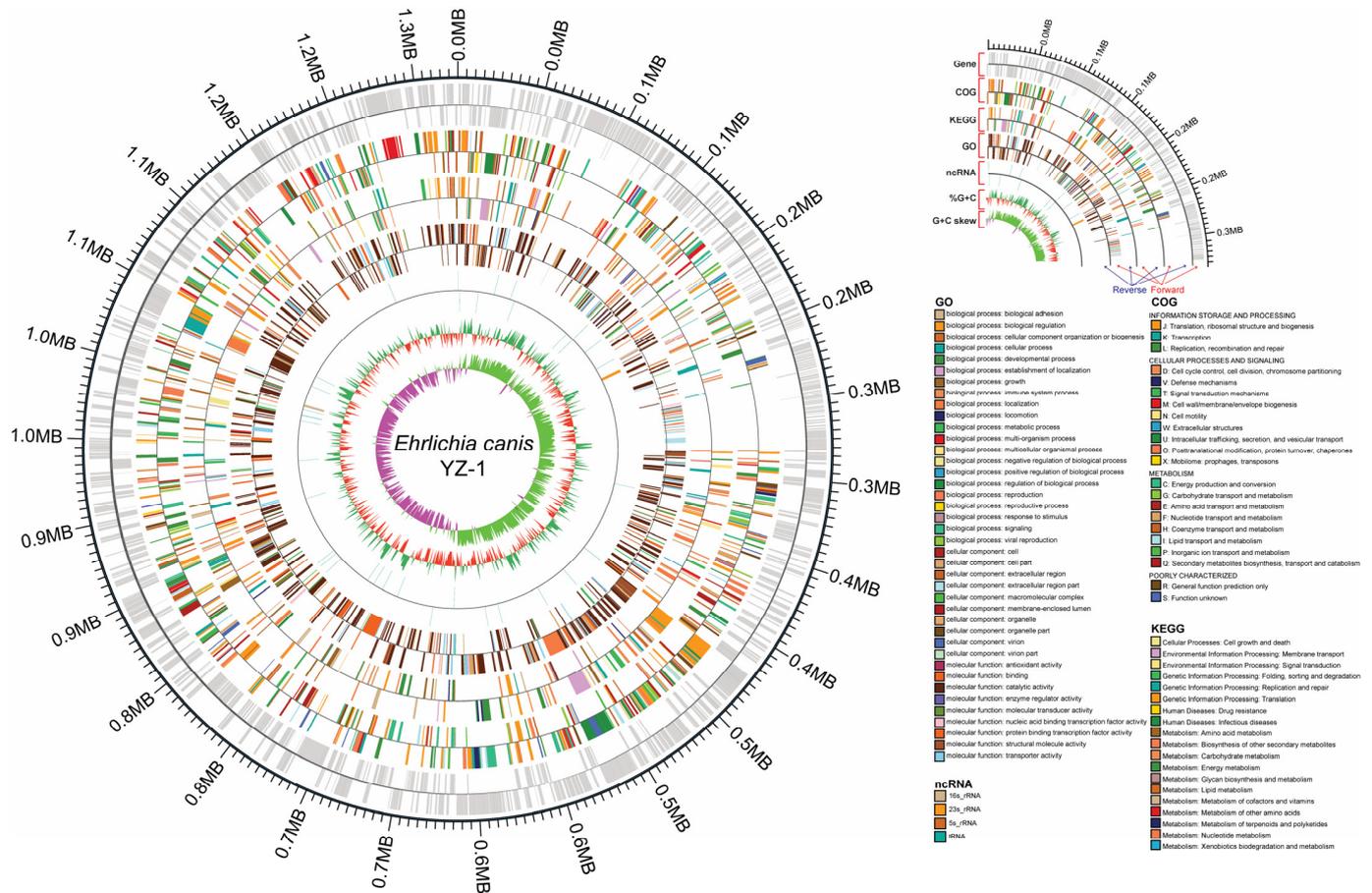
#### 3.1. *E. canis* YZ-1 Genome Assemblies and Characteristics

Utilizing next-generation sequencing technologies, we successfully sequenced, assembled, and characterized the whole genome of *E. canis* strain YZ-1 [4]. This enabled a comprehensive pairwise comparison with other *E. canis* and *Ehrlichia* strains available in publicly accessible databases. The whole genome of *E. canis* YZ-1 spans 1,314,789 bp, harboring 1022 genes, exhibiting an overall GC content of 29%, and comprising 73% coding bases (Table 1, Figures 1 and S1). In our comparative analysis, 900 out of 1022 genes demonstrated identical sequences to those of *E. canis*, albeit with little similarity to other species, including *E. chaffeensis* (6/1022), *E. muris* (2/1022), *Wolbachia* WPWAWU\_1221 (1/1022), *Ehrlichia* spp. (1/1022), and *Bacillus* sp. (1/1022) (Figure S1).

**Table 1.** Basic information on the *Ehrlichia* strains used for analysis in this study.

Organism Name	GenBank #	No. of Genes	%G + C	No. of Bases	No. of Coding Bases	Coding Bases/Total Bases
<i>Ehrlichia canis</i> YZ-1	CP025749	1022	29.00	1,314,789	956,238	72.73
<i>Ehrlichia canis</i> str. Jake	NC_007354	985	28.96	1,315,030	959,246	72.94
<i>Ehrlichia chaffeensis</i> str. Arkansas	NC_007799	1158	30.10	1,176,248	945,019	80.34
<i>Ehrlichia</i> spp. HF	NZ_CP007474	988	29.65	1,148,904	879,236	76.53
<i>Ehrlichia muris</i> AS145	NC_023063	964	29.66	1,196,717	904,005	75.54
<i>Ehrlichia ruminantium</i> Welgevonden	NC_005295	976	27.48	1,512,977	959,837	63.44
<i>Ehrlichia minasensis</i> strain UFMG-EV	NZ_CDBGH01000070	1119	29.89	1,366,818	982,241	71.86

Data obtained from IMG (v1.1; <http://img.jgi.doe.gov> accessed on 1 December 2023).



**Figure 1.** The complete genome of *Ehrlichia canis* YZ-1 from Yangzhou, China. As indicated in the right panel of the figure, there are six tracks of the *E. canis* YZ-1 genome from inside to outside: gene coding sequences in 1000 bp windows, COG annotation, KEGG annotation, GO annotation, noncoding (nc) RNA location, percentage of nucleotides G + C in the genome, and G + C skew values. Each piece of genomic information was shown in forward (upside) and reverse (downside) directions. Moreover, the color map of the GO, COG, KEGG, and ncRNA was included in the figure.

Repetitive sequences, and interspersed and tandem repeats, are vital in genome structure maintenance. For *E. canis* YZ-1, we identified 34 long-terminal repeats with a total length of 2405 bp, 14 transposons with a combined size of 1339 bp, and 105 tandem repeats, constituting approximately 2.6% of the entire genome (Table 2) [40].

We thoroughly analyzed noncoding RNA, specifically rRNA, which is essential for ribosome composition and protein translation. Our analysis identified one 16S rRNA (2738 bp), one 5S rRNA (115 bp), and one 23S rRNA (2791 bp), along with 36 tRNA species totaling 2794 bp (Figure 1). Additionally, a genomic island spanning 44,886 bp was identified at positions 283,341 bp to 328,226 bp in the *E. canis* YZ-1 genome (Figure 1).

**Table 2.** Nuclear repeats analysis on the *Ehrlichia canis* YZ-1.

Repeat Type	Number	Total Length (bp)	Percentage in the Genome (%)
<b>Interspersed repeats</b>			
Long-terminal repeats	34	2405	0.1829
DNA transposons	14	1339	0.1018
Long interspersed nuclear elements	8	808	0.0615
Short interspersed nuclear elements	4	225	0.0171
Rolling circle	3	288	0.0219
Unknown	2	152	0.0116
<b>Tandem repeats</b>			
Tandem repeats	105	33,841	2.5739
Minisatellite DNA	52	2500	0.1901
Microsatellite DNA	1	28	0.0021

DNA methylation, a crucial biological process involving adding methyl groups to DNA molecules, can influence the activity of DNA segments without altering the sequence. In *E. canis* YZ-1, we identified 3730 DNA methylation locations, including 205 (5.5%) N6-methyladenine (6 mA), 391 (10.5%) N4-methylcytosine (4 mC), and an overwhelming majority of unknown methylation patterns (84.0%, 3134/3730) (Figure S1). Notably, 5-methylcytosine (5 mC) was not detected in the analyzed methylation types. (Figure S1) [41–43].

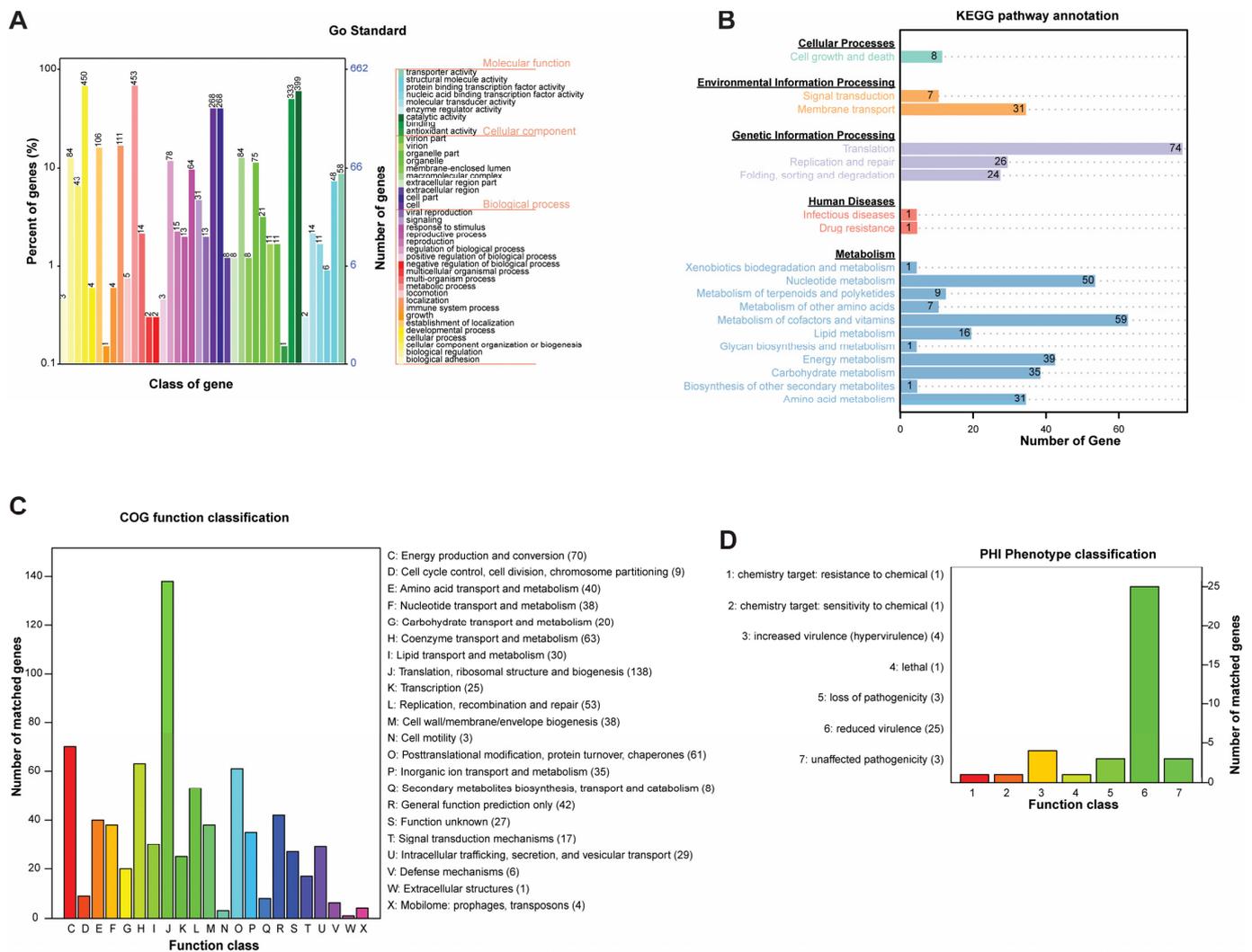
### 3.2. Genomic Functional Analysis

The Gene Ontology (GO) project offers structured, controlled vocabularies classifying molecular and cellular biology into three nonoverlapping domains: molecular function, cellular component, and biological process, with further subdomains. In our analysis of 1022 genes from *E. canis* YZ-1, 2959 annotations were obtained through GO analysis (Figure 2A). Notably, a substantial number of genes were associated with cellular processes (450 genes) and metabolic processes (453 genes) within the biological process domain. Similarly, the cellular component domain revealed a predominant presence of genes related to cell (268) and cell part (268) (Figure 2A). In the molecular function domain, the most represented functions were binding (333) and catalytic activity (399) (Figure 2A) [44].

KEGG, a database integrating genomic and higher-order functional information, provides graphical representations of cellular processes. More than half of the genes in *E. canis* YZ-1 (537 out of 1022) were associated with 97 pathways in KEGG. Further analysis of subpathways revealed categories with over 50 genes, including translation-related pathways (74 genes) in genetic information processing, nucleotide metabolism (50 genes), and metabolism of cofactors and vitamins (59 genes) in metabolism (Figure 2B) [45].

The COGs database, based on the orthology concept, classifies proteins from completely sequenced genomes. In the *E. canis* YZ-1 genome, proteins were classified using the COG database, enhancing the genome sequence's utility for functional and evolutionary analysis. The highest enrichment was observed in genes coding for proteins related to translation, ribosomal structure, and biogenesis (138 genes), followed by energy production and conversion (70 genes), coenzyme transport and metabolism (63 genes), post-translational modification, protein turnover, and chaperones (61 genes), and replication, recombination, and repair (53 genes) (Figure 2C) [46,47].

As a zoonotic pathogen, *E. canis* threatens animals, humans, ecosystems, and regional economies. The PHI database was consulted to understand the dynamic interactions between pathogens and their hosts. In *E. canis* YZ-1, 25 genes were related to reduced virulence, the most prevalent among identified PHI phenotypes (Figure 2D). Additionally, four genes related to increased virulence (hypervirulence) were identified, showing varying similarities to genes from other organisms (Figure 2D, Supplementary Table S1).



**Figure 2.** Genomic functional analysis of *Ehrlichia canis* YZ-1. (A) The functional analysis of genes, both number (Y-axis on the right) and percentage of total genes (Y-axis on the left), from *E. canis* YZ-1 was performed using the GO database (X-axis), which is established based on the biological process. (B) The number of genes (X-axis) from the *E. canis* YZ-1 in each metabolic pathway (Y-axis) was analyzed using the KEGG database. (C) The number of genes (Y-axis) of *E. canis* YZ-1 in biological functions (X-axis) was shown based on the COG database. (D) The number of genes from *E. canis* YZ-1 in each Pathogen–Host Interactions (PHI) category was shown here.

Moreover, six genes in *E. canis* YZ-1 were associated with virulence genes *CcmC*, *Hsp60*, *ClpC*, *SodB*, *LPS*, and *ClpP*, albeit with low similarity percentages ranging from 40.8% to 58.4%. Analyzing the SignalP server and TMHMM revealed 35 signal peptides and 300 transmembrane helices proteins. Interestingly, 35 of these proteins possessed signal peptides but lacked transmembrane helices. VirB genes related to the type IV secretion system were identified, consisting of two clusters containing virB8/virB9/virB10/virB11/virD4 and virB3/virB4/three large virB6, along with three virB9, virB8, and virB4 located separately (Supplementary Table S2). Unfortunately, no gene related to antibiotic resistance was found in *E. canis* YZ-1 [48].

### 3.3. Similarity between *E. canis* YZ-1 and Other *Ehrlichia* Species/Strains

#### 3.3.1. Synteny Comparative Genome Analysis

Synteny analysis, employing orthologous genes with similar functions, elucidates homologous gene conservation and order across genomes. In our study, we conducted a com-

prehensive synteny analysis comparing *E. canis* YZ-1 with various *Ehrlichia* species/strains, including *E. chaffeensis*, *Ehrlichia* spp., *E. muris*, *E. ruminantium*, *E. mineirensis*, and *E. canis* Jake [49] (Table 1, Supplementary Table S3).

When aligned with *E. chaffeensis* strain Arkansas (CP000236), which was isolated from a patient resident in Arkansas [10,11], there were 76.7% (1,008,696/1,314,789 bp) of *E. canis* YZ-1 and 85.9% (1,010,920/1,176,248) of *E. chaffeensis* involved in both bidimensional (Figure S2) and parallel (Figure 3) comparative analysis. In all the 337 collinear alignment blocks, there were 309 collinearities with a similarity between 48.8% and 91.1%, nine inversions with a similarity between 70.0% and 84.3%, 12 translocations with a similarity between 47.9% and 81.8%, and six inversions/translocations with a similarity between 64.5% and 87.8% (Figures 3A and S2A).

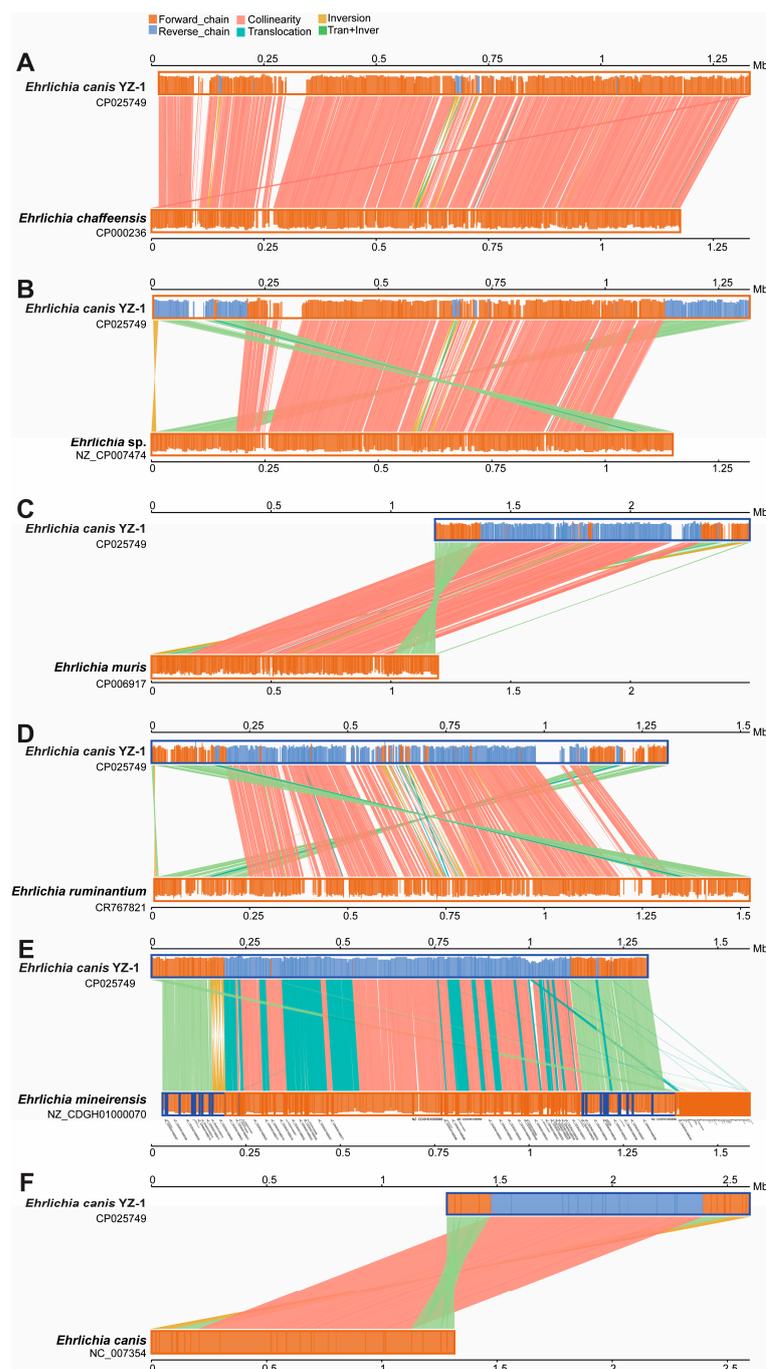
When aligned with *Ehrlichia* spp. HS strain (*Ehrlichia japonica*) (NZ\_CP007474) [50], there were 77.48% (1,018,665/1,314,789 bp) of *E. canis* YZ-1 and 89.0% (1,022,729/1,148,904) of *Ehrlichia* spp. involved in this analysis. In all the 294 collinear alignment blocks, there were 183 collinearities with a similarity between 45.6% and 90.8%, nine inversions with a similarity between 75.0% and 86.8%, 12 translocations with a similarity between 59.0% and 82.6%, and 92 inversions/translocations with a similarity between 54.2% and 88.1% (Figures 3B and S2B).

When aligned with *E. muris* strain AS145 (CP006917), which was isolated from a wild mouse in Japan [51], there were 77.1% (1,013,862/1,314,789 bp) of *E. canis* YZ-1 and 85.0% (1,017,665/1,196,717) of *E. muris* involved in this analysis. In all the 317 collinear alignment blocks, there were 201 collinearities with a similarity between 46.2% and 93.1%, 15 inversions with a similarity between 67.1% and 84.8%, ten translocations with a similarity between 51.4% and 88.9%, and 91 inversions/translocations with a similarity between 54.8% and 89.7% (Figures 3C and S2C).

When aligned with the *E. ruminantium* Welgevonden-type strain (CP006917), which was isolated from cattle in Welgevonden, South Africa [52], there were 68.1% (895,013/1,314,789 bp) of *E. canis* YZ-1 and 59.3% (898,925/1,516,355) of *E. ruminantium* involved in this analysis. In all the 434 collinear alignment blocks, there were 241 collinearities with a similarity between 50.0% and 86.7%, 17 inversions with a similarity between 62.3% and 81.0%, 28 translocations with a similarity between 45.5% and 96.4%, and 148 inversions/translocations with a similarity between 35.0% and 91.4% (Figures 3D and S2D).

*E. mineirensis* is a new *Ehrlichia* species isolated from *Rhipicephalus (Boophilus) microplus* from Minas Gerais, Brazil, and has been molecularly characterized recently [14,53]. When aligned with this strain (NZ\_CDGH01000070), there were 95.9% (1,260,937/1,314,789 bp) of *E. canis* YZ-1 and 89.1% (1,261,326/1,414,910) of *E. mineirensis* involved in this analysis. In all the 335 collinear alignment blocks, there were 80 collinearities with a similarity between 62.8% and 92.9%, 13 inversions with a similarity between 46.6% and 90.6%, 114 translocations with a similarity between 56.7% and 96.1%, and 128 inversions/translocations with a similarity between 39.5% and 93.3% (Figures 3E and S2E).

When aligned with *E. canis* strain Jake (NC\_007354) [54], there were 99.3% (1,305,712/1,314,789 bp) of *E. canis* YZ-1 and 99.3% (1,305,822/1,315,030) of *E. canis* Jake involved in this analysis. In all the 38 collinear alignment blocks, there were 17 collinearities with a similarity between 91.0% and 100%, five inversions with a similarity between 79.3% and 99.96%, zero translocation, and 16 inversions/translocations with a similarity between 77.7% and 100%, which again proved that our strain is *Ehrlichia canis*, not another *Ehrlichia* species (Figures 3F and S2F).



**Figure 3.** Parallel synteny analysis of *Ehrlichia canis* YZ-1 with the *Ehrlichia* reference sequences. The parallel synteny analysis was performed on *E. canis* YZ-1 from this study with the *Ehrlichia* reference sequences from NCBI, including *E. chaffeensis* (GenBank accession number CP000236) (A), *Ehrlichia* spp. (NZ\_CP007474) (B), *E. muris* (CP006917) (C), *E. ruminantium* (CR767821) (D), *E. mineirensis* (NZ\_CDBGH01000070) (E), and *E. canis* (NC\_007354) (F). The genomic differences/similarities in collinearity, inversion, translocation, and translocation + inversion are shown in height and color, as indicated at the top of the figure.

### 3.3.2. Insertion–Deletion Mutations Analysis

Indels, or insertion–deletion mutations, are the insertions and/or deletions of nucleotides into and/or out of genomic DNA with less than 1kb length. They are supremely critical in clinical next-generation sequencing, due to their driving mechanism underlying many constitutional and oncological diseases [55]. We identified these mutations by

comparing our *E. canis* strain YZ-1 to other *Ehrlichia* species or strains. The indels mainly happened inside the CDS (coding sequencings), with a higher number in *E. chaffeensis* (13 insertions; 20 deletions) and *E. canis* (17 insertions; 16 deletions) compared with the other species/strains (Supplementary Figure S3 and Table S4). However, these insertion/deletion mutations were extremely short (less than 1 bp) and mostly frameshift mutations or without impacts on the open reading frame (Supplementary Figure S3 and Table S5).

### 3.3.3. Structural Variation (SV) Analysis

Structural variations (SVs) are generally defined as a region of DNA approximately 50 bp long or longer and can include inversions and balanced translocations or genomic imbalances, such as insertions and deletions. These SVs often overlap with segmental duplications, DNA regions present more than once, or copies of which more than 90% of genes are identical [56–58]. Here, we performed the SVs analysis on our *E. canis* YZ-1 and the reference *Ehrlichia* species one-to-one to identify the regions and types of SVs (Figure 4 and Supplementary Table S6). Compared with *E. canis* YZ-1, *E. chaffeensis* (334), *Ehrlichia* spp. (393), *E. muris* (402), and *E. ruminantium* (525) have the most SVs, mainly complex iIndels, which are not comparable due to the significant mutations that occurred in that region (Figure 4A–D and Supplementary Table S6). Meanwhile, *E. mineirensis* has 147 SVs, mainly insertions and inversions (Figure 4E and Supplementary Table S6). However, the *E. canis* Jake strain only has 42 SVs, mainly insertions and deletions (Figure 4F and Supplementary Table S6).

Lorem ipsum

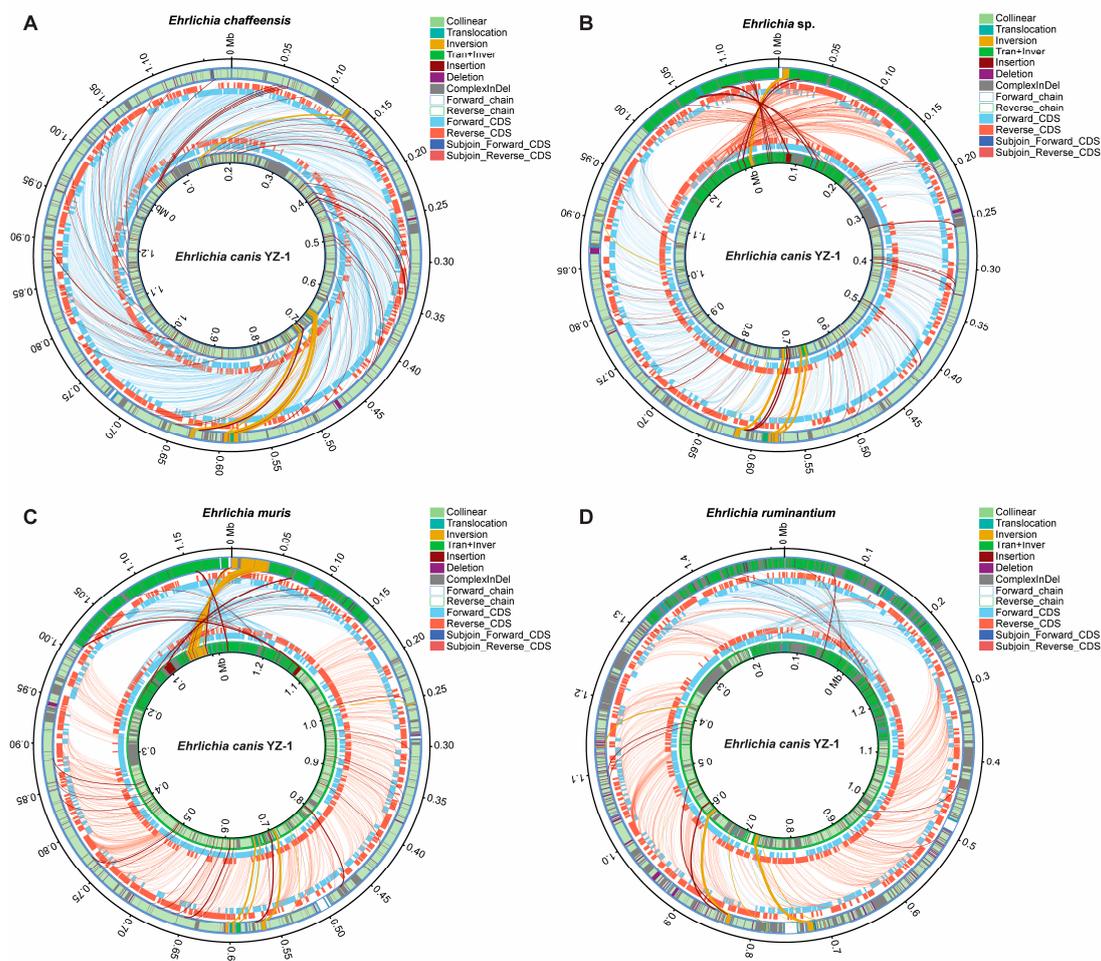
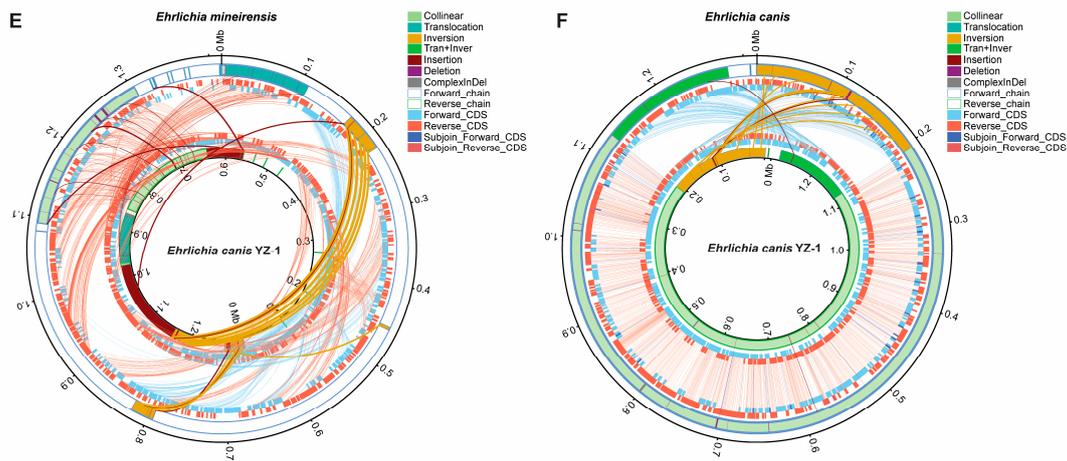


Figure 4. Cont.



**Figure 4.** Comparative mutations analysis of *Ehrlichia canis* YZ-1 with the *Ehrlichia* reference sequences. The potential structural mutations were analyzed by arranging *E. canis* YZ-1 from this study with *Ehrlichia* reference sequences published before, including *E. chaffeensis* (CP000236) (A), *Ehrlichia* spp. (NZ\_CP007474) (B), *E. muris* (CP006917) (C), *E. ruminantium* (CR767821) (D), *E. mineirensis* (NZ\_CDGH01000070) (E), and *E. canis* (NC\_007354) (F). The mutation types identified here were indicated in the figure accordingly.

#### 4. Discussion

In this study, we conducted next-generation sequencing and comparative genomic analysis on *E. canis* YZ-1, an isolate which was cultured from an infected dog from a commercial canine farm. Our investigation aimed to unravel pathogenesis at the gene level and establish genetic relatedness with other *Ehrlichia* species/strains. Notably, our findings contribute to a deeper understanding of this zoonotic pathogen's evolutionary dynamics and functional aspects.

Our genomic analysis confirmed *E. canis* YZ-1's identity, revealing substantial genomic congruence with the *E. canis* Jack strain and a close relationship with *E. chaffeensis* and *E. ruminantium*. Intriguingly, we observed a notable reduction in virulence-associated genes, suggesting an adaptive strategy favoring prolonged survival within the host, aligning with observations of asymptomatic infections in our experimental dog model [4]. The presence of genes such as *ClpP* and *ClpX*, involved in protease activity and ATP-dependent processes, underscores the pathogen's intricate regulation of virulence [59].

Pathogen–host interactions revealed a plethora of genes with low similarity to those in the database. While the mechanisms remain elusive, these findings align with our previous observations of infected dogs recovering without treatment, indicating potential adaptations for a prolonged host survival period [4]. Identifying a type IV secretion system (T4SS) suggests a role in translocating bacterial effectors into host cells, further highlighting the intricate nature of *E. canis* YZ-1's interactions with its host. Similar to the *E. canis* reported before, we found two clusters of Vir homologous proteins in *E. canis* YZ-1. One contains *virB8/virB9/virB10/virB11/virD4*, and one includes *virB3/virB4/virB6/virB6/virB6* [54]. But we also found *virB9*, *virB8*, and *virB6* located between and downstream of these two clusters, which was also reported in the *Ehrlichia* spp. HF recently [60].

Comparative genetic analysis showcased *E. canis* YZ-1's almost complete synteny with the *E. canis* Jake strain, affirming its species identity, which was reported to be almost identical to *Ehrlichia ruminantium* (Welgevonden and Gardel strains) before [54]. However, *E. canis* YZ-1 has the lowest similarity with the *E. ruminantium* Welgevonden strain, with only 68.1% of *E. canis* YZ-1 and 59.3% of *E. ruminantium* involved in the analysis. Instead, *E. canis* YZ-1 is closer to *E. mineirensis* (95.9% of *E. canis* YZ-1 and 89.1% of *E. mineirensis*), a new *Ehrlichia* strain reported recently [14,53], possibly influenced by host specificity and environmental factors.

One limitation of our analysis was not to include the first *E. canis* strain in Australia [12]. The genomic comparison in this study was conducted before we had access to the genome sequences of the Australian strain. Notably, Neave et al. conducted a comprehensive genomic analysis of the isolates from Australia, the Jake strain, and the YZ-1, and concluded that the Australian *E. canis* genomes were highly conserved, and the Australian genomes were most similar to *E. canis* YZ-1 from China [12].

Interestingly, our analysis revealed a closer relationship between *E. canis* YZ-1 and *E. canis* strains from Australia than those from other regions. Phylogenetic analysis using different genes demonstrated varying relationships, highlighting the importance of comprehensive analyses involving multiple genetic markers (Figure S4). Notably, the inclusion of the variable *trp36* gene pointed to the Australian strains belonging to the Taiwanese genotype, emphasizing the need for a nuanced understanding of strain diversity [61].

This close relationship between *E. canis* YZ-1 and *E. ruminantium* may also be supported by the finding that multiple membrane transporter genes in our *E. canis* YZ-1 are originally from *E. ruminantium* when the analysis was performed based on the Transporter Classification Database, which is a database with a comprehensive IUBMB approved classification system for membrane transport proteins known as the transporter classification (TC) system. There are 62 membrane transport proteins that were identified from *E. canis* YZ-1, and 11 of them (accession #: Q5HB83; Q9R425) are from *E. ruminantium* (Supplemental Table S2). Interestingly, 10 out of the 11 proteins were major antigenic proteins (Q9R425). However, the reason behind this is still unclear. One of the possibilities is that both pathogens have infected the same host (ticks or animals) and shared their genomic information. For instance, an *Ehrlichia* species close to *E. canis* and *E. ruminantium* has been collected from the same host camel [62]. Moreover, there are 11 proteins in *E. canis* YZ-1 related to ubiquinone (ubiquitous or coenzyme Q in humans and animals, respectively) biosynthesis. These proteins or coding genes were also reported as originally coming from *E. ruminantium* and *E. chaffeensis* [58]. Further studies on the divergence or characterization of these genes and proteins should be performed in establishing therapeutic interventions in the ongoing battle against tick-borne pathogens. Moreover, a more in-depth exploration of evolutionary history and potential reasons behind these observations requires further studies.

## 5. Conclusions

The comprehensive genomic sequencing of *E. canis* YZ-1 has furnished essential resources for an in-depth exploration of this pathogen, offering valuable insights into host-pathogen interactions, potential evolutionary trajectories, and genetic distinctions from globally prevalent *E. canis* strains. The abundance of genes associated with virulence reduction, potentially aiding evasion of the host immune system, coupled with its close relationship to *E. ruminantium*, presents intriguing avenues for future research and analysis. Unraveling the mechanisms underlying these genetic attributes is imperative for a nuanced understanding of *E. canis* YZ-1's adaptive strategies. The proteins implicated in these processes and those mediating pathogen-host interactions emerge as promising candidates for developing vaccines and therapeutic interventions against ehrlichiosis. Further studies are warranted to decipher these genes' functional significance and role in the intricate dynamics of host-pathogen relationships. By unraveling the molecular intricacies of *E. canis* YZ-1, this study lays the groundwork for advancing our comprehension of ehrlichial pathogenesis and enhancing disease prevention and control strategies. Continued investigations into the multifaceted aspects of *E. canis* YZ-1's biology will contribute to our understanding of ehrlichiosis and offer potential avenues for targeted therapeutic interventions in the ongoing battle against tick-borne pathogens.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms12010125/s1>. Supplementary Table S1. Pathogen-host interactions analysis of *E. canis* YZ-1; Supplementary Table S2. All the genes of *E. canis* YZ-1 identified in KEGG and TC system analysis; Supplementary Table S3. SNPs (single nucleotide

polymorphisms) analysis of *E. canis* YZ-1 and related *Ehrlichia* strains; Supplementary Table S4. Indels (insertions and deletions) analysis of *E. canis* YZ-1 and related *Ehrlichia* strains; Supplementary Table S5. Mutation type caused by indels (insertions and deletions) in *E. canis* YZ-1 and related *Ehrlichia* strains; Supplementary Table S6. The SV types of *E. canis* YZ-1 and related *Ehrlichia* strains; Figure S1. The epigenetic modifications of *Ehrlichia canis* YZ-1; Figure S2. Bidimensional synteny analysis of *Ehrlichia canis* YZ-1 with the reference sequences; Figure S3. The comparative analysis of indels between *Ehrlichia canis* YZ-1 and reference sequences; Figure S4. Phylogenetic relationships of *Ehrlichia canis* YZ-1 and references.

**Author Contributions:** Conceptualization, investigation, methodology, formal analysis, and writing, J.Z., J.W., and C.W.; investigation, review, and manuscript writing, J.Z. and C.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by a grant from the National Key Research Project of China (grant 2016YFD0500804) and the National Natural Science Foundation of China (grant 31472225). The funders had no role in study design, data collection, interpretation, or the decision to submit the work for publication.

**Data Availability Statement:** The *E. canis* strain YZ-1 whole genome sequence has been deposited in GenBank under the accession number CP025749.

**Acknowledgments:** The authors are grateful for the technical support from Xiaofeng Zhen, Jiangsu Agri-animal Husbandry Vocational College, Taizhou, China.

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

## References

1. Yu, X.J.; McBride, J.W.; Walker, D.H. Restriction and expansion of Ehrlichia strain diversity. *Vet. Parasitol.* **2007**, *143*, 337–346. [[CrossRef](#)] [[PubMed](#)]
2. Esemu, S.N.; Ndip, L.M.; Ndip, R.N. Ehrlichia species, probable emerging human pathogens in sub-Saharan Africa: Environmental exacerbation. *Rev. Environ. Health* **2011**, *26*, 269–279. [[CrossRef](#)] [[PubMed](#)]
3. Zhang, J.; Kelly, P.; Guo, W.; Xu, C.; Wei, L.; Jongejan, F.; Loftis, A.; Wang, C. Development of a generic Ehrlichia FRET-qPCR and investigation of ehrlichioses in domestic ruminants on five Caribbean islands. *Parasites Vectors* **2015**, *8*, 506. [[CrossRef](#)] [[PubMed](#)]
4. Zhang, J.; Wang, J.; Kelly, P.J.; Zhang, Y.; Li, M.; Li, J.; Zhang, R.; Wang, Y.; Huang, K.; You, J.; et al. Experimental infection and co-infection with Chinese strains of *Ehrlichia canis* and *Babesia vogeli* in intact and splenectomized dogs: Insights on clinical, hematologic and treatment responses. *Vet. Parasitol.* **2023**, *323*, 110032. [[CrossRef](#)] [[PubMed](#)]
5. Stiles, J. Canine rickettsial infections. *Vet. Clin. N. Am. Small Anim. Pract.* **2000**, *30*, 1135–1149. [[CrossRef](#)]
6. Aziz, M.U.; Hussain, S.; Song, B.; Ghauri, H.N.; Zeb, J.; Sparagano, O.A. Ehrlichiosis in Dogs: A Comprehensive Review about the Pathogen and Its Vectors with Emphasis on South and East Asian Countries. *Vet. Sci.* **2022**, *10*, 21. [[CrossRef](#)]
7. Harrus, S.; Waner, T. Diagnosis of canine monocytotropic ehrlichiosis (*Ehrlichia canis*): An overview. *Vet. J.* **2011**, *187*, 292–296. [[CrossRef](#)]
8. Conrad, M.E. *Ehrlichia canis*: A tick-borne rickettsial-like infection in humans living in the southeastern United States. *Am. J. Med. Sci.* **1989**, *297*, 35–37. [[CrossRef](#)]
9. Forero-Becerra, E.; Patel, J.; Martínez-Díaz, H.C.; Betancourt-Ruiz, P.; Benavides, E.; Durán, S.; Olaya-Másmela, L.A.; Bolaños, E.; Hidalgo, M.; McBride, J.W. Seroprevalence and Genotypic Analysis of *Ehrlichia canis* Infection in Dogs and Humans in Cauca, Colombia. *Am. J. Trop. Med. Hyg.* **2021**, *104*, 1771–1776. [[CrossRef](#)]
10. Miura, K.; Rikihisa, Y. Virulence potential of Ehrlichia chaffeensis strains of distinct genome sequences. *Infect. Immun.* **2007**, *75*, 3604–3613. [[CrossRef](#)]
11. Dunning Hotopp, J.C.; Lin, M.; Madupu, R.; Crabtree, J.; Angiuoli, S.V.; Eisen, J.A.; Seshadri, R.; Ren, Q.; Wu, M.; Utterback, T.R.; et al. Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet.* **2006**, *2*, e21. [[CrossRef](#)]
12. Neave, M.J.; Mileto, P.; Joseph, A.; Reid, T.J.; Scott, A.; Williams, D.T.; Keyburn, A.L. Comparative genomic analysis of the first *Ehrlichia canis* detections in Australia. *Ticks Tick-Borne Dis.* **2022**, *13*, 101909. [[CrossRef](#)]
13. Cheng, C.; Ganta, R.R. Laboratory maintenance of Ehrlichia chaffeensis and Ehrlichia canis and recovery of organisms for molecular biology and proteomics studies. *Curr. Protoc. Microbiol.* **2008**, *9*, 3A.1.1–3A.1.21. [[CrossRef](#)]
14. Cabezas-Cruz, A.; Zweygarth, E.; Broniszewska, M.; Passos, L.M.; Ribeiro, M.F.; Manrique, M.; Tobes, R.; de la Fuente, J. Complete Genome Sequence of Ehrlichia mineirensis, a Novel Organism Closely Related to Ehrlichia canis with a New Host Association. *Genome Announc.* **2015**, *3*, e01450-14. [[CrossRef](#)] [[PubMed](#)]
15. Zhang, J.; Wang, J.; Wang, C. Complete Genome Sequence of Ehrlichia canis Strain YZ-1, Isolated from a Beagle with Fever and Thrombocytopenia. *Genome Announc.* **2018**, *6*, 10–1128. [[CrossRef](#)] [[PubMed](#)]
16. Zhang, J.; Lu, G.; Kelly, P.; Zhang, Z.; Wei, L.; Yu, D.; Kayizha, S.; Wang, C. First report of Rickettsia felis in China. *BMC Infect. Dis.* **2014**, *14*, 682. [[CrossRef](#)]

17. Zhang, J.; Lu, G.; Li, J.; Kelly, P.; Li, M.; Wang, J.; Huang, K.; Qiu, H.; You, J.; Zhang, R.; et al. Molecular Detection of *Rickettsia felis* and *Rickettsia bellii* in Mosquitoes. *Vector Borne Zoonotic Dis.* **2019**, *19*, 802–809. [[CrossRef](#)]
18. Besemer, J.; Lomsadze, A.; Borodovsky, M. GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **2001**, *29*, 2607–2618. [[CrossRef](#)]
19. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)]
20. Saha, S.; Bridges, S.; Magbanua, Z.V.; Peterson, D.G. Empirical comparison of ab initio repeat finding programs. *Nucleic Acids Res.* **2008**, *36*, 2284–2294. [[CrossRef](#)]
21. Gardner, P.P.; Daub, J.; Tate, J.G.; Nawrocki, E.P.; Kolbe, D.L.; Lindgreen, S.; Wilkinson, A.C.; Finn, R.D.; Griffiths-Jones, S.; Eddy, S.R.; et al. Rfam: Updates to the RNA families database. *Nucleic Acids Res.* **2009**, *37*, D136–D140. [[CrossRef](#)] [[PubMed](#)]
22. Lagesen, K.; Hallin, P.; Rødland, E.A.; Staerfeldt, H.H.; Rognes, T.; Ussery, D.W. RNAMmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **2007**, *35*, 3100–3108. [[CrossRef](#)] [[PubMed](#)]
23. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964. [[CrossRef](#)] [[PubMed](#)]
24. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)] [[PubMed](#)]
25. Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **2004**, *32*, D277–D280. [[CrossRef](#)]
26. Kanehisa, M.; Goto, S.; Hattori, M.; Aoki-Kinoshita, K.F.; Itoh, M.; Kawashima, S.; Katayama, T.; Araki, M.; Hiraoka, M. From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res.* **2006**, *34*, D354–D357. [[CrossRef](#)]
27. Magrane, M. UniProt Knowledgebase: A hub of integrated protein data. *Database J. Biol. Databases Curation* **2011**, *2011*, bar009. [[CrossRef](#)]
28. Vargas, W.A.; Martín, J.M.; Rech, G.E.; Rivera, L.P.; Benito, E.P.; Díaz-Minguez, J.M.; Thon, M.R.; Sukno, S.A. Plant defense mechanisms are activated during biotrophic and necrotrophic development of *Colletotricum graminicola* in maize. *Plant Physiol.* **2012**, *158*, 1342–1358. [[CrossRef](#)]
29. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786. [[CrossRef](#)]
30. Jehl, M.A.; Arnold, R.; Rattai, T. Effective—A database of predicted secreted bacterial proteins. *Nucleic Acids Res.* **2011**, *39*, D591–D595. [[CrossRef](#)]
31. Chen, L.; Xiong, Z.; Sun, L.; Yang, J.; Jin, Q. VFDB 2012 update: Toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res.* **2012**, *40*, D641–D645. [[CrossRef](#)] [[PubMed](#)]
32. Liu, B.; Pop, M. ARDB—Antibiotic Resistance Genes Database. *Nucleic Acids Res.* **2009**, *37*, D443–D447. [[CrossRef](#)]
33. Carver, T.; Berriman, M.; Tivey, A.; Patel, C.; Böhme, U.; Barrell, B.G.; Parkhill, J.; Rajandream, M.A. Artemis and ACT: Viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **2008**, *24*, 2672–2676. [[CrossRef](#)] [[PubMed](#)]
34. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [[CrossRef](#)] [[PubMed](#)]
35. Darling, A.E.; Mau, B.; Perna, N.T. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **2010**, *5*, e11147. [[CrossRef](#)]
36. Katoh, K.; Standley, D.M. MAFFT: Iterative refinement and additional methods. *Methods Mol. Biol.* **2014**, *1079*, 131–146. [[CrossRef](#)] [[PubMed](#)]
37. Alikhan, N.F.; Petty, N.K.; Ben Zakour, N.L.; Beatson, S.A. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genom.* **2011**, *12*, 402. [[CrossRef](#)]
38. Sullivan, M.J.; Petty, N.K.; Beatson, S.A. Easyfig: A genome comparison visualizer. *Bioinformatics* **2011**, *27*, 1009–1010. [[CrossRef](#)]
39. Carver, T.; Thomson, N.; Bleasby, A.; Berriman, M.; Parkhill, J. DNAPlotter: Circular and linear interactive genome visualization. *Bioinformatics* **2009**, *25*, 119–120. [[CrossRef](#)]
40. de Koning, A.P.; Gu, W.; Castoe, T.A.; Batzer, M.A.; Pollock, D.D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **2011**, *7*, e1002384. [[CrossRef](#)]
41. Ehrlich, M.; Gama-Sosa, M.A.; Carreira, L.H.; Ljungdahl, L.G.; Kuo, K.C.; Gehrke, C.W. DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. *Nucleic Acids Res.* **1985**, *13*, 1399–1412. [[CrossRef](#)] [[PubMed](#)]
42. Vanyushin, B.F.; Tkacheva, S.G.; Belozersky, A.N. Rare bases in animal DNA. *Nature* **1970**, *225*, 948–949. [[CrossRef](#)] [[PubMed](#)]
43. Dunn, D.B.; Smith, J.D. The occurrence of 6-methylaminopurine in deoxyribonucleic acids. *Biochem. J.* **1958**, *68*, 627–636. [[CrossRef](#)] [[PubMed](#)]
44. Harris, M.A.; Clark, J.; Ireland, A.; Lomax, J.; Ashburner, M.; Foulger, R.; Eilbeck, K.; Lewis, S.; Marshall, B.; Mungall, C.; et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **2004**, *32*, D258–D261. [[CrossRef](#)] [[PubMed](#)]
45. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)]
46. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36. [[CrossRef](#)]
47. Tatusov, R.L.; Koonin, E.V.; Lipman, D.J. A genomic perspective on protein families. *Science* **1997**, *278*, 631–637. [[CrossRef](#)]

48. Urban, M.; Cuzick, A.; Seager, J.; Wood, V.; Rutherford, K.; Venkatesh, S.Y.; Sahu, J.; Iyer, S.V.; Khamari, L.; De Silva, N.; et al. PHI-base in 2022: A multi-species phenotype database for Pathogen-Host Interactions. *Nucleic Acids Res.* **2022**, *50*, D837–D847. [[CrossRef](#)]
49. Cheng, F.; Wu, J.; Fang, L.; Wang, X. Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species. *Front. Plant Sci.* **2012**, *3*, 198. [[CrossRef](#)]
50. Bekebrede, H.; Lin, M.; Teymournejad, O.; Rikihisa, Y. Discovery of in vivo Virulence Genes of Obligatory Intracellular Bacteria by Random Mutagenesis. *Front. Cell Infect. Microbiol.* **2020**, *10*, 2. [[CrossRef](#)]
51. Thirumalapura, N.R.; Qin, X.; Kuriakose, J.A.; Walker, D.H. Complete Genome Sequence of Ehrlichia muris Strain AS145T, a Model Monocytotropic Ehrlichia Strain. *Genome Announc.* **2014**, *2*, e01234–13. [[CrossRef](#)] [[PubMed](#)]
52. Collins, N.E.; Liebenberg, J.; de Villiers, E.P.; Brayton, K.A.; Louw, E.; Pretorius, A.; Faber, F.E.; van Heerden, H.; Josemans, A.; van Kleef, M.; et al. The genome of the heartwater agent Ehrlichia ruminantium contains multiple tandem repeats of actively variable copy number. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 838–843. [[CrossRef](#)]
53. Cabezas-Cruz, A.; Vancová, M.; Zweygarth, E.; Ribeiro, M.F.; Grubhoffer, L.; Passos, L.M. Ultrastructure of Ehrlichia mineirensis, a new member of the Ehrlichia genus. *Vet. Microbiol.* **2013**, *167*, 455–458. [[CrossRef](#)] [[PubMed](#)]
54. Mavromatis, K.; Doyle, C.K.; Lykidis, A.; Ivanova, N.; Francino, M.P.; Chain, P.; Shin, M.; Malfatti, S.; Larimer, F.; Copeland, A.; et al. The genome of the obligately intracellular bacterium Ehrlichia canis reveals themes of complex membrane structure and immune evasion strategies. *J. Bacteriol.* **2006**, *188*, 4015–4023. [[CrossRef](#)]
55. Sehn, J.K. Chapter 9—Insertions and Deletions (Indels). In *Clinical Genomics*; Kulkarni, S., Pfeifer, J., Eds.; Academic Press: Boston, MA, USA, 2015; pp. 129–150. [[CrossRef](#)]
56. Sharp, A.J.; Locke, D.P.; McGrath, S.D.; Cheng, Z.; Bailey, J.A.; Vallente, R.U.; Pertz, L.M.; Clark, R.A.; Schwartz, S.; Segreaves, R.; et al. Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **2005**, *77*, 78–88. [[CrossRef](#)] [[PubMed](#)]
57. Freeman, J.L.; Perry, G.H.; Feuk, L.; Redon, R.; McCarroll, S.A.; Altshuler, D.M.; Aburatani, H.; Jones, K.W.; Tyler-Smith, C.; Hurles, M.E.; et al. Copy number variation: New insights in genome diversity. *Genome Res.* **2006**, *16*, 949–961. [[CrossRef](#)] [[PubMed](#)]
58. Yokoyama, T.T.; Kasahara, M. Visualization tools for human structural variations identified by whole-genome sequencing. *J. Hum. Genet.* **2020**, *65*, 49–60. [[CrossRef](#)] [[PubMed](#)]
59. Makovets, S.; Titheradge, A.J.; Murray, N.E. ClpX and ClpP are essential for the efficient acquisition of genes specifying type IA and IB restriction systems. *Mol. Microbiol.* **1998**, *28*, 25–35. [[CrossRef](#)]
60. Lin, M.; Xiong, Q.; Chung, M.; Daugherty, S.C.; Nagaraj, S.; Sengamalay, N.; Ott, S.; Godinez, A.; Tallon, L.J.; Sadzewicz, L.; et al. Comparative Analysis of Genome of Ehrlichia sp. HF, a Model Bacterium to Study Fatal Human Ehrlichiosis. *BMC Genom.* **2021**, *22*, 11. [[CrossRef](#)]
61. Nikolaev, S.; Montoya-Burgos, J.I.; Margulies, E.H.; Rougemont, J.; Nyffeler, B.; Antonarakis, S.E. Early history of mammals is elucidated with the ENCODE multiple species sequencing data. *PLoS Genet.* **2007**, *3*, e2. [[CrossRef](#)]
62. Younan, M.; Ouso, D.O.; Bodha, B.; Keitany, E.K.; Wesonga, H.O.; Sitawa, R.; Kimutai, J.; Kuria, W.; Sake, W.S.; Svitek, N.; et al. Ehrlichia spp. close to Ehrlichia ruminantium, Ehrlichia canis, and “Candidatus Ehrlichia regneryi” linked to heartwater-like disease in Kenyan camels (Camelus dromedarius). *Trop. Anim. Health Prod.* **2021**, *53*, 147. [[CrossRef](#)]

**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.