Molecular characteristics of *Yersinia enterocolitica* and related species in Ningxia, China based on whole-genome sequencing

Yuan Yue
Ningxia University https://orcid.org/0000-0003-1418-8051
Mei Shen
CDC: Centers for Disease Control and Prevention
Xiang Liu
CDC: Centers for Disease Control and Prevention
Qiong Hao
CDC: Centers for Disease Control and Prevention
Shuai Xu
NIEH China CDC: Chinese Center for Disease Control and Prevention National Institute of Environmental Health
Zhiguo Liu
NIEH China CDC: Chinese Center for Disease Control and Prevention National Institute of Environmental Health
Xuexin Hou
NIEH China CDC: Chinese Center for Disease Control and Prevention National Institute of Environmental Health
Huaiqi Jing
NIEH China CDC: Chinese Center for Disease Control and Prevention National Institute of Environmental Health
Xuezhang Zhou
Ningxia University
Zhenjun Li (lizhenjun@icdc.cn)
NIEH China CDC: Chinese Center for Disease Control and Prevention National Institute of Environmental Health

Research Article

Keywords: Yersinia enterocolitica, Whole-Genome Sequencing, Molecular Characteristics, Sequence types (STs), Core Genome multilocus sequence typing (cgMLST)

Posted Date: April 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1558815/v1

Abstract

Background: *Yersinia enterocolitica* is a common zoonotic pathogen and has emerged as an important foodborne pathogen worldwide. *Y. enterocolitica* has been sporadically recovered from animals, foods, and human clinical samples in various regions of Ningxia. Molecular characteristics of *Y. enterocolitica* in the Ningxia Hui Autonomous Region was not well described in current studies.

Methods: The molecular characteristics of 271 *Yersinia* strains isolated from the Ningxia Hui Autonomous Region were analyzed using SNP and core gene multilocus sequence typing (cgMLST). The isolates were taken from animals (n = 208), food (n = 50), and patients (n = 13) between 2007 and 2019.

Results: The SNP results indicated that the 271 isolates were divided into 12 species, *Y. enterocolitica* (n=187), *Y. intermedia* (n=31), *Y. massiliensis* (n=30), *Y. mollaretii* (n=7), *Y. pekkanenii* (n=5), *Y. proxima* (n=4), *Y. alsatica* (n=2), *Y. frederiksenii* (n=1), *Y. kristensenii* (n=1), *Y. hibernica* (n=1), *Y. canariae* (n=1), and *Y. rochesterensis* (n=1). Of the 187 *Y. enterocolitica* isolates, 81.28% (n = 152) were of animal origin, the food source was 12.30% (n = 23) and the patient source was 6.42% (n = 12). The clustering of *Y. enterocolitica* strains were closely related to the biotype. In total, 187 *Y. enterocolitica* strains were classified into 54 sequence types (STs) and 125 cgMLST types (CTs). The most common STs were ST429, which were significantly associated with serotype O:3. 4/O:3 was the dominant bioserotype in the Ningxia region, which differed from most of China. The 125 CTs clustered to form 55 microclades of the HC100 (hierarchical clustering).

Conclusion: The results indicated that the distribution of *Y. enterocolitica* in the Ningxia Hui Autonomous Region was both general and specific and differed to some extent from the distribution elsewhere in China and abroad. The difference in the distribution of virulence genes between pathogenic and non-pathogenic isolates was attributed to the presence or absence of the type III secretion system. Swine farming and meat food cause human infection and transmission.

Background

Yersinia enterocolitica is a common zoonotic pathogen that is widely found in soil, water, animals, and various foods [1, 2]. It is also an essential foodborne pathogen, transmitted mainly through direct contact between humans and animals or the environment [3]. Enteric yersiniosis does not require antibiotic treatment, and symptoms of the gastrointestinal tract usually resolve within 2–3 weeks after infection of the body. However, *Y. enterocolitica* occasionally causes primary or secondary infections, including erythema nodosum, *Yersinia* enteritis, arthritis, glomerulonephritis, mesenteric lymphadenitis, and sepsis [4].

Y. enterocolitica is heterogeneous and consists of six biotypes: 1A, 1B, 2, 3, 4, and 5. Historically, biotype 1B was regarded as highly pathogenic, and biotype 1A was regarded as being non-pathogenic in humans; other biotypes are considered to have low pathogenicity [5]. O:3, O:8, O:9, and O:5, 27 are the most common clinical serotypes [6]. Furthermore, *Y. massiliensis, Y. frederiksenii*, *Y. intermedia*, *Y. kristensenii*, *Y. bercovieri*, *Y. mollaretii*, *Y. rohdei*, *Y. ruckeri*, and *Y. aldovae* are sometimes referred to as *Y. enterocolitica*-like bacteria [7]. Their biochemical phenotype—which is distinguishable based on fermentation of sucrose, L-rhamnose, raffinose, and melibiose—is extremely similar to that of *Y. enterocolitica* [8].

Enteric yersiniosis has been reported globally, with the most severe epidemic in Europe, where it was the third most common zoonotic disease in 2019 [9]. The United Kingdom Standards for Microbiology Investigations of Faecal

Specimens for Enteric Pathogens state that testing of fecal specimens for *Yersinia* species is only recommended when clinical suspicion has been raised [10]. As a consequence, the incidence of gastrointestinal disorders caused by *Y. enterocolitica* in the UK was underestimated [11]. In the mid-1980s, there were 2 outbreaks of *Y. enterocolitica* in China, resulting in more than 500 infections [12]. The cessation of surveillance of the disease in China for nearly 20 years has resulted in *Y. enterocolitica* infection rarely being considered by clinicians in the diagnosis of gastrointestinal diseases, and testers rarely being able to provide a basis for accurate diagnosis [13]. It was not until 2016 that *Y. enterocolitica* was included in the national food contaminant surveillance network [14]. Such cases reflect a general lack of attention to the disease at a domestic and international level, which can lead to untimely treatment and misdiagnosis, resulting in chronic and prolonged infection. As such, research on *Y. enterocolitica* is essential and urgent.

The surveillance survey results of diseases caused by *Y. enterocolitica* in China through consecutive years have indicated that its hosts include livestock, poultry, and rodents [15]. The geographical distribution varies significantly, and its prevalence is closely related to the distribution and activity of local animals [15]. Pigs and dogs are an important source of *Y. enterocolitica* infections in humans due to the work involved in pig farming and the close relationship between dogs and humans [16, 17]. The processing and consumption of pork food can also contribute to the infection and transmission of *Y. enterocolitica* in the population. However, information about the epidemiology of *Y. enterocolitica* in Ningxia is lacking. *Y. enterocolitica* has been sporadically recovered from animals, foods, and human clinical samples in this area. Consequently, it is urgent and necessary to conduct similar studies in Ningxia to enhance prevention and control.

The current molecular typing methods used for *Yersinia* include pulsed-field gel electrophoresis (PFGE) [18], multiple-locus variable-number tandem-repeat analysis (MLVA) [19], and multilocus sequence typing (MLST) [20]. Since the above-listed typing methods are based on segmental sequence information, they limit the discriminatory ability and phylogenetic information to ensure accuracy. CgMLST is a critical standard for the use of genomic sequence data in clinical microbiology [21]. It has been used for the classification and identification of a wide range of pathogenic bacteria because of its advantages of automation, standardization, and reproducibility [21–23]. Through SNP and cgMLST, this study aims to analyze the distribution and molecular characteristics of *Y. enterocolitica* and related species isolated in Ningxia from 2007 to 2019, with a view to laying the foundation for the traceability and control of related diseases.

Methods

Isolate Collection

The 271 isolates in this experiment were obtained from the Centre for Disease Control and Prevention of the Ningxia Hui Autonomous Region. Strain enrichment has been performed in phosphate-buffered saline with sorbitol and bile salts (PSB) at 4°C for 21 days, and *Yersinia* has been inoculated onto selective agar (CIN Agar; Oxoid, Basingstoke, UK/HKM, Guangzhou, PRC). A typical bulls-eye appearance (deep red centers surrounded by outer transparent zones) on CIN-selective agar plates has been inoculated onto Kligler iron and urea media. We have been performed the identification of the strains using the API20E biochemical identification system. The strains were further distinguished by serotyping (*Y. enterocolitica* antisera set from the Institute of Chinese Biomedicine) [24] and biotyping (Bile Aesculin Agar; Oxoid, Basingstoke, UK. Brain Heart Infusion Agar; Oxoid, Basingstoke, UK. Tween 80; Amresco, USA. $CaCl_2$; Sinopharm Chemical Reagent Co., Ltd, PRC. Biochemical Reaction Tablets; Rosco, Denmark) [25]. The biotype was determined by biochemical experiments on lipase activity, salicin, esculin hydrolysis, xylose,

trehalose, indole production, ornithine decarboxylase, Voges-Proskauer test, pyrazinamidase activity, sorbose, inositol and nitrate reduction [25, 26].

Whole-Genome Sequencing and Assembly

Genomic DNA was extracted using a Wizard® Genomic DNA Purification Kit (Promega, United States) according to the manufacturer's protocol. DNA concentration, quality, and integrity were determined using a 5400 Fragment Analyzer System (Agilent, United States) and a NanoDrop 2000 Spectrophotometer (Thermo, United States). All 271 isolates were sequenced using Illumina Nova technology. Raw reads were quality filtered and assembled. The conditions for filtering were as follows:1) reads containing more than 40% low-quality bases (mass value ≤20) were removed;2) reads with N beyond a certain proportion (the default was 10%) were removed; 3) reads that overlapped with the adapter sequence by more than 15 bp with fewer than 3 mismatches between them were removed. Assemblies were performed using SOAP denovo (http://soapdenovo2.sourceforge.net/) [27], SPAdes (http://bioinf.spbau.ru/spades), and Abyss (https://www.bcgsc.ca/resources/software/abyss) software. Assembly results were integrated using CISA software (http://sb.nhri.org.tw/CISA/en/CISA). The preliminary assembly results were complemented with gapclose software to filter reads with low sequencing depths (less than 0.35 of the average depth) to remove homolane contamination, resulting in the final assembly results. The reference genes of *Yersinia* were *Y. enterocolitica* (GCA_000009345.1), *Y. intermedia* (GCF_009730055.1), *Y. massiliensis* (GCF_000312485.1), *Y. mollaretii* (GCF_013282725.1), *Y. frederiksenii* (GCF_000754805.1), *Y. kristensenii* (GCA_900460525.1), *Y. hibernica* (GCF_004124235.1), *Y. canariae* (GCF_009831415.1), *Y. rochesterensis* (GCF_003600645.1), *Y. nurmii* (GCF_001112925.1), *Y. pestis* (GCF_000222975.1), *Y. ruckeri* (GCF_017498685.1), *Y. aldovae* (GCA_000834395.1), *Y. aleksiciae* (GCF_001319845.1), *Y. bercovieri* (GCA_001319545.1), *Y. entomophaga* (GCF_001656035.1), *Y. pekkanenii* (GCF_001152565.1), *Y. pseudotuberculosis* (GCA_900637475.1), *Y. rohdei* (GCF_000834455.1), and *Y. wautersii* (GCA_001319825.1), *Y. thracica* (GCF_902170565.1), *Y. artesiana* (GCF_902726545.1), *Y. vastinensis* (GCF_902726565.1), *Y. alsatica* (GCF_902170305.1), *Y. proxima* (GCA_902170785.1), and *Y. similis* (GCA_000582515.1).

Average Nucleotide Identity (ANI) Estimation

The ANI was estimated for each pair of isolates using the 271 *Yersinia* genomes (see above) and fastANI v1.1 [28]. The reference genes for the *Yersinia* were the same as those listed in the previous section.

Clustering Analyses

Y. enterocolitica 8081 (GCA_000009345.1) was used as the reference sequence, and the whole-genome sequences of 271 isolates were compared with the reference sequence using MUMmer 4.0 software (http://mummer.sourceforge.net/) [29]. All SNP loci of each strain were obtained. The following protocol was used: Integrate the position data of all SNP loci in the reference genome, filter SNPs from duplicated regions, and remove low-quality SNPs for subsequent analysis [30,31]. Maximum Likelihood trees were constructed using iQ-TREE (http://www.iqtree.org/) with 1000 bootstrap replicates. Then, isolates identified as *Y. enterocolitica* (n=187) by ANI analysis were used to construct a maximum likelihood tree using *Y. enterocolitica* 8081 (GCA_000009345.1) as the reference genome. The trees were then visualized and annotated in iTOL (using their associated scripts that were obtained after processing the previous series) [31].

Virulence Profiles

The assembled genomes of the isolates in this study were inserted into the Virulence Factor Database (VFDB; MOH Key Laboratory of Systems Biology of Pathogen, Institute of Pathogen Biology, Beijing, http://www.mgc.ac.cn).

ST Determination

Sequence types (STs) of *Y. enterocolitica* were assigned using EnteroBase (http://enterobase.warwick.ac.uk/species/index/yersinia) [32]. We chose the McNally 7 Gene scheme, which contained *aar*F, *dfp*, *gal*R, *gln*S, *hem*A, *rfa*E, and *spe*A [33]. Sequence data of the isolates were extracted from their genome data. The Minimum Spanning Trees (MST) of 187 *Y. enterocolitica* isolates were constructed based on STs, biotype, and serotype. In addition, the whole genome sequences of 87 publicly available *Y. enterocolitica* strains were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/). These strains were isolated from 10 countries: the United Kingdom (n = 36), New Zealand (n = 19), France (n = 10), Germany (n = 8), the United States (n = 4), Italy (n = 4), Ireland (n = 2), Spain (n = 1), Greece (n = 1), and Australia (n = 1); and 1 unspecified as shown in Table S2. MSTs were constructed using strains from the Ningxia region and those publicly available in the NCBI database.

Core Genome MLST *in silico* Subtyping

Y. enterocolitica isolates were typed using cgMLST implemented in EnteroBase (http://enterobase.warwick.ac.uk/species/index/yersinia): cgMLST V1+HeirCC V1 (Hierarchical cgMLST clustering). It is based on the profiles of 1553 coding loci in the EnteroBase *Yersinia* cgMLST scheme. For cgMLST analysis, generation and annotation of an NJ tree or an MST were based on HierCC data with GrapeTree.

Data availability

FASTQ datafiles have been deposited in NCBI-SRA and public archives under the project's accession number PRJNA993921 (Centers for Disease Control and Prevention). The accession numbers of all isolates are provided in Table S1.

Results

Review of the Historical Data for Y. enterocolitica and Related Species

The strains were obtained from five prefectures in Ningxia Hui Autonomous Region [Yinchuan (n = 112), Shizuishan (n = 7), Wuzhong (n = 1), Guyuan (n = 3), and Zhongwei (n = 148)], from 2007 to 2019 (Fig. 1 and Table S1).

In this study, 271 *Y. enterocolitica* and related species isolated from animals, foods, and human clinical samples were analyzed. In total, 208 (76.75%) were of animal origin, 50 (18.45%) were of food origin, and 13 (4.78%) were of patient origin. The source of animal samples was mainly feces (n = 102), pharyngeal swabs (n = 39), anal swabs (n = 7), and intestinal contents (n = 59) of animals. The food samples were obtained mostly from meat. The origin of patient samples was feces.

Using traditional phenotypic methods, 130/271 (47.97%) isolates were serotyped, and the most common serotypes were O:3 (n = 73), O:5 (n = 30), O:8 (n = 14), and O:9 (n = 9). In total, 148 isolates were reported as O: unidentifiable because the O-antigen reacted with more than one antiserum or with none of the antisera. Animal hosts included pig (n = 150), sheep (n = 32), rat (n = 15), cattle (n = 6), chicken (n = 3), and hamster (n = 2). Food was derived from meat products, comprising beef (n = 25), pork (n = 11), chicken (n = 9), lamb (n = 3), and fish (n = 2). Food (n = 50)

comprised fresh meat (n = 19) and frozen meat (n = 31). Human samples were of fecal origin, the majority of which were from children (n = 9) (Table S1).

Of the 187 *Y. enterocolitica* isolates, 81.28% (n = 152) were of animal origin, the food source was 12.30% (n = 23) and the patient source was 6.42% (n = 12). Isolates of animal origin included 42.76% biotype 1A (n = 65), 50.0% biotype 4 (n = 76), 2.63% biotype 3 (n = 4) and 4.61% biotype 5 (n = 7). Of these, all isolates of biotype 5 were from sheep. In total, biotype 4 of isolates (n = 84) were from pig hosts. The dominant serotype of the isolates of animal origin was O:3. Food-derived isolates were 86.96% biotype 1A (n = 20) and 13.04% biotype 2 (n = 3). Patient-origin strains included 33.33% biotype 1A (n = 4) and 66.67% biotype 4 (n = 8) (Table 1).

Table 1
Serotype and biotype distribution of *Y.enterocolitica* isolates

		Animal (n = 152)					Food	Human
		Pig (n = 121)	**Sheep (n = 17)**	**Cattle (n = 6)**	**Rat (n = 5)**	**Chicken (n = 3)**	(n = 23)	(n = 12)
Biotype	**1A**	41/33.88%	10/58.82%	6/100%	5/100%	3/100%	20/86.96%	4/33.33%
	2	——	——	——	——	——	3/13.04%	——
	3	4/3.31%	——	——	——	——	——	——
	4	76/62.81%	——	——	——	——	——	8/66.67%
	5	——	7/41.18%	——	——	——	——	——
Serotype	**O:3**	72/59.50%	——	——	——	——	——	1/8.33%
	O:5	17/14.05%	2/11.76%	2/33.33%	——	2/66.67%	3/13.04%	——
	O:8	1/0.83%	8/47.06%	——	——	——	4/17.39%	1/8.33%
	O:9	3/2.48%	2/11.76%	——	——	——	2/8.70%	1/8.33%
	O:53	——	——	——	——	——	1/4.35%	——
	O:1,2,5	——	——	——	——	——	——	——
	O:5,8,9	——	1/5.88%	——	——	——	——	1/8.33%
	NA	28/23.14%	4/23.53	4/66.67%	5/100%	1/33.33%	13/56.52%	8/66.67%

Note: NA, not applicable and nonagglutinative.

Average Nucleotide Identity (ANI) Estimation

The 271 *Yersinia* genomes were evaluated according to the suggested 95–96% ANI [34]. 12 species were delineated using a 95% ANI cut-off value: *Y. enterocolitica, Y. intermedia, Y. massiliensis, Y. mollaretii, Y. pekkanenii*, *Y. proxima*, *Y. alsatica, Y. frederiksenii, Y. kristensenii, Y. hibernica, Y. canariae, and Y. rochesterensis*.

Biotype identification

The biotypes of *Y. enterocolitica* were identified by traditional biochemical experiments, which revealed biotype 1A (n = 89), biotype 2 (n = 3), biotype 3 (n = 4), biotype 4 (n = 84), and biotype 5 (n = 7). There was no biotype 1B (Table 3S).

Clustering Analyses

After filtering and screening, we constructed a maximum likelihood (ML) tree based on 1,563,073 SNPs. The best-fitting model of the ML tree was GTR + F + ASC + G4, chosen according to BIC (Bayesian information criterion). The ML tree clearly presented 271 isolates comprising 187 *Y. enterocolitica* (69.0%), 31 *Y. intermedia* (11.44%), 30 *Y. massiliensis* (11.07%), 7 *Y. mollaretii* (2.58%), 5 *Y. pekkanenii* (1.85%), 4 *Y. proxima* (1.48%), 2 *Y. alsatica* (0.74%), 1 *Y. frederiksenii* (0.37%), 1 *Y. kristensenii* (0.37%), *Y. hibernica* (0.37%), *Y. canariae* (0.37%), and *Y. rochesterensis* (0.37%) isolates (Fig. 2a).

Notably, the 4 isolates closest to the *Y. proxima* reference sequence and the *Y. artesiana* reference sequence clustered together with *Y. enterocolitica*. *Y. artesiana* and *Y. proxima* are subspecies of *Y. enterocolitica*, first identified by Savin et al. (Savin et al., 2019) and named NEW3 and NEW4, respectively. In this study *Y. enterocolitica* and the reference genome (GCA_000009345.1) refer to *Y. enterocolitica* subsp. *enterocolitica*. The results of its ANI analysis with both *Y. artesiana* and *Y. proxima* were < 95%.

The ML tree showed accurate clustering separation of the *Yersinia* species, consistent with the results of the ANI analysis and identical separation into 12 distinct species as determined by BAPS (Fig. 2a). The identification of *Yersinia* species before whole genome sequence techniques was largely based on limited biochemical data. The information presented by the ML tree revealed the inadequacy of the resolution provided by biochemical tests and emphasizes the need to use modern molecular methods for the classification of bacterial species.

In accordance with *Y. enterocolitica*, the ML tree was broadly divided into two clades. Biotype1A and 4 formed discrete clusters, whereas biotypes 2–3 consisted of closely related but distinct lineages, confirmed by BAPS clustering. Biotype 5 was divided into a separate branch relative to biotype 1A and biotype 4. The reference sequence *Y. enterocolitica* 8081 (biotype 1B) was divided into a separate branch as there was no biotype 1B among the 187 *Y. enterocolitica* isolates. The distribution of the strains was not directly related to the time, place, or host of isolation. It was evident from the ML tree that clustering was related to the distribution of the biotypes. This result was identical to the findings of Reuter et al. (Reuter et al., 2014). They classified *Y. enterocolitica* into two species clusters (SC), SC6 (biotypes 1A, 1B) and SC7 (biotypes 2, 3, 4, and 5). Biotype 4 was correlated with serotype O:3, and biotype 1A was associated with serotypes O:5, O:8, and O:9. Biotypes 3, 4, and 5 displayed tight clusters with short terminal branches compared with biotypes 1A and 2. One possible explanation is that they were the product of one or more recent population bottlenecks or population expansions (Fig. 2b).

Virulence Profiles

In total, 187 *Y. enterocolitica* isolates were annotated to 99 virulence genes in 4 categories: motility, invasion, immune modulation, and effector delivery system (Figure 3). As can be seen from Figure 3, the distribution of virulence genes was closely related to the biotype. Obviously, biotypes 3, 4, and 5 had virulence genes associated with the type III secretion system (T3SS) compared to biotypes 1A.

Type III secretion system

The effector delivery system was T3SS, which includes Ysc T3SS and Ysa T3SS. T3SS is a pathogenic strategy common to numerous gram-negative pathogens in plants and animals to inject virulence proteins into the cytoplasm of target eukaryotic cells [37]. When members of *Yersinia* cause disease in humans, they invariably employ the type III secretion machinery to inhibit the primary line of immunological defense in vertebrates, the professional phagocytic cells [38].

Yersinia Outer Proteins and Other Proteins Encoded by Low-Calcium Response Stimulons In very early work on *Yersinia* physiology, a phenomenon known as the low-calcium response (LCR) was identified, whereby bacteria grown in rich media at elevated temperatures (37°C) would show a growth defect when calcium ions were chelated. However, at 26°C, growth would continue logarithmically regardless of the calcium content, thus indicating a strong dependence of temperature on the regulation of LCR. *Y. enterocolitica* contains a highly conserved region of low calcium response-stimulating proteins (LCRS). Virulence genes in the LCRS region include the LcrV genes encoding the V antigen (LcrV) [39] and yadA [40]. These genes are most abundantly expressed at low calcium concentrations, resulting in growth restriction in vitro [38]. The secretion of *Yersinia* outer proteins (Yops) involves up to 22 proteins (YscA-L, YscN-U, LCRD, LCRP/YscM) [41]. Yops are especially important in enteropathogenic *Yersiniae* and require effective bacterial adherence to target host cells for expression and functional deployment [41].

Heat-stable enterocolitica toxin The heat-stable *enterocolitica* toxin (Yst) is chromosomally mediated [42] and acts by stimulating guanylate cyclase in intestinal epithelial cells [43]. Delor and Cornelis [44] demonstrated that Yst may be an important factor involved in *Yersinia enterocolitica*-associated diarrhea in young rabbits.

FliA Motility genes were used for flagellar assembly. Before *Y. enterocolitica* establishes an intimate contact with the intestinal epithelium, flagella and motility play an important role in initiating host cell invasion [37]. The flagellar regulatory genes *flh*DC or *fli*A are both required for motile expression. The *fli*A gene, present in *E. coli* and *Salmonella typhimurium*, encodes another sigma factor, Ko^{28}, also known as RPOF [45]. Inhibition of motility at 37°C involves changes in DNA topology [42]. Iriarte et al. demonstrated loss of motility but no effect on pathogenicity in *fli*A mutants [45].

InvA

The invasive factor is the *inv* gene, which plays a crucial role in the initial stages of an intestinal mucosal invasion. It encodes the 103 kDa Inv protein invasin [46]. Invasin, an outer membrane protein that coats the bacteria, binds directly to integrin receptors on the cell surface and mediates cell entry and internalization of the bacterium. This process has been likened to a “zipper mechanism”[46]. Upon binding, the invasin ligand may induce a conformational change in the receptor that initiates a cascade of intracellular signals leading to uptake [47]. However, invasin promotes bacterial attachment to the extracellular matrix proteins fibronectin and collagen by binding with high affinity to β1 integrin [40]. Moreover, integrin receptors appear to be ideally suited for the process of polymerization of microfilaments during the entry process.

Immune modulation

The major gene of the immune modulation is the O antigen, the primary component of the lipopolysaccharide contained in the outer membrane of Gram-negative bacteria that is required for the correct expression or function of other outer membrane virulence factors [48]

ST Typing

Post-WGS implementation, the STs were derived from the genome data. Sequence typing data were available for all 187 *Y. enterocolitica* isolates. Multilocus sequence typing detected 54 STs in 187 isolates identified in this study. The most common STs were ST429 at 42.25% (79/187); ST3 at 4.81% (9/187); ST13 at 3.74% (7/187); ST278 at 3.21% (6/187); and ST178 and ST637 at 2.67% (5/187) (Table S3). As shown in Fig. 4, MST presented two main centers, ST429 and ST13. This was consistent with the results of *Y. enterocolitica* ML tree. In addition, biotype 1A was more abundantly polymorphic than biotype 4. ST429 included biotype 4 (n = 79) and biotype 3 (n = 2) and was closely related to serotype O:3. ST3 (n = 9), ST278 (n = 6), ST178 (n = 5), ST637 (n = 5), ST640 (n = 4), ST643 (n = 4), and ST216 (n = 3) were biotype 1A and were associated with serotypes O:5, O:8, O:9. ST13 (n = 7) was biotype 5, again associated with serotypes O:5, O:8, and O:9. These results were consistent with the findings of Hunter et al. (2019). In a survey of *Y. enterocolitica* and *Y. pseudotuberculosis* isolated from clinical specimens in the UK from 2004 to 2018, Hunter et al. (2019) found that certain STs correlated with specific serogroups, including ST18/O:3, ST12/O:9, ST184/O:6,30, and ST192/O:8. ST12 and ST4 were the two sources in the MST constructed from strains from the Ningxia region and those publicly available in the database (Table S2 and Fig. 4b). Bioserotypes 1A/O:6,30 and 2,3,4/O:9 were the biotypes and serotypes corresponding to these two sources. The strong correlation between phylogenetic signals and serotypes may suggest that the early evolution of *Y. enterocolitica* was dominated by ecological specialization.

Core Genome MLST

A neighbor-joining (NJ) tree and an MST based on cgMLST analysis of 187 *Y. enterocolitica* isolates were constructed. These 1,553 cgMLST target genes were randomly distributed across the genome, encoding functional enzymes and proteins. *Y. enterocolitica* isolates were divided into 125 cgMLST types (CTs) (Table S3 and Fig. 5a). CgMLST analysis revealed the core genome diversity of strains with the same ST from 0 to 84 allelic differences. The same cgMLST distribution was present between the different STs. The NJ tree of 187 *Y. enterocolitica* isolates from Ningxia indicated the names of two microclades of the HC1490 cluster (Fig. 5a). HC1490_10 and HC1490_2 were the primary microclades. These two microclades of the HC1490 cluster were consistent with the results of *Y. enterocolitica* ML tree. HC1490_2 were strongly associated with biotypes 3, 4, and 5; HC1490_10 was closely related to biotype 1A and 2.

Hierarchical clustering (HC) of CgMLST (HierCC) defines clusters based on cgMLST. Distances between genomes are calculated using the number of shared cgMLST alleles, and genomes are linked on a single-linkage clustering criterion. These clusters were assigned stable cluster group numbers at different, fixed cgMLST allele distances. *Yersinia*, for instance, had cut-offs such as 0, 2, 5, 10, 20, 50, 100, and 200, etc. After comparison of the clustering of HC50, HC100, HC200, HC400, and HC600, HC100 was determined as the criterion for CT clustering. HC100 indicates that the clusters included all strains with links no more than 100 alleles apart. The 125 CTs present in the 187 isolates from the Ningxia region clustered to form 54 microclades of HC100. Of these, HC100_2571, HC100_150, HC100_1273, and HC100_4570 were the principle microclades. The NJ tree constructed from strains from the Ningxia region and those publicly available in the database showed that several microclades of the HC100 cluster were significantly associated with serotypes such as: HC100_397/O:5,27; HC100_111/O:9; HC100_4570 and HC100_466/O:8; and HC100_150/O:5. O:3 was related to three microclades of the HC100 cluster: HC100_406, HC100_7, and HC100_2571 (Fig. 5b). Biotypes 1B and 5 were located in relatively separate microclades. Biotypes 2, 3, and 4 were divided into several microclades. Interestingly, individual microclades were host-specific, for instance, HC100_4570/sheep, HC100_111/pig and human, HC100_2571/pig and human, biotype 1B/human. The NJ tree suggested that cgMLST significantly improved the phenotyping and identification of *Y. enterocolitica* compared to

STs. This method allowed for more accurate discrimination of sample information such as the source compared to ST typing.

Discussion

In the present research, we used SNP and cgMLST to study the molecular and phylogenetic characteristics of *Y. enterocolitica* and related species. ANI analysis was used to confirm strains that were misclassified as *Y. enterocolitica* through biochemical tests. Clustering analysis of *Y. enterocolitica* and related species was investigated by SNP. Identification and typing of *Y. enterocolitica* at the subspecies biotype level were performed using cgMLST. The results showed that the Ningxia isolates of biotype 1A had more abundant genetic polymorphisms than biotypes 2, 3, 4, and 5. Phylogenetic analysis and molecular typing were closely related to biotypes and serotypes.

As evidenced by the *Y. enterocolitica* ML tree, clustering was tightly linked to biotype. *Y. enterocolitica* is clinically important in distinguishing between non-pathogenic organism carriers, self-limiting infections with low pathogenic strains, or more aggressive, highly pathogenic strain types of infection. Correct clustering is therefore essential in both epidemiological tracing and subtype identification. Among the pathogenic isolates in this study, 4/O:3 was the dominant biotype, which differed from the findings of Mu et al. [49] and Xiao et al. [26] (3/O:3). In addition, Mu et al. [49] demonstrated that pathogenic strains in China were only found in serotypes O:3 and O:9. In the present study, however, pathogenic strains were found to be distributed in serotype O:5 in addition to serotypes O:3 and O:9.

There were two MLST protocols for *Yersinia* in the EnteroBase. One scheme is the Achtman 7 Gene, which included the seven housekeeping genes *adk*, *arg*A, *aro*A, *gln*A, *thr*A, *tmk*, and *trp*E [35]. The other is the McNally 7 Gene, which contained *aar*F, *dfp*, *gal*R, *gln*S, *hem*A, *rfa*E, and *spe*A [33]. We chose the McNally 7 Gene scheme because it contained more STs. Since traditional typing methods are based on segmental sequence information, they limit the discriminatory ability and phylogenetic information to ensure accuracy.

CgMLST is a critical standard for the use of genomic sequence data in clinical microbiology. It has been used for the classification and identification of a wide range of pathogenic bacteria because of its advantages of automation, standardization, and reproducibility [50.51]. We selected the prospective set of 1,553 core genes from the EnteroBase for evaluation and demonstrated the ability of this approach to identify *Y. enterocolitica* at the intraspecific and subspecific bioserotype levels. CgMLST utilized next-generation sequencing technology to analyze hundreds of genes, which greatly increased discriminatory power. The crucial element of this technology was the selection of the core genes. Generally, core genes were common to all strains within a species. On average, all isolates in the study, except NX19106, had 99.68% (98.6–100%) good core genome targets of the defined cgMLST scheme [52]. Although NX19106 had 91.50% of the core genes, it was available because, in practice, the cgMLST assays allowed for a certain level of gene deletion. The HierCC developed by EnteroBase was optimized for missing genes, and 10% deletion will not affect the analysis significantly.

It is a known fact that pigs serve as the primary reservoir, having a high prevalence of pathogenic *Y. enterocolitica* [14]. There were few allelic differences and a close genetic distance between human-derived strains and strains isolated from food and livestock (especially pigs and sheep) (Fig. 5). This suggested that people involved in pig farming and meat processing were at high risk of *Y. enterocolitica* infection, while those working in cattle, sheep and chicken farming and processing were at low risk of infection, however cross contamination during production and processing was not excluded. These crowds should have enhanced protective measures. Government departments need to strengthen livestock quarantine and health supervision. Meanwhile, pathogenic strains were present in

samples of food origin, indicating that meat products were another leading cause of human infection. As such, we should avoid improperly cooked meat and be alert to the safety risks of meat products in our daily consumption.

A further issue that should be noted was that the five samples isolated from patients with diarrhea were biotype 1A. Virulence genes for flagellar assembly and O antigen formation were annotated from these samples. Generally, that biotype 1A is acknowledged as being non-pathogenic. However, in recent years, some studies have found that 1A is somewhat pathogenic as well [53], which demands attention.

Since Ningxia Hui Autonomous Region is an area with a high concentration of Muslims, pig breeding farms are usually established outside Muslim settlements and are relatively concentrated. The number of fecal samples from patients with diarrhea was low as a consequence of the improvement in farming conditions in recent years, the frequent use of antibiotics, and the high proportion of halal meals in Ningxia. This was also a limitation of this study. The distribution of *Y. enterocolitica* in Ningxia is both universal and specific, as shown by the STs and CTs. The Ningxia Hui Autonomous Region has contributed a highly significant role in the evolution as well as genetic diversity. Consequently, alertness is required with regards to the risk of human infection from domestic animals and food.

Conclusions

CgMLST was a critical standard for the use of genomic sequence data in clinical microbiology. It significantly improved the identification and typing of pathogenic microorganisms. Furthermore, the distribution of *Y. enterocolitica* in the Ningxia Hui Autonomous Region was both general and specific. The difference in the distribution of virulence genes between pathogenic and non-pathogenic isolates was attributed to the presence or absence of the type III secretion system. Swine farming and meat food cause human infection and transmission. Therefore, it is necessary to strengthen efforts such as animal quarantine and food supervision to reduce the risk of *Y. enterocolitica* infection.

Declarations

Acknowledgments

We thank all the staff of the Bacteriology Section of the Ningxia Hui Autonomous Region Centre for Disease Control and Prevention for their help in collecting the strains.

Funding

This work was supported by grants from the National Key R&D Program of China (2019YFC1200700, 2019YFC1200601-6) and the Ningxia Hui Autonomous Region Natural Genetic Project (2022AAC03719).

Author contributios

YY performed the preparation of DNA samples, identification of biotypes, clustering analysis, cgMLST analysis, and drafted the manuscript; SM, LX and HQ mainly performed the strain isolation; XS performed the data analysis; LZG, HXX and JHQ revised the manuscript; ZXZ and LZJ participated in the design of this study and reviewed the manuscript; LZJ managed the project. All authors read and approved the final manuscript.

Ethics approval and consent to participate

This study was carried out according to the principles of the Declaration of Helsinki, the ARRIVE guidelines and the 3Rs.The research protocol was approved by the Ethics Committees of the Chinese Center for Disease Control and Prevention (ICDC-2019015). All strains from humans were collected as part of a standard clinical investigation of patients with diarrhea. These patients were anonymous and had signed an informed consent form. All animal samples used in this study were approved by the Research Animal Ethics and Ethical Committee of the Chinese Centre for Disease Control and Prevention (2020-025). Written informed consent was obtained from the owners for the participation of their animals in this study. The *Yersinia* strains used in this study were obtained from academic articles published in PubMed.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article. The supplementary information files are freely available to any scientist wishing to use them for non-commercial purposes upon request via e-mail with the corresponding author.

Competing interests

No potential conflict of interest was reported by the author(s).

References

1. Ahmed HA, Tahoun A, Elez RA, El-Hamid MA, Ellatif SA. Prevalence of *Yersinia enterocolitica* in milk and dairy products and the effects of storage temperatures on survival and virulence gene expression. International Dairy Journal. 2019;94:16-21.
2. Bari ML, Hossain MA, Isshiki K, Ukuku D. Review Article Behavior of *Yersinia enterocolitica* in Foods. 2013. doi:10.4061/2011/420732
3. Bancerz-Kisiel A, Szweda W. Yersiniosis–zoonotic foodborne disease of relevance to public health. Annals of agricultural and environmental medicine: AAEM. 2015;22(3):397-402.
4. Fallon RJ. Yersiniosis. Lab Diagnosis, Clinical Features, Epidemiology. Journal of Clinical Pathology. 1987;40(4):473.
5. Valentin-Weigand P, Heesemann J, Dersch P. Unique virulence properties of *Yersinia enterocolitica* O:3–An emerging zoonotic pathogen using pigs as preferred reservoir host. International Journal of Medical Microbiology. 2014; 304(7): 824-834.
6. Garzetti D, Susen R, Fruth A, Tietze E, Heesemann J, Rakin A. A molecular scheme for *Yersinia enterocolitica* patho-serotyping derived from genome-wide analysis. International Journal of Medical Microbiology.2014;304(3-4): 275-283.
7. Sulakvelidze A. Yersiniae other than *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*: the ignored species. Microbes & Infection. 2000;2(5):497-513.
8. Bottone EJ. *Yersinia enterocolitica*: the charisma continues. Clinical microbiology reviews.1997;10(2):257.
9. European Food Satety Authority. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016. EFSA Journal. 2017;15(12).

10. UK Standards for Microbiology Investigations B 30: Investigation of faecal specimens for enteric pathogens. Public Health England. 2014. https://www.gov.uk/government/publications/smi-b-30-investigation-of-faecal-specimens-for-enteric-patho-gens
11. Hunter E, Greig DR, Schaefer U,Wright MJ, Jenkins C. Identification and typing of *Yersinia enterocolitica* and *Yersinia pseudotuberculosis isolated from human clinical specimens in England between 2004 and 2018. Journal of Medical Microbiology. 2019; 68(4): 538-548.*
12. Wang X, Qiu H, Dong J, Cui ZG, Kan B, Xiao YC, et al. O:8 serotype *Yersinia enterocolitica* strains in China. International Journal of Food Microbiology. 2008;125(3):259-266.
13. Guo BC, Liu X, Hao Q, Yan L, Xie MY, Xue XY, et al. Monitoring and analysis of *Yersinia enterocolitica* in Ningxia area. Modern medicine and health. 2011;27(018):2724-2726
14. Peng ZX, Zou MY, Xu J, Guan WY, Li Y, Liu DR, et al. Antimicrobial susceptibility and drug-resistance genes of *Yersinia* spp. of retailed poultry in 4 provinces of China. Chin J Prev Med. 2018;52(4):358-363.
15. Liang J, Duan R, Xia S, Hao Q, Yang JC, Xiao YC, et al. Ecology and geographic distribution of *Yersinia enterocolitica* among livestock and wildlife in China. Veterinary Microbiology. 2015;178(1-2):125-131.
16. Duan R, Liang J, Shi G, Cui ZG, Hai R, Wang P, et al. Homology Analysis of Pathogenic Yersinia Species *Yersinia enterocolitica*, *Yersinia pseudotuberculosis*, and *Yersinia pestis* Based on Multilocus Sequence Typing. Journal of Clinical Microbiology. 2014; 52(1):20-29.
17. Stamm I, Hailer M, Depner B, Kopp PA, Rau J. *Yersinia enterocolitica* in diagnostic fecal samples from European dogs and cats: identification by fourier transform infrared spectroscopy and matrix-assisted laser desorption ionization-time of flight mass spectrometry. Journal of Clinical Microbiology. 2013;51(3):887-893.
18. Wang X, Cui ZG, Wang H, Tang LY, Yang JC, Gu L, et al. Pathogenic Strains of *Yersinia enterocolitica* Isolated from Domestic Dogs (Canis familiaris) Belonging to Farmers Are of the Same Subtype as Pathogenic *Y.enterocolitica* Strains Isolated from Humans and May Be a Source of Human Infection in Jiangsu Province, China. 2010; 48(5): 1604-1610.
19. Wang X, Gu WP, Cui ZG, Luo L, Cheng P. Multiple-Locus Variable-Number Tandem-Repeat Analysis of Pathogenic *Yersinia enterocolitica* in China. Plos One. 2012; 7(5):e37309.
20. Laukkanen-Ninios R, Didelot X, Jolley KA, Morelli G, Sangal V, Kristo P, et al. Population structure of the *Yersinia pseudotuberculosis* complex according to multilocus sequence typing. Environmental Microbiology. 2011;13(12):3114-3127.
21. Zhou H, Liu W, Qin T, Chen L, Ren H. Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of *Klebsiella pneumoniae*. Front Microbiol. 2017; 8:371.
22. De BM, Pinholt M, Top J, Bletz S, Mellmann A, Van SW, et al. Core Genome Multilocus Sequence Typing Scheme for High-Resolution Typing of *Enterococcus faecium*. Journal of Clinical Microbiology. 2015; 53(12):3788.
23. Zhang H, Chen W, Wang J, Xu BY. 10-Year Molecular Surveillance of *Listeria monocytogenes* Using Whole-Genome Sequencing in Shanghai, China, 2009–2019. Frontiers in Microbiology. 2020; 11:551020.
24. Wang X, Cui Z, Jin D, Tang L, Xia S, Wang H, et al. Distribution of pathogenic *Yersinia enterocolitica* in China. Eur J Clin Microbiol Infect Dis. 2009; 28(10):1237-1244.
25. Bottone EJ. *Yersinia enterocolitica*: the charisma continues. Clinical microbiology reviews.1997; 10(2):257.
26. Xiao YC, Wang X, Qiu HY, et al. Study of Biotyping for pathogenic *Y.enterocolitica* strains in China. Chinese Journal of Zoonoses. 2010;26(07):651-653.

27. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009;25(15):1966-1967.
28. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications.2018;9:5114.
29. Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to Identify Similar Regions in Large Sequence Sets. Current Protocols in Bioinformatics. 2003;1:10.13.11-10.13.18.
30. Gary B. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research.1999;2:573-580.
31. Ivica L, Peer B. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research. 2016;W1:W242-W245.
32. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Achtman M. The EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny and *Escherichia* core genomic diversity.Genome Research. 2019;30(1):gr.251678.251119.
33. Hall M, Chattaway MA, Reuter S, Savin C, Mcnally A. Use of Whole-Genus Genome Sequence Data To Develop a Multilocus Sequence Typing Tool That accurately Identifies *Yersinia*-January 2015.j.am.chem.soc. 2015;53(1):35-42.
34. Richter M, Rosselló-Móra R, Ckner F, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics. 2015;32(6):929-931.
35. Savin C, Criscuolo A, Guglielmini J, Guern ASL, Carniel E, Pizarro-Cerdá J, et al. Genus-wide *Yersinia* core-genome multilocus sequence typing for species identification and strain characterization. Microbial Genomics. 2019;5(10): e0000301.
36. Reuter S, Connor TR, Barquist L, Walker D, Thomson NR. Parallel independent evolution of pathogenicity within the genus *Yersinia*. Proceedings of the National Academy of Sciences. 2014;111(18):6768.
37. Fàbrega A , Vila J . *Yersinia enterocolitica*: Pathogenesis, virulence and antimicrobial resistance. Enfermedades Infecciosas Y Microbiología Clínica. 2012;30(1):24-32.
38. Ramamurthi KS, Schneewind O. Type III Protein Secretion in *Yersinia* Specie. Annual Review of Cell & Developmental Biology. 2003;18(18):107.
39. Brubaker RR. The V antigen of yersiniae: an overview.Contributions in Microbiology and Immunology. 1991;12: 127-133.
40. Roggenkamp A, Neuberger HR, Flugel A, Schmoll T, Heesemann J, et al. Substitution of two histidine residues in YadA protein of *Yersinia enterocolitica* abrogates collagen binding, cell adherence and mouse virulence. Molecular Microbiology. 1995;16:1207-1219.
41. Straley S, Skrzypek E, Piano GV, Bliska JB. Yops of *Yersinia* spp pathogenic to humans. Infection and Immunity. 1993; 61: 3105-3110.
42. Straley SC, Perry RD. Environmental modulation of gene expression and pathogenesis in *Yersinia*. Trends in Microbiology. 1995;3:310-317.
43. Robins-Browne RM, Still CS, Miliotis MD, Koornhof HJ. Mechanism of action of *Yersinia enterocolitica* enterotoxin. Infection and Immunity. 1979;25:680-684.
44. Delor I, Cornelis R. Role of *Yersinia enterocolitica* Yst toxin in experimental infection of young rabbits. Infection and Immunity. 1992;60:4269-4277.
45. Iriarte M, Stainier I, Mikulskis AV, Cornelis GR. The *fle*A gene encoding σ28 in *Yersinia enterocolitica*. Journal of Bacteriology. 1995;177:2299-2304.

46. Koornhof HJ, Smego RA, Nicol M. Yersiniosis II: The Pathogenesis of *Yersinia* Infections. European Journal of Clinical Microbiology & Infectious Diseases. 1999;18(2):87-112.
47. Isberg RR, Leong JM. Multiple β1 chain integrins are receptors for invasin, a protein that promotes bacteria penetration into mammalian cells. Cell.1990; 60: 861-871.
48. Bengoechea JA, Najdenski H, Skurnik M. Lipopolysaccharide O antigen status of *Yersinia enterocolitica* O:8 is essential for virulence and absence of O antigen affects the expression of other *Yersinia* virulence factors. Molecular Microbiology. 2010;52(2):451-469.
49. Mu H, Duan R, Qin S, Xiao M, He ZK, Lv DY, et al. Distribution and molecular epidemiological characteristics of *Yersinia enterocolitica* in China. Chinese Journal of Zoonoses. 2021;37(07):571-577.
50. Achtman, M. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(24):14043-14048.
51. De BM, Pinholt M, Top J, Bletz S, Mellmann A, Van SW, et al. Core Genome Multilocus Sequence Typing Scheme for High-Resolution Typing of *Enterococcus faecium*. Journal of Clinical Microbiology. 2015; 53(12):3788.
52. Lepuschitz S, Hauser K, Schriebl A, Schlagenhaufen C, Ruppitsch W. Fecal *Klebsiella pneumoniae* Carriage Is Intermittent and of High Clonal Diversity. Frontiers in Microbiology. 2020;11:581081.
53. Singh I, Virdi JS. Interaction of *Yersinia enterocolitica* biotype 1A strains of diverse origin with cultured cells in vitro. Japanese Journal of Infectious Diseases. 2005;58(1): 31-33.

Figures

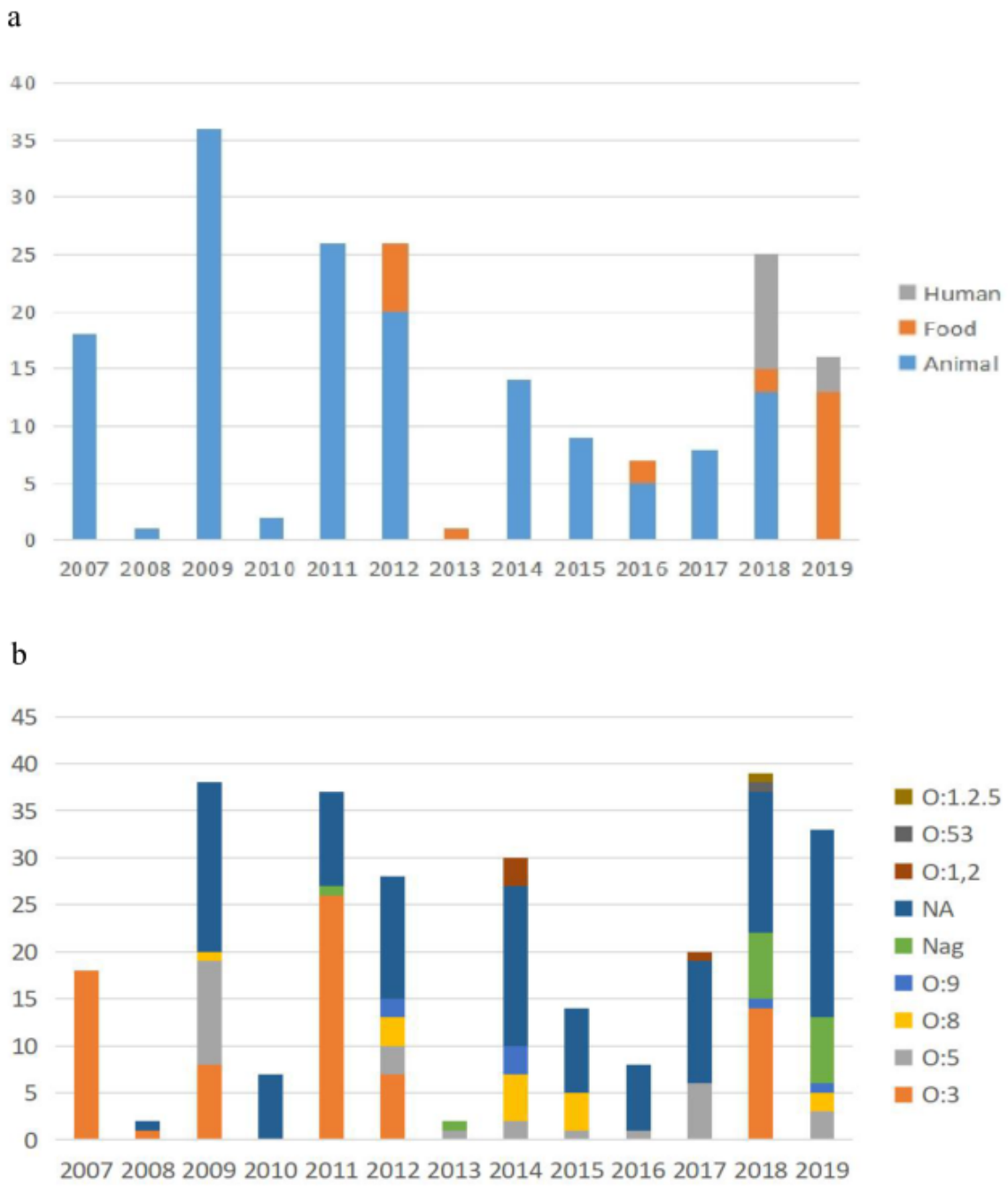

Figure 1

Number of *Yersinia* spp. strains isolated from the Ningxia Hui Autonomous Region between 2007 and 2019 by the host (**a**) and serotype (**b**). NA, not applicable; Nag, nonagglutinative.

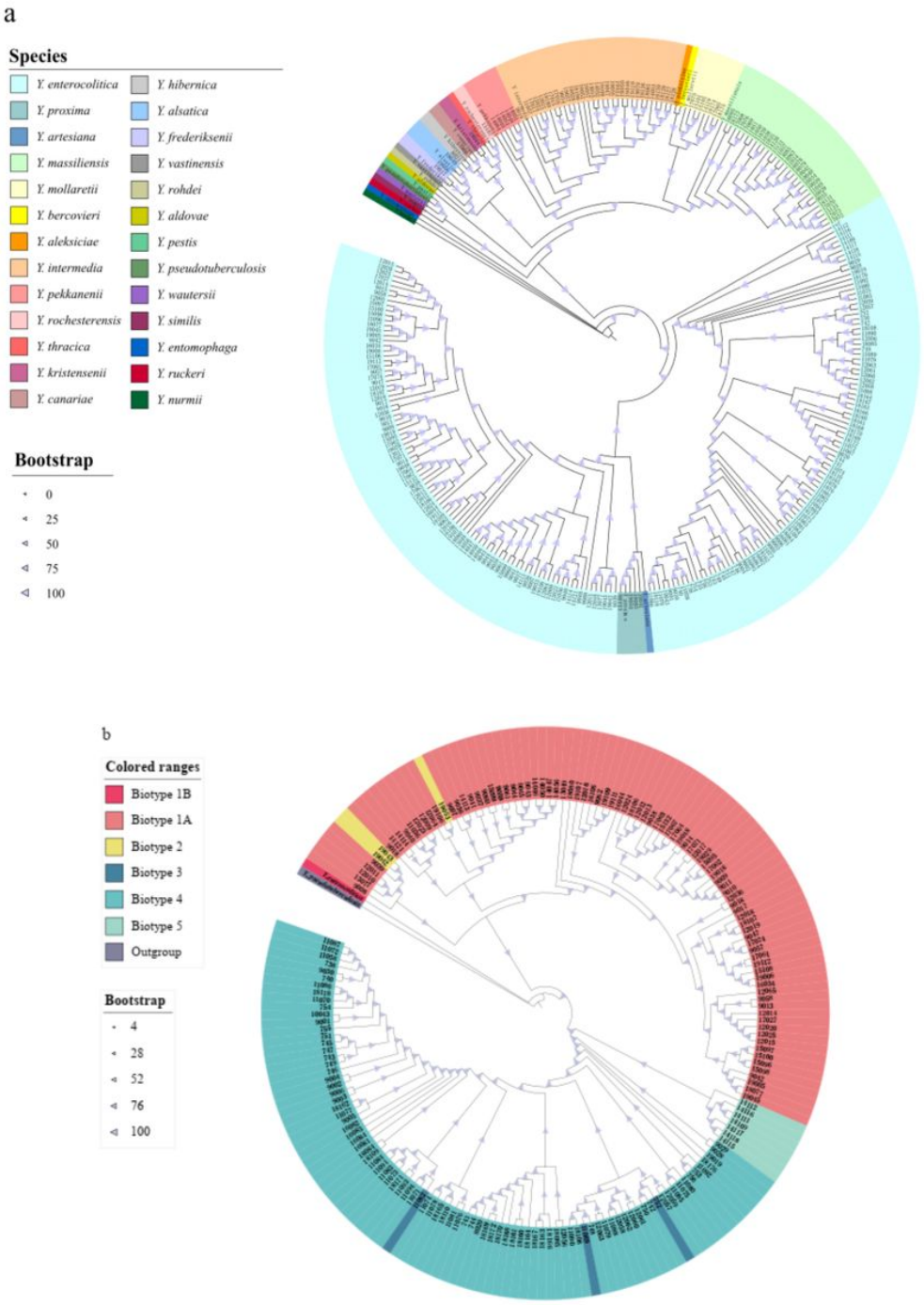

Figure 2

SNP-based maximum likelihood (ML) trees of *Yersinia* spp. strains. **a** The SNP-based ML tree was built from a recombination-filtered alignment of the whole genome SNP (wgSNP) present in 271 isolates. The ML tree was built using the GTR+F+ASC+G4 model, with 1,000 bootstraps based on 1,563,073 SNPs, with *Y. enterocolitica* 8081 (GCA_000009345.1) was used as the reference sequence. **b** The SNP-based ML tree was built from a recombination-filtered alignment of the 499,550 SNPs present in 187 *Y. enterocolitica* isolates. *Y. enterocolitica* 8081 (GCA_000009345.1) was used as the reference sequence. *Y. pseudotuberculosis* (GCA_900637475.1) was used as outgroups.

Figure 3

Distribution of virulence genes in 187 *Y. enterocolitica* strains.

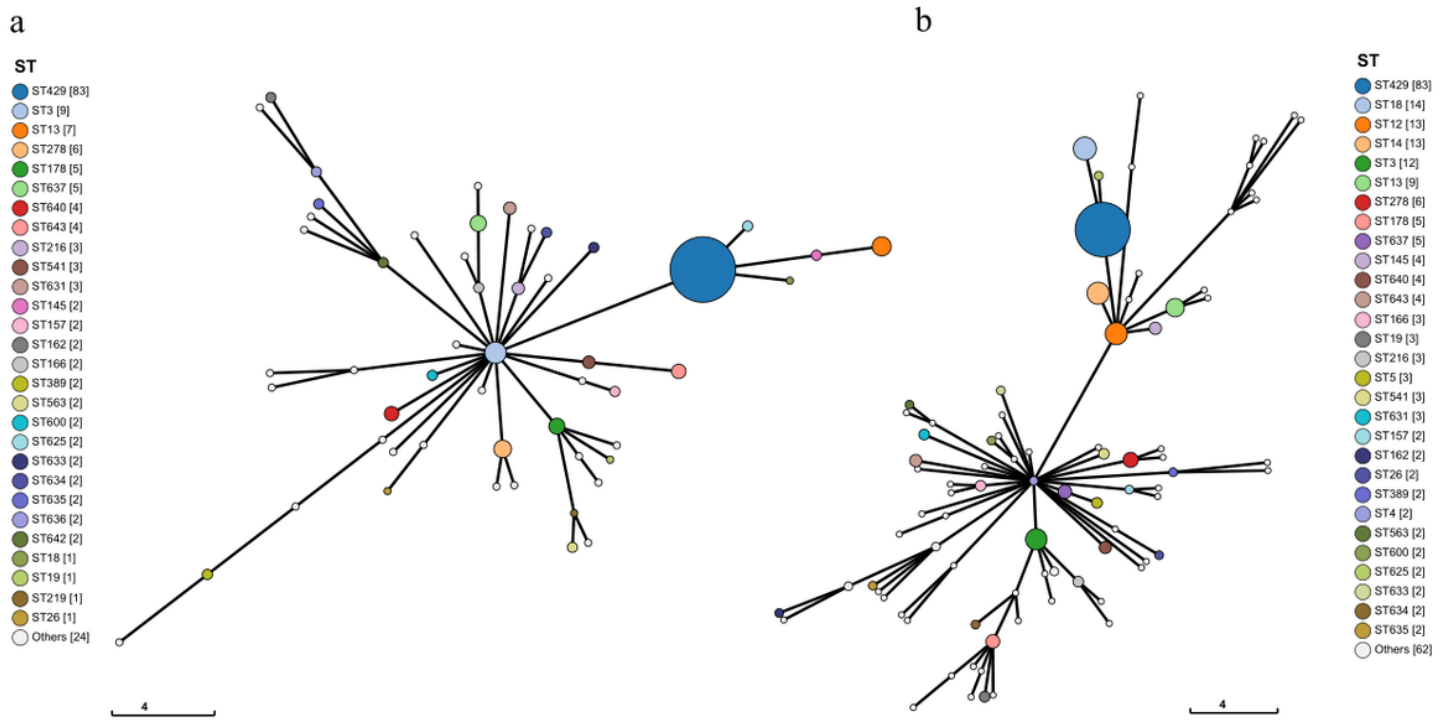

Figure 4

The Minimum Spanning Trees (MST) of *Y. enterocolitica* isolates. The circle size was proportional to the number of isolates. Links between circles were represented according to the number of allelic differences between STs. **a** showed isolates from this study (187). **b** showed isolates from this study (187) and isolates from the public database (87).

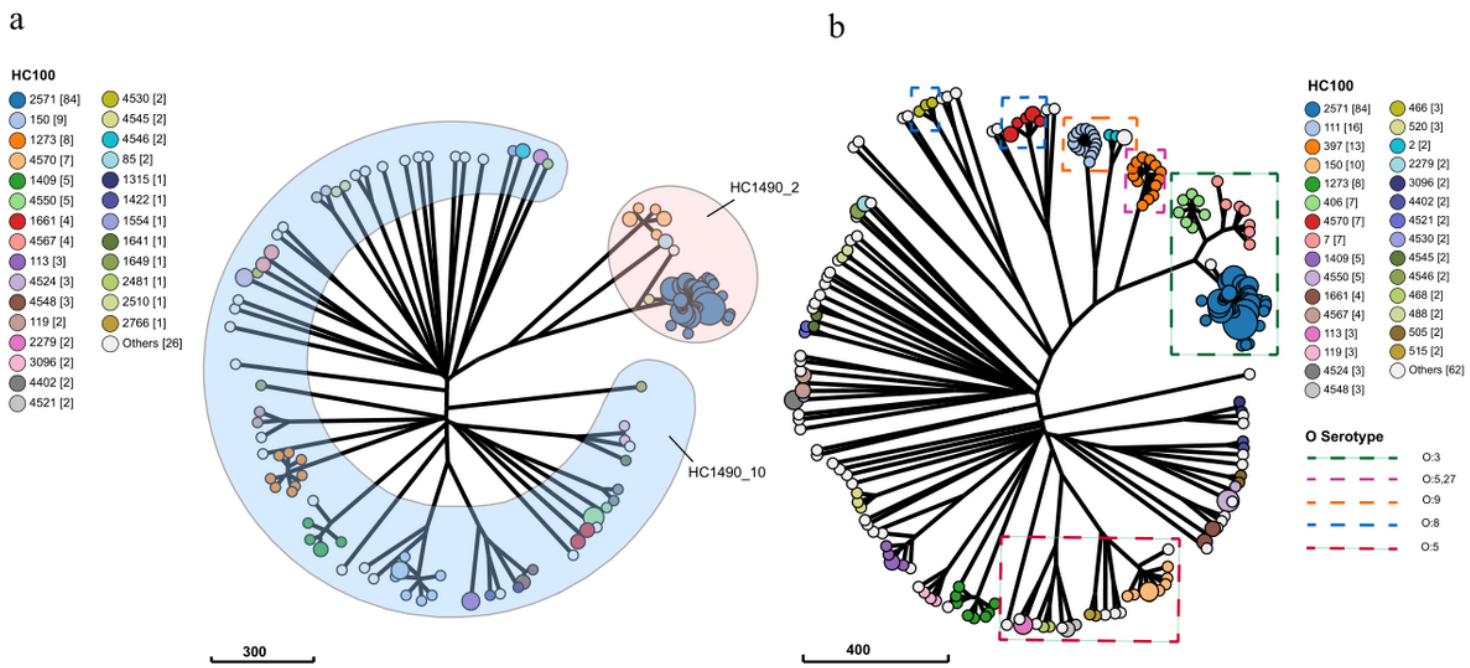

Figure 5

The Neighbor-joining (NJ) tree of *Y. enterocolitica* isolates based on cgMLST. The circle size was proportional to the number of isolates. **a** showed isolates from this study (187). **b** showed isolates from this study (187) and isolates from the public database (87). **a and b** Clusters generated using the hierarchical clustering method from EnteroBase and using a 100 cgMLST allele distance (HC100) were represented by circles.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS.docx