Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparative Genomics Reveals Divergent Adaptive Strategies in Classical Pathogenic and Environmental/Opportunistic Brucella Lineages

Zhao Caihong
Shihezi University
Dang Xinyu
Shihezi University
Ahmad Ali
Shihezi University
Rakhshanda Rani
Shihezi University
Xingmei Deng
Shihezi University
Zhang Hui
Shihezi University
Jia Guo
Shihezi University
Yong Wei
Xinjiang Shawan Tianrun Co., Ltd., Shawan
Sun Zhihua
zhihuasun918@shzu.edu.cn

Shihezi University

Research Article

Keywords: Brucella, Ochrobactrum, comparative genomics, pangenome analysis, virulence-associated homologs, antimicrobial resistance genes, mobile genetic elements

Posted Date: September 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10919344/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Comparative Genomics Reveals Divergent Adaptive Strategies in Classical Pathogenic and Environmental/Opportunistic *Brucella* Lineages

Zhao Caihong[1]†, Dang Xinyu[1]†, Ahmad Ali[1], Rakhshanda Rani[1], Xingmei Deng[1], Zhang Hui[1], Jia Guo[1,3], Yong Wei[2]*, Sun Zhihua[1,2]*

[1]State International Joint Research Center for Animal Health Breeding, College of Animal Science and Technology, Shihezi University, Shihezi 832003, China

[2]Xinjiang Shawan Tianrun Co., Ltd., Shawan 832100, China

[3]Xinjiang Aheqi Agriculture and Animal Husbandry Development Co., Ltd., Aheqi 843500, China

*Correspondence: zhihuasun918@shzu.edu.cn; 772410807@qq.com;

†These authors contributed equally to this work.

Abstract

Background: Classical pathogenic Brucella species are host-adapted facultative intracellular bacteria, whereas several species formerly classified within Ochrobactrum are predominantly environmental organisms and opportunistic pathogens. The genomic basis of the contrasting ecological and pathogenic phenotypes of these lineages remains incompletely understood. This study compared classical pathogenic and environmental/opportunistic Brucella lineages to identify genomic features associated with their contrasting ecological strategies.

Results: We analysed 1,557 publicly available RefSeq assemblies representing four classical pathogenic species (B. abortus, B. melitensis, B. suis and B. canis) and four environmental/opportunistic species reclassified from Ochrobactrum (B. anthropi, B. intermedia, B. pseudogrignonensis and B. haematophila); 1,159 genomes passed quality control. Genome architecture, pangenome structure, VFDB-associated homologues, predicted antimicrobial resistance genes (ARGs), functional annotations and mobile genetic elements were compared. Classical pathogenic Brucella showed compact and comparatively conserved genomes, whereas the environmental/opportunistic group had larger genomes and a substantially broader accessory-gene repertoire. The groups also differed in the distribution of virulence-associated homologues, predicted ARGs, metabolic annotations and mobile genetic elements.

Conclusions: The observed genomic contrasts are consistent with host specialization in classical pathogenic Brucella and adaptation to heterogeneous external niches in environmental/opportunistic lineages. Database-derived virulence, resistance and pathway annotations remain sequence-based predictions and should not be interpreted as direct evidence of phenotype. These findings provide a genomic framework for experimentally testable hypotheses on ecological adaptation, host

association and the evolution of pathogenicity within the Brucella/Ochrobactrum complex.

Keywords: Brucella; Ochrobactrum; comparative genomics; pangenome analysis; virulence-associated homologs; antimicrobial resistance genes; mobile genetic elements

Background

Brucella species are Gram-negative bacteria best known as the etiological agents of brucellosis in humans and animals [1]. The classical pathogenic species *B. abortus*, *B. melitensis*, *B. suis*, and *B. canis* exhibit distinct host associations and virulence phenotypes [2,3]. *B. melitensis* is a major cause of zoonotic brucellosis and is most commonly associated with sheep and goats, whereas *B. abortus*, *B. suis*, and *B. canis* are primarily associated with cattle, swine, and dogs, respectively [3,4]. Brucellosis remains an important One Health problem because of its consequences for human health, livestock productivity, reproductive performance, and international trade [5,6]. The low infectious dose and potential for aerosol transmission also make several classical *Brucella* species important biosafety concerns [7].

Species historically assigned to the genus *Ochrobactrum* are predominantly environmental or free-living Gram-negative bacteria, although several species can cause opportunistic infections, particularly in immunocompromised or otherwise vulnerable patients [8,9]. Genome-based taxonomic analyses led to the transfer of *Ochrobactrum* species to the genus *Brucella*, and several of the resulting combinations are validly published and used by major taxonomic and clinical resources [10–12]. Nevertheless, this taxonomic merger remains debated because classical *Brucella* and the former *Ochrobactrum* lineages differ substantially in ecology, physiology, pathogenic potential, clinical significance, and biosafety implications [44,45]. To minimize ambiguity, we use the currently accepted *Brucella* combinations while explicitly distinguishing "classical pathogenic *Brucella*" (*B. abortus*, *B. melitensis*, *B. suis*, and *B. canis*) from the "environmental/opportunistic group" (*B. anthropi*, *B. intermedia*, *B. pseudogrignonensis*, and *B. haematophila*; formerly *Ochrobactrum* spp.).

The genomic determinants underlying the pronounced phenotypic divergence between these groups remain incompletely resolved. Previous studies have implicated flagellar-system genes in host-cell interactions [14], the urease locus in acid resistance and virulence [5], iron-acquisition systems in intracellular fitness [15], and resistance-nodulation-division (RND) efflux systems in resistance to host-derived antimicrobial compounds and bacterial survival [16]. However, most previous investigations have focused on individual strains, specific loci, or experimental models rather than large-scale comparative analyses across multiple genomic dimensions. In the present study, we analyzed 1,557 publicly available RefSeq

assemblies and retained 1,159 genomes after quality control, based on the species-level counts summarized in Table 1. We compared genome architecture, pangenome composition, virulence-associated homologs, predicted ARG profiles, functional annotations, and mobile genetic elements to identify genomic features associated with the contrasting ecological strategies of classical pathogenic and environmental/opportunistic *Brucella* lineages.

Methods

Study design

This study used a comparative genomics design based on publicly available NCBI RefSeq assemblies. Eight Brucella species were analysed, comprising four classical pathogenic species and four environmental/opportunistic species formerly classified within Ochrobactrum. The analysis compared genome architecture, pangenome composition, virulence-associated homologues, predicted antimicrobial resistance genes, functional annotations and mobile genetic elements.

Genome retrieval, quality control and reference-guided scaffolding

Whole-genome assemblies were retrieved from the National Center for Biotechnology Information (NCBI) RefSeq collection. The initial dataset comprised 426 *B. abortus*, 793 *B. melitensis*, 122 *B. suis*, 33 *B. canis*, 82 *B. anthropi* (formerly *O. anthropi*), 77 *B. intermedia* (formerly *O. intermedium*), 21 *B. pseudogrignonensis* (formerly *O. pseudogrignonense*), and 3 *B. haematophila* (formerly *O. haematophilum*) assemblies, yielding a total of 1,557 genomes. For downstream comparisons, the four species formerly classified as *Ochrobactrum* were pooled into the environmental/opportunistic analytical group.

Initial quality control was performed using SeqKit v2.13.0 [17]. Each species was evaluated relative to a designated reference genome: GCF_000369945.1 (*B. abortus*), GCF_000007152.1 (*B. melitensis*), GCF_000007505.1 (*B. suis*), GCF_000018525.1 (*B. canis*), GCF_015326295.1 (*B. anthropi*), GCF_900454225.1 (*B. intermedia*), GCF_022024335.1 (*B. pseudogrignonensis*), and GCF_005938105.1 (*B. haematophila*). The initial filtering criteria required genome length to fall within ±0.1 Mb of the corresponding species-specific reference, GC content within ±1 percentage point, and N50 ≥1,000 bp. Environmental/opportunistic assemblies that failed the initial screen were processed with Kraken 2 v2.17.1 [18], using default parameters and the k2_standard_16_GB_20260226 database, to remove sequences classified as contaminants. A second quality-control screen was then applied using a relaxed genome-length threshold of ±0.3 Mb while retaining the GC-content and N50 criteria. Quality-control summaries were visualized in RStudio (v2025.05). Assemblies that passed the final filters were subjected to reference-guided scaffolding using RagTag v2.1.0 [19].

Genome annotation and pangenome analysis

Reference-guided FNA assemblies were annotated using Prokka v1.15.6 [20]. The resulting GFF3 files were used for pangenome analysis with Roary v3.11.2 [21] at a protein-sequence identity threshold of 95%. Roary outputs were visualized using the accompanying roary_plots.py script. Pangenome analyses were performed separately for each classical pathogenic species and for the pooled environmental/opportunistic group.

Annotation of virulence-associated homologues and antimicrobial resistance genes

Protein sequences generated by Prokka (FAA format) were queried against the Virulence Factor Database (VFDB) [23] using BLAST+ v2.17.0 [22]. Both the VFDB core and full datasets, together with the corresponding virulence-factor description files, were used for annotation with the following thresholds: E-value ≤1e-5, query coverage ≥70%, sequence identity ≥50%, and alignment length ≥70. Antimicrobial resistance determinants were annotated by querying the protein sequences against the Comprehensive Antibiotic Resistance Database (CARD; version 4.0.1) [24] using BLAST+ v2.17.0 [22], with thresholds of E-value ≤1e-5, query coverage ≥70%, sequence identity ≥80%, and alignment length ≥70. Group-level presence/absence patterns were summarized using UpSet plots in RStudio. Because sequence similarity to a curated virulence or resistance determinant does not, by itself, establish the corresponding phenotype, matches are described throughout this manuscript as virulence-associated homologs or predicted ARGs unless supported by experimental validation.

Association between virulence-associated and resistance-gene densities

Virulence-associated and resistance annotations were summarized at the genome level. Gene density was calculated as the number of annotated virulence-associated homologs or predicted ARGs divided by the total number of coding sequences (CDSs) in each genome. Linear regression was used to assess the association between virulence-associated gene density and ARG density within each analytical group, and the coefficient of determination (R^2) and corresponding P value were reported. Analyses were visualized in RStudio (v2025.05). These analyses quantify genome-level co-occurrence and should not be interpreted as evidence of transcriptional co-regulation or functional coupling.

Functional annotation and pathway profiling

Prokka-derived protein sequences were used for functional annotation. KEGG Ortholog (KO) assignments were generated using KofamScan v1.3.0 [25] with KOfam profile data, and KO identifiers were mapped to KEGG pathway categories. In parallel, eggNOG-mapper v2.1.13 [26] was used to assign orthologous groups and Clusters of Orthologous Groups (COG) functional categories. Group-level functional profiles were summarized and visualized in RStudio. Because the workflow did not include a formal

over-representation test or an explicit background model, these analyses are described as functional annotation and pathway profiling rather than statistical enrichment analysis.

Mobile genetic element annotation

Mobile genetic elements were characterized using Prokka-derived nucleotide assemblies (FNA) and GenBank annotation files (GBK). Insertion sequences (ISs) were identified using ISEScan v1.7.3 [27], genomic islands (GIs) using IslandPath-DIMOB v1.0.0 [28], and prophage regions using PhiSpy v5.0.10 [29]. Plasmid replicons were screened using PlasmidFinder v3.0.3 [30] against the plasmidfinder_db reference collection. Accordingly, the absence of a PlasmidFinder hit was interpreted as the absence of a recognized plasmid replicon rather than definitive evidence for the absence of all plasmid DNA.

Results

Genome quality control and dataset composition

Quality control excluded assemblies with excessive fragmentation, atypical genome size, or deviant GC content relative to the designated species-specific reference. Of the 1,557 downloaded assemblies, 1,159 passed the final quality-control criteria: 319 *B. abortus*, 609 *B. melitensis*, 80 *B. suis*, 27 *B. canis*, 54 *B. anthropi*, 54 *B. intermedia*, 14 *B. pseudogrignonensis*, and 2 *B. haematophila* genomes (Table 1). The retained classical pathogenic genomes were generally compact (approximately 3.2–3.4 Mb), whereas the environmental/opportunistic genomes were larger (approximately 4.6–5.6 Mb), consistent with a broader gene repertoire in the latter group. Figure 1 summarizes the quality-control metrics before and, where applicable, after contaminant removal.

Figure 1. Quality-control profiles of classical pathogenic and environmental/opportunistic *Brucella* genomes. Panels A–D show *B. abortus*, *B. melitensis*, *B. suis*, and *B. canis*, respectively; panels E–H show *B. anthropi*, *B. intermedia*, *B. pseudogrignonensis*, and *B. haematophila*; and panels I–K show post-decontamination quality-control profiles for the indicated environmental/opportunistic groups.

Table 1. Genome retrieval and quality-control summary.

Species	Downloaded	Reference genome	QC criteria	Retained
B. abortus	426	GCF_000369945.1	3.2–3.4 Mb; GC 56–58%; N50 ≥1,000 bp	319
B. melitensis	793	GCF_000007152.1	3.2–3.4 Mb; GC 56–58%; N50 ≥1,000 bp	609
B. suis	122	GCF_000007505.1	3.2–3.4 Mb; GC 56.5–58.5%; N50 ≥1,000 bp	80
B. canis	33	GCF_000018525.1	3.2–3.4 Mb; GC 56–58%; N50 ≥1,000 bp	27
B. anthropi	82	GCF_015326295.1	4.8–5.0 Mb; GC 55–57%; N50 ≥1,000 bp	54

Species	Downloaded	Reference genome	QC criteria	Retained
B. haematophila	3	GCF_005938105.1	5.4–5.6 Mb; GC 55.5–57.5%; N50 ≥1,000 bp	2
B. intermedia	77	GCF_900454225.1	4.6–4.8 Mb; GC 56.5–58.5%; N50 ≥1,000 bp	54
B. pseudogrignonensis	21	GCF_022024335.1	4.9–5.1 Mb; GC 53–55%; N50 ≥1,000 bp	14

Pangenome structure and intergroup gene-repertoire diversity

Pangenome analysis revealed marked differences in gene-repertoire diversity among the five analytical groups (Figure 2). The four classical pathogenic species had substantially smaller gene-cluster repertoires, with totals ranging from 3,141 to 3,917 clusters. In contrast, the pooled environmental/opportunistic group contained 34,875 gene clusters. The pangenomes of the classical pathogenic species were comparatively conserved, whereas the environmental/opportunistic group showed continued accumulation of accessory genes, a pattern consistent with a more open and heterogeneous pangenome. However, because the environmental/opportunistic group combines four species whereas each classical pathogenic group represents a single species, the absolute difference in pangenome size should be interpreted cautiously.

Figure 2. Pangenome profiles of the five analytical groups. Panels A–D show *B. abortus*, *B. melitensis*, *B. suis*, and *B. canis*, respectively; panel E shows the pooled environmental/opportunistic group.

Gene presence–absence patterns further demonstrated more restricted accessory-gene repertoires among the classical pathogenic species, whereas the environmental/opportunistic group exhibited pronounced heterogeneity (Figure 3). These patterns are compatible with contrasting ecological strategies, including greater genome conservation in host-adapted classical *Brucella* and a broader accessory-gene repertoire among lineages exposed to heterogeneous environmental niches. Nevertheless, this comparison is descriptive and is influenced by differences in taxonomic breadth and sample composition among the analytical groups.

Figure 3. Gene presence–absence variation across classical pathogenic and environmental/opportunistic *Brucella* genomes.

Distribution of virulence-associated homologues

VFDB-based annotation identified distinct distributions of virulence-associated homologs among the five analytical groups (Figure 4). Using the VFDB core dataset, 366, 405, 404, 383, and 713 virulence-associated matches were detected in *B. abortus*, *B. melitensis*, *B. suis*, *B. canis*, and the environmental/opportunistic group, respectively. The corresponding totals obtained with the VFDB full dataset were 3,571, 3,810, 3,814, 3,739, and 6,642. A set of 159 VFDB-associated homologs was shared across all five groups. Ten homologs (acrB, adeG, basI, flgS, fliF, icaR, lgtE, mbtI, mtrD, and ureH) were detected only in the combined classical pathogenic dataset, whereas

the environmental/opportunistic group contained 173 group-specific VFDB-associated homologs. These group-specific homologs included annotations related to iron acquisition and siderophore systems, adhesion and biofilm-associated functions, cell-envelope biogenesis, secretion systems, regulation, metabolism, and transport (Table 2). These database matches indicate sequence similarity to known virulence factors but do not establish that the corresponding homologs perform equivalent pathogenic functions in *Brucella*.

Figure 4. Distribution of VFDB-associated homologs across *Brucella* groups. (A) Number of annotated virulence-associated matches; (B) shared and group-specific annotations visualized by UpSet analysis.

Table 2. Environmental/opportunistic group-specific VFDB-associated homologs.

Functional category	No.	Virulence-associated homologs
Strain-specific loci	16	ACICU_RS00395, ACICU_RS00500, ACICU_RS04575, BPS_RS26880, CBUD_RS11275, CD0873, IlpA, KP1_RS17240, KP1_RS29180, LPG_RS03850, LPG_RS03855, LPG_RS04900, LPG_RS11000, PA2383, Rv2951c, YE_RS15465
Iron & siderophore	24	asbB, bauA, bauF, fepB, fpvA, fpvI, fpvR, fslB, fslD, fyuA, fyuA/psn, hmbR, hmuR, hpuB, hxuB, hxuC, irp2, isdE, iucB, iucC, iutA, mbtG, mbtJ, shuA
Adhesion & biofilm	31	aafC, aatC, bap, bapA, bapC, bepB, bmpA, bmpD, bopD, bplE, bplF, brkA, caf1A, fimC, fimD, fimK, fimZ, flpF, focD, lfgH, lpcC, lpfC, nspA, papC, pefC, pgaC, psaC, sfaF, tadA, tadB, tadC
Flagella & motility	3	cheR, fliI, fliR
Secretion systems	39	cesH, clbA, clbF, clbL, clbP, clpC, csuD, entC, exlB, icsP/sopA, lsgC, lspE, msrA/B(pilB), pchR, pilK, pilQ, pilR, pilT, pilU, pseC, pvdS, rtxD, rvhB11, rvhD4, ssrB, tapB, trwF, tssL1, tviC, vexC, virB11, virB4-2, virB8-1, virB9-2, virB9/cagX, virD4, virD4/cag5, vpsC, vscR
Toxins & hemolysins	5	cyaE, cylF, dep/capD, hlyD, sodC
Capsule & cell wall	21	cap8J, cps4H, fcl, gtfB, gtfC, gtfD, hasA, hifC, kpsF, mmaA4, rfaE, rfaJ, wbcJ, wbdA, wbkA, wbpE, wbtD, wcbE, wcbK, wcbP, wzm
Transcription & regulation	26	abaR, adr2, adsA, algU, argP, ascR, bfmR, bspR2, bspR5, btrS, bvgA, chpD, coxFIC1, draC, gadC, htrB, lipF, lgtA, mprA, mtrE, mucD, phoP, rcpA, sigE, siiE, ysrR
Metabolism & transport	7	bioJ, eptC, fptA, lgtA, manB/yhxB, ppaD, pomB2

Antimicrobial resistance gene profiles

CARD-based analysis revealed a narrower predicted resistome in the four classical pathogenic species than in the environmental/opportunistic group (Figure 5). *B. melitensis*, *B. suis*, and *B. canis* each carried four annotated ARGs (mprF, FosXCC, adeF, and qacG), whereas *B. abortus* additionally carried APH(3')-Ia. The environmental/opportunistic group contained 21 ARG annotations, including aminoglycoside-modifying enzymes, OCH-family β-lactamases, OXA-919, cmlA9, floR, sul1, and sul2 (Table 3). Thus, both the number and diversity of predicted resistance determinants were greater in the environmental/opportunistic group. This

pattern is consistent with, but does not independently demonstrate, ecological differences in antimicrobial exposure or horizontal gene acquisition.

Figure 5. Predicted antimicrobial resistance gene profiles. (A) Number of CARD-associated ARG annotations; (B) shared and group-specific ARG distributions.

Table 3. CARD-associated antimicrobial resistance gene annotations.

Group	No.	Predicted ARGs
B. abortus	5	APH(3')-Ia; mprF; FosXCC; adeF; qacG
B. melitensis	4	mprF; FosXCC; adeF; qacG
B. suis	4	mprF; FosXCC; adeF; qacG
B. canis	4	mprF; FosXCC; adeF; qacG
Environmental/opportunistic group	21	AAC(3)-IIc; AAC(6')-Il; ANT(2")-Ia; ANT(9)-Ic; APH(3")-Ib; APH(6)-Id; OCH-1; OCH-2; OCH-3; OCH-4; OCH-5; OCH-6; OCH-7; OCH-8; OXA-919; adeF; cmlA9; floR; qacG; sul1; sul2

Genome-level association between virulence-associated and resistance-gene densities

Genome-level regression analysis revealed positive associations between virulence-associated gene density and ARG density in *B. abortus* (R^2 = 0.176, P < 0.0001), *B. melitensis* (R^2 = 0.176, P < 0.0001), and *B. canis* (R^2 = 0.615, P < 0.0001) (Figure 6A,B,D). *B. suis* showed a weak negative association (R^2 = 0.064, P = 0.0244; Figure 6C), whereas the environmental/opportunistic group showed no statistically supported association (R^2 = 0.010, P = 0.2957; Figure 6E). These results describe patterns of genomic co-occurrence only and do not establish co-regulation, shared transcriptional control, or causal interactions between virulence-associated and resistance determinants.

Figure 6. Association between virulence-associated gene density and ARG density. Panels A–E correspond to *B. abortus*, *B. melitensis*, *B. suis*, *B. canis*, and the environmental/opportunistic group, respectively.

Functional annotation and pathway profiles

KEGG-based functional profiling showed broadly similar core pathway repertoires across groups, together with a larger number of pathway assignments in the environmental/opportunistic group (Figures 7 and 8A). The environmental/opportunistic group was associated with 377 KEGG pathway categories, compared with 352, 351, 356, and 350 categories for *B. abortus*, *B. melitensis*, *B. suis*, and *B. canis*, respectively. A total of 343 pathway categories were shared between the environmental/opportunistic and classical pathogenic groups and included central metabolic functions, secondary-metabolite biosynthesis, microbial metabolism in diverse environments, and ABC transporters. Twenty-three pathway categories were detected only in the environmental/opportunistic group, including assignments related to galactose and phenylalanine metabolism. Some KO mappings

also corresponded to host-oriented KEGG pathway labels, such as antigen processing and presentation and C-type lectin receptor signaling; these labels should be interpreted as shared ortholog mappings rather than evidence that the bacteria encode complete eukaryotic immune pathways.

Figure 7. KEGG pathway annotation profiles of the five *Brucella* analytical groups. Panels A–E correspond to *B. abortus*, *B. melitensis*, *B. suis*, *B. canis*, and the environmental/opportunistic group, respectively.

eggNOG/COG classification (Figure 8B) assigned the four classical pathogenic species to 23 COG functional categories (A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, S, T, U, V, W, and Z). The environmental/opportunistic group was assigned to 22 of these categories and lacked category W (extracellular structures) in the summarized output. This difference is reported descriptively and should not be interpreted as evidence for the complete absence of extracellular structures without direct gene-level or phenotypic validation.

Figure 8. Intergroup functional comparisons. (A) KEGG pathway-category comparison; (B) eggNOG/COG functional-category comparison.

Mobile genetic elements

ISEScan, IslandPath-DIMOB, and PhiSpy identified insertion sequences, genomic islands, and prophage regions, respectively (Figure 9). IS5- and IS3-family elements were prominent, with marked variation in abundance and composition among groups. Predicted genomic islands also differed in number and size distribution, whereas prophage carriage was generally low and varied among species. No plasmid replicons represented in the PlasmidFinder reference database were detected in the analyzed genomes.

Figure 9. Mobile genetic element profiles across *Brucella* groups. Panels A–D summarize insertion-sequence abundance, completeness, family distribution, and overall family composition; panels E–H summarize genomic-island abundance and distribution; and panels I–J summarize prophage abundance and carriage status.

Discussion

This comparative genomic analysis revealed a clear contrast between the compact, conserved genomes of classical pathogenic *Brucella* and the larger, more heterogeneous genomes of the environmental/opportunistic lineages formerly classified as *Ochrobactrum*. Classical pathogenic genomes remained close to 3.2–3.4 Mb and displayed relatively restricted pangenome sizes, whereas environmental/opportunistic genomes were approximately 4.6–5.6 Mb and collectively contained a much larger accessory-gene repertoire. This pattern is consistent with the general principle that stable host-associated niches can favor genome conservation

or reduction, whereas heterogeneous external environments can maintain broader metabolic and accessory functions [31]. However, the environmental/opportunistic group pooled four species, whereas each classical pathogenic group represented a single species. Consequently, part of the observed pangenome contrast may reflect greater taxonomic breadth and should not be attributed solely to ecological lifestyle.

The VFDB analysis likewise revealed substantial differences in the distribution of virulence-associated homologs. Importantly, the environmental/opportunistic group contained more VFDB matches overall, underscoring that the number of database-annotated "virulence genes" is not a direct measure of pathogenicity. Virulence databases contain many functions—such as iron acquisition, adhesion, secretion, cell-envelope biogenesis, motility, stress responses, and transport—that can also contribute to environmental fitness. Conversely, several homologs detected only in the classical pathogenic group, including fliF/flgS-associated flagellar functions, urease-associated loci, siderophore-related annotations, and RND-family efflux systems, represent biologically plausible candidates for host adaptation [14,32–37]. Nevertheless, several cited functional studies were conducted in bacterial species other than *Brucella*. These annotations should therefore be regarded as hypotheses requiring experimental validation rather than as evidence of *Brucella*-specific virulence mechanisms. The same caution applies to lipopolysaccharide- and biofilm-associated homologs such as lgtE and icaR [38–40].

The predicted resistome was broader in the environmental/opportunistic group, which contained 21 CARD-associated ARGs compared with four or five in each classical pathogenic species. Exposure to diverse antimicrobial compounds, microbial competitors, and mobile genetic elements in environmental settings provides a plausible ecological context for this pattern [41]. In classical pathogenic *Brucella*, retention of genes such as mprF may also contribute to resistance against cationic host-defense peptides, although the phenotypic contribution of individual annotations requires experimental testing [42]. The genome-level associations between virulence-associated and resistance-gene densities should likewise be considered hypothesis-generating. Positive slopes in *B. abortus*, *B. melitensis*, and *B. canis* indicate co-occurrence across genomes but do not demonstrate co-regulation; moreover, the modest R^2 values observed for *B. abortus* and *B. melitensis* indicate limited explanatory power despite statistical significance.

Functional profiling supported a broader metabolic repertoire in the environmental/opportunistic group, including additional assignments related to carbon and amino-acid metabolism. These observations are compatible with adaptation to variable nutrient conditions. However, KEGG pathway labels derived from shared orthologs can include eukaryotic or host-oriented categories; consequently, assignments such as antigen processing and presentation or C-type

lectin receptor signaling should not be interpreted as complete bacterial pathways. Similarly, the absence of COG category W from the summarized environmental/opportunistic profile does not establish the loss of extracellular structures. Mobile genetic element analysis revealed variation in insertion sequences, genomic islands, and prophages, suggesting that genome plasticity may differ among lineages. The absence of PlasmidFinder hits indicates that no recognized plasmid replicons were detected, but it does not exclude unrepresented or highly divergent plasmids. This distinction is particularly relevant because the *B. suis* 1330 reference genome consists of two chromosomes and does not contain a plasmid [43].

Taken together, these findings support a model in which ecological niche and host association are accompanied by distinct genomic architectures and accessory-gene repertoires within the *Brucella*/*Ochrobactrum* complex. The results also emphasize the importance of distinguishing sequence-based functional annotation from experimentally demonstrated phenotype. This distinction is particularly important in the context of the ongoing taxonomic debate surrounding the merger of classical *Brucella* with species formerly assigned to *Ochrobactrum* [44,45]. Comparative genomics can identify candidate determinants of host adaptation and environmental fitness, but causal inference will require targeted genetic, transcriptomic, phenotypic, and infection-model validation.

Several limitations should be considered when interpreting these findings. First, the study relies on publicly available assemblies generated using different sequencing platforms, laboratories, assembly pipelines, and sampling schemes, which may introduce technical heterogeneity and batch effects. Second, the analytical design is unbalanced: each classical pathogenic group represents a single species, whereas the environmental/opportunistic group pools four species, and sample sizes differ substantially among groups. Third, the virulence, resistance, pathway, and mobile-element results are computational predictions and were not experimentally validated. Finally, several environmental/opportunistic species are represented by very small numbers of genomes, particularly *B. haematophila* (n = 2 after quality control), limiting inference regarding within-species diversity and generalizability.

Conclusions

Comparative genomic analysis of classical pathogenic and environmental/opportunistic *Brucella* lineages revealed substantial differences in genome size, pangenome structure, virulence-associated homologs, predicted resistomes, functional annotations, and mobile genetic elements. The environmental/opportunistic group exhibited larger genomes and more diverse accessory-gene repertoires, whereas the classical pathogenic species displayed more compact and conserved genomic architectures. These patterns are consistent with contrasting ecological strategies but should be interpreted in light of unequal

sampling, pooled-species comparisons, and the predictive nature of database-based annotations. The present study therefore provides a genomic framework for prioritizing candidate determinants of environmental adaptation and host association while underscoring the need for experimental validation before assigning causal roles in virulence, antimicrobial resistance, or immune evasion.

List of abbreviations

ARG: antimicrobial resistance gene; CARD: Comprehensive Antibiotic Resistance Database; CDS: coding sequence; COG: Cluster of Orthologous Groups; GI: genomic island; IS: insertion sequence; KEGG: Kyoto Encyclopedia of Genes and Genomes; KO: KEGG Ortholog; MGE: mobile genetic element; NCBI: National Center for Biotechnology Information; RND: resistance-nodulation-division; VFDB: Virulence Factor Database.

Declarations

Ethics approval and consent to participate

Not applicable. This study analysed publicly available bacterial genome assemblies and did not involve human participants, human data, human tissue or animals.

Consent for publication

Not applicable.

Availability of data and materials

The genome assemblies analysed in this study are publicly available through the NCBI RefSeq database. Data supporting the findings of this study are included within the article.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Scientific and Technological Tackling Plan for Key Fields of the Corps (2024AB034), Tacheng Prefecture Science and Technology Planning Project (2026BB4014), the Science and Technology Action Project for Rural Revitalization and Industrial Development in the Autonomous Region (2026NC033), Tianchi Talent Project (CZ004311), and The Shihezi University International Science and Technology Cooperation Promotion Plan Project (GJHZ202402).

Authors' contributions

C.Z. and X.D. designed the study. C.Z., Y.W. and X.D. performed data analysis, ran computational scripts and generated the figures. C.Z., X.D., A.A., R.R., and Z.S. drafted the manuscript. Z.C., H.Z., J.G. and X.D. reviewed and edited the manuscript. The

remaining authors contributed resources, supervision and project administration. All authors read and approved the final manuscript.

Acknowledgements

The authors thank their supervisor for valuable suggestions and guidance and all personnel who contributed to this study.

References

1. Cloeckaert A, Vergnaud G, Zygmunt MS. Omp2b porin alteration in the course of evolution of Brucella spp. Front Microbiol. 2020;11:284. https://doi.org/10.3389/fmicb.2020.00284

2. Rajendhran J. Genomic insights into Brucella. Infect Genet Evol. 2021;87:104635. https://doi.org/10.1016/j.meegid.2020.104635

3. Marzetti S, Carranza C, Roncallo M, et al. Recent trends in human Brucella canis infection. Comp Immunol Microbiol Infect Dis. 2013;36(1):55-61. https://doi.org/10.1016/j.cimid.2012.09.002

4. Cloeckaert A, Roop RM 2nd, Scholz HC, et al. Editorial: Pathogenomics of the genus Brucella and beyond, volume II. Front Microbiol. 2024;15:1370330. https://doi.org/10.3389/fmicb.2024.1370330

5. Sangari FJ, Seoane A, Rodríguez MC, et al. Characterization of the urease operon of Brucella abortus and assessment of its role in virulence of the bacterium. Infect Immun. 2007;75(2):774-780. https://doi.org/10.1128/iai.01244-06

6. Hao M, Wang M, Tang T, et al. Regulation of the gene for alanine racemase modulates amino acid metabolism with consequent alterations in cell wall properties and adhesive capability in Brucella spp. Int J Mol Sci. 2023;24(22):16145. https://doi.org/10.3390/ijms242216145

7. Doganay GD, Doganay M. Brucella as a potential agent of bioterrorism. Recent Pat Antiinfect Drug Discov. 2013;8(1):27-33. https://doi.org/10.2174/1574891x11308010006

8. Kettaneh A, Weill FX, Poilane I, et al. Septic shock caused by Ochrobactrum anthropi in an otherwise healthy host. J Clin Microbiol. 2003;41(3):1339-1341. https://doi.org/10.1128/jcm.41.3.1339-1341.2003

9. Ryan MP, Pembroke JT. The genus Ochrobactrum as major opportunistic pathogens. Microorganisms. 2020;8(11):1797. https://doi.org/10.3390/microorganisms8111797

10. Leclercq SO, Cloeckaert A, Zygmunt MS. Taxonomic organization of the family Brucellaceae based on a phylogenomic approach. Front Microbiol. 2019;10:3083. https://doi.org/10.3389/fmicb.2019.03083

11. Hördt A, López MG, Meier-Kolthoff JP, et al. Analysis of 1,000+ type-strain genomes substantially improves taxonomic classification of Alphaproteobacteria. Front Microbiol. 2020;11:468. https://doi.org/10.3389/fmicb.2020.00468

12. Centers for Disease Control and Prevention. 12/19/2022: lab update: reclassification of Ochrobactrum species into the Brucella genus [Internet]. Atlanta: CDC; 2022 Dec 19 [cited 2024 Apr 3]. https://www.cdc.gov/locs/2022/12-19-2022-Lab-Update-Reclassification_Ochrobactrum_species_Brucella_genus.html

13. Aljanazreh B, Shamseye AA, Abuawad A, et al. Genomic distribution of the insertion sequence IS711 reveal a potential role in Brucella genome plasticity and host preference. Infect Genet Evol. 2023;112:105457. https://doi.org/10.1016/j.meegid.2023.105457

14. Lestrate P, Delrue RM, Danese I, et al. Attenuated signature-tagged mutagenesis mutants of Brucella melitensis identified during the acute phase of infection in mice. Infect Immun. 2003;71(12):7053-7060. https://doi.org/10.1128/IAI.71.12.7053-7060.2003

15. Bellaire BH, Elzer PH, Baldwin CL, Roop RM 2nd. Genetic organization and iron-responsive regulation of the Brucella abortus 2,3-dihydroxybenzoic acid biosynthesis operon, a cluster of genes required for wild-type virulence in pregnant cattle. Infect Immun. 2003;71(4):1794-1803. https://doi.org/10.1128/IAI.71.4.1794-1803.2003

16. Posadas DM, Martín FA, Sabio y García JV, et al. The TolC homologue of Brucella suis is involved in resistance to antimicrobial compounds and virulence. Infect Immun. 2007;75(1):379-389. https://doi.org/10.1128/IAI.01349-06

17. Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One. 2016;11(10):e0163962. https://doi.org/10.1371/journal.pone.0163962

18. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20(1):257. https://doi.org/10.1186/s13059-019-1891-0

19. Alonge M, Lebeigle L, Kirsche M, et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 2022;23(1):258. https://doi.org/10.1186/s13059-022-02823-7

20. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-2069. https://doi.org/10.1093/bioinformatics/btu153

21. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691-3693. https://doi.org/10.1093/bioinformatics/btv421

22. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. https://doi.org/10.1186/1471-2105-10-421

23. Zhou S, Liu B, Zheng D, et al. VFDB 2025: an integrated resource for exploring anti-virulence compounds. Nucleic Acids Res. 2025;53(D1):D871-D877. https://doi.org/10.1093/nar/gkae968

24. Alcock BP, Huynh W, Chalil R, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2023;51(D1):D690-D699. https://doi.org/10.1093/nar/gkac920

25. Aramaki T, Blanc-Mathieu R, Endo H, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020;36(7):2251-2252. https://doi.org/10.1093/bioinformatics/btz859

26. Cantalapiedra CP, Hernández-Plaza A, Letunic I, et al. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38(12):5825-5829. https://doi.org/10.1093/molbev/msab293

27. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. Bioinformatics. 2017;33(21):3340-3347. https://doi.org/10.1093/bioinformatics/btx433

28. Bertelli C, Brinkman FSL. Improved genomic island predictions with IslandPath-DIMOB. Bioinformatics. 2018;34(13):2161-2167. https://doi.org/10.1093/bioinformatics/bty095

29. Akhter S, Aziz RK, Edwards RA. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. Nucleic Acids Res. 2012;40(16):e126. https://doi.org/10.1093/nar/gks406

30. Carattoli A, Hasman H. PlasmidFinder and in silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). Methods Mol Biol. 2020;2075:285-294. https://doi.org/10.1007/978-1-4939-9877-7_20

31. Moran NA. Microbial minimalism: genome reduction in bacterial pathogens. Cell. 2002;108(5):583-586. https://doi.org/10.1016/s0092-8674(02)00665-7

32. Léonard S, Ferooz J, Haine V, et al. FtcR is a new master regulator of the flagellar system of Brucella melitensis 16M with homologs in Rhizobiaceae. J Bacteriol. 2007;189(1):131-141. https://doi.org/10.1128/JB.01349-08

33. Fong YH, Wong HC, Yuen MH, et al. Structure of UreG/UreF/UreH complex reveals how urease accessory proteins facilitate maturation of Helicobacter pylori urease. PLoS Biol. 2013;11(10):e1001678. https://doi.org/10.1371/journal.pbio.1001678

34. Harrison AJ, Yu M, Gårdenborg T, et al. The structure of MbtI from Mycobacterium tuberculosis, the first enzyme in the biosynthesis of the siderophore mycobactin, reveals it to be a salicylate synthase. J Bacteriol. 2006;188(17):6081-6091. https://doi.org/10.1128/JB.00472-06

35. Shafer WM, Qu XD, Waring AJ, Lehrer RI. Modulation of Neisseria gonorrhoeae susceptibility to vertebrate antibacterial peptides due to a member of the resistance/nodulation/division efflux pump family. Proc Natl Acad Sci U S A. 1998;95(4):1829-1833. https://doi.org/10.1073/pnas.95.4.1829

36. Warner DM, Levy SB. Different effects of transcriptional regulators MarA, SoxS and Rob on susceptibility of Escherichia coli to cationic antimicrobial peptides (CAMPs): Rob-dependent CAMP induction of the marRAB operon. Microbiology. 2010;156(Pt 2):570-578. https://doi.org/10.1099/mic.0.033437-0

37. Yoon EJ, Courvalin P, Grillot-Courvalin C. RND-type efflux pumps in multidrug-resistant clinical isolates of Acinetobacter baumannii: major role for AdeABC overexpression and AdeRS mutations. Antimicrob Agents Chemother. 2013;57(7):2989-2995. https://doi.org/10.1128/AAC.02430-12

38. Chen HD, Groisman EA. The biology of the PmrA/PmrB two-component system: the major regulator of lipopolysaccharide modifications. Annu Rev Microbiol. 2013;67:83-112. https://doi.org/10.1146/annurev-micro-092412-155626

39. Wakarchuk W, Schur MJ, St Michael F, et al. Biochemical properties of Neisseria gonorrhoeae LgtE. J Bacteriol. 2002;184(23):6410-6416. https://doi.org/10.1128/JB.184.23.6410-6416.2002

40. Conlon KM, Humphreys H, O'Gara JP. icaR encodes a transcriptional repressor involved in environmental regulation of ica operon expression and biofilm formation in Staphylococcus epidermidis. J Bacteriol. 2002;184(16):4400-4408. https://doi.org/10.1128/JB.184.16.4400-4408.2002

41. Gaze WH, Krone SM, Larsson DG, et al. Influence of humans on evolution and mobilization of environmental antibiotic resistome. Emerg Infect Dis. 2013;19(7):1127-1134. https://doi.org/10.3201/eid1907.120871

42. Samant S, Hsu FF, Neyfakh AA, et al. The Bacillus anthracis protein MprF is required for synthesis of lysylphosphatidylglycerols and for resistance to cationic antimicrobial peptides. J Bacteriol. 2009;191(4):1311-1319. https://doi.org/10.1128/JB.01345-08

43. Paulsen IT, Seshadri R, Nelson KE, et al. The Brucella suis genome reveals fundamental similarities between animal and plant pathogens and symbionts. Proc Natl Acad Sci U S A. 2002;99(20):13148-13153. https://doi.org/10.1073/pnas.192319099

44. Moreno E, Blasco JM, Letesson JJ, Gorvel JP, Moriyón I. Pathogenicity and its implications in taxonomy: the Brucella and Ochrobactrum case. Pathogens. 2022;11(3):377. https://doi.org/10.3390/pathogens11030377

45. Moreno E, Middlebrook EA, Altamirano-Silva P, et al. If you're not confused, you're not paying attention: Ochrobactrum is not Brucella. J Clin Microbiol. 2023;61(8):e00438-23. https://doi.org/10.1128/jcm.00438-23

Figures

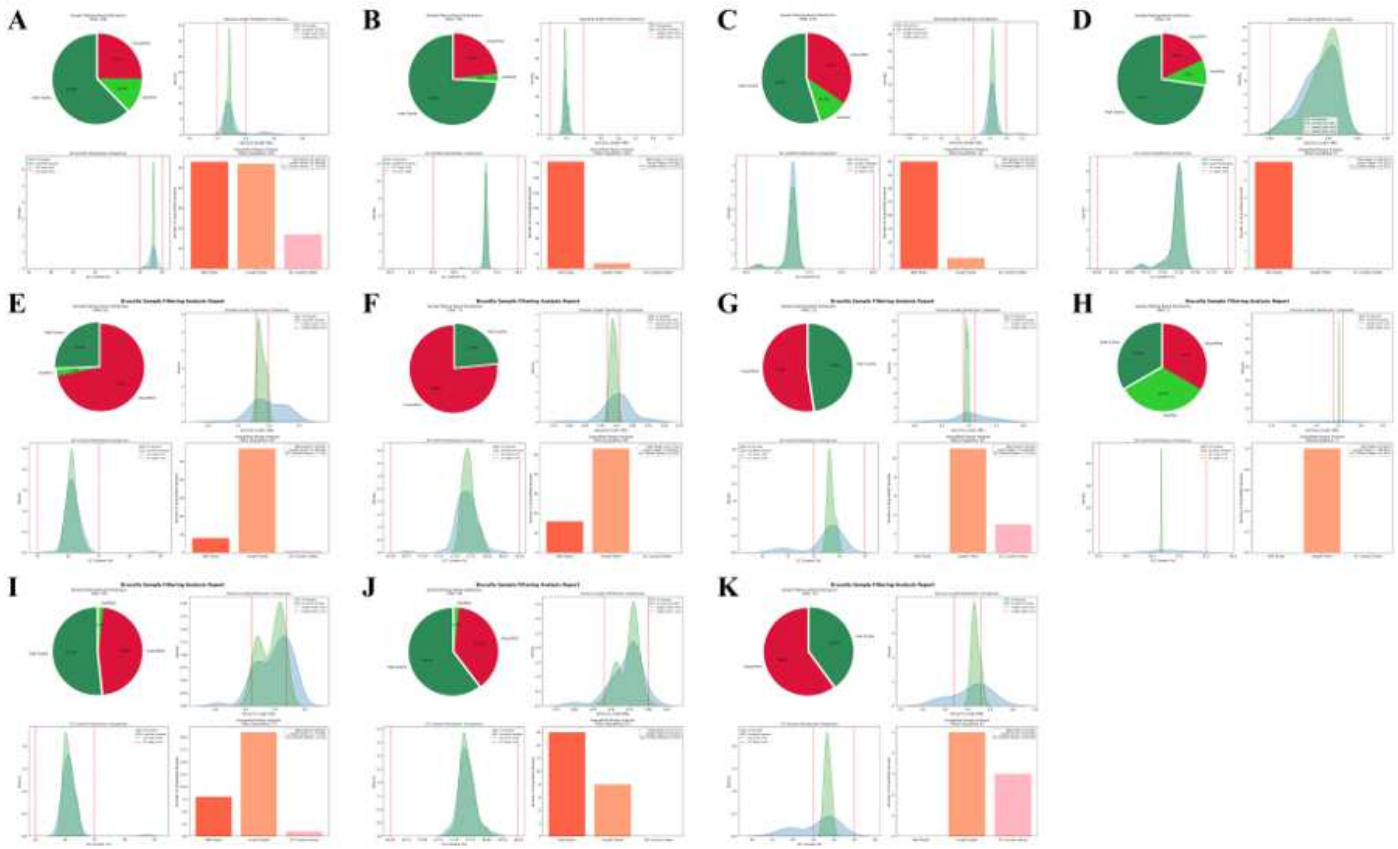


Figure 1

Quality-control profiles of classical pathogenic and environmental/opportunistic *Brucella*genomes. Panels A–D show *B. abortus*, *B. melitensis*, *B. suis*, and *B. canis*, respectively; panels E–H show *B. anthropi*, *B. intermedia*, *B. pseudogrignonensis*, and *B. haematophila*; and panels I–K show post-decontamination quality-control profiles for the indicated environmental/opportunistic groups.

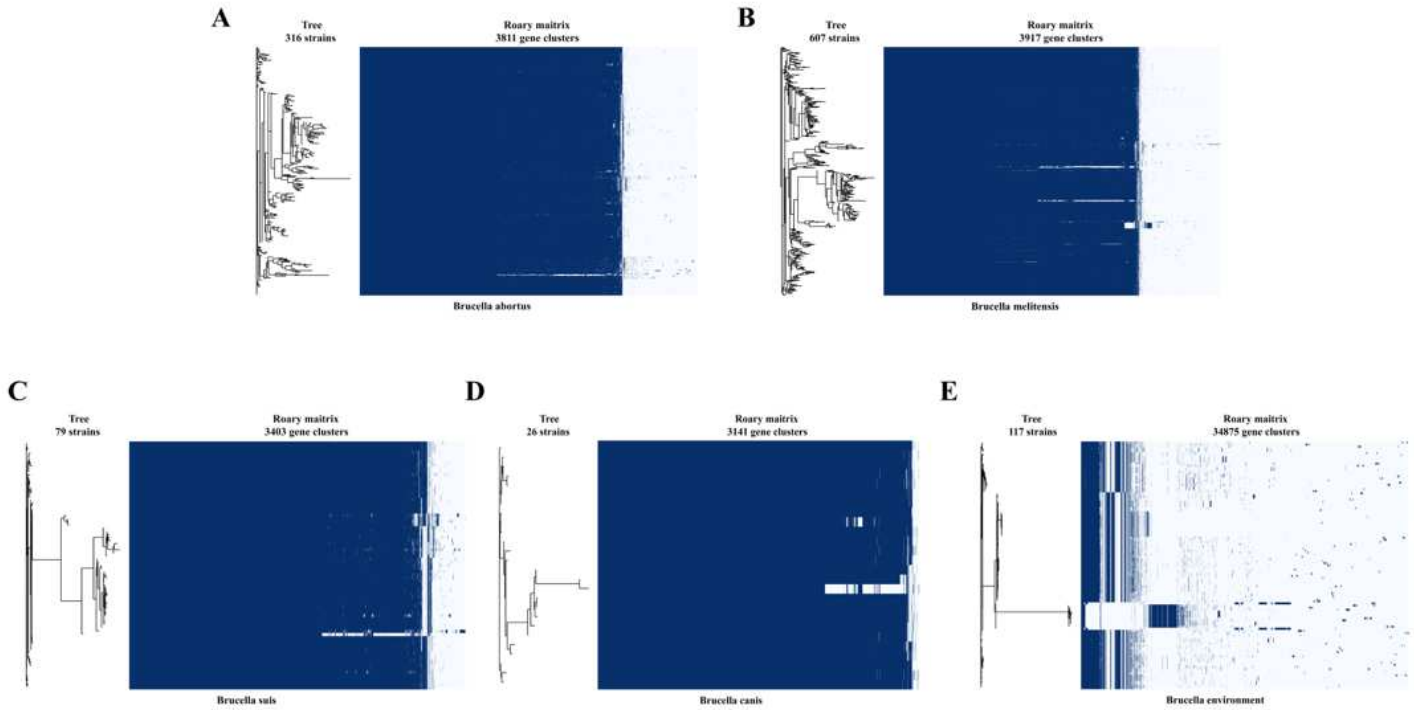


Figure 2

Pangenome profiles of the five analytical groups. Panels A–D show *B. abortus*, *B. melitensis*, *B. suis*, and *B. canis*, respectively; panel E shows the pooled environmental/opportunistic group.

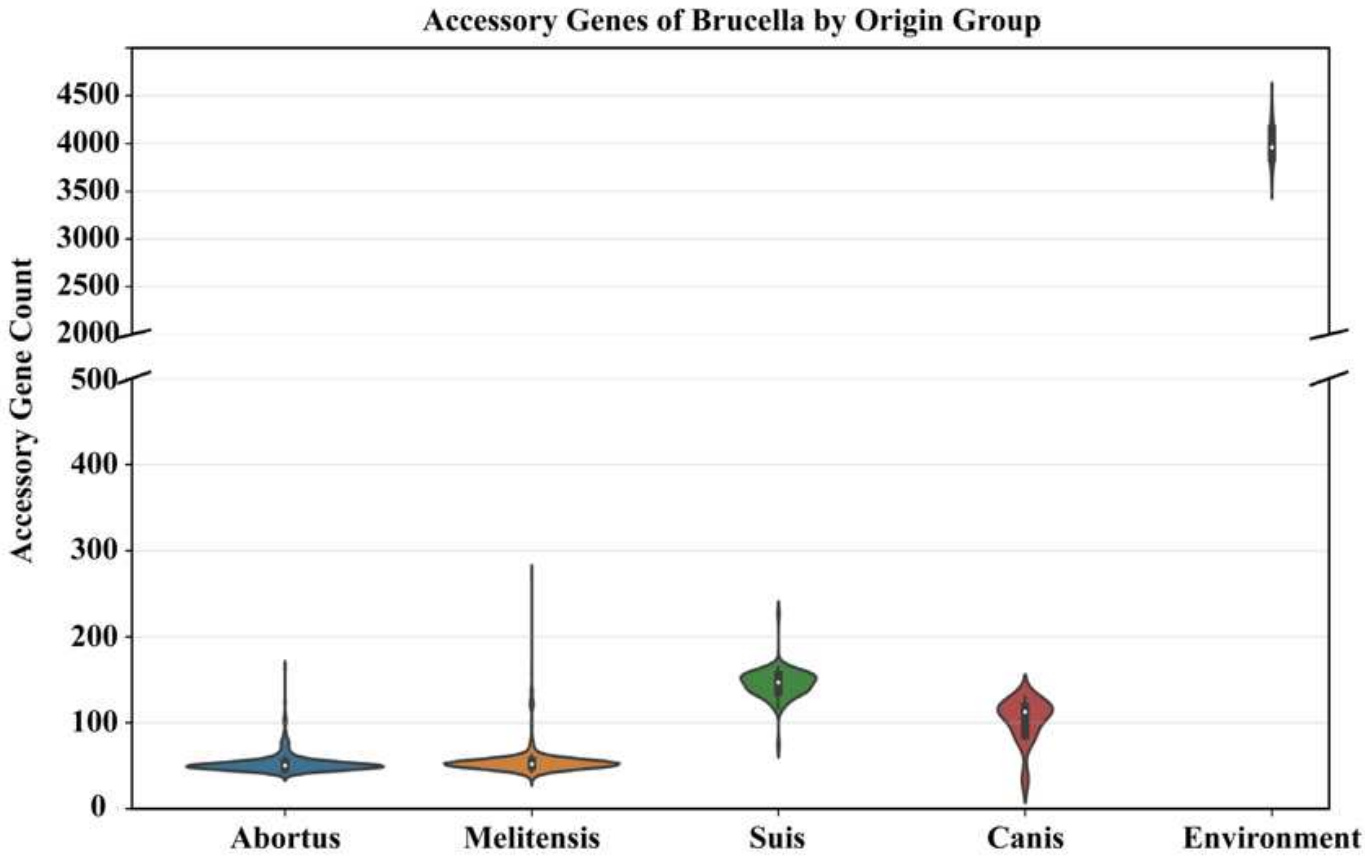


Figure 3

Gene presence–absence variation across classical pathogenic and environmental/opportunistic *Brucella*genomes.

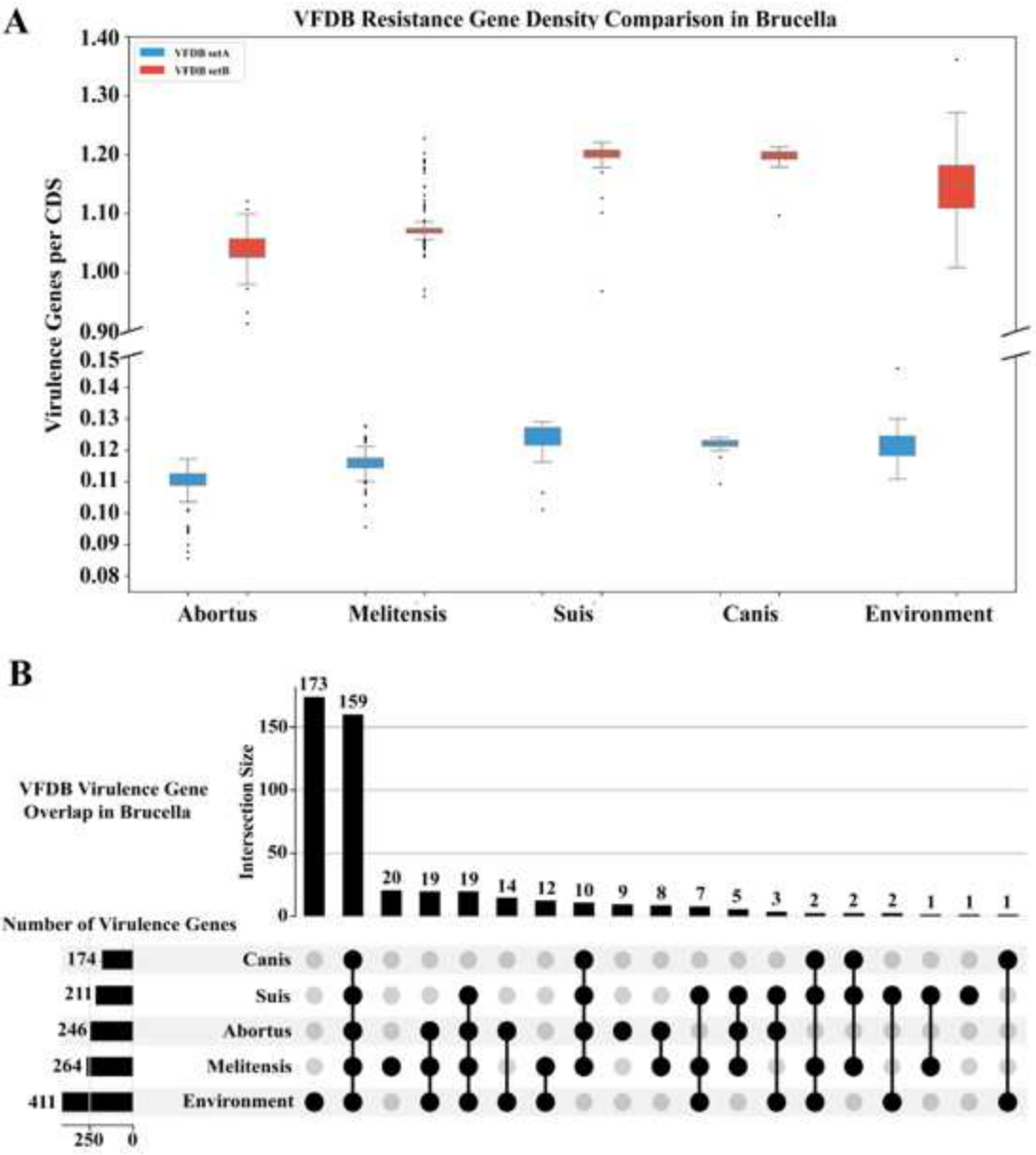


Figure 4

Distribution of VFDB-associated homologs across *Brucella* groups. (A) Number of annotated virulence-associated matches; (B) shared and group-specific annotations visualized by UpSet analysis.

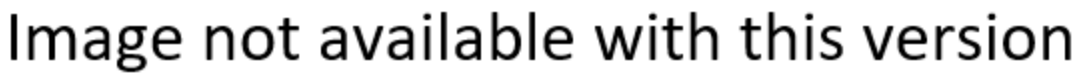


Figure 5

Predicted antimicrobial resistance gene profiles. (A) Number of CARD-associated ARG annotations; (B) shared and group-specific ARG distributions.

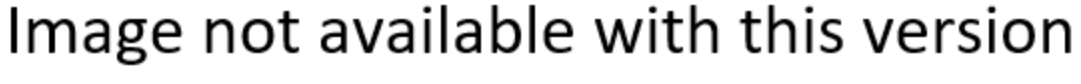


Figure 6

Association between virulence-associated gene density and ARG density. Panels A–E correspond to *B. abortus*, *B. melitensis*, *B. suis*, *B. canis*, and the environmental/opportunistic group, respectively.

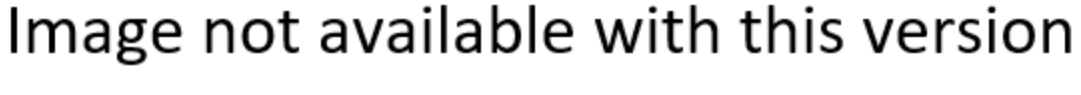


Figure 7

KEGG pathway annotation profiles of the five *Brucella* analytical groups. Panels A–E correspond to *B. abortus*, *B. melitensis*, *B. suis*, *B. canis*, and the environmental/opportunistic group, respectively.

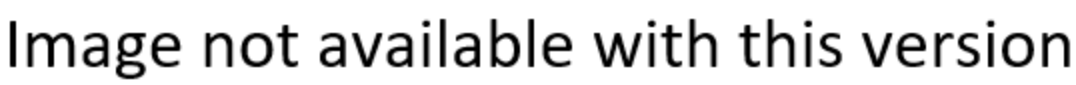


Figure 8

Intergroup functional comparisons. (A) KEGG pathway-category comparison; (B) eggNOG/COG functional-category comparison.

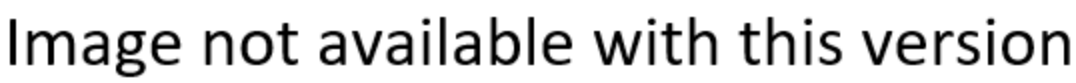


Figure 9

Mobile genetic element profiles across *Brucella* groups. Panels A–D summarize insertion-sequence abundance, completeness, family distribution, and overall family composition; panels E–H summarize genomic-island abundance and distribution; and panels I–J summarize prophage abundance and carriage status.