Preplanned Studies

Comparative Transcriptome Analysis of *Bartonella* Species from Diverse Hosts Reveals Evolutionary Insights — Global, 2023–2025

Min Chen[1]; Na Han[1]; Wen Zhang[1]; Yujun Qiang[1]; Qiyong Liu[1]; Dongmei Li[1,#]

Summary

What is already known about this topic?

Transcriptomic approaches have been used in phylogenetic studies of eukaryotes, yet their application to prokaryotic organisms remains limited. To date, research on species differentiation and phylogenetic relationships within *Bartonella* spp. has relied primarily on genomic data, and the taxonomic relationships among species in this genus remain poorly resolved.

What is added by this report?

The transcriptome of *Bartonella* spp. varies based on species and host origins and exhibits systematic differences. Transcriptome-based phylogenetic analysis reveals that strains cluster by species and host origin. This pattern indicates that differentiation by species and genetic evolution of *Bartonella* is predominantly shaped by the host origin.

What are the implications for public health practice?

Our study confirms that transcriptomics is an effective tool for determining differentiation and evolutionary relationships among *Bartonella* spp. and demonstrates its potential applicability to studies of other prokaryotic species.

ABSTRACT

Introduction: To investigate transcriptional differences and their implications for evolutionary relationships among *Bartonella* species from diverse host origin.

Methods: Illumina high-throughput sequencing technology was used to sequence the transcriptomes of eighty-seven *Bartonella* strains. The differences in gene expression among strains from different species and hosts were analyzed, and the results of the genome phylogenetic analysis were compared to explore the role and influencing factors of transcription levels in *Bartonella* species differentiation.

Results: The transcriptomes of *Bartonella* strains varied systematically by species and host origin, with numerous differentially expressed genes (DEGs) identified among strains from different sources. Experimental verification confirmed that the differences in expression of *bepC*, *secB*, *secDF*, and *ftsY* played a key role in host-specific recognition. Furthermore, phylogenetic analysis based on transcriptomic data clearly reflected the taxonomic relationships among *Bartonella* species, indicating that their genetic evolution was primarily driven by host-related factors, a finding consistent with genome-based analysis.

Conclusions: Transcriptome data provides a powerful approach for clarifying species differentiation and evolutionary relationships within *Bartonella* spp., with potential applicability to other prokaryotic species. These findings provided critical insights for resolving taxonomic uncertainties and advancing systematic research.

Bartonella spp. are aerobic, fastidious intracellular parasites transmitted by arthropod vectors or through animal scratches (*1*). These pathogens exhibit broad host ranges and cause a spectrum of zoonotic diseases in humans, ranging from self-limiting infections to fatal outcomes (*1*). RNA sequencing (RNA-seq) has emerged as a powerful tool for studying the genetic evolution of species, particularly in eukaryotic phylogenetics (*2*). However, its application to phylogenetic and species differentiation studies in prokaryotes remains limited. Previous phylogenetic analyses of *Bartonella* spp. have relied on single-gene or multi-gene approaches (*3*–*4*), yet the taxonomic criteria employed in these studies remain poorly defined.

Eighty-seven *Bartonella* strains obtained from the Chinese Center for Disease Control and Prevention were analyzed (Supplementary Table S1, available at https://weekly.chinacdc.cn/). Total RNA was extracted following cultivation and used for cDNA library

construction. Paired-end sequencing was performed on the Illumina NovaSeq 6000 platform, followed by quality control, mRNA enrichment, and de novo transcriptome assembly. The assembled transcripts were functionally annotated against the Swiss-Prot, eggNOG, GO, and KEGG databases. Differentially expressed genes (DEGs) were identified using DESeq2 (|log2FoldChange|≥2, adjusted P<0.05) and subsequently subjected to GO and KEGG enrichment analyses. Single nucleotide polymorphisms (SNPs) in core genes were called from both transcriptome and genome data using Snippy (version 4.6.0, The University of Melbourne, Melbourne, Australia) and Gubbins (version 3.3.5, Wellcome Sanger Institute, Cambridge, UK), with *B. grahamii* (ATCC 700132) as the reference. Maximum likelihood (ML) and Bayesian inference (BI) phylogenetic trees were constructed using IQ-TREE (version 2.3.6, Center for Integrative Bioinformatics Vienna, Vienna, Austria) and MrBayes (version 3.2.7, Department of Biodiversity Informatics, Swedish Museum of Natural History, Stockholm, Sweden), respectively. DEGs were validated by quantitative polymerase chain reaction (qPCR) using *gltA* as the reference gene.

After quality control, 2.18 billion clean reads were obtained. De novo assembly yielded a reference transcriptome comprising 51,639 transcripts, of which 34,069 unigenes were functionally annotated. Hierarchical clustering revealed divergent gene expression patterns that correlated with host origin (Figure 1A). Strains of canine and monkey origin formed distinct clusters characterized by upregulation of class I and II genes. In contrast, rodent-origin strains exhibited heterogeneous expression profiles. Notably, numerous DEGs were identified between different species and host orders, whereas far fewer were detected within the same species or host order (Supplementary Tables S2–S3, available at https://weekly.chinacdc.cn/). Species-level DEGs were enriched in broad metabolic and information-processing categories; however, host-order-level DEGs were strikingly enriched in functions related to host interaction and pathogenesis. Specifically, significant GO terms involved host cell components, the type IV secretion system (T4SS), and immune defense mechanisms. KEGG analysis further confirmed enrichment in key pathways, including bacterial secretion and prokaryotic defense systems. Given this enrichment in host-recognition-related secretion systems, we examined the expression of these genes in detail. Key genes encoding T4SS structural components, including *virB2–4*, were significantly upregulated in strains from the order Carnivora (Table 1). qPCR confirmed significant differential expression of *bepC*, *secB*, and *secDF* (P<0.05), with trends consistent with the RNA-seq data (Figure 1B–E, Supplementary Table S4, available at https://weekly.chinacdc.cn/).

Phylogenetic analyses were conducted independently using transcriptomic and genomic datasets. For the transcriptomic and genomic analyses, 19 and 31 reference genomes were incorporated, respectively, yielding 3,498 and 24,395 core gene SNPs. Despite the difference in reference genome numbers, comparative analysis confirmed that this did not affect reconstruction reliability (Figure 2A–B). Transcriptome trees were fully resolved and exhibited strong nodal support (ML≥75%, BI≥85%). All species-level branches displayed clear species-specific clustering, revealing closely related pairs such as *B. koehlerae–B. henselae* and *B. doshiae–B. bacilliformis*. The genome-based tree also demonstrated clear species clustering with even higher support values (ML=100%, BI=100%). While interspecific relationships were largely congruent between tree types, discrepancies were observed; notably, *B. bacilliformis* clustered more closely with *B. clarridgeiae* in the genome tree. Despite these minor topological differences, species-level clade structures were highly consistent across datasets. Both ML trees exhibited clear host-associated clustering at the order level (Rodentia, Primates, and Carnivora) and at finer host-species resolution. For example, dog-associated *B. vinsonii* subsp. *berkhoffii*, cat-associated *B. koehlerae* and *B. henselae*, and monkey-associated *B. quintana* formed distinct clusters. In contrast, geographical distribution had limited influence on

TABLE 1. Enrichment of differentially expressed genes in the bacterial secretion system pathway.

Groups	Status	Genes
Primates vs. Carnivora	Up	*virB6*, *virB11*, *virB2*, *virB4*, *yidC*, *virB8*, *virD4*, *sdhB*, *virB9*, *secDF*, *secY*, *tatC*, *bepC*
	Down	*lktB*, *pyrG*, *virB10*, *secA*, *ftsY*, *hrpI*, *xpsE*, *ffh*, *virB3*, *yajC*, *secB*
Primates vs. Rodentia	Up	*virB8*, *virB11*, *virD4*, *tatC*, *yajC*, *virB10*, *tatA*, *bepC*, *secDF*, *virB3*, *fliF*, *gltX1*, *secY*, *virB3*, *virB6*, *yidC*
	Down	*virB4*, *ftsY*, *pyrG*, *yidC*, *secY*, *secB*, *virB9*

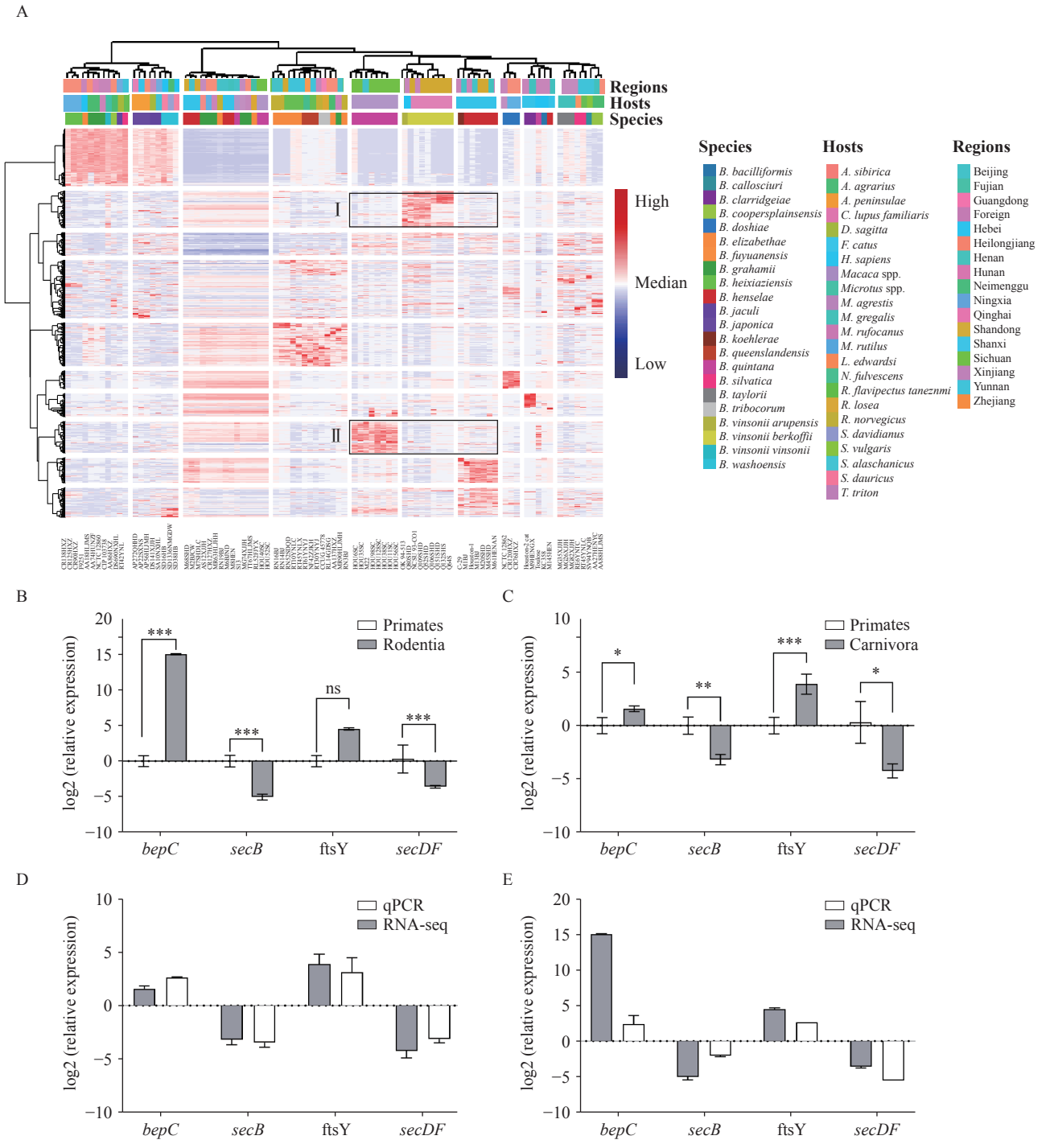

FIGURE 1. Transcriptome analysis results. (A) Gene expression clustering heatmap of 87 *Bartonella* strains. (B) qPCR validation of selected DEGs from the "Primates *vs.* Rodentia" comparison. (C) qPCR validation of selected DEGs from the "Primates *vs.* Carnivora" comparison. (D) Comparison of qPCR and RNA-seq fold-change values for the "Primates *vs.* Rodentia" comparison. (E) Comparison of qPCR and RNA-seq fold-change values for the "Primates *vs.* Carnivora" comparison.
Abbreviation: ns=not significant; qPCR=quantitative polymerase chain reaction; DEG=differentially expressed genes.
* $P<0.05$, ** $P<0.01$, *** $P<0.001$.

clustering patterns. Apart from some strains from Sichuan and Shandong provinces forming clusters, the remaining strains showed a mosaic distribution without strong geographical signal.

DISCUSSION

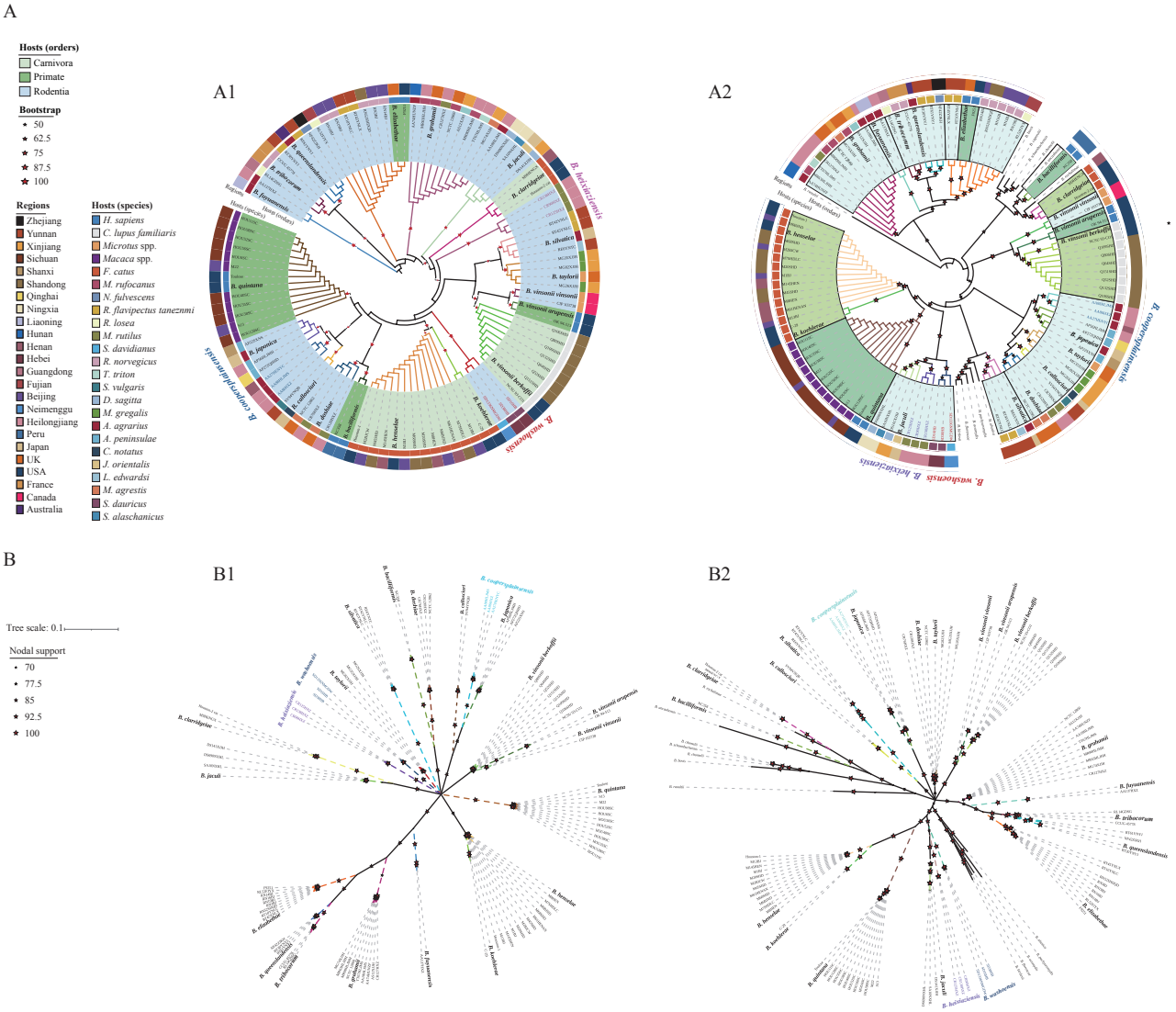

FIGURE 2. Phylogenetic trees based on core-gene SNPs. (A) ML trees constructed from transcriptomic (A1) and genomic (A2) data; (B) BI trees constructed from transcriptomic (B1) and genomic (B2) data.
Note: Branch colors indicate *Bartonella* species. Reference genomes are shown in bold; species lacking reference genomes are labeled on an outer ring with matching colors.
Abbreviation: SNP=Single nucleotide polymorphism; ML=Maximum likelihood; BI=Bayesian inference.

Consistent with the role of gene regulation in adaptation, hierarchical clustering revealed that *Bartonella* gene expression was conserved within species but variable between species (*5*). Furthermore, transcriptome composition differed significantly across species and host origin. Host specificity has been linked to the ability of pathogens to inject effector proteins into host cells via specialized secretion systems (*6*). Previous studies have demonstrated that *Bartonella* spp. rely primarily on the VirB/VirD4 T4SS to transport effector proteins into host cells, thereby modulating immune responses and disrupting cellular structures to promote colonization (*7*). In the present study, differential expression of *virB* and *bepC* was confirmed, and several novel host-associated DEGs were identified, including *tatC*, *tatA*, and *ftsY*. Given their implicated roles in virulence and adaptation in other pathogens, these genes may contribute to host-specific recognition in *Bartonella* spp. (*8*).

Traditional phylogenetic analyses have relied on mitochondrial and a limited number of nuclear gene markers, which are susceptible to stochastic errors (*9*). Advances in high-throughput RNA-seq now enable the application of comprehensive transcriptomic data to phylogenetic studies. In the present analysis, phylogenetic reconstruction based on core-gene SNPs

from both transcriptomic and genomic data revealed highly congruent tree topologies, validating the utility of transcriptomes for resolving taxonomic relationships within *Bartonella* spp. These analyses further revealed significant clustering of strains according to host origin, indicating that host association acted as a primary driver of species differentiation in *Bartonella* spp. In contrast to previous studies which had emphasized the role of geographic distribution in the genetic differentiation of *Bartonella*, our findings suggested that its influence was comparatively limited (*4*). This pattern could be attributed to increased cross-regional movement of host animals driven by modern socioeconomic development, which likely attenuated the impact of geographic barriers on the genetic evolution of *Bartonella* spp.

This study has several limitations. First, the limited number of strains representing certain *Bartonella* species, host types, and geographic regions may affect the generalizability of these findings. Future studies should expand the strain collection to determine whether expression profiles exhibit stronger host-specific or geographic clustering patterns. Second, the molecular mechanisms underlying the identified host-associated genes remain uncharacterized. Subsequent investigations should employ transcriptional regulation experiments to define the functional roles of these genes in host adaptation.

In conclusion, the transcriptome of *Bartonella* spp. varied systematically by species and host origin, supporting a model of host-driven speciation and evolution. This study demonstrates the utility of transcriptomics for elucidating evolutionary relationships in *Bartonella* spp. and potentially in other prokaryotes.

Conflicts of interest: No conflicts of interest.

Funding: Supported by the National Natural Science Foundation of China (31970005).

doi: 10.46234/ccdcw2026.039

[#] Corresponding author: Dongmei Li, lidongmei@icdc.cn.

[1] National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine, Beijing, China.

Submitted: October 28, 2025
Accepted: February 23, 2026
Issued: February 27, 2026

REFERENCES

1. McCormick DW, Rassoulian-Barrett SL, Hoogestraat DR, Salipante SJ, SenGupta D, Dietrich EA, et al. *Bartonella* spp. infections identified by molecular methods, United States. Emerg Infect Dis 2023;29(3):467 – 76. https://doi.org/10.3201/eid2903.221223.
2. Jin WT, Gernandt DS, Wehenkel C, Xia XM, Wei XX, Wang XQ. Phylogenomic and ecological analyses reveal the spatiotemporal evolution of global pines. Proc Natl Acad Sci USA 2021;118(20):e2022302118. https://doi.org/10.1073/pnas.2022302118.
3. Xu AL, Lu L, Zhang W, Song XP, Li GC, Miao Y, et al. Microevolution of *Bartonella grahamii* driven by geographic and host factors. mSystems 2024;9(10):e01089 – 24. https://doi.org/10.1128/msystems.01089-24.
4. Al-Hassani MKA, Kadhim HM, Al-Galebi AAS, Gharban HAJ. First molecular phylogeny of *Bartonella bovis* in Iraqi cattle. Open Vet J 2024;14(9):2361 – 7. https://doi.org/10.5455/OVJ.2024.v14.i9.24.
5. Faure E, Kwong K, Nguyen D. *Pseudomonas aeruginosa* in chronic lung infections: how to adapt within the host?. Front Immunol 2018;9:2416. https://doi.org/10.3389/fimmu.2018.02416.
6. Wang M, Guan QX, Wang CY, Hu LY, Hu XY, Xu ML, et al. Anchorage of bacterial effector at plasma membrane *via* selective phosphatidic acid binding to modulate host cell signaling. PLoS Pathog 2024;20(11):e1012694. https://doi.org/10.1371/journal.ppat.1012694.
7. Fromm K, Ortelli M, Boegli A, Dehio C. Translocation of YopJ family effector proteins through the VirB/VirD4 T4SS of *Bartonella*. Proc Natl Acad Sci USA 2024;121(20):e2310348121. https://doi.org/10.1073/pnas.2310348121.
8. Yi X, Chen YN, Cai HY, Wang JJ, Zhang YY, Zhu ZQ, et al. The temperature-dependent expression of type II secretion system controls extracellular product secretion and virulence in mesophilic *Aeromonas salmonida* SRW-OG1. Front Cell Infect Microbiol 2022;12:945000. https://doi.org/10.3389/fcimb.2022.945000.
9. Ran JH, Shen TT, Wu H, Gong X, Wang XQ. Phylogeny and evolutionary history of Pinaceae updated by transcriptomic analysis. Mol Phylogenet Evol 2018;129:106 – 16. https://doi.org/10.1016/j.ympev.2018.08.011.

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY TABLE S1. The background information of the *Bartonella* strains used in this study.

Species	Hosts	Regions	Strains
Bartonella callosciuri	*Sciurus vulgaris*	Yunnan, China	SV94YNQB
Bartonella clarridgeiae	*Felis catus*	Henan, China	M9HNGX
		USA	Houston-2 cat (ATCC 700693)
Bartonella coopersplainsensis	*Apodemus agrarius*	Henan, China	AA27HENYC
		Heilongjiang, China	AA86HXZ, AA88HLJMS
Bartonella doshiae	*Myodes rutilus*	Heilongjiang, China	CR120HXZ, CR76HXZ
	Microtus agrestis	UK	NCTC12862 (ATCC 700133)
Bartonella elizabethae	*Rattus losea*	Fujian, China	RL52FJYX
	Rattus norvegicus	Beijing, China	RN14BJ, RN16BJ, RN19BJ, RN3BJ
		Shandong, China	RN52SDQD
	Rattus tanezumi	Yunnan, China	RT10YNLC, RT45YNLX
	Homo sapiens	USA	F9251 (ATCC 49927)
Bartonella grahamii	*Apodemus agrarius*	Heilongjiang, China	AA18HLJMS
		Hunan, China	AA74HUNZF
	Allactaga sibirica	Xinjiang, China	AS12XJJH
	Myodes rutilus	Heilongjiang, China	CR127HXZ
	Microtus agrestis	Xinjiang, China	MG74XJJH
	Myodes rufocanus	Heilongjiang, China	MR63HLJHH, MR90HLJMH
	Tscherskia triton	Heilongjiang, China	TT67HLJMS
Bartonella heixiaziensis	*Myodes rufocanus*	UK	NCTC 12860 (ATCC 700132)
	Myodes rutilus	Heilongjiang, China	CR125HXZ, CR138HXZ, CR90HXZ
Bartonella henselae	*Felis catus*	Beijing, China	M13BJ, M1BJ, M2BJCW, M6BJND
		Henan, China	M145HEN, M61HENAN, M8HEN
		Shandong, China	M20SHD, M45SHDLC, M68SHDLC, M7SHDLC
		USA	Houston-1 (ATCC 49882)
Bartonella jaculi	*Dipus sagitta*	Xijiang, China	DS141XJJH
		Ningxia, China	DS690NXHL
	Spermophilus alaschanicus	Ningxia, China	SA10NXHL
Bartonella japonica	*Apodemus peninsulae*	Shanxi, China	AP22SXNS
		Qinghai, China	AP272QHHD
		Heilongjiang, China	AP56HLJMH
Bartonella queenslandensis	*Rattus tanezumi*	Yunnan, China	RT30YNYJ, RT61YNYJ
	Niviventer fulvescens	Zhejiang, China	NF42ZJKH
Bartonella quintana	*Macaca mulatta*	Sichuan, China	HOU11SC, HOU128SC, HOU35SC, HOU38SC, HOU40SC, HOU52SC, HOU56SC, HOU6SC, HOU98SC
		Beijing, China	M22, S13
Bartonella silvatica	*Homo sapiens*	USA	Toulouse (ATCC VR-358)
	Leopoldamys edwardsi	Yunnan, China	RE6YNTC
	Rattus tanezumi	Yunnan, China	RT42YNLC, RT45YNLC
Bartonella taylorii	*Microtus gregalis*	Xinjiang, China	MG26XJJH, MG35XJJH, MG82XJJH
Bartonella fuyuanensis	*Apodemus agrarius*	Heilongjiang, China	AA137HXZ

Continued

Species	Hosts	Regions	Strains
Bartonella tribocorum	*Rattus norvegicus*	France	CCUG 45778 (CIP 105476)
	Rattus losea	Guangdong, China	RL14GDSG
Bartonella vinsonii berkhoffii	*Canis lupus familiaris*	Shandong, China	Q106SHD, Q109SHD, Q132SHD, Q151SHD, Q52SHD, Q64SHD, Q80SHD
		USA	NCSU 93-CO1 (ATCC 51672)
Bartonella washoensis	*Sciurotamias davidianus*	Inner Mongolia, China	SD1336NMGDW
	Spermophilus dauricus	Hebei, China	SD16HB, SD38HB
Bartonella vinsonii arupensis	*Homo sapiens*	USA	OK 94-513 (ATCC 700727)
Bartonella vinsonii vinsonii	*Microtus* spp.	Canada	CIP 103738 (ATCC VR-152)
Bartonella bacilliformis	*Homo sapiens*	Peru	KC358 (ATCC 35685)
Bartonella koehlerae	*Felis catus*	USA	C-29 (ATCC 700693)

SUPPLEMENTARY TABLE S2. Comparative results of differentially expressed genes for strains from different *Bartonella* species.

Groups	Up	Down	Total
B. grahamii vs. B. fuyuanensis	183	176	359
B. grahamii vs. B. vinsonii berkhoffii	3,547	6,512	10,059
B. grahamii vs. B. bacilliformis	2,162	1,303	3,465
B. grahamii vs. B. callosciuri	1,316	1,656	2,972
B. grahamii vs. B. clarridgeiae	1,205	4,113	5,318
B. grahamii vs. B. coopersplainsensis	2,867	2,733	5,600
B. grahamii vs. B. doshiae	2,901	3,287	6,187
B. grahamii vs. B. elizabethae	1,991	1,783	3,774
B. grahamii vs. B. henselae	2,533	5,239	7,772
B. grahamii vs. B. quintana	2,149	4,815	6,964
B. grahamii vs. B. heixiaziensis	3,428	2,574	6,002
B. grahamii vs. B. jaculi	3,381	2,303	5,684
B. grahamii vs. B. japonica	3,316	2,479	5,795
B. grahamii vs. B. queenslandensis	968	1,676	2,644
B. grahamii vs. B. silvatica	2,378	2,474	4,852
B. grahamii vs. B. taylorii	3,888	2,324	6,212
B. grahamii vs. B. washoensis	3,889	2,425	6,314
B. grahamii vs. B. tribocorum	1,141	1,720	2,861
B. grahamii vs. B. koehlerae	2,043	2,227	4,270
B. grahamii vs. B.vinsonii arupensis	2,099	2,370	4,469
B. grahamii vs. B. vinsonii vinsonii	2,241	2,079	4,320

SUPPLEMENTARY TABLE S3. Comparative results of differentially expressed genes for *Bartonella* strains from different host origins.

Groups	Conditions	Up	Down	Total
	Orders			
Primates vs. Rodentia	/	5,863	3,068	8,931
Primates vs. Carnivora	/	4,039	3,314	7,353
Rodentia vs. Carnivora	/	3,070	10,845	13,915
	Species			
C. lupus familiaris vs. *F. catus*	Carnivora	2,699	3,291	5,990
H. sapiens vs. *M. mulatta*	Primates	432	5,364	5,796
A. agrarius vs. *M. rutilus*		2,323	3,189	5,512
A. agrarius vs. *Rattus tanezumi*	Rodentia	696	2,072	2,768
A. agrarius vs. *R. norvegicus*		1,512	3,031	4,543
	Restriction of Bartonella species			
M. agrestis vs. *M. rutilus*	*B. doshiae*	959	740	1,699
S. alaschanicus vs. *D. sagitta*	*B. jaculi*	355	177	532
N. fulvescens vs. *Rattus tanezumi*	*B. queenslandensis*	570	241	811
H. sapiens vs. *M. mulatta*	*B. quintana*	904	1,105	2,009
L. edwardsi vs. *Rattus tanezumi*	*B. silvatica*	493	1,074	1,567
S. davidianus vs. *S. dauricus*	*B. washoensis*	118	249	367
R. losea vs. *R. norvegicus*		212	11	223
R. losea vs. *Rattus tanezumi*		260	218	478
R. norvegicus vs. *Rattus tanezumi*	*B. elizabethae*	1,132	425	1557
H. sapiens vs. *R. norvegicus*		320	124	444
H. sapiens vs. *Rattus tanezumi*		1,943	210	2,153
A. agrarius vs. *A. sibirica*		1,074	493	1,567
A. agrarius vs. *M. rufocanus*		358	293	651
A. agrarius vs. *Myodes rutilus*		167	1,055	1,222
M. rufocanus vs. *Myodes rutilus*		11	441	452
M. rufocanus vs. *M. agrestis*	*B. grahamii*	13	345	358
M. rufocanus vs. *A. sibirica*		22	493	515
A. agrarius vs. *M. agrestis*		165	624	789
A. agrarius vs. *T. triton*		212	623	835
M. rufocanus vs. *T. triton*		53	364	417

SUPPLEMENTARY TABLE S4. Primer sequences for validating transcriptome reliability.

Gene name	Forward primer (5'-3')	Reverse primer (5'-3')
gltA	GGGGACCAGCTCATGGTGG	AATGCAAAAAGAACAGTAAACA
bepC	AAAGCAGCAGAGGCAGTGTATCG	TTGACGGGAAAATCTTCGGCTACG
secDF	CGCTAAGTCTTCTCGGTGCTAC	TCTTCACGGATACGCTCATTGATTAG
secB	GTGAAATTAACAACAATGGTGGAGAAC	TCCCGTGGACGCAGTGAAC
fstY	TTATGCTTGCCGCCTGTGATATG	GCATCTGCACCTAATTTTGTTGAAAC