

Supplementary information for manuscript:

Horizontal Transfers and Gene Losses in the phospholipid pathway of *Bartonella* reveal clues about early ecological niches

Authors: Qiyun Zhu, Michael Kosoy, Kevin J. Olival, Katharina Dittmar

Contents

Figure S1 Core genome phylogeny of *Bartonella*

Figure S2 Standard phylogenetic analysis of *Bartonella*

Figure S3 Phylogeny of the Ugp system in *Bartonella*

Figure S4 Phylogeny of *Bartonella plsX* gene

Table S1 Additional genomic data assessed in this study

Table S2 Selection test results

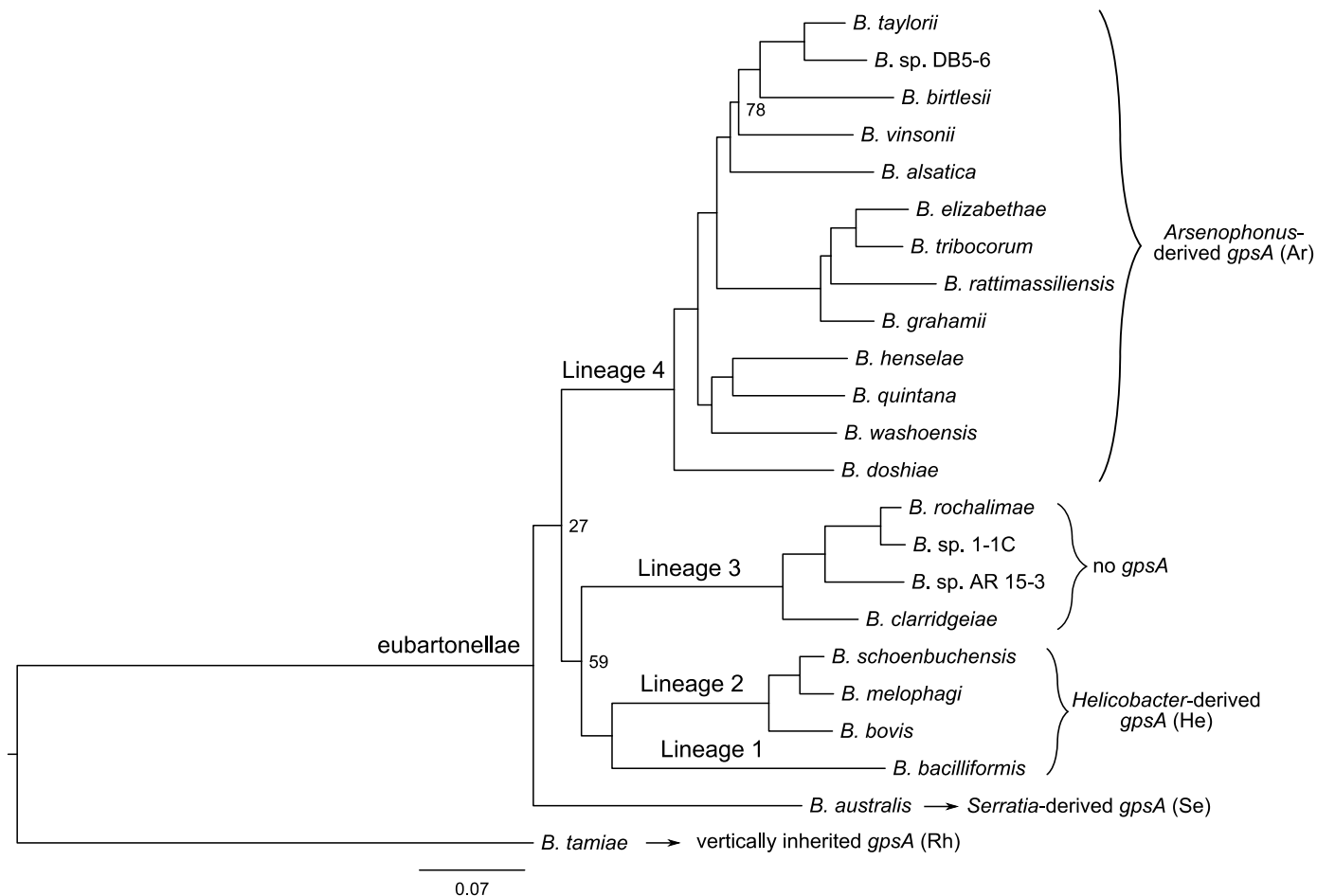


Figure S1 Core genome phylogeny of *Bartonella*

The tree was reconstructed based on 516 single-copy genes shared by 23 *Bartonella* species using the maximum likelihood (ML) method as implemented in RAxML. Major *Bartonella* lineages according to Engel et al. (2011) are indicated. Node labels are bootstrap support values (out of 100 replicates). Nodes without labels are fully supported (bootstrap = 100).

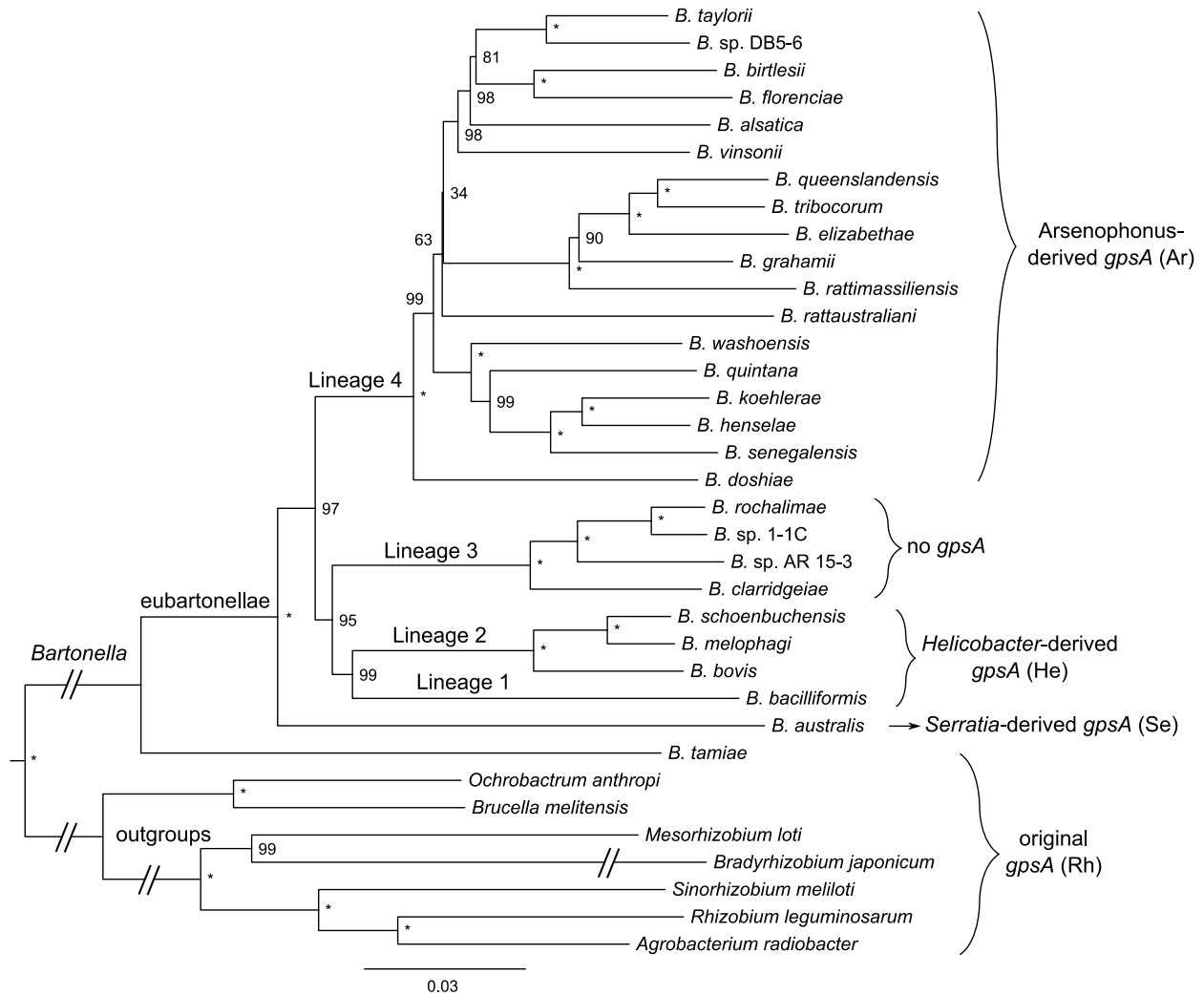


Figure S2 Standard phylogenetic analysis of *Bartonella*

Tree topology based on commonly used genetic markers of 28 *Bartonella* species, rooted to Rhizobiales. *Bartonella* species are ingroup. Major *Bartonella* lineages according to Engel et al (2011) are indicated. Node labels are bootstrap support values (out of 100 replicates). Asterisks (*) indicate 100% support.

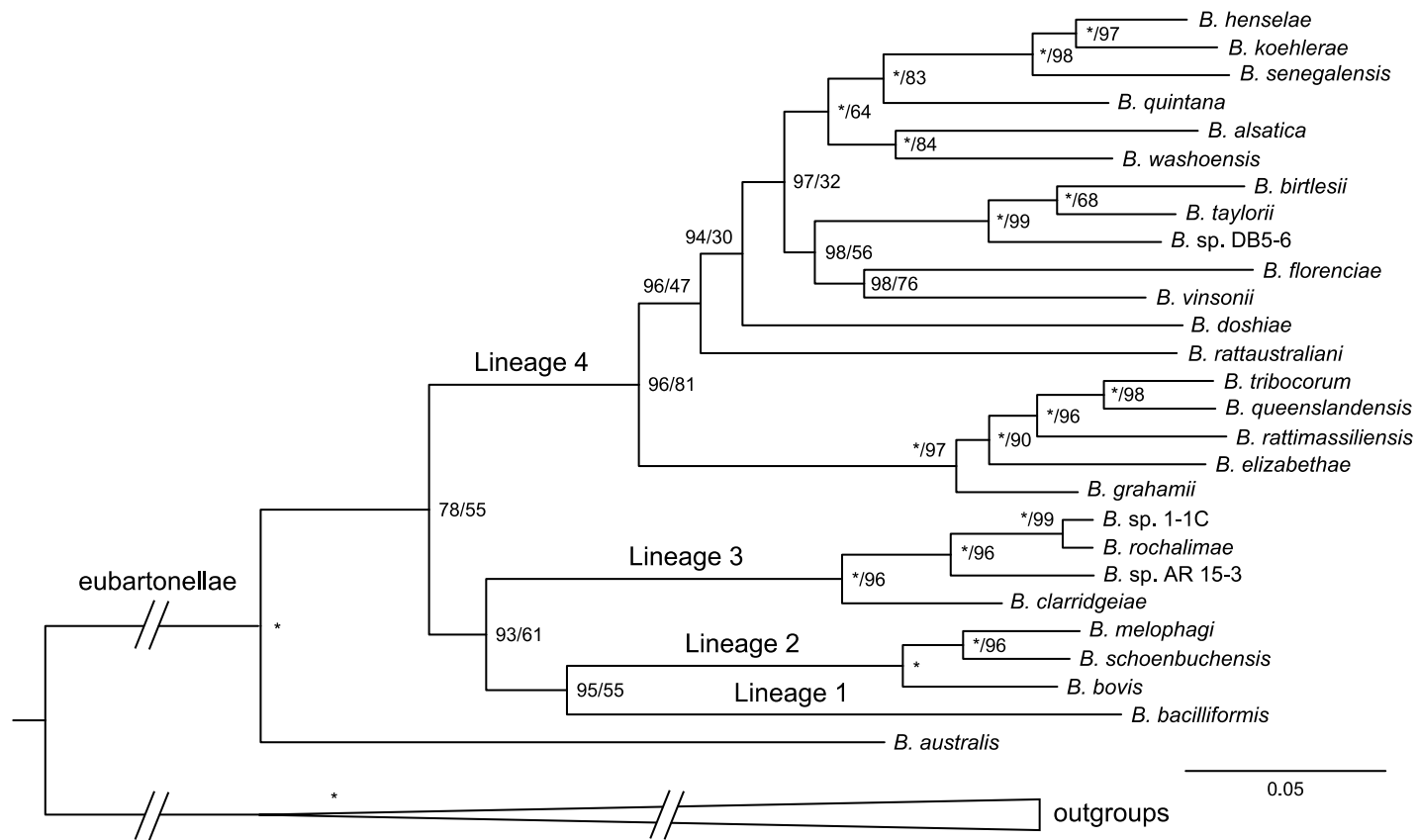


Figure S3 Phylogeny of the Ugp system in *Bartonella*

Four genes that constitute the *ugp* operon: *ugpB*, *A*, *E* and *C*, were concatenated, a phylogenetic tree was inferred in MrBayes. Node labels are Bayesian posterior probabilities / maximum-likelihood bootstrap support values. Asterisks (*) indicate 100% support. The tree is rooted to Rhizobiales outgroups (Table S1), which are collapsed into a triangle.

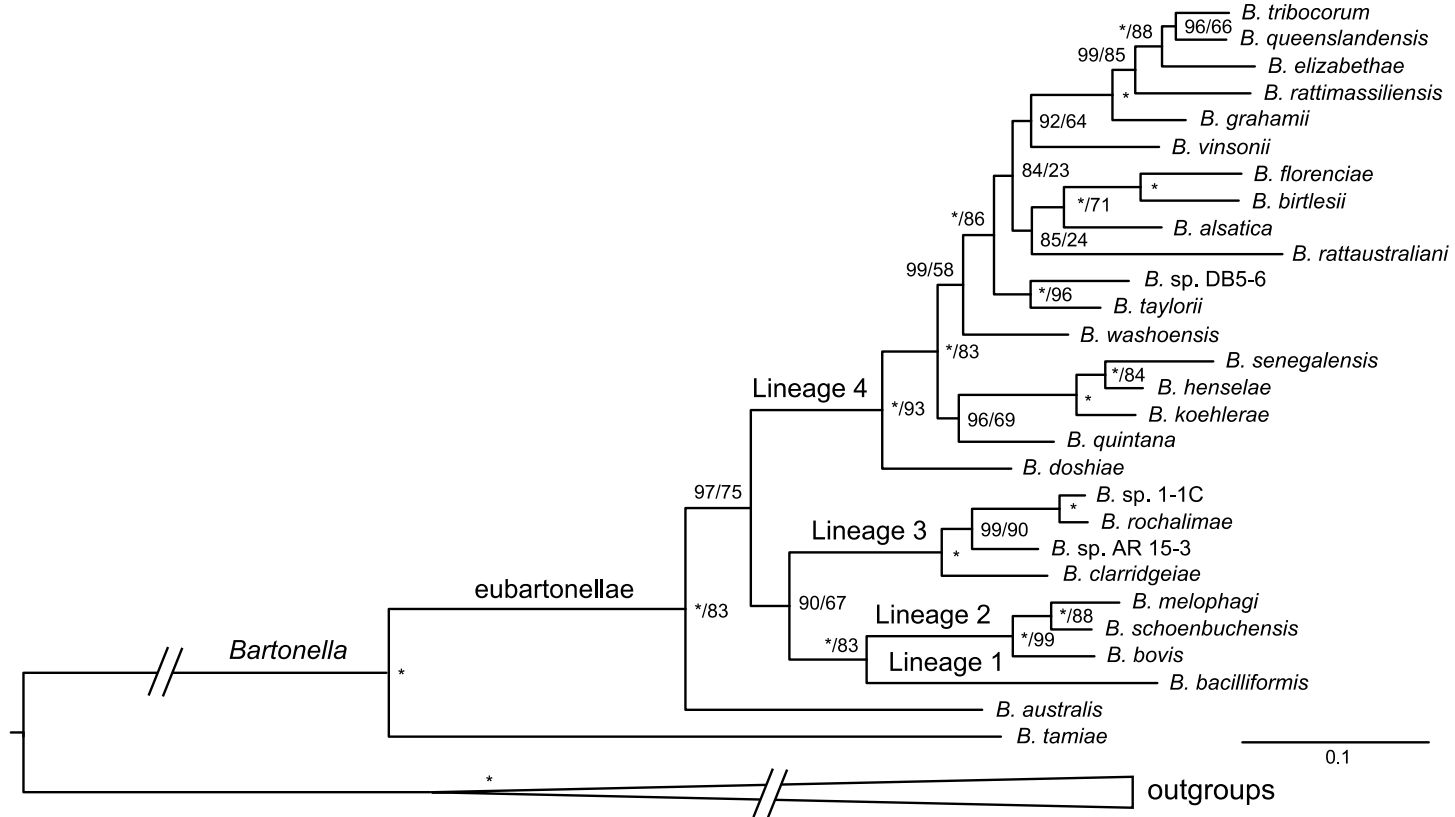


Figure S4 Phylogeny of *Bartonella plsX* gene

Node labels are Bayesian posterior probabilities / maximum-likelihood bootstrap support values. Asterisks (*) indicate 100% support. The tree is rooted to Rhizobiales outgroups (Table S1), which are collapsed into a triangle.

Table S1 Additional genomic data assessed in this study

Genus	Species	Strain	NCBI Assembly
Outgroups of <i>Bartonella</i> (Figure 2, S1-4)			
<i>Ochrobactrum</i>	<i>O. anthropi</i>	ATCC 49188	GCF_000017405
<i>Brucella</i>	<i>B. melitensis</i>	bv. 1 str. 16M	GCF_000007125
<i>Mesorhizobium</i>	<i>M. loti</i>	MAFF303099	GCF_000009625
<i>Sinorhizobium</i>	<i>S. meliloti</i>	1021	GCF_000006965
<i>Rhizobium</i>	<i>R. leguminosarum</i>	bv. viciae 3841	GCF_000009265
<i>Agrobacterium</i>	<i>A. radiobacter</i>	K84	GCF_000016265
<i>Bradyrhizobium</i>	<i>B. japonicum</i>	USDA 110	GCF_000011365
Other Rhizobiales (Figure 3A)			
<i>Bartonella</i>	<i>B. tamiae</i>	Th307	GCF_000279995
<i>Brucella</i>	<i>B. abortus</i>	S19	GCF_000018725
	<i>B. microti</i>	CCM 4915	GCF_000022745
	<i>B. pinnipedialis</i>	B2/94	GCF_000221005
	<i>B. ovis</i>	ATCC 25840	GCF_000016845
	<i>B. canis</i>	ATCC 23365	GCF_000018525
	<i>B. suis</i>	1330	GCF_000007505
<i>Ochrobactrum</i>	<i>O. sp. CDB2</i>	-	GCF_000344725
	<i>O. intermedium</i>	M86	GCF_000332835
<i>Mesorhizobium</i>	<i>M. ciceri</i>	bv. biserrulae WSM1271	GCF_000185905
	<i>M. opportunistum</i>	WSM2075	GCF_000176035
	<i>M. australicum</i>	WSM2073	GCF_000230995
<i>Nitratireductor</i>	<i>N. aquibiodomus</i>	RA22	GCF_000265055
	<i>N. indicus</i>	C115	GCF_000300515
	<i>N. pacificus</i>	pht-3B	GCF_000300335
<i>Rhizobium</i>	<i>R. etli</i>	CFN 42	GCF_000092045
	<i>R. tropici</i>	CIAT 899	GCF_000330885
<i>Agrobacterium</i>	<i>A. fabrum</i>	str. C58	GCF_000092025
	<i>A. sp. H13-3</i>	-	GCF_000192635
	<i>A. vitis</i>	S4	GCF_000016285
<i>Sinorhizobium</i>	<i>S. medicae</i>	WSM419	GCF_000017145
	<i>S. fredii</i>	NGR234	GCF_000018545
<i>Bradyrhizobium</i>	<i>B. sp. ORS 278</i>	-	GCF_000026145
<i>Rhodopseudomonas</i>	<i>R. palustris</i>	CGA009	GCF_000195775
<i>Xanthobacter</i>	<i>X. autotrophicus</i>	Py2	GCF_000017645
<i>Methylobacterium</i>	<i>M. extorquens</i>	PA1	GCF_000018845
Helicobacteraceae and other Campylobacterales (Figure 3B)			
<i>Helicobacter</i>	<i>H. pylori</i>	G27	GCF_000021165
	<i>H. pylori</i>	26695	GCF_000307795
	<i>H. pylori</i>	B38	GCF_000091345
	<i>H. pylori</i>	J99	GCF_000008785
	<i>H. acinonychis</i>	str. Sheeba	GCF_000009305
	<i>H. bizzozeronii</i>	CIII-1	GCF_000237285
	<i>H. felis</i>	ATCC 49179	GCF_000200595
	<i>H. mustelae</i>	12198	GCF_000091985
	<i>H. cinaedi</i>	CCUG 18818	GCF_000155475
	<i>H. hepaticus</i>	ATCC 51449	GCF_000007905
	<i>H. canadensis</i>	MIT 98-5491	GCF_000162575
	<i>H. winghamensis</i>	ATCC BAA-430	GCF_000158455
	<i>Wolinella</i>	<i>W. succinogenes</i>	DSM 1740
<i>Campylobacter</i>	<i>C. concisus</i>	13826	GCF_000017725
	<i>C. fetus</i>	subsp. fetus 82-40	GCF_000015085
<i>Arcobacter</i>	<i>A. butzleri</i>	JV22	GCF_000185325

Genus	Species	Strain	NCBI Assembly	
Enterobacteriales and other Gammaproteobacteria (Figure 3C)				
<i>Arsenophonus</i>	<i>A. nasoniae</i>	DSM 15247	GCA_000429565	
<i>Providencia</i>	<i>P. alcalifaciens</i>	DSM 30120	GCF_000173415	
	<i>P. rettgeri</i>	DSM 1131	GCF_000158055	
	<i>P. rustigianii</i>	DSM 4541	GCF_000156395	
	<i>P. stuartii</i>	ATCC 25827	GCF_000154865	
	<i>P. asymbiotica</i>	ATCC 43949	GCF_000196475	
<i>Photorhabdus</i>	<i>P. luminescens</i>	subsp. laumondii TTO1	GCF_000196155	
	<i>Xenorhabdus</i>	<i>X. bovienii</i>	SS-2004	GCF_000027225
<i>Proteus</i>	<i>X. nematophila</i>	ATCC 19061	GCF_000252955	
	<i>P. mirabilis</i>	HI4320	GCF_000069965	
	<i>P. penneri</i>	ATCC 35198	GCF_000155835	
<i>Morganella</i>	<i>P. vulgaris</i>	CCUG 19017	GU254061	
	<i>M. morganii</i>	subsp. morganii KT	GCF_000286435	
	<i>Serratia</i>	<i>S. proteamaculans</i>	568	GCF_000018085
		<i>S. liquefaciens</i>	ATCC 27592	GCF_000422085
<i>S. plymuthica</i>		AS9	GCF_000214235	
<i>Yersinia</i>	<i>S. symbiotica</i>	str. 'Cinara cedri'	GCF_000238975	
	<i>S. marcescens</i>	WW4	GCF_000336425	
	<i>Y. enterocolitica</i>	subsp. palearctica 105.5R(r)	GCF_000192105	
	<i>Y. kristensenii</i>	ATCC 33638	GCF_000173715	
	<i>Y. pestis</i>	A1122	GCF_000222975	
<i>Candidatus Hamiltonella</i>	<i>Y. rohdei</i>	ATCC 43380	GCF_000173775	
	<i>H. defensa</i>	5AT	GCF_000021705	
	<i>Dickeya</i>	<i>D. dadantii</i>	Ech703	GCF_000023545
<i>D. zeae</i>		Ech1591	GCF_000023565	
<i>Pectobacterium</i>	<i>P. atrosepticum</i>	SCRI1043	GCF_000011605	
<i>Candidatus Blochmannia</i>	<i>B. floridanus</i>	-	GCF_000043285	
	<i>B. pennsylvanicus</i>	str. BPEN	GCF_000011745	
	<i>B. vafer</i>	str. BVAF	GCF_000185985	
<i>Wigglesworthia</i>	<i>W. glossinidia</i>	endosymbiont of <i>Glossina morsitans morsitans</i> (Yale colony)	GCF_000247565	
<i>Baumannia</i>	<i>B. cicadellinicola</i>	str. Hc (<i>Homalodisca coagulata</i>)	GCF_000013185	
<i>Sodalis</i>	<i>S. glossinidius</i>	str. 'morsitans'	GCF_000010085	
<i>Escherichia</i>	<i>E. albertii</i>	albertii TW07627	GCF_000155105	
	<i>E. coli</i>	2534-86	GCF_000225065	
	<i>E. fergusonii</i>	ATCC 35469	GCF_000026225	
	<i>Shigella</i>	<i>S. boydii</i>	Sb227	GCF_000012025
		<i>S. dysenteriae</i>	Sd197	GCF_000012005
<i>S. flexneri</i>		K-315	GCF_000268165	
<i>Citrobacter</i>	<i>S. sonnei</i>	Ss046	GCF_000092525	
	<i>C. koseri</i>	ATCC BAA-895	GCF_000018045	
	<i>C. rodentium</i>	ICC168	GCF_000027085	
<i>Salmonella</i>	<i>S. bongori</i>	NCTC 12419	GCF_000252995	
	<i>S. enterica</i>	subsp. enterica serovar Typhi str. P-stx-12	GCF_000245535	
<i>Enterobacter</i>	<i>E. cloacae</i>	subsp. dissolvens SDM	GCF_000235765	
	<i>E. radicincitans</i>	DSM 16656	GCF_000280495	
	<i>E. sp. 638</i>	-	GCF_000016325	
<i>Yokenella</i>	<i>Y. regensburgei</i>	ATCC 43003	GCF_000239335	
<i>Klebsiella</i>	<i>K. oxytoca</i>	E718	GCF_000276705	
	<i>K. pneumonia</i>	KCTC 2242	GCF_000220485	
	<i>K. variicola</i>	At-22	GCF_000025465	
<i>Cronobacter</i>	<i>C. sakazakii</i>	ES15	GCF_000263215	
	<i>C. turicensis</i>	z3032	GCF_000027065	
<i>Erwinia</i>	<i>E. amylovora</i>	ATCC 49946	GCF_000027205	

Genus	Species	Strain	NCBI Assembly
	<i>E. billingiae</i>	Eb661	GCF_000196615
	<i>E. pyrifoliae</i>	DSM 12163	GCF_000026985
	<i>E. tasmaniensis</i>	ET1/99	GCF_000026185
<i>Pantoea</i>	<i>P. ananatis</i>	AJ13355	GCF_000270125
	<i>P. sp. GM01</i>	-	GCF_000282675
	<i>P. sp. YR343</i>	-	GCF_000282695
	<i>P. vagans</i>	C9-1	GCF_000148935
<i>Actinobacillus</i>	<i>A. pleuropneumoniae</i>	serovar 3 str. JL03	GCF_000016685
<i>Haemophilus</i>	<i>H. sputorum</i>	CCUG 13788	GCF_000238795
<i>Pasteurella</i>	<i>P. dagmatis</i>	ATCC 43325	GCF_000163475
	<i>P. multocida</i>	subsp. multocida str. Pm70	GCF_000006825
<i>Vibrio</i>	<i>V. cholera</i>	O395	GCF_000021625
	<i>V. cyclitrophicus</i>	ZF14 subcontig_64	GCF_000256265
	<i>V. fischeri</i>	SR5	GCF_000241785
	<i>V. parahaemolyticus</i>	K5030	GCF_000182465

Table S2 Selection test results

Gene	Tree	M0	Site models		Foreground branch	Branch model			Branch site models	
		ω	LRT (M2a vs. M1a)	LRT (M8a vs. M8)		ω_0	ω_1	LRT (one-ratio vs. two-ratio)	LRT (A1 vs. A)	Positive sites (BEB)
<i>gpsA</i> (Rh)	Figure 3A	0.0909	0.00	0.00	<i>B. tami</i> ae clade	0.0939	0.0035	16.90***	5.05*	26C, 80Y, 83L, 146E, 150K, 258N*, 271A, 284R, 302K, 314R*, 328S
					<i>Agrobacterium</i> + <i>Rhizobium</i> clade	0.0867	0.0991	1.45	-	
					<i>Brucella</i> clade	0.0913	0.0810	0.14	-	
					<i>Ochrobactrum</i> + <i>Brucella</i> clade	0.0919	0.0865	0.22	-	
					<i>Mesorhizobium</i> clade	0.0917	0.0856	0.23	-	
					<i>Nitratireductor</i> clade	0.0974	0.0525	13.81***	-	
					<i>Sinorhizobium</i> clade	0.0911	0.0891	0.02	-	
<i>gpsA</i> (He)	Figure 3B	0.0730	0.00	0.00	<i>Bartonella</i> (L1+L2) node	0.0754	0.0012	8.37***	3.19	62S, 102K*, 230Q
					<i>Bartonella</i> (L1+L2) clade	0.0298	0.1563	52.9***	0.00	102K, 118T, 124H, 230Q
<i>gpsA</i> (En)	Figure 3C	0.0887	0.00	0.00	<i>Bartonella</i> (L4) node	0.0886	0.0889	0.00	0.88	102N*, 150I, 319G, 333T*
					<i>Bartonella</i> (L4) clade	0.0694	0.1742	240.00***	0.00	24A**, 37Y**, 38N**, 49H**, 54A**, 60Q*, 77A*, 79R**, 91H**, 95N*, 97Q**, 102N**, 104H**, 116H**, 121L*, 166S*, 176D*, 239L**, 267I**, 275I**, 282I*, 296R**, 297M**, 314H**, 319G**, 331S*, 333T**
					<i>Arsenophonus</i> node	0.0889	0.0369	0.83	-	
					<i>Arsenophonus</i> clade	0.0883	0.1194	1.42	-	
					<i>Arsenophonus-Bartonella</i> (L4) node	0.0893	0.0498	5.03*	-	
					<i>B. australis</i> tip	0.0877	0.2050	10.49**	0.00	40S, 107V, 108L, 123W*, 218V*, 284R, 301H, 303D
<i>ugpB-A-E-C</i>	Figure S3	0.0844	-	-	<i>B. australis</i> tip	0.0846	0.0825	0.04	-	
					<i>Bartonella</i> (L1) tip	0.0837	0.0986	1.55	-	
					<i>Bartonella</i> (L2) clade	0.0816	0.1160	12.64***	-	
					<i>Bartonella</i> (L3) clade	0.0856	0.0756	1.86	-	
					<i>Bartonella</i> (L4) clade	0.0921	0.0806	4.94*	-	
<i>plsX</i>	Figure S4	0.0826	0.00	2.50	<i>Bartonella</i> (L3) clade	0.0794	0.1100	3.83	-	

The table reports ω (dN/dS value), LRT (likelihood ratio test value, = $2(\ln L_1 - \ln L_0)$) and its significance level (indicated by *), as well as putative positively selected sites with significance level (indicated by *) estimated by the BEB method. Branches of interest were labeled as foreground branches in the analyses. "Node" refers to a single branch. "Clade" refers to the selected branch and all sub-branches after it. ω under the one-ratio model (M0) is the global ω for the whole tree. ω_1 and ω_0 under the branch model refer to the ω values for the foreground branch and the tree background, respectively. Coordinates and residues of putative positively selected sites refer to the protein sequence alignment in Figure 4.