Horizontal Transfers and Gene Losses in the Phospholipid Pathway of *Bartonella* Reveal Clues about Early Ecological Niches

Qiyun Zhu[1,*], Michael Kosoy[2], Kevin J. Olival[3], and Katharina Dittmar[1,4,*]

[1]Department of Biological Sciences, University at Buffalo, State University of New York

[2]Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Foothills Research Campus, Fort Collins, Colorado

[3]EcoHealth Alliance, New York, New York

[4]Graduate Program of Evolution, Ecology, and Behavior, University at Buffalo, State University of New York

*Corresponding author: E-mail: qiyunzhu@buffalo.edu; kd52@buffalo.edu.

Accepted: July 28, 2014

Data deposition: This project has been deposited at the EMBL/GenBank data libraries under the accession KJ606299–KJ606321.

Abstract

Bartonellae are mammalian pathogens vectored by blood-feeding arthropods. Although of increasing medical importance, little is known about their ecological past, and host associations are underexplored. Previous studies suggest an influence of horizontal gene transfers in ecological niche colonization by acquisition of host pathogenicity genes. We here expand these analyses to metabolic pathways of 28 *Bartonella* genomes, and experimentally explore the distribution of bartonellae in 21 species of blood-feeding arthropods. Across genomes, repeated gene losses and horizontal gains in the phospholipid pathway were found. The evolutionary timing of these patterns suggests functional consequences likely leading to an early intracellular lifestyle for stem bartonellae. Comparative phylogenomic analyses discover three independent lineage-specific reacquisitions of a core metabolic gene—NAD(P)H-dependent glycerol-3-phosphate dehydrogenase (*gpsA*)—from Gammaproteobacteria and Epsilonproteobacteria. Transferred genes are significantly closely related to invertebrate *Arsenophonus*-, and *Serratia*-like endosymbionts, and mammalian *Helicobacter*-like pathogens, supporting a cellular association with arthropods and mammals at the base of extant *Bartonella* spp. Our studies suggest that the horizontal reacquisitions had a key impact on bartonellae lineage specific ecological and functional evolution.

Key words: *Bartonella*, horizontal gene transfer, gene loss, phospholipid pathway, intracellularity, host association.

Introduction

The increased availability of whole-genome data is providing more comprehensive insights into microbial evolution (Toft and Andersson 2010). One phenomenon of bacterial evolution concerns a process known as horizontal gene transfer (HGT), where bacteria transfer genetic material to related or to unrelated bacterial lineages (Doolittle 1999; Doolittle et al. 2003). From a biological perspective, HGT is vital for the origination of new bacterial functions, including virulence, pathogenicity, or antibiotic resistance (Koonin et al. 2001; Gophna et al. 2004; Barlow 2009).

Instances of HGT also contain important information about evolutionary events in the bacterial lineage. Specifically, uneven distribution patterns of genes across lineages speak not only to the potential presence of HGT but also to its frequency throughout lineage evolution. Depending on the evolutionary history posttransfer, an HGT event may be informative about the directionality and mode of transfer, allowing identification of donor and recipient genomes.

Because many bacteria have niche preferences (e.g., intracellular or extracellular habitat, host species range, tissue tropism, etc.), the identification of the donating lineage may provide specific information about the nature of the environment the exchange took place in. This is particularly interesting for intracellular pathogens, as it implies HGT to have occurred in a specific environment—the host cells. The taxonomic

© The Author(s) 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

identification of putative bacterial donors therefore allows inferences about the ancestral bacterial community composition at the time of exchange, although the extant host range and microbiome diversity may have changed.

With this in mind, we analyzed 28 currently available genomes of the bacteria *Bartonella* for HGT events. *Bartonella* species are Gram-negative Alphaproteobacteria and are thought to persist mainly as facultative intracellular invaders. They have been classified as emergent pathogens and are ubiquitously associated with mammals, where they parasitize erythrocytes and endothelial cells (Pulliainen and Dehio 2012). More than half of the known *Bartonella* species are pathogenic to humans, and clinical manifestations vary from acute intraerythrozytic bacteremia to vasoproliferative tumor growth (Kaiser et al. 2011; Harms and Dehio 2012). Although it is known that bartonellae readily straddle the boundary between mammals and invertebrates, their ecological past remains obscure (Chomel et al. 2009). Phylogenetically, bartonellae form a derived monophyletic clade within the mostly plant associated Rhizobiales (Gupta and Mok 2007; Engel et al. 2011; Guy et al. 2013). Bartonellae have been increasingly detected in a broad range of blood-feeding or biting insects, and recent research on their diversity in blood-feeding insects suggests an early association to fleas (Tsai et al. 2011; Morick et al. 2013), but the full range of invertebrate associations is still underexplored.

Evidence for HGT in bartonellae has been found in previous studies, which mainly concentrated on the identification of gene transfer agents involved in the spread of known host adaptability and pathogenicity genes, including the T4SS secretion system (VirB, Trw, and Vbh) (Berglund et al. 2009; Saisongkorh et al. 2010). Surprisingly, little information exists on the horizontal transfer of other, more fundamental operational genes (i.e., metabolic genes), which may also have implications for host-adaptation. Specifically, bartonellae are thought to be in the early stages of a transition to stable intracellularity (Toft and Andersson 2010). Although they are stealthy pathogens in their mammalian hosts, and can survive and reproduce intra- and extracellularly, they have also been discussed as intracellular endosymbionts in their insect hosts (e.g., fleas, ked flies, bat flies) (Tsai et al. 2011). This transitional lifestyle has genomic ramifications, which have been associated with processes of gene loss, HGT, and recombination that specifically affect genes coding for cell membrane formation (outer surface structures), or intermediate metabolism (Zientz et al. 2004; Toft and Andersson 2010). Consequently, exploring the signature, provenance, and order (timing) of HGT events in metabolic pathways may be crucial in understanding the particular steps involved in the development of an intracellular lifestyle in bartonellae.

We here present horizontal and vertical patterns in the evolution of the core metabolic phospholipid pathway in bartonellae. Specifically, we employed an initial discovery screen for HGT, followed by comparative genomic analyses and validation, phylogenetics and experimental approaches to explore the evolutionary successions of gains and losses of genes, with the goal to elucidate the ancestral and extant biological associations of bartonellae on organismal and cellular levels.

Materials and Methods

Taxon Sampling

Genomic data of *Bartonella* and other bacterial organisms related to this study were downloaded from the National Center for Biotechnology Information (NCBI) GenBank (http://www.ncbi.nlm.nih.gov/, last accessed August 15, 2014) or from the website of the *Bartonella* Group Sequencing Project, Broad Institute of Harvard and the Massachusetts Institute of Technology (http://www.broadinstitute.org/, last accessed August 15, 2014). *Bartonella* species were grouped into four lineages (L1–L4) plus *B. tamiae* and *B. australis*, following the current taxonomy (Engel et al. 2011; Pulliainen and Dehio 2012; Guy et al. 2013). For the purpose of this study, we will refer to all bartonellae except *B. tamiae* as eubartonellae. This is based on the recognition that *B. tamiae* has been described as clearly distinct from all other currently known bartonellae lineages (Kosoy et al. 2008; Guy et al. 2013). A total of 28 *Bartonella* species were examined in this study (table 1).

BLAST Hit Distribution Analysis of *Bartonella* Genomes—Initial Discovery Screen

Initial discovery analysis of putative HGT events in metabolic pathways was assisted by an automated pipeline (available in the Dittmar Lab: https://github.com/DittmarLab/HGTector, last accessed August 15, 2014). This pipeline is based on a computational method of rapid, exhaustive, and genome-wide detection of HGT, featuring the systematic analysis of BLAST hit distribution patterns combined with a priori defined hierarchical evolutionary categories (Zhu et al. 2014). Batch BLASTP of *Bartonella* protein-coding genes was performed against the entire NCBI nr database (E value cutoff = 1×10^{-5}, other parameters remain default). Genes that have less than a statistically relevant threshold of the expected number of hits based on known close relatives of *Bartonella*, but meanwhile show multiple top hits from taxonomically distant organisms (non-Rhizobiales groups), were considered to be candidates of HGT-derived genes and were subject to further phylogenetic analyses (see below) (see Zhu et al. 2014 for details on pipeline). Particular attention was paid to genes involved in the core central intermediate metabolism and cell wall formation, which have been identified in previous studies on bacterial metabolism (Zientz et al. 2004).

Phylogenetic Analyses and Validation of Horizontally Transferred Genes

Phylogenetic analyses were employed to validate the putative horizontal and vertical histories of the genes identified in the

Table 1
Basic Information of *Bartonella* Genomes and Corresponding Samples Assessed in This Study

Lineage	Species	Strain	Size (Mb)	Host	Country	Year	PubMed	GenBank Acc. No.
L4	*B. alsatica*	IBS 382	1.67	Rabbit (*Oryctolagus cuniculus*)	France	1998	10028274	AIME01000000
	B. florenciae	R4 (2010)	2.05	Shrew (*Crocidura russula*)	France	2010	—[a]	CALU00000000
	B. birtlesii	LL-WM9	1.92	Mouse (*Apodemus sylvaticus*)	The United States	2002	20395436	AIMC00000000
	B. sp. DB5-6	—	2.15	Shrew (*Sorex araneus*)	Sweden	1999	12613756	AILT00000000
	B. taylorii	8TBB	2.02	Vole (*Microtus agrestis*)	The United Kingdom	2001	17096870	AIMD00000000
	B. vinsonii	Pm136co	1.86	Squirrel (*Spermophilus beecheyi*)	The United States	1999	12574261	AIMH00000000
	B. grahamii	as4aup	2.37	Mouse (*Apodemus sylvaticus*)	Sweden	1999	12613756	NC_012846-47
	B. rattimassiliensis	15908	2.17	Rat (*Rattus norvegicus*)	France	2002	15297537	AILY00000000
	B. queenslandensis	AUST/NH15	2.38	Rat (*Rattus leucopus*)	Australia	1999	19628592	CALX00000000
	B. rattaustraliani	AUST/NH4	2.16	Rat (*Rattus tunneyi*)	Australia	1999	19628592	CALW00000000
	B. elizabethae	F9251	1.98	Human (*Homo sapiens*)	The United States	1986	7681847	AIMF00000000
	B. tribocorum	CIP 105476	2.64	Rat (*Rattus norvegicus*)	France	1997	9828434	NC_010160-61
	B. koehlerae	C29	1.75	Cat (*Felis catus*)	The United States	1999	10074535	—[b]
	B. henselae	Houston-1	1.93	Human (*Homo sapiens*)	The United States	1990	1371515	NC_005956
	B. quintana	Toulouse	1.58	Human (*Homo sapiens*)	France	1993	15210978	NC_005955
	B. senegalensis	OS02	1.97	Soft tick (*Ornithodoros sonrai*)	Senegal	2008	23991259	CALV00000000
	B. washoensis	Sb944nv	1.97	Squirrel (*Spermophilus beecheyi*)	The United States	2002	12574261	AILU00000000
	B. doshiae	NCTC 12862	1.81	Vole (*Microtus agrestis*)	The United Kingdom	1993	7857789	AILV00000000
L3	*B. rochalimae*	ATCC BAA-1498	1.54	Human (*Homo sapiens*)	The United States[c]	2007	17554119	FN645455-67
	B. sp. 1-1C	—	1.57	Rat (*Rattus norvegicus*)	Taiwan	2006	19018019	FN645486-505
	B. sp. AR 15-3	—	1.59	Squirrel (*Tamiasciurus hudsonicus*)	Japan[d]	2009	19331727	FN645468-85
	B. clarridgeiae	73	1.52	Cat (*Felis catus*)	France	1995	9163438	NC_014932
L2	*B. bovis*	91-4	1.62	Cattle (*Bos taurus*)	France	1998	11931146	AGWA00000000
	B. melophagi	K-2C	1.57	Sheep ked (*Melophagus ovinus*)	The United States	2003	—[e]	AIMA00000000
	B. schoenbuchensis	R1	1.67	Deer (*Capreolus capreolus*)	Germany	1999	11491358	FN645506-24
L1	*B. bacilliformis*	KC583	1.45	Human (*Homo sapiens*)	Peru	1963	1715879	NC_008783
	B. australis	Aust/NH1	1.6	Kangaroo (*Macropus giganteus*)	Australia	1999	18258063	NC_020300
	B. tamiae	Th239	2.26	Human (*Homo sapiens*)	Thailand	2004	18077632	AIMB00000000

Note.—Data were collected from NCBI or original publications, or calculated in Geneious.
[a]DOI:10.4056/sigs.4358060. Not available in PubMed.
[b]Downloaded from the Broad Institute website. Not available in GenBank.
[c]After traveling to Peru.
[d]Imported from United States.
[e]Feldgarden M, et al. Unpublished data.

initial discovery screen. Phylogenetic patterns nesting a *Bartonella* gene within a homologous gene clade of a candidate donor group, or as strongly supported sister group of a candidate donor group, were considered significant evidence supporting the horizontal transfer from this particular donor to *Bartonella* (Koonin et al. 2001; Nelson-Sathi et al. 2012; Husnik et al. 2013; Schonknecht et al. 2013). Nucleotide sequences of metabolic genes of interest (i.e., phospholipid pathway) were extracted from *Bartonella* genomes as well as genomes of selected organisms that represent the putative donor group and its sister groups. Sequences were aligned in MAFFT version 7 (Katoh and Standley 2013), using the L-INS-i algorithm. The MAFFT program was called from the "Translational Align" panel of Geneious 6.1 (Biomatters 2013). Alignment edges were trimmed manually, if needed. The phylogenies of single-gene families were reconstructed based on nucleotide and amino acid sequence alignments (to check for congruence) in a Bayesian Markov chain Monte Carlo (MCMC) statistical framework using MrBayes 3.2 (Ronquist et al. 2012), as well as a maximum likelihood (ML) method implemented in RAxML 7.7 (Stamatakis 2006). The Bayesian MCMC runs had a chain length of 20 million generations, with the sample frequency set as 1,000. The optimal nucleotide substitution models for all three codon positions were computed in PartitionFinder 1.1 (Lanfear et al. 2012). Three independent runs were performed for each data set to ensure consistency among runs. Trace files were analyzed in Tracer 1.5 (Drummond and Rambaut 2007) to check for convergence in order to determine a proper burn-in value for each analysis. A consensus tree was built from the retained tree-space, and posterior probabilities are reported per clade. The ML was run implementing the GTR + G model (for all codon positions) and a bootstrap analysis was performed to gauge clade support.

Survey of Genomic Environments

In order to determine the frequency, components, and boundaries of the putatively horizontally transferred genetic material, genomic environments were manually examined in

Geneious 6.1 (Biomatters 2013). Our assumptions are that multiple independent transfers of a gene would likely result in different gene environments being affected. Likewise, if different *Bartonella* species share the same gene environment adjacent to horizontally transferred genetic material, and the transferred genes follow the previously detected vertical evolutionary pattern for bartonellae, presumably a single ancestral HGT event can be inferred for all species in that lineage. Putatively HGT-derived genes and their adjacent genomic elements were identified in recipient and donor genomes and compared across species and within lineages. Results from this analysis were mapped onto the *Bartonella* species tree (see below).

Molecular Evolution Analyses

Selection analyses were carried out to gage selective pressures operating on all genes in the phospholipid pathway. Selection was assessed using the ML method in the Codeml program of the Phylogenetic Analysis by Maximum Likelihood (PAML) 4.7 package (Yang 2007). As the first step, an analysis under the one-ratio model (M0) was performed to estimate a global ω value (d*N*/d*S* ratio) across the phylogenetic tree. Global selective pressures were assessed using the site models (M1a, M2a, M8, and M8a). Evolutionary rates of particular branches of interest (ω_1) versus the background ratio (ω_0) were computed using the branch model (model = 2). Selective pressures operating on subsets of sites of these branches were calculated using the branch-site models (model A and A1). The significance of change of ω value and evidence of positive selection was assessed using the likelihood ratio test. Positive sites were identified using the Bayes Empirical Bayes (BEB) analysis (Yang et al. 2005).

The tertiary structure of the GpsA (NAD(P)H-dependent glycerol-3-phosphate dehydrogenase) protein in *Coxiella burnetii* (Gammaproteobacteria: Legionellales) was used to model the position of the identified sites with positive selection in horizontally transferred *gpsA* genes (Seshadri et al. 2003; Minasov et al. 2009).

The possibility that the horizontally acquired *gpsA* genes underwent convergent evolution in *Bartonella,* relative to their ancestors was explored. Potential ancestral states of the *gpsA* genes before HGT were reconstructed under model A in Codeml of PAML (see above). The sequence was then compared with the consensus sequence of extant *Bartonella* species. A statistical approach recently introduced by Parker et al. (2013) was applied to identify the signatures of convergent evolution of *gpsA* versions after horizontal acquisition. In brief, this method is based on the significance of differences in site-wise log-likelihood supports among a commonly accepted species tree and given alternative convergent topologies under the same substitution model.

Phylogenetic Analysis of *Bartonella* Species

In order to parsimoniously map HGT events to the evolutionary history of bartonellae, phylogenetic relationships among *Bartonella* species were inferred using standard phylogenetic and phylogenomic approaches as follows:

1. *Phylogenomic analysis*: The proteomes of 23 annotated *Bartonella* genomes were downloaded, from which orthologous groups (OGs) were identified using OrthoMCL 2.0 (Li et al. 2003) with default parameters (BLASTP *E* value cutoff = 1×10^{-5}, percent match cutoff = 80%, MCL inflation parameter = 1.5). OGs that have exactly one member in each and every genome were isolated, resulting in 516 OGs. Members of each of these OGs were aligned in MAFFT (Katoh and Standley 2013) and refined in Gblocks 0.91b (Castresana 2000) to remove problematic regions. An optimal amino acid substitution model for each OG was computed in ProtTest 3.3 (Darriba et al. 2011) using the Bayesian information criterion. The 516 alignments were concatenated into one data set, based on which a phylogenetic tree was reconstructed using the ML method as implemented in RAxML (Stamatakis 2006) with 100 fast bootstrap replicates. *Bartonella tamiae* was used as an outgroup, with all other bartonellae treated as ingroup.
2. *Standard phylogenetic analysis*: Five additional species could not be included in above approach, as their proteomes are not available from GenBank. To explore and confirm the phylogenetic positions of these *Bartonella* species, a separate analysis following previously outlined approaches (see above) was performed using six commonly used gene markers (Inoue et al. 2010; Sato et al. 2012; Mullins et al. 2013) from 28 *Bartonella* genomes (*B. tamiae* included in ingroup) and seven outgroup genomes, which represent close sister genera of *Bartonella* (supplementary table S1, Supplementary Material online) (Gupta and Mok 2007; Guy et al. 2013).

Experimental Genotyping

In order to further explore the distribution of *Bartonella* clades with HGT-derived metabolic genes in blood-feeding insects, we screened a global sampling of 21 species of Siphonaptera and Hippoboscoidea for *gpsA* sequences (table 2). All of these samples had been positive for *Bartonella gltA* gene and 16S rRNA detection by polymerase chain reaction (PCR) in previous analyses (Morse et al. 2012, 2013). Genomic DNA was extracted from each individual specimen, using the DNeasy Blood & Tissue Kit (Qiagen Sciences Inc., Germantown, MD), following the animal tissue protocol. The quality and concentrations of DNA were assessed with a NanoDrop spectrometer (Thermo Fisher Scientific, Wilmington, DE). Bacterial *gpsA* diversity was assessed by amplification of *gpsA* genes from each sample using specific primers and reaction conditions: *Helicobacter*-derived *gpsA* (He) (see Results) forward: 5′-ATG AAA ATA ACA RTT TTT GGW GGY GG-3′, reverse: 5′-TTA

Table 2
PCR-Verified *gpsA* Types in *Bartonella*-Positive Insect Samples

Family	Species	Mammalian Host Species	Location	Detected *gpsA*	
				Origin	GenBank Acc. No.
Keds					
Hippoboscidae	*Lipoptena cervi*	*Odocoileus* sp.(deer)	The United States	He	KJ606299
Bat flies					
Streblidae	*Trichobius frequens*	Unknown (glue trap)	Puerto Rico	None	—
Streblidae	*Paradyschiria lineata*	*Noctilio leporinus* (bulldog bat)	Panama	Ar	KJ606300
Streblidae	*Trichobius corynorhinus*	*Corynorhinus townsendii* (vesper bat)	The United States	Ar	KJ606301
Streblidae	*Trichobius adamsi*	*Macrotus waterhousii* (leaf-nosed bat)	Dominican Republic	Ar	KJ606302
Nycteribiidae	*Leptocyclopodia* sp. nov.	*Harpionycteris whiteheadi* (fruit bat)	Philippines	Ar	KJ606303,KJ606321
Nycteribiidae	*Eucampsipoda africana*	*Rousettus aegyptiacus* (Egyptian fruit bat)	Kenya	Ar	KJ606304
Nycteribiidae	*Phthiridium* sp., scissa group	*Rhinolophus pearsoni* (Pearson's Horseshoe bat)	Laos	Ar	KJ606305
Streblidae	Unidentified bat fly	Unknown	French Guiana	Ar	KJ606320
Fleas					
Ischnopsyllidae	*Ischnopsyllus variabilis*	*Myotis daubontoni* (vesper bat)	Switzerland	Ar	KJ606306
Ischnopsyllidae	*Ischnopsyllus indicus*	*Pipistrellus javanicus* (vesper bat)	Philippines	Ar	KJ606307
Ischnopsyllidae	*Dampfia grahami grahami*	*Eptesicus matroka* (vesper bat)	Madagascar	Ar	KJ606308
Ceratophyllidae	*Kohlsia* sp.	—	Mexico	Ar	KJ606309
Ceratophyllidae	*Eumolpianus* sp.	—	—	Ar	KJ606310
Ceratophyllidae	*Megabothris walkeri*	*Microtus agrestis* (vole)	Finland	Ar	KJ606311
Ceratophyllidae	*Nosopsyllus laeviceps ellobii*	*Meriones unguiculatus* (gerbil)	Mongolia	Ar	KJ606312
Ceratophyllidae	unidentified ceratophyllid flea	—	The United States	Ar	KJ606313
Leptopsyllidae	*Pectinoctenus lauta*	*Cricetulus migratorius* (hamster)	Xinjiang, China	Ar	KJ606314
Leptopsyllidae	*Leptopsylla nana*	*Cricetulus migratorius* (hamster)	Xinjiang, China	Ar	KJ606315
Leptopsyllidae	*Ophthalmopsylla kiritschenkovi*	*Phodopus roborovski* (hamster)	Mongolia	Ar	KJ606316
Leptopsyllidae	*Mesopsylla hebes clara*	*Allactaga bullata* (jerboa)	Mongolia	Ar	KJ606317
Rhopalopsyllidae	*Ectinorus onychius onychius*	*Loxodontomys micropus* (mouse)	Argentina	Ar	KJ606318
Rhopalopsyllidae	*Ectinorus lareschiae*	*Phyllotis xanthopygus* (mouse)	Argentina	None	—
Rhopalopsyllidae	*Polygenis* sp.	—	—	None	—
Stephanocircidae	*Craneopsylla minerva wolffheuglia*	*Ctenomys* sp. (tuco-tuco)	Argentina	None	—
Pulicidae	*Xenopsylla conformis conformis*	*Meriones meridianus* (gerbil)	Mongolia	Ar	KJ606319

Note.—"None" indicates negative for *gpsA*, whereas positive for *Bartonella* (=lineage 3 bartonellae).

ATA CCT TCW GCY ACT TCG CC-3′; Enterobacteriales-derived *gpsA* (Ar/Se) forward: 5′-GGT TCT TAT GGY ACY GCW TTA GC-3′, reverse: 5′-TAR ATT TGY TCG GYA ATT GGC ATT TC-3′. Subsequent TA cloning (if applicable) was performed to isolate amplicons. Based on previous studies of the microbial diversity of bat flies, we expect a subset of species to harbor *Arsenophonus* and like organisms (ALOs) as endosymbionts (Morse et al. 2013; Duron et al. 2014). In these species, we specifically targeted *Arsenophonus*-type *gpsA* for comparative purposes. Sequence analysis and phylogenetic analysis followed the standard protocols described above.

Results

The initial discovery screen revealed several candidates for possible horizontal transfer in a variety of metabolic pathways (e.g., peptidoglycan biosynthesis; glutamate/aspartate transport). However, the phospholipid pathway stood out, in that several fundamental genes involved in this pathway show patterns of repeated homologous replacements from identifiable sources outside Alphaproteobacteria, and/or gene loss (fig. 1). These genes are: 1) The *gpsA* gene, which is a chromosomal minimal core gene encoding NAD(P)H-dependent glycerol-3-phosphate (G3P) dehydrogenase, an enzyme that is essential to the synthesis of bacterial membrane lipids; 2) the *glpK* gene (glycerol kinase), which encodes an enzyme that is located in the cell membrane and catalyzes the Mg^{2+}-ATP-dependent phosphorylation of glycerol to G3P; and 3) the Glp system (encoded by genes *glpS-T-P-Q-U-V*), an ATP-binding cassette transporter (ABC transporter) that is responsible for importing extracellular glycerol (Ding et al. 2012).

Loss of *glpK* and Glp System Precedes Loss of *gpsA*

Results of BLAST-based and phylogenetic approaches reveal a pattern of additional gene losses in a core

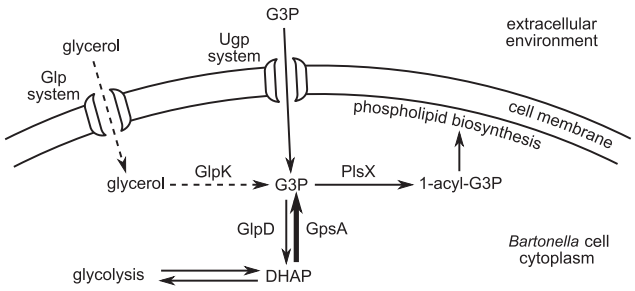

Fig. 1.—Role of GpsA in *Bartonella* phospholipid biosynthesis. Part of the alphaproteobacterial phospholipid biosynthesis pathway is illustrated based on Cronan (2003), Pereto et al. (2004), Spoering et al. (2006), Yeh et al. (2008), and the KEGG pathway entry bhe00564 (glycerophospholipid metabolism in *B. henselae*). The illustration highlights the three possible paths of obtaining G3P. The dashed lines represent the reactions affected by ancient gene losses; and the bold line represents the reaction affected by one ancient gene loss followed by three independent horizontal regains in the evolutionary history of *Bartonella*.

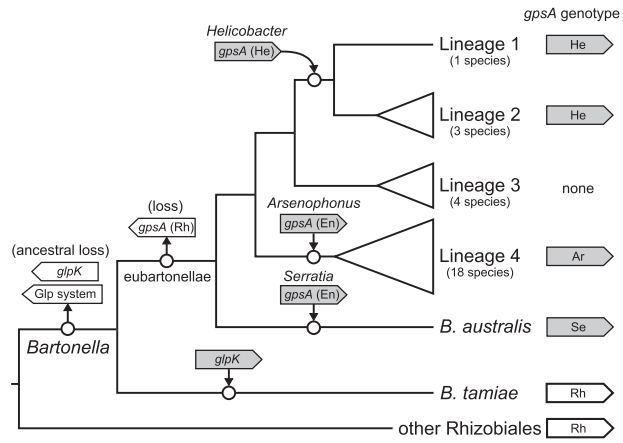

Fig. 2.—Losses and gains of *gpsA* and metabolically related genes in the evolutionary history of *Bartonella*. Schematically illustrated relationships of *Bartonella* lineages based on phylogenomic and phylogenetic analyses of the 28 *Bartonella* species (supplementary fig. S1, Supplementary Material online). Topology is congruent with a recent phylogenomic study (Guy et al. 2013). Major monophyletic lineages (table 1) were collapsed into triangles. Branch lengths are not drawn to scale. The presence and origin of *gpsA* is indicated to the right of corresponding lineages. Horizontally acquired genes are indicated by gray boxes, whereas vertically inherited genes are indicated by white boxes. HGT events are represented by incoming arrows, with the putative donor groups (if identifiable) labeled. Gene loss events are represented by outgoing arrows and boxes with dashed outlines. Phylogenetic positions of losses and gains are indicated by circles.

alphaproteobacterial metabolic pathway. Specifically, key genes involved in the glycerol pathway are ancestrally lost in the bartonellae. Only the *B. tamiae* genome, the most basal *Bartonella* species, contains *glpK*. However, phylogenetic analysis places this copy closely related to Enterobacteriaceae, which is suggestive of a horizontal origin. This, together with the complete absence of *glpK* and the Glp system in extant eubartonellae, supports a loss of the glycerol pathway at the base of all currently known bartonellae, preceding the *gpsA* loss (fig. 2). The absence of GlpK and the Glp system precludes the ability of eubartonellae to utilize extracellular glycerol as a source of G3P (fig. 1). No other functional homologs of these genes are known, or have been found in our genomic analysis.

Multiple Origins of *Bartonella gpsA* Genes

BLAST-based and phylogenetic analyses reveal four origins of *gpsA* genes among *Bartonella* genomes (fig. 2 and supplementary figs. S1 and S2, table S2, Supplementary Material online). Known functional equivalents (not homologs) of bacterial GpsA (G3PDH), such as archaeal EgsA (G1PDH), are not present in any of the genomes (Koga et al. 1998). No *Bartonella* species has more than one copy of *gpsA*. Only the earliest diverging *B. tamiae* (Guy et al. 2013) contains a *gpsA* gene close to those of other Rhizobiales (fig. 3*A*). Specifically, *B. tamiae* is placed as a sister group to Brucellaceae (*Brucella* and *Ochrobactrum*), within the Rhizobiaceae (*Rhizobium*, *Agrobacterium*, and *Sinorhizobium*). This topology mirrors our current knowledge of Rhizobiales and *Bartonella* evolution (Gupta and Mok 2007; Munoz et al. 2011).

Bartonella bacilliformis (lineage 1), and all members of lineage 2 (*B. bovis*, *B. melophagi*, and *B. schoenbuchensis*) possess a *gpsA* that nests strongly supported within a genus of Epsilonproteobacteria, namely *Helicobacter* (fig. 3*B*). In the phylogenetic analysis all *Helicobacter*-derived *gpsA* (He) genes form a strict monophyletic group, with lineage 1 and lineage 2 split at the base. Its immediate sister group is a clade of four *Helicobacter* species (fig. 3*B*). The general structure of the Helicobacteraceae clade resembles the species tree of this family from previous studies (Dewhirst et al. 2005; Gupta 2006; Munoz et al. 2011).

Lineage 3 bartonellae (*B. clarridgeiae*, *B. rochalimae*, *B.* sp. 1-1C, and *B.* sp. AR 15-3) lack any identifiable homolog of *gpsA*.

Bartonella australis and all members of lineage 4 (table 1) possess *gpsA* genes, which were captured from the gammaproteobacterial Enterobacteriales. These *gpsA* genes were transferred in separate instances, as *B. australis gpsA* has high sequence similarity and phylogenetic affiliation with *Serratia* species [*gpsA* (Se)], whereas all available representatives of lineage 4 contain an *gpsA* gene [*gpsA* (Ar)] that is closely related to that of *Arsenophonus* type bacteria (ALOs) (fig. 3*C* and *D*). Specifically, analyses reveal that all lineage 4 *Bartonella gpsA* genes form a monophyletic group that is sister to extant ALOs. Together, they are nested within a clade including ALOs' closest sister groups: *Providencia*,

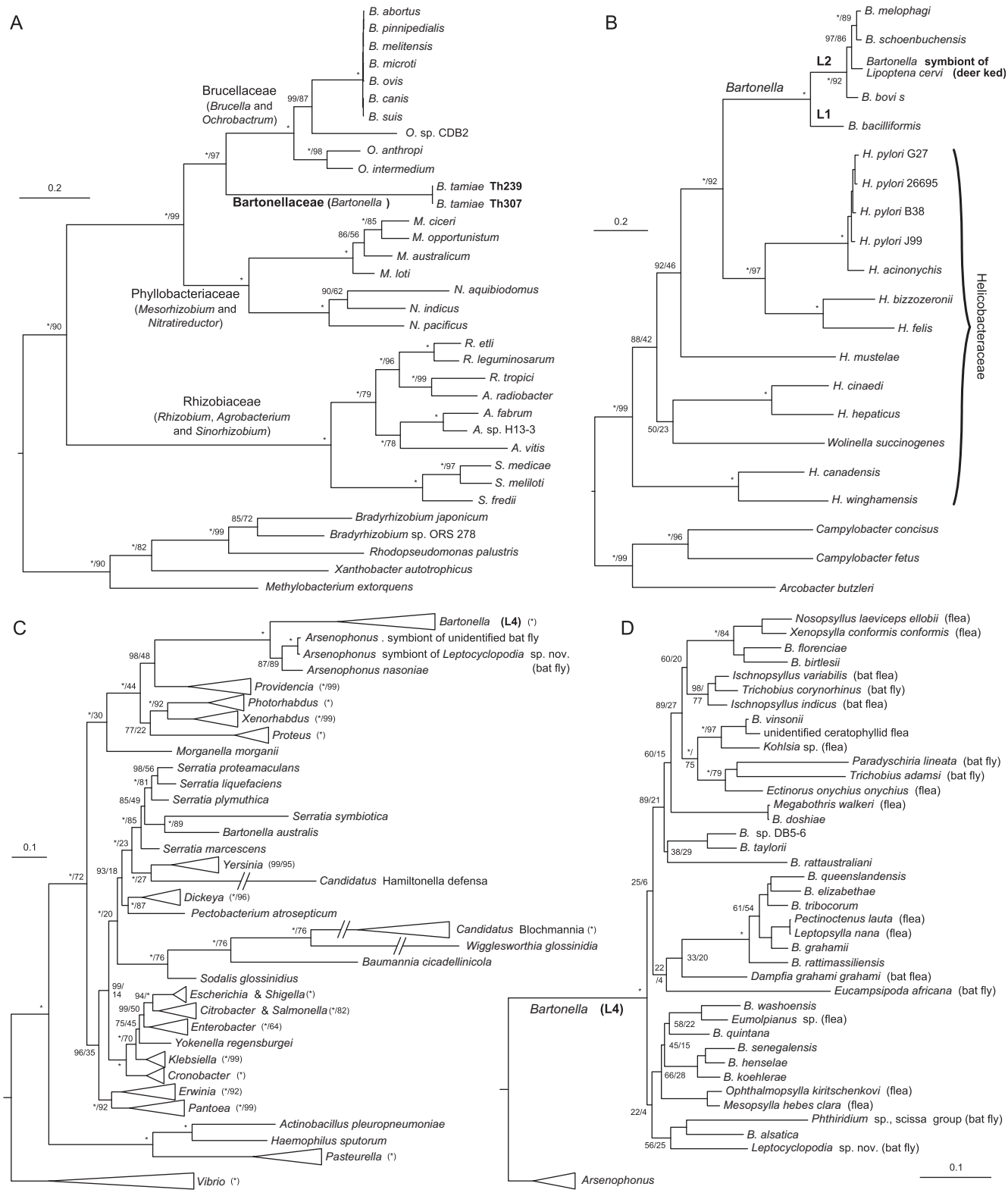

Fig. 3.—Phylogenies of different versions of *gpsA*. Trees were reconstructed using Bayesian inference as implemented in MrBayes. Node labels (*x*/*y*) represent Bayesian posterior probabilities (*x*%) computed in MrBayes and ML bootstrap support values (*y*% out of 1,000 replicates) computed in RAxML. Asterisks (*) indicate 100% support. *Bartonella* clades are denoted in bold font. (*A*) Vertical inheritance history of *gpsA* (Rh) in Rhizobiales. Families Bartonellaceae, Brucellaceae, Phyllobacteriaceae, and Rhizobiaceae are placed as ingroups and the other Rhizobiales organisms as outgroups, according

(continued)

Photorhabdus, *Xenorhabdus*, and *Proteus*. On the other hand, *B. australis gpsA* nests within the *Serratia* clade, with its closest sister group being *Serratia symbiotica*.

Genomic Environments of the *gpsA* Genes Support One Ancestral Loss and Three Individual Transfers at the Base of Major *Bartonella* Lineages

Rhizobiales-Derived gpsA (Rh)

The *gpsA* (Rh) gene (supplementary fig. S3*A*, Supplementary Material online) is located within a gene cassette that typically contains five tandemly arranged genes in the genome of *B. tamiae* and those of close sister groups of *Bartonella*, including *Brucella*, *Ochrobactrum*, *Mesorhizobium*, *Agrobacterium*, *Rhizobium*, and *Sinorhizobium*. In all other *Bartonella* genomes, this cassette is still present, but consistently rhizobial *gpsA* and its immediate downstream open reading frame (ORF) (Ycil-like protein coding sequences) are absent in all eubartonellae. Instead, this space is occupied by sequences without identifiable ORFs. No sequence similarity can be detected between those sequences and the original contents. The above-described pattern is consistent with a single ancestral loss of *gpsA* followed by three gains (see below; fig. 2), each of which coincides with the current lineage classification of bartonellae (Engel et al. 2011). From all known bartonellae, lineage 3 is the only clade in which all species are not only missing the *glpK* and the *glp* genes, but it also never regained *gpsA* (fig. 2).

Helicobacter-Derived gpsA (He)

The original *gpsA* (He) (fig. 4) is residing in a genomic environment that is highly variable among *Helicobacter* species. In most cases, it is upstream of the *glyQ* (glycyl-tRNA synthetase subunit alpha) gene. In lineage 1 and 2 *Bartonella* genomes, only the *gpsA* gene seems to have been transferred (fig. 2), without its upstream and downstream neighbors from *Helicobacter*. The gene is located in a genomic locus, where the upstream side is a group of four ORFs ending with the *hisS* (histidyl-tRNA synthetase) gene. The downstream side of *gpsA* in lineage 2 *Bartonella* genomes is an rRNA operon, which is typically present in all *Bartonella* genomes as two to three copies (Viezens and Arvand 2008; Guy et al. 2013). The horizontal transfer of *gpsA* (He) into lineages 1 and 2 seems to have interrupted an ORF present in all bartonellae, which in lineage 2 bartonellae is still present with a residual sequence (fig. 4). Phylogenetic analysis of this ORF sampled across all bartonellae mirrors current hypotheses of *Bartonella* species evolution (Guy et al. 2013).

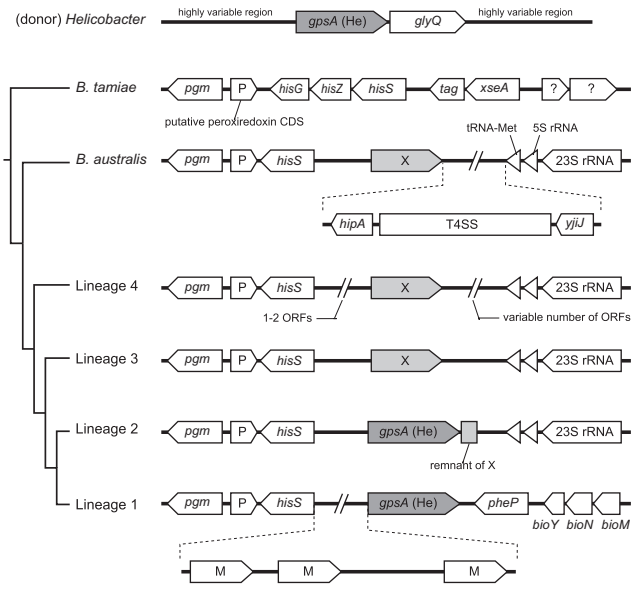

Fig. 4.—Genomic context of *gpsA* (He) in *Bartonella* and other bacterial groups. Genes are represented by boxes. ORFs annotated as hypothetical genes are either indicated by "?," or by single letters (e.g., "M" and "X," see below). Lengths of genes and intergenic regions are not drawn to scale. "X" represents an ORF that is disrupted by the insertion of *gpsA* (He). "M" represents a multicopy ORF that exists only in *B. bacilliformis* and *B. australis* genomes.

Arsenophonus-Derived gpsA (Ar)

In the genomes of *Arsenophonus* and other Enterobacteriaceae, the original *gpsA* (Ar) gene (supplementary fig. S3*B*, Supplementary Material online) resides within a cassette of four genes (*secB–gpsA–cysE–cspR*) right downstream of the O-antigen gene cluster, a frequently horizontally transferred structure (Wildschutte et al. 2010; Ovchinnikova et al. 2013). In lineage 4 *Bartonella* genomes, *gpsA* (Ar) seems to have been transferred singularly into a genomic region that is present in all bartonellae. Upstream of it is a *cyo* operon (*cyoA, B, C, D*) encoding the cytochrome *o* ubiquinol oxidase,

Fig. 3.—Continued
to Gupta and Mok (2007). (*B*) Horizontal transfer of *gpsA* (He) from *Helicobacter* to L1 and L2 *Bartonella* (including an experimentally verified deer ked sample). The tree is rooted at the common ancestor of Helicobacteraceae and Campylobacteraceae, according to Gupta (2006). (*C*) Horizontal transfers of *gpsA* (Ar) from *Arsenophonus*-like bacteria to L4 *Bartonella*, and that of *gpsA* (Se) from *Serratia* to *B. australis*. The tree includes recipient *Bartonella* species, representative Enterobacteriales groups, and two *Arsenophonus*-positive bat fly samples sequenced in this study. It is rooted to Vibrionales according to Williams et al. (2010). Monophyletic groups are collapsed in triangles with nodal support values labeled to the right. Long branches are truncated and indicated by two slashes (//). (*D*) Posttransfer evolutionary history of *Arsenophonus*-derived *gpsA* (Ar) in L4 *Bartonella*. This is an expansion of the L4 *Bartonella* clade in (*C*). Experimentally verified insect samples are indicated by the insect names. Nodal support values of derived clades are omitted.

a component of the aerobic respiratory chain (Reva et al. 2006). Downstream is a cluster of three genes (*fabI1–fabA–fabB*) that are essential in fatty acid biosynthesis (Campbell and Cronan 2001).

Serratia-Derived gpsA (Se)

In *Serratia*, *gpsA* is located in the homologous gene cassette to other Enterobacteriaceae. In the *B. australis* genome, *gpsA* was cotransferred with two other genes of the donor cassette (*grxC–secB–gpsA*) and resides in a region that is highly variable among *Bartonella* species. However, the structures upstream and downstream of this highly variable region are relatively constant in all *Bartonella* genomes (supplementary fig. S3C, Supplementary Material online). Notably, several house-keeping genes involved in lipid metabolism (*plsX*, *fabH*, *accB*, *accC*, and *glpD*) (Campbell and Cronan 2001; Cronan 2003) are located proximal to this region.

Other Genes in the Phospholipid Pathway

Meanwhile, an alternative route of G3P acquisition is intact: The Ugp system (encoded by the operon *ugpB-A-E-C*), an ABC transporter that imports G3P into cells (Brzoska et al. 1994), is present in all extant *Bartonella* species. Phylogenetic analysis reveals that this operon is vertically transmitted in the eubartonella (supplementary fig. S4 and table S2, Supplementary Material online).

G3P's utilization in the phospholipid biosynthesis pathway is mediated by PlsX (G3P acyltransferase), which converts G3P into 1-acyl-G3P for subsequent steps. PlsX-deficient *Escherichia coli* strains cannot synthesize a cell membrane (Bell 1974). All *Bartonella* species maintain one copy of this gene. The phylogenetic tree of *plsX* mirrors the species tree of *Bartonella* (supplementary fig. S5 and table S2, Supplementary Material online).

Molecular Evolution Analyses

All genes tested in this analysis maintain an ORF (regardless of horizontal or vertical origin). Analyses testing for selective pressure along branches and among sites were carried out on trees of individual *gpsA* gene families and other metabolically related genes (supplementary table S3, Supplementary Material online). The following general patterns that apply to all three *gpsA* families were observed: There are clear signatures of global stabilizing selection operating on *gpsA* genes, including on branches leading up to the nodes representing the transfers from putative donor groups (*Helicobacter*, *Arsenophonus* [ALOs], and *Serratia*) to stem-*Bartonella* of the major lineages (L1 + L2, L4, and *B. australis*, respectively). Strong stabilizing selection was also observed in horizontally acquired and subsequently vertically transmitted copies of *Bartonella gpsA*. Within the *Bartonella* clades (L1 + L2, L4, and *B. australis* alone) that represent the evolutionary history of *gpsA* after acquisition, the ω values are significantly elevated compared with the tree backgrounds (supplementary table S3, Supplementary Material online), suggesting an accelerated rate of evolution after horizontal acquisition.

Representative protein sequences of each version of *gpsA* were aligned to the *C. burnetii* sequence, whose tertiary structure has been experimentally verified (fig. 5). The sequence similarity among versions is generally low. However, functional motifs and their adjacent sites exhibit strong conservation across all *gpsA* sequences. None of these sites were predicted to be under positive selection. The majority of sites that were detected to be under significant positive (diversifying) selection (27/31, $P > 95\%$) have been identified in the *gpsA* of lineage 4 *Bartonella* (supplementary table S3, Supplementary Material online, and fig. 5).

Applying the statistical methods outlined by Parker et al. (2013) resulted in no significant signature of convergent evolution in horizontally transferred *gpsA* versions across *Bartonella*.

All other genes in the phospholipid pathway are under strong stabilizing selection across all lineages in the bartonellae (supplementary table S3, Supplementary Material online).

Phylogenetic Analyses of *Bartonella*

Phylogenomic analysis of *Bartonella* core genomes (23 species) and of selected genes (28 species) recovers previously identified clades and relationships. Topologies of ingroup bartonellae (eubartonellae) mirror results from the recent analysis of Guy et al. (2013) (table 1 and supplementary fig. S1, Supplementary Material online), challenging current *Bartonella* classification, and supporting the idea of a derived evolutionary position of *B. bacilliformis* (lineage 1, fig. 1). Although there is some controversy about the relationship of L1 and L2 bartonellae to each other (Engel et al. 2011; Guy et al. 2013), the observed pattern of ORF disruption by the *Helicobacter*-derived *gpsA* (see Results; fig. 4) provides a solid piece of evidence of a single transfer event, and a shared derived ancestry of L1 and L2 (Guy et al. 2013). *Bartonella tamiae* occupies a strongly supported ancestral position to all other bartonellae in the standard phylogenetic analysis. We recovered a basal position for *B. australis* in eubartonellae in both analyses, and the evolutionary sequence of lineage-specific diversification is strongly supported on all nodes.

Experimental Genotyping of *Bartonella gpsA*

GpsA sequences were successfully recovered from most of our samples of blood-feeding insects (table 2). One bat fly sample contained two *gpsA* copies—one copy of the known, and expected *Arsenophonus* [ALO] endosymbiont of this group (Morse et al. 2013), and one copy of *Bartonella gpsA* (Ar). *Helicobacter*-type *gpsA* (He) was detected in a deer ked (*Lipoptena cervi*) sample, and is phylogenetically nested within the L2 *Bartonella* clade (fig. 3*B*). *Arsenophonus*-type

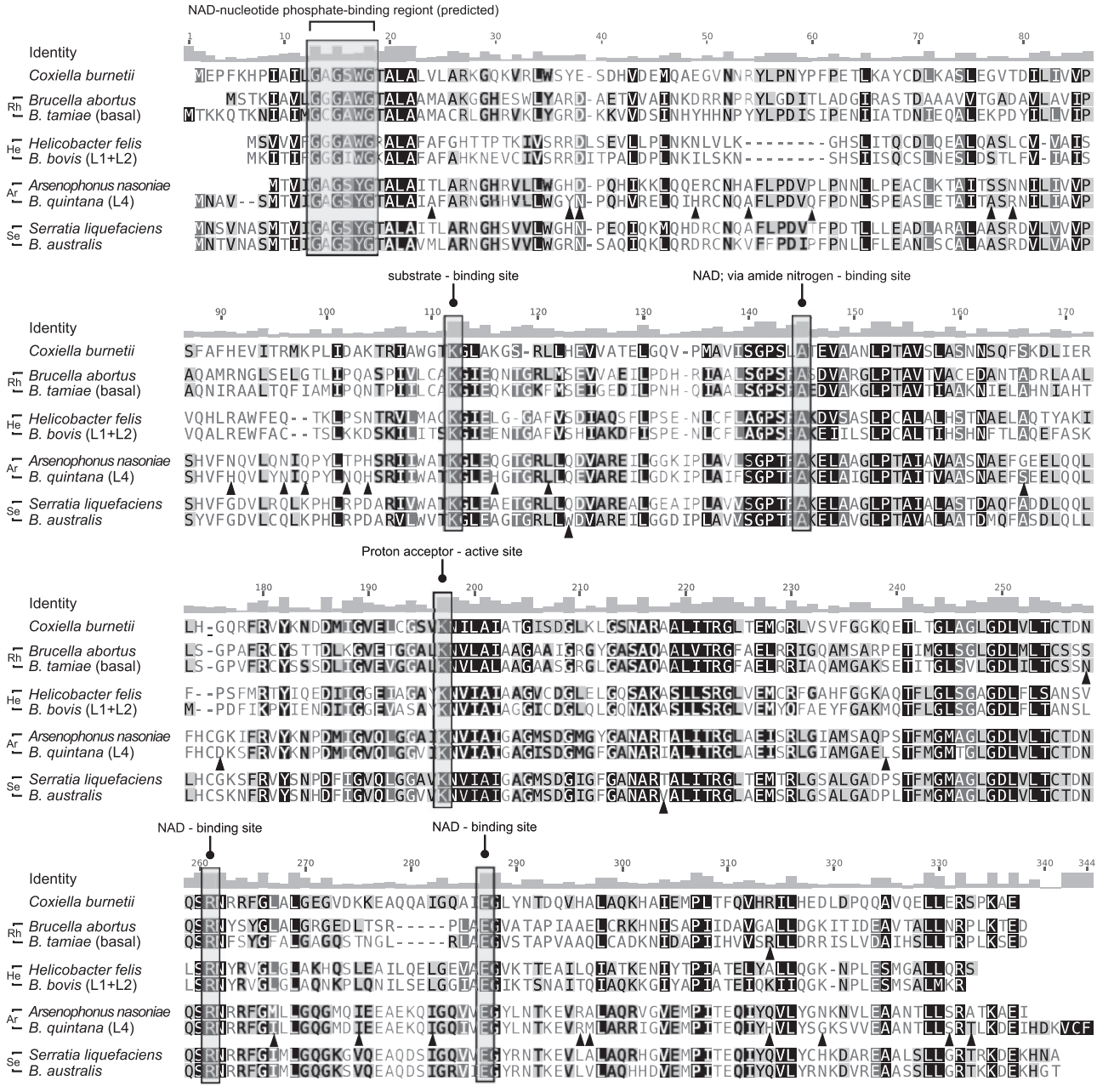

Fig. 5.—Comparison of protein sequences of different GpsA versions. Alignment of full-length GpsA protein sequences to *Coxiella burnettii*. Nucleotide positions are shaded by similarity from low (light) to high (dark) on a grayscale. GpsA proteins are aligned in pairs with a representative sequence from the donor group and its *Bartonella* counterpart. Functional sites and motifs are boxed, as recorded in UniProtKB. Significantly positively selected sites predicted by BEB are indicated by solid triangles.

gpsA (Ar) was found in most fleas (Siphonaptera) and in all bat flies (Nycteribiidae and Streblidae), increasing the number of known bartonellae vector species for both fleas, and bat flies. All of these are distributed within the L4 *Bartonella* clade, which shows species groupings generally consistent with previous analyses (supplementary fig. S1, Supplementary Material online). Flea and bat fly host affiliations scatter throughout previously known subclades. No *Serratia*-type *gpsA* (Se) was detected in any of our samples (*B. australis*). The currently known distribution of *B. australis* is restricted to Australia (Fournier et al. 2007), and does not overlap with our sampling. Several fleas did not yield any *gpsA* gene, despite being positive for *Bartonella gltA* and 16S rRNA pertaining to lineage 3.

Discussion

Ancestral Intracellularity in *Bartonella*

Given the conserved nature of the bacterial phospholipid pathway, the ancestral loss of *glpK* and the Glp system in stem bartonellae followed by a *gpsA* loss at the base of eubartonellae likely created an ancestral population of bartonellae unable to use either glycerol or glucose metabolites (fig. 1). Our analyses therefore suggest that the ancestors of extant eubartonellae relied directly on G3P, which can be imported into the cytoplasm by the Ugp system that remains intact in all bartonellae. G3P is known to be an intermediate metabolite of a strictly intracellular biochemical pathway in prokaryotic and eukaryotic cells, and does not occur stably in the extracellular environment (e.g., blood) (Cronan 2003; Spoering et al. 2006). Therefore, G3P capture and utilization by ancestral bartonellae were likely accomplished by cytoplasmic associations to a living cell. Previous research has shown that in the absence of readily available G3P, the ability of *gpsA* mutant bacteria to form cell membranes is severely compromised, resulting in the cessation of cell growth (Bell 1974; Cronan and Bell 1974). This functional peculiarity may explain the slow extracellular growth of the *gps*A-less lineage 3 bartonellae on blood agar and their more successful isolation in living cell lines (Heller et al. 1997; Podsiadly et al. 2007). Specifically, the four extant representatives of lineage 3 bartonellae (*B. rochalimae, B. clarridgeiae*, *B*. sp. 1-1C, and *B*. sp. AR 15-3) are possibly the surviving representatives of an ancient lineage, as they are still lacking *glpK*, the Glp system, as well as *gpsA*. The ubiquitous loss of these important genes in the phospholipid pathway prior to the evolution of extant eubartonellae certainly suggests that bartonellae had an early intracellular beginning.

Functional Importance of the Acquired *gpsA* Genes

Our results provide strong evidence that bartonellae *gpsA* was acquired from three independent prokaryotic sources outside of alpha-proteobacteria after a single initial loss at the base of the eubartonellae lineage. The repeated retentions of HGT-derived *gpsA* in the *Bartonella* genomes confirm the functional importance of *gpsA*, in the context of the loss-and-regain hypothesis of Doolittle et al. (2003). Based on an array of studies related to HGT, it has been hypothesized that genes that are selectively advantageous in the new organisms have a higher probability of being retained (Kuo and Ochman 2009). Results suggest vertical inheritance and global stabilizing selection after *gpsA* transfer in *Bartonella* lineages, as well as the maintenance of ORFs in all transferred genes. Taken together with the previously confirmed expression of *gpsA* in bartonellae (Saenz et al. 2007; Omasits et al. 2013), the above evidence supports the functionality of the *gpsA* genes after transfer. Furthermore, the protein sequence alignment shows that all functionally important sites are conserved among the HGT-derived *gpsA* versions (fig. 3). This, combined with the overall stabilizing selection on horizontally transferred *gpsA* genes (see Results), implies that the acquired genes are likely to have inherited their original functional role in the biosynthesis of bacterial membrane lipids. However, the significant elevation of evolutionary rates in all three acquired genes and the detection of specific sites under positive selection suggest that amid the overall stabilizing selection, the genes may still undergo functional evolution to adapt to bartonellae-specific metabolic pathways. Moreover, although different *gpsA* genes were inserted into distinct genomic loci, it is notable that their genomic contexts typically include clusters of genes that are involved in bacterial lipid metabolism (see Results; fig. 4). This suggests that they have been integrated into the existing transcriptional regulation machinery of lipid metabolic genes, which may have facilitated their retention (Lercher and Pal 2008). Bartonellae *gpsA* mutants are known to result in an abacteremic phenotype (Saenz et al. 2007), pointing to the importance of functional *gpsA* in pathogenicity and hematogenous spread. Therefore, it is possible that the horizontal reacquisitions of functional *gpsA* facilitated the hematogenous spread of bartonellae to diverse hosts through blood-feeding vectors, as confirmed for the majority of extant bartonellae.

Ancestral Host-Associations of Bartonellae

For two prokaryotic organisms to exchange genetic material likely requires interactions in micro- or nanospace, either directly between bacteria (e.g., conjugation) or with transfer agents (e.g., bacteriophages) in an appropriate environment (Frost et al. 2005; Polz et al. 2013). Based on the high likelihood of their intracellular beginnings (as inferred from the ancestral losses in the phospholipid pathway), the identified cases of HGT provide not only clues to infer past shared ecological connections between *Bartonella* and other bacteria but also point to putative ancestral hosts.

In line with this argumentation, we suggest that the two transfers from Gammaproteobacteria are more likely to have occurred in an arthropod. Specifically, the *Arsenophonus*-derived transfer likely stems from endosymbionts exclusively associated with arthropods. ALOs, the immediate well-supported sister group of the bartonellae *gpsA,* is ecologically versatile and widely distributed among arthropods, including blood-feeders, such as ticks (Ixodidae), keds (Hippoboscidae), and bat flies (Nycteribiidae) (Trowbridge et al. 2006; Novakova et al. 2009; Morse et al. 2013; Duron et al. 2014). Among blood-feeding parasites, *Arsenophonus* and like species are still primary endosymbionts of extant hippoboscoid flies, which are among the confirmed insect vectors of bartonellae of lineages 4 and 2 (Halos et al. 2004; Morse et al. 2013). Therefore, we propose that hippoboscoid flies may have already been among the ancestral blood-feeding arthropod hosts of bartonellae providing a biological reservoir conducive for horizontal transfer of genes. Interestingly, the

gammaproteobacterial ALOs have never been detected in extant Siphonaptera, the insect order with the most common and speciose *Bartonella* vectors (Chomel et al. 2009; Tsai et al. 2011; Pulliainen and Dehio 2012). Yet all lineage 4 bartonellae transmitted by fleas carry the *Arsenophonus*-derived *gpsA* (Ar) (fig. 3*C*). These facts imply that the *Arsenophonus*-derived transfer of *gpsA* (Ar) to stem L4 *Bartonella* likely did not occur in fleas, but that use of Siphonaptera are likely evolutionarily derived vectors of lineage 4 bartonellae.

The *gpsA* transfer from the enterobacterial *Serratia* involves *B. australis*, which at present seems to be the only representative of its lineage, although this may change with better sampling coverage. *Serratia* species colonize diverse habitats, including plants, insects, and vertebrates (Grimont F and Grimont PAD 2006). Given this wide, and largely underexplored host range, it is difficult to pinpoint a specific ancestral host for the HGT exchange of the *Serratia*-derived *gpsA* in *Bartonella*. However, the *Serratia*-derived *gpsA* of *B. australis* strongly nests within a monophyletic clade containing *Serratia symbiotica*, a known endosymbiont of aphids (Aphidoidea) (fig. 3*C*). In insects *Serratia* may function as pathogen or symbiont, and have been shown to invade the hemocoel and the intestinal tract (Grimont F and Grimont PAD 2006). In mammals, infection often is opportunistic and rarely systemic (unless previous immunosuppression exists) (Grimont F and Grimont PAD 2006; Mahlen 2011). Therefore, it is conceivable that an insect host was involved in this transfer too.

In contrast, we suggest that the horizontal integration of *Helicobacter*-derived *gpsA* into a *Bartonella* genome likely took place in a mammalian host, especially given that bartonellae *gpsA* (He) is firmly nested within the *Helicobacter* clade. *Helicobacter* are predominantly mammalian pathogens (Whary and Fox 2004; Rogers 2012), whose hosts overlap well with the known host range of *Bartonella* species. In their hosts helicobacter-type bacteria typically colonize the gastrointestinal tract and liver, causing peptic ulcers, chronic gastritis, cancer, and other diseases. Meanwhile, blood is a secondary site for some species, where they adhere to erythrocytes, which is also the dominant infection site of bartonellae (Whary and Fox 2004; Dubois and Boren 2007).

It is important to note that the timing of each inferred horizontal transfer coincides with the subsequent speciation of extant bartonellae lineages, yet it is difficult to assess exactly when these events happened on an evolutionary scale. However, the transferred *gpsA* genes exclusively affect eubartonellae (=every lineage after *B. tamiae*), and the transferred genes are still closely related to extant prokaryotic groups, allowing donor identification. This could be in part due to strong functional constraints, but may also point to an evolutionarily recent HGT relative to the total lineage age. This would support a picture of a more recent diversification with invertebrates and mammals, as suggested by the hypothesis of "explosive radiation" of bartonellae by other authors (Engel et al. 2011; Guy et al. 2013).

Extant Host Associations

Experimental *Bartonella* genotyping and subsequent phylogenetic analyses recover expected lineages given currently known host and vector ranges (Halos et al. 2004; Morse et al. 2012) (table 2), and confirm predictions of *gpsA* origin based on phylogenomic analyses. Specifically, bartonellae of deer keds (*L. cervi*, Hippoboscidae), which are known vectors of lineage 2 bartonellae, contain *Helicobacter*-derived *gpsA* (He) (fig. 3*B*). A diverse sampling of flea and bat fly species shows *Arsenophonus*-derived *gpsA* (Ar), as is expected for vectors of L4 *Bartonella* species (fig. 3*D*). Some samples (e.g., *Megabothris walkeri*) closely cluster with known L4 *Bartonella* species (e.g., *B. doshiae*). Others, such as bartonellae from *Trichobius* species, appear to be phylogenetically distant from any established *Bartonella* subclades, suggesting putative novel species in bat flies, which has been proposed previously (Billeter et al. 2012; Morse et al. 2012). These findings call for more in-depth studies to characterize extant *Bartonella* diversity.

The overall topology of the L4 *gpsA* (Ar) tree (fig. 3*D*) does not mirror the phylogeny of either fleas (Whiting et al. 2008) or bat flies (Dittmar et al. 2006; Petersen et al. 2007), suggesting the absence of *Bartonella*-insect coevolution within these two groups. For bat flies, this is in contrast to their primary endosymbionts, which exhibit notable coevolutionary patterns with their fly hosts (Hosokawa et al. 2012; Morse et al. 2013). Flea and bat fly bartonellae are interspersed among each other in the tree, implying frequent horizontal transmission of L4 bartonellae between insect vectors, and low host specificity.

More detailed coevolutionary analyses of mammal–*Bartonella* and insect–*Bartonella* relationships are warranted given these findings.

Conclusions

Our study shows that the phospholipid pathway in *Bartonella* has been affected by gene losses and gains throughout their evolution. Specifically, *glpK*, the Glp system, and *gpsA* were lost, but only *gpsA* genes were reacquired in eubartonellae by three independent horizontal transfers from Gamma- and Epsilonproteobacteria. Results from this discovery-based study indicate a key impact of HGT on the ecological and functional evolution in bartonellae.

Supplementary Material

Supplementary figures S1–S5 and tables S1–S3 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the National Science Foundation, DEB 1050793, to K.D. The authors declare that they have no conflict of interest.

Literature Cited

Barlow M. 2009. What antimicrobial resistance has taught us about horizontal gene transfer. Methods Mol Biol. 532:397–411.

Bell RM. 1974. Mutants of *Escherichia coli* defective in membrane phospholipid synthesis: macromolecular synthesis in an sn-glycerol 3-phosphate acyltransferase Km mutant. J Bacteriol. 117:1065–1076.

Berglund EC, et al. 2009. Run-off replication of host-adaptability genes is associated with gene transfer agents in the genome of mouse-infecting *Bartonella grahamii*. PLoS Genet. 5:e1000546.

Billeter SA, et al. 2012. Bartonella species in bat flies (Diptera: Nycteribiidae) from western Africa. Parasitology 139:324–329.

Biomatters. 2013. Geneious version 6.1. Available from: http://www.geneious.com/.

Brzoska P, Rimmele M, Brzostek K, Boos W. 1994. The pho regulon-dependent Ugp uptake system for glycerol-3-phosphate in *Escherichia coli* is trans inhibited by Pi. J Bacteriol. 176:15–20.

Campbell JW, Cronan JE Jr. 2001. Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. Annu Rev Microbiol. 55:305–332.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17: 540–552.

Chomel BB, et al. 2009. Ecological fitness and strategies of adaptation of *Bartonella* species to their hosts and vectors. Vet Res. 40:29.

Cronan JE. 2003. Bacterial membrane lipids: where do we stand? Annu Rev Microbiol. 57:203–224.

Cronan JE Jr, Bell RM. 1974. Mutants of *Escherichia coli* defective in membrane phospholipid synthesis: mapping of the structural gene for L-glycerol 3-phosphate dehydrogenase. J Bacteriol. 118:598–605.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165.

Dewhirst FE, et al. 2005. Discordant 16S and 23S rRNA gene phylogenies for the genus *Helicobacter*: implications for phylogenetic inference and systematics. J Bacteriol. 187:6106–6118.

Ding H, Yip CB, Geddes BA, Oresnik IJ, Hynes MF. 2012. Glycerol utilization by *Rhizobium leguminosarum* requires an ABC transporter and affects competition for nodulation. Microbiology 158:1369–1378.

Dittmar K, Porter ML, Murray S, Whiting MF. 2006. Molecular phylogenetic analysis of nycteribiid and streblid bat flies (Diptera: Brachycera, Calyptratae): implications for host associations and phylogeographic origins. Mol Phylogenet Evol. 38:155–170.

Doolittle WF. 1999. Phylogenetic classification and the universal tree. Science 284:2124–2129.

Doolittle WF, et al. 2003. How big is the iceberg of which organellar genes in nuclear genomes are but the tip? Philos Trans R Soc Lond B Biol Sci. 358:39–57; discussion 57–58.

Drummond AJ, Rambaut A. 2007. Tracer, Available from: http://tree.bio.ed.ac.uk/software/tracer/.

Dubois A, Boren T. 2007. *Helicobacter pylori* is invasive and it may be a facultative intracellular organism. Cell Microbiol. 9:1108–1116.

Duron O, et al. 2014. Origin, acquisition and diversification of heritable bacterial endosymbionts in louse flies and bat flies. Mol Ecol. 23: 2105–2117.

Engel P, et al. 2011. Parallel evolution of a type IV secretion system in radiating lineages of the host-restricted bacterial pathogen Bartonella. PLoS Genet. 7:e1001296.

Fournier PE, et al. 2007. *Bartonella australis* sp. nov. from kangaroos, Australia. Emerg Infect Dis. 13:1961–1962.

Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol. 3: 722–732.

Gophna U, Charlebois RL, Doolittle WF. 2004. Have archaeal genes contributed to bacterial virulence? Trends Microbiol. 12:213–219.

Grimont F, Grimont PAD. 2006. The genus *Serratia*. In: Dworkin M, editor.The prokaryotes. Vol. 6: Proteobacteria: gamma subclass. New York: Springer. p. 219–244.

Gupta RS. 2006. Molecular signatures (unique proteins and conserved indels) that are specific for the epsilon proteobacteria (Campylobacterales). BMC Genomics 7:167.

Gupta RS, Mok A. 2007. Phylogenomics and signature proteins for the alpha proteobacteria and its main groups. BMC Microbiol. 7:106.

Guy L, et al. 2013. A gene transfer agent and a dynamic repertoire of secretion systems hold the keys to the explosive radiation of the emerging pathogen *Bartonella*. PLoS Genet. 9:e1003393.

Halos L, et al. 2004. Role of Hippoboscidae flies as potential vectors of *Bartonella* spp. infecting wild and domestic ruminants. Appl Environ Microbiol. 70:6302–6305.

Harms A, Dehio C. 2012. Intruders below the radar: molecular pathogenesis of *Bartonella* spp. Clin Microbiol Rev. 25:42–78.

Heller R, et al. 1997. Prevalence of *Bartonella henselae* and *Bartonella clarridgeiae* in stray cats. J Clin Microbiol. 35:1327–1331.

Hosokawa T, et al. 2012. Reductive genome evolution, host-symbiont cospeciation and uterine transmission of endosymbiotic bacteria in bat flies. ISME J. 6:577–587.

Husnik F, et al. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell 153:1567–1578.

Inoue K, et al. 2010. *Bartonella japonica* sp. nov. and *Bartonella silvatica* sp. nov., isolated from *Apodemus* mice. Int J Syst Evol Microbiol. 60: 759–763.

Kaiser PO, Riess T, O'Rourke F, Linke D, Kempf VA. 2011. *Bartonella* spp.: throwing light on uncommon human infections. Int J Med Microbiol. 301:7–15.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780.

Koga Y, Kyuragi T, Nishihara M, Sone N. 1998. Did archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomeric glycerophosphate backbones caused the separation of the two lines of descent. J Mol Evol. 46:54–63.

Koonin EV, Makarova KS, Aravind L. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol. 55: 709–742.

Kosoy M, et al. 2008. *Bartonella tamiae* sp. nov., a newly recognized pathogen isolated from three human patients from Thailand. J Clin Microbiol. 46:772–775.

Kuo CH, Ochman H. 2009. The fate of new bacterial genes. FEMS Microbiol Rev. 33:38–43.

Lanfear R, Calcott B, Ho SY, Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol. 29:1695–1701.

Lercher MJ, Pal C. 2008. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. Mol Biol Evol. 25:559–567.

Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13: 2178–2189.

Mahlen SD. 2011. Serratia infections: from military experiments to current practice. Clin Microbiol Rev. 24:755–791.

Minasov G, et al. 2009. 2.1 Angstrom resolution crystal structure of glycerol-3-phosphate dehydrogenase (gpsA) from *Coxiella burnetii*.

Available from: http://www.ncbi.nlm.nih.gov/Structure/mmdb/mmdbsrv.cgi?uid=3k96.
Morick D, et al. 2013. Vertical nontransovarial transmission of *Bartonella* in fleas. Mol Ecol. 22:4747–4752.
Morse SF, et al. 2012. Global distribution and genetic diversity of *Bartonella* in bat flies (Hippoboscoidea, Streblidae, Nycteribiidae). Infect Genet Evol. 12:1717–1723.
Morse SF, et al. 2013. Evolution, multiple acquisition, and localization of endosymbionts in bat flies (Diptera: Hippoboscoidea: Streblidae and Nycteribiidae). Appl Environ Microbiol. 79:2952–2961.
Mullins KE, et al. 2013. Molecular typing of Candidatus *Bartonella ancashi*, a new human pathogen causing verruga peruana. J Clin Microbiol. 51: 3865–3868.
Munoz R, et al. 2011. Release LTPs104 of the all-species living tree. Syst Appl Microbiol. 34:169–170.
Nelson-Sathi S, et al. 2012. Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of haloarchaea. Proc Natl Acad Sci U S A. 109:20537–20542.
Novakova E, Hypsa V, Moran NA. 2009. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. BMC Microbiol. 9:143.
Omasits U, et al. 2013. Directed shotgun proteomics guided by saturated RNA-seq identifies a complete expressed prokaryotic proteome. Genome Res. 23:1916–1927.
Ovchinnikova OG, Rozalski A, Liu B, Knirel YA. 2013. O-antigens of bacteria of the genus *Providencia*: structure, serology, genetics, and biosynthesis. Biochemistry (Moscow) 78:798–817.
Parker J, et al. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. Nature 502:228–231.
Pereto J, Lopez-Garcia P, Moreira D. 2004. Ancestral lipid biosynthesis and early membrane evolution. Trends Biochem Sci. 29:469–477.
Petersen FT, Meier R, Kutty SN, Wiegmann BM. 2007. The phylogeny and evolution of host choice in the Hippoboscoidea (Diptera) as reconstructed using four molecular markers. Mol Phylogenet Evol. 45: 111–122.
Podsiadly E, Chmielewski T, Marczak R, Sochon E, Tylewska-Wierzbanowska S. 2007. *Bartonella henselae* in the human environment in Poland. Scand J Infect Dis. 39:956–962.
Polz MF, Alm EJ, Hanage WP. 2013. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. Trends Genet. 29:170–175.
Pulliainen AT, Dehio C. 2012. Persistence of *Bartonella* spp. stealth pathogens: from subclinical infections to vasoproliferative tumor formation. FEMS Microbiol Rev. 36:563–599.
Reva ON, et al. 2006. Functional genomics of stress response in *Pseudomonas putida* KT2440. J Bacteriol. 188:4079–4092.
Rogers AB. 2012. Gastric *Helicobacter* spp. in animal models: pathogenesis and modulation by extragastric coinfections. Methods Mol Biol. 921: 175–188.
Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61: 539–542.
Saenz HL, et al. 2007. Genomic analysis of *Bartonella* identifies type IV secretion systems as host adaptability factors. Nat Genet. 39: 1469–1476.
Saisongkorh W, Robert C, La Scola B, Raoult D, Rolain JM. 2010. Evidence of transfer by conjugation of type IV secretion system genes between *Bartonella* species and *Rhizobium radiobacter* in amoeba. PLoS One 5: e12666.
Sato S, et al. 2012. Isolation and phylogenetic analysis of *Bartonella* species from wild carnivores of the suborder Caniformia in Japan. Vet Microbiol. 161:130–136.
Schonknecht G, et al. 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. Science 339: 1207–1210.
Seshadri R, et al. 2003. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. Proc Natl Acad Sci U S A. 100: 5455–5460.
Spoering AL, Vulic M, Lewis K. 2006. GlpD and PlsB participate in persister cell formation in *Escherichia coli*. J Bacteriol. 188:5136–5144.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.
Toft C, Andersson SG. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. Nat Rev Genet. 11:465–475.
Trowbridge RE, Dittmar K, Whiting MF. 2006. Identification and phylogenetic analysis of *Arsenophonus*- and *Photorhabdus*-type bacteria from adult Hippoboscidae and Streblidae (Hippoboscoidea). J Invertebr Pathol. 91:64–68.
Tsai YL, Chang CC, Chuang ST, Chomel BB. 2011. *Bartonella* species and their ectoparasites: selective host adaptation or strain selection between the vector and the mammalian host? Comp Immunol Microbiol Infect Dis. 34:299–314.
Viezens J, Arvand M. 2008. Simultaneous presence of two different copies of the 16S rRNA gene in *Bartonella henselae*. Microbiology 154: 2881–2886.
Whary MT, Fox JG. 2004. Natural and experimental *Helicobacter* infections. Comp Med. 54:128–158.
Whiting MF, Whiting AS, Hastriter MW, Dittmar K. 2008. A molecular phylogeny of fleas (Insecta: Siphonaptera): origins and host associations. Cladistics 24:677–707.
Wildschutte H, Preheim SP, Hernandez Y, Polz MF. 2010. O-antigen diversity and lateral transfer of the wbe region among *Vibrio splendidus* isolates. Environ Microbiol. 12:2977–2987.
Williams KP, et al. 2010. Phylogeny of gammaproteobacteria. J Bacteriol. 192:2305–2314.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. 22: 1107–1118.
Yeh JI, Chinte U, Du S. 2008. Structure of glycerol-3-phosphate dehydroenase, an essential monotopic membrane enzyme involved in respiration and metabolism. Proc Natl Acad Sci U S A. 105: 3280–3285.
Zhu Q, Kosoy M, Dittmar K. 2014. HGTector: an automated method facilitating genome-wide discovery of putative horizontal gene transfers, bioRxiv. Posted April 2, 2014.
Zientz E, Dandekar T, Gross R. 2004. Metabolic interdependence of obligate intracellular bacteria and their insect hosts. Microbiol Mol Biol Rev. 68:745–770.

Associate editor: Purificación López-García