The louse-borne human pathogen *Bartonella quintana* is a genomic derivative of the zoonotic agent *Bartonella henselae*

Cecilia M. Alsmark*[†‡], A. Carolin Frank*[†], E. Olof Karlberg*[†], Boris-Antoine Legault*, David H. Ardell*[§], Björn Canbäck*[¶], Ann-Sofie Eriksson*, A. Kristina Näslund*, Scott A. Handley*[‖], Maxime Huvet*, Bernard La Scola*[,**], Martin Holmberg[††], and Siv G. E. Andersson*[‡‡]

*Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, 752 36 Uppsala, Sweden; and [††]Department of Medical Sciences, Section for Infectious Diseases, Uppsala University Hospital, 752 85 Uppsala, Sweden

Edited by Stanley Falkow, Stanford University, Stanford, CA, and approved February 19, 2004 (received for review September 4, 2003)

We present the complete genomes of two human pathogens, *Bartonella quintana* (1,581,384 bp) and *Bartonella henselae* (1,931,047 bp). The two pathogens maintain several similarities in being transmitted by insect vectors, using mammalian reservoirs, infecting similar cell types (endothelial cells and erythrocytes) and causing vasculoproliferative changes in immunocompromised hosts. A primary difference between the two pathogens is their reservoir ecology. Whereas *B. quintana* is a specialist, using only the human as a reservoir, *B. henselae* is more promiscuous and is frequently isolated from both cats and humans. Genome comparison elucidated a high degree of overall similarity with major differences being *B. henselae* specific genomic islands coding for filamentous hemagglutinin, and evidence of extensive genome reduction in *B. quintana*, reminiscent of that found in *Rickettsia prowazekii*. Both genomes are reduced versions of chromosome I from the highly related pathogen *Brucella melitensis*. Flanked by two rRNA operons is a segment with similarity to genes located on chromosome II of *B. melitensis*, suggesting that it was acquired by integration of megareplicon DNA in a common ancestor of the two *Bartonella* species. Comparisons of the vector–host ecology of these organisms suggest that the utilization of host-restricted vectors is associated with accelerated rates of genome degradation and may explain why human pathogens transmitted by specialist vectors are outnumbered by zoonotic agents, which use vectors of broad host ranges.

The genus *Bartonella* contains three major human pathogens, all of which are facultative intracellular bacteria. *Bartonella quintana* is the causative agent of trench fever, a disease that affected more than 1 million soldiers during World War I and is currently reemerging among homeless and alcoholic individuals in urban areas (1). The trench fever agent is a human specialist that is transmitted by the human body louse, *Pediculus humanus*, which also serves as the vector for the typhus pathogen, *Rickettsia prowazekii*. The agent of Carrión's disease, *Bartonella bacilliformis*, is likewise a human pathogen specialist. In contrast, *Bartonella henselae* infects both humans and cats; 30–60% of domestic cats in the U.S. are infected with *B. henselae*. Transmission among cats is mediated by the cat flea, *Ctenocephalides felis*, and to humans by cat scratches or cat bites (1).

Natural *B. henselae* infections produce no clinical symptoms in the cat, whereas cat-scratch disease in humans causes disease manifestations in lymphatic organs, the skin, the liver, and the cardiovascular and the nervous systems (1). *B. henselae* and *B. quintana* are unique among bacterial pathogens in that they may induce tumor-like lesions of the skin (bacillary angiomatosis), the liver, and the spleen (bacillary peliosis) in immunocompromised individuals, predominantly AIDS patients (2, 3). Pathological angiogenesis manifested as skin lesions may also be induced by *B. bacilliformis* (verruga peruana). Both *B. henselae* and *B. quintana* are able to invade endothelial cells and replicate within these cells (4, 5). Adhesion to endothelial cells by *B. henselae* has mainly been studied in human umbilical vein endothelial cells, and internalization has been shown to occur both by conventional phagocytosis and by an invasome-mediated mechanism (5). Colonization of endothelial cells is important in both reservoir and incidental mammalian hosts and is considered essential for the establishment and maintenance of the angioproliferative lesions (4, 6–8).

Of the 19 described *Bartonella* species that infect a wide variety of domestic and wild animals such as cats, dogs, mice, rats, squirrels, deer, and moose, 7 have been associated with human disease (9, 10). Experimental animal infection models suggest that the initial infection in the host is in a yet unknown niche, from which erythrocytes are periodically seeded (10–12). The VirB/VirD type IV secretion system is required at an early stage of the infection, possibly for colonization of the primary niche (11). Bacteria can persist in mature erythrocytes for several weeks, with transmission among hosts being mediated by blood-sucking arthropods. In human infections, *B. quintana*, but not *B. henselae*, shows intraerythrocytic presence. The human body louse (*B. quintana*), the cat flea (*B. henselae*), and the sand fly (*B. bacilliformis*) have all been implicated in the eco-epidemiology of *Bartonella* infections. Several *Bartonella* species, including *B. henselae*, have been detected by PCR in ticks (13), which feed on a large variety of vertebrate hosts.

The expanding number of species identified, the novel disease symptoms described, and the potential for genetic exchange across species have long-term consequences for the epidemiology and virulence of *Bartonella* infections. To study the evolution of closely related vector-borne pathogens with different host preference patterns, we compared the complete genome sequence of the louse-borne human specialist *B. quintana* with that of its close relative, the zoonotic agent *B. henselae*. Our analysis suggests that specialization to a single vector and/or host is associated with the loss of extrachromosomal DNA and phage genes, possibly making these species less able to accommodate change and invade novel environmental niches.

This paper was submitted directly (Track II) to the PNAS office.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. BX897699 (*B. henselae*) and BX897700 (*B. quintana*)].

[†]C.M.A., A.C.F., and E.O.K. contributed equally to this work.

[‡]Present address: Department of Zoology, Natural History Museum, London SW7 5BD, United Kingdom.

[§]Present address: Linnaeus Centre for Bioinformatics, Biomedical Center, Uppsala University, 752 36 Uppsala, Sweden.

[¶]Present address: Department of Microbial Ecology, Lund University, 223 62 Lund, Sweden.

[‖]Present address: Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110.

**Present address: Unité des Rickettsies, Faculté de Médecine, 133 85 Marseille, France.

[‡‡]To whom correspondence should be addressed. E-mail: siv.andersson@ebc.uu.se.

www.pnas.org/cgi/doi/10.1073/pnas.0305659101

Materials and Methods

Genome Sequencing and Assembly. Starting from single colonies, *B. henselae* (strain Houston-1; original isolate) and *B. quintana* (strain Toulose) were cultured on blood agar plates at 35°C under 5% CO_2/95% air for 5 days. DNA preparation, library construction, and sequencing were as described previously (14). A total of 36,328 and 22,731 shotgun sequences were obtained from the M13 libraries of *B. henselae* and *B. quintana*, respectively. Additionally, 5,469 pUC18 clones and 3,166 PCR products from *B. henselae* and 5,217 pUC18 clones and 2,303 PCR products from *B. quintana* were sequenced. After assembly with the PHRED and PHRAP software, physical gaps were closed by direct sequencing of both ends of gap-covering pUC18 clones as well as by long-range PCR by using primers from contig ends. The location and order of sequences covering long repeated regions in *B. henselae* were resolved by hybridization of PCR products to Southern blots of *Not*I digests separated by pulse-field gel electrophoresis. The organization of these and other repetitive regions was confirmed by ordinary PCR and long-range PCR by using single locus primers. All ambiguous sites were manually edited with CONSED by resequencing of PCR products when necessary. In the final version of the genome sequence, each nucleotide had been sequenced from both directions at least once until assigned the highest quality value by PHRED.

Genome Analysis. Annotation was accomplished with the help of DANS, an annotation system developed in-house (H. H. Fuxelius, B.C., E.O.K., A.C.F., and S.G.E.A., unpublished data). DANS allows annotation of unfinished microbial genome sequence data, with the annotations constantly being propagated into the most recent version of the sequence assembly. The system is based on software and standard algorithms for bioinformatic analyses such as the ORF-prediction software GLIMMER (15), BLAST (16), CLUSTALW (17), PAUP (18), and the tRNA-prediction software ARAGORN (19). tmRNAs were identified by using BRUCE (20), and initiator tRNAs were identified by using TFAM (D.H.A. and S.G.E.A., unpublished data). Comparisons of metabolic pathways in α-proteobacteria were performed with the help of COLOR PATHWAY (21).

Criteria used for discriminating true genes among ORFs without homologs in other species were based on estimates of G+C content values, strand-specific mutation biases, ratios of nonsynonymous and synonymous substitutions for orthologs present in the two genomes, length, and genomic locations. Within ORFs, base composition ratios of the second and third codon positions were statistically compared under the hypothesis of equality using the normal approximation of the χ^2 test (22). Both the proportions of G+C and of G+T were used, and candidates that did not have significant differences in base composition ($\alpha = 0.05$) were examined further. In addition, we compared the distribution of [T, (not T)] at third codon positions for each ORF to that of the genomic strand in which the ORF resides, to search for strand-biased signals of mutation associated with transcription. This test (χ^2, $\alpha = 0.01$) sensitively discriminated ORFs with significant BLAST hits to known homologs and achieved the expected false-positive rate with ORFs not annotated as true genes.

An ortholog table for *B. henselae* and *B. quintana* ORFs was constructed based on the *E* values obtained in BLAST searches (16), in which one genome was used as a query against the other. A conservative cutoff value ($<1e^{-80}$) was initially used to identify a set of ORFs present in both *B. henselae* and *B. quintana*. This resulted in an ortholog backbone table in which genes had been sorted according to their order in the two genomes. Orthologs associated with a less conservative E-value ($<1e^{-20}$) were subsequently incorporated into this table if the backbone gene order was preserved. The homologous pairs were aligned through their inferred protein translations with CLUSTALW (17), using the BLOSUM matrix series and default settings. The CODEML program of the PAML package (v. 3.0) (23) was then used to calculate pairwise d_N/d_S, ratios assuming homogeneity of selection over codons and transition–transversion bias as a free parameter (initial estimation value 2).

Proteins in repeat families were clustered together by amino acid identity (>50% over >50% of the length of the protein) as inferred by BLAST. Sequences with sequence similarities to a functional gene and/or an ORF in *B. henselae* and/or *B. quintana* but not spanning an entire ORF with initiation and termination codons were considered pseudogenes. Short fragments of sequences with sequence similarity to a functional gene and/or an ORF in the *B. henselae* and/or *B. quintana* genomes were treated as extensively degraded gene remnants.

Similarity searches of all *B. quintana* and *B. henselae* protein sequences against the protein sequences encoded by *Bartonella suis* chromosomes I and II were made separately by using BLASTP ($E < 1e^{-20}$). For *B. henselae* proteins with hits against proteins encoded on both *B. suis* chromosomes, the top hit in each chromosome was retrieved for phylogenetic analysis, along with top BLASTP hits ($E < 1e^{-10}$) from *Agrobacterium tumefaciens*, *Sinorhizobium meliloti*, *Escherichia coli*, *Xylella fastidiosa*, and *Caulobacter crescentus*. The protein sequences were aligned by using CLUSTALW, and phylogenetic trees were made by using the neighbor-joining algorithm (24) as implemented in NEIGHBOR from the PHYLIP package. Pair-wise maximum likelihood distances were calculated with TREE-PUZZLE (25) using the WAG model of protein evolution.

Results and Discussion

General Features of the *Bartonella* Genomes. *B. quintana* contains a single circular chromosome that comprises 1,581,384 bp, whereas the *B. henselae* genome has a larger size of 1,931,047 bp (Fig. 1). A total of 1,116 orthologous protein-coding genes were identified among the 1,143 and 1,491 genes present in *B. quintana* and *B. henselae*, respectively (Table 1, which is published as supporting information on the PNAS web site). The overall coding fractions of the *B. quintana* and *B. henselae* genomes were estimated to be as low as 72.7% and 72.3%, respectively. Both genomes show strong strand-specific mutation biases, with large excesses of G and T on the leading strands, a feature that was here used to infer the putative origin of replication (*oriC*). Nucleotide sequence divergences for orthologous genes were 0.073 nonsynonymous and 0.620 synonymous substitutions per site on average.

Phage Integrations/Excisions and Chromosomal Rearrangements. The backbone of the two *Bartonella* genomes is colinear with the exception of a few rearrangements (Fig. 2*b*) that correspond to three symmetric translocation/inversion events or a series of inversions around the terminus of replication. In addition, there is a small, symmetric translocation of a 20-kb fragment. The breakpoints for rearrangements coincide with long repeated sequences encompassing both genes and noncoding sequences located on genomic islands that are uniquely present in *B. henselae* (Fig. 2*a*). Present at the breakpoint positions in *B. quintana* are pseudogenes and gene remnants of the *B. henselae* islands, indicating that the islands were present in their common ancestor (Fig. 2*c*). Comparative analyses with the *Brucella melitensis* genome suggest an evolutionary scenario that involves excision of sequence fragments associated with rearrangements in the *B. quintana* genome through recombination at repeated sequences in the islands.

In total, we identified 301 genes unique to *B. henselae*, but only 26 unique *B. quintana* genes. A majority, 62%, of the genes solely present in *B. henselae* are located on four main segments; a

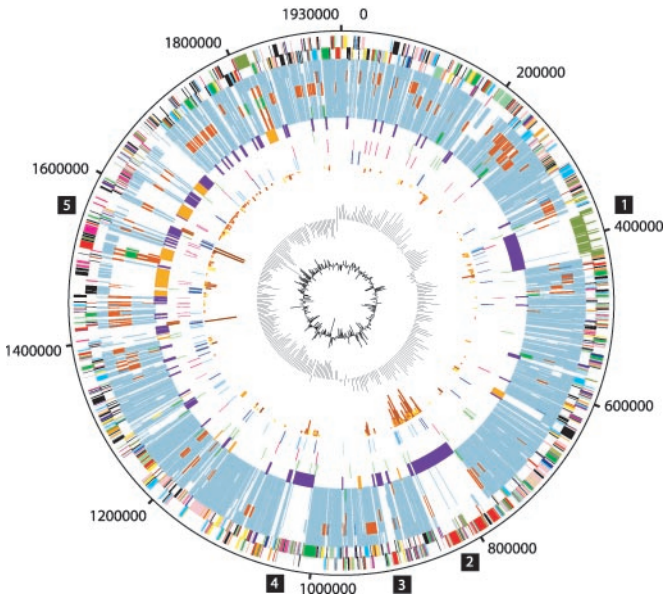

Fig. 1. The *B. henselae* and the *B. quintana* genome maps and the location of genes with homologs in other α-proteobacterial genomes. The outer circle shows predicted coding regions on the plus strand in *B. henselae* color-coded by role categories. The second circle shows predicted coding regions on the minus strand color-coded by role categories. Third circle, genes with orthologs in *B. quintana*. Fourth circle, top hits to *B. melitensis* according to replicon: blue, main chromosome (I); red, chromosome II. Fifth circle, top hits to *A. tumefaciens* according to replicon: blue, circular chromosome; red, linear chromosome, green, plasmid pAtC58, dark green, plasmid pTiC58. Sixth circle, top hits to *Mesorhizobium loti* according to replicon: blue, main chromosome; red, plasmid pMLa; green, plasmid pMLb. Seventh circle, top hits to *S. meliloti* according to replicon: blue, main chromosome; red, plasmid pSymA; green, plasmid pSymB. Eighth circle, *B. henselae* islands and islets in gray, *Bartonella* islands in orange. Ninth circle, tRNAs in green, rRNAs in red. Tenth circle, integrase remnants in pink. Eleventh circle, pseudogenes in blue and extensively degraded gene remnants in light blue. Twelfth circle, repeats, the length of the line is proportional to the length of the repeated region, and the color gradient represents percent similarity ranging from 100% (red) to 75% (yellow). Thirteenth circle, GC skew in sliding window of 15 kb, step size 7 kb. Innermost circle, deviation from average GC content in sliding windows of 15 kb, step size 7 kb. Numbers refer to prophage (1), *B. henselae* specific islands of 72 kb (2), 9 kb (3), 34 kb (4), and *Bartonella* specific island (5).

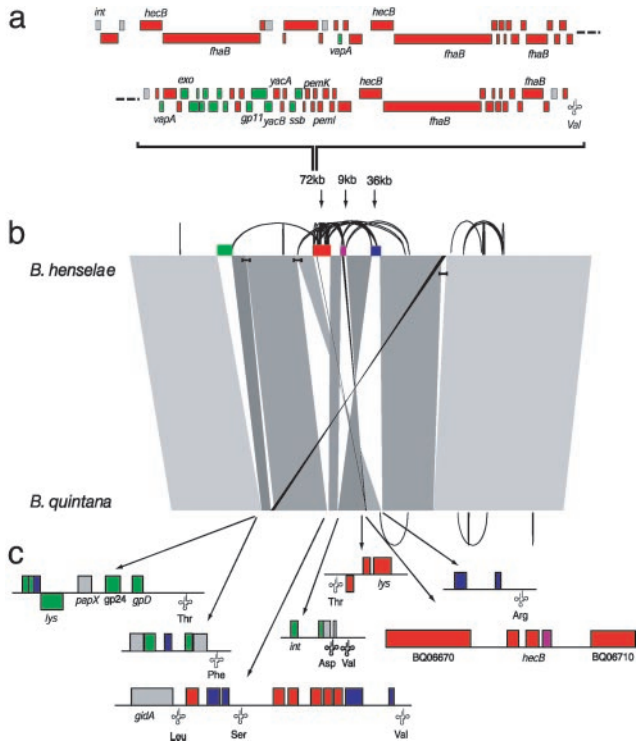

Fig. 2. Schematic illustration of the rearrangements observed between *B. henselae* and *B. quintana*. (*a*) Gene order structure in the 72-kb island of *B. henselae*. (*b*) The breakpoints for rearrangements coincide with species-specific islands and repeated sequences. *B. henselae*-specific islands are color-coded (green, the prophage; red, the 72-kb island; purple, the 9-kb island; blue, the 34-kb island). Repeated regions longer than 1,000 bp are drawn as arcs between the two repeat copies. Black horizontal bars represent gene-order synteny with *B. melitensis*. (*c*) Pseudogenes and remnants of genes in these islands at the *B. quintana* breakpoints are marked with squares in the corresponding colors. Pseudogenes other than phage and integrase are shown in gray color.

prophage region of 55 kb and three genomic islands of 72 , 34, and 9 kb (Fig. 1, regions 1–4). The prophage is an evolutionary mosaic with genes of different origins interspersed with genes showing sequence similarities to a putative prophage in *Wolbachia pipientis* (26). The integration site is flanked on one side by a gene coding for $tRNA^{Leu}$ and on the other by an integrase gene. Present at the corresponding position in *B. quintana* is 3.5-kb noncoding sequences consisting of prophage gene remnants (Fig. 2*c*), suggesting loss of the prophage in this species.

Each of the three *B. henselae* islands is flanked by tRNAs and integrase genes, as expected if acquired by means of phage integrations (Fig. 1, circles 9 and 10). Present on the 34- and 72-kb islands (Fig. 2*a*) are multiple copies of *fhaC*/*hecB* and *fhaB* that form a two-partner secretion system (27), where the *fhaC*/*hecB* gene product mediates transport of filamentous hemagglutinin encoded by *fhaB*. A phylogenetic analysis of two-partner domains suggests that these may be horizontally transmitted across species (28), which is consistent with their location on genomic islands in *B. henselae*. The *fhaB* homologs are flanked by sequences putatively coding for proteins of 138 aa that are repeated 20 times in the *B. henselae* genome. In addition to the three main islands, a short region of unique DNA in *B. henselae* displays similarity to the pathogenicity island of *Photorabdus luminescens* (29).

One of the few segments present in *B. quintana* but not in *B. henselae* is located next to a gene for prophage lysozyme and a tRNA gene and contains a gene with similarity to *yopP* in *Yersinia enterocolitica* that codes for a secreted effector molecule causing apoptosis in macrophages (30). Another such segment, also located next to a tRNA gene, contains a gene coding for a putative toxin/hemolysin secretion protein that is present in *S. meliloti* but not in other α-proteobacteria. These are the only two genes in *B. quintana* that are not present in *B. henselae* but show sequence similarity to genes from other species. Taken together, these observations suggest that the *B. quintana* genome is derived from a larger *B. henselae*-like ancestral genome by reductive genome evolution, including loss of sequences acquired by phage integrations.

Repeat Families. The *B. henselae* genome contains an unusually high fraction of repeated or partially repeated genes (>50% amino acid identity over >50% of the gene length) with 78 repeat families of 2–14 members in each group, 41 of which are two-member families. In the 34-kb and 72-kb islands (Fig. 2*a*), there are 4 and 14 genes, respectively, that show similarity to those found at the 3′ end of the prophage with up to 100% identity at the amino acid level (Fig. 1, circle 12). In total, 46% of the identified genes in the three main islands are members of repeat families, and 62% of the 78 repeat families contain genes

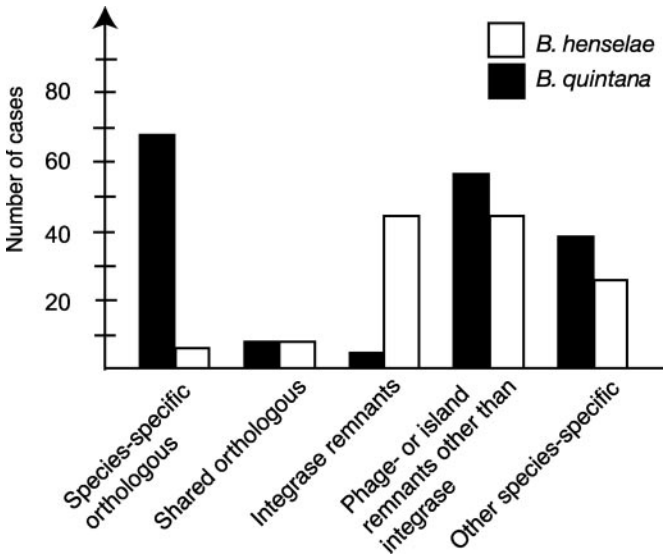

Fig. 3. Pseudogenes. The number of pseudogenes in *B. henselae* (open bars) and *B. quintana* (filled bars) is shown for different classes of genes.

from one of these segments. In contrast, *B. quintana* contains only 18 repeat families, 12 of which are two-member families. The largest family in *B. quintana* contains seven tandemly repeated copies of the *trwL* gene located in one of the two operons coding for putative type IV secretion systems (12), and the second largest is a family of hemin-binding proteins (31).

Pseudogenes. In total, we identified 128 pseudogenes and extensively degraded gene sequences in *B. henselae* (excluding 23 tandem duplication remnants), as compared with 175 fragmented genes in *B. quintana* (Fig. 3). As many as 67 pseudogenes in *B. quintana* are represented by intact orthologs in *B. henselae*, where they code for proteins with basic metabolic functions. In contrast, only six *B. henselae* pseudogenes are present as intact genes in *B. quintana*, a few of which are found in the operon *fatBCD* that codes for proteins involved in iron transport. As many as 57 and 75 pseudogenes in *B. quintana* and *B. henselae* represent degraded fragments of genes present in the phage and the islands.

Notable are 4 and 43 integrase gene remnants in *B. quintana* and *B. henselae*, respectively (Fig. 1, circle 10), many of which reside in intergenic regions spanning in length from 1 to 4 kb that have otherwise no association with genomic islands, phage remnants, tRNA genes, or pseudogenes (Fig. 4*a*). A majority are extensively degraded, with only the 5′-terminal and/or central part of the integrase left in the genome (Fig. 4*b*). Because integrase genes often contain promoter sequences at their 5′ terminus, these sequences may function as promoter cassettes. The homologous spacer regions in *B. quintana* do not contain integrase genes but are significantly longer than other spacers in this species ($P = 1.6e^{-6}$, Wilcoxon signed rank test), perhaps indicating that they represent the remains of highly degraded integrase sequences (Fig. 4*a*). Related integrase genes are also present in other α-proteobacterial genomes, but none of these show the same broad distribution as in *B. henselae*.

Integration of Megareplicon Sequences. Members of the genus *Brucella* are facultative intracellular pathogens that cause abortions in pregnant animals. *Bartonella* and *Brucella* are phylogenetic sister clades (32), but despite the shared ancestry and many lifestyle similarities they differ remarkably in genome sizes and structures. *Brucella* species have genomes of 3.3 megabases (Mb)

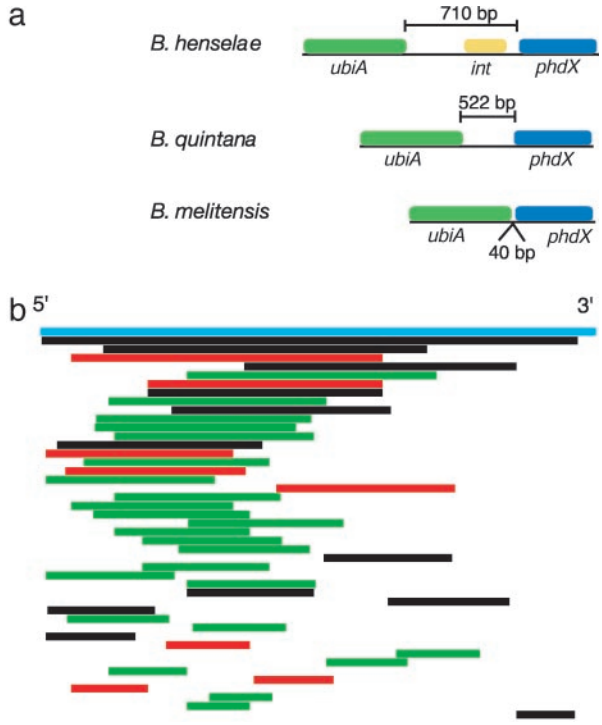

Fig. 4. Integrase genes. (*a*) Size of the spacer region flanked by *ubiA* and *phdX* in *B. henselae*, *B. quintana*, and *B. melitensis*. (*b*) Size of integrase gene remnants in *B. henselae* sorted according to length. Sequence similarity against the full-length integrase gene was inferred by using the tBLASTn algorithm (black, $E < 1e^{-30}$; red, $E < 1e^{-20}$; green, $E < 1e^{-4}$).

that normally are split into two chromosomes of 2.1 Mb (Chr I) and 1.1 Mb (Chr II) (33, 34). In total, we identified 760 *B. henselae* genes for which homologs are present on Chr I of *B. suis* ($E < 1e^{-20}$). Only 122 *B. henselae* genes have homologs solely on Chr II and another 119 have homologs on both chromosomes. To trace the origin of *B. henselae* genes with similarity to sequences on Chr II, we reconstructed phylogenetic trees for 103 of the 119 *B. henselae* genes with homologs on both chromosomes of *B. suis* and for which an outgroup was available. A majority of these, 75 trees, support a clustering of the *B. henselae* gene with its homolog on Chr I to the exclusion of the gene duplicate on Chr II. Only two genes support a clustering of the two *B. suis* homologs, as expected for recent gene duplication and transfer events.

In the remaining 26 trees, the *B. henselae* gene clusters with its homolog on the second megareplicon of *B. suis*. A closer inspection of gene order structures reveals that as many as 18 of the 26 *B. henselae* genes that cluster with *B. suis* genes from Chr II are flanked on one or both sides by genes with homologs solely on Chr II in *B. suis*. These small islands, which encompass 4–20 genes with hits solely or predominantly to Chr II, are dispersed around the *B. henselae* genome. Additional short stretches of consecutive genes with homologs solely on Chr II were identified, although for these no paralogs are present on Chr I to verify their origin. Both the prophage and the 72-kb island are flanked on one or both sides by genes with top hits to Chr II of *B. suis*.

The most notable of these *B. suis* Chr II-like segments is located in the fourth quarter of the *B. henselae* chromosome and is flanked by the two rRNA operons (Fig. 1, circle 9). As many as 20 genes in the intervening segment display top hits or cluster with genes on Chr II, as compared with only two genes with top hits to Chr I (Table 2, which is published as supporting information on the PNAS web site). Further downstream of the rRNA

EVOLUTION

operons is another segment encompassing some 200 genes that stand out by having a remarkably low coding potential and many genes with no homologs to other genomes (Fig. 1, region 5). The coding content of a segment spanning 200 kb in *B. quintana* and 282 kb in *B. henselae* is only 50% and 60%, respectively (Table 2). Flanking this large segment on the 3′ side is a fragment homologous to that located in between the rRNA operons on Chr I in *Brucella*.

Genes with assigned functions in this region include those coding for surface proteins, a type IV secretion system, as well as plasmid and phage genes. As many as 21.4% of the identified genes in this region of the *B. henselae* genome are genus-specific, as compared with 5.5% in the remaining part of the genome (excluding islands). The fraction of species-specific genes is also high, amounting to 11.5% in *B. quintana* as compared with 0.3% in the rest of the genome. Located in this region are *yopP* and the other few *B. quintana* genes without homologs in *B. henselae* but with similarity to other genomes. This region is also atypical in that it has a slightly elevated G+C content at third codon positions, many integrase remnants, and numerous short repeated sequences in tandem, which in other species are implicated in phase variable expression of surface-exposed antigens (36).

The high noncoding content, the many *Bartonella*-specific genes, and the high density of genes that cluster with Chr II in *Brucella* support an evolutionary scenario that includes one or more integration events of genes from a second replicon by homologous recombination at the rRNA operons, followed by sequence loss and rearrangements. In this context, it is interesting to note that *B. suis* biovar 3 contains a single circular chromosome in which chromosome II is integrated into the main chromosome by recombination at the rRNA operons (35). Some of the genes acquired by phage and plasmid insertions at a second replicon that later integrated into the main chromosome may have played an important role for the evolution of lifestyle features unique to *Bartonella*, such as vector-borne transmission pathways and the colonization of erythrocytes.

Host-Integrated Metabolism. Vector-borne intracellular lifestyles have evolved twice in the α-proteobacteria, in the lineages leading to *Rickettsia* and *Bartonella.* These two genera are not sister clades, and there is no conservation of gene order structures beyond that of individual operons. Both lineages show signs of reductive genome evolution, including small genome sizes and low genomic coding contents (14, 37). The lower gene contents in the obligate intracellular *Rickettsia* species (14, 37) are due to a broader substitution of bacterial gene functions for host-supplied compounds such as amino acids, nucleoside monophosphate, and cofactors than in *Bartonella*. The presence of genes encoding enzymes in pathways for the production of pyruvate in *Bartonella* (38) contrasts with the absence of such genes in *Rickettsia*, which rely on the import of pyruvate from the host cell cytoplasm. Additionally, *Rickettsia* imports ATP from the cytoplasm with the aid of genes encoding transport systems for ATP and ADP that are absent from *Bartonella*.

There are only a few examples of the converse; although both species are capable of aerobic growth, genes for cytochrome oxidase subunits I, II, and III (*coxABC*) were not identified in *Bartonella*. Likewise, genes involved in hemin biosynthesis are present in *Rickettsia* but absent from *Bartonella.* Also, these differences correlate with lifestyle characteristics; *Bartonella* parasitizes erythrocytes (10–12) and is suggested to acquire essential iron molecules from the heme moiety of hemoglobin. Consistently, *B. quintana* and *B. henselae* have the highest reported hemin requirement for bacterial growth *in vitro* (39), and no genes for hemin biosynthesis were identified in either genome. A hemin-binding protein was characterized in *B. quintana*, and four additional members of this gene family were identified, all of which are expressed under normal growth conditions (31). A large number of genes coding for iron and hemin transporters, and two genes putatively coding for heme-dependent transcriptional regulators, were identified in both the *B. henselae* and the *B. quintana* genomes. These may be essential for the import of iron from the gut of the louse and the cat flea, respectively, as well as from erythrocytes of cats and humans.

Genome Evolution of Louse-Borne Human Specialists *R. prowazekii* and *B. quintana*. Evolutionary theory predicts that single-host pathogens (specialists) should be more successful than multihost pathogens (generalists), due to functional trade-offs that limit the fitness of generalists and because evolution is expected to proceed faster in narrower niches (40, 41). The two louse-borne human specialists *R. prowazekii* and *B. quintana* are reduced genomic versions of their closest relatives, the tick-borne *Rickettsia conorii* and the cat flea-borne *B. henselae*, respectively, with no examples of shared, acquired genes that would explain the adaptation to their unique lifestyle and growth niche. *Rickettsia* normally invade the tissues of ticks (and fleas) and are vertically transmitted to their progeny, with mammals being accidental hosts and serving as vectors between ticks. *Bartonella* has a different ecology; here, mammals are the reservoirs and fleas are the vectors, within which bacteria survive in the gut without invading tissues. Despite these ecological differences, transmission by the human body louse is in both *Rickettsia* and *Bartonella* associated with the inactivation and loss of ancestral genes that are retained in their contemporary closest relatives with a different host range size. This is reminiscent of the finding that the human specialist pathogen *Bordetella pertussis* is a genomic derivative of *Bordetella bronchiseptica*, which infects a wide range of animals (42). As in the case of *B. henselae*, large regions of unique DNA were identified in *B. bronchiseptica*, several of which are derived from prophages (42).

For vector-borne pathogens, the host-specificity and lifestyle of the vector is a major determinant of the potential for cross-species transmission. In contrast to the narrow host range of the human body louse (43), fleas and ticks can feed on a broad range of mammals (44). Furthermore, lice have a generation time of less than 1 month (45), cat fleas of less than 3 months (46), whereas ticks can have a lifespan of several years (47). Short generation times and extreme host-restriction for lice may restrict opportunities for future transmission to other hosts. However, it is interesting to note that *R. prowazekii* has been identified in flying squirrels with other species of fleas and lice as probable vectors (48), possibly suggesting that *R. prowazekii* spread into the human population from the flying squirrel cycle. Indeed, *R. prowazekii* seems to be well adapted to squirrels, as no mortality or morbidity has been observed among infected animals, whereas mortality in infected human body lice is close to 100%. Likewise, *B. quintana* infections in humans may reflect recent cross-species transmission events from other, related host–vector systems.

The human body louse is estimated to have diverged from the head louse only about 72,000 years ago, possibly associated with the use of clothing and the spread of humans out of Africa (49). Although there is no calibrated molecular clock for *Bartonella*, the level of sequence divergence for *B. henselae* and *B. quintana* ($d_N = 0.073$; $d_S = 0.62$) is inconsistent with such a recent divergence date for the split between these two species. This observation suggests either that the transmission to humans occurred before the emergence of the human body louse or alternatively that *R. prowazekii* and *B. quintana* represent highly diverged lineages that became isolated from their relatives upon entering the human–louse–human cycle or other enzoonotic cycles. Preliminary studies of strain diversity in *B. henselae* reveal substantial variation in genome sizes and presence/absence of genomic islands. However, none of the *B. henselae* strains studied

so far are as highly reduced in size as *B. quintana*, and some strains contain partial genomic islands, which supports the theory that the islands were present in a common ancestor of the two species and have been lost from *B. quintana* (H. Lindroos, A. Mira, D. Repsilber, A.K.N., and S.G.E.A., unpublished data).

Thus, our results suggest that adaptations to vector-borne transmission pathways and pathogen specialization to a unique host has predictable consequences. Initially, the status as a specialist or a generalist may reflect the host range size of the vector and the expression of virulence traits that enhance the survival potential in novel hosts. With time, however, narrow growth niches will make specialists less able to cope with environmental changes because of extensive gene loss and limited opportunities for novel gene acquisitions. Unless stabilized by host-selected functions, species specialists may in the long term be out-competed by invading generalists with improved phenotypes acquired during selective sweeps in their broader host–vector systems. Although the rate of genome evolution is enhanced for pathogens that are transmitted by specialist vectors, as predicted (40, 41), it is the rate of gene loss, rather than the acquisition of novel virulence traits, that proceeds more rapidly. Future studies of sequence polymorphism and genome size diversity among louse-borne human pathogens and comparisons with isolates from their natural sources will provide information about the timing of the sequence elimination events.

This research was supported by grants from the Swedish Foundation for Strategic Research (to S.G.E.A.), the Swedish Research Council (to S.G.E.A.), the Knut and Alice Wallenberg Foundation (to S.G.E.A.), the 5th European Union-framework program Quality of Life (to S.G.E.A.), and the National Science Foundation (to D.H.A.). For specific author contributions, see *Supporting Text*, which is published as supporting information on the PNAS web site.

1. Koehler, J. E. (1996) *Adv. Pediatr. Infect. Dis.* **11,** 1–27.
2. Regnery, R. L., Anderson, B. E., Clarridge, J. E., III, Rodriguez-Barradas, M. C., Jones, D. C. & Carr, J. H. (1992) *J. Clin. Microbiol.* **30,** 265–274.
3. Anderson, B. E. & Neuman, M. A. (1997) *Clin. Microbiol. Rev.* **10,** 203–219.
4. Brouqui, P. & Raoult, D. (1996) *Res. Microbiol.* **147,** 719–731.
5. Dehio, C., Meyer, M., Berger, J., Schwarz, H. & Lanz, C. (1997) *J. Cell Sci.* **110,** 2141–2151.
6. Kirby, J. E. & Nekorchuk, D. M. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 4656–4661.
7. Kempf, V. A., Volkman, B., Schaller, M., Sander, C. A., Alitalo, K., Riess, T. & Autentriech, I. B. (2001) *Cell Microbiol.* **3,** 623–632.
8. Resto-Ruiz, S. I., Schmiederer, M., Sweger, D., Newton, C., Klein, T. W., Friedman, H. & Anderson, B. (2002) *Infect. Immun.* **70,** 4564–4570.
9. Breitschwerdt, E. B. & Kordick, D. L. (2000) *Clin. Microbiol. Rev.* **13,** 428–438.
10. Dehio, C. (2001) *Trends Microbiol.* **9,** 279–285.
11. Schulein, R. & Dehio, C. (2002) *Mol. Microbiol.* **46,** 1053–1067.
12. Schulein, R., Seubert, A., Gille, G., Lanz, C., Hansmann, Y., Piémont, Y. & Dehio, C. (2001) *J. Exp. Med.* **193,** 1077–1086.
13. Sanego, Y. O., Zeaiter, Z., Caruso, G., Merola, F., Shpynov, S., Brouqui, P. & Raoult, D. (2003) *Emerg. Infect. Dis.* **9,** 329–332.
14. Andersson, S. G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C. M., Podowski, R. M., Näslund, K., Eriksson, A.-S., Winkler, H. H. & Kurland, C. G. (1998) *Nature* **396,** 133–140.
15. Saltzberg, S. L., Delcher, A. L., Kasif, S. & White, O. (1998) *Nucleic Acids Res.* **26,** 544–548.
16. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Liplman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
17. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680.
18. Swofford, D. L. (1998) PAUP, *Phylogenetic Analysis Using Parsimony* (Sinauer, Sunderland, MA).
19. Laslett, D. & Canbäck, B. (2004) *Nucleic Acids Res.* **32,** 11–16.
20. Laslett, D., Canbäck, C. & Andersson, S. G. E. (2002) *Nucleic Acids Res.* **30,** 3449–3453.
21. Lindroos, H. & Andersson, S. G. E. (2002) *Proc. IEEE* **90,** 1793–1802.
22. Yang, Z. & Nielsen, R. (2000) *Mol. Biol. Evol.* **17,** 32–43.
23. Zar, J. J. (1994) *Biostatistical Analysis* (Prentice Hall, Upper Saddle River, NJ), 4th Ed.
24. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
25. Strimmer, K. & von Haesler, A. (1996) *Mol. Biol. Evol.* **13,** 964–969.
26. Masui, S., Kamoda, S., Sasaki, T. & Ishikawa, H. (2000) *J. Mol. Evol.* **51,** 491–497.
27. Jacob-Dubuisson, F., Locht, C. & Antoine, R. (2001) *Mol. Microbiol.* **40,** 306–313.
28. Rojas, C. M., Ham, J. H., Deng, W. L., Doyle, J. J. & Collmer, A. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 13142–13147.
29. Daborn, P. J., Waterfield, N., Silva, C. P. Au, C. P., Sharma, S. & ffrench-Constant, R. H. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 10742–10747.
30. Monack, D. M., Mecsas, J., Ghori, N. & Falkow, S. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10385–10390.
31. Minnick, M. F., Sappington, K. N., Smitherman, L. S., Andersson, S. G. E., Karlberg, O. & Carroll, J. A. (2003) *Infect. Immun.* **71,** 814–821.
32. Moreno, E., Stackebrandt, E., Dorsch, M., Wolters, J., Busch, M. & Mayer, H. (1990) *J. Bacteriol.* **172,** 3569–3576.
33. DelVecchio, V. G., Kapatral, V., Redkar, R. J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 443–448.
34. Paulsen, I. T., Seshadri, R., Nelson, K. E., Eisen, J. A., Heidelberg, J. F., Read, T. D., Dodson, R. J., Umayam, L., Brinkac, L. M., Beanan, M. J., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 13148–13153.
35. Jumas-Bilak, E., Michaux-Charachon, S., Bourg, G., O′Callaghan, D. & Ramuz, M. (1998) *Mol. Microbiol.* **27,** 99–106.
36. Hallet, B. (2001) *Curr. Opin. Microbiol.* **4,** 570–581.
37. Ogata, S., Audic, S., Renesto-Audiffren, P., Fournier, P. E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J. M. & Raoult, D. (2001) *Science* **293,** 2093–2098.
38. Canbäck, B., Andersson, S. G. E. & Kurland, C. G. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 6097–6102.
39. Myers, W. F. (1969) *J. Bacteriol.* **97,** 663.
40. Woolhouse, M. E. J., Taylor, L. H. & Haydon, D. T. (2001) *Science* **292,** 1109–1112.
41. Whitlock, M. C. (1996) *Am. Nat.* **148,** S65.
42. Parkhill, J., Sebaiha, M., Preston, A., Murphy, L. D., Thomson, N., Harris, D. E., Golden, M. T. G., Churcher, C. M., Bentley, S. D., Mungall, K. L., *et al.* (2003) *Nat. Genet.* **35,** 32–40.
43. Page, R. D., Lee, P. L., Becher, S. A., Griffiths, R. & Clayton, D. H. (1998) *Mol. Phylogenet. Ecol.* **9,** 276–293.
44. Rust, M. K. & Dryden, M. W. (1997) *Annu. Rev. Entomol.* **42,** 451–473.
45. Buxton, P. A. (1947) *The Louse: An Account of the Lice Which Infest Man: Their Medical Importance and Control*. (E. Arnold, London), 2nd Ed.
46. Krämer, F. & Mencke, N. (2001) *Flea Biology and Control: The Biology of the Cat Flea, Control and Prevention with Imidaclopid in Comparison with Small Animals* (Springer, Berlin).
47. Sonenshine, D. E. (1999) *Biology of Ticks* (Am. Soc. Microbiol., Washington, DC), Vol. 1.
48. Bozeman, F. M., Masiello, S. A., Williams, M. S. & Elisberg, B. L. (1975) *Nature* **255,** 545–547.
49. Kittler, R., Kayser, M. & Stoneking, M. (2003) *Curr. Biol.* **13,** 1414–1417.