

# Genomic analysis of *Bartonella* identifies type IV secretion systems as host adaptability factors

Henri L Saenz<sup>1,7</sup>, Philipp Engel<sup>1,7</sup>, Michèle C Stoeckli<sup>1</sup>, Christa Lanz<sup>2</sup>, Günter Raddatz<sup>2,6</sup>, Muriel Vayssier-Taussat<sup>1,3</sup>, Richard Birtles<sup>4</sup>, Stephan C Schuster<sup>2,5</sup> & Christoph Dehio<sup>1</sup>

The bacterial genus *Bartonella* comprises 21 pathogens causing characteristic intraerythrocytic infections. *Bartonella bacilliformis* is a severe pathogen representing an ancestral lineage, whereas the other species are benign pathogens that evolved by radial speciation. Here, we have used comparative and functional genomics to infer pathogenicity genes specific to the radiating lineage, and we suggest that these genes may have facilitated adaptation to the host environment. We determined the complete genome sequence of *Bartonella tribocorum* by shotgun sequencing and functionally identified 97 pathogenicity genes by signature-tagged mutagenesis. Eighty-one pathogenicity genes belong to the core genome (1,097 genes) of the radiating lineage inferred from genome comparison of *B. tribocorum*, *Bartonella henselae* and *Bartonella quintana*. Sixty-six pathogenicity genes are present in *B. bacilliformis*, and one has been lost by deletion. The 14 pathogenicity genes specific for the radiating lineage encode two laterally acquired type IV secretion systems, suggesting that these systems have a role in host adaptability.

Several severe bacterial pathogens, such as the agents of plague, typhoid fever and whooping cough, have evolved from rather benign progenitors by a process of genome degradation, which was facilitated by changes in the human population structure over the last 1,500–20,000 years<sup>1–5</sup>. Little is known, however, about how prototypic pathogens that cause massive host damage through adoption of new infection strategies may evolve into host-adapted pathogens with attenuated virulence. This lack of knowledge might primarily be due to the extinction of prototypical pathogens as result of competition with their epidemiologically more successful host-adapted descendants. However, an exception is found in the genus *Bartonella*, which, besides 20 host-adapted pathogens of low virulence potential, encompasses *B. bacilliformis* as the ‘missing link’<sup>6,7</sup>. With a restricted distribution in the Andes region, this ‘living fossil’ is a severe human pathogen causing up to 85% mortality<sup>8</sup>. In contrast, the other bartonellae commonly cause benign but widespread infections in their specific mammalian reservoirs: for example, *B. quintana* causes human trench fever, and *B. henselae* causes asymptomatic feline infections<sup>7</sup>. Phylogenetic analysis indicate an isolated position of *B. bacilliformis* as sole representative of a deep-branching ancestral lineage, whereas the other bartonellae evolved more recently in a lineage that radiates as a result of their adaptation to different mammalian reservoirs<sup>6,9,10</sup> (Fig. 1).

Notably, *B. bacilliformis* and the species of the radiating lineage share a similar lifestyle in the blood of their mammalian reservoirs,

characterized by a marked tropism for endothelial cells and erythrocytes that results in long-lasting intraerythrocytic infections. This unique infection strategy has been described in most detail for *B. tribocorum* infections in rats<sup>11,12</sup>. Genetic studies in this model have identified pathogenicity factors required for the infection of endothelial cells and erythrocytes—the type IV secretion systems (T4SSs) VirB and Trw, respectively<sup>13,14</sup>. T4SSs are macromolecular transporters ancestrally related to bacterial conjugation systems. Several mammalian pathogens have adopted T4SSs to translocate effector proteins into host cells<sup>15</sup>; for example, the VirB T4SS of *B. henselae* injects seven effectors into human endothelial cells to subvert cellular functions<sup>16</sup>. Whether VirB, Trw or any other pathogenicity factor<sup>11</sup> identified in *Bartonella* has contributed to the remarkable evolutionary success of the radiating lineage is unknown. However, during separation from the ancient *B. bacilliformis* lineage, the last common ancestor (LCA) of the radiating lineage should have acquired pathogenicity factors facilitating host adaptation and virulence attenuation. These pathogenicity factors have likely adopted essential functions during infection and should thus be encoded by the core genome of contemporary species of the radiating lineage. On the basis of these two assumptions, we carried out an integrated genomics approach to infer pathogenicity factors critical for host adaptability and virulence attenuation within the radiating lineage. We first determined a consolidated set of pathogenicity genes for the model

<sup>1</sup>Focal Area Infection Biology, Biozentrum, University of Basel, CH-4056 Basel, Switzerland. <sup>2</sup>Max Planck Institute for Developmental Biology, D-72076 Tuebingen, Germany. <sup>3</sup>Unité de Biologie Moléculaire et Immunologie Parasitaire et Fongique, Institut Scientifique de Recherche Agronomique, Ecole Nationale Veterinaire d'Alfort, F-94700 Maisons-Alfort, France. <sup>4</sup>Department of Veterinary Pathology, Faculty of Veterinary Science, University of Liverpool, Liverpool L69 7LB, UK. <sup>5</sup>Center for Infectious Disease Dynamics and Center for Comparative Genomics and Bioinformatics, Penn State University, University Park, Pennsylvania 16802, USA. <sup>6</sup>Present address: Magnetic Resonance Center, Max Planck Institute for Biological Cybernetics, D-72076 Tuebingen, Germany. <sup>7</sup>These authors contributed equally to this work. Correspondence should be addressed to C.D. (christoph.dehio@unibas.ch) or S.C.S. (scs@bx.psu.edu).

Received 11 April; accepted 15 September; published online 25 November 2007; doi:10.1038/ng.2007.38

pathogen *B. tribocorum*, that is, the set of genes required for causing intraerythrocytic bacteremia in the rat model<sup>12</sup>. To this end, we determined the complete genome sequence of *B. tribocorum* and carried out a genome-wide mutant screen for loss of infectivity in rats. Pathogenicity genes of *B. tribocorum* present in the core genome of the three sequenced species of the radiating lineage (*B. tribocorum*, *B. henselae*<sup>17</sup> and *B. quintana*<sup>17</sup>) were then tested for their presence or absence in the genome sequence of *B. bacilliformis* recently released from The Institute of Genomic Research (see Methods). Pathogenicity genes absent from *B. bacilliformis* were subsequently analyzed for their presence in as yet unsequenced species of the radiating lineage. By this approach, we show here that the acquisition of VirB-like T4SSs by lateral gene transfer (LGT) is the only trackable change in pathogenicity gene composition that seems critical for the remarkable host adaptation and virulence attenuation characteristic of species of the radiating *Bartonella* lineage.

## RESULTS

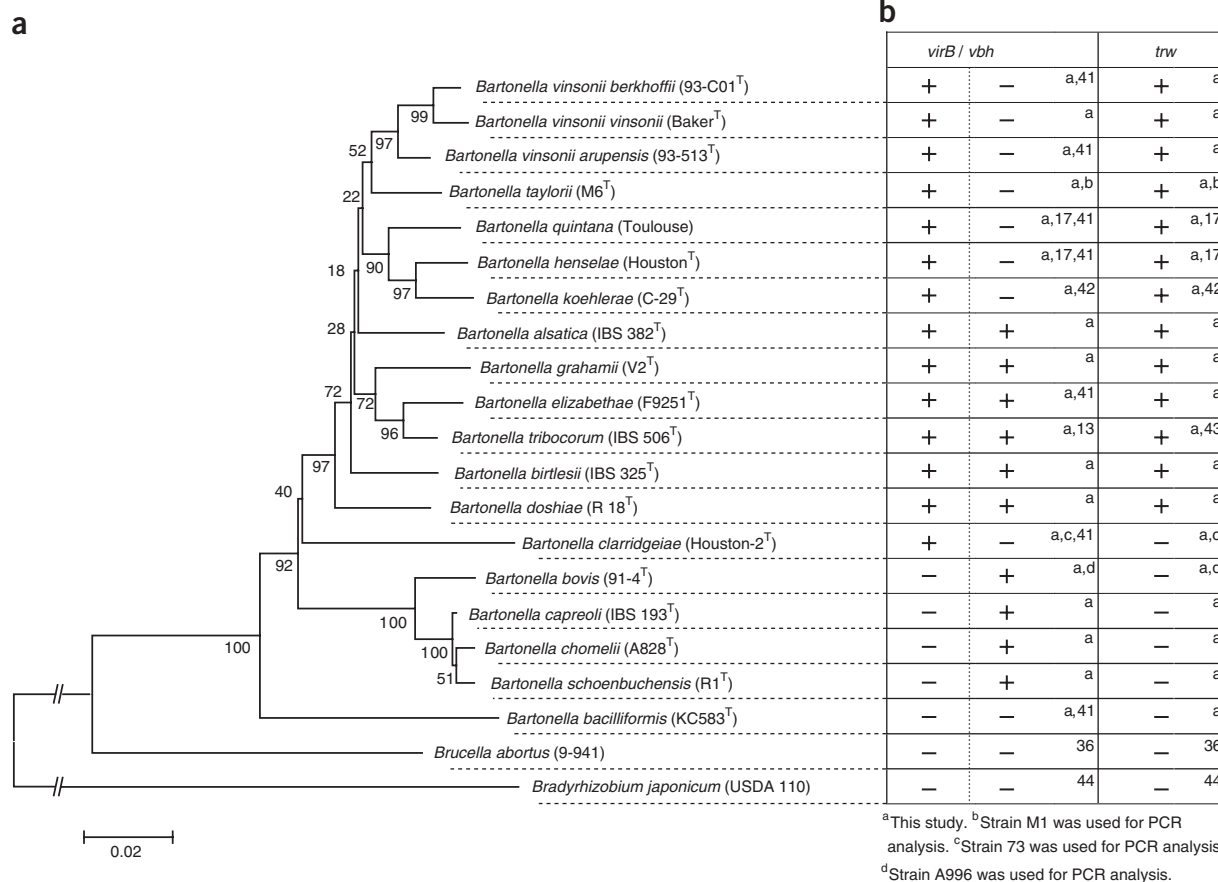
### Genome sequence of the model pathogen *B. tribocorum*

We sequenced the genome of *B. tribocorum* strain 506<sup>T</sup> by the Sanger method at 11-fold coverage using the shotgun approach<sup>18</sup> and closed the remaining gaps by sequencing of PCR fragments or directly on

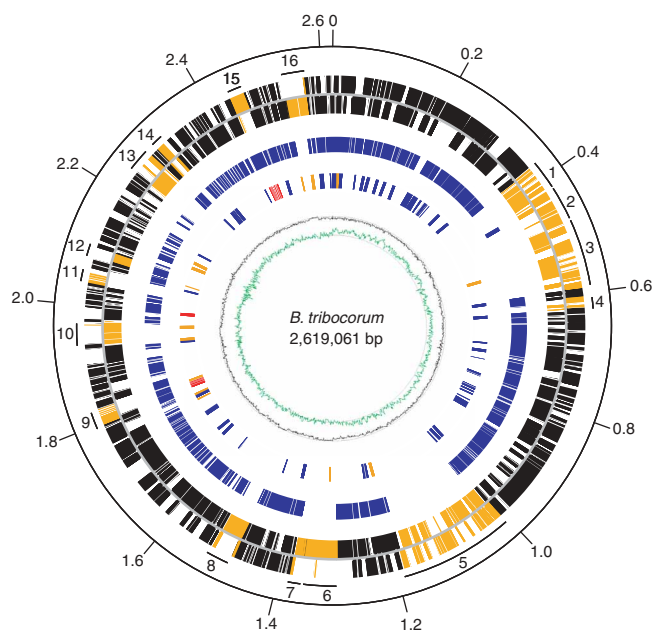
chromosomal DNA. The genome consists of a circular chromosome of 2,619,061 bp (Fig. 2) and a plasmid of 23,343 bp. Compared to the genomes of *B. henselae*<sup>17</sup> (1.93 Mb) and *B. quintana*<sup>17</sup> (1.58 Mb), the *B. tribocorum* genome is 36% and 66% larger, respectively (Table 1), which is reflected by a 45% and 89% greater number of protein-coding genes. Despite these differences, the three genomes compared have a similar coding density (Table 1) and a high degree of collinearity (Supplementary Fig. 1 online).

### Core genome of the radiating lineage

We determined the set of orthologous genes shared by *B. tribocorum*, *B. henselae* and *B. quintana*, and we refer to this here as the 'core genome' of the radiating lineage. Ortholog mapping was carried out by reciprocal whole-genome BLAST searches of the translated ORFs<sup>19</sup>. To take gene identity, similarity and size into account, we normalized BLAST scores by dividing them by the BLAST score of each query ORF against itself. From the distribution of these normalized BLAST scores<sup>20</sup> in each pair of compared genomes (Supplementary Fig. 1), we defined an arbitrary BLAST score threshold of 0.7, above which genes were considered orthologous, comparable to 70% identity over the entire gene length. Genes with a lower normalized BLAST score were also considered orthologous if they were flanked by orthologous



**Figure 1** Phylogenetic tree of *Bartonella* based on multilocus sequence analysis and summary table of the presence or absence of loci encoding the T4SSs VirB, VirB-homolog (Vbh) and Trw in the different *Bartonella* species. (a) The phylogenetic tree shown was calculated on the basis of concatenated alignments of protein sequences of *rpoB*, *groEL*, *ribC* and *gltA* using the Kimura algorithm as the distance method and neighbor joining as the tree-construction method. The organism names and the strain designations are shown (for details and reference see Supplementary Table 5). The value of 0.02 on the scale bar indicates 1 amino acid substitution per 50 sites. Numbers at the nodes of the tree indicate bootstrap values (1,000 replicates). (b) The table indicates the presence or absence of the T4SS loci *virB*, *vbh* and *trw*. References are given in the upper right of each box.



**Figure 2** Circular genome map of *B. tribocorum*. From the outside, the first two circles indicate genes on the + and – strands. Genomic islands are highlighted in orange and numbered according to **Supplementary Table 2**. The third circle (blue) shows the core genome of the radiating lineage as determined in **Figure 3**. The fourth circle shows the pathogenicity genes encoded by *B. tribocorum* as presented in detail in **Supplementary Table 3**. The pathogenicity genes falling into the core genome of the radiating lineage are highlighted in blue if present in *B. bacilliformis* and in red if absent from *B. bacilliformis*. The pathogenicity genes of the accessory genome of *B. tribocorum* are shown in orange. The two inner circles display the genomic island content (black) and the genomic island skew (green).

genes showing gene order conservation (microsynteny). From this analysis, the core genome of *B. tribocorum*, *B. henselae* and *B. quintana* comprises 1,097 genes (**Fig. 3** and **Supplementary Table 1** online). The accessory genome composed of the nonorthologous genes includes 1,057 genes in *B. tribocorum* (96% of core genome), 391 genes in *B. henselae* (36% of core) and 45 genes in *B. quintana* (4% of core) (**Fig. 3**). Within the *B. tribocorum* genome, these nonorthologous genes encompass a high proportion of horizontally acquired phages and other genomic islands (highlighted in orange in the outer circle of **Fig. 2**; see also **Supplementary Table 2** online). Together with the large number of cryptic phage integrases and other genomic island remnants in *B. henselae* and *B. quintana*<sup>17</sup>, these data suggest a functional role for LGT in genome evolution within the radiating lineage.

### Search for pathogenicity genes in *B. tribocorum*

To identify a comprehensive set of genes associated with pathogenicity, we adapted signature-tagged mutagenesis (STM)<sup>21</sup> to *B. tribocorum*. Individually tagged mutants assembled in pools were tested in the rat-infection model<sup>12</sup> for the capacity to cause intraerythrocytic bacteremia, the hallmark of *Bartonella* infection in the mammalian reservoir. Each pool of 36 mutants was inoculated in two rats (input pool), and bacteremic blood was cultivated on days 7 and 14 after infection (output pools). Among 3,084 mutants screened, 359 were present in the input pool but absent from the output pools. These abacteremic mutant candidates were retested in newly assembled mutant pools inoculated in at least four rats. Finally, a total of 130

**Table 1** General genomic features of *B. tribocorum* in comparison with *B. henselae* and *B. quintana*

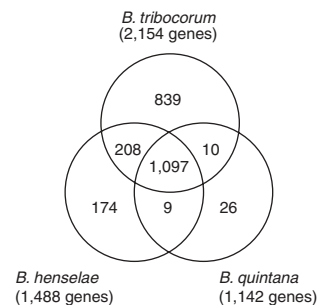
	<i>B. tribocorum</i> <sup>a</sup>	<i>B. henselae</i>	<i>B. quintana</i>
Chromosome size (bp)	2,619,061	1,931,047	1,581,384
G+C content	38.8% (35.0%)	38.2%	38.8%
Total number of PCG	2,154 (18)	1,488	1,142
Average length of PCG (bp)	906	942	999
Integrase remnants	47 (0)	43	4
Number of rRNA operons	2 (0)	2	2
Number of tRNA genes	42 (0)	44	44
Percentage coding	71.6% (70.0%)	72.3%	72.7%
Plasmids	1 (23,343 bp)	0	0

<sup>a</sup>Numbers in parentheses refer to the plasmid. PCG, protein-coding genes.

mutants showed a consistently abacteremic phenotype, accounting for 4% of the 3,084 mutants screened. This result is in the range of avirulent mutant frequencies reported for STM studies in other bacteria<sup>22</sup>. The *in vitro* growth of all abacteremic mutants was comparable to wild-type bacteria, indicating that the abacteremic phenotypes were not due to general growth defects. We determined the transposon-insertion sites of the 130 abacteremic mutants by direct genomic sequencing. Thereby, 113 mutants were shown to carry insertions inside a gene, and 15 out of 17 intergenic insertions occurred maximally 350 bp upstream of a gene. Taken together, 97 different protein-coding genes were found to be essential for colonization of the mammalian host and thus were termed pathogenicity genes (**Supplementary Table 3** online).

### Pathogenicity genes specific to the radiating lineage

From the 97 pathogenicity genes identified in *B. tribocorum*, 81 are encoded by the core genome of *B. tribocorum*, *B. henselae* and *B. quintana* (**Table 2**). Despite the similar size of the core and accessory genomes in *B. tribocorum*, the core genome of the radiating lineage thus includes the majority of the identified pathogenicity genes. From the 81 pathogenicity genes encoded by this core genome of the radiating lineage, 15 are absent from *B. bacilliformis* (**Table 2** and **Supplementary Table 4** online). Fourteen of these 15 pathogenicity genes are part of the T4SS loci *virB* and *trw* (**Fig. 4**). These T4SS loci have been acquired by LGT and are thus referred to as genomic islands or pathogenicity islands (**Supplementary Table 2**)<sup>14,16</sup>. The remaining one pathogenicity gene shared among *B. tribocorum*, *B. henselae* and *B. quintana* but absent from *B. bacilliformis* encodes the membrane protein BT1873, which is conserved in *Brucella*



**Figure 3** Core genome and accessory genomes of three species of the radiating *Bartonella* lineage determined on the basis of ortholog gene sets (**Supplementary Table 1**). Numbers in parentheses indicate the total number of protein-coding genes.

**Table 2** Presence of pathogenicity genes identified by STM in *B. tribocorum*, and their presence within the radiating lineage and *B. bacilliformis*

Category	Total genes	Pathogenicity genes
<i>B. tribocorum</i> genome	2,154	97
Core genome of the radiating lineage:	1,097	81
with homolog in <i>B. bacilliformis</i>	959	66
without homolog in <i>B. bacilliformis</i>	138	15

Data for the core genome of the radiating lineage are derived from a comparison of the three available genomes of *B. henselae*, *B. quintana* and *B. tribocorum*.

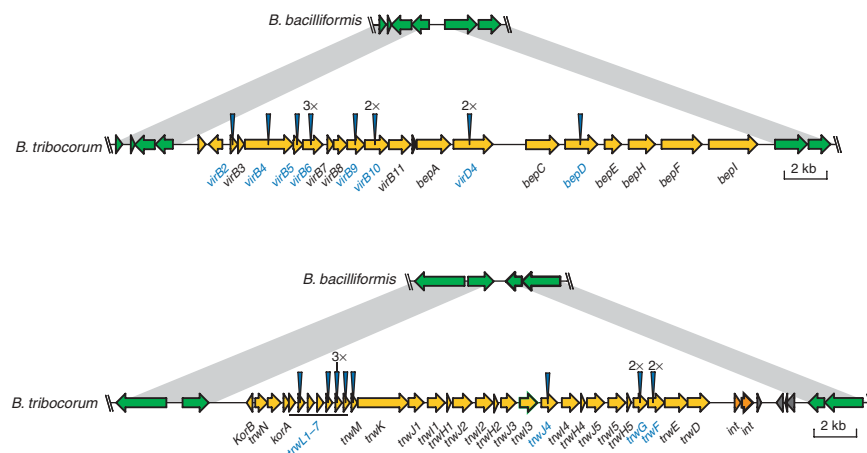
(Supplementary Table 4) and other bacteria of the  $\alpha$ -proteobacterial lineage. Comparison of corresponding chromosomal regions in *B. tribocorum* and *B. bacilliformis* indicates deterioration of a three-gene locus composed of BT1873 (deleted), BT1874 (deleted) and BT1875 (pseudogene) in *B. bacilliformis* (data not shown). BT1873 was thus lost by the *B. bacilliformis* lineage after separation from the radiating lineage.

Homologous sequences flanking the insertion sites of the *virB* and *trw* genomic islands in *B. tribocorum* are short intergenic regions without signatures of gene deletion or deterioration in *B. bacilliformis* (Fig. 4), suggesting that these T4SS loci were acquired after separation from the *B. bacilliformis* lineage rather than deleted within the *B. bacilliformis* lineage. To exclude the possibility of deletion of these genomic islands in the highly passaged *B. bacilliformis* strain KC583 used for genome sequencing, we analyzed five low-passage clinical isolates of *B. bacilliformis* (T2, Monz269, ER-Cha, LA6.3 and Cusco407)<sup>23</sup> by PCR. For all tested *B. bacilliformis* strains, we could amplify short PCR fragments spanning the flanking regions of the T4SS genomic island integration sites in *B. tribocorum*. In addition, amplification of conserved genes of both genomic islands (*virB4* and *trwK*) did not give any positive results in all tested *B. bacilliformis* strains (data not shown). Taken together, these results confirm the absence of *virB* and *trw* T4SS loci in the *B. bacilliformis* lineage and indicate acquisition of these genomic islands by LGT after separation from the *B. bacilliformis* lineage.

### Distribution of T4SS loci in the radiating lineage

Pathogenicity factors associated with host adaptation should be present in all species of the radiating lineage but absent from the ancient *B. bacilliformis* lineage. We thus examined the as yet unsequenced species of the radiating lineage for the presence of the T4SS loci *virB* and *trw*. We amplified conserved genes of these loci (*virB4* for *virB*, *trwK* for *trw*) by PCR and validated the identity of the

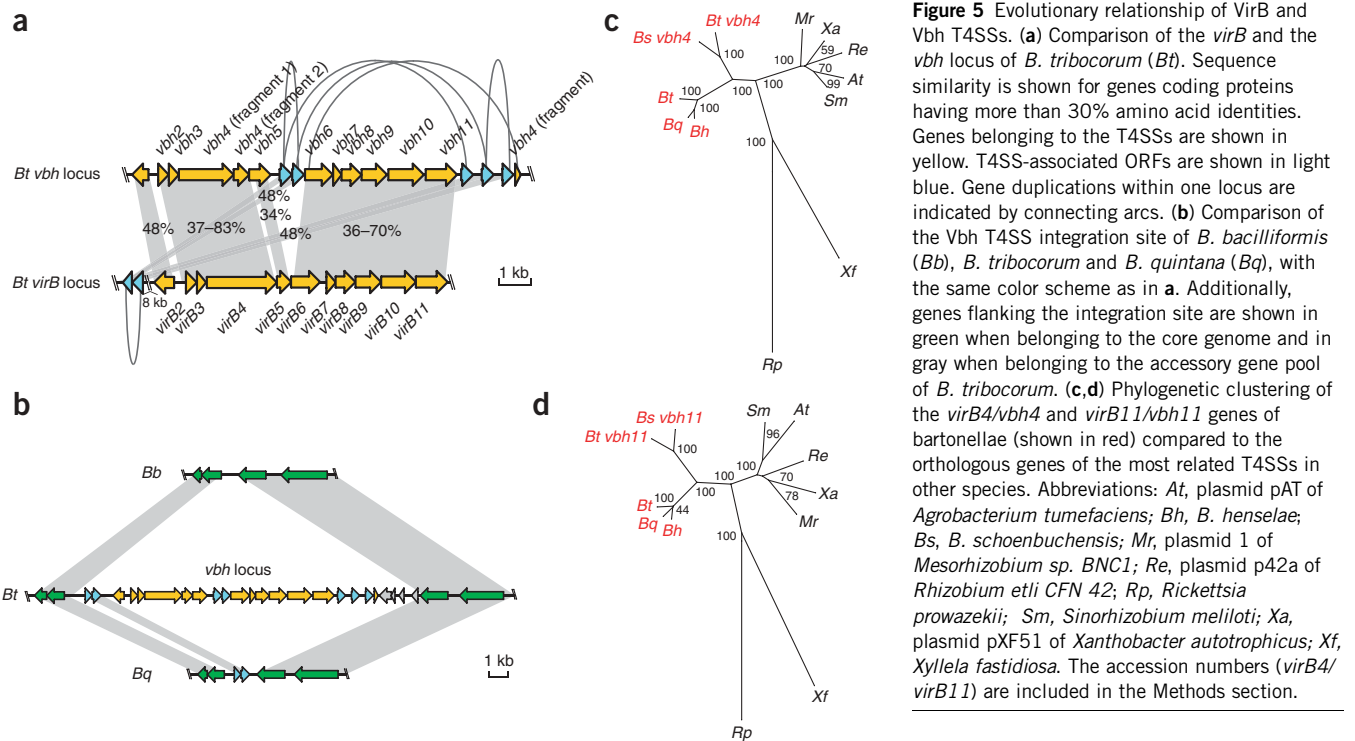
**Figure 4** STM mutant insertion sites in the T4SS loci *virB* and *trw* of *B. tribocorum* and comparison of the flanking region with *B. bacilliformis*. Genes belonging to one of the two T4SSs are shown in yellow. STM mutant insertion sites are indicated as blue triangles and the corresponding gene names colored in blue. The number of STM mutant insertion sites for each pathogenicity gene is given if more than one insertion was identified (Supplementary Table 3). The genes of the flanking regions are shown in green. Hypothetical or putative genes are shown in gray. The two integrase remnants flanking the *trw* locus are colored in orange.



obtained PCR fragments by DNA sequencing. Together with DNA blot analysis (data not shown) and literature mining, this analysis confirmed the presence of *virB* and *trw* loci in subsets of the radiating lineage. To illustrate the distribution of these T4SSs among the bartonellae in relation to their phylogeny, we inferred a phylogenetic tree by multilocus sequence analysis (MLSA)<sup>24,25</sup> (Fig. 1). Based on concatenated protein sequences of the housekeeping genes *groEL*, *gltA*, *rpoB* and *ribC*, this rooted Kimura neighbor-joining tree is highly congruent with trees calculated by other algorithms (that is, maximum evolution, maximum parsimony, and unweighted pair-group method with arithmetic mean; data not shown), and most nodes are supported by high bootstrap values. We found the *virB* locus present in the radiating lineage in all species except those infecting ruminants, including *Bartonella bovis*, *Bartonella capreoli*, *Bartonella chomelii* and *Bartonella schoenbuchensis*. The *trw* locus showed a similar distribution pattern, except that it was absent from *Bartonella clarridgeiae*, the modern species most closely related to the ruminant-specific sublineage.

The absence of both of these T4SS loci in the ruminant-specific sublineage, which diverged first within the radiating lineage (Fig. 1), may call into question whether T4SSs have a central role in host adaptation and virulence attenuation as shared characteristics of species belonging to the radiating lineage. However, annotation of the *B. tribocorum* genome revealed a third T4SS locus, which we termed *virB*-homolog (*vbh*) to reflect the high degree of sequence similarity with the *virB* locus (Fig. 5). Indeed, a phylogenetic analysis of the ATPase genes *virB4* and *vbh4* and *virB11* and *vbh11* with their respective homologs from related T4SSs showed that *vbh* and *virB* homologs cluster in one clade (Fig. 5), illustrating their close evolutionary relatedness. Notably, as indicated by PCR and DNA blot analysis, the *vbh* locus is absent from the sequenced *B. bacilliformis* strain KC583 (Fig. 5) as well as from five low-passage *B. bacilliformis* isolates (T2, Monz269, ER-Cha, LA6.3 and Cusco407; data not shown)<sup>23</sup>, whereas it is present in the ruminant-specific sublineage and several other species of the radiating lineage (Fig. 1). The radiating lineage thus acquired the *vbh* locus during or soon after separation from the ancient *B. bacilliformis* lineage. The sublineage infecting nonruminants, which encodes a functional VirB T4SS in every species, has lost the Vbh T4SS in *B. clarridgeiae* as well as the phylogenetic subclade comprising *B. henselae*, *B. quintana*, *Bartonella koehlerae*, *Bartonella taylorii* and *Bartonella vinsonii* subspecies.

Closer inspection of the *vbh* and *virB* genomic islands showed marked similarities beyond the encoded T4SS components: for example, a conserved nuclease gene encoded upstream of *vbh2* and

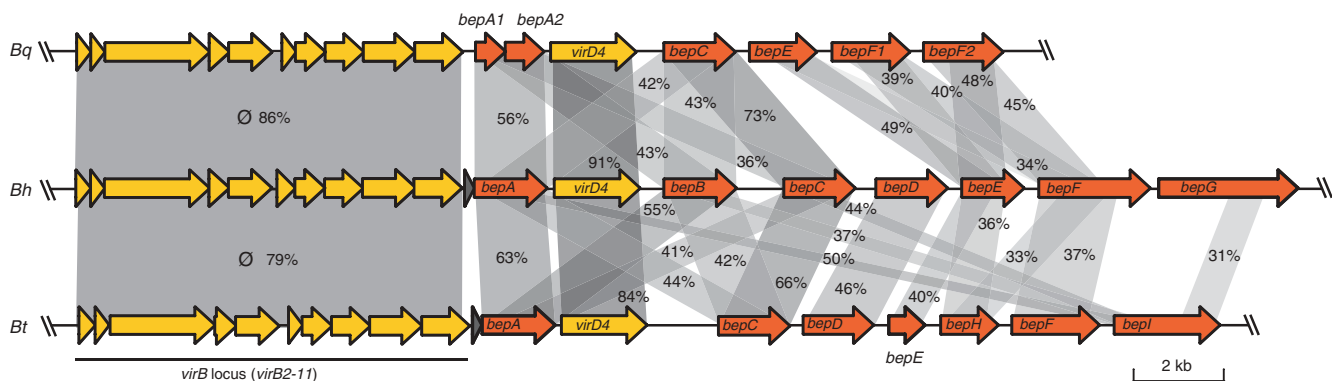


**Figure 5** Evolutionary relationship of VirB and Vbh T4SSs. **(a)** Comparison of the *virB* and the *vbh* locus of *B. tribocorum* (*Bt*). Sequence similarity is shown for genes coding proteins having more than 30% amino acid identities. Genes belonging to the T4SSs are shown in yellow. T4SS-associated ORFs are shown in light blue. Gene duplications within one locus are indicated by connecting arcs. **(b)** Comparison of the Vbh T4SS integration site of *B. bacilliformis* (*Bb*), *B. tribocorum* and *B. quintana* (*Bq*), with the same color scheme as in **a**. Additionally, genes flanking the integration site are shown in green when belonging to the core genome and in gray when belonging to the accessory gene pool of *B. tribocorum*. **(c,d)** Phylogenetic clustering of the *virB4/vbh4* and *virB11/vbh11* genes of bartonellae (shown in red) compared to the orthologous genes of the most related T4SSs in other species. Abbreviations: *At*, plasmid pAT of *Agrobacterium tumefaciens*; *Bh*, *B. henselae*; *Bs*, *B. schoenbuchensis*; *Mr*, plasmid 1 of *Mesorhizobium* sp. BNC1; *Re*, plasmid p42a of *Rhizobium etli* CFN 42; *Rp*, *Rickettsia prowazekii*; *Sm*, *Sinorhizobium meliloti*; *Xa*, plasmid pXF51 of *Xanthobacter autotrophicus*; *Xf*, *Xylella fastidiosa*. The accession numbers (*virB4/virB11*) are included in the Methods section.

*virB2* that has not been associated with any other T4SS locus (Fig. 5). These findings suggest that *virB* might have been derived from duplication of *vbh* during an early stage of evolution of the radiating lineage. Alternatively, the *virB* genomic island might have been acquired by LGT from a similar external source as the *vbh* genomic island, that is, a conjugative plasmid. Subsequent to the duplication or LGT event that gave rise to *virB*, the *vbh* locus deteriorated in at least some species of the non-ruminant-specific sublineage. This is indicated by frameshift mutations in the *vbh4* gene of *B. tribocorum*, and by an almost complete reduction of the *vbh* genomic island in *B. quintana* and *B. henselae* (Fig. 5). In contrast, a preliminary analysis of the *vbh* locus of *B. schoenbuchensis* did not show any deleterious mutation and, moreover, it indicated that this locus encodes translocatable T4SS substrates (P.E. and C.D., unpublished observations), suggesting that Vbh represents a functional substrate-translocating T4SS in the ruminant-specific sublineage. These findings imply that Vbh and VirB are functionally and evolutionarily related, and that

since the acquisition of VirB, this T4SS has functionally replaced the previously acquired Vbh. As an alternative scenario to a sequential acquisition of VirB-like T4SSs, both Vbh and VirB might have coexisted as functional T4SS in the LCA of the radiating lineage. In this case, VirB would have been reduced in the ruminant-specific sublineage, whereas Vbh would have been under deterioration in the non-ruminant-specific sublineage.

Taken together, these data suggest that two VirB-like T4SSs were acquired by the radiating lineage after separation from the ancient *B. bacilliformis* lineage. These functionally redundant T4SSs may have evolved differently in the ruminant-specific and non-ruminant-specific sublineages, preserving at least one functional VirB-like T4SS (Vbh and/or VirB) in every species. In contrast, the distantly related Trw T4SS seems to have been acquired by the radiating lineage after separation from the sublineage composed of the ruminant-specific species and *B. clarridgeiae*. The VirB-like T4SS thus represent the only pathogenicity factors analyzed in this study that match our criteria for



**Figure 6** Comparison of the *virB* T4SS loci of *B. henselae*, *B. quintana* and *B. tribocorum*. Sequence similarity is shown for genes having more than 30% amino acid identities over a sequence stretch of more than 250 amino acids. VirB T4SS genes are shown in yellow and VirB T4SS effector genes in red.

mediating host adaptability and virulence attenuation, although we cannot exclude the possibility that other pathogenicity factors (such as Trw) or even factors without direct roles in pathogenicity may further contribute to these adaptive traits.

### Fast evolution of effectors translocated by the VirB T4SS might reflect host adaptation

The *virB* loci of *B. henselae*, *B. tribocorum* and *B. quintana* show a degree of sequence conservation in the range of other core genome-encoded genes. In contrast, the genes encoding VirB-translocated effectors (*bepA-I*) are much less conserved and show multiple incidences of gene duplication, conversion and deletion (Fig. 6). This genetic plasticity suggests that the *bep* genes evolved faster than most other genes, probably as result of positive selection during radial speciation. The diversity in the translocated effectors may therefore reflect the specific host adaptations and concomitant virulence attenuation by species of the radiating lineage in their respective mammalian reservoirs.

### DISCUSSION

Bacterial conjugation systems are known to play a dual role in bacterial infection. First, they facilitate the spread within the bacterial population of mobile DNA elements, which may encode pathogenicity factors, antibiotic resistances or other factors that increase bacterial fitness during infection. Second, the conjugation systems themselves may adopt prominent functions in the infection process; in particular, they may act to facilitate the translocation of macromolecular pathogenicity factors into host cells. Such adapted conjugation systems—better known as T4SSs—have been described as essential pathogenicity factors in several genetically amenable pathogens of mammals (including *Helicobacter pylori*, *Legionella pneumophila*, *Brucella* spp. and *Bartonella* spp.) and plants (for example, *Agrobacterium tumefaciens*)<sup>15</sup>. Moreover, data from genome sequencing projects have shown that many obligate intracellular pathogens of mammals (for example, *Rickettsia* spp., *Coxiella* spp., *Anaplasma marginale* and *Ehrlichia* spp.) and also mutualistic endosymbionts of insects (for example, *Wolbachia* spp.)<sup>26</sup> and amoebae (for example, environmental *Chlamydia* spp.)<sup>27</sup> encode T4SSs. Although genetic evidence for the role of T4SSs in the biology of these obligate intracellular bacteria has been lacking, the conservation of T4SS loci in their often highly reduced genomes suggests functional roles of T4SSs in establishing intracellular replication niches. The T4SSs of pathogenic as well as mutualistic bacteria are thus considered to facilitate host cell interactions and in particular to mediate the translocation of macromolecular effectors such as proteins or DNA-protein complexes into host cells<sup>15</sup>.

The comparative and functional genomic analysis of host adaptability in the genus *Bartonella* reported here allowed us to propose an additional role for T4SSs in the evolution of host-associated bacteria. We provide evidence that the acquisition of VirB-like T4SSs within the radiating *Bartonella* lineage was associated with increased host adaptability, as manifested by virulence attenuation in a given host and increased adaptation to new hosts as compared to the ancestrally related, highly pathogenic species *B. bacilliformis*. As our approach was limited to the analysis of pathogenicity factors that are conserved in the radiating *Bartonella* lineage but absent from *B. bacilliformis*, we cannot exclude the possibility that additional bacterial factors may contribute to host adaptability, including nonconserved pathogenicity factors or factors without a primary role in pathogenesis. Notably, the VirB-translocated effector proteins of *Bartonella* (Beps) show an atypically high degree of sequence variation among different species, suggesting an increased rate of evolution as the result of positive

selection for adaptive functions in the infected host. VirB-translocated effectors are known to subvert multiple physiological functions within their target host cells, such as actin dynamics, innate immune responses and apoptosis<sup>16,28</sup>. Additional studies will be required to examine how these physiological changes may contribute to T4SS-dependent host adaptation in the specific mammalian reservoirs of the diverse bartonellae. It is also presently unknown whether the role of VirB-like T4SSs in host adaptability inferred here for bartonellae might be a shared feature of effector-translocating T4SSs in other host-associated bacteria.

### METHODS

**Bacterial strains and growth conditions.** *Bartonella* and *Escherichia coli* strains were grown as described previously<sup>13</sup>. **Supplementary Table 5** online lists all bacterial strains and plasmids used in this study.

**Genome sequencing.** Genomic DNA from *B. tribocorum* IBS 506<sup>T</sup> (ref. 29) was isolated using the QIAGEN Genomic DNA Isolation kit (Qiagen). The following DNA libraries were constructed and end-sequenced to 11-fold sequence coverage: one library of 3–5 kb (TOPO Shotgun subcloning kit, Invitrogen), three libraries of 35–43 kb (Epicentre Technologies) and one library of 100–180 kb (Bio S&T). Remaining gaps were closed by direct sequencing on genomic DNA or by PCR amplification and subsequent sequencing. We calculated the final sequencing error rate to be  $0.014 \times 10^{-5}$  using the PHRED/PHRAP/CONSED software package<sup>30–32</sup>.

**Annotation and genome analysis.** For the automated annotation and the manual curation, we used the annotation package GENDB<sup>33</sup> as described in the **Supplementary Methods** online. For the comparative genome analysis, we used the program MUMmer<sup>34</sup> and the Artemis comparison tool (ACT)<sup>35</sup>. We determined the orthologous genes by carrying out reciprocal best-match BLAST comparisons followed by manual curation. Comparisons with the genomes of *B. bacilliformis* (CP000524) and *Brucella abortus*<sup>36</sup> were done by using BLASTP (see **Supplementary Methods**).

**Construction of phylogenetic trees.** For the phylogenetic tree of bartonellae, a multilocus sequence analysis (MLSA) approach<sup>24,25</sup> was used by aligning five different housekeeping genes (*groEL*, *ribC*, *rpoB* and *gltA*) as described in **Supplementary Methods**. The accession numbers of the sequences of the different species are given in **Supplementary Table 6** online. All protein sequences were aligned with CLUSTAL W software version 1.82<sup>37</sup>, and overhanging ends were cut and phylogenetic trees calculated using different algorithms included in the MEGA 3.1 software<sup>38</sup>. The trees of the *virB* and *vbh4* genes and *virB* and *vbh11* genes were built by using the same software tools as before (see **Supplementary Methods**).

**Transposon vector and signature-tag construction.** The suicide transposon vector pHS006 contains an origin of transfer for conjugative transfer, the *Himar1* transposon, carrying a kanamycin resistance marker and a hyperactive transposase<sup>39,40</sup>. Construction details are described in **Supplementary Methods**. To construct the signature tags, we generated a pool of degenerated single-stranded 120-bp DNA molecules (STM oligo) containing a central stretch of 50 random base pairs ([NK]<sub>25</sub>) flanked by two constant sequences by oligonucleotide synthesis (Microsynth). After amplification by PCR they were introduced into pHS006 (**Supplementary Methods**).

**STM library construction.** A total of 42 individually tagged transposon vectors were separately transferred to *B. tribocorum* by two-parental mating as described previously<sup>12</sup>. Thirty-six tags were chosen on the basis of reproducible detection (**Supplementary Methods**) and used to produce kanamycin-resistant *B. tribocorum* transconjugants. From each mating, 96 single kanamycin-resistant *B. tribocorum* colonies were transferred to a 96-well plate and labeled with the tag number for storage at  $-70^{\circ}\text{C}$ .

**Animal infections.** Infections were done using the *B. tribocorum* rat-infection model described previously<sup>12</sup>. We carried out animal care and ensured animal well-being in accordance with the Swiss Act on Animal Protection and Good

Animal Care Practice. For infection, 36 differently tagged mutants were grown separately from the transposon library for each input pool. The same amounts of each mutant were pooled in PBS directly before infection and used to infect two rats with  $10^9$  bacteria (0.3 ml of culture with  $OD_{595} = 1$ ) each intravenously in the tail vein. An aliquot of the input pool was used as template for the detection PCR. Blood was taken from the tail veins of the infected rats after 7 and 14 d after infection, serially diluted in PBS and plated on Columbia blood agar (CBA) plates. Grown bacterial colonies (the output pool) were counted, harvested in PBS and used as template for the detection PCR. Rescreening was done following the same protocol using four rats per input pool.

**PCR detection of STM mutants.** For each input and output pool, 36 tag-specific PCR reactions were done. For PCR, we used tag-specific primers together with a generic primer (**Supplementary Table 5**) to yield a product of approximately 600 bp (**Supplementary Methods**).

**Identification and analysis of transposon insertion sites.** Genomic DNA from single mutants, regrown from the mutant library, was isolated and used for sequencing with transposon-specific primers (**Supplementary Table 5**).

**PCR screening for T4SS loci.** Genomic DNA of the species listed in **Supplementary Table 5** was used as template for PCR. By aligning the genome sequences of *B. tribocorum*, *B. henselae* and *B. quintana*, we designed primers on conserved genes of the VirB and Trw T4SSs. We tested several primer combinations for each species (**Supplementary Methods**). The absence of the *virB*, *trw* and *vbh* T4SS loci in different *B. bacilliformis* strains (KC583, T2, Monz269, ER-Cha, LA6.3 and Cusco407) was shown by amplifying the chromosomal integration sites of the three T4SSs. In addition, the PCR amplification of conserved genes of the VirB, Vbh and Trw T4SSs was tested for the different *B. bacilliformis* strains (**Supplementary Methods**).

**DNA blot analysis.** To further show the presence or absence of the T4SSs (*VirB*, *Trw* and *Vbh*) in different *Bartonella* species (*Bartonella alsatica*, *B. bacilliformis* KC583, *B. clarridgeiae*, *B. henselae*, *B. quintana*, *B. tribocorum*, *B. vinsonii berkhoffii*, *B. vinsonii arupensis*, *Bartonella birtlesii* and *B. koehlerae*), we carried out DNA blot analysis using the Digoxigenin hybridization system (Roche; see **Supplementary Methods**).

**Accession codes.** EMBL Nucleotide Sequence Database: data have been deposited with accession codes AM260525 and AM260524 (*B. tribocorum* genome and plasmid sequence, respectively), AM690314 and AM690315 (*ribC* and *groEL* of *B. birtlesii*), AM690317 and AM690316 (*ribC* and *groEL* of *B. chomelii*), and AM420307 and AM420308 (sequences of the *vbh4* and *vbh11* genes of *B. schoenbuchensis*, respectively). NCBI accession codes were as follows: sequence of *B. bacilliformis* recently released from The Institute of Genomic Research (CP000524), accession numbers (*virB4/virB11*) for plasmid pAT of *Agrobacterium tumefaciens* (NP\_396095/NP\_396102); *B. henselae* (YP\_034053/YP\_034060); *B. quintana* (YP\_032622/YP\_032629); *vbh* of *B. schoenbuchensis* (AM420307/420308); *virB* of *B. tribocorum* (genes BT1691/BT1698 of AM260525); *vbh* of *B. tribocorum* (genes BT2334/BT2345 of AM260525); plasmid 1 of *Mesorhizobium* sp. *BNC1* (YP\_665963/YP\_665955); *Rickettsia prowazekii* (NP\_220495/NP\_220676); plasmid p42a of *Rhizobium etli* *CFN 42* (YP\_471641/NP\_659884); plasmid pSymA of *Sinorhizobium meliloti* (NP\_435962/NP\_435955); plasmid pXF51 of *Xanthobacter autotrophicus* (ZP\_01200724/ ZP\_01200716); *Xyllella fastidiosa* (NP\_061663/NP\_061671).

Note: Supplementary information is available on the Nature Genetics website.

#### ACKNOWLEDGMENTS

We thank E.J. Rubin for providing plasmids containing the *Himar1* transposon and transposase. We acknowledge the use of the MiGenAS system and excellent support by MiGenAS team members, especially M. Rampp. We thank G. Schroeder for critical reading of the manuscript. This work was supported by grant 3100A0-109925/1 from the Swiss National Science Foundation (C.D.), by grant 55005501 from the Howard Hughes Medical Institute (C.D.), by a generous donation from the Freiwillige Akademische Gesellschaft Basel (C.D.) and by the Center for Systems Bacterial Infection (C-SBI) of SystemsX, the Swiss Initiative in Systems Biology.

#### AUTHOR CONTRIBUTIONS

C.D., H.L.S., P.E. and S.C.S. designed the research; H.L.S., P.E., M.C.S., M.V.-T. and C.L. performed research; R.B. contributed analytical tools and materials; H.L.S., P.E., G.R., S.C.S. and C.D. analyzed data; and H.L.S., P.E. and C.D. wrote the paper.

Published online at <http://www.nature.com/naturegenetics>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

1. Wren, B.W. The yersiniae—a model genus to study the rapid evolution of bacterial pathogens. *Nat. Rev. Microbiol.* **1**, 55–64 (2003).
2. Parkhill, J. *et al.* Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523–527 (2001).
3. McClelland, M. *et al.* Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of *Salmonella enterica* that cause typhoid. *Nat. Genet.* **36**, 1268–1274 (2004).
4. Parkhill, J. *et al.* Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* **35**, 32–40 (2003).
5. Preston, A., Parkhill, J. & Maskell, D.J. The bordetellae: lessons from genomics. *Nat. Rev. Microbiol.* **2**, 379–390 (2004).
6. Birtles, R.J. & Raoult, D. Comparison of partial citrate synthase gene (*gltA*) sequences for phylogenetic analysis of *Bartonella* species. *Int. J. Syst. Bacteriol.* **46**, 891–897 (1996).
7. Jacomo, V., Kelly, P.J. & Raoult, D. Natural history of *Bartonella* infections (an exception to Koch's postulate). *Clin. Diagn. Lab. Immunol.* **9**, 8–18 (2002).
8. Ihler, G.M. *Bartonella bacilliformis*: dangerous pathogen slowly emerging from deep background. *FEMS Microbiol. Lett.* **144**, 1–11 (1996).
9. Houpiqian, P. & Raoult, D. Molecular phylogeny of the genus *Bartonella*: what is the current knowledge? *FEMS Microbiol. Lett.* **200**, 1–7 (2001).
10. Marston, E.L., Sumner, J.W. & Regnery, R.L. Evaluation of intraspecies genetic variation within the 60 kDa heat-shock protein gene (*groEL*) of *Bartonella* species. *Int. J. Syst. Bacteriol.* **49**, 1015–1023 (1999).
11. Dehio, C. Molecular and cellular basis of *Bartonella* pathogenesis. *Annu. Rev. Microbiol.* **58**, 365–390 (2004).
12. Schulein, R. *et al.* Invasion and persistent intracellular colonization of erythrocytes. A unique parasitic strategy of the emerging pathogen *Bartonella*. *J. Exp. Med.* **193**, 1077–1086 (2001).
13. Schulein, R. & Dehio, C. The VirB/VirD4 type IV secretion system of *Bartonella* is essential for establishing intraerythrocytic infection. *Mol. Microbiol.* **46**, 1053–1067 (2002).
14. Seubert, A., Hiestand, R., de la Cruz, F. & Dehio, C. A bacterial conjugation machinery recruited for pathogenesis. *Mol. Microbiol.* **49**, 1253–1266 (2003).
15. Cascales, E. & Christie, P.J. The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.* **1**, 137–149 (2003).
16. Schulein, R. *et al.* A bipartite signal mediates the transfer of type IV secretion substrates of *Bartonella henselae* into human cells. *Proc. Natl. Acad. Sci. USA* **102**, 856–861 (2005).
17. Alsmark, C.M. *et al.* The louse-borne human pathogen *Bartonella quintana* is a genomic derivative of the zoonotic agent *Bartonella henselae*. *Proc. Natl. Acad. Sci. USA* **101**, 9716–9721 (2004).
18. Fleischmann, R.D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
19. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
20. Eppinger, M. *et al.* Who ate whom? Adaptive *Helicobacter* genomic changes that accompanied a host jump from early humans to large felines. *PLoS Genet.* **2**, 1097–1110 (2006).
21. Hensel, M. *et al.* Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**, 400–403 (1995).
22. Saenz, H.L. & Dehio, C. Signature-tagged mutagenesis: technical advances in a negative selection method for virulence gene identification. *Curr. Opin. Microbiol.* **8**, 612–619 (2005).
23. Birtles, R.J. *et al.* Identification of *Bartonella bacilliformis* genotypes and their relevance to epidemiological investigations of human bartonellosis. *J. Clin. Microbiol.* **40**, 3606–3612 (2002).
24. Maiden, M.C. Multilocus sequence typing of bacteria. *Annu. Rev. Microbiol.* **60**, 561–588 (2006).
25. Maiden, M.C. *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* **95**, 3140–3145 (1998).
26. Backert, S. & Meyer, T.F. Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr. Opin. Microbiol.* **9**, 207–217 (2006).
27. Horn, M. *et al.* Illuminating the evolutionary history of chlamydiae. *Science* **304**, 728–730 (2004).
28. Schmid, M.C. *et al.* A translocated bacterial protein protects vascular endothelial cells from apoptosis. *PLoS Pathog.* **2**, e115 (2006).
29. Heller, R. *et al.* *Bartonella tribocorum* sp. nov., a new *Bartonella* species isolated from the blood of wild rats. *Int. J. Syst. Bacteriol.* **48**, 1333–1339 (1998).

30. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998).
31. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
32. Ewing, B., Hillier, L., Wendl, M.C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
33. Meyer, F. *et al.* GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* **31**, 2187–2195 (2003).
34. Delcher, A.L. *et al.* Alignment of whole genomes. *Nucleic Acids Res.* **27**, 2369–2376 (1999).
35. Carver, T.J. *et al.* ACT: the Artemis Comparison Tool. *Bioinformatics* **21**, 3422–3423 (2005).
36. Halling, S.M. *et al.* Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis*. *J. Bacteriol.* **187**, 2715–2726 (2005).
37. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
38. Kumar, S., Tamura, K. & Nei, M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**, 150–163 (2004).
39. Lampe, D.J., Akerley, B.J., Rubin, E.J., Mekalanos, J.J. & Robertson, H.M. Hyperactive transposase mutants of the *Himar1* *mariner* transposon. *Proc. Natl. Acad. Sci. USA* **96**, 11428–11433 (1999).
40. Lampe, D.J., Churchill, M.E. & Robertson, H.M. A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. *EMBO J.* **15**, 5470–5479 (1996).
41. Sweger, D. *et al.* Conservation of the 17-kilodalton antigen gene within the genus *Bartonella*. *Clin. Diagn. Lab. Immunol.* **7**, 251–257 (2000).
42. Lindroos, H.L. *et al.* Characterization of the genome composition of *Bartonella koehlerae* by microarray comparative genomic hybridization profiling. *J. Bacteriol.* **187**, 6155–6165 (2005).
43. Seubert, A., Schulein, R. & Dehio, C. Bacterial persistence within erythrocytes: a unique pathogenic strategy of *Bartonella* spp. *Int. J. Med. Microbiol.* **291**, 555–560 (2002).
44. Kaneko, T. *et al.* Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.* **9**, 189–197 (2002).