A Genomic Approach to the Role of Type IV Secretion Systems in *Bartonella* Host Adaptation

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von
Philipp Engel
aus Essen, Deutschland

Basel, 2011

creative commons

Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 2.5 Schweiz

---

Sie dürfen:

das Werk vervielfältigen, verbreiten und öffentlich zugänglich machen

Zu den folgenden Bedingungen:

Namensnennung. Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen (wodurch aber nicht der Eindruck entstehen darf, Sie oder die Nutzung des Werkes durch Sie würden entlohnt).

Keine kommerzielle Nutzung. Dieses Werk darf nicht für kommerzielle Zwecke verwendet werden.

Keine Bearbeitung. Dieses Werk darf nicht bearbeitet oder in anderer Weise verändert werden.

- Im Falle einer Verbreitung müssen Sie anderen die Lizenzbedingungen, unter welche dieses Werk fällt, mitteilen. Am Einfachsten ist es, einen Link auf diese Seite einzubinden.
- Jede der vorgenannten Bedingungen kann aufgehoben werden, sofern Sie die Einwilligung des Rechteinhabers dazu erhalten.
- Diese Lizenz lässt die Urheberpersönlichkeitsrechte unberührt.

Quelle: http://creativecommons.org/licenses/by-nc-nd/2.5/ch/ Datum: 3.4.2009

Genemigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Christoph Dehio

Prof. Urs Jenal

Basel, den 08.12.2009

Prof. Dr. Eberhard Parlow

(Dekan)

for Fabienne

Statement to my Thesis

This work was carried out in the group of Prof. Christoph Dehio in the Focal area Infection Biology at the Biozentrum of the University of Basel. My PhD thesis committee consisted of:

Prof. Christoph Dehio
Prof. Urs Jenal
Prof. Guy Cornelis
Prof. Walter Salzburger

My thesis is written in a cumulative format. It consists of a synopsis covering several aspects related to my work. This is followed by result chapters presenting my research consisting of a published research article, a manuscript in preparation, a submitted manuscript, unpublished results, and a published review article. Finally, I recapitulate the major findings of my thesis and discuss some aspects and open questions of this work.

Table of Contents

1. Introduction

1.1 Studying bacterial evolution in the genomic era

How bacteria and their traits are shaped by evolution has always been an important question. Particularly, the understanding of the evolutionary mechanisms underlying the occurrence and emergence of bacterial pathogens is of general importance. Since evolution occurs through changes in heritable traits encoded on the DNA, a prerequisite for studying bacterial evolution is the access to DNA sequence information. For long time, DNA sequences were available for only a few genes or taxa, and systematic or representative data which would allow inference of evolutionary aspects was missing (Seifert and DiRita 2006). The appearance of the first complete genome sequence in 1995, the one of *Haemophilus influenzae,* has dramatically changed the possibilities to examine evolutionary patterns (Fleischmann et al. 1995). In this study, the complete sequence of the attenuated laboratory strain Rd was compared to the sequences of known virulence genes cloned from more pathogenic clinical isolates providing insights into the molecular evolution of this pathogen.

In the following, new genome projects were initiated and the continuous improvement of existing as well as the establishment of new sequencing technologies (Schuster 2008) has resulted in more and more genomic sequences deposited in databases (Fig. 1) . At the time of writing, 1001 complete microbial genomes (archea and bacteria) were found to be stored in public databases (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi). The availability of genomic sequence data did not only provide more than enough information to infer comprehensive phylogenic relationships among bacterial species, but was also the onset for a new way of studying bacterial evolution (Lawrence 2005). Comparative genomics - the comparison of different genome sequences across biological species or strains - is nowadays an important research field mainly aiming at the understanding of molecular mechanisms underlying evolution of biological organisms (Koonin and Wolf 2008). By these means, certain biological questions about evolution could be addressed the first time and unexpected genomic characteristics revealed.

Bacterial chromosomes have been viewed for long time as collections of a defined set of genes (Lawrence and Hendrickson 2005). However, the unprecedented view into the genomes of bacteria discovered a high degree of

plasticity reflected by the difference in size among sequenced bacterial genomes ranging from 180 kb in the intracellular symbiont *Carsonella rudii* to 13 Mb in the soil bacterium *Sorangium cellulosum* (Fig. 1). The high abundance and importance of horizontal gene transfer (HGT), i.e. the lateral acquisition of foreign DNA, might be the biggest conceptual novelty brought about by comparative genomics of prokaryotes (Lawrence 1999; Ochman et al. 2000). In the pre-genomic area, HGT was viewed as a marginal phenomenon responsible for specific evolutionary events such as the spread of resistances (Koonin and Wolf 2008). Nowadays it is clear that horizontal transfer of genes plays a central role in the evolution of bacteria, and together with the phenomenon of gene loss represents the major factors contributing to the observed differences in genome size among bacteria. This high degree of genomic plasticity found among bacteria seems to reflect their distinct lifestyles and their adaptation to a wide range of ecological niches (Pallen and Wren 2007).

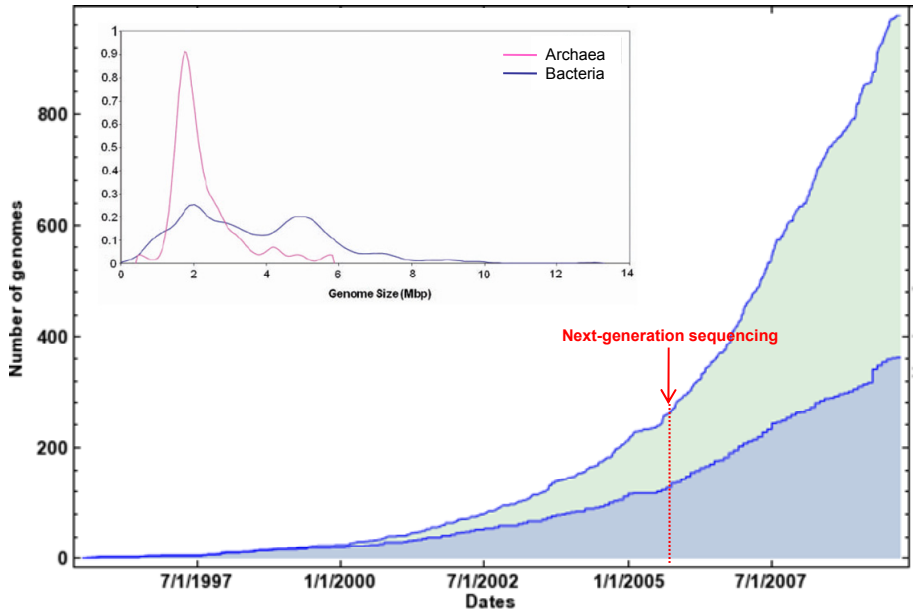


Figure 1: Exponential growth of microbial genome sequences in public databases since 1995 (green: genomes/ blue: genera). The onset of next-generation sequencing by the landmark publication of the sequencing-by-synthesis technology (Margulies et al. 2005) resulted in a boost of available genome sequences. Nowadays, genomes of most microbial genera are available. (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_growth.html). The inset shows the distribution of genome sizes among bacteria and archaea. Sequenced genomes of bacteria show a wide distribution with two peaks at 2 and 5 Mb, whereas archeal genomes sharply peak at 2Mb. This difference could be a sequencing bias, as the representation of archaeal genomes in the current databases is much less complete than the representation of bacterial genomes (adapted from (Koonin and Wolf 2008)).

1.2 Genome dynamics: Common strategies for adaptation in bacteria

Bacterial genomes are shaped by three main forces: gene change, gene loss, and gene gain (Pallen and Wren 2007). In the following each of the three mechanisms will be described and their consequences in regard to bacterial evolution and adaptation discussed. It is to note that most adaptive changes result from the interplay of these three forces. For pathogens, the most important adaptive changes underlying genomic dynamics are summarized in Fig. 2.

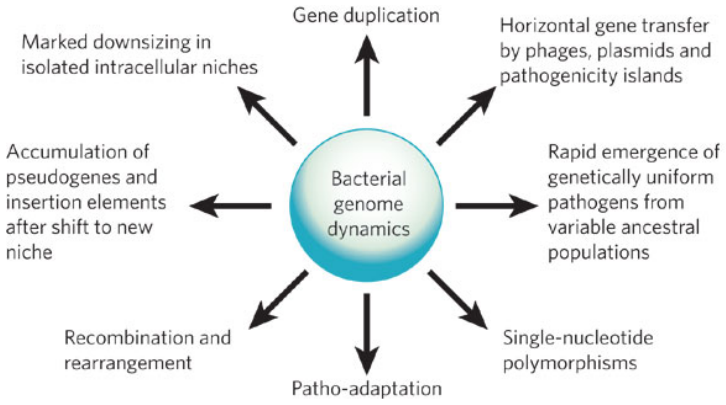


Figure 2: Representation of different genome dynamics occurring in bacterial pathogens. The three main forces shaping a genome (gene gain, gene loss, and gene change) can take place in a single bacterium. Most adaptive changes result from the interplay of these forces (Pallen and Wren 2007).

1.2.1 Gene change

Gene change describes mechanisms affecting the sequence or the order of existing genes (Pallen and Wren 2007). These mechanisms include **nucleotide substitutions, insertions, and deletions**, as well as **rearrangements** of single genes or larger genomic regions.

1.2.1.1 Rearrangements

Bacterial chromosomes experience a high degree of organization in regard to gene distribution and sequence composition. The maintenance of these structural features seems to be under strong selection (Rocha 2004). Therefore, the generation of

genetic variability by genomic recombination might be in an evolutionary conflict with the conservation of genomic organization. Still, the occurrence of genetic variability by intra-genomic recombination is frequently observed in bacteria (Hacker et al. 2003; Rocha 2004), even though varying a lot among bacteria adapted to different lifestyles (Mira et al. 2002). Genome stability correlates reciprocally with the degree of gene content and the presence of various types of repeated sequences (Mira et al. 2002). Obligate intracellular bacteria (e.g. *Rickettsia* or *Buchnera*) have the smallest genomes and display the highest degree of stability which reflects their protected and highly constant ecological niches (Moran 2002; Tamas et al. 2002; Moran et al. 2009). In contrast, free-living species as well as facultative intracellular pathogens such as *Salmonella* (Liu and Sanderson 1996), *Shigella* (Yang et al. 2005), *Strepptococcus* (Nakagawa et al. 2003), or *Helicobacter pylori* (Aras et al. 2003) undergo frequent genomic rearrangements due to the presence of repeated sequence elements. In these species, the periodic stresses imposed e.g. by the immune system or the environment probably resulted in the relaxed selection on organizational features of their chromosomes, thereby allowing fast adaptation via repeat-mediated rearrangements (Rocha 2004). The analysis of genomic organization between pairs of bacterial genomes by gene position plots (dot plots) revealed that genomic rearrangements often show symmetry around the origin or terminus of replication (Fig. 3). This phenomenon was thought to be the result of high recombination frequencies at the open replication forks (Andersson 2000; Tillier and Collins 2000).

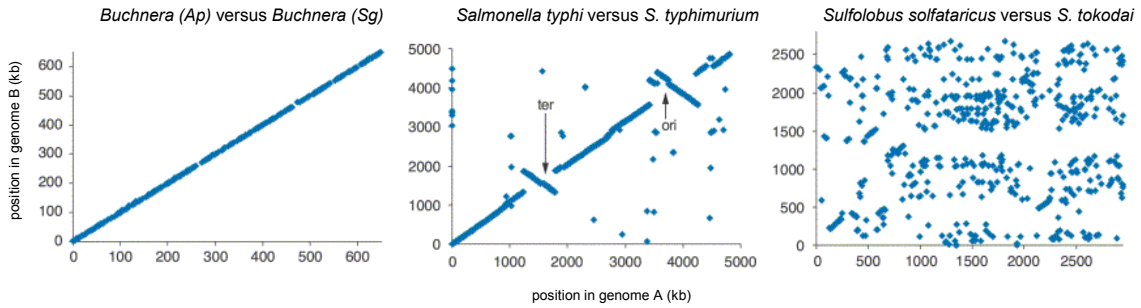


Figure 3: Dot plot comparisons of related species with different degree of genome organization (adapted from (Mira et al. 2002)]).

Beside large genomic rearrangements single-gene translocations and inversions were shown to be abundant in genomes harboring insertion sequence (IS)

elements. In *Sulfolobus spp.*, the atypical high content of these elements resulted in a highly variable genome organization as seen from the scattered non-linear dot plot in Fig. 3 (She et al. 2001).

Gene change can also occur by recombination between homologous genes. This mechanism, also called gene conversion, is best described for *Neisseria* ssp. which increase antigenic variation of different surface-exposed proteins, e.g. pilus subunits, by genetic recombination between expressed genes and silent loci (Palmer and Brayton 2007; Hill and Davies 2009).

1.2.1.2 Nucleotide substitutions, insertions, and deletions

Nucleotide insertions/deletions in genes frequently result in the occurrence of premature stop codons. Therefore, they are under strong negative selection or display the first step towards gene loss (see 1.2.3). Exceptions are the insertions/deletions of nucleotide triplets leading to the preservation of the reading frame. In some bacterial pathogens, a stochastical change in expression of surface structures produces a heterogenic phenotype which plays an important role for the infectivity of these pathogens. This phenomenon is called phase variation and can be achieved by e.g. slipped-strand mispairing (SSM) of contiguous DNA repeat units during replication. SSM results in the expansion or contraction of the number of repeats and ultimately to the phase-variable expression of a protein, if the regulatory or coding sequence is affected (van der Woude and Baumler 2004).

Nucleotide substitutions display the most abundant type of gene change and continuously produce genetic variability which evolution can act on (Seifert and DiRita 2006; Barrick et al. 2009). The type and number of mutations that become fixed in a population can tremendously vary within and between genomes. If mutations have no significant consequences for fitness, they accumulate stochastically in direct proportion to the mutation rate (Rocha 2008). Nucleotide changes in intergenic regions or those that do not change the amino acid sequence of a gene often are considered to have no fitness effect and therefore are not affected by natural selection. In contrast, mutations which have a positive or negative fitness effect on the organism (e.g. mutations changing the amino acid sequence) are quickly fixed (positive selection) or eliminated from the population (purifying

selection), respectively (Jordan et al. 2002). Proteins responsible for conserved cellular core functions are under severe functional constraints and therefore evolve slowly by accumulation of mainly neutral mutations. Most mutations affecting the protein sequence of these genes are deleterious and strong purifying selection is acting on them. As neutral mutations are directly proportional to the mutation rate, conserved core genome genes can be used to infer the relative time which has passed since divergence of species, thereby allowing to infer evolutionary relationships (Seifert and DiRita 2006). Although most genes evolve under purifying selection, the evolutionary rate can significantly vary across different functional classes of proteins (Jordan et al. 2002). Proteins which adopt new functions or are involved in the specific interaction with the environment were shown to carry substantially more mutations affecting the protein sequence (Murphy 1993; Moxon and Thaler 1997; Zheng et al. 2004; Chen et al. 2006; Bergthorsson et al. 2007; Petersen et al. 2007). These proteins are subject to adaptive evolution, and positive selection has resulted in fixation of mutations increasing the fitness of the organism in its ecological niche. For uropathogenic *E.coli* strains (UPEC), it was shown that proteins known to be important for causing urinary tract infection are under positive selection compared to non-UPEC strains. This reflects the adaptation of the UPEC strains to their specific niche within the host (Fig. 4).

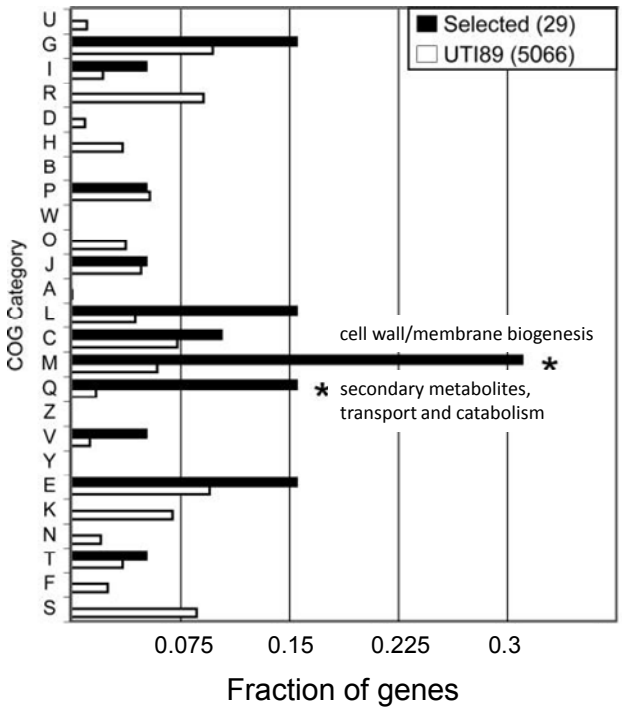


Figure 4: Genes under positive selection in uropathogenic *E.coli* strains. The analyzed genes are categorized according to functional classes. Black bars represent positively selected genes and white bars display total number of genes. Genes under positive selection are enriched in two functional categories (indicated by asterisk): These genes encoded adhesins, outer membrane proteins, proteins involved in iron acquisition and DNA repair systems, and proteins for regulation of LPS structures (adapted from (Chen et al. 2006)).

Another example are the amino acid variations in the fimbrial adhesins of *E. coli* and *Salmonella* which were shown to determine host specificity and tissue tropism allowing the transition from a commensal to a pathogenic lifestyle (Sokurenko et al. 1998; Boddicker et al. 2002; Hacker et al. 2003; Weissman et al. 2003). These so-called pathoadaptive processes also include the fast evolution of host-interacting proteins of pathogenic bacteria as a result of the selection pressure imposed by the host immune system. In this co-evolutionary arms race, both sides, the bacteria and the host, try to continuously improve their fitness by adaptive evolution (Stavrinides et al. 2008; Boller and He 2009).

Beside intra-genomic differences, the degree of accumulation of mutations varies extensively among genomes from different bacterial strains and species (Tamas et al. 2002; Hacker et al. 2003; Denamur and Matic 2006). Small effective population sizes and increased mutation rates seem to be the two main reasons leading to an increased occurrence of mutations in certain bacterial organisms. For example, endosymbionts of the genus *Buchnera* revealed mutation rates 2-4-fold greater than their close relatives *E. coli* or *Salmonella* (Moran 1996; Clark et al. 1999). In contrast to free-living prokaryotes with an effective population size estimated at > 10^9 for *E. coli* (Moran 1996), endsymbionts have tiny populations which mainly evolved clonal. This is caused (i) by the bottleneck occurring in each host generation when progeny are inoculated and (ii) by the strict separation of endosymbiont lineages in different hosts (Moran 1996). According to Muller's ratchet (Felsenstein 1974), smaller populations accumulate faster mildly deleterious mutations and therefore show an increased rate of sequence evolution.

A general increase in the mutation rate by changing the fidelity of DNA replication or repair can result in a more rapid accumulation of mutations. This fine-tuning of the mutation rate is used by bacteria to increase their adaptability to changing environmental conditions and results in so-called mutator strains (Denamur and Matic 2006). However, since deleterious mutations are appearing at a much higher rate than beneficial, increased mutation rates are not always of advantage and can be counter-selected at one point of evolution (Giraud et al. 2001). Bacteria with a generally increased mutation rate were detected in many natural bacterial populations of *E. coli*, *Salmonella*, or *Neisseria*. In particular among pathogens, they frequently occur resulting in efficient variation of surface antigens and facilitating the acquisition of virulence or resistance determinants (Taddei et al. 1997). As an

example, among strains of *Pseudomonas aeruginosa* isolated from lungs of cystic fibrosis patients, a correlation between multiple-antibiotic resistance and high mutation rates was found (Oliver et al. 2000).

1.2.2 Gene gain

One of the greatest surprises uncovered by genome comparison was the degree of genomic variability within many bacterial species. The fact that two *E. coli* strains can differ by almost one quarter of their genomes (Hayashi et al. 2001; Welch et al. 2002) has reflated the debate about the definition of a prokaryotic species (Gevers et al. 2005; Fraser et al. 2009). For several species, the number of genes common to all strains of that species (i.e. core genome) may display only a small fraction of its entire gene pool (i.e. pan-genome) (Medini et al. 2008). It was even proposed that for some species the pan-genome may be of unlimited size (Tettelin et al. 2005). As an example, the analysis of 17 *Streptococcus pneumoniae* genomes revealed an almost 4-fold larger pan-genome than core genome (Hiller et al. 2007).

Where are these genes coming from? Two main mechanisms are known to result in gain of genetic material: **gene duplication** and **horizontal gene transfer** (Pallen and Wren 2007).

1.2.2.1 Gene duplication

The duplication of genes displays a highly relevant biological process. An increase in gene dosage may result in the higher expression of a given gene. Furthermore, subsequent diversification of one of the two gene copies frequently leads to emergence of a new function, thereby facilitating adaptation to changing environments (Gevers et al. 2004; Andersson and Hughes 2009). Gene duplications belong to the most common type of mutations, and it has been estimated that at least 10% of all cells in growing bacterial culture contain one duplication somewhere in the genome (Anderson and Roth 1981; Andersson and Hughes 2009). Therefore, it is not surprising that in some larger bacterial genomes paralogous genes (i.e. related genes emerged by duplications) can make up as much as 50 % of the entire coding content (Fraser-Liggett 2005). The number of paralogous genes strongly correlates with genome size (Fig. 5), and duplicated genes mainly belong to functional classes

involved in the adaptation to constantly changing environments (such as amino acid, inorganic ion, or carbohydrate metabolism, as well as transcription, defense mechanisms, energy production and conversion) (Gevers et al. 2004). This is in line with the small fraction of paralogous genes found in the genomes of obligate intracellular organisms which colonize highly constant and nutrient-rich environments (Fig. 5).

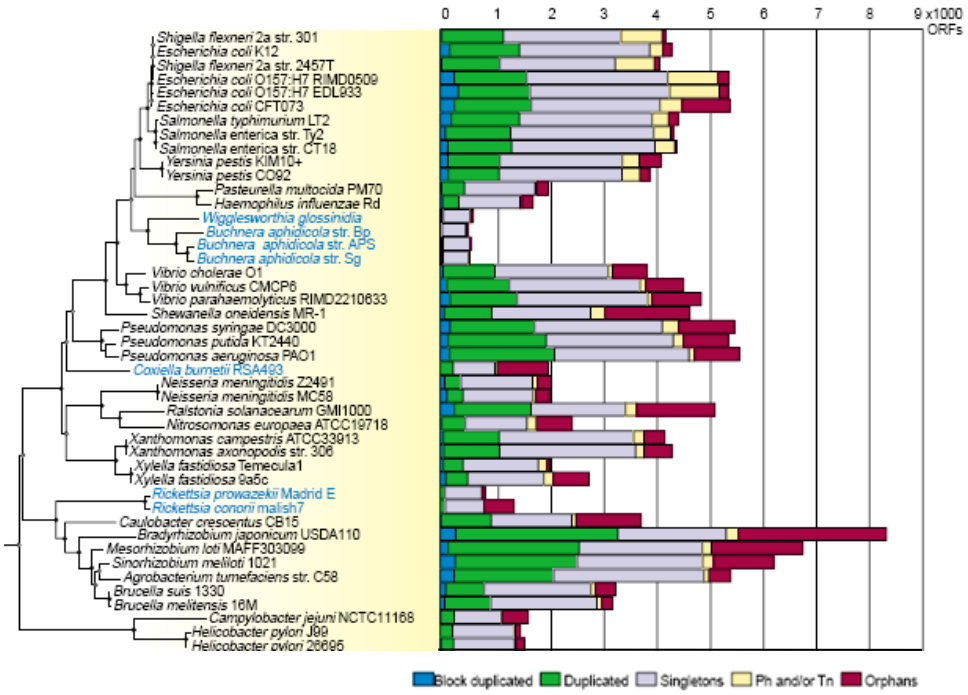


Figure 5: Prevalence of duplicated genes in proteobacteria. The category “Block duplicated” refers to genes belonging to a large duplicated DNA segment. Singletons are genes present only once in the corresponding genome. Ph and/or Tn are phage- and transposon-related genes. Orphans are genes without any homolog in other genomes. Obligate intracellular species are depicted in blue (Gevers et al. 2004).

For many bacteria, gene duplications display an adaptive response to the selective pressure acting on the organism. In *Borrelia burgdorferi*, duplicated motility and chemotaxis genes comprise more than 6% of the proteome. These genes seem to confer the ability to migrate to distant sites under different physiological conditions in the tick as well as in the mammalian host (Fraser et al. 1997). In agreement with the complex nature of the mycobacterial cell wall, genes involved in fatty acid metabolism are duplicated in *Mycobacterium*, thus, reflecting the adaptive evolution of the bacterial cell surface (Gevers et al. 2004). Many other examples are known

where the amplification of genes results in a dosage effect, thereby displaying direct adaptive response to a selective pressure. These types of duplications are often found in genes conferring resistance to antibiotics emphasizing the importance of this type of gene gain (Sandegren and Andersson 2009).

1.2.2.2 Horizontal gene transfer

The process of horizontal gene transfer (HGT) is the transmission of genetic information from one bacterial genome to another (Ochman et al. 2000). Albeit it was thought for a long time to be restricted to the spreading of resistance genes among bacterial populations, comparative genomics has revealed that HGT displays the most potential and abundant evolutionary path in the microbial world for generating biological diversity (Koonin and Wolf 2008). Although cases of HGT between bacteria and their hosts are reported, as for example in *Legionella pneumophila* (Bruggemann et al. 2006), it generally occurs between different bacterial strains and species. Mechanisms known to mediate HGT are conjugative transfer of plasmids or transposons, transduction of bacteriophages, or transformation of naked DNA (Salyers et al. 1995; Thomas and Nielsen 2005). Once acquired by a recipient cell, the foreign genetic information can either be maintained on an extra-chromosomal replicon or integrated into the chromosome (Burrus and Waldor 2004). The former case accounts for plasmids acquired by conjugation, although they may get assimilated into the recipient chromosome at one point of evolution.

The integration of the horizontally acquired DNA mostly occurs as one syntenic block and is typically maintained as a so-called genomic island (GI) in the recipient genome (Juhas et al. 2009). GIs can be identified by different characteristic features (Fig. 6). Normally, they are relatively large segments of DNA (10-200 kb) present among closely related strains but absent from others. Smaller segments are also referred to as genomic islets (Hacker and Kaper 2000). Due to their foreign source, the nucleotide composition of GIs can differ from the rest of the chromosome measurable e.g. by the GC content. GIs often are inserted at conserved sites in the recipient chromosome as for example in tRNA genes. The site-specific integration results in the presence of perfect direct repeats of about 20 bp flanking the GI (Schmidt and Hensel 2004). Depending on the evolutionary stage, GIs can harbor genes or gene remnants associated with their mobilization or horizontal transfer such

as integrases, insertion sequence elements, transposases, or plasmid and phage genes (Buchrieser et al. 1998; Gal-Mor and Finlay 2006). Most importantly, GIs often carry genes offering a selective advantage to the host bacteria which results in the fixation within the bacterial population (Schmidt and Hensel 2004). It is to note that many of the identified horizontally acquired islands are lacking one or more of the above-described features. This can be explained by the fact that many GIs are in a state of evolutionary regression, which results in the deterioration of mobilization genes and other signatures of their integration, thereby only maintaining genes conferring adaptation. Another reason is that GIs can be of many different origins harboring very distinct features, and their exact mechanism of chromosomal integration often is not fully understood (Juhas et al. 2009). Depending on the mediated function, a GI is referred to as pathogenicity, symbiosis, resistance, or metabolic island (Dobrindt et al. 2004).

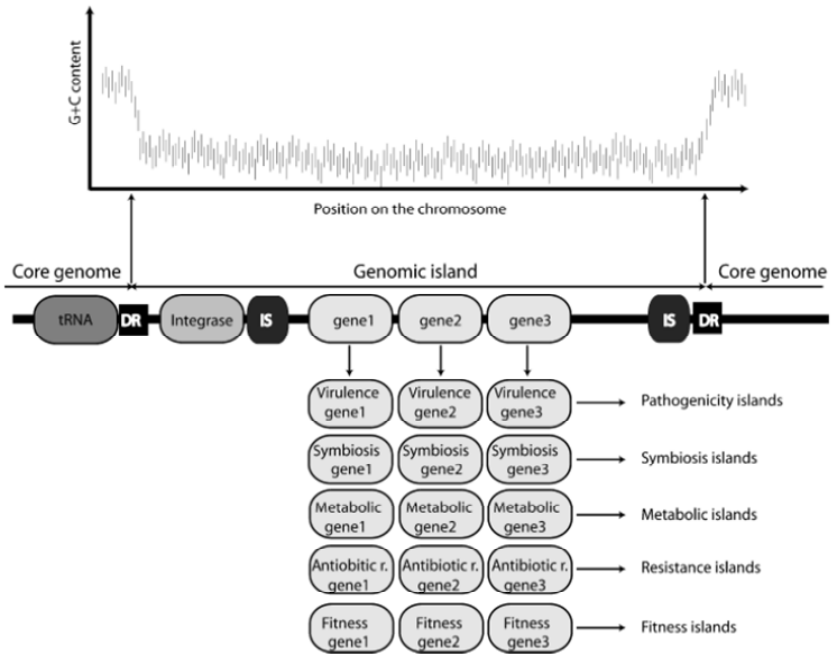


Figure 6: Typical features of GIs. According to their gene content, GIs can be described as pathogenicity, symbiosis, metabolic, resistance, or fitness islands. DR, directed repeat and IS, insertion sequence element (Juhas et al. 2009).

In the evolution of pathogens, HGT is of particular importance since virulence-related traits can be acquired in a single step resulting in a so-called evolutionary “quantum-leap” (Groisman and Ochman 1996), i.e. the rapid alteration of the life-style of the recipient bacterium and the emergence of high-virulence strains. Herein, the

spread and dissemination of antibiotic resistance genes among bacterial populations displays a serious threat to public health. This is particularly the case for pathogenic variants of gram-positive cocci (e.g. *Staphylococcus aureus*) causing severe sepsis and catheter-associated infections (Deurenberg et al. 2007). Due to the constant application of antibacterial drugs in hospitals, many of these strains carry GIs or plasmids which are conferring resistance to a variety of antibiotics (Chambers and Deleo 2009; Fischbach and Walsh 2009). The change from the almost universal susceptibility to increasing antibiotic resistance in a few decades is may be the most illustrative example of the remarkable capacity of HGT in conferring bacterial adaptation.

Beside antibiotic resistances, frequently genes involved in host-interaction are disseminated via HGT allowing unrelated pathogens to use very similar strategies to subvert their hosts (Gal-Mor and Finlay 2006). Table 1 provides a list of known virulence factors found on mobile genetic elements in different bacterial pathogens. Functionally versatile factors, easily adaptable for different purposes, are often acquired by HGT. Type III secretion systems (T3SS) and type IV secretion systems (T4SS) display good examples, as these nanomachines represent molecular needles dedicated to the injection of different effector molecules into host cells (Christie et al. 2005; Coburn et al. 2007; Mueller et al. 2008). By switching the number and type of translocated molecules the host niche can be modulated for the specific purpose of the bacteria (Hueck 1998; Cascales and Christie 2003).

Table 1: Mobile genetic elements that encode virulence factors present in human pathogens (Pallen and Wren 2007)

Type of mobile element	Pathogen	Virulence factor
Plasmid	*Bacillus anthracis*	Anthrax toxin
	Clostridium tetani	Tetanus toxin
	Enterotoxigenic *Escherichia coli*	Heat-stabile toxin, heat-labile toxin and fimbriae
	Mycobacterium ulcerans	Polyketide toxin
	Salmonella enterica serovar Typhimurium	SpvR, SpvA, SpvB, SpvC and SpvD proteins*
	Shigella spp.	Type III secretion system
	Staphylococcus aureus	Exfoliatin B
	Pathogenic *Yersinia* spp.	Type III secretion system
Prophage	*Corynebacterium diphtheriae*	Diphtheria toxin
	Enterohaemorrhagic *E. coli*	Shiga toxin and type III secretion effectors
	S. aureus	Staphylococcal enterotoxin A, exfoliatin A and Panton–Valentine leukocidin
	Streptococcus pyogenes	Streptococcal pyrogenic exotoxins, DNases and streptococcal phospholipase A_2 (Sla)
	Vibrio cholerae	Cholera toxin
Pathogenicity island	*Clostridium difficile*	Clostridial enterotoxin and clostridial cytotoxin
	Enteropathogenic and enterohaemorrhagic *E. coli*	Type III secretion system
	Uropathogenic *E. coli*	Fimbriae, iron-uptake systems, the capsular polysaccharide and α-haemolysin
	Helicobacter pylori	Cag antigen
	S. enterica	Type III secretion systems
	S. aureus	Toxic-shock toxin, staphylococcal enterotoxin B, enterotoxin C, enterotoxin K and enterotoxin L

*Involved in intracellular survival.

Strikingly, the same traits adopted by one species for virulence can be used by another to establish a symbiotic interaction. This was demonstrated for T3SS which were found to be present in the tsetse fly-symbiont *Sodalis glossinidius* and in various plant-symbiotic species of the rhizobiae (Dale et al. 2001; Marie et al. 2001).

Although GIs and HGT have mostly been associated with the dissemination of host interacting factors or antibiotic resistance genes, it is obvious that free-living organisms are using the same strategies to adapt to their ever-changing environment which offers vast amounts of different ecologically niches (Davison 1999). The life-style of these species in microbial communities exposes them to large amounts of foreign DNA in contrast to their host-associated relatives (Seifert and DiRita 2006). Accordingly, free-living bacteria generally comprise of larger genomes and harbor plasmids or GIs containing metabolic functions of several hundred kbs (Konstantinidis and Tiedje 2005; Sallstrom and Andersson 2005). In the anaerobic bacterium *Geobacter sulfurreducens*, a GI of 300 kb was detected encoding genes implicated in anaerobic metabolism, thus, reflecting the adaptive evolution to this specific life-style. Other environmental bacteria were found to carry mobile genetic elements responsible for the degradation of xenobiotics, i.e. anthropogenic compounds released into the environment. Analysis of the enzymatic pathways encoded on these elements indicated that they display an adaptive response to these environmental pollutants by circumventing their potential toxic effects and using them as alternative nutrient sources (Springael and Top 2004).

Finally, it is to note that the identification of horizontally transferred genes by genomic analyses has certain limitations. The continuous amelioration of the foreign DNA to the genetic composition of the recipient chromosome blurs the signatures of HGT (Lawrence and Ochman 1997). Still, there is clear evidence for ancient horizontal gene transfer. It was shown that highly conserved genes were horizontally transferred between all three domains of life, such as glutamine or methionyl-tRNA synthetase (Brown 2003). A striking example represents the archeal species *Methanosacrina mazei*, which has laterally acquired nearly 30% of its genetic content from an eubacterial source (Deppenmeier et al. 2002). Although, the extent of ancient HGT in the microbial world is not clear yet, the question about the feasibility of constructing a universal 'Tree of Life' has become a hot topic (Wolf et al. 2002).

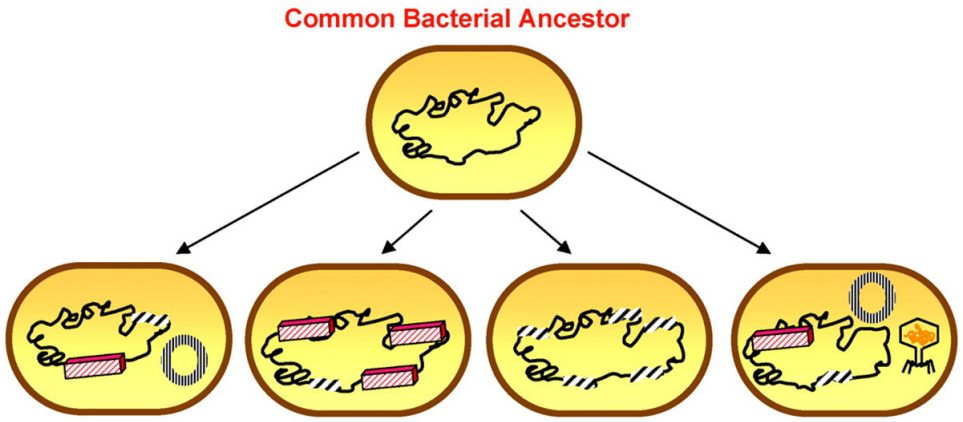
1.2.3 Gene loss

Since bacterial genomes are not ever-expanding (Koonin and Wolf 2008), it is obvious that the acquisition of genetic material by HGT or gene duplication is counterbalanced by loss of DNA. Furthermore, the lack of extensive non-coding sequences indicates that there is a bias towards small genomes in bacteria. This is in clear contrast to eukaryotic genomes which can differ by almost 300-fold in size with only a sixfold difference in gene number (Mira et al. 2001).

In fact, for many bacteria, it has been shown that gene loss is linked to the specific adaptation to different ecological niches. When a bacterium is adapting to a new environment, the selective pressure acting on it will be different than before, and former beneficial traits may become incompatible with growth in the new niche. Consequently, this leads to the down-regulation or elimination of the incompatible genes (Maurelli 2007). For example, strains of the pathogenic species *Shigella* lack *ompT*. This gene encodes a surface protease present in closely related non-pathogenic *E. coli*. Apparently, the deletion of *ompT* was a prerequisite for the adaptation to the pathogenic life-style of *Shigella*, as the experimental introduction of this gene was shown to suppress its intracellular spread (Fig. 7) (Nakata et al. 1993; Hacker et al. 2003).

In contrast to these pathoadaptive processes which are resulting in the loss of one or a few genes, several bacterial species exhibit a much more extreme reduction of their coding content. Across different evolutionary lineages, the transition from a free-living to strictly intracellular life-style was associated with an extensive loss of DNA (Casadevall 2008). Albeit, the evolutionary forces driving genome reduction are of different nature, yet they are clearly related to the restricted intracellular life-style of these organisms. At first, the highly constant environment within the host cell, rich in metabolic intermediates, results in the elimination of the selective pressure acting on most biosynthetic genes. Second, intracellular bacteria normally have small population sizes resulting in the fixation of slightly deleterious mutations (see 1.2.1.2). Thereby potentially beneficial genes enhancing the efficiency of important cellular functions, such as DNA replication and repair, can be lost, which consequently leads to further increase in the accumulation of deleterious mutations. The resulting pseudogenization is then followed by successive deletion of the non-functional genes (Casadevall 2008).

As bacteria have adapted to the intracellular lifestyle at various points of evolution, their genomes provide snapshots of the different stages of genome reduction. The recently evolved severe human pathogens *Yersinia pestis*, *Salmonella enteric* serovar Typhi, or *Bordetella pertussis* consistently show large-scale loss and inactivation of genes due to adaptation to a human-specific life-style (Wren 2003; McClelland et al. 2004; Preston et al. 2004). In these genomes, a high number of IS elements and repeats are found which were implicated as the source of chromosomal rearrangements, extensive deletions, and gene inactivation (Moran and Plague 2004). Ongoing genome reduction has also been found in the genomes of *Rickettsia prowazekii and Mycobacterium leprae* which harbor a high number of pseudogenes (Andersson et al. 1998; Cole et al. 2001). The most extreme examples of genome reductions are found among obligate intracellular pathogens and symbionts, such as *Buchnera*, *Wolbachia*, *Chlamydia* or *Mycoplasma* (Casadevall 2008). In *Buchnera*, the absence of pseudogenes, repetitive elements, and the highly conserved genomic structure (Fig. 3) reflects their ancient host-association which was established more than 150 million years ago (Tamas et al. 2002). The lack of opportunities to acquire foreign DNA in the protected niche of these organisms results in the irreversibility to a different lifestyle (Fig. 7). The discovery of the so far smallest bacterial genome, the one of the endosymbiont *Carsonella rudii* having a size of only ~180 kb, posed the questions whether obligate intracellular organisms are facing a dead end from where is no escape (Andersson 2006).



Acquisition of	PAIs, plasmid	PAIs	- -	PAIs, plasmid, phages
Loss of	Non-essential genes, including *cad*A, *omp*T	Non-essential genes	Non-essential genes, ca. 70% of the genome	Non-essential genes
Lifestyle	Facultative intracellular pathogen (Cytoplasm)	Extracellular pathogen	Obligate intracellular symbiont (Bacteriocyte)	Facultative intracellular pathogen (Vacuole)

Figure 7: Bacterial evolution driven by the interplay of different genome dynamics. Different variants of pathogenic and symbiotic gammaproteobacteria have evolved by the acquisition and loss of genetic information from a common bacterial ancestor. *cadA*, lysine decarboxylase-encoding gene; *ompT*, outer membrane protein T-encoding gene; PAI, pathogenicity island; EIEC, enteroinvasive *E. coli*; UPEC, uropathogenic *E. coli* (Hacker et al. 2003).

1.3 Alphaproteobacterial evolution

1.3.1 Genomic plasticity: Driving force of adaptability

The alphaproteobacteria, which the genus *Bartonella* belongs to, represents one of the most diverse bacterial subdivisions. They display great variability not only in metabolic capacity, morphology, or life cycle (Batut et al. 2004), but they are also colonizing most imaginable habitats. Free-living alphaproteobacteria have been found in water and soil; others form intra- and extracellular associations with eukaryotes, such as unicellular organisms, nematodes, arthropods, plants, and mammals. In many of these different ecologically niches alphaproteobacteria display the predominant bacterial species (Nirgianaki et al. 2003; Venter et al. 2004; Giovannoni and Stingl 2005). The high prevalence of these microorganisms indicates their great capability to specifically adapt to a wide range of different habitats. In this respect, it was proposed that the alphaproteobacteria represent the bacterial equivalents of Darwin's finches (Ettema, 2009).

The high degree of adaptability in the alpha subdivision is reflected by the variation in genomic content: (i) their genome sizes are ranging from less 1 Mb to more than 9 Mb (Kaneko et al. 2002; McCutcheon et al. 2009), (ii) only 33%-97% of all genes of a given alphaproteobacterial genome are also from an alphaproteobacterial origin (Esser et al. 2007), and (iii) in some genomes, species-specific genes are counting more than a thousand (Boussau et al. 2004). In a quantitative analysis of the flux of genes during evolution, it was estimated that the last common ancestor of the alpha subdivision consisted of 3'000-5'000 genes and represented a free-living, aerobic, and motile bacterium with surface proteins for environmental and host interaction (Boussau et al. 2004). Massive genome expansion was observed at branches of soil- and plant-associated bacteria, whereas genome reduction mostly occurred within lineages of facultative or obligate intracellular bacteria of mammals (Fig. 8). Functional classes which show increased abundance in expanded genomes include energy metabolism, transport, and regulation; all which are processes involved in adaptability. Strikingly, environmental shifts have occurred several times during evolution in different lineages of the alphaproteobacteria, and they were always associated with the same fluctuation in genome size (Boussau et al. 2004).

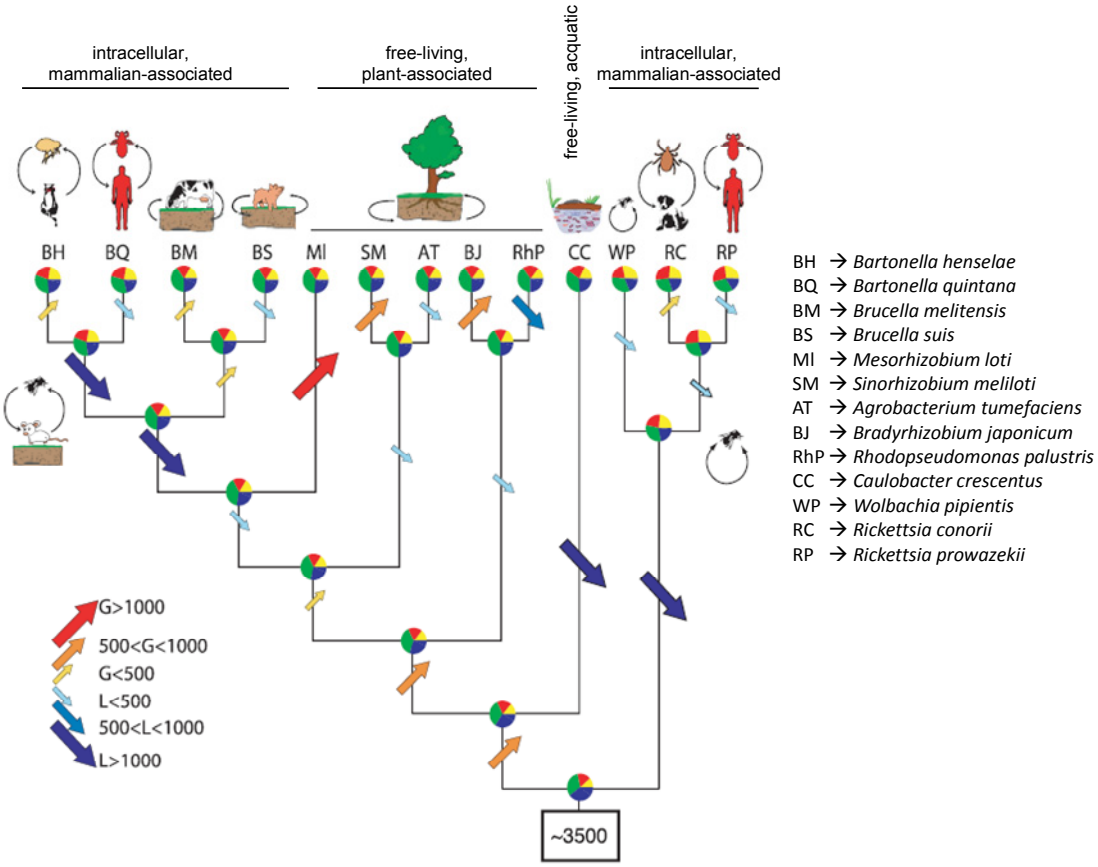


Figure 8: Net gene loss or gain throughout the evolution of the alphaproteobacterial species. Arrows pointing upward indicate net gains of genes (G), and arrows pointing downward indicate net losses of genes (L). Colors of circles refer to the relative fraction of genes assigned to the different functional groups in the modern and inferred genome at the node. Yellow, information storage and processing; green, metabolism; red, cellular processes; blue, poorly characterized (adapted from (Boussau et al. 2004).

Genome reduction is observed in the early branching lineage of the host-associated Rickettsiales, as well as in the lineage separating the facultative intracellular species of *Bartonella* and *Brucella* from the other Rhizobiales (Fig. 8). In both lineages, the high abundance of pseudogenes indicates a parallel ongoing trend towards genome reduction, although there is also evidence for lineage-specific horzontal acquisition of plasmid and phage-derived genes (Alsmark et al. 2004; Wu et al. 2004; Cho et al. 2007; Darby et al. 2007). For *Bartonella*, specialization to a single vector and host has been associated with the loss of genetic content. The louse-borne human pathogen *Bartonella quintana* was shown to display a reduced genomic variant of the closely related zoonotic species *Bartonella henselae* (Alsmark et al. 2004). Interestingly, *Rickettsia prowazekii*, as *B. quintana* transmitted by the

human body louse, also shows extensive genome reduction compared to its relatives (Fig. 8). This suggests that the massive genome decay in these species might be related to the biology of their common arthropod vector (Andersson et al. 1998; Alsmark et al. 2004).

In many alphaproteobacterial species auxiliary replicons of up to several megabases are present. Horizontally acquired genes located on these extrachromosomal segments count for most of the observed genome expansion in the alpha subdivision. Some of these elements contain plasmid-like replication systems indicating their plasmid-derived evolution (Batut et al. 2004). Large genetic variability is found on these auxiliary replicons which mostly contain niche-associated traits (Giuntini et al. 2005). In some cases, they are integrated into the chromosome and maintained as GIs as shown for *Bartonella* (Alsmark et al. 2004) and *Bradyrhizobium* (Viprey et al. 2000). Recently, it was revealed that parts of the integrated auxiliary replicon of *Bartonella* are massively amplified from an alternative origin of replication and packaged into phage particles. The authors proposed that this mechanism displays a novel strategy for increased lateral exchange of the many host-interacting genes contained in this region (Berglund et al. 2009). Further, in *Mesorhizobium loti*, an integrated symbiosis island of 500 kb was shown to be excised and transferred across species (Sullivan and Ronson 1998).

Summarizing, alphaproteobacterial genomes seem to harbor hot spots or so called “plasticity zones” where accessory gene content involved in niche specific adaptation is assimilated to and maintained. By the physical separation of adaptive traits and conserved housekeeping functions the evolutionary adaptability might be substantially enhanced explaining the ecological success of the alphaproteobacteria (Ettema and Andersson 2009).

1.3.2 Persistence: Common theme in infection strategies

Many alphaproteobacterial species are interacting with diverse eukaryotic hosts, either as symbionts (e.g. most Rhizobiales, *Wolbachia*), commensals (e.g. *Sinorhizobium*) or pathogens (e.g. *Agrobacterium*, *Anaplasma*, *Rickettsia*, *Bartonella*, and *Brucella*). Despite these differences in hosts-association and the above-described variation in genomic content, they share some properties in regard of their infection strategies. The common theme is not to kill the host, but to cause a chronic infection resulting in long-term persistence. Although the molecular mechanisms may differ, all host-associated alphaproteobacteria follow some general concepts to establish a chronic infection (Batut et al. 2004). The different steps towards a chronic infection are summarized in Fig. 9.

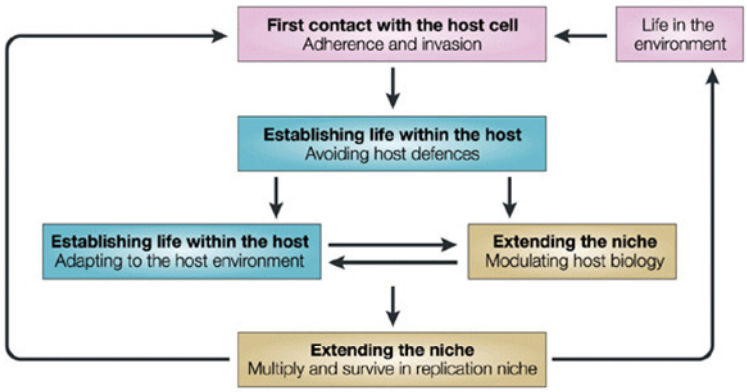


Figure 9: The road to a chronic infection. After the first contact is establish the bacteria have to adapt to the host environment and establish their niche. A crucial role plays the avoidance or suppression of the host's defense response. This requires a fine-tuned interaction with the host cells in which the secretion of effector proteins often plays an important role (Batut et al. 2004).

Most important for causing chronic infections, and therefore a recurrent feature in the interaction between alphaproteobacteria and their hosts, is the maintenance and extension of the intracellular niches (Batut et al. 2004). The induction of cell proliferation by preventing host cells from undergoing apoptosis is one way how niche expansion can be achieved. *Bartonella*, *Brucella*, and *Rickettsia* are following this strategy by activating different anti-apoptotic pathways (Gross et al. 2000; Kirby and Nekorchuk 2002; Joshi et al. 2003; Schmid et al. 2006). Another type of niche extension is observed in the plant-associated species of *Agrobacterium* or *Rhizobia*. These bacteria induce massive cell proliferation resulting in the formation of tumour-

like structures, thereby providing optimal conditions for bacterial growth and persistence (Zhu et al. 2000; Endre et al. 2002). Tumour-like manifestations resulting from vasoproliferation of the endothelium can also occur from infections with different *Bartonella* species (Koehler et al. 1992; Conley et al. 1994; Schmid et al. 2004). The hypertrophy (i.e. increase in cell size) of the ovaries induced by Wolbachia in the isopod hosts is another example of niche expansion among the alphaproteobacteria (Azzouna et al. 2004).

Also, important for long-term persistence is the fine-tuned modulation of the targeted host and the avoidance of a strong immune response (Batut et al. 2004). In this process, regulation plays a central role. In alphaproteobacteria, two-component regulatory systems are typically used to control the expression of host-interacting factors as an adaptive response to an outside stimulus. Also, the lipopolysaccharides (LPS) of many of the alpha subdivision species carry adaptive properties (Batut et al. 2004). In *Brucella* and *Bartonella*, the LPS has reduced endotoxic and immunostimulatory properties displaying a typical feature of stealth pathogens (Merrell and Falkow 2004; Zahringer et al. 2004; Barquero-Calvo et al. 2007). Many of the eukaryote-interacting species are using secretion systems to inject effector molecules into target cells resulting in a very specific modulation of the different niches. Most abundant are T4SSs ancestrally related to conjugation machineries. These versatile nanomachineries display key factors for host interaction in *Agrobacterium*, *Brucella*, *Anaplasma* and *Bartonella*. Recently, a putative role in pathogenicity was also proposed for T4SSs found in *Wolbachia*, *Erlichia*, and *Rickettsia* (Rances et al. 2008; Bao et al. 2009; Gillespie et al. 2009). Also, other secretion systems are found in alphaproteobacteria, as for example T3SS. In some of the rhizobial species these systems are involved in modulation of host range and host defense (Bartsev et al. 2004).

Although alphaproteobacterial species use clearly different molecular mechanisms for host interaction, and their underlying genetic factors are of different origin or only distantly related, the conceptual strategy seems to be the same. The persistent nature of alphaproteobacterial host-interactions, together with the high degree of adaptability, seems to render the evolutionary success of this bacterial subdivision.

1.4 Type IV secretion systems: Versatile nanomachines

1.4.1 Diversifying evolution of conjugation machineries

Type IV secretion systems (T4SSs) are multisubunit membrane-spanning nanomachineries which are dedicated to the transfer of bacterial effector molecules into cells of prokaryotes and eukaryotes (Christie et al. 2005). Among all secretion systems, T4SSs are unique by the ability to not only transfer proteins, but also DNA and DNA-protein complexes across different kingdoms. Presumably, the evolutionary origin as conjugation machineries explains this unique characteristic (Lawley et al. 2003; Christie 2004). T4SSs play an important role for niche adaptation of host-associated bacteria. First, by their ability to mediate HGT of various DNA molecules among bacteria T4SSs contribute to genome plasticity and the dissemination of antibiotic resistance as well as virulence traits (Hamilton et al. 2005; Juhas et al. 2008). Second, many T4SSs of pathogens and symbionts contribute directly to host interaction by translocating effector molecules into targeted host cells. This results in the modulation of a wide range of essential host cell functions in favor of the bacteria (Schmid et al. 2004; Schulein et al. 2005; Segal et al. 2005; Backert and Meyer 2006).

T4SSs are encoded by multiple genes often organized in a single block in the chromosome. As expected from their functional versatility, the type, number, and order of the T4SS-encoding genes can significantly differ. By constructing protein homology networks it was shown that T4SSs consist of conserved core components which are complemented by non-conserved proteins. This has resulted in a step-wise diversification of T4SSs. The genetic modularity seems to contribute essentially to the functional specialization of these nanomachines (Medini et al. 2006).

Based on the genetic organization, shared homologies, and evolutionary relationships, T4SSs have been classified into different types using two different classification schemes (Lawley et al. 2003; Christie et al. 2005). Originally, three different classes were based on the incompatibility group of conjugative plasmids: F (incompatibility group IncF), P (incompatibility group IncP), and I (incompatibility group IncI) (Lawley et al. 2003). Alternatively, type F and P were grouped together to the type IVA systems and type I was referred to as type IVB systems (Christie et al. 2005). The recent identification of a novel T4SS with only limited homology to the so

far existing groups indicated the presence of another type. Due to its function in the HGT of GIs, this type was referred to as GI-T4SS (Juhas et al. 2007). An overview of all types and their evolutionary relationship is given in Fig. 10.

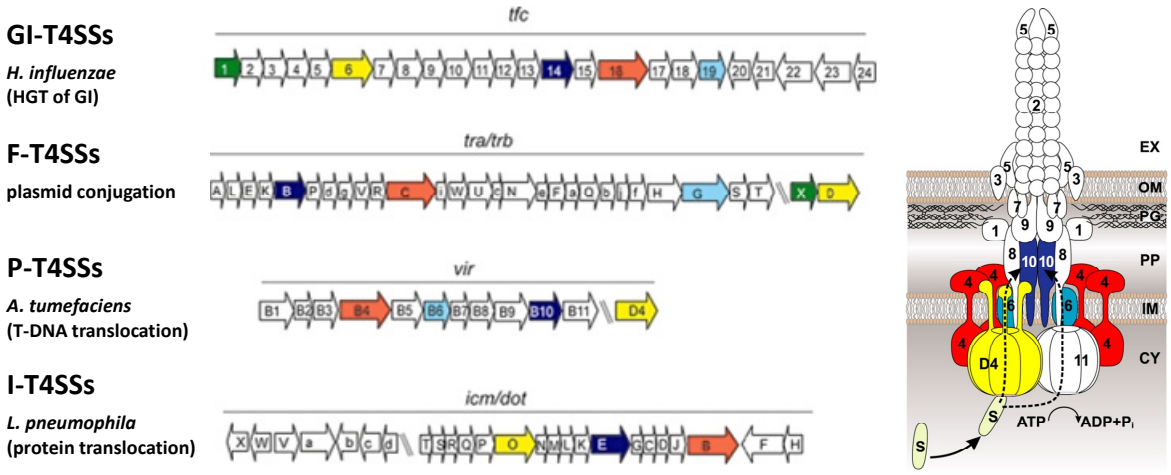


Figure 10: Different groups of T4SSs and a model of the VirB/D4 T4SS of *A. tumefaciens*. In the illustration depicting the genetic organization, genes homologous across T4SS groups are highlighted with the same color. *virB4* (red), *virB10* (blue), and *virD4* (yellow) are the only genes conserved across all T4SS groups. *tfc, tra/trb, vir*, and *icm/dot* refer to the gene names of the given loci. The same color scheme was used for the model of the VirB/D4 T4SS (adapted from (Schroder and Dehio 2005; Juhas et al. 2008)]).

Most knowledge regarding evolution, structure, and function of T4SSs comes from the type IVA systems, in particular from the prototypical VirB/VirD4 system of *A. tumefaciens* (Fig. 10). Also, type IVA are the T4SSs found in many alphaproteobacteria including *Bartonella*. Therefore, for the rest of this chapter, the focus will lie on this class of T4SSs.

The **VirB/VirD4 T4SS** of the plant pathogen *A. tumefaciens* mediates the transfer of oncogenic genes into host cells resulting in tumour-formation known as crown gall disease (Zhu et al. 2000). This system is encoded by the ~10 kb *virB* cluster comprising of 11 genes, *virB1* to *virB11*. The extracellular pilus is built up by VirB2 and VirB5 subunits. VirB4, VirB11, and VirD4 are ATPases which energize substrate secretion and are possibly involved in the structural assembly of the entire system. The localization and function of VirB3 is still unclear (Fronzes et al. 2009). VirB6 to VirB10 are structural components being part of the secretion channel (Juhas et al. 2008). As shown by crystallization and cryo-electron microscopy for the homologous T4SS of plasmid pKM191, the proteins VirB7, VirB9, and VirB10 build

up a core complex of this channel, spanning from the inner to the outer membrane (Chandran et al. 2009; Fronzes et al. 2009). VirB1 which is absent from many type IVA systems encodes a transglycosylase with a bifunctional role. It lyses the peptidoglycan cell wall and assists in the assembly of the entire T4SS within the membrane (Zupan et al. 2007). Another component of T4SSs represents VirD4 which is encoded separately from the *virB* gene cluster in *A. tumefaciens*. This inner-membrane-associated component is also referred to as coupling-protein, as it mediates the interaction between translocated effector molecules and the T4SS machinery (Fig. 10). The VirB/D4 system of *A. tumefaciens* translocates the so called T-DNA protein complex into plant cells which consists of the effector proteins VirD2 and VirE2. They are bound to a linear DNA segment that encodes the oncogenic proteins (Vergunst et al. 2000; Grange et al. 2008). Other translocated proteins are VirD5, VirE3, and VirF. Although their function is not yet understood in detail, they seem to assist in DNA-transfer and nuclear import (Lacroix et al. 2006).

Horizontal transfer has dominated the evolution of T4SSs, as congruence between tree topologies inferred from conserved housekeeping genes and different components of T4SS is mostly lacking (Frank et al. 2005). It was proposed that T4SSs dedicated to host-interaction have been derived from conjugation systems many times independently during evolution. Their high prevalence, the ease of transmission by HGT, and their functional flexibility provide the basis for a rapid diversification resulting in the adoption of a wide range of different functions (Frank et al. 2005).

1.4.2 VirB and Trw T4SSs of *Bartonella*

In *Bartonella* the two T4SSs **VirB and Trw** represent striking examples for the adaptive capabilities of these nanomachineries. Both systems belong to the type IVA systems and are built up by the same subunits as the prototypical T4SS of *A. tumefaciens* (expect for the *virB1* gene). In the *B. tribocorum*-rat model, both T4SS were shown to be essential for causing intraerythrocytic infections, a hallmark of *Bartonella* pathogenicity (Schulein and Dehio 2002; Seubert et al. 2003).

The **VirB T4SS** of *Bartonella* translocates different effector proteins into endothelial host cells to subvert cellular functions critical for establishing a chronic

infection (Schulein and Dehio 2002). Evolutionary analyses strongly imply that the VirB T4SS was horizontally acquired during evolution of *Bartonella* (Frank et al. 2005). In agreement with this, the VirB T4SS and its effector proteins are integrated into the chromosome of *B. henselae*, *B. quintana*, and *B. tribocorum* as one block bearing characteristics of a pathogenicity island (Schulein and Dehio 2002; Alsmark et al. 2004; Schulein et al. 2005). Most homology is found with conjugation systems located on conjugative plasmids from *A. tumefaciens* (pAT), *Sinorhizobium meliloti* (pSymA), and *Rhizobium etli* (p42d) (Frank et al. 2005).

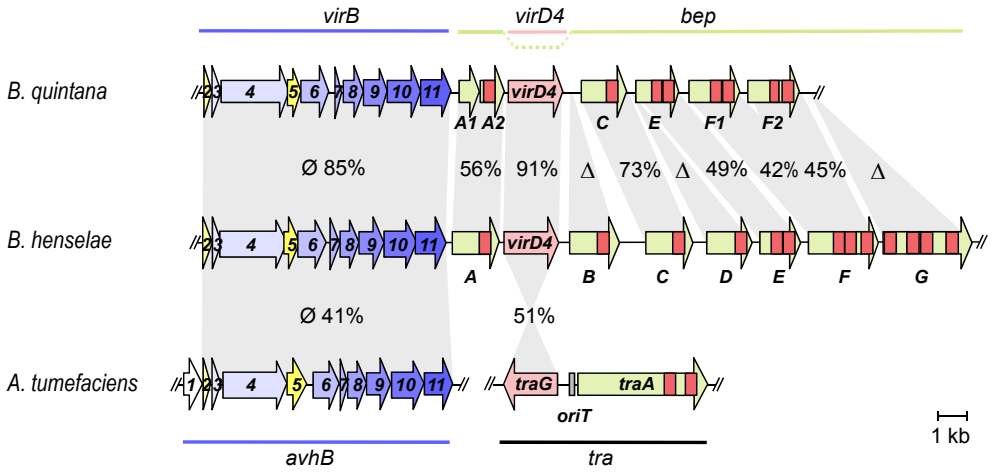


Figure 11: Genetic organization the VirB T4SS locus of *Bartonella*. Homology to the phylogenetically related conjugation system of plasmid pAT of *A. tumefaciens* is shown for comparison. Machinery encoding genes are depicted in yellow and blue, coupling proteins in pink and translocated effector molecules (relaxases and *bep* genes) in light green color. The BID domains of the relaxase and the *bep* genes are shown in red color. The degree of conservation for individual genes or groups of genes (in average, Ø) is indicated in percentages of amino acid identity of the aligned translated protein sequences (Schroder and Dehio 2005)

The transformation of the VirB T4SS from a conjugation into a protein translocation system dedicated to host adaptation required several steps (Fig. 11). Conjugative plasmids typically encode a relaxase which cleaves the plasmid DNA and interacts with the coupling protein of the T4SS, thereby enabling the plasmid transfer into the recipient cell (Schroder and Lanka 2005). This gene is not necessary for protein translocation and was deleted during the adaptive evolution of the T4SS of *Bartonella*. However, the C-terminal sequence motif of the relaxase which is sufficient for interaction with the coupling protein was adopted by *Bartonella* to target the

proteins for translocation to the machinery. All *Bartonella* effector proteins (Beps) harbor a sequence motif in their C-terminal moiety with homology to the ones found in relaxases (Fig. 11). This motif consists of the so-called *Bartonella*-intracellular delivery (BID) domain and a positively charged C-terminus (Schulein et al. 2005). Beside this secretion signal, *bep* genes were found to harbor additional BID domains, tandem-repeated tyrosine-phosphorylation motifs or FIC (filamentation-induced by cAMP) domains in their N-terminal moiety. The homology found among different Beps suggests that they evolved by duplications, diversification and reshuffling from a single ancestral effector gene which had been derived from the C-terminal moiety of the relaxase gene and the FIC domain present in many bacterial genomes (Schulein et al. 2005).

As the VirB system, the **Trw T4SS** is essential for the infectivity of different *Bartonella* species. However, its exact function is not clear yet. As neither effector proteins nor a coupling protein are found in the locus encoding the machinery genes, the Trw system seems not to translocate any substrates into host cells. Instead, it was hypothesized that this system could function as an adherence factor mediating the invasion of erythrocytes (Dehio 2008). The *trw* genes of *Bartonella* share a remarkably high similarity to the conjugation machinery located on the broad-host range plasmid R388 which was originally isolated from *E.coli* (Fig. 12). Apart from the genes encoding the T4SS structure, the two systems have also a negative heterodimeric regulator system in common. A specific feature of the Trw T4SS of *Bartonella*, and in general unique among T4SSs, is the amplification of genes coding for several structural components (Seubert et al. 2003). The gene *trwL* is present in 7 to 8 tandem-repeated copies and *trwH*, *I*, and *J*, corresponding to *virB5*, *6*, and *7*, were found to be co-amplified 2 to 3 times. The amplified copies of *trwL* as well as *trwJ* show a high degree of sequence diversification (Nystedt et al. 2008). This goes in line with the fact that both genes are encoding pilus subunits implying that their amplification and diversification may reflect antigenic variation or allow binding to different surface proteins of the erythrocytes.

In summary, these findings indicate a recent HGT of the Trw T4SS in the lineage of *Bartonella* followed by an adaptive evolution marked by deletion of genes important for substrate translocation (as relaxase and coupling protein) and the amplification-diversification of pilus subunits.

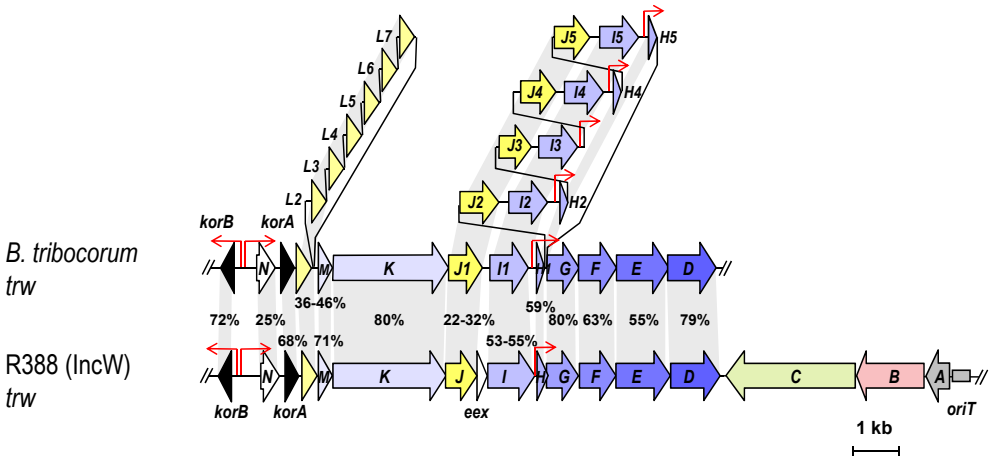


Figure 12: Genetic organization of the Trw T4SS locus of *Bartonella*. Homology to the phylogenetically related conjugation systems of plasmid R388 is shown for comparison. Machinery encoding genes are depicted in yellow and blue, coupling proteins in pink and the relaxase gene in light green color. The degree of conservation for individual genes is indicated in percentages of amino acid identity of the aligned translated protein sequences. The genes *korA* and *korB*, enccoding the heterodimeric regulator system, are depicted in black color (Schroder and Dehio 2005).

.

1.5 References

Alsmark, C. M., A. C. Frank, et al. (2004). "The louse-borne human pathogen Bartonella quintana is a genomic derivative of the zoonotic agent Bartonella henselae." Proc Natl Acad Sci U S A 101(26): 9716-9721.

Anderson, P. and J. Roth (1981). "Spontaneous tandem genetic duplications in Salmonella typhimurium arise by unequal recombination between rRNA (rrn) cistrons." Proc Natl Acad Sci U S A 78(5): 3113-3117.

Andersson, D. I. and D. Hughes (2009). "Gene Amplification and Adaptive Evolution in Bacteria." Annu Rev Genet.

Andersson, S. G. (2000). "The genomics gamble." Nat Genet 26(2): 134-135.

Andersson, S. G. (2006). "Genetics. The bacterial world gets smaller." Science 314(5797): 259-260.

Andersson, S. G., A. Zomorodipour, et al. (1998). "The genome sequence of Rickettsia prowazekii and the origin of mitochondria." Nature 396(6707): 133-140.

Aras, R. A., J. Kang, et al. (2003). "Extensive repetitive DNA facilitates prokaryotic genome plasticity." Proc Natl Acad Sci U S A 100(23): 13579-13584.

Azzouna, A., P. Greve, et al. (2004). "Sexual differentiation traits in functional males with female genital apertures (male symbol fga) in the woodlice Armadillidium vulgare Latr. (Isopoda, Crustacea)." Gen Comp Endocrinol 138(1): 42-49.

Backert, S. and T. F. Meyer (2006). "Type IV secretion systems and their effectors in bacterial pathogenesis." Curr Opin Microbiol 9(2): 207-217.

Bao, W., Y. Kumagai, et al. (2009). "Four VirB6 paralogs and VirB9 are expressed and interact in Ehrlichia chaffeensis-containing vacuoles." J Bacteriol 191(1): 278-286.

Barquero-Calvo, E., E. Chaves-Olarte, et al. (2007). "Brucella abortus uses a stealthy strategy to avoid activation of the innate immune system during the onset of infection." PLoS One 2(7): e631.

Barrick, J. E., D. S. Yu, et al. (2009). "Genome evolution and adaptation in a long-term experiment with Escherichia coli." Nature 461(7268): 1243-1247.

Bartsev, A. V., W. J. Deakin, et al. (2004). "NopL, an effector protein of Rhizobium sp. NGR234, thwarts activation of plant defense reactions." Plant Physiol 134(2): 871-879.

Batut, J., S. G. Andersson, et al. (2004). "The evolution of chronic infection strategies in the alpha-proteobacteria." Nat Rev Microbiol 2(12): 933-945.

Berglund, E. C., A. C. Frank, 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(7): e1000546.

Bergthorsson, U., D. I. Andersson, et al. (2007). "Ohno's dilemma: evolution of new genes under continuous selection." Proc Natl Acad Sci U S A 104(43): 17004-17009.

Boddicker, J. D., N. A. Ledeboer, et al. (2002). "Differential binding to and biofilm formation on, HEp-2 cells by Salmonella enterica serovar Typhimurium is dependent upon allelic variation in the fimH gene of the fim gene cluster." Mol Microbiol 45(5): 1255-1265.

Boller, T. and S. Y. He (2009). "Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens." Science 324(5928): 742-744.

Boussau, B., E. O. Karlberg, et al. (2004). "Computational inference of scenarios for alpha-proteobacterial genome evolution." Proc Natl Acad Sci U S A 101(26): 9722-9727.

Brown, J. R. (2003). "Ancient horizontal gene transfer." Nat Rev Genet 4(2): 121-132.

Bruggemann, H., C. Cazalet, et al. (2006). "Adaptation of Legionella pneumophila to the host environment: role of protein secretion, effectors and eukaryotic-like proteins." Curr Opin Microbiol 9(1): 86-94.

Buchrieser, C., R. Brosch, et al. (1998). "The high-pathogenicity island of Yersinia pseudotuberculosis can be inserted into any of the three chromosomal asn tRNA genes." Mol Microbiol 30(5): 965-978.

Burrus, V. and M. K. Waldor (2004). "Shaping bacterial genomes with integrative and conjugative elements." Res Microbiol 155(5): 376-386.

Casadevall, A. (2008). "Evolution of intracellular pathogens." Annu Rev Microbiol 62: 19-33.

Cascales, E. and P. J. Christie (2003). "The versatile bacterial type IV secretion systems." Nat Rev Microbiol 1(2): 137-149.

Chambers, H. F. and F. R. Deleo (2009). "Waves of resistance: Staphylococcus aureus in the antibiotic era." Nat Rev Microbiol 7(9): 629-641.

Chandran, V., R. Fronzes, et al. (2009). "Structure of the outer membrane complex of a type IV secretion system." Nature 462(7276): 1011-1015.

Chen, S. L., C. S. Hung, et al. (2006). "Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach." Proc Natl Acad Sci U S A 103(15): 5977-5982.

Cho, N. H., H. R. Kim, et al. (2007). "The Orientia tsutsugamushi genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes." Proc Natl Acad Sci U S A 104(19): 7981-7986.

Christie, P. J. (2004). "Type IV secretion: the Agrobacterium VirB/D4 and related conjugation systems." Biochim Biophys Acta 1694(1-3): 219-234.

Christie, P. J., K. Atmakuri, et al. (2005). "Biogenesis, architecture, and function of bacterial type IV secretion systems." Annu Rev Microbiol 59: 451-485.

Clark, M. A., N. A. Moran, et al. (1999). "Sequence evolution in bacterial endosymbionts having extreme base compositions." Mol Biol Evol 16(11): 1586-1598.

Coburn, B., I. Sekirov, et al. (2007). "Type III secretion systems and disease." Clin Microbiol Rev 20(4): 535-549.

Cole, S. T., K. Eiglmeier, et al. (2001). "Massive gene decay in the leprosy bacillus." Nature 409(6823): 1007-1011.

Conley, T., L. Slater, et al. (1994). "Rochalimaea species stimulate human endothelial cell proliferation and migration in vitro." J Lab Clin Med 124(4): 521-528.

Dale, C., S. A. Young, et al. (2001). "The insect endosymbiont Sodalis glossinidius utilizes a type III secretion system for cell invasion." Proc Natl Acad Sci U S A 98(4): 1883-1888.

Darby, A. C., N. H. Cho, et al. (2007). "Intracellular pathogens go extreme: genome evolution in the Rickettsiales." Trends Genet 23(10): 511-520.

Davison, J. (1999). "Genetic exchange between bacteria in the environment." Plasmid 42(2): 73-91.

Dehio, C. (2008). "Infection-associated type IV secretion systems of Bartonella and their diverse roles in host cell interaction." Cell Microbiol 10(8): 1591-1598.

Denamur, E. and I. Matic (2006). "Evolution of mutation rates in bacteria." Mol Microbiol 60(4): 820-827.

Deppenmeier, U., A. Johann, et al. (2002). "The genome of Methanosarcina mazei: evidence for lateral gene transfer between bacteria and archaea." J Mol Microbiol Biotechnol 4(4): 453-461.

Deurenberg, R. H., C. Vink, et al. (2007). "The molecular evolution of methicillin-resistant Staphylococcus aureus." Clin Microbiol Infect 13(3): 222-235.

Dobrindt, U., B. Hochhut, et al. (2004). "Genomic islands in pathogenic and environmental microorganisms." Nat Rev Microbiol 2(5): 414-424.

Endre, G., A. Kereszt, et al. (2002). "A receptor kinase gene regulating symbiotic nodule development." Nature 417(6892): 962-966.

Esser, C., W. Martin, et al. (2007). "The origin of mitochondria in light of a fluid prokaryotic chromosome model." Biol Lett 3(2): 180-184.

Ettema, T. J. and S. G. Andersson (2009). "The alpha-proteobacteria: the Darwin finches of the bacterial world." Biol Lett 5(3): 429-432.

Felsenstein, J. (1974). "The evolutionary advantage of recombination." Genetics 78(2): 737-756.

Fischbach, M. A. and C. T. Walsh (2009). "Antibiotics for emerging pathogens." Science 325(5944): 1089-1093.

Fleischmann, R. D., M. D. Adams, et al. (1995). "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd." Science 269(5223): 496-512.

Frank, A. C., C. M. Alsmark, et al. (2005). "Functional divergence and horizontal transfer of type IV secretion systems." Mol Biol Evol 22(5): 1325-1336.

Fraser-Liggett, C. M. (2005). "Insights on biology and evolution from microbial genome sequencing." Genome Res 15(12): 1603-1610.

Fraser, C., E. J. Alm, et al. (2009). "The bacterial species challenge: making sense of genetic and ecological diversity." Science 323(5915): 741-746.

Fraser, C. M., S. Casjens, et al. (1997). "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi." Nature 390(6660): 580-586.

Fronzes, R., P. J. Christie, et al. (2009). "The structural biology of type IV secretion systems." Nat Rev Microbiol 7(10): 703-714.

Fronzes, R., E. Schafer, et al. (2009). "Structure of a type IV secretion system core complex." Science 323(5911): 266-268.

Gal-Mor, O. and B. B. Finlay (2006). "Pathogenicity islands: a molecular toolbox for bacterial virulence." Cell Microbiol 8(11): 1707-1719.

Gevers, D., F. M. Cohan, et al. (2005). "Opinion: Re-evaluating prokaryotic species." Nat Rev Microbiol 3(9): 733-739.

Gevers, D., K. Vandepoele, et al. (2004). "Gene duplication and biased functional retention of paralogs in bacterial genomes." Trends Microbiol 12(4): 148-154.

Gillespie, J. J., N. C. Ammerman, et al. (2009). "An anomalous type IV secretion system in Rickettsia is evolutionarily conserved." PLoS One 4(3): e4833.

Giovannoni, S. J. and U. Stingl (2005). "Molecular diversity and ecology of microbial plankton." Nature 437(7057): 343-348.

Giraud, A., I. Matic, et al. (2001). "Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut." Science 291(5513): 2606-2608.

Giuntini, E., A. Mengoni, et al. (2005). "Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of Sinorhizobium meliloti natural strains." BMC Genomics 6: 158.

Grange, W., M. Duckely, et al. (2008). "VirE2: a unique ssDNA-compacting molecular machine." PLoS Biol 6(2): e44.

Groisman, E. A. and H. Ochman (1996). "Pathogenicity islands: bacterial evolution in quantum leaps." Cell 87(5): 791-794.

Gross, A., A. Terraza, et al. (2000). "In vitro Brucella suis infection prevents the programmed cell death of human monocytic cells." Infect Immun 68(1): 342-351.

Hacker, J., U. Hentschel, et al. (2003). "Prokaryotic chromosomes and disease." Science 301(5634): 790-793.

Hacker, J. and J. B. Kaper (2000). "Pathogenicity islands and the evolution of microbes." Annu Rev Microbiol 54: 641-679.

Hamilton, H. L., N. M. Dominguez, et al. (2005). "Neisseria gonorrhoeae secretes chromosomal DNA via a novel type IV secretion system." Mol Microbiol 55(6): 1704-1721.

Hayashi, T., K. Makino, et al. (2001). "Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12." DNA Res 8(1): 11-22.

Hill, S. A. and J. K. Davies (2009). "Pilin gene variation in Neisseria gonorrhoeae: reassessing the old paradigms." FEMS Microbiol Rev 33(3): 521-530.

Hiller, N. L., B. Janto, et al. (2007). "Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome." J Bacteriol 189(22): 8186-8195.

Hueck, C. J. (1998). "Type III protein secretion systems in bacterial pathogens of animals and plants." Microbiol Mol Biol Rev 62(2): 379-433.

Jordan, I. K., I. B. Rogozin, et al. (2002). "Microevolutionary genomics of bacteria." Theor Popul Biol 61(4): 435-447.

Joshi, S. G., C. W. Francis, et al. (2003). "Nuclear factor kappa B protects against host cell apoptosis during Rickettsia rickettsii infection by inhibiting activation of apical and effector caspases and maintaining mitochondrial integrity." Infect Immun 71(7): 4127-4136.

Juhas, M., D. W. Crook, et al. (2007). "Novel type IV secretion system involved in propagation of genomic islands." J Bacteriol 189(3): 761-771.

Juhas, M., D. W. Crook, et al. (2008). "Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence." Cell Microbiol 10(12): 2377-2386.

Juhas, M., J. R. van der Meer, et al. (2009). "Genomic islands: tools of bacterial horizontal gene transfer and evolution." FEMS Microbiol Rev 33(2): 376-393.

Kaneko, T., Y. Nakamura, et al. (2002). "Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110." DNA Res 9(6): 189-197.

Kirby, J. E. and D. M. Nekorchuk (2002). "Bartonella-associated endothelial proliferation depends on inhibition of apoptosis." Proc Natl Acad Sci U S A 99(7): 4656-4661.

Koehler, J. E., F. D. Quinn, et al. (1992). "Isolation of Rochalimaea species from cutaneous and osseous lesions of bacillary angiomatosis." N Engl J Med 327(23): 1625-1631.

Konstantinidis, K. T. and J. M. Tiedje (2005). "Genomic insights that advance the species definition for prokaryotes." Proc Natl Acad Sci U S A 102(7): 2567-2572.

Koonin, E. V. and Y. I. Wolf (2008). "Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world." Nucleic Acids Res 36(21): 6688-6719.

Lacroix, B., J. Li, et al. (2006). "Will you let me use your nucleus? How Agrobacterium gets its T-DNA expressed in the host plant cell." Can J Physiol Pharmacol 84(3-4): 333-345.

Lawley, T. D., W. A. Klimke, et al. (2003). "F factor conjugation is a true type IV secretion system." FEMS Microbiol Lett 224(1): 1-15.

Lawrence, J. G. (1999). "Gene transfer, speciation, and the evolution of bacterial genomes." Curr Opin Microbiol 2(5): 519-523.

Lawrence, J. G. (2005). "Common themes in the genome strategies of pathogens." Curr Opin Genet Dev 15(6): 584-588.

Lawrence, J. G. and H. Hendrickson (2005). "Genome evolution in bacteria: order beneath chaos." Curr Opin Microbiol 8(5): 572-578.

Lawrence, J. G. and H. Ochman (1997). "Amelioration of bacterial genomes: rates of change and exchange." J Mol Evol 44(4): 383-397.

Liu, S. L. and K. E. Sanderson (1996). "Highly plastic chromosomal organization in Salmonella typhi." Proc Natl Acad Sci U S A 93(19): 10303-10308.

Margulies, M., M. Egholm, et al. (2005). "Genome sequencing in microfabricated high-density picolitre reactors." Nature 437(7057): 376-380.

Marie, C., W. J. Broughton, et al. (2001). "Rhizobium type III secretion systems: legume charmers or alarmers?" Curr Opin Plant Biol 4(4): 336-342.

Maurelli, A. T. (2007). "Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens." FEMS Microbiol Lett 267(1): 1-8.

McClelland, M., K. E. Sanderson, et al. (2004). "Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid." Nat Genet 36(12): 1268-1274.

McCutcheon, J. P., B. R. McDonald, et al. (2009). "Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont." PLoS Genet 5(7): e1000565.

Medini, D., A. Covacci, et al. (2006). "Protein homology network families reveal step-wise diversification of Type III and Type IV secretion systems." PLoS Comput Biol 2(12): e173.

Medini, D., D. Serruto, et al. (2008). "Microbiology in the post-genomic era." Nat Rev Microbiol 6(6): 419-430.

Merrell, D. S. and S. Falkow (2004). "Frontal and stealth attack strategies in microbial pathogenesis." Nature 430(6996): 250-256.

Mira, A., L. Klasson, et al. (2002). "Microbial genome evolution: sources of variability." Curr Opin Microbiol 5(5): 506-512.

Mira, A., H. Ochman, et al. (2001). "Deletional bias and the evolution of bacterial genomes." Trends Genet 17(10): 589-596.

Moran, N. A. (1996). "Accelerated evolution and Muller's rachet in endosymbiotic bacteria." Proc Natl Acad Sci U S A 93(7): 2873-2878.

Moran, N. A. (2002). "Microbial minimalism: genome reduction in bacterial pathogens." Cell 108(5): 583-586.

Moran, N. A., H. J. McLaughlin, et al. (2009). "The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria." Science 323(5912): 379-382.

Moran, N. A. and G. R. Plague (2004). "Genomic changes following host restriction in bacteria." Curr Opin Genet Dev 14(6): 627-633.

Moxon, E. R. and D. S. Thaler (1997). "Microbial genetics. The tinkerer's evolving tool-box." Nature 387(6634): 659, 661-652.

Mueller, C. A., P. Broz, et al. (2008). "The type III secretion system tip complex and translocon." Mol Microbiol 68(5): 1085-1095.

Murphy, P. M. (1993). "Molecular mimicry and the generation of host defense protein diversity." Cell 72(6): 823-826.

Nakagawa, I., K. Kurokawa, et al. (2003). "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution." Genome Res 13(6A): 1042-1055.

Nakata, N., T. Tobe, et al. (1993). "The absence of a surface protease, OmpT, determines the intercellular spreading ability of Shigella: the relationship between the ompT and kcpA loci." Mol Microbiol 9(3): 459-468.

Nirgianaki, A., G. K. Banks, et al. (2003). "Wolbachia infections of the whitefly Bemisia tabaci." Curr Microbiol 47(2): 93-101.

Nystedt, B., A. C. Frank, et al. (2008). "Diversifying selection and concerted evolution of a type IV secretion system in Bartonella." Mol Biol Evol 25(2): 287-300.

Ochman, H., J. G. Lawrence, et al. (2000). "Lateral gene transfer and the nature of bacterial innovation." Nature 405(6784): 299-304.

Oliver, A., R. Canton, et al. (2000). "High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection." Science 288(5469): 1251-1254.

Pallen, M. J. and B. W. Wren (2007). "Bacterial pathogenomics." Nature 449(7164): 835-842.

Palmer, G. H. and K. A. Brayton (2007). "Gene conversion is a convergent strategy for pathogen antigenic variation." Trends Parasitol 23(9): 408-413.

Petersen, L., J. P. Bollback, et al. (2007). "Genes under positive selection in Escherichia coli." Genome Res 17(9): 1336-1343.

Preston, A., J. Parkhill, et al. (2004). "The bordetellae: lessons from genomics." Nat Rev Microbiol 2(5): 379-390.

Rances, E., D. Voronin, et al. (2008). "Genetic and functional characterization of the type IV secretion system in Wolbachia." J Bacteriol 190(14): 5020-5030.

Rocha, E. P. (2004). "Order and disorder in bacterial genomes." Curr Opin Microbiol 7(5): 519-527.

Rocha, E. P. (2004). "The replication-related organization of bacterial genomes." Microbiology 150(Pt 6): 1609-1627.

Rocha, E. (2008). "Evolutionary patterns in prokaryotic genomes." Curr Opin Microbiol 11(5): 454-460.

Sallstrom, B. and S. G. Andersson (2005). "Genome reduction in the alpha-Proteobacteria." Curr Opin Microbiol 8(5): 579-585.

Salyers, A. A., N. B. Shoemaker, et al. (1995). "Conjugative transposons: an unusual and diverse set of integrated gene transfer elements." Microbiol Rev 59(4): 579-590.

Sandegren, L. and D. I. Andersson (2009). "Bacterial gene amplification: implications for the evolution of antibiotic resistance." Nat Rev Microbiol 7(8): 578-588.

Schmid, M. C., F. Scheidegger, et al. (2006). "A translocated bacterial protein protects vascular endothelial cells from apoptosis." PLoS Pathog 2(11): e115.

Schmid, M. C., R. Schulein, et al. (2004). "The VirB type IV secretion system of Bartonella henselae mediates invasion, proinflammatory activation and antiapoptotic protection of endothelial cells." Mol Microbiol 52(1): 81-92.

Schmidt, H. and M. Hensel (2004). "Pathogenicity islands in bacterial pathogenesis." Clin Microbiol Rev 17(1): 14-56.

Schroder, G. and C. Dehio (2005). "Virulence-associated type IV secretion systems of Bartonella." Trends Microbiol 13(7): 336-342.

Schroder, G. and E. Lanka (2005). "The mating pair formation system of conjugative plasmids-A versatile secretion machinery for transfer of proteins and DNA." Plasmid 54(1): 1-25.

Schulein, R. and C. Dehio (2002). "The VirB/VirD4 type IV secretion system of Bartonella is essential for establishing intraerythrocytic infection." Mol Microbiol 46(4): 1053-1067.

Schulein, R., P. Guye, et al. (2005). "A bipartite signal mediates the transfer of type IV secretion substrates of Bartonella henselae into human cells." Proc Natl Acad Sci U S A 102(3): 856-861.

Schuster, S. C. (2008). "Next-generation sequencing transforms today's biology." Nat Methods 5(1): 16-18.

Segal, G., M. Feldman, et al. (2005). "The Icm/Dot type-IV secretion systems of Legionella pneumophila and Coxiella burnetii." FEMS Microbiol Rev 29(1): 65-81.

Seifert, H. S. and V. J. DiRita (2006). Evolution of microbial pathogens. Washington, D.C., ASM Press.

Seubert, A., R. Hiestand, et al. (2003). "A bacterial conjugation machinery recruited for pathogenesis." Mol Microbiol 49(5): 1253-1266.

She, Q., R. K. Singh, et al. (2001). "The complete genome of the crenarchaeon Sulfolobus solfataricus P2." Proc Natl Acad Sci U S A 98(14): 7835-7840.

Sokurenko, E. V., V. Chesnokova, et al. (1998). "Pathogenic adaptation of Escherichia coli by natural variation of the FimH adhesin." Proc Natl Acad Sci U S A 95(15): 8922-8926.

Springael, D. and E. M. Top (2004). "Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies." Trends Microbiol 12(2): 53-58.

Stavrinides, J., H. C. McCann, et al. (2008). "Host-pathogen interplay and the evolution of bacterial effectors." Cell Microbiol 10(2): 285-292.

Sullivan, J. T. and C. W. Ronson (1998). "Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene." Proc Natl Acad Sci U S A 95(9): 5145-5149.

Taddei, F., I. Matic, et al. (1997). "To be a mutator, or how pathogenic and commensal bacteria can evolve rapidly." Trends Microbiol 5(11): 427-428; discussion 428-429.

Tamas, I., L. Klasson, et al. (2002). "50 million years of genomic stasis in endosymbiotic bacteria." Science 296(5577): 2376-2379.

Tettelin, H., V. Masignani, et al. (2005). "Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome"." Proc Natl Acad Sci U S A 102(39): 13950-13955.

Thomas, C. M. and K. M. Nielsen (2005). "Mechanisms of, and barriers to, horizontal gene transfer between bacteria." Nat Rev Microbiol 3(9): 711-721.

Tillier, E. R. and R. A. Collins (2000). "Genome rearrangement by replication-directed translocation." Nat Genet 26(2): 195-197.

van der Woude, M. W. and A. J. Baumler (2004). "Phase and antigenic variation in bacteria." Clin Microbiol Rev 17(3): 581-611, table of contents.

Venter, J. C., K. Remington, et al. (2004). "Environmental genome shotgun sequencing of the Sargasso Sea." Science 304(5667): 66-74.

Vergunst, A. C., B. Schrammeijer, et al. (2000). "VirB/D4-dependent protein translocation from Agrobacterium into plant cells." Science 290(5493): 979-982.

Viprey, V., X. Perret, et al. (2000). "Host-plant invasion by rhizobia." Subcell Biochem 33: 437-456.

Weissman, S. J., S. L. Moseley, et al. (2003). "Enterobacterial adhesins and the case for studying SNPs in bacteria." Trends Microbiol 11(3): 115-117.

Welch, R. A., V. Burland, et al. (2002). "Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli." Proc Natl Acad Sci U S A 99(26): 17020-17024.

Wolf, Y., I. B. Rogozin, et al. (2002). "Genome trees and the tree of life." Trends Genet 18(9): 472-479.

Wren, B. W. (2003). "The yersiniae--a model genus to study the rapid evolution of bacterial pathogens." Nat Rev Microbiol 1(1): 55-64.

Wu, M., L. V. Sun, et al. (2004). "Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements." PLoS Biol 2(3): E69.

Yang, F., J. Yang, et al. (2005). "Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery." Nucleic Acids Res 33(19): 6445-6458.

Zahringer, U., B. Lindner, et al. (2004). "Structure and biological activity of the short-chain lipopolysaccharide from Bartonella henselae ATCC 49882T." J Biol Chem 279(20): 21046-21054.

Zheng, Y., R. J. Roberts, et al. (2004). "Segmentally variable genes: a new perspective on adaptation." PLoS Biol 2(4): E81.

Zhu, J., P. M. Oger, et al. (2000). "The bases of crown gall tumorigenesis." J Bacteriol 182(14): 3885-3895.

Zupan, J., C. A. Hackworth, et al. (2007). "VirB1* promotes T-pilus formation in the vir-Type IV secretion system of Agrobacterium tumefaciens." J Bacteriol 189(18): 6551-6563.

2. Aim of the Thesis

2. Aim of the Thesis

When the project was started in March 2006, the primary aim of this thesis was to understand functional and evolutionary aspects of the infection strategy adopted by the genus *Bartonella*. The main focus was the analysis of different type IV secretion systems of *Bartonella* involved in the interaction with the mammalian host. To this end, I generated genome-wide data sets, applied methods of comparative genomics and molecular evolution, and I used *in vivo* and *in vitro* approaches to corroborate findings from dry-lab analyses.

3. Results

3.1 Research article I

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

Henri L Saenz*, Philipp Engel*, Michèle C Stoeckli, Christa Lanz, Günter Raddatz, Muriel Vayssier-Taussat, Richard Birtles, Stephan C Schuster, and Christoph Dehio

*These authors contributed equally to this work

Nature Genetics, Volume 39, Number 12, December 2007, 1469-1476.

3.1.1 Summary

Evolutionary processes underlying the transformation of prototypical pathogens causing massive host damage into host-adapted pathogens with attenuated virulence are poorly understood. As niche competition normally leads to the extinction of prototypical pathogens, information necessary to understand the evolutionary path by which successfully adapted descendants were selected is lost.

In this study, we have inferred that the horizontal acquisition of different type IV secretion systems (T4SSs) facilitated *Bartonella* to adjust its characteristic infection strategy, which is reflected by an increased host adaptability of the descendants.

Our phylogenetic analysis based on four housekeeping genes implicated that bartonellae with different virulence potential form distinct phylogenetic clades (Saenz et al. 2007). The analysis identified a deep-branching ancestral lineage, which solely comprises the highly virulent human pathogen *B. bacilliformis*. Most other *Bartonella* species causing mild or asymptomatic intraerythrocytic infections in various mammalian reservoir hosts are descendants of rapidly multiplying lineages. These occurred after the divergence from the ancestral lineage of *B. bacilliformis*. To identify pathogenicity genes (PGs) that could have facilitated host adaptability within these lineages, an integrated approach of comparative and functional genomics was used. We applied signature-tagged mutagenesis to screen for abacteremic mutants in the *B. tribocorum*-rat model. By this means, 97 PGs of *B. tribocorum*, which are essential for the intraerythrocytic colonization of the rat host, were identified. Complete sequencing of the genome of *B. tribocorum* allowed chromosomal mapping of the PGs and comparison to the publicly available genomes of *B. bacilliformis*, *B. quintana*, and *B. henselae*. Most of the identified PGs were found to be conserved in all four genomes, reflecting the common infection strategies used by bartonellae to colonize their different reservoir hosts. However, 15 PGs of *B. tribocorum* were present in *B. quintana* and *B. henselae*, but absent from *B. bacilliformis*. Except for one, these PGs encode different components of two horizontally acquired T4SSs, VirB and Trw, both having adopted essential but different functions during infection. A comprehensive screening for T4SS genes by PCR amplification and DNA hybridization showed that T4SS are present in all of the analyzed species belonging to the lineages with increased host adaptability, but not found in any tested strain of

B. bacilliformis. Further, the absence of T4SS gene remnants from the genome of *B. bacilliformis* suggested a horizontal acquisition of these factors after the divergence of the ancestral lineage. Interestingly, in the sub-clade of ruminant-infecting *Bartonella* species none of the *virB* or *trw* genes could be detected. However, we identified components of a VirB-homologous (Vbh) T4SS. By phylogenetic analysis, we showed that the VirB and Vbh T4SSs have presumably been derived from each other suggesting similar functional roles in different *Bartonella* lineages.

Summarizing, our results showed that T4SSs, horizontally acquired during evolution, resulted in the fine-tuning of an infection strategy already adopted by the common ancestor of *Bartonella*. Whereas the Trw T4SS is only present in a subset of species, a VirB-like T4SS (either VirB or Vbh) was found in all analyzed strains with increased host adaptability. By providing a nanomachine dedicated to the translocation of a cocktail of different effector proteins into mammalian target cells, the VirB T4SS represents the appropriate molecular tool to confer host adaptability. This is reflected by the high degree of variation found between the effector protein-encoding genes of *Bartonella* species adapted to different hosts. The fine-tuning of the interaction with the mammalian host has presumably enabled *Bartonella* species to adjust their virulence and broaden their dissemination. Consequently, this resulted in the colonization of a much wider range of different hosts.

Statement of the own participation

I contributed to this publication by analyzing different *Bartonella* species for presence/absence of T4SSs. I performed the comparative genome analysis of *B. tribocorum*, *B. henselae*, *B. quintana*, and *B. bacilliformis*. Moreover, I carried out all phylogenetic analyses and also discovered the Vbh T4SS in the ruminant-infecting *Bartonella* species. The genome of *B. tribocorum* was sequenced and annotated by the genome project team consisting of Christoph Dehio, Christa Lanz, Günter Raddatz, Henri Saenz, Stephan Schuster, and me. The assembly of the *B. tribocorum* genome was achieved during my Master studies. The genome-wide mutant screen by signature-tagged mutagenesis was carried out by Henri Saenz. The manuscript was written by me, Henri Saenz, and Christoph Dehio.

nature genetics

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

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

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[1–5]. 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'[6,7]. With a restricted distribution in the Andes region, this 'living fossil' is a severe human pathogen causing up to 85% mortality[8]. 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[7]. 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[6,9,10] (**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[11,12]. 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[13,14]. T4SSs are macromolecular transporters ancestrally related to bacterial conjugation systems. Several mammalian pathogens have adopted T4SSs to translocate effector proteins into host cells[15]; for example, the VirB T4SS of *B. henselae* injects seven effectors into human endothelial cells to subvert cellular functions[16]. Whether VirB, Trw or any other pathogenicity factor[11] 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

[1]Focal Area Infection Biology, Biozentrum, University of Basel, CH-4056 Basel, Switzerland. [2]Max Planck Institute for Developmental Biology, D-72076 Tuebingen, Germany. [3]Unité de Biologie Moléculaire et Immunologie Parasitaire et Fongique, Institut Scientifique de Recherche Agronomique, Ecole Nationale Veterinaire d'Alfort, F-94700 Maisons-Alfort, France. [4]Department of Veterinary Pathology, Faculty of Veterinary Science, University of Liverpool, Liverpool L69 7LB, UK. [5]Center for Infectious Disease Dynamics and Center for Comparative Genomics and Bioinformatics, Penn State University, University Park, Pennsylvania 16802, USA. [6]Present address: Magnetic Resonance Center, Max Planck Institute for Biological Cybernetics, D-72076 Tuebingen, Germany. [7]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[12]. 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*[17] and *B. quintana*[17]) 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^T by the Sanger method at 11-fold coverage using the shotgun approach[18] 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*[17] (1.93 Mb) and *B. quintana*[17] (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[19]. 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[20] 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

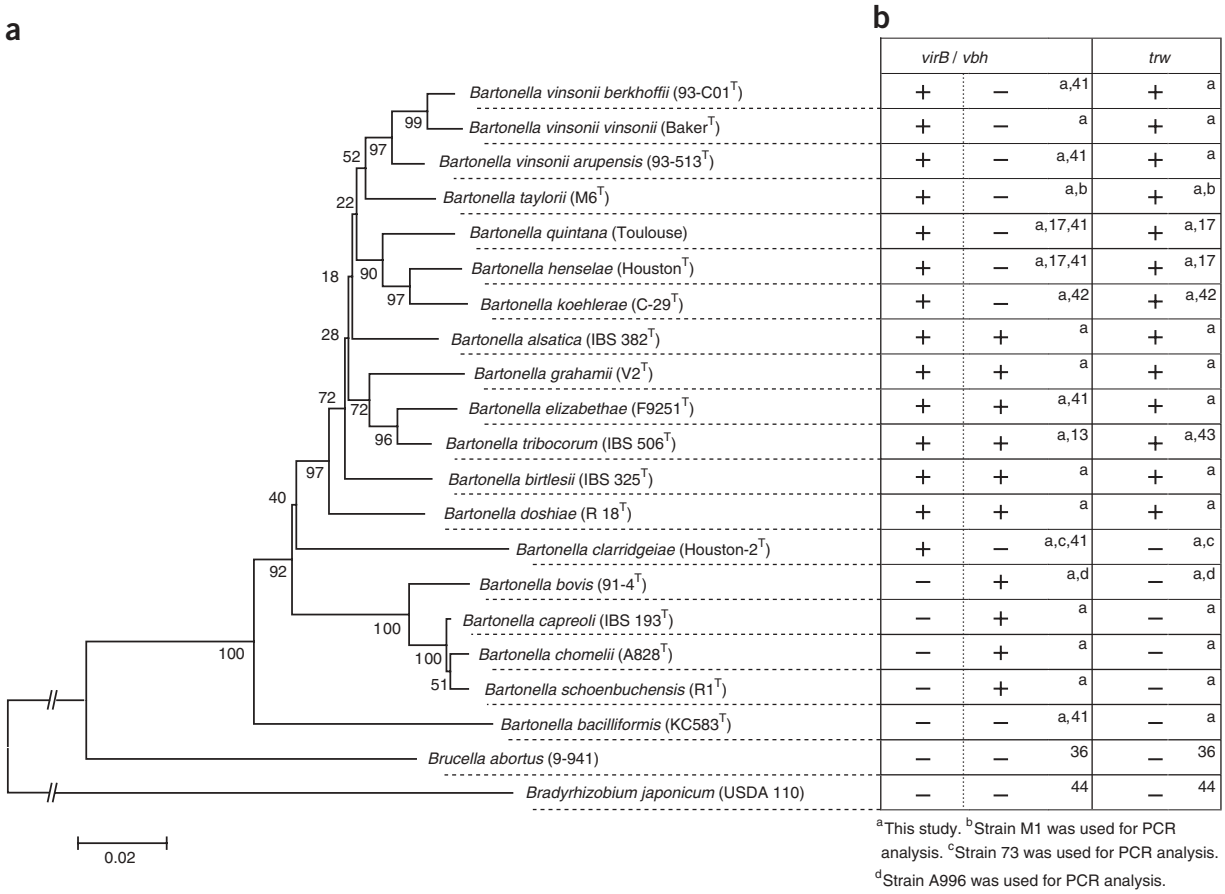


Figure 1 Phylogenetic tree of *Bartonella* based on multilocus sequence analysis and summary table of the presence or absence of loci encoding the T4SSs VirB, VirB-homolog (Vbh) and Trw in the different *Bartonella* species. (**a**) The phylogenetic tree shown was calculated on the basis of concatenated 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.

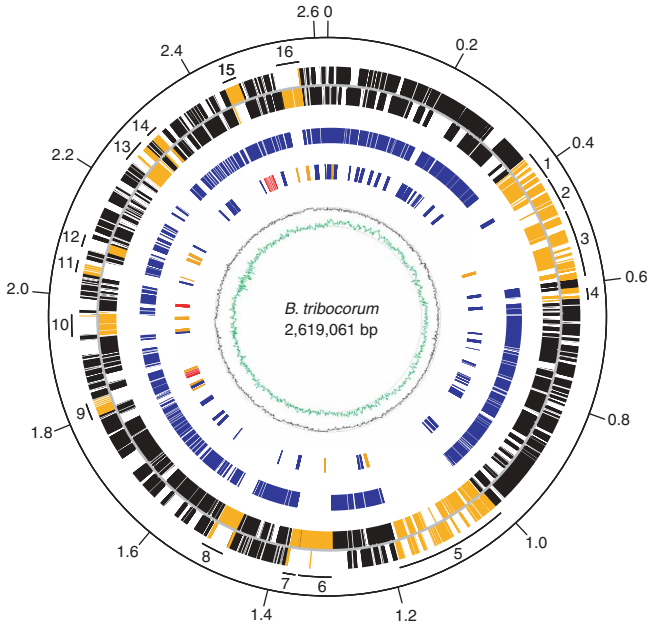


Figure 2 Circular genome map of *B. tribocorum*. From the outside, the first two circles indicate genes on the + and – strands. Genomic islands are highlighted in orange and numbered according to **Supplementary Table 2**. The third circle (blue) shows the core genome of the radiating lineage as determined 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*[17], 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)[21] to *B. tribocorum*. Individually tagged mutants assembled in pools were tested in the rat-infection model[12] 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*

	B. tribocorum[a]	*B. henselae*	*B. quintana*
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

[a]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[22]. 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**)[14,16]. 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*

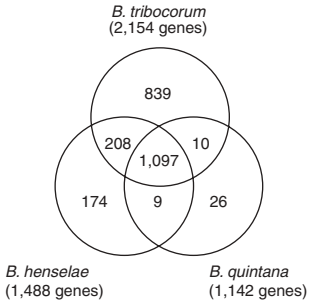


Figure 3 Core genome and accessory genomes of three species of the radiating *Bartonella* lineage determined on the basis of ortholog gene sets (**Supplementary Table 1**). Numbers in parentheses indicate the total number of protein-coding genes.

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
B. tribocorum genome	2,154	97
Core genome of the radiating lineage:	1,097	81
with homolog in *B. bacilliformis*	959	66
without homolog in *B. bacilliformis*	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 α-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)[23] 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 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)[24,25] (**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)[23], 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

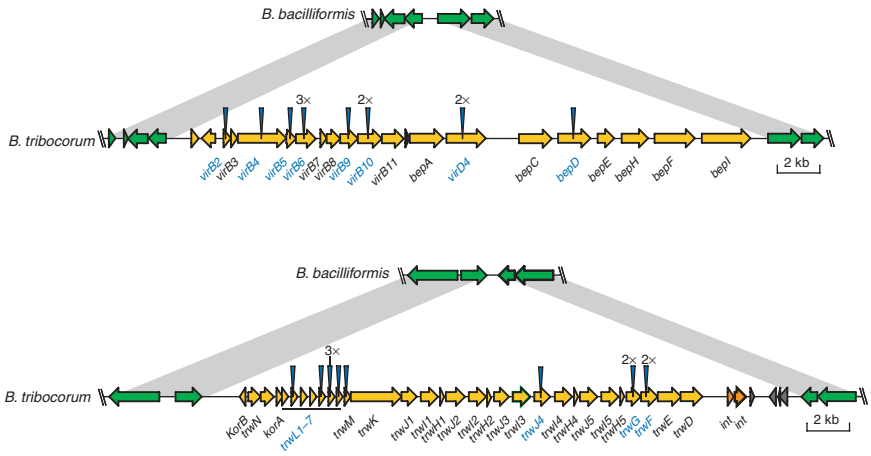


Figure 4 STM mutant insertion sites in the T4SS loci *virB* and *trw* of *B. tribocorum* and comparison of the flanking region with *B. bacilliformis*. Genes belonging to one of the two T4SSs are shown in yellow. STM mutant insertion sites are indicated as blue triangles and the corresponding gene names 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.

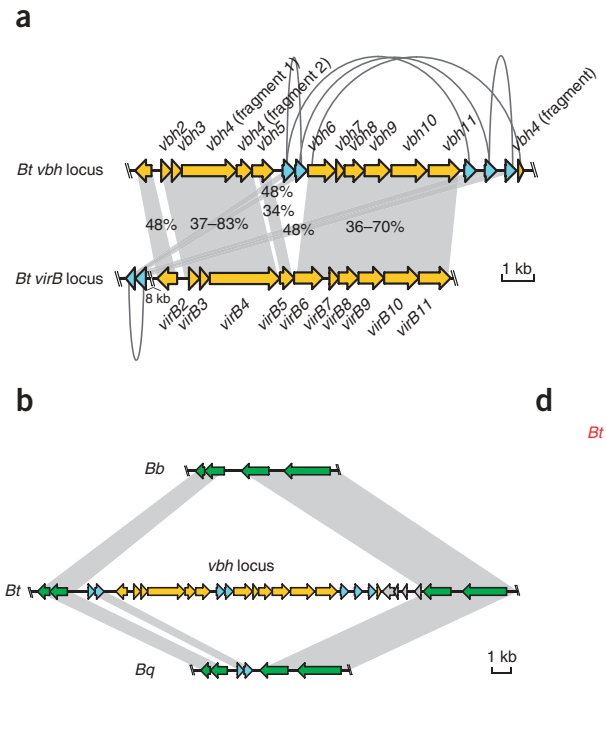


Figure 5 Evolutionary relationship of VirB and Vbh T4SSs. (**a**) Comparison of the *virB* and the *vbh* locus of *B. tribocorum* (*Bt*). Sequence similarity is shown for genes coding proteins having more than 30% amino acid identities. Genes belonging to the T4SSs are shown in yellow. T4SS-associated 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, Xyllela 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

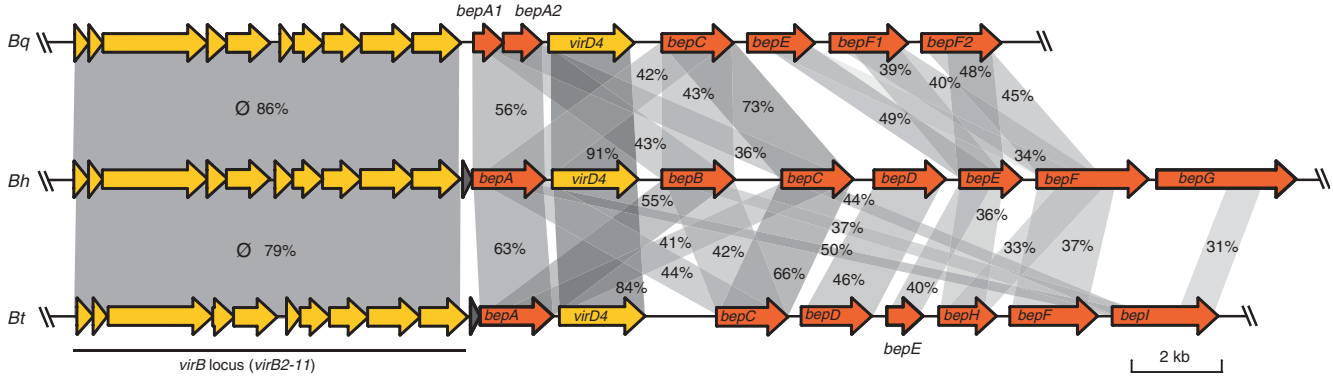


Figure 6 Comparison of the *virB* T4SS loci of *B. henselae*, *B. quintana* and *B. tribocorum*. Sequence similarity is shown for genes having more than 30% amino acid identities over a sequence stretch of more than 250 amino acids. VirB T4SS genes are shown in yellow and VirB T4SS effector genes in 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*)[15]. 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.)[26] and amoebae (for example, environmental *Chlamydia* spp.)[27] 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[15].

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[16,28]. 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[13]. **Supplementary Table 5** online lists all bacterial strains and plasmids used in this study.

Genome sequencing. Genomic DNA from *B. tribocorum* IBS 506^T (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 0.014×10^{-5} using the PHRED/PHRAP/CONSED software package[30–32].

Annotation and genome analysis. For the automated annotation and the manual curation, we used the annotation package GENDB[33] as described in the **Supplementary Methods** online. For the comparative genome analysis, we used the program MUMmer[34] and the Artemis comparison tool (ACT)[35]. 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*[36] were done by using BLASTP (see **Supplementary Methods**).

Construction of phylogenetic trees. For the phylogenetic tree of bartonellae, a multilocus sequence analysis (MLSA) approach[24,25] 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[37], and overhanging ends were cut and phylogenetic trees calculated using different algorithms included in the MEGA 3.1 software[38]. 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[39,40]. 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 ($[NK]_{25}$) 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[12]. 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 °C.

Animal infections. Infections were done using the *B. tribocorum* rat-infection model described previously[12]. 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); *Xyllela 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. Houpikian, 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).

3.2 Research article II

Parallel adaptive radiations in a bacterial pathogen

Philipp Engel, Walter Salzburger, Marius Liesch, Chao-Chin Chang, Soichi Maruyama, Christa Lanz, Alexandra Clateau, Aurélie Lajus, Claudine Médigue, Stephan C. Schuster, and Christoph Dehio

Manuscript in preparation.

Format: Report for Science

3.2.1 Summary

Our previously published results show that most contemporary species of *Bartonella* have arisen after lateral acquisition of type IV secretion systems (T4SSs). These colonization factors appear to have enabled *Bartonella* species to adapt to different mammalian reservoir hosts (3.1. Research article I). This diversification resembles an evolutionary process called adaptive radiation which is the rapid diversification of an ancestral lineage into an array of species by specific adaptation to different ecological niches. So far, adaptive radiations were mostly described for the evolution of metazoan eukaryotes. Although several studies implicate the occurrence of adaptive radiations in bacteria, the underlying molecular mechanisms are poorly understood. Adaptive radiations can be driven by the sudden availability of unoccupied niches or by key innovations allowing the exploration of new niches.

Here, we analyzed different *Bartonella* lineages for the presence of adaptive radiation with the aim of understanding the molecular mechanisms underlying host adaptation. We used 454-pyrosequencing for generating draft genome sequences of five *Bartonella* species isolated from different hosts (*B. clarridgeiae*, *B. rochalimae*, *B.* sp. AR15-3, *B.* sp. 1-1C, and *B. schoenbuchensis)*. For *B. clarridgeiae,* we completed the genome assembly by end-sequencing of several hundred clones of a Fosmid library. Including publicly available genome sequences, we inferred a genome-wide phylogenetic tree of *Bartonella*. This phylogenetic analysis was further complemented with available gene sequences from non-sequenced *Bartonella* species. The resulting supertree revealed that two lineages of *Bartonell*a, both harboring the VirB T4SS, have evolved by separate adaptive radiations. Remarkably, in both lineages, different *Bartonella* species have independently adapted to same or similar mammalian host(s).

To detect genes important for niche (i.e. host) adaptation, we comprehensively detected genes which have undergone adaptive evolution by measuring non-synonymous in relation to synonymous mutation rates. Strikingly, in both lineages the translocated effector proteins of the VirB T4SS were found to carry strong signal of adaptive evolution. This was in agreement with our previously published study, where the VirB T4SS was identified as an important host adaptability factor (3.1. Research article I). Further, it implicated that this horizontally acquired T4SS displays a key innovation driving the parallel adaptive radiations in the two lineages.

To understand the independent occurrence of these radiations, we assessed the evolutionary history of the VirB T4SS. In line with the evolutionary parallelism found on the ecological level, our analyses detected a highly parallel molecular evolution of the VirB T4SS in the two radiating lineages. First, genomic comparisons revealed that the VirB T4SS is integrated at different chromosomal sites in the two lineages. Second, in one of the two lineages the T4SS gene cluster was found to be partially triplicated and the effector genes disseminated over the chromosome. Third, phylogenetic analyses revealed that the effector genes of the two lineages have evolved from a single ancestral gene by lineage-specific amplification and subsequent diversification. Our data implies that these diversifying processes of the effector genes occurred prior to the radiations. Strikingly, in both lineages, we identified independently evolved effector proteins harboring tandem-repeated tyrosine-phosphorylation motifs. We subsequently showed that these motifs indeed are phosphorylated in eukaryotic cells.

The high degree of parallelism found in the molecular evolution of the VirB T4SS is evidence for comparable, but independent, responses to similar selective pressure, and it reflects the parallel occurrence of adaptive radiations in the two different *Bartonella* lineages. Hence, we propose that the chromosomal fixation of *virB* T4SS genes and the amplification-diversification of genes encoding the effector proteins displayed a two-step key innovation driving adaptive radiations in *Bartonella*. This is the first study which demonstrates adaptive radiations for bacteria in their natural habitats, and moreover, assesses the underlying molecular mechanisms.

Statement of the own participation

Except for the sequencing, library construction, and automated gene annotation, all data reported in this manuscript was generated by me. Stephan Schuster was responsible for the 454-sequencing. The Fosmid library of the *B. clarridgeiae* genome was generated and sequenced by Christa Lanz. The automated annotation was performed at Genoscope, the French National Sequencing Center, under supervision of Alexandra Calteau and Claudine Médigue. Walter Salzburger was intensively involved providing me with support and advise for the evolutionary analyses. The manuscript was written by me, Walter Salzburger, and Christoph Dehio.

3.2.2 Manuscript

Parallel adaptive radiations in a bacterial pathogen

Engel P.[1], Salzburger W.[2], Liesch M.[1], Chang, C.C. [3], Maruyama, S.[4], Lanz C.[5], Clateau A.[6], Lajus A.[6], Médigue C.[6], Schuster S.C.[7], and Dehio C.[1*]

[1]Biozentrum, University of Basel, Switzerland, [2]Zoological Institute, University of Basel, Switzerland, [3]College of Veterinary Medicine, National Chung Hsing University, Taichung, Taiwan, [4]Nihon University, Fujisawa, Kanagawa, Japan, [5]Max Planck Institute for Developmental Biology, Tübingen, Germany, [6]Laboratoire de Génomique Comparative, CNRS UMR8030, Institut de Génomique, CEA, Génoscope, Evry, France, [7]Center for Infectious Disease Dynamics, PennState University, Pennsylvania, USA

[*]Corresponding author:
Prof. Christoph Dehio
Biozentrum, University of Basel
Klingelbergstrasse 70
CH-4056 Basel, Switzerland
Tel.: +41-61-267 2140
Fax.: +41-61-267 2118
e-mail: christoph.dehio@unibas.ch

Running title: *Adaptive radiations in bacteria*

Abstract

Organismal diversification commonly occurs through adaptive radiation, the evolution of ecological and phenotypic diversity within a rapidly multiplying lineage. Bacteria evolved enormous biological diversity by exploring disparate environmental niches, yet the demonstration of bacterial speciation driven by adaptive radiation has been challenging. Here we report the first compelling example for an adaptive radiation in bacteria in nature. Based on the genome-wide analyses of ten species of the alphaproteobacterial genus *Bartonella* we describe a parallel adaptive radiation within two sister clades of this mammalian pathogen. Moreover, we inferred a horizontally acquired and subsequently diversified protein secretion system as an evolutionary key innovation driving these parallel adaptive radiations, thereby establishing a molecular paradigm for adaptive radiation in bacteria.

Main text

Adaptive radiation, i.e. the rapid evolution of a multitude of species from a single ancestor as a consequence of the adaptation to divergent environments (Schluter 2000), is thought to be responsible for the genesis of a great portion of the diversity of life. Adaptive radiations typically occur after the arrival of an organism in an environment with unoccupied niches (ecological opportunity) or by the acquisition of a novel trait (evolutionary key innovation) allowing the exploitation of so far unavailable niches (*1, 2*). The most famous textbook examples of adaptive radiations are the Darwin's finches on the Galapagos archipelago (*3*), the Caribbean *Anolis* lizards (*4*) and the species flocks of cichlid fishes in the East African Great lakes Victoria, Malawi and Tanganyika (*5*). Surprisingly, as of yet no convincing case of adaptive radiation has been described for bacteria in their natural environment – despite their great genetic and phenotypic diversity, their abundance in nearly any environment on Earth, and their capability to rapidly occupy distinct niches as shown in experimental evolution studies (*6*-*10*).

Four features are characteristic of an adaptive radiation (*2*): (i) common ancestry of the radiating group of organisms; (ii) rapid rates of speciation; (iii) the correlation between newly evolved phenotypes and the available ecological niches; and (iv) the utility of phenotypic traits to exploit the environment. While it is rather obvious how these criteria are fulfilled in the 13 species of Darwin's finches with their distinctive beak morphologies, anoles with their characteristic limbs correlated to twig diameter, and hundreds of cichlid species with their diverse mouth and teeth structures, the inference of phenotype-environment correlation and trait utility remains a challenge in bacteria due to their much simpler *bauplan*. Uncertainties with

the delineation of species boundaries and the unresolved issue of defining ecological niches in microorganisms (*11-13*) pose additional problems with correlating divergent phenotypic traits with niche-specific adaptations.

Here we present the alphaproteobacterial pathogens of the genus *Bartonella* as a first compelling, natural example of adaptive radiation in bacteria. Our genome-wide supertree-phylogenetic analysis based on a backbone of 478 conserved core genome genes (515,751 bp in total) retrieved from five available genome sequences plus five newly sequenced genomes (Figs. 1a and S1) reveals four major clades in the monophyletic bartonellae: (i) an ancestral lineage 1 represented by the highly virulent human pathogen *B. bacilliformis (14)* ; (ii) lineage 2 comprising four ruminant-infecting species with overlapping host niches *(15-17)*; (iii) lineage 3 consisting of two species and two strains recently isolated from different hosts; and (iv) the most species-rich lineage 4 with 13 described species. With short internal branches and longer branches at the tips, lineages 2, 3 and 4 show signs of rapid speciation in our phylogeny.

The more than 20 described *Bartonella* species are specifically adapted to cause a long-lasting intra-erythrocytic infection in their respective mammalian reservoir hosts (*18, 19*) (see Fig. 1a). Whereas closely related species of lineage 2 have overlapping hosts (*15-17*), the niche-restriction found in lineage 3 and 4 displays a species-specific phenotype as demonstrated by the exclusive retrieval of a given *Bartonella* spp. from the blood of naturally or experimentally infected reservoir hosts, but not from any non-reservoir hosts (*20-22*). Our own infection experiments confirm the niche-specificity for the species of lineage 3 including the recently isolated strains 1-1C and AR15 (Fig. S2) (*21*). In case of lineage 3 and 4, the mammalian reservoir hosts thus represent the equivalent to the distinct ecological niches, and niche specificity constitutes the distinct phenotypic property of these

bacterial pathogen species. Hence, there exists a strong phenotype-environment correlation in *Bartonella*.

With the overall and respective monophyly of the radiating lineages 3 and 4, the rapid cladogenesis in these lineages (Fig.1a), and the phenotype – environment correlation (i.e. exclusive infection of the specific reservoir host), the bartonellae fulfill the first three criteria of an adaptive radiation. It is less obvious, though, how the last criterion, trait utility, can be substantiated in *Bartonella*. First, a candidate trait would have to be involved in species – environment (i.e. reservoir host) interaction. For *Bartonella*, molecular factors that are essential for causing bacteremia in their reservoir hosts display such candidate traits. Second, in analogy to the morphological modulation of the phenotypic or adaptive traits in Darwin's finches, anoles lizards or cichlid fishes, any molecular factor used to exploit the distinct environments (i.e. the reservoir hosts) in a specific manner should be divergent among niche-specialized species. Molecular evolutionary analyses provide the means to identify such traits as these are expected to show signs of adaptive evolution at the molecular level, i.e. an excess of non-synonymous (*d*n) over synonymous substitutions (*d*s) (*23*). Hence, we performed a genome-wide *d*n/*d*s analysis in the available genomes of lineage 3 and 4 (Fig. S3 and Supporting Online Material [SOM]). This unbiased approach revealed that among the few positively selected genes are those coding for two different type IV secretion systems (T4SS), VirB and Trw, as well as several autotransporter proteins (Table S2, S3, S4). The two T4SS have recently been identified as so-called host adaptability factors that are essential for establishing intraerythrocytic infections in *Bartonella* (*24*). While Trw is exclusively found in the most species-rich lineage 4, VirB is shared among the two radiating lineages 3 and 4 (Fig. 1). The VirB system constitutes a multisubunit nanomachine (*25*) translocating a set of different *Bartonella* effector proteins (Beps) into mammalian host cells where they subvert a wide range

of different cellular functions (*26-30*). Beps consist of a modular domain structure and have evolved by gene duplications and functional diversification from a single ancestral effector gene (Fig. S4). It is exactly these Beps that show strong signs of positive selection – both when applying site tests (Table S4) as well as branch tests (Figs. 2, S6, and S7). Thus, our analyses together with previous work on the molecular mechanisms of infection (*24*) corroborate that the VirB T4SS displays an adaptive trait in *Bartonella* used to exploit distinct niches (i.e. the reservoir host). Hereby, also the last criterion of an adaptive radiation, trait utility, is fulfilled – at least for lineage 3 and 4. At the same time, the VirB T4SS represents the likely evolutionary key innovation facilitating adaptive radiation in these *Bartonella* lineages.

The VirB T4SS has presumably been acquired in the common ancestor of lineage 3 and 4 (*24*). T4SSs are ancestrally related to plasmid-encoded conjugation machineries mediating the horizontal transfer of these extra-chromosomal replicons among bacteria (*31, 32*). It is thus tempting to speculate that the common ancestor of the two radiating lineages acquired the VirB T4SS as conjugation system of a conjugative plasmid and was subsequently integrated into the chromosome. This integration occurred in different ways in the two lineages (SOM): In the sequenced genomes of lineage 4, a single gene cluster encodes the different secretion system components (*virB2-virB11* and *virD4*) as well as the *bep* genes suggesting a single integration event. On the contrary, the VirB T4SS components in lineage 3 are encoded by two to three paralogous gene clusters, which became integrated at two chromosomal sites different from the one found in lineage 4 (Fig. 1b and S5). Beside the adjacent location to *virB* T4SS genes (as in lineage 4), *bep* genes are found at six additional genomic loci in lineage 3 (Figs. 1b and S5).

The independent evolution of the VirB T4SS in the two lineages also becomes evident from our gene trees of the effector genes based on different sequence data sets (Figs. 2, S6, and S7). In all analyses, the *bep* genes of lineage 3 and 4 formed two separate clades. Apparently, after the integration and divergence of the VirB T4SS in lineage 3 and 4 independent duplications of the single ancestor-gene in each lineage resulted in the parallel emergence of two distinct arsenals of T4SS effector genes. Furthermore, the *bep* gene tree reveals that these duplication events preceded the radiations in both lineages, since distinct sub-clades (Bep clades in Fig. 2) often contain positional orthologs present in all analyzed genomes of the corresponding lineage. Subsequent to these duplications, the different effector gene copies have undergone extensive adaptive diversification in both lineages. Our branch tests of positive selection (Figs. 2, S6, and S7) detected adaptive evolution on many internal branches suggesting the emergence of new effector variants prior to the radiations. This is in line with the various functions experimentally identified for the distinct Bep clades of lineage 4 (*27*-*30*). The C-terminal BID (*Bartonella* intracellular delivery) domain, conserved among all *bep* genes (Fig. S5), is essential for T4SS translocation, but was specifically adapted by effector proteins of the BepA/B subclade (Fig. 2) for modulation of the apoptotic fate of host cells (*29*). In the N-terminal part, a substantial number of *bep* genes encode so called FIC (filamentation-induced by cAMP) domains, which were recently shown to mediate a new post-translational modification of proteins by covalent attachment of an AMP moiety (*33*-*35*). This specific modification of target proteins called AMPylation seems to be used by several bacterial pathogens to tune distinct host cellular functions (*36*). *Bartonella* species of lineage 3 and 4 harbor an arsenal of divergent FIC-containing effector proteins presumably resulting in the modulation of a wide range of host cell functions by the specific AMPylation of various target proteins. Effector proteins

comprising an N-terminal FIC domain and a C-terminal BID domain represent the most conserved *bep* genes across lineages 3 and 4, thus they likely constitute the domain architecture of the common ancestor gene (SOM). However, many *bep* genes of lineage 4 (*bepD, bepE, bepF, bepG, and bepH*) are missing a FIC domain, and often harbour additional BID domains (Fig. S4). In our gene tree (Fig 2), they display the most distant clades with respect to lineage 3 implying that they derived from the ancestral domain structure by secondary recombination events. Some of these derived effectors (*bepD, bepE, bepF* of *B. henselae*) were shown to become phosphorylated by host cell kinases at conserved tyrosine phosphorylation motifs (Fig. 3 and S4). Thereby, these effector proteins recruit cellular binding partners and interfere with signalling pathways of the host cell (*37*). The presence of these tandemly repeated motifs in four out of seven Bep clades indicates their important role for VirB T4SS-mediated host cell interactions in lineage 4.

One of the most fascinating features of adaptive radiations is the frequent occurrence of evolutionary parallelisms (*5*). This is also the case for *Bartonella*. The independent adaptive radiations of lineage 3 and 4 have occurred in overlapping or highly similar niches (i.e. hosts). For example, species of both clades have adapted to cats or rats as reservoir host (Fig. 1). In line with these 'ecological' parallelisms at the species level, our comparative analyses of the eight available genomes of the radiating lineages 3 and 4 uncovered evolutionary parallelisms at the molecular level of the VirB T4SS and its effector genes. Separate mechanisms have fixed the VirB T4SS in the chromosomes of the two lineages. Still, the assemblage of T4SS effector genes has been shaped by gene duplication and positive selection in both lineages independently (see above), even though the radiations occurred at different times (lineage 3 shows shorter genetic distances and is, hence, younger [see Fig. 1]). Moreover, both lineages have convergently evolved a novel class of effectors with a

derived domain structure. In lineage 3 as well as in lineage 4 (see above), we find the effector proteins comprising derived domain structures with tyrosine-phosphorylation motifs (Bep clade 9, Figs. 2 and S4). As the sub-clades harbouring these effector proteins in the two lineages hold distant phylogenetic positions (Fig. 2), they must have evolved separately from each other. *In silico* predictions of tyrosine-phosphorylation motifs consistently detected high numbers of tandem-repeated phosphorylation sites in the derived effector genes of lineage 3 (Fig. 3, Table S5). By ectopic expression in HEK293 cells, we show that these effector proteins become indeed tyrosine-phosphorylated within eukaryotic cells (Fig. 3). Interestingly, the repeated motifs detected in effector proteins of lineage 3 are clearly different from the ones in lineage 4 and are also less conserved among each other (Fig. 3). This implicates that parallel evolution at the molecular level in the two lineages led to the emergence of new effector variants with altered domain architecture and newly acquired functional modules comprising tandemly-repeated tyrosine-phosphorylation motifs.

Taken together, in both lineages, the parallel evolution of the effector genes follows gene-duplication-diversification processes resulting in the generation of extensive genetic variation on which adaptive evolution can act. The modularity of the effector proteins consisting of versatile units (as phosphorylation motifs, BID domains, or FIC domains) provides the framework for the adaptive modulation of these traits in different ecological niches (i.e. hosts). The parallelism in the molecular evolution of the VirB T4SS prior to their independent adaptive radiations provides compelling evidence for similar selective conditions acting in both lineages. We propose that the key innovation for the adaptive radiations evolved by a two-step process, (i) by the independent chromosomal fixation of the VirB T4SS and an

effector ancestor-gene, and (ii) by the subsequent parallel amplification and diversification of effector genes.

This is the first study demonstrating adaptive radiation in bacteria and assessing the underlying molecular mechanisms. The distinguishable ecological niches of *Bartonella* provide a suitable setting to assess adaptive radiation, yet the phenomenon might be much more frequent in bacteria that evolved as specialists to explore diverse arrays of ecological environments.

Acknowledgments

We thank Rusudan Okujava and Arto Pulliainen for providing plasmids pRO1100 and pAP015, respectively.This work was supported by grant 3100-061777 from the Swiss National Science Foundation (to C.D.), grant 55005501 from the Howard Hughes Medical Institute (to C.D.), and grant 51RT-0_126008 (InfectX) in the frame of the SystemsX.ch Swiss Initiative for Systems Biology (to C.D.).

References

1. S. Gavrilets, A. Vose, *Proc Natl Acad Sci U S A* 102, 18040 (Dec 13, 2005).
2. D. Schluter, *The ecology of adaptive radiation*, Oxford series in ecology and evolution (Oxford University Press, Oxford, 2000), pp. viii, 288 p.
3. P. R. Grant, B. R. Grant, *How and why species multiply : the radiation of Darwin's finches*, Princeton series in evolutionary biology (Princeton University Press, Princeton, 2008), pp. xix, 218 p., [32] p. of plates.
4. M. A. Butler, S. A. Sawyer, J. B. Losos, *Nature* 447, 202 (May 10, 2007).
5. W. Salzburger, *Mol Ecol* 18, 169 (Jan, 2009).
6. Z. D. Blount, C. Z. Borland, R. E. Lenski, *Proc Natl Acad Sci U S A* 105, 7899 (Jun 10, 2008).
7. R. Kassen, *Ann N Y Acad Sci* 1168, 3 (Jun, 2009).
8. R. C. MacLean, *J Evol Biol* 18, 1376 (Nov, 2005).
9. P. B. Rainey, M. Travisano, *Nature* 394, 69 (Jul 2, 1998).
10. J. Sikorski, E. Nevo, *Proc Natl Acad Sci U S A* 102, 15924 (Nov 1, 2005).
11. F. M. Cohan, *Annu Rev Microbiol* 56, 457 (2002).
12. C. Fraser, E. J. Alm, M. F. Polz, B. G. Spratt, W. P. Hanage, *Science* 323, 741 (Feb 6, 2009).
13. K. T. Konstantinidis, A. Ramette, J. M. Tiedje, *Philos Trans R Soc Lond B Biol Sci* 361, 1929 (Nov 29, 2006).
14. G. M. Ihler, *FEMS Microbiol Lett* 144, 1 (Oct 15, 1996).
15. C. C. Chang *et al.*, *Emerg Infect Dis* 6, 306 (May-Jun, 2000).
16. C. Dehio *et al.*, *Int J Syst Evol Microbiol* 51, 1557 (Jul, 2001).

17. J. M. Rolain, E. Rousset, B. La Scola, R. Duquesnel, D. Raoult, *Ann N Y Acad Sci* 990, 236 (Jun, 2003).

18. Central themes in the lifestyle of the facultative intracellular pathogen *Bartonella* are the transmission route via blood-sucking arthropods (such as ticks, fleas, and flies), the ability to invade different cell types, in particular endothelial cells and erythrocytes resulting in a persistent intra-erythrocytic bacteremia, and their specific adaptation to one or a few closely related host(s). Expect for *B. bacilliformis*, bartonellae cause only mild or even asympotmatic infections in their corresponding natural hosts. Seven species have been recognized as zoonotic pathogens with humans representing an incidental host resulting in a wide range of different disease manifestations *(19)*.

19. C. Dehio, *Nat Rev Microbiol* 3, 621 (Aug, 2005).

20. B. B. Chomel *et al.*, *Vet Res* 40, 29 (Mar-Apr, 2009).

21. B. B. Chomel *et al.*, *Vet Res* 40, 27 (Mar 10, 2009).

22. M. Y. Kosoy *et al.*, *Comp Immunol Microbiol Infect Dis* 23, 221 (Oct, 2000).

23. Natural selection of orthologous genes can be inferred by calculating the ratio of non-synonymous (*d*n, amino acid change) to synonymous (*d*s, amino acid conservation) substitution (ω = *d*n/*d*s). Generally, ω < 1, ω = 1, ω > 1 represent purifying, neutral, and positive selection (= adaptive evolution), respectively (Yang and Nielsen, 2000). Since positive selection may act episodically during evolution on only a few sites, the *d*n/*d*s ratio calculated over the entire orthologous genes often results in ω < 1. Therefore, an incrased ω indicates relaxed purifying selection suggesting that certain sites might be under positive selection. To test for sites or branches under positive

selection, maximum-likelihood based methods are used (implemented e.g. in PAML *[48]* and HyPhy *[45]*).

24. H. L. Saenz *et al.*, *Nat Genet* 39, 1469 (Dec, 2007).
25. R. Fronzes *et al.*, *Science* 323, 266 (Jan 9, 2009).
26. C. Dehio, *Cell Microbiol* 10, 1591 (Aug, 2008).
27. T. A. Rhomberg, M. C. Truttmann, P. Guye, Y. Ellner, C. Dehio, *Cell Microbiol* 11, 927 (Jun, 2009).
28. F. Scheidegger *et al.*, *Cell Microbiol* 11, 1088 (Jul, 2009).
29. M. C. Schmid *et al.*, *PLoS Pathog* 2, e115 (Nov, 2006).
30. R. Schulein *et al.*, *Proc Natl Acad Sci U S A* 102, 856 (Jan 18, 2005).
31. A. C. Frank, C. M. Alsmark, M. Thollesson, S. G. Andersson, *Mol Biol Evol* 22, 1325 (May, 2005).
32. R. Schulein, C. Dehio, *Mol Microbiol* 46, 1053 (Nov, 2002).
33. C. A. Worby *et al.*, *Mol Cell* 34, 93 (Apr 10, 2009).
34. M. L. Yarbrough *et al.*, *Science* 323, 269 (Jan 9, 2009).
35. M. L. Yarbrough, K. Orth, *Nat Chem Biol* 5, 378 (Jun, 2009).
36. C. R. Roy, S. Mukherjee, *Sci Signal* 2, pe14 (2009).
37. M. Selbach *et al.*, *Cell Host Microbe* 5, 397 (Apr 23, 2009).
38. E. C. Berglund *et al.*, *PLoS Genet* 5, e1000546 (Jul, 2009).
39. C. M. Alsmark *et al.*, *Proc Natl Acad Sci U S A* 101, 9716 (Jun 29, 2004).
40. K. Inoue *et al.*, *Emerg Infect Dis* 15, 526 (Apr, 2009).
41. J. W. Lin, C. Y. Chen, W. C. Chen, B. B. Chomel, C. C. Chang, *J Med Microbiol* 57, 1496 (Dec, 2008).
42. J. C. Wilgenbusch, D. Swofford, *Curr Protoc Bioinformatics* Chapter 6, Unit 6 4 (Feb, 2003).
43. J. P. Huelsenbeck, F. Ronquist, *Bioinformatics* 17, 754 (Aug, 2001).

44. D. Vallenet *et al.*, *Nucleic Acids Res* 34, 53 (2006).

45. S. L. Pond, S. D. Frost, S. V. Muse, *Bioinformatics* 21, 676 (Mar 1, 2005).

46. N. Blom, S. Gammeltoft, S. Brunak, *J Mol Biol* 294, 1351 (Dec 17, 1999).

47. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, *Genome Res* 14, 1188 (Jun, 2004).

48. Z. Yang, *Mol Biol Evol* 24, 1586 (Aug, 2007).

Fig. 1. (A) Genome-wide supertree-phylogeny of *Bartonella* and (B) genomic organization of *virB* T4SS and effector gene loci in lineage 3 and lineage 4. (A) Maximum likelihood analysis based on 478 genes (515,751 sites) of ten sequenced *Bartonella* species (indicated by bold and underlined type): the available genomes of *B. bacilliformis* (CP000524), *B. grahamii* (*38*), *B. henselae*, *B. quintana (39)*, and *B. tribocorum (24)* as well as the complete genome of *B. clarridgeiae* and the draft genomes of *B. schoenbuchensis*, *B. rochalimae*, *B. sp.* AR 15-3, and *B. sp.* 1-1C sequenced in this study (for genome features see Table S1). *B. sp.* AR 15-3 and *B. sp.* 1-1C were recently isolated from Amercian red squirrel and rat, respectively (*40, 41*). Non-sequenced *Bartonella* species were included in the analysis based on sequence data from *rpoB*, *glt*a, *ribC*, and *groEL* genes. Numbers above the branches represent maximum-likelihood bootstraps obtained with PAUP (*42*); numbers below represent values from Bayesian inference obtained with the MrBayes (*43*). The two monophyletic clades harboring the VirB T4SS (*24*) are marked by the shaded area. The different lineages are depicted in red, orange, blue, and green color (l1, lineage 1). Mammalian hosts are indicated for each species. The same monophyletic clades were obtained when using more than one species as outgroup. (B) Synteny plot of *B. clarridgeiae* (lineage 3) and *B. quintana* (lineage 4) generated in MaGe (*44*). Syntenic relationships comprising at least 5 genes are indicated by violet and blue lines for genes found on the same and the opposite strand, respectively. The genomic integration sites of *virB* T4SS loci are indicated. For lineage 3, the six loci encoding additional effector genes are marked by arrows. Genes are connected by gray boxes if they are orthologs of each other. For *bep* genes, connections are drawn if they belong to the same monophyletic clade (Fig. 2) or if they are top blast hits of each other. Bq, *B. quintana*, Bh, *B. henselae*, Bt, *B. tribocorum*, Bg, *B. grahamii*, B15, *B. sp.* AR15-3, Bc, *B. clarridgeiae*, Br, *B. rochalimae*, B1-1, *B. sp.* 1-1C.

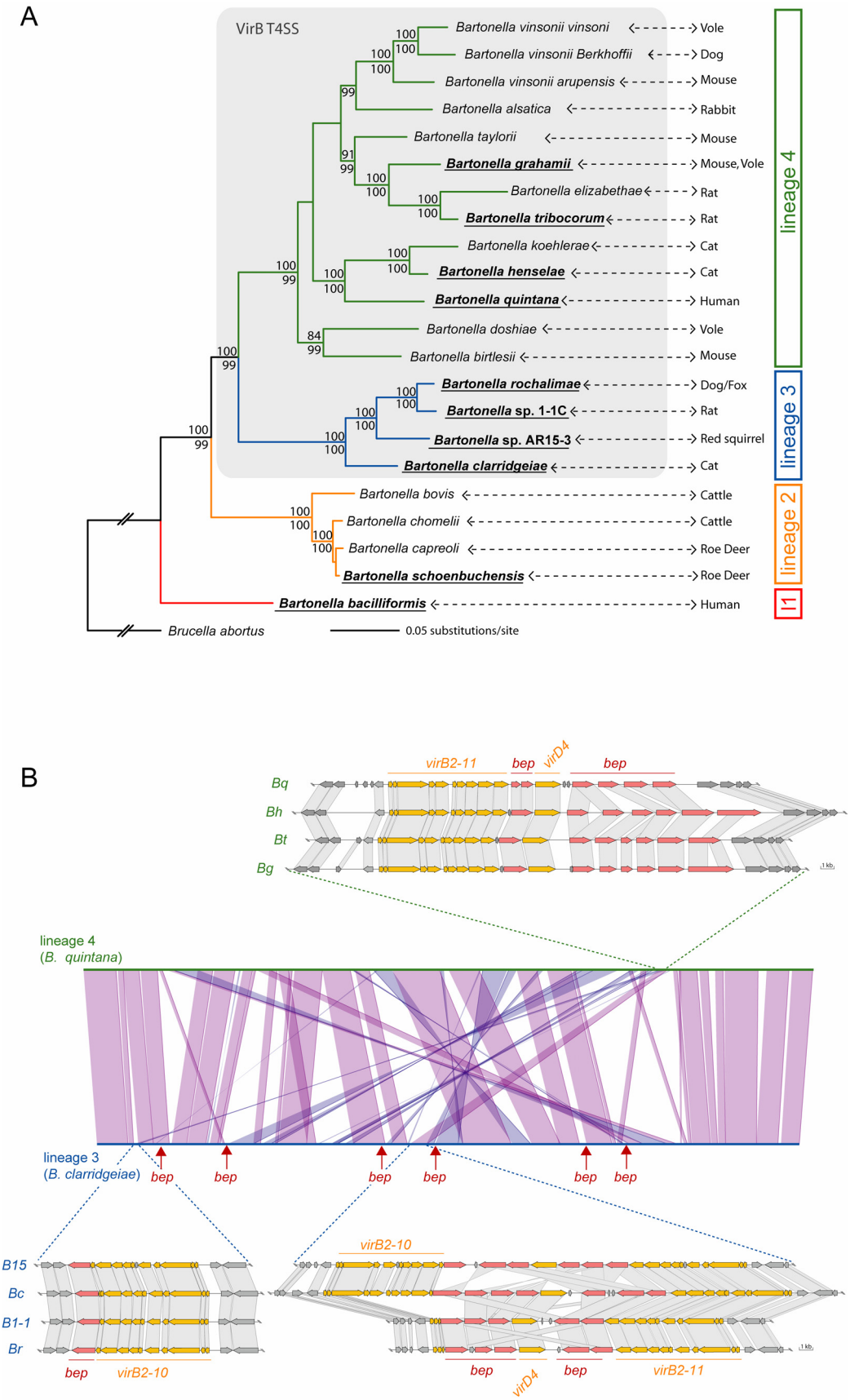

A
VirB T4SS
Bartonella vinsonii vinsoni
Vole
Bartonella vinsonii Berkhoffii
Dog
Bartonella vinsonii arupensis
Mouse
Bartonella alsatica
Rabbit
Bartonella taylorii
Mouse
Bartonella grahamii
Mouse, Vole
Bartonella elizabethae
Rat
Bartonella tribocorum
Rat
Bartonella koehlerae
Cat
Bartonella henselae
Cat
Bartonella quintana
Human
Bartonella doshiae
Vole
Bartonella birtlesii
Mouse
Bartonella rochalimae
Dog/Fox
Bartonella sp. 1-1C
Rat
Bartonella sp. AR15-3
Red squirrel
Bartonella clarridgeiae
Cat
Bartonella bovis
Cattle
Bartonella chomelii
Cattle
Bartonella capreoli
Roe Deer
Bartonella schoenbuchensis
Roe Deer
Bartonella bacilliformis
Human
Brucella abortus
0.05 substitutions/site
lineage 4
lineage 3
lineage 2
I1
B
virB2-11
bep
virD4
bep
Bq
Bh
Bt
Bg
lineage 4
(B. quintana)
lineage 3
(B. clarridgeiae)
bep
bep
bep
bep
bep
bep
virB2-10
B15
Bc
B1-1
Br
bep
virB2-10
bep
virD4
bep
virB2-11

Figure 1

Fig. 2. Gene tree of *Bartonella* effector proteins. Maximum-likelihood analysis (general time-reversible model with gamma-correction [GTR+I+G]) based on the BID domain and the C-terminus of VirB T4SS effector genes identified in lineage 3 and lineage 4. Numbers above the branches represent maximum-likelihood bootstraps with PAUP (*42*)); numbers below represent posterior probabilities obtained with the MrBayes (*43*). Values above 80% are shown. Locus_tag and gene name (if existing) is given for each *bep* gene, and color-coded according to species. The dashed line in red color indicates the independent clustering of the *bep* genes from lineages 3 and lineage 4. Bold lines in black color depict phylogenetic branches with *d*n/*d*s values > 1 as calculated with HYPHY (*45*). For each branch under positive selection, the estimated *d*n/*d*s value is indicated. The later onset of the radiation in lineage 3 (see Fig. 1) explains the more frequent detection of positive selection in this lineage (particularly, on tip branches). Monophyletic sub-clades comprising orthologous *bep* genes are indicated. In lineage 3, BARCLv2_0629 and BARCLv2_0635 are not included in any clade, as their phylogenetic positions are neither supported by bootstrapping nor posterior probabilities. Shaded areas indicate clades of *bep* genes harboring tandem-repeated tyrosine-phosphorylation motifs. Trees inferred from data sets based on the C-terminal FIC domain or the entire effector gene sequence reveal the same two lineage-specific monophyletic clades (Figs. S6 and S7).

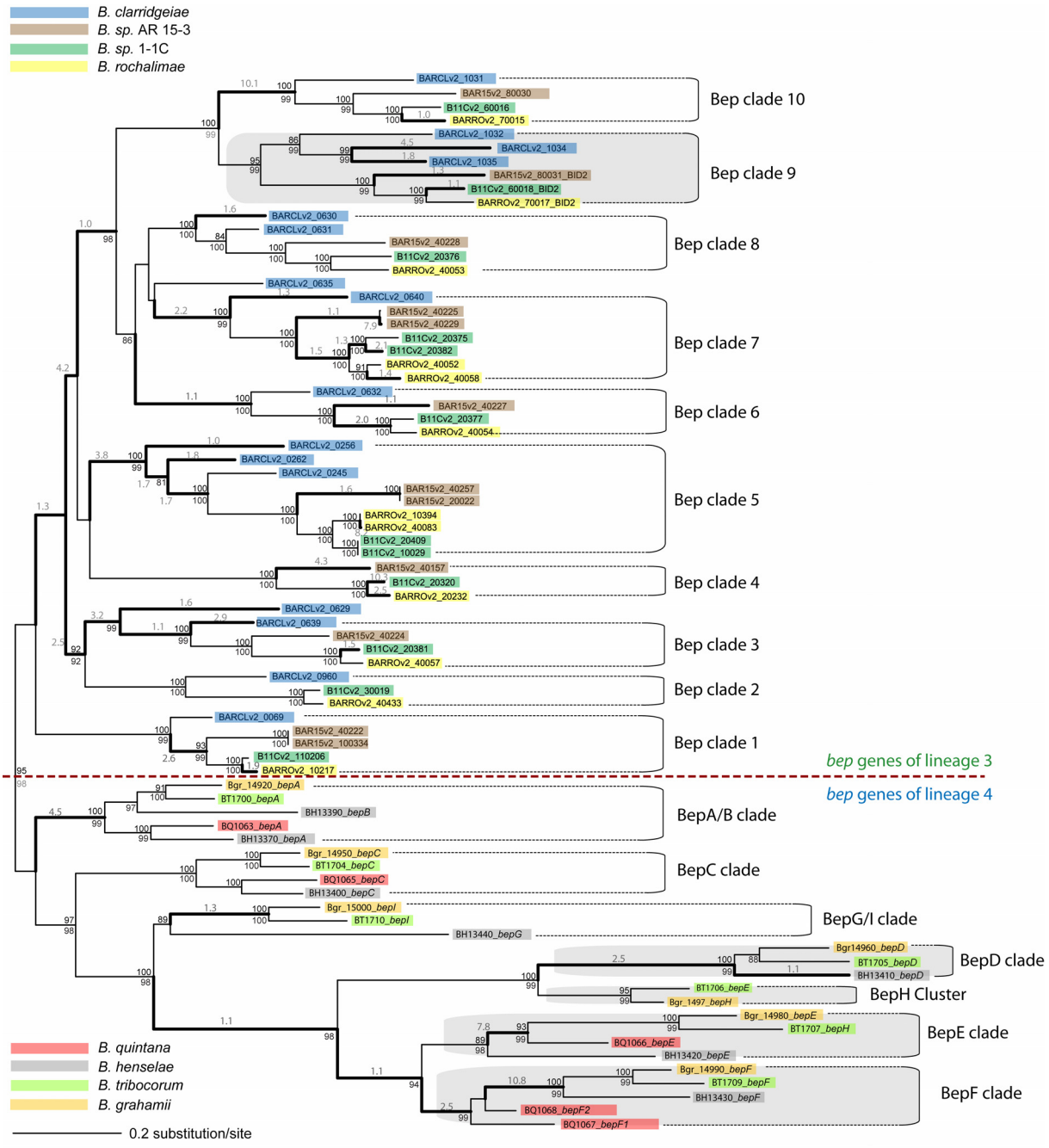

B. clarridgeiae
B. sp. AR 15-3
B. sp. 1-1C
B. rochalimae
Bep clade 10
Bep clade 9
Bep clade 8
Bep clade 7
Bep clade 6
Bep clade 5
Bep clade 4
Bep clade 3
Bep clade 2
Bep clade 1
bep genes of lineage 3
bep genes of lineage 4
BepA/B clade
BepC clade
BepG/I clade
BepD clade
BepH Cluster
BepE clade
BepF clade
B. quintana
B. henselae
B. tribocorum
B. grahamii
0.2 substitution/site

Figure 2

Fig. 3. Tyrosine-phosphorylation of Beps of lineage 3. (A) Domain structure of four orthologous effector genes of Bep clade 9 (lineage 3) and predicted tyrosine-phosphorylation motifs. The motifs depicted by green bars were identified with the program NetPhos2.0 (*46*) using a threshold of 0.8. The consensus sequence of tandem-repeated motifs is shown for each of the four Beps by WebLogos (*47*). (B) WebLogo generated from all predicted tyrosine-phosphorylation motifs of Bep clade 9. (C) Weblogo generated from all predicted tyrosine-phosphorylation motifs of Bep clades BepD*,* BepE, BepF, and BepH (lineage 4). The motifs were identified with the program NetPhos2.0 using the same threshold as for Bep clade 9 of lineage 3. (D) and (E) Immunoprecipitation/Western blot analysis of HA-GFP-Bep fusion proteins ectopically expressed in HEK293T cells. Names refer to the locus_tags of the corresponding *bep* genes. Immunoprecipitations (IP) were performed with HA antibody-coated agarose beads. In (D), Western blot (WB) analysis was performed with anti-phosphotyrosine antibody. In (E), Western blot analysis was performed with anti-GFP antibody.

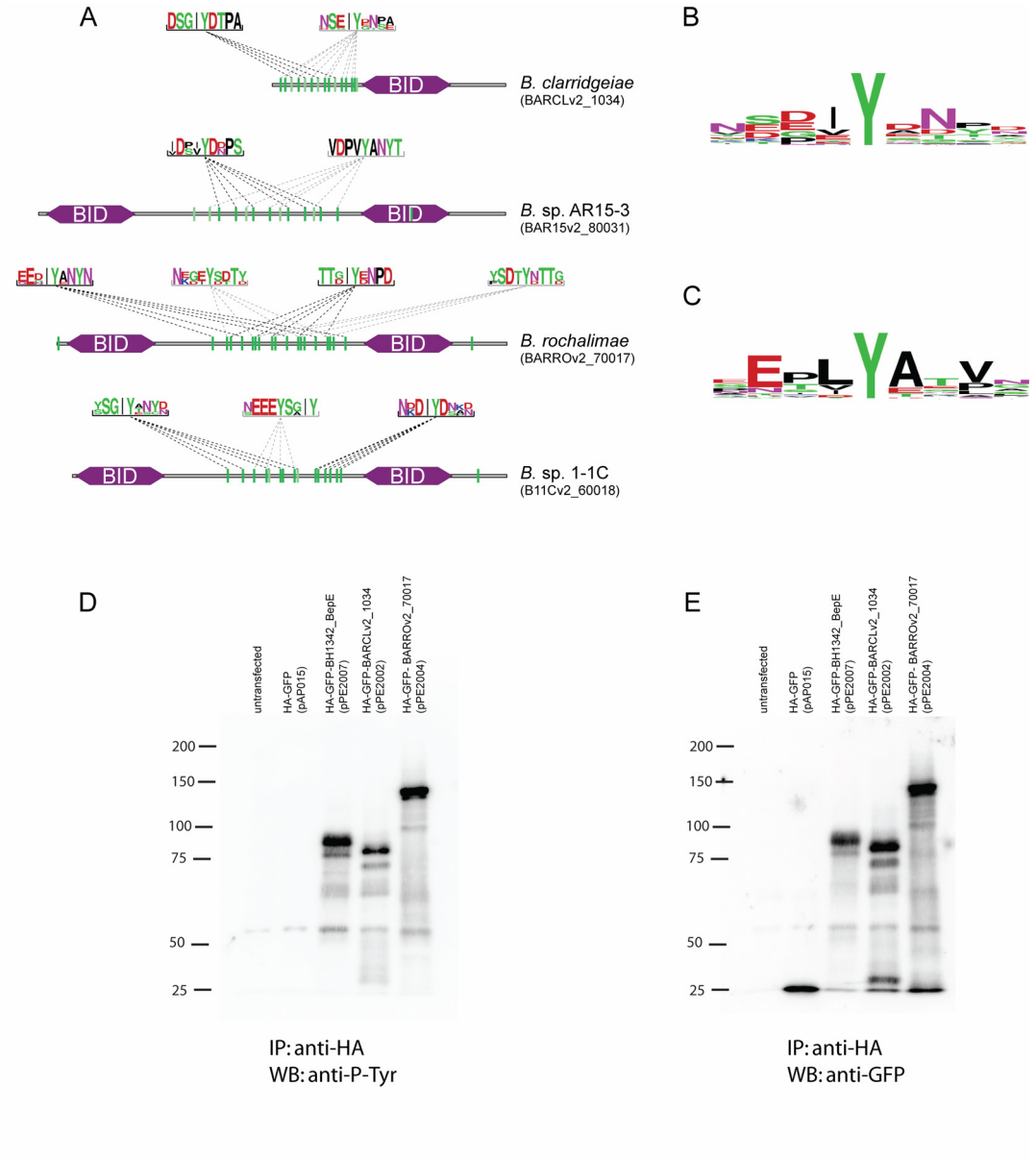

A
BID
B. clarridgeiae
(BARCLv2_1034)
BID
BID
B. sp. AR15-3
(BAR15v2_80031)
BID
BID
B. rochalimae
(BARROv2_70017)
BID
BID
B. sp. 1-1C
(B11Cv2_60018)
B
C
D
untransfected
HA-GFP (pAP015)
HA-GFP-BH1342_BepE (pPE2007)
HA-GFP-BARCLv2_1034 (pPE2002)
HA-GFP- BARROv2_70017 (pPE2004)
200
150
100
75
50
25
IP: anti-HA
WB: anti-P-Tyr
E
untransfected
HA-GFP (pAP015)
HA-GFP-BH1342_BepE (pPE2007)
HA-GFP-BARCLv2_1034 (pPE2002)
HA-GFP- BARROv2_70017 (pPE2004)
200
150
100
75
50
25
IP: anti-HA
WB: anti-GFP

Figure 3

3.2.3 Supporting Online Material

Parallel adaptive radiations in a bacterial pathogen

Engel P.[1], Salzburger W.[2], Liesch M.[1], Chang, C.C. [3], Maruyama, S.[4], Lanz C.[5], Clateau A.[6], Lajus A.[6], Médigue C.[6], Schuster S.C.[7], and Dehio C.[1]

[1]Biozentrum, University of Basel, Switzerland, [2]Zoological Institute, University of Basel, Switzerland, [3]College of Veterinary Medicine, National Chung Hsing University, Taichung, Taiwan, [4]Nihon University, Fujisawa, Kanagawa, Japan, [5]Max Planck Institute for Developmental Biology, Tübingen, Germany, [6]Laboratoire de Génomique Comparative, CNRS UMR8030, Institut de Génomique, CEA, Génoscope, Evry, France, [7]Center for Infectious Disease Dynamics, PennState University, Pennsylvania, USA

Supporting Online Material includes:

Supporting Results

Genome-wide natural selection analysis reveals the VirB and Trw type IV secretion system (T4SSs) as well as different autotransporters as adaptive traits in the radiating lineage 3 and lineage 4 of *Bartonella*

In order to screen for adaptive traits among *Bartonella* species, we analyzed all genes from the available genomes of the radiating lineage 3 (*B. henselae*, *B. quintana, B. tribocorum,* and *B. grahamii*) and lineage 4 (*B. clarridgeiae*, *B. sp.* AR 15-3, *B. sp.* 1-1C, and *B. rochalimae*) for signs of adaptive evolution. To this end, we inferred the natural selection of orthologous genes by estimation of ω, the ratio of non-synonymous (*d*n, amino acid change) to synonymous (*d*s, amino acid conservation) substitution rates (ω = *d*n/*d*s). Generally, ω < 1, ω = 1, ω > 1 represent purifying, neutral, and positive selection (= adaptive evolution), respectively (*1*).

Based on the eight available genome sequences, our “gene-wide” *d*n/*d*s estimates revealed strong purifying selection (ω < 0.25) for most orthologous genes identified in the two lineages (lineage 3: 964 [88%] and lineage 4: 1004 [92%]). The mean value of ω obtained for lineage 3 and lineage 4 was ω = 0.19 and ω = 0.12, respectively. Only a small fraction of genes (lineage 3: 133 [12%] and lineage 4: 86 [8%]) displayed increased *d*n/*d*s values (ω ≥ 0.25) which might indicate signs of adaptive evolution (Fig. S3, Tables S2 and S3).

As adaptive evolution is often not affecting the entire gene sequence but only a few sites (*2*), we additionally analyzed the 86 and 133 genes for positive selection (ω > 1) acting on individual codons. To this end, we used a maximum likelihood method implemented in the software package PAML (see Materials and Methods), which fits different models of evolution on the analyzed sets of orthologous genes.

These models allow ω to vary among sites by assigning codons to different classes of natural selection (*3*). Likelihood ratio tests (LRTs) were then used to compare the log likelihood values of two nested models and to determine which model of evolution is fitting the data significantly better. We used two pairs of models which can be tested against each other: M1a (NearlyNeutral) versus M2 (Selection) and M7 (beta) versus M8 (beta + ω) (*4*). Model M2 as well as M8 are models including classes for sites under positive selection. The results of this analysis (Tables S2 and S3) show that in both radiating lineages positive selection was detected for a subset of genes with ω ≥ 0.25 (lineage 3: 29 [M1a vs M2] and 31 [M7 vs M8] genes; lineage 4: 14 [M1a vs M2] and 20 [M7 vs M8] genes. As non-synonymous mutations accumulate over time, the later onset of the radiation in lineage 3 (reflected by the shorter distances between descendant taxa [Fig. 1]) might explain the larger fraction of genes under positive selection. Hence, the data set derived from the genomes of lineage 3 appears to be more sensitive to the detection of adaptive traits by natural selection analysis.

A substantial fraction of genes, exhibiting ω ≥ 0.25, were found to be present in both lineages (Table S4). Interestingly, autotransporters and different components of the VirB T4SS displayed a large fraction of these common genes with ω ≥ 0.25. The VirB T4SS and several autotransporters are known to be important for host colonization *(5-8)* and, thus, might display adaptive traits of *Bartonella*.

Autotransporters are outermembrane-anchored proteins which are used by *Bartonella* for uptake of essential metabolic resources from the host environments, as e.g. hemin via the hemin-binding proteins (*9, 10*). In our maximum likelihood analysis testing for sites under positive selection, most of the identified autotransporters revealed very strong signal of positive selection (Table S4). As

these genes tend to recombine and the assignment of direct orthologs is difficult, false positive detection of positive selection has to be considered and might partially explain the strong signal detected in our analysis. However, recombination by itself displays an efficient way to confer adaptation to changing environments and might support that these factors represent adaptive traits.

The VirB T4SS constitutes a multisubunit nanomachine (*11*) translocating a set of different *Bartonella* effector proteins (Beps) into mammalian host cells. These effector proteins were shown to subvert a wide range of different cellular functions of the host (*8, 12-15*). Strikingly, all analyzed *bep* genes of lineage 3 (nine) as well as lineage 4 (four) were present among the genes with $\omega \geq 0.25$ (Tables S2 and S3). In lineage 4, all nine VirB T4SS effectors exhibited $\omega \geq 0.4$, thereby belonging to the genes with the highest *d*n/*d*s values detected. Our maximum likelihood analysis revealed eight of these nine effectors to be under positive selection, whereas only one out of four analyzed *bep* genes in lineage 4 exhibited statistically significant positive selection (Table S4). Due to the more recent radiation of lineage 3, the stronger signal of positive selection in the effector genes of this lineage was expected (see above). In addition to the effector proteins, two structural components of the VirB T4SS machinery were present among the common genes with $\omega \geq 0.25$. These are *virB*5 and *virB7* encoding a putative pilus subunit and an outermembrane-associated protein, respectively (*16*).

Other factors known to be essential for host colonization *(5)* revealing in both lineages $\omega \geq 0.25$ represent a TonB-like protein and a conserved protein of unknown functions (BQ02810, BARCLv2_0226). Whereas the TonB-like homolog was found to be under positive selection in lineage 4, the protein of unknown function showed no signs of positive selection in neither of the two lineages (Table S4).

Among the genes exclusively exhibiting ω ≥ 0.25 in lineage 4, several components of another T4SS, Trw, were detected (Table S3). The Trw T4SS is only present in species of lineage 4 and was shown to be essential for intra-erythrocytic colonization *in vivo* (*17*). We detected positive selection in several components of the Trw T4SS, particularly, in pilus subunits, which have recently been reported to have evolved by amplification and diversification (*18*). As the Trw T4SS is supposed to be involved in erythrocyte binding, the adaptive evolution detected in surface-exposed genes could reflect the adaptation to divergent hosts. Among the remaining genes of lineage 4 exhibiting ω ≥ 0.25, only three proteins of unknown function (BQ11480, BQ03410, and BQ13290) are known to be essential for host colonization(*5*). In one of these three genes, our analysis detected significant adaptive evolution.

Beside the genes in common with lineage 4, no other genes of lineage 3 with ω ≥ 0.25 encode known host colonization factors. Many candidate genes found in lineage 3 code for small proteins (~ 100 aa) with unknown function. The only interesting factors, exclusively detected in lineage 3, are genes homologous to type III secretion system (T3SS) effector genes, which belong to the widely distributed YopJ family. In many bacterial pathogens, these proteins display important factors for host interaction (*19*). In lineage 3, three of these genes were found to be under positive selection. With the exception of *B. henselae,* YopJ homologs are also present in the analyzed species of lineage 4. However, in case of *Bartonella,* these proteins are not known to be involved in host interaction and T3SSs enabling translocation of these effector proteins into host cells were not yet identified.

Taken together, the analysis revealed a limited set of molecular factors of *Bartonella*, which appear to present adaptive traits as (i) they exhibit signs of adaptive evolution and (ii) are known to be essential for niche (i.e. host) colonization.

These adaptive traits comprise the VirB T4SS, various autotransporters, and the Trw T4SS. As autotransporters are also present in non-radiating *Bartonella* lineages, and the Trw T4SS is restricted to lineage 4, the horizontally acquired VirB T4SS displays the only detectable adaptive trait common to the radiating lineage 3 and 4 but absent from non-radiating lineages.

Differences in genomic organization indicates an independent evolutionary history of the VirB T4SS in lineage 3 and lineage 4

In the published genomes of lineage 4 (*B. henselae*, *B. quintana, B. tribocorum,* and *B. grahamii*), the VirB T4SS genes, *virB2-virB11* as well as the coupling protein gene *virD4*, are encoded at the same chromosomal locus (Figs. 1 and S5). Also, the genes of the *Bartonella* effector proteins (Beps), which are translocated by the VirB T4SS machinery into the host cell, are encoded in this region, with *bepA* being located directly downstream of *virB11,* followed by *virD4* and the remaining *bep* genes. Interestingly, the genome sequences of lineage 3 (*B. clarridgeiae*, *B. rochalimae, B. sp.* AR 15-3 and *B. sp.* 1-1C) revealed marked differences in organization, copy number, and chromosomal localization of the *virB* and *bep* genes (Figs. 1 and S5). In the completely assembled genome of *B. clarridgeiae*, we find three copies of the *virB2-virB10* genes encoded at two different chromosomal loci. Two copies, encoded at the same locus, are facing each other and belong to inverted repeats of ~10kb which are separated by several *bep* genes and the coupling protein-encoding gene *virD4*. A third copy of the *virB2-virB10* gene cluster and an additional *bep* gene are encoded in a genomic region highly conserved across different *Bartonella* lineages (Fig. 1). The copy number and genomic organization of the T4SS of *B. clarridgeiae* was confirmed by end-sequencing of fosmid clones spanning the corresponding chromosomal regions. In contrast to VirB2-VirB10, the ATPase VirB11 and the coupling protein VirD4 are only encoded by a single gene copy. In the three additional draft genomes of lineage 3, the same chromosomal integration and amplification of the *virB* T4SS genes were found. However, one of the three copies must have been partially deleted in a common ancestor of *B. rochalimae* and *B. sp.* 1-1C, as only *virB2, virB3*, and a remnant of the *virB4* gene were found in the corresponding region of these two genomes (Figs. 1 and S5). Interestingly, the

different copies of *virB2-virB10* are identical within each of the four genomes, but clearly divergent across each other. This implicates an intragenomic concerted evolution of these regions driven by a homogenization process such as e.g. gene conversion. The fact that duplicated components of another T4SS, Trw, also evolved in concert, and the finding of several other identical genes or gene clusters in different *Bartonella* genomes (*18*), indicates that sequence homogenization is a common mechanism in *Bartonella* to conserve paralogous gene copies.

In *B. clarridgeiae*, the regions evolving in concert do not only include the *virB2-virB10* gene cluster but also an N-terminal part of the single gene copy *virB11*, which is only present as remnant in the other two copies of the T4SS. Two additional open reading frames are part of the homogenized regions of the inverted *virB* gene clusters. One of these coding sequences comprises a C-terminal part of the glutamine syntethase I gene (*glnA*), which is a vertically inherited housekeeping gene located downstream of the entire T4SS locus. The presence of this inverted C-terminal *glnA* fragment upstream of the locus is evident of a VirB T4SS duplication event subsequent to the integration of a first copy (Fig. S5). Moreover, this inverted duplication could also have resulted in the mirrored organization of the *bep* genes located between the two T4SS copies.

Beside the adjacent location to *virB* T4SS genes (as in lineage 4), a blast analysis revealed further *bep* genes in the genomes of lineage 3. In total, we found six additional loci encoding one or several *bep* genes. Some of these effector genes are missing in one or the other genome of lineage 4, and the existence of gene remnants provides evidence of their deletion. Interestingly, at the intragenomic level, some of the *bep* genes are highly similar or even identical to each other indicating concerted evolution of these paralogous gene copies (Fig. 2). Altogether, we

identified 12 to 16 *bep* genes in the genomes of lineage 3 (excluding remnants not harboring the C-terminal translocation signal), whereas in lineage 4 only five to seven *bep* genes are present.

Although incomplete synteny at corresponding loci may hinder the comparison between the two different lineages, no gene remnants could be found at the different integration sites across the two lineages, neither for *bep* nor *virB* genes. It is difficult to determine whether massive genomic recombination events resulted in the different chromosomal locations and the lineage-specific dissemination of the *virB* and *bep* genes. However, such a scenario appears not to be likely, as the overall genomic backbone is mostly conserved (Fig. 1) and the VirB T4SS integrations sites do not share adjacently-located genomic regions across the two lineages.

Another possibility could be that the VirB T4SS was assimilated into the chromosome by two independent lineage-specific events. This is supported by the fact that T4SS are ancestrally related to conjugation machineries which are encoded on plasmids and responsible for the horizontal dissemination of these extra-chromosomal replicons (*20*). It could well be that the VirB T4SS was plasmid-encoded when the two lineages diverged from each other and was integrated into the chromosome in a lineage-specific manner. The fact that related T4SSs, as the AvhB or the VirB system of pAT (*Agrobacterium tumefaciens*) and pSymB (*Sinorhizobium meliloti*) are plasmid-encoded supports such a scenario (*21, 22*). Furthermore, the highly related VirB-homologous (Vbh) T4SS (*5*), which was found to be encoded by two copies in *B. grahamii*, one located on a plasmid and the other one integrated into the chromosome, indicates that VirB-like T4SSs may episodically be maintained on extrachromosmal replicons in *Bartonella* (*23*).

The ancestral effector gene consisted of a N-terminal FIC and a C-terminal BID domain

The phylogenetic trees inferred from the different sequence data sets of the *Bartonella* effector proteins (*bep)* genes (Figs. 2, S6 and S7) revealed two monophyletic clades separating the effector genes of lineages 3 and lineage 4. This implicates a lineage-specific amplification-diversification process which resulted in the evolution of two different effector gene sets. As some *bep* genes consist of a modular domain structure which has resulted from extensive duplication and reshuffling (Fig S4) *(8)*, the domain architecture of the ancestral effector gene, present before the two lineages diverged, remains elusive. However, several findings suggest that the effector genes harboring an N-terminal FIC and a C-terminal BID domain (FIC-BID) represent the ancestral domain structure. Most *bep* genes of the two lineages harbor the FIC-BID structure. In lineage 3, only effector genes of Bep clade 9 consist of domain architectures different than FIC-BID (Fig. S4). Further, our gene tree based on BID domain and C-terminus (Fig. 2) shows that *bep* genes with the shortest evolutionary distance across the two lineages are the ones harboring the FIC-BID structure. *Bep* genes of the two lineages with different domain structures constitute more distant clades in respect to each other indicating that they must have been derived by independent recombination events. Strikingly, the presence of a putative effector gene harboring the FIC-BID structure in the ancestrally related VirB-homologous T4SS of *B. grahamii* or *B. schoenbuchensis* consolidates our hypothesis *(22)*.

Materials and Methods

Bacterial strains and growth conditions. *B. clarridgeiae* strain 73 (*24*), *B.* sp. 1-1C (*25*), and *B. rochalimae* ATCC BAA-1498 (*26*) were grown routinely for 3–5 days on tryptic soy agar containing 5% defibrinated sheep-blood in a water-saturated atmosphere with 5% CO2 at 35 °C. *B.* sp. AR15-3 (*27*) and *B. schoenbuchensis* R1 (*28*) were grown for the same time span under the same conditions on heart infusion agar and colombian base agar, respectively.

Genome sequencing, assembly, and annotation. Strains originating from single colonies were grown under the conditions described before. Genomic DNA was isolated using the QIAGEN Genomic DNA Isolation kit (Qiagen). For 454-sequencing, genomic DNA was prepared with an appropriate kit supplied by Roche Applied Science for sequencing on a Roche GS-FLX (*29*). For the assembly of the reads, the newbler standard running parameters with ace file output were used. A summary of number of reads, contigs, average read length, and genome coverage is presented in Table S6 for each of the five genomes. Newbler assemblies were considerably improved by linking overlapping contigs on the basis of the "_to" and "_from" information appended to the read name in the ace files generated with Newbler. Also, repeats were identified by analyzing the average coverage of each of the Newbler contigs. If the link between two contigs was ambiguous, PCR and long-range PCR were used to confirm contig joins. For *B. clarridgeiae*, a library of about 35 kb was generated using the CopyControl^{TM} Fosmid Library Production Kit (Epicentre). By the end-sequencing of clones from this library with Sanger

technology, 983 high-quality reads were obtained and mapped onto the assembly based on 454-sequencing reads. The remaining sequence gaps were closed by PCR. The final singular contig displaying the circular chromosome of *B. clarridgeiae* was fully covered by staggered fosmid clones indicating a correct assembly. Gene predictions of the genome of *B. clarridgeiae* as well as the draft genomes of *B. schoenbuchensis, B. rochalimae, B*. sp AR15-3, and *B.* sp. 1-1C were performed using AMIGene software (*30*). Automated functional gene annotation was conducted with the microbial genome annotation system MaGe (*31*). For orthologous genes, the annotation was adopted from the manually annotated genome of *B. tribocorum*. Manual validation of the annotation was performed for the *virB* T4SS and *bep* genes only. By using the "FusionFission" tool of MaGe, fragmented genes were identified and the corresponding sequence subsequently examined for 454-sequencing errors due to homopolymeric runs. After correcting these errors, the updated sequences were re-annotated as described before. The data is stored on the web-based interface MaGe (Bartonella2Scope) and publicly available.

Rat infections. Animal care was carried out in accordance with the Swiss Act on Animal Protection and Good Animal Care Practice. Ten weeks old female WISTAR rats obtained from RCC-Füllinsdorf were housed in an S2-animal facility for two weeks prior to infection allowing acclimatization. For inoculation, bacterial strains were grown as described before, harvested in phosphate-buffered saline (PBS), and diluted to OD_{595}=1. Rats were anesthetized with a 2-3% Isuflurane/O_2 mixture and infected with 10 µl of the bacterial suspension in the dermis of the right ear. Blood samples were taken at the tail vein and immediately mixed with PBS containing 3.8% sodium-citrate to avoid coagulation. After freezing to -70°C and subsequent thawing,

undiluted and diluted blood samples were plated on tryptic soy agar and heart infusion agar containing 5% defibrinated sheep-blood. CFUs were counted after 8-12 days of growth.

Phylogenetic analyses. Phylogenetic trees were based on nucleotide sequence data. Alignments were generated on protein sequences with ClustalW (*32*) and back-translated into aligned DNA sequences. To calculate tree topologies, we used maximum-likelihood and Bayesian inference methods as implemented in the programs Paup (*33*) and MrBayes (*34*), respectively. The genome-wide phylogeny of *Bartonella* was calculated on the basis of 478 genes of the ten sequenced genomes of *Bartonella* and the genome of *Brucella abortus* (bv. 1 str. 9-941). This orthologous gene set was determined by using the “PhyloProfile Synteny” tool of MaGe (*31*). The threshold for ortholog assignment was set to 60% protein identity over at least 80% of the length of proteins being directional best hits of each other. The alignments of the 478 identified genes were concatenated resulting in a total of 515,751 aligned nucleotide sites. Tree topology and branch lengths were obtained by maximum-likelihood analysis using the HKY85 model. Bootstrap support values were calculated for 100 replicates. To obtain the posterior probabilites by the Bayesian inference method the mcmc command of MrBayes (*34*) was run for one million iterations with standard parameters. To complement the genome-wide *Bartonella* phylogeny with non-sequenced *Bartonella* species, we include available sequence data for the *gltA*, *groEL*, *ribC*, and *ropB* genes (which all were present in the genome-wide data set) displaying 7731 nucleotide sites. Tree topology, branch lengths, and bootstraps were obtained as described before. To obtain the posterior probabilites by the Bayesian inference method the mcmc command of MrBayes (*34*) was run for five million

iterations with standard parameters. *Bep* gene trees were inferred from nucleotide alignments of either the most C-terminal BID domain including the C-terminal tail (948 sites), the FIC domain including the N-terminal extension (1305 sites), or the entire *bep* sequence of genes harboring the ancestral domain structure (3972 sites). For the latter data set, *bepI* of *B. grahamii* and *B. tribococrum* were included, although they harbor a central sequence insert not present in the other *bep* genes. To select an appropriate substitution model, the Akaike information criterion of Modeltest 3.7 (*35*) and MrModeltest 2.0 (*36*) was used for the maximum-likelihood and Bayesian inference analysis, respectively. For the alignments based on the BID domain and C-terminus, as well as for the data set based on entire *bep* gene sequences, we obtained the GTR+G+I model with both programs. For the alignments based on the FIC domain and N-terminus, the TVM+I+G model (Modeltest 3.7) and GTR+G+I (MrModeltest 2.0) were selected. With the parameters provided by these models, trees were inferred by the maximum-likelihood and Bayesian inference analyses as described before. Bootstrap support values were calculated for 100 replicates, and posterior probabilites obtained with the mcmc command of MrBayes (*34*) for one million iterations with model-specific parameters.

Natural selection analysis. Orthologous genes for each of the two phylogenetic lineages were determined by using the “PhyloProfile Synteny” tool of MaGe (Vallenet et al., 2006). The threshold for ortholog assignment was set to 30% protein identity of at least 60% over the length of proteins being directional best hits of each other. The same tool was used to detect genes without homologs in each of the two lineages. By comparing these automatically identified orthologs and non-orthologs, genes present in neither of the two lists were detected and manually assigned to one of the

two. The orthologous genes were subsequently aligned as described before. To obtain the average *d*n/*d*s value (ω) of each gene, the arithmetic mean of pair-wise *d*n/*d*s values calculated by the method of Yang and Nielsen (2000) as implemented in PAML 1.41 (Yang, 2007) was used. Site tests of positive selection for genes exhibiting ω ≥0.25 were performed with PAML 1.41 (Yang, 2007) by using the codeml module. Models M0 (one-ratio), M1 (neutral), M2 (selection), M3 (discrete), M7 (beta), and M8 (beta + ω) were analyzed. Paup (*33*) was used to infer maximum-likelihood trees of each gene. For the codeml control file the standard parameters were used. The models M1a (nearlyNeutral) and M2a (positiveSelection) as well as the models M7 (beta) and M8 (beta + ω) were compared in a Likelihood-ratio test using two degrees of freedom to determine relative significance of models. Phylogenetic branches were tested for positive selection by using the TestBranchDNDS.bf module implemented as standard analysis in HyPhy (*37*).

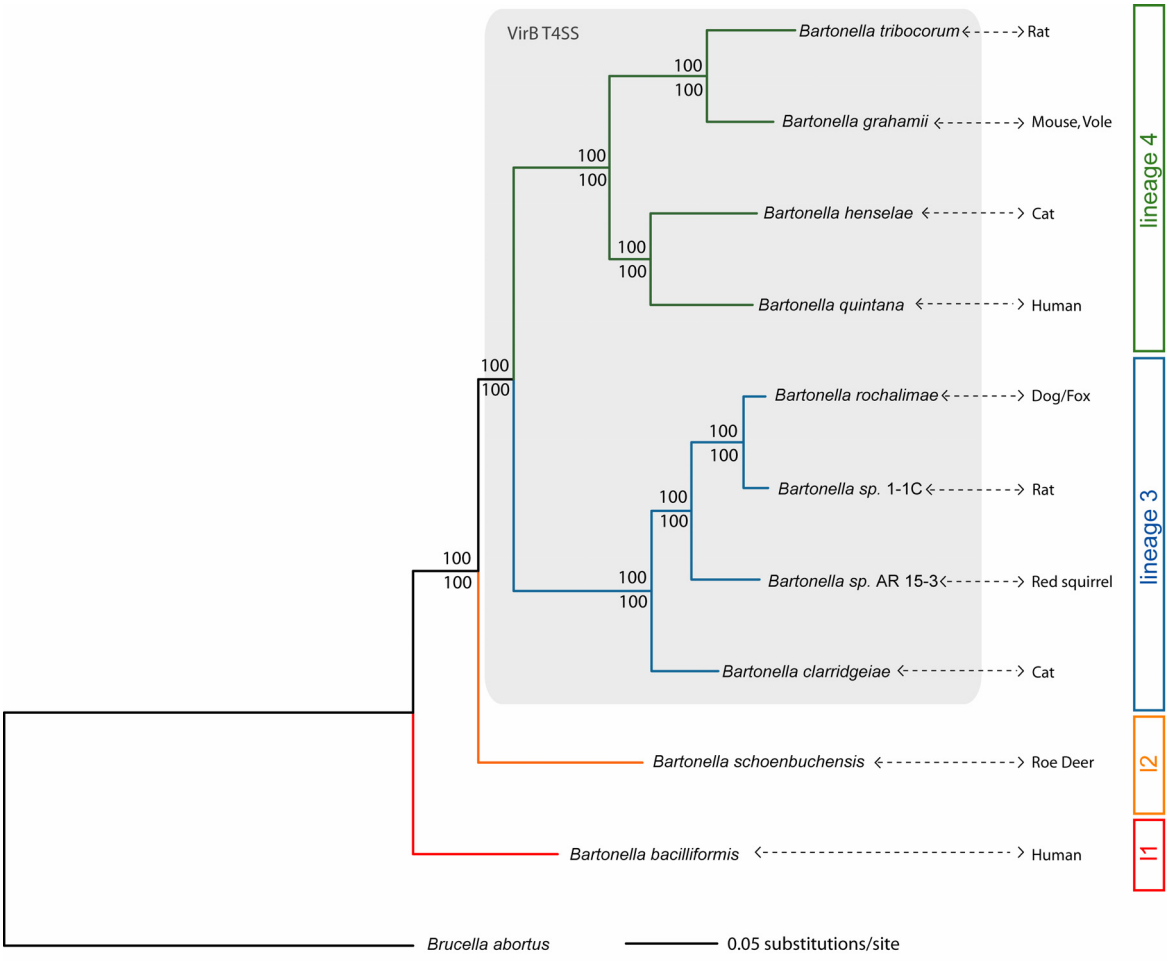
Cloning of plasmids for expression of HA-GFP-Bep fusion proteins. To construct the plasmids pPE2002 and pPE2004, *bep* genes BARCLv2_1034 (*B. clarridgeiae*) and BARROv3_80017 (*B. rochalimae*) were amplified from genomic DNA with primer pairs containing flanking *Bam*HI/*Not*I sites: prPE453 (ATAAGAATGCGGCCGCGATGAAAACCCATAACACTCCTG) / prPE454 (CGGGA-TCCTTAATGTGTTATAACCATCGTTC) and prPE455 (ATAAGAATGCGGCCGCGA-TGAATTTTGGAGAAAAGAAA-AAAATG) / prPE456 (CGGGATCCTTAAATAGCTAC-AGCTAACGATTTTTTC), respectively. PCR products were digested with enzymes *BamH*I and *Not*I and ligated into the *Bam*HI/*Not*I sites of the backbone of plasmid pAP013 (kindly provided by Arto Pulliainen). The resulting constructs pPE2001 (BARCLv2_1034) and pPE2003 (BARROv3_80017) were cut with Not*I* and ligated with a GFP fragment obtained from *NotI* digested pAP013. The plasmid pPE2007 was constructed by cutting *bepE* of *B. henselae* from plasmid pRO1100 (kindly

provided by Rusudan Okujava) with *Not*I and BamH*I* and ligating it into pAP013. All plasmid DNA isolations and PCR purifications were performed with Macherey-Nagel and Promega columns according to manufacturer's instruction.

Immunoprecipitation of HA-GFP-Bep fusion proteins. The protocol for growth and transfection of HEK293T is described in (*14*). 36 h after transfection, cells were incubated for 10 minutes with 10 ml Pervanadate medium (5 ml PBS containing 100 mM orthovanadate, 200 mM H_2O_2, incubated for 10 min with 500 µl Catalase (2 mg/ml in PBS), and 45 ml cell media added). After washing three times with 7 ml of PBS at room temperature cells were scraped off and resuspended in 1 ml of ice-cold PBS containing 1 mM EDTA, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM orthovanadate, 1 mM leupeptin, and 1 mM pepstatin and collected by centrifugation (3,000g at 4°C for 60 sec). The resulting pellet was lysed in 300 ml of ice cold modified RIPA buffer (50 mM Tris·HCl [pH 7.4], 75 mM NaCl, 1 mM EDTA, 1 mM orthovanadate, 1 mM leupeptin, 1 mM pepstatin) for 1 hour at 4°C. The lysate was centrifuged (16,000g at 4°C for 15 min) and 12 ml of anti-HA-agarose (Sigma) added to the supernatant. After 150 min of incubation at 4°C on a slowly turning rotation shaker, the agarose was washed three times with 300 ml of modified RIPA buffer (3,000g for 10 sec). The affinity-gel pellet was then resuspended in 20 ml of modified RIPA buffer, 20 ml of SDS-sample buffer (2´) were added, and the sample was heated for 5 min at 95°C. Proteins were separated on a 10% SDS-polyacrylamide gel, blotted on a nitrocellulose membrane (Hybond-C Extra, Amersham Pharmacia), and examined for tyrosine phosphorylation by using monoclonal antibody 4G10 (Millipore) and anti-mouse IgG-horseradish peroxidase (HRP) afterwards. The HRP-conjugated antibody was visualized by enhanced chemiluminescence (PerkinElmer).

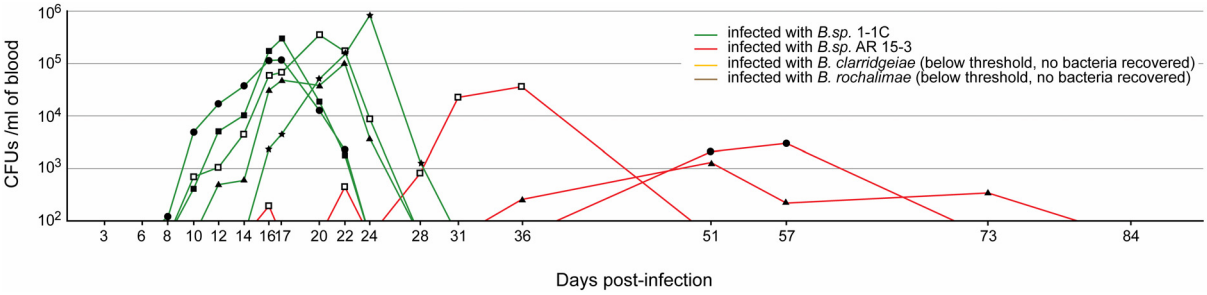
For visualization of the signal from GFP-fusion proteins the membrane was subsequently incubated in 4% PBS-Tween containing 0.02% NaN_3 and anti-GFP antibody (Invitrogen), followed by anti-mouse IgG-HRP and visualized by enhanced chemiluminescence.

Supplementary Fig. 1. Phylogenetic tree of ten sequenced *Bartonella* species based on the nucleotide alignment of 478 genes. As outgroup the species *Brucella abortus* (bv. 1 str. 9-941) was used. The tree topology and branch lengths were obtained by maximum-likelihood analysis (as implemented in PAUP(*33*)). Numbers above the branches represent maximum-likelihood bootstraps of 100 replicates, numbers below the branches represent posterior probabilities obtained with the MrBayes program (one million iterations, (*34*)). The two monophyletic clades harboring the VirB T4SS (*5*) are marked by the shaded area. The mammalian reservoir hosts are given for each species, and the four different lineages of *Bartonella* are indicated by red, orange, blue, and green color (l1, lineage 1, l2, lineage 2).



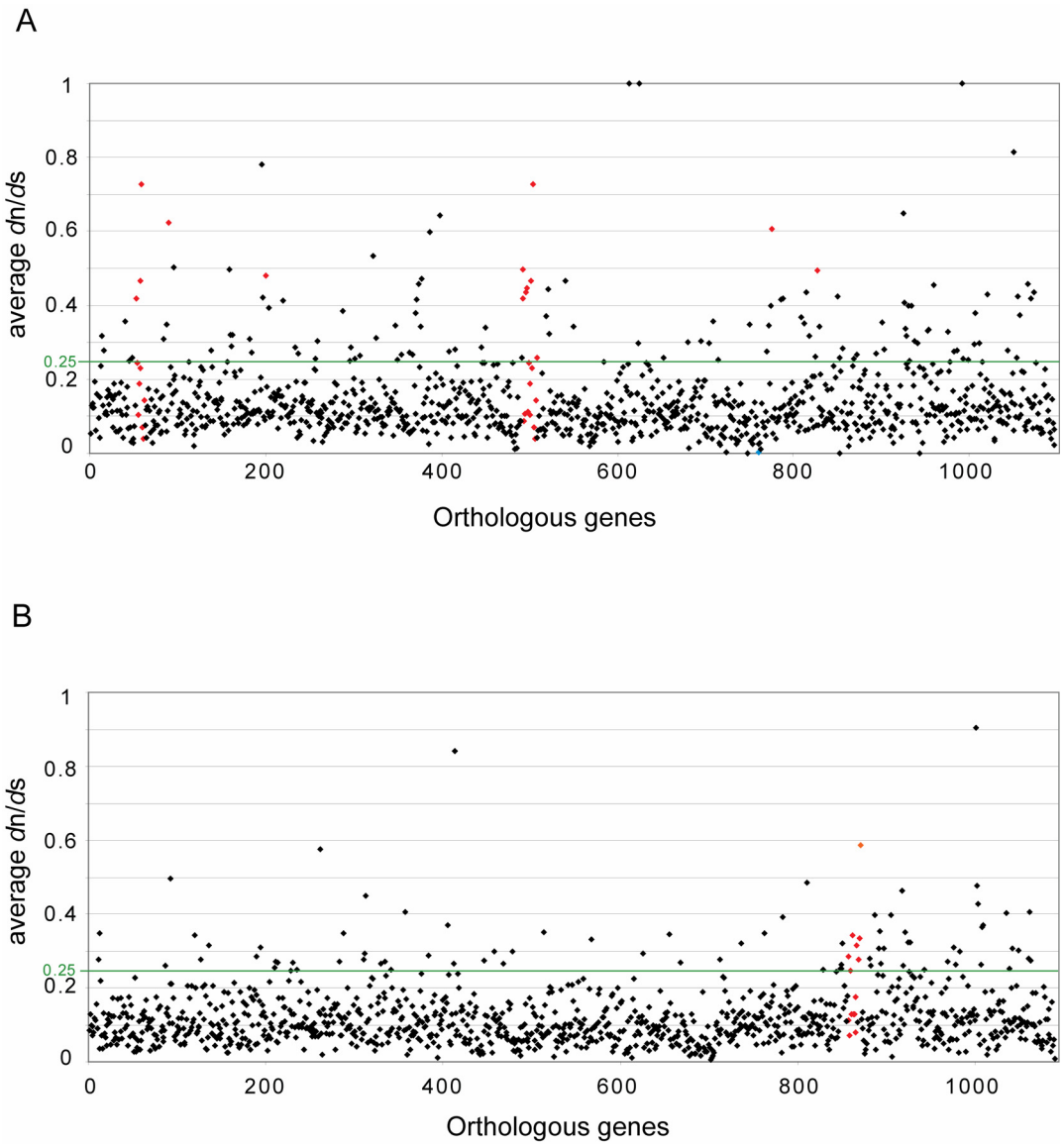
Supplementary Figure 1

Supplementary Fig. 2. Experimental infections of rats with *Bartonella* species of lineage 3. Groups of rats (n = 5) were inoculated intradermally with 10 µl of phosphate-buffered saline (PBS) solution containing bacteria of one of the four different species *B. clarridgeiae*, *B. rochalimae*, *B.* sp. 1-1C, or *B.* sp. AR 15-3 (OD_{595} = 1). At the time points indicated in the graphs, blood was drawn, diluted, and plated for counting of colony-forming units (CFUs). Plotted graphs represent single animals from which bacteria could be recovered at a given time point post-infection. In all five animals infected with the rat-specific strain *B.* sp. 1-1C, bacteria were detected between day 8 and 28 post-infection (green graphs). In only three of five animals infected with the red squirrel-specific strain *B.* sp. AR 15-3, bacteria could be detected (red graphs). In these three animals, bacteria came up much later and did not reach the same blood titer as *B.* sp. 1-1C. From none of the animals infected with either *B. rochalimae* or *B. clarridgeiae,* bacteria could be recovered (threshold of detection ≤ 10^2 CFU/ml). Additionally, a recent study (*38*) showed that *B. rochalimae* is reliably infecting dogs, its natural reservoir host, but not cats or guinea pigs. These independent studies confirm the host specificity among *Bartonella* species.



Supplementary Figure 2

Supplementary Fig. 3. Gene-wide *d*n/*d*s analysis of the core genomes of (A) lineage 3 and (B) lineage 4. The mean values plotted on the y-axis were obtained from *d*n/*d*s values of pairs of orthologs calculated with the method of Yang and Nielsen (2000) implemented in PAML (*3*). The genes are ordered according to their position in the genome of (A) *B. clarridgeiae* and (B)*B. quintana*. v*irB* T4SS and *bep* genes are indicated in red color. For lineage 3 (A), 133 genes where detected to have an average *d*n/*d*s value ≥ 0.25, whereas for lineage 4 (B) 86 genes exhibited *d*n/*d*s values ≥ 0.25. In lineage 3, three genes exhibit average *d*n/*d*s values >> 1 (see also Table S2). This is explained by the fact that between *B. rochalimae* and *B. sp* 1-1C, only non-synonymous mutations but no synonymous mutations have been detected in these three genes. For the other pair-wise comparisons, all three genes revealed ω < 1. This is indicated by the high standard deviations for average *d*n/*d*s values of these genes (see Table S2).

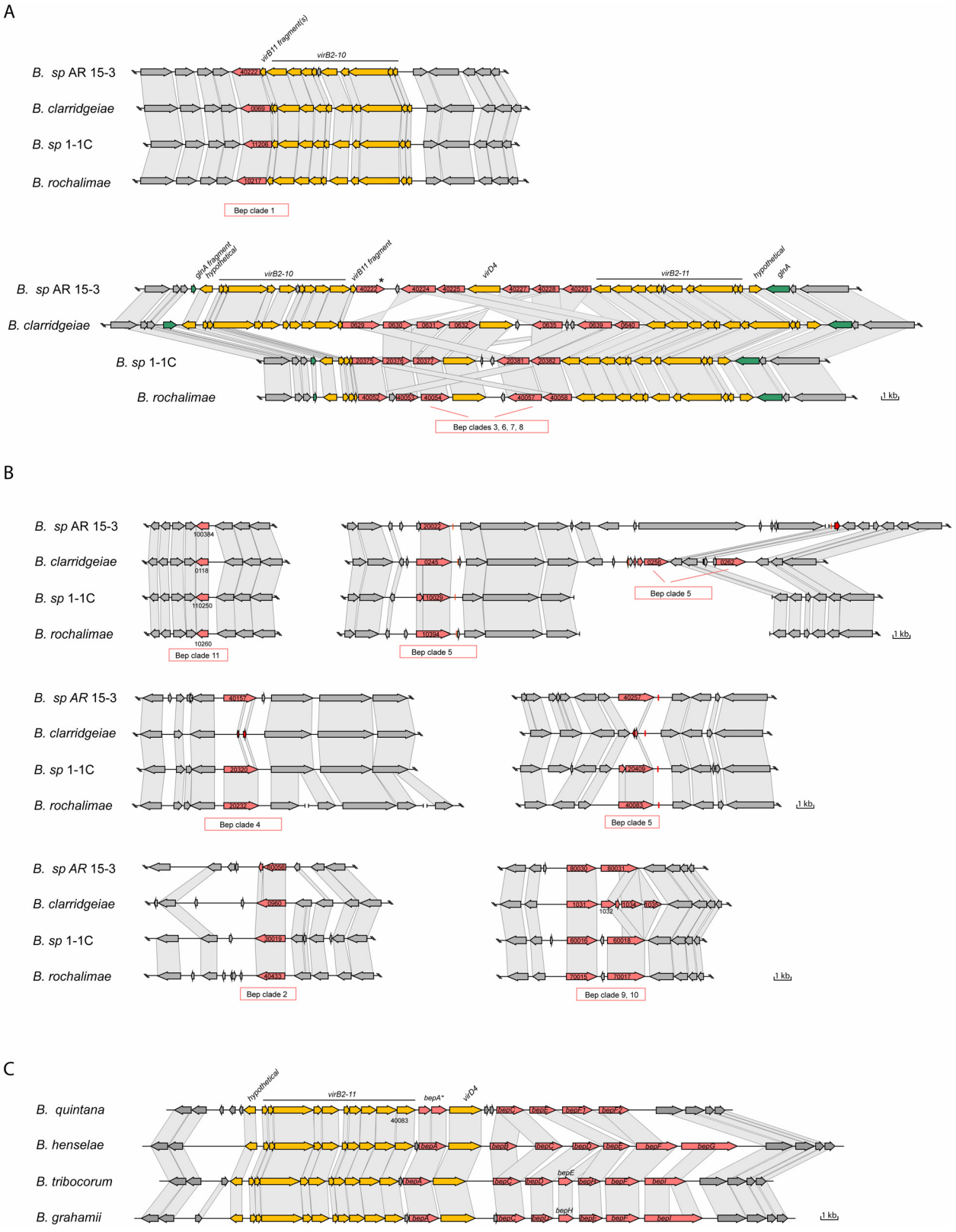


Supplementary Figure 3

Supplementary Fig 4. Domain structure of Beps from (A) lineage 3 and (B) lineage 4. The FIC and BID domains are shown in orange and violet, respectively. The FIC motif (HPFxxGNG) is depicted. Motifs differing from the consensus are indicated. Beps are grouped according to the monophyletic clades indicated in Fig. 2. BARCLv2_0629 and BARCLv2_0635 are excluded from groups, because their phylogenetic positions are not well supported in the inferred gene trees (Figs. 2, S6 and S7). Tyrosine-phosphorylation motifs depicted for Bep clade 9, BepD, BepE, BepF, and BepH were identified with the program NetPhos2.0 (*39*) using a threshold of 0.8.

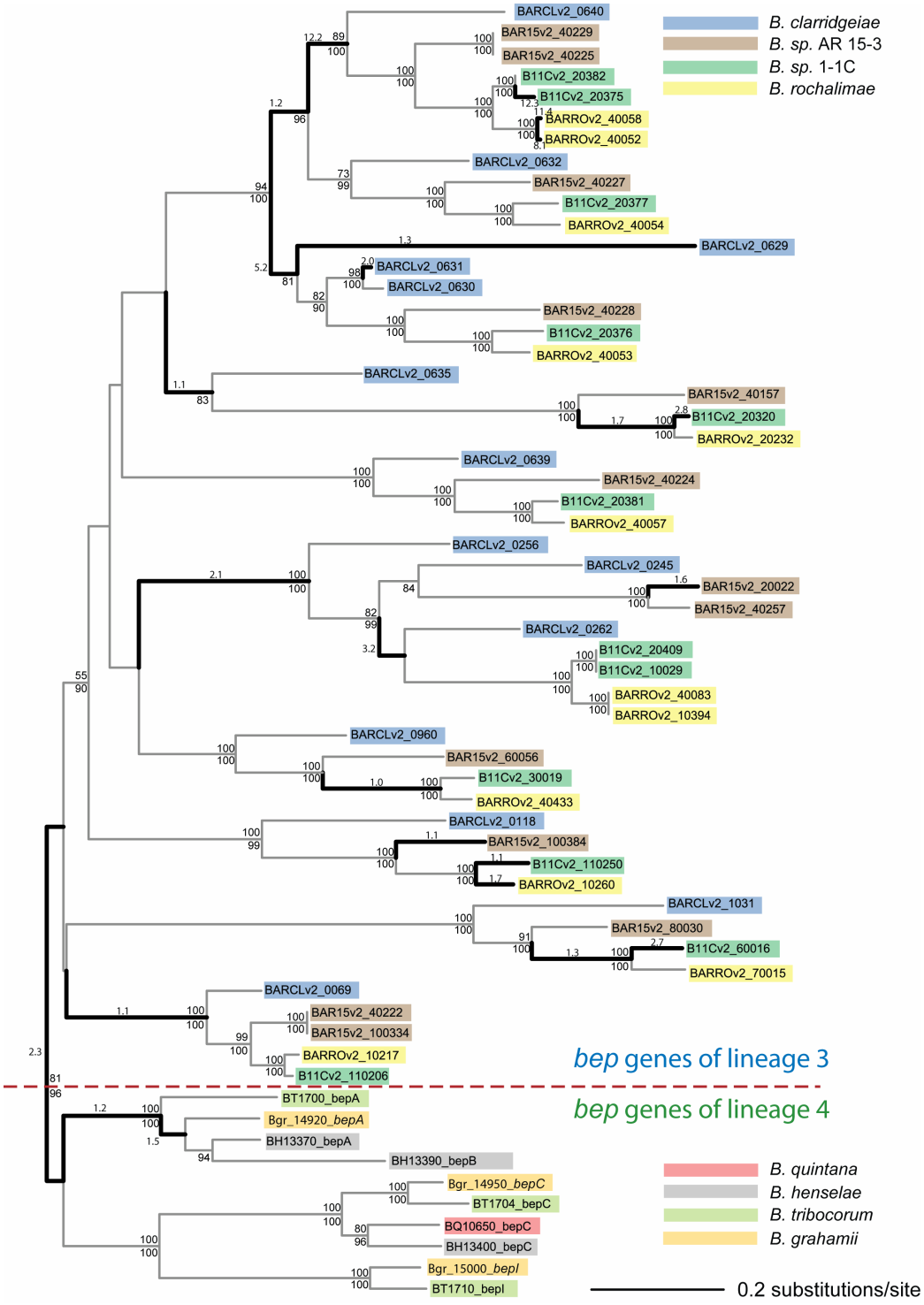
Supplementary Figure 4

Supplementary Fig. 5. Comparison of *virB* T4SS- and *bep*-encoding loci of lineage 3 (A and B) and lineage 4 (C). (A) Comparison of the two *virB* T4SS loci of *B. clarridgeiae, B. sp.* AR 15-3*, B. sp.* 1-1C*,* and *B. rochalimae.* (B) Comparison of the genomic regions of *B. clarridgeiae, B. sp* AR 15-3*, B. s*p. 1-1C*,* and *B. rochalimae* encoding additional *bep* genes. (C) Comparsion of the *virB* T4SS loci of *B. quintana, B. henselae, B. grahamii,* and *B. tribocorum*. *virB* T4SS genes are colored in yellow, *bep* genes in red, as well as flanking regions and putative coding sequences in gray. The monophyletic clades of different *bep* genes (see Fig. 2) are indicated and show that most genes, belonging to the same monophyletic clade, constitute positional orthologs of each other. Bep clade 11 does not appear in Fig. 2, since these *bep* genes consist of only the FIC domain. Genes are connected by gray boxes, if they are orthologs of each other. For *bep* genes, connections are drawn, if genes belong to the same monophyletic clade (see Fig. 2), or if they are top blast hits of each other. The *bep* gene tagged by an asterisk (BAR15v2_400222) belongs to the Bep-clade 1. In (A), the glutamine syntethase I (*glnA*) gene and its fragment flanking the two inverted *virB* T4SS copies are colored in green. In (B), contig breaks are indicated by the interruption of the sequence line.



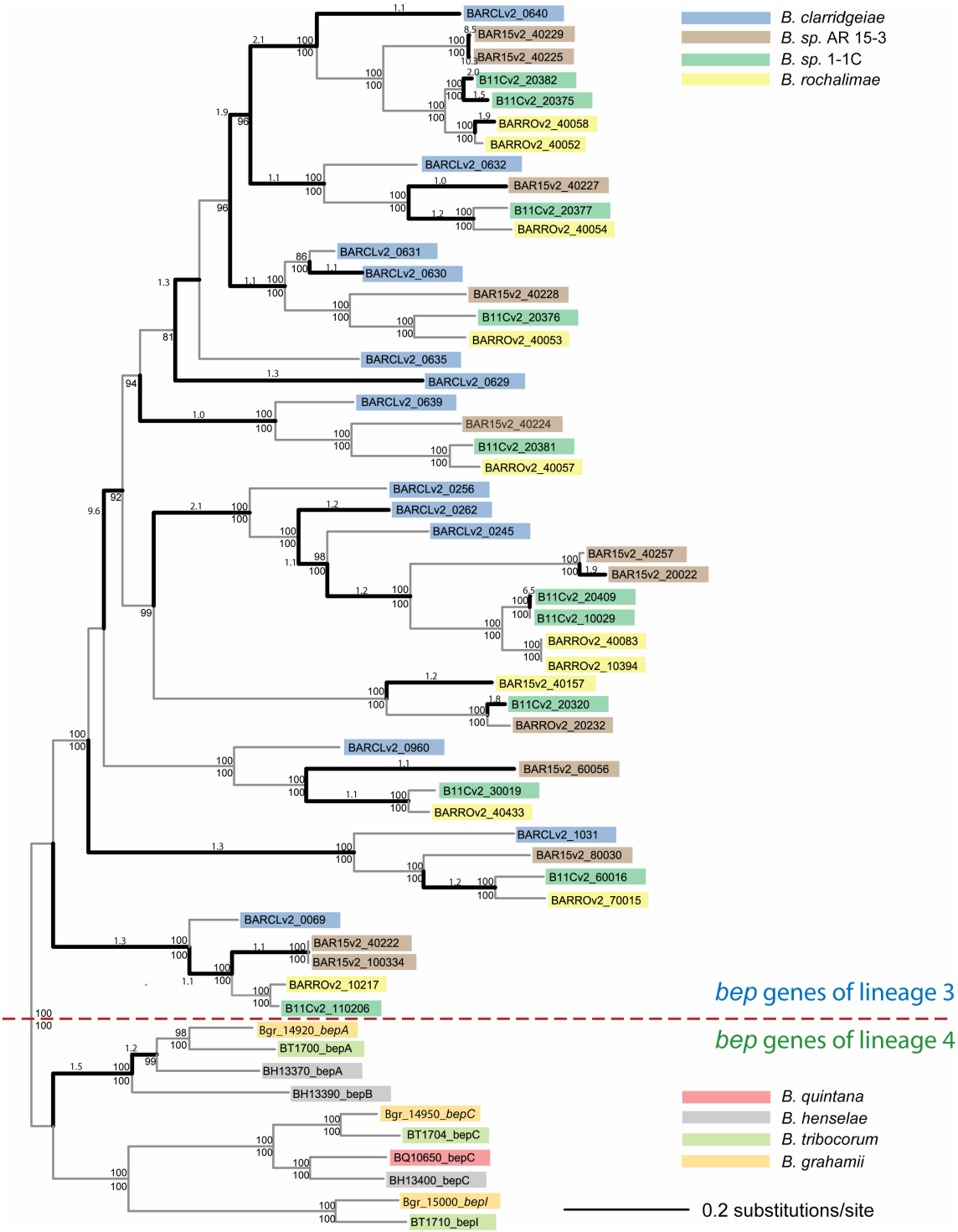
Supplementary Figure 5

Supplementary Fig. 6. *bep* gene tree based on the FIC domain-containing N-terminus. The topology and branch lengths are according to maximum likelihood analysis. Numbers above the branches represent maximum likelihood bootstraps obtained with PAUP (*33*), numbers below the branches represent posterior probabilities obtained with MrBayes (*34*). Locus_tag and gene names (if existing) are given. Beps are color-coded according to the species. The dashed line indicates the independent clustering of the Bep genes from lineage 3 and lineage 4. Bold lines in black color depict phylogenetic branches with *d*n/*d*s values > 1 as calculated with HYPHY (*37*). For each branch under positive selection, the estimated *d*n/*d*s value is indicated.



Supplementary Figure 6

Supplementary Fig. 7. (B) *bep* gene trees based on the entire sequence of genes harboring a N-terminal FIC and C-terminal BID domain. The topology and branch lengths are according to maximum likelihood analysis. Numbers above the branches represent maximum likelihood bootstraps obtained with PAUP (*33*), numbers below the branches represent posterior probabilities obtained with MrBayes (*34*). Locus_tag and gene names (if existing) are given. Beps are color-coded according to the species. The dashed line indicates the independent clustering of the Bep genes from lineage 3 and lineage 4. Bold lines in black color depict phylogenetic branches with *d*n/*d*s values > 1 as calculated with HYPHY (*37*). For each branch under positive selection, the estimated *d*n/*d*s value is indicated.



Supplementary Figure 7

Supplementary Table 1. Genomic features of analyzed *Bartonella* species and information about their reservoir host and phylogenetic lineage (see Fig. 1).

	B. bacilliformis[a]	*B. schoenbuchensis*[b]	*B. clarridgeiae*[b]	*B. rochalimae*[b]	*B. sp AR15-3*[b]	*B. sp 1-1C*[b]	*B. grahamii*[c]	*B. henselae*[a]	*B. quintana*[a]	*B. tribocorum*[a]
Reservoir host	Human	Roe deer	Cat	Dog	Red squirrel	Rat	Mouse	Cat	Human	Rat
Lineage (see Fig. 1)	1	2	3	3	3	3	4	4	4	4
Genome status	complete	draft	complete	draft	draft	draft	complete	complete	complete	complete
Genome size	1,445,021 bp	1,685,713 bp	1,522,743 bp	1,552,290 bp	1,607,248 bp	1,585,739 bp	2,341,328 bp	1,931,047 bp	1,581,384 bp	2,619,061bp
Plasmid	no	yes	no	no	no	no	yes	no	no	yes
Number of contigs	1	18	1	13	17	19	1 (1)[d]	1	1	1 (1)[d]
GC content	38,2 %	37.60%	35.70%	35.60%	35.70%	35.80%	38.10%	38.20%	38.80%	38.80%
CDS	1,283	1.584	1,342	1,405	1,487	1,465	1,737 (31)[d]	1,488	1,142	2,136 (18)[d]
rRNA	2x3	2x3	2x3	3x3	2x3	2x3	2x3	2x3	2x3	2x3
tRNA	44	42	41	44	41	41	42	44	44	42
Average CDS length	909 bp	863 bp	924 bp	878 bp	883 bp	880 bp	not analyzed	942 bp	999 bp	906 bp

[a] information adopted from *(40)*, [b] species sequenced in this study, [c] information adopted from *(23)*,

[d] Numbers in brackets refer to plasmids present in *B. tribocorum* and *B. grahamii*

Supplementary Table 2. Genes of lineage 3 with average *d*n/*d*s values (ω) ≥ 0.25. Genes are ordered according to ω. *virB* T4SS and *bep* genes are indicated by gray color. Orthologous genes present in genomes of other lineages and Likelihood ratio test (LRT) results of the PAML selection analysis are given.

			Orthologs in other *Bartonella* genome[a]					PAML results[b]	
Gene	**Product**	**Length (aa)**	***Brucella abortus***	***B. bacilliformis***	***B. schoenbuchensis***	***B. quintana***	**ω**	**M1a vs. M2**	**M7 vs. M8**
BARCLv2_1245	conserved protein of unknown function	91	BruAb1_1710	BARBAKC583_0136	BARSCv2_10259	BQ12030	16.73 ±40.3	0.00	0.00
BARCLv2_0768	conserved exported protein of unknown function	77	no	no	BARSCv2_40129	BQ06010	16.69 ±40.32	0.00	0.00
BARCLv2_0782	ATP-binding protein of ABC-transport system	245	BruAb2_0027	no	BARSCv2_40051	BQ06320	16.56 ±40.39	1.28	3.00
BARCLv2_1321	conserved protein of unknown function	85	BruAb1_1804	no	BARSCv2_10329	BQ12930	0.82 ±0.22	1.36	1.77
BARCLv2_0240	conserved exported protein of unknown function	49	no	BARBAKC583_0299	BARSCv2_20051	no	0.78 ±0.22	5.92	5.96
BARCLv2_0077	VirB T4SS protein, VirB5	150	no	no	no	BQ10560[c]	0.73 ±0.09	11.02*	12.06*
BARCLv2_0647	VirB T4SS protein, VirB5	150	no	no	no	BQ10560[c]	0.73 ±0.09	11.02*	12.06*
BARCLv2_1163	conserved protein of unknown function	97	no	no	no	BQ11040	0.65 ±0.34	2.11	2.27
BARCLv2_0505	glycine cleavage system T protein	373	BruAb2_0504	BARBAKC583_1096	BARSCv2_40500	BQ10130	0.64 ±1.18	1.18	1.41
BARCLv2_0118	VirB T4SS effector protein, Bep (fragment)	223	no	no	no	BQ10650[c]	0.62 ±0.21	13.84**	14.52**
BARCLv2_0960	VirB T4SS effector protein, Bep	430	no	no	no	BQ10650[c]	0.61 ±0.18	11.42*	14.62**
BARCLv2_0491	protein of unknown function	120	no	no	BARSCv2_10073	no	0.6 ±0.09	0.09	0.18
BARCLv2_0416	TonB protein	244	BruAb1_0622	BARBAKC583_0461	BARSCv2_20204	BQ04170	0.53 ±0.17	0.37	0.53
BARCLv2_0125	Heme exporter protein CcmD	50	no	BARBAKC583_1287	BARSCv2_10492	BQ01050	0.5 ±0.34	0.99	1.06
BARCLv2_0194	exported protein of unknown function	90	no	no	no	no	0.5 ±0.16	14.42**	18.29**
BARCLv2_0632	VirB T4SS effector protein, Bep	528	no	no	no	BQ10650[c]	0.5 ±0.12	6.51*	11.62*
BARCLv2_1031	VirB T4SS effector protein, Bep	547	no	no	no	BQ10650[c]	0.49 ±0.06	16.29**	22.75**
BARCLv2_0245	VirB T4SS effector protein, Bep	497	no	no	no	BQ10650[c]	0.48 ±0.11	10.19*	18.9*
BARCLv2_0481	putative flagellar motor protein	405	BruAb2_1079	BARBAKC583_1125	BARSCv2_10086	no	0.47 ±0.08	0.02	0.15
BARCLv2_0075	VirB T4SS protein, VirB7	103	no	no	no	BQ10580[c]	0.47 ±0.09	0.43	0.8
BARCLv2_0645	VirB T4SS protein, VirB7	103	no	no	no	BQ10580[c]	0.47 ±0.09	0.43	0.8
BARCLv2_0690	conserved protein of unknown function	77	BruAb1_0849	BARBAKC583_0864	BARSCv2_40153	BQ07510	0.47 ±0.21	6.46*	6.53*
BARCLv2_1342	protein of unknown function	556	no	no	no	no	0.46 ±0.08	35.82**	55.96**
BARCLv2_0478	conserved protein of unknown function	242	no	BARBAKC583_1128	BARSCv2_10089	no	0.46 ±0.08	0.00	0.02
BARCLv2_1207	DsbB domain protein	184	BruAb1_1630	BARBAKC583_0182	BARSCv2_10215	BQ11650	0.46 ±0.09	3.33	4.32
BARCLv2_0640	VirB T4SS effector protein, Bep	539	no	no	no	BQ10650[c]	0.45 ±0.06	18.16**	23.35**
BARCLv2_0669	conserved protein of unknown function	77	no	no	no	BQ07720	0.44 ±0.21	0.00	0.00
BARCLv2_1013	conserved protein of unknown function	158	no	no	no	BQ09010	0.43 ±0.07	2.71	3.46
BARCLv2_1355	protein of unknown function	32	no	no	no	no	0.43 ±0.35	0.00	0.00
BARCLv2_0639	VirB T4SS effector protein, Bep	625	no	no	no	BQ10650[c]	0.43 ±0.01	19.73**	23.36**
BARCLv2_1288	conserved protein of unknown function	225	BruAb1_1970	BARBAKC583_0098	BARSCv2_10295	BQ12360	0.43 ±0.11	0.00	0.00
BARCLv2_1327	conserved protein of unknown function	387	BruAb1_1862	BARBAKC583_0063	BARSCv2_10335	BQ13000	0.42 ±0.08	5.28	5.85
BARCLv2_1059	conserved protein of unknown function	141	BruAb1_0406	BARBAKC583_0375	BARSCv2_20121	BQ03120	0.42 ±0.09	5.04	5.44
BARCLv2_0241	conserved protein of unknown function	131	BruAb2_0470	BARBAKC583_0300	BARSCv2_20052	BQ02980	0.42 ±0.35	0.00	0.00
BARCLv2_0982	conserved protein of unknown function	464	no	no	no	no	0.42 ±0.1	59.2**	94.79**
BARCLv2_0631	VirB T4SS effector protein, Bep	538	no	no	no	BQ10650[c]	0.42 ±0.06	15.09**	22.57**
BARCLv2_1351	conserved protein of unknown function	221	BruAb1_2010	no	no	no	0.42 ±0.13	0.00	-0.07
BARCLv2_0069	VirB T4SS effector protein, Bep	547	no	no	no	BQ10650[c]	0.42 ±0.1	2.46	3.4
BARCLv2_0475	Inducible Bartonella autotransporter	823	BruAb1_0895	BARBAKC583_1132[c]	BARSCv2_10091[c]	BQ10380[c]	0.42 ±0.17	35.06**	54.3**
BARCLv2_0978	conserved protein of unknown function	40	BruAb1_1401	BARBAKC583_0931	BARSCv2_40323	BQ08690	0.41 ±0.37	0.92	1.20

			Orthologs in other *Bartonella* genome[a]					PAML results[b]	
Gene	**Product**	**Length (aa)**	***Brucella abortus***	***B. bacilliformis***	***B. schoenbuchensis***	***B. quintana***	**ω**	**M1a vs. M2**	**M7 vs. M8**
BARCLv2_0279	conserved protein of unknown function	218	no	BARBAKC583_1049	BARSCv2_40463	BQ09660	0.41 ±0.12	0.00	0.00
BARCLv2_1164	conserved protein of unknown function	139	BruAb1_1640	BARBAKC583_0248	BARSCv2_10151	BQ11050	0.41 ±0.07	3.14	3.26
BARCLv2_1170	conserved protein of unknown function	74	no	BARBAKC583_0239	BARSCv2_10161	no	0.4 ±0.25	0.33	0.63
BARCLv2_1173	conserved protein of unknown function	152	no	BARBAKC583_0234	BARSCv2_10165	BQ11080	0.4 ±0.08	2.39	2.47
BARCLv2_0957	Effector protein yopJ	249	no	no	no	BQ11580[c]	0.4 ±0.07	5.23	5.99*
BARCLv2_0250	Effector protein yopJ	279	no	no	no	BQ11580[c]	0.39 ±0.1	6.20*	8.93*
BARCLv2_0367	conserved protein of unknown function	164	BruAb1_0516	BARBAKC583_0416	BARSCv2_20162	BQ03740	0.38 ±0.34	0.00	0.20
BARCLv2_1270	ATP synthase epsilon chain	138	BruAb1_1778	BARBAKC583_0115	BARSCv2_10281	BQ12220	0.38 ±0.08	1.00	1.30
BARCLv2_0473	Inducible Bartonella autotransporter E protein (fragment)	673	no	BARBAKC583_1132[c]	BARSCv2_10091[c]	BQ10380[c]	0.38 ±0.11	108.28**	128.21**
BARCLv2_1329	Antioxidant, AhpC/TSA family	143	BruAb1_1956	BARBAKC583_0053	BARSCv2_10347	BQ13100	0.37 ±0.22	-9.41	6.16
BARCLv2_0664	conserved protein of unknown function	210	no	BARBAKC583_0744	BARSCv2_40233	no	0.37 ±0.17	0.78	1.60
BARCLv2_1005	conserved protein of unknown function	134	BruAb1_1433	BARBAKC583_0954	BARSCv2_40346	BQ08930	0.37 ±0.08	0.83	0.83
BARCLv2_0054	conserved protein of unknown function	242	BruAb1_2088	BARBAKC583_1339	BARSCv2_10433	BQ00450	0.36 ±0.19	1.53	2.04
BARCLv2_0882	conserved protein of unknown function	249	BruAb1_1170	BARBAKC583_0582	BARSCv2_20329	BQ07020	0.36 ±0.1	0.00	0.00
BARCLv2_1123	Effector protein yopJ	339	no	no	no	BQ01158[c]	0.35 ±0.13	6.19*	8.68*
BARCLv2_1011	conserved protein of unknown function	107	BruAb1_1447	no	no	no	0.35 ±0.34	9.59*	10.29*
BARCLv2_0115	FxsA cytoplasmic membrane protein	157	BruAb1_2048	BARBAKC583_1295	BARSCv2_10481	BQ00960	0.35 ±0.12	2.27	3.00
BARCLv2_0929	50S ribosomal protein L29	66	BruAb1_1230	BARBAKC583_0674	BARSCv2_40270	BQ08150	0.35 ±0.21	0.16	0.38
BARCLv2_0951	conserved protein of unknown function	225	BruAb1_0953	BARBAKC583_0651	BARSCv2_40294	BQ08380	0.35 ±0.29	3.82	3.79
BARCLv2_0446	conserved protein of unknown function	125	no	BARBAKC583_1158	BARSCv2_10119	no	0.34 ±0.15	0.00	0.09
BARCLv2_0480	putative flagellar motor protein	426	BruAb2_1080	BARBAKC583_1126	BARSCv2_10087	no	0.34 ±0.09	0.18	0.50
BARCLv2_1038	conserved protein of unknown function	144	BruAb1_0895	no	no	BQ04750	0.34 ±0.1	2.25	2.88
BARCLv2_0699	conserved protein of unknown function	85	BruAb1_1043	BARBAKC583_0854	BARSCv2_40141	BQ07410	0.34 ±0.07	0.01	0.26
BARCLv2_0566	conserved exported protein of unknown function	875	no	no	no	no	0.34 ±0.12	48.71**	63.89**
BARCLv2_1165	conserved protein of unknown function	228	BruAb1_1844	BARBAKC583_0247	no	BQ11000	0.34 ±0.08	4.53	5.07
BARCLv2_1200	conserved protein of unknown function	126	no	BARBAKC583_0193	BARSCv2_10208	BQ11450	0.33 ±0.15	1.08	1.44
BARCLv2_1199	conserved protein of unknown function	63	no	BARBAKC583_0195	BARSCv2_10207	BQ11430	0.33 ±0.15	0.00	0.00
BARCLv2_1225	conserved exported protein of unknown function	352	BruAb1_1693	BARBAKC583_0156	BARSCv2_10240	BQ11930	0.33 ±0.11	0.01	0.76
BARCLv2_0670	conserved protein of unknown function	813	no	BARBAKC583_0748	BARSCv2_40172	BQ07710	0.32 ±0.03	4.85	6.04*
BARCLv2_0198	Hemin binding protein	291	BruAb1_0115[c]	BARBAKC583_1214[c]	BARSCv2_40484[c]	BQ02410[c]	0.32 ±0.04	9.37*	17.43*
BARCLv2_0196	Hemin binding protein	277	BruAb1_0115[c]	BARBAKC583_1214[c]	BARSCv2_10559[c]	BQ02410[c]	0.32 ±0.08	8.84*	8.47*
BARCLv2_0016	conserved exported protein of unknown function	199	BruAb1_0152	BARBAKC583_1369	BARSCv2_10401	BQ00150	0.32 ±0.21	1.79	1.93
BARCLv2_1166	conserved protein of unknown function	297	no	no	BARSCv2_10157	BQ11020	0.32 ±0.1	3.22	4.86
BARCLv2_1020	conserved protein of unknown function	837	no	BARBAKC583_1109	no	BQ10280	0.32 ±0.13	0.00	14.52
BARCLv2_0420	conserved protein of unknown function	175	BruAb1_0433	BARBAKC583_0391	BARSCv2_20137	BQ03450	0.31 ±0.07	0.00	0.00
BARCLv2_0103	3-phosphoshikimate 1-carboxyvinyltransferase	442	BruAb1_0025	BARBAKC583_1303	BARSCv2_10474	BQ00880	0.31 ±0.06	7.81*	9.68*
BARCLv2_0221	Protein slyX homolog	50	no	no	no	no	0.31 ±0.12	0.00	0.00
BARCLv2_0867	ABC transporter, permease protein	232	no	BARBAKC583_0597	BARSCv2_20344	BQ06870	0.3 ±0.1	2.19	2.56
BARCLv2_1177	conserved protein of unknown function	283	no	BARBAKC583_0229	BARSCv2_10169	BQ11120	0.3 ±0.07	0.88	2.30
BARCLv2_0327	conserved protein of unknown function	84	BruAb1_1516	BARBAKC583_0993	BARSCv2_40420	BQ03350	0.3 ±0.12	0.20	0.27
BARCLv2_0847	Lysophospholipase l2	293	BruAb1_0046	no	no	BQ05220	0.3 ±0.07	0.00	0.00
BARCLv2_1180	conserved protein of unknown function	87	no	BARBAKC583_0226	BARSCv2_10177	BQ11160	0.3 ±0.06	0.25	1.15
BARCLv2_0877	Phosphatidate cytidylyltransferase	269	BruAb1_1163	BARBAKC583_0587	BARSCv2_20334	BQ06970	0.3 ±0.09	0.39	1.13
BARCLv2_0781	permease protein of ABC transporter	296	BruAb1_1771	no	BARSCv2_40052	BQ06310	0.3 ±0.42	2.31	3.25
BARCLv2_1274	ATP synthase delta subunit	194	BruAb1_1782	BARBAKC583_0111	BARSCv2_10285	BQ12260	0.3 ±0.04	2.17	3.83
BARCLv2_1261	conserved protein of unknown function	95	no	no	BARSCv2_10278	BQ12190	0.3 ±0.13	0.00	0.00
BARCLv2_0197	Hemin binding protein	241	BruAb1_0115[c]	BARBAKC583_1214[c]	BARSCv2_40297[c]	BQ02410[c]	0.29 ±0.04	35.92**	50.26**
BARCLv2_0381	conserved protein of unknown function	172	BruAb1_0603	no	no	BQ03880	0.29 ±0.1	0.41	0.84
BARCLv2_0292	conserved protein of unknown function	215	no	no	no	BQ09490	0.29 ±0.1	2.26	3.5
BARCLv2_0559	conserved membrane protein of unknown function	363	BruAb1_0705	BARBAKC583_0507	BARSCv2_20252	BQ04620	0.29 ±0.04	6.34*	7.89*
BARCLv2_1065	mutator MutT protein	136	BruAb1_1915	BARBAKC583_0364	BARSCv2_20107	BQ01900	0.28 ±0.19	3.35	4.25

			Orthologs in other *Bartonella* genome[a]					PAML results[b]	
Gene	**Product**	**Length (aa)**	***Brucella abortus***	***B. bacilliformis***	***B. schoenbuchensis***	***B. quintana***	**ω**	**M1a vs. M2**	**M7 vs. M8**
BARCLv2_0524	pantoate--beta-alanine ligase	281	BruAb1_0355	BARBAKC583_0476	BARSCv2_20219	BQ04310	0.28 ±0.09	0.00	0.00
BARCLv2_1125	50S ribosomal protein L21	159	BruAb1_1829	BARBAKC583_0309	BARSCv2_20058	BQ01320	0.28 ±0.06	1.80	3.18
BARCLv2_0024	lysyl-tRNA synthetase	549	BruAb2_0442	BARBAKC583_1360	BARSCv2_10409	BQ00220	0.28 ±0.44	2.84	4.69
BARCLv2_1236	TolA protein	377	BruAb1_1683	BARBAKC583_0145	BARSCv2_10253	BQ11830	0.28 ±0.05	0.20	0.78
BARCLv2_1313	SurF1 family protein	255	BruAb1_0496	BARBAKC583_0076	BARSCv2_10321	BQ12840	0.28 ±0.05	0.00	0.22
BARCLv2_0171	Ribosomal-protein-alanine acetyltransferase	160	BruAb1_2125	BARBAKC583_1242	BARSCv2_10535	BQ02150	0.28 ±0.13	2.03	2.25
BARCLv2_0518	conserved membrane protein of unknown function	179	BruAb1_0675	BARBAKC583_0470	BARSCv2_20213	BQ04250	0.28 ±0.03	6.16*	9.17*
BARCLv2_1233	conserved protein of unknown function	152	BruAb1_1686	BARBAKC583_0148	BARSCv2_10250	BQ11860	0.28 ±0.07	0.08	0.55
BARCLv2_0949	conserved protein of unknown function	181	BruAb1_0955	BARBAKC583_0653	BARSCv2_40292	BQ08360	0.27 ±0.07	0.00	0.00
BARCLv2_0226	conserved protein of unknown function	473	no	BARBAKC583_0286	BARSCv2_20039	BQ02810	0.27 ±0.06	0.00	0.51
BARCLv2_0464	Flagellar hook-basal body complex protein fliE	112	BruAb2_0149	BARBAKC583_1142	BARSCv2_10102	no	0.27 ±0.16	2.26	3.12
BARCLv2_1107	Filament-A percursor	391	BruAb1_1817	BARBAKC583_0323	BARSCv2_20069	BQ01520	0.27 ±0.04	0.62	2.17
BARCLv2_0296	conserved protein of unknown function	128	BruAb1_1587	BARBAKC583_1018	BARSCv2_40443	BQ09450	0.27 ±0.06	3.57	4.40
BARCLv2_1082	glycerol-3-phosphate ABC transporter permease	283	BruAb2_0569	BARBAKC583_0347	BARSCv2_20091	BQ01740	0.27 ±0.5	0.00	0.00
BARCLv2_0468	conserved protein of unknown function	210	BruAb2_0153	BARBAKC583_1138	BARSCv2_10098	no	0.27 ±0.07	1.35	2.53
BARCLv2_0454	Flagellar motor switch protein fliY	170	BruAb2_1067	BARBAKC583_1151	BARSCv2_10111	no	0.27 ±0.03	0.44	1.02
BARCLv2_0397	Methyltranferase	162	no	no	BARSCv2_20184	BQ04020	0.26 ±0.06	0.00	0.16
BARCLv2_1029	LysM/M23 peptidase domain protein	362	BruAb1_0900	BARBAKC583_0521	BARSCv2_20269	BQ04790	0.26 ±0.04	0.28	1.27
BARCLv2_0813	conserved protein of unknown function	126	BruAb1_1079	BARBAKC583_0643	BARSCv2_20395	BQ05530	0.26 ±0.2	0.00	0.00
BARCLv2_0063	conserved protein of unknown function	144	no	BARBAKC583_1330	BARSCv2_10443	BQ00580	0.26 ±0.08	0.00	0.00
BARCLv2_1062	ATP synthase B chain (modular protein)	210	BruAb1_0409	BARBAKC583_0378	BARSCv2_20124	BQ03150	0.26 ±0.06	0.36	1.04
BARCLv2_0618	conserved protein of unknown function	246	no	no	no	BQ10520[c]	0.26 ±0.04	0.00	0.97
BARCLv2_0651	VirB T4SS-associated conserved protein of unknown function	246	no	no	no	BQ10520[c]	0.26 ±0.04	0.00	0.46
BARCLv2_1326	conserved protein of unknown function	481	BruAb1_1861	BARBAKC583_0064	BARSCv2_10334	BQ12990	0.26 ±0.05	2.91	4.32
BARCLv2_1078	ABC transporter, periplasmic binding protein	300	no	BARBAKC583_0350	BARSCv2_20094	BQ01770	0.26 ±0.08	3.07	5.27
BARCLv2_1246	conserved protein of unknown function	121	BruAb1_1709	BARBAKC583_0135	BARSCv2_10260	BQ12040	0.25 ±0.08	0.71	1.19
BARCLv2_0390	Cytochrome c-type biogenesis protein ccmH	125	BruAb1_0629	BARBAKC583_0433	BARSCv2_20179	BQ03960	0.25 ±0.04	0.00	0.00
BARCLv2_0325	conserved protein of unknown function	350	BruAb1_1518	BARBAKC583_0995	BARSCv2_40422	BQ03330	0.25 ±0.08	0.00	0.00
BARCLv2_0892	Transcriptional regulator	128	BruAb1_1185	no	no	no	0.25 ±0.08	0.00	0.00
BARCLv2_0449	conserved exported protein of unknown function	147	BruAb2_1064	BARBAKC583_1155	BARSCv2_10115	no	0.25 ±0.07	3.02	3.88
BARCLv2_1254	conserved protein of unknown function	506	BruAb1_1754	BARBAKC583_0125	BARSCv2_10271	BQ12130	0.25 ±0.08	6.03*	8.46*
BARCLv2_1171	putative phage protein	328	no	BARBAKC583_0238	BARSCv2_10163	BQ11060	0.25 ±0.03	20.56**	23.36**
BARCLv2_0059	heat shock protein GrpE	219	BruAb1_0167	BARBAKC583_1334	BARSCv2_10438	BQ00500	0.25 ±0.16	9.59*	12.51*
BARCLv2_0380	conserved protein of unknown function	162	BruAb1_0602	no	no	BQ03870	0.25 ±0.07	2.26	2.39
BARCLv2_0145	conserved exported protein of unknown function	608	no	no	no	no	0.25 ±0.05	3.24	7.43
BARCLv2_0735	conserved protein of unknown function	96	BruAb2_0644	BARBAKC583_0797	BARSCv2_40109	BQ05830	0.25 ±0.16	5.73	5.83
BARCLv2_0191	Formamidopyrimidine-DNA glycosylase	291	BruAb1_2156	BARBAKC583_1221	BARSCv2_10554	BQ02370	0.25 ±0.17	1.54	1.94
BARCLv2_1278	conserved protein of unknown function	190	BruAb1_1786	BARBAKC583_0105	BARSCv2_10290	BQ12300	0.25 ±0.08	0.00	0.00
BARCLv2_1227	OpgC protein	373	BruAb1_1691	BARBAKC583_0154	BARSCv2_10242	BQ11910	0.25 ±0.06	3.84	4.08
BARCLv2_0565	conserved exported protein of unknown function	469	no	BARBAKC583_0512	no	no	0.25 ±0.05	0.00	2.35
BARCLv2_0605	conserved protein of unknown function	182	no	no	no	no	0.25 ±0.24	0.47	0.09
BARCLv2_1373	Transmembrane protein	150	BruAb2_0497	no	no	BQ13370	0.25 ±0.05	0.70	1.34

[a] for ortholog assignment the “Phylo Profile Synteny” tool provided on the MaGe platform (*30*) was used (parameters: Bidirectional Blast Hit, 30 % identities, 0.6 for minLrap and maxLrap), automated ortholog assignments were manually curated, [b] LRT statistics for model 1a versus model 2 and model 7 versus model 8, values indicating significant positive selection are depicted by asterisks (*: p <0.05, **: p <0.001); [c] more than one orthologous gene is present in the corresponding genome, no clear ortholog assignment possible

Supplementary Table 3. Genes of Lineage 4 with average *d*n/*d*s values (ω) ≥ 0.25. Genes are ordered according to ω. *virB* T4SS and *bep* genes are indicated by gray color. Orthologous genes present in genomes of other lineages and Likelihood ratio test (LRT) results of the PAML selection analysis are given.

			Orthologs in other genomes[a]					CodeML results[b]	
Gene	Product	Length (aa)	*Brucella abortus*	*B. bacilliformis*	*B. schoenbuchensis*	*B. clarridgeiae*	ω	M1a vs M2	M7 vs M8
BQ12460	conserved protein of unknown function	203	no	no	no	no	0.9 ±0.08	6.18*	7.35*
BQ04750	conserved protein of unknown function	134	no	no	no	BARCLv2_1038	0.84 ±0.09	4.26	5.16
BQ10680	VirB T4SS effector protein, BepF	555	no	no	no	BARCLv2_0069[c]	0.58 ±0.1	9.67*	21.70**
BQ03120	conserved protein of unknown function	135	BruAb1_0406	BARBAKC583_0375	BARSCv2_20121	BARCLv2_1059	0.57 ±0.08	2.13	3.60
BQ01050	conserved protein of unknown function	83	no	BARBAKC583_1287	BARSCv2_10492	BARCLv2_0125	0.5 ±0.24	0.00	0.00
BQ09660	conserved protein of unknown function	228	no	no	no	BARCLv2_0279	0.49 ±0.11	1.72	3.54
BQ12470	Trw T4SS protein, TrwN	189	no	no	no	no	0.48 ±0.1	13.35*	16.79**
BQ11430	conserved protein of unknown function	62	no	BARBAKC583_0195	BARSCv2_10207	BARCLv2_1199	0.46 ±0.18	2.47	3.88
BQ03700	conserved exported protein of unknown function	111	no	no	no	no	0.45 ±0.1	0.00	0.07
BQ12560	Trw T4SS protein, TrwL5	102	no	no	no	no	0.43 ±0.18	7.42*	8.76*
BQ13300	conserved exported protein of unknown function	129	no	no	no	no	0.41 ±0.13	1.20	2.10
BQ04170	TonB protein	248	BruAb1_0622	BARBAKC583_0461	BARSCv2_20204	BARCLv2_0416	0.41 ±0.1	10.25*	15.14**
BQ12930	Chorismate mutase	110	BruAb1_1804	no	BARSCv2_10329	BARCLv2_1321	0.4 ±0.09	2.55	3.34
BQ11180	conserved protein of unknown function	83	BruAb1_0275	BARBAKC583_0213	no	BARCLv2_1184	0.4 ±0.18	2.45	4.83
BQ10970	conserved protein of unknown function	67	no	BARBAKC583_0251	BARSCv2_10148	BARCLv2_1160	0.4 ±0.16	0.00	0.00
BQ09320	conserved exported protein of unknown function	99	BruAb1_1503	BARBAKC583_0985	no	BARCLv2_0336	0.39 ±0.37	0.00	0.00
BQ12610	Trw T4SS protein, TrwH1	47	no	no	no	no	0.37 ±0.09	5.66	6.75*
BQ04650	conserved exported protein of unknown function	87	no	BARBAKC583_0511	BARSCv2_20256	no	0.37 ±0.14	0.00	0.12
BQ12600	Trw T4SS protein, TrwI1	281	no	no	no	no	0.36 ±0.05	0.00	0.00
BQ11050	conserved protein of unknown function	148	BruAb1_1640	BARBAKC583_0248	BARSCv2_10151	BARCLv2_1164	0.35 ±0.13	0.17	0.63
BQ06010	conserved protein of unknown function	103	no	no	BARSCv2_40129	BARCLv2_0768	0.35 ±0.14	1.27	0.00
BQ11480	conserved protein of unknown function	290	no	BARBAKC583_0188	BARSCv2_10210	BARCLv2_1202	0.35 ±0.08	6.81*	12.38*
BQ00140	conserved exported protein of unknown function	213	BruAb1_0151	BARBAKC583_1370	BARSCv2_10400	BARCLv2_0015	0.35 ±0.05	0.15	1.95
BQ03410	conserved exported protein of unknown function	205	BruAb1_1507	BARBAKC583_0987	BARSCv2_40414	no	0.35 ±0.16	0.00	3.56
BQ09010	conserved protein of unknown function	176	no	no	no	BARCLv2_1013	0.35 ±0.09	2.67	3.54
BQ07710	conserved protein of unknown function	819	no	BARBAKC583_0748	BARSCv2_40172	BARCLv2_0670	0.35 ±0.05	11.50*	16.37**
BQ01360	conserved protein of unknown function	73	no	no	no	no	0.34 ±0.13	0.00	0.29
BQ10580	VirB T4SS protein, VirB7	100	no	no	no	BARCLv2_0075[c]	0.34 ±0.08	1.13	2.54
BQ10660	VirB T4SS effector protein, BepD	414	no	no	no	BARCLv2_0069[c]	0.33 ±0.07	0.00	0.87
BQ06720	conserved exported protein of unknown function	109	no	no	no	no	0.33 ±0.08	0.00	1.27
BQ11620	conserved protein of unknown function	70	no	BARKC0786	no	no	0.32 ±0.06	0.00	0.00
BQ11600	conserved exported protein of unknown function	74	no	BARBAKC583_1056[c]	BARSCv2_20032[c]	no	0.32 ±0.19	0.89	2.96
BQ10410	conserved exported protein of unknown function	792	no	BARBAKC583_1132[c]	BARSCv2_10091	BARCLv2_0476[c]	0.32 ±0.09	6.74*	20.36**
BQ08690	conserved protein of unknown function	60	BruAb1_1401	BARBAKC583_0931	BARSCv2_40323	BARCLv2_0978	0.32 ±0.27	5.19	5.20
BQ01620	conserved exported protein of unknown function	353	BruAb2_0057	no	no	no	0.32 ±0.05	0.00	0.00
BQ10630	VirB T4SS effector protein, BepA	272	no	no	no	BARCLv2_0069[c]	0.32 ±0.09	0.00	2.47
BQ02210	conserved exported protein of unknown function	134	no	no	no	no	0.31 ±0.08	4.14	7.18*
BQ12190	conserved protein of unknown function	93	no	no	BARSCv2_10278	BARCLv2_1261	0.31 ±0.14	0.00	0.56
BQ11040	conserved protein of unknown function	114	no	no	no	BARCLv2_1163	0.31 ±0.08	0.00	0.00
BQ11100	conserved protein of unknown function	231	no	no	BARSCv2_10167	BARCLv2_1175	0.31 ±0.07	4.52	9.25*

			Orthologs in other genomes[a]					CodeML results[b]	
Gene	Product	Length (aa)	*Brucella abortus*	*B. bacilliformis*	*B. schoenbuchensis*	*B. clarridgeiae*	ω	M1a vs M2	M7 vs M8
BQ13000	conserved protein of unknown function	389	BruAb1_1862	BARBAKC583_0063	BARSCv2_10335	BARCLv2_1327	0.31 ±0.06	1.57	4.80
BQ13100	Thiol 3-disulfide interchange protein	226	BruAb1_1956	BARBAKC583_0053	BARSCv2_10347	BARCLv2_1329	0.3 ±0.06	0.13	0.70
BQ05360	conserved protein of unknown function	584	no	no	BARSCv2_20378	BARCLv2_0831	0.3 ±0.06	0.00	0.95
BQ12260	ATP synthase delta chain	194	BruAb1_1782	BARBAKC583_0111	BARSCv2_10285	BARCLv2_1274	0.3 ±0.07	0.13	1.18
BQ05570	conserved protein of unknown function	164	no	BARBAKC583_0647	BARSCv2_20400	BARCLv2_0810	0.3 ±0.13	22.22**	25.47**
BQ07410	conserved protein of unknown function	157	no	BARBAKC583_0854	BARSCv2_40141	BARCLv2_0699	0.29 ±0.05	5.81	13.24*
BQ03690	conserved exported protein of unknown function	118	no	no	no	no	0.29 ±0.08	1.17	1.65
BQ04440	colicin V production protein	189	BruAb1_0470	BARBAKC583_0487	BARSCv2_20233	BARCLv2_0541	0.29 ±0.04	0.22	2.08
BQ10530	VirB T4SS protein, VirB2	104	no	no	no	BARCLv2_0080[c]	0.29 ±0.09	0.00	0.00
BQ02150	ribosomal-protein-alanine acetyltransferase	160	BruAb1_2125	BARBAKC583_1242	BARSCv2_10535	BARCLv2_0171	0.29 ±0.07	0.00	0.00
BQ13290	conserved exported protein of unknown function	127	BruAb1_0462	BARBAKC583_0036	BARSCv2_10358	BARCLv2_1361	0.28 ±0.06	1.23	3.51
BQ10840	orotidine 5'-phosphate decarboxylase	280	no	no	no	no	0.28 ±0.27	0.00	2.83
BQ01520	filament-A percursor	424	BruAb1_1817	BARBAKC583_0323	BARSCv2_20069	BARCLv2_1107	0.28 ±0.04	0.00	0.65
BQ08360	conserved protein of unknown function	184	BruAb1_0955	BARBAKC583_0653	BARSCv2_40292	BARCLv2_0949	0.28 ±0.03	5.59	7.19*
BQ10650	VirB T4SS effector protein, BepC	532	no	no	no	BARCLv2_0069[c]	0.28 ±0.02	0.00	5.30
BQ03680	conserved exported protein of unknown function	117	no	no	no	no	0.28 ±0.05	0.00	2.18
BQ00120	conserved exported protein of unknown function	89	BruAb1_0148	BARBAKC583_1372	BARSCv2_10398	BARCLv2_0013	0.28 ±0.11	2.10	4.03
BQ05220	lysophospholipase L2	300	BruAb2_0046	no	no	BARCLv2_0847	0.28 ±0.08	0.00	1.49
BQ13310	Na+/H+ antiporter	133	BruAb2_0717	BARBAKC583_0035	BARSCv2_10359	no	0.27 ±0.05	0.85	2.19
BQ12110	conserved exported protein of unknown function	480	BruAb1_1746	BARBAKC583_0127	BARSCv2_10269	BARCLv2_1252	0.27 ±0.05	0.09	2.72
BQ03350	conserved protein of unknown function	85	BruAb1_1516	BARBAKC583_0993	BARSCv2_40420	BARCLv2_0327	0.27 ±0.13	3.11	4.24
BQ02430	Hemin binding protein, HbpB	382	no	no	BARSCv2_40088	no	0.27 ±0.08	67.57**	82.87**
BQ03940	Cytochrome c maturation protein E, Heme chaperone ccmE	159	BruAb1_0627	BARBAKC583_0431	BARSCv2_20177	BARCLv2_0388	0.27 ±0.06	0.00	0.00
BQ07900	conserved exported protein of unknown function	181	no	BARBAKC583_0732	BARSCv2_40244	no	0.27 ±0.06	0.00	0.28
BQ02450	conserved exported protein of unknown function	160	BruAb2_0198	no	BARSCv2_10563	BARCLv2_0200	0.27 ±0.05	1.81	3.65
BQ02710	predicted membrane protein	212	BruAb1_1363	BARBAKC583_1194	BARSCv2_10008	BARCLv2_0216	0.27 ±0.06	1.08	2.68
BQ05460	conserved protein of unknown function	183	BruAb1_0790	BARBAKC583_0636	BARSCv2_20388	BARCLv2_0821	0.27 ±0.08	0.00	0.02
BQ04740	sec-independent protein translocase	103	BruAb1_0894	no	no	BARCLv2_1039	0.27 ±0.07	0.54	1.59
BQ03900	conserved protein of unknown function	77	BruAb1_0622	BARBAKC583_0427	BARSCv2_20172	BARCLv2_0383	0.26 ±0.12	0.00	0.00
BQ11060	Phage tail protein	260	no	BARBAKC583_0238	BARSCv2_10163	BARCLv2_1171	0.26 ±0.04	0.00	0.00
BQ10290	autotransporter	1753	no	no	no	no	0.26 ±0.05	282.17**	345.48**
BQ12220	ATP synthase epsilon chain	138	BruAb1_1778	BARBAKC583_0115	BARSCv2_10281	BARCLv2_1270	0.26 ±0.07	0.00	0.70
BQ12590	Trw T4SS protein, TrwJ1	240	no	no	no	no	0.26 ±0.05	13.80*	20.03**
BQ10850	conserved membrane protein of unknown function	357	no	no	no	no	0.26 ±0.06	0.00	1.24
BQ00960	conserved membrane protein of unknown function	163	BruAb1_2048	BARBAKC583_1295	BARSCv2_10481	BARCLv2_0115	0.26 ±0.05	1.63	4.89
BQ11450	conserved protein of unknown function	129	no	BARBAKC583_0193	BARSCv2_10208	BARCLv2_1200	0.26 ±0.1	0.00	2.28
BQ02420	Hemin binding protein, HbpA	267	BruAb1_0115[c]	BARBAKC583_1214[c]	BARSCv2_10560[c]	BARCLv2_0196[c]	0.25 ±0.03	19.27**	24.17**
BQ12970	conserved protein of unknown function	327	no	no	no	no	0.25 ±0.05	0.00	0.00
BQ10280	autotransporter	799	no	BARBAKC583_1109	BARSCv2_10011	BARCLv2_1020	0.25 ±0.06	0.00	14.65*
BQ10380	Inducible Bartonella autotransporter	759	no	BARBAKC583_1132[c]	BARSCv2_10092	BARCLv2_0475[c]	0.25 ±0.04	0.00	0.00
BQ02810	conserved protein of unknown function	447	no	BARBAKC583_0286	BARSCv2_20039	BARCLv2_0226	0.25 ±0.03	0.16	2.96
BQ11830	TolA protein	381	BruAb1_1683	BARBAKC583_0145	BARSCv2_10253	BARCLv2_1236	0.25 ±0.03	0.00	5.10
BQ04010	Hemin binding protein, HbpD	270	BruAb1_0115[c]	BARBAKC583_1214[c]	BARSCv2_40297[c]	BARCLv2_0196[c]	0.25 ±0.05	17.49**	26.87**
BQ09930	conserved membrane protein of unknown function	361	no	no	BARSCv2_40486	BARCLv2_1127	0.25 ±0.06	0.00	0.00
BQ10560	VirB T4SS protein, VirB5	146	no	no	no	BARCLv2_0077[c]	0.25 ±0.07	1.08	2.57
BQ02680	conserved exported protein of unknown function	106	no	no	no	BARCLv2_0220	0.25 ±	1.52	3.58

[a] for ortholog assignment the "Phylo Profile Synteny" tool provided on the MaGe platform (*31*) was used (parameters: Bidirectional Blast Hit, 30 % identities, 0.6 for minLrap and maxLrap), automated ortholog assignments were manually curated, [b] LRT statistics for model 1a versus model 2 and model 7 versus model 8, values indicating significant positive selection are depicted by asterisks (*: p <0.05, **: p <0.001); [c] more than one orthologous gene is present in the corresponding genome, no clear ortholog assignment possible

Supplementary Table 4. Common genes of lineage 3 and lineage 4 with average *d*n/*d*s values (ω) ≥ 0.25 and results of the PAML analysis. Paralogous genes are grouped together.

Lineage 3:

			PAML results							**Lineage 4:**			PAML results					
			Model 1 vs Model 2[a]			Model 7 vs Model 8[b]							Model 1 vs Model 2[a]			Model 7 vs Model 8[b]		
Gene	**Length**	**ω**	**LRT**	**% sites > 1**	**dn/ds**	**LRT**	**% sites > 1**	**dn/ds**	**Protein product**	**Gene**	**Length**	**ω**	**LRT**	**% sites >1**	**dn/ds**	**LRT**	**% sites >1**	**dn/ds**
BARCLv2_0075	103	0.47	0.43			0.80			VirB T4SS proteins, VirB7[c]	BQ10580	100	0.34	1.13	4.7	3.3	2.5	4.7	3.5
BARCLv2_0645	103	0.47	0.43			0.80												
BARCLv2_0077	150	0.73	11.02*	25.0	3.4	12.06*	31.7	3.0	VirB T4SS proteins, VirB5[c]	BQ10560	146	0.25	1.08	1.8	7.3	2.6	2.6	6.4
BARCLv2_0647	150	0.73	11.02*	25.0	3.4	12.06*	31.7	3.0										
BARCLv2_0118	223	0.62	13.84**	12.2	4.8	14.52**	13.5	4.5	VirB T4SS effector proteins, Beps[c]	BQ10630	272	0.32	0			2.5		
BARCLv2_0069	547	0.42	2.46			3.40				BQ10650	532	0.28	0			5.3		
BARCLv2_0245	497	0.48	10.19*	9.5	3.0	18.90*	18.7	2.3		BQ10660	414	0.33	0			0.9		
BARCLv2_0631	538	0.42	15.09**	6.7	4.5	22.57**	9.8	3.6		BQ10680	555	0.58	9.67*	26.3	1.9	21.7**	45.7	1.6
BARCLv2_0632	528	0.50	6.51*	3.2	4.6	11.62*	7.7	3.1										
BARCLv2_0639	625	0.43	19.73**	3.1	8.8	23.36**	3.6	7.4										
BARCLv2_0640	539	0.45	18.16**	4.0	6.8	23.35**	6.2	5.1										
BARCLv2_0960	430	0.61	11.42*	7.1	4.5	14.62**	13.3	3.2										
BARCLv2_1031	547	0.49	16.29**	8.1	4.5	22.75**	11.8	3.5										
BARCLv2_0196	277	0.32	8.84*	7.4	4.4	8.47*	9.3	3.7	Hemin binding proteins[c]	BQ02420	267	0.25	19.27**	11.1	39.4	24.17**	12.4	25.2
BARCLv2_0197	241	0.29	35.92**	9.8	8.3	50.26**	11.7	6.9		BQ02430	382	0.27	67.57**	12.7	27.7	82.87**	14.4	16.9
BARCLv2_0198	291	0.32	9.37*	15.1	2.8	17.43*	18.1	2.7		BQ04010	270	0.25	17.49**	7.5	6.6	26.87**	11.8	4.3
BARCLv2_0473	673	0.38	108.28**	8.1	25.9	128.21**	12.3	12.1	Inducible Bartonella autotransporters[c]	BQ10380	759	0.25	0	5.2	1.0	0.0		
BARCLv2_0475	823	0.42	35.06**	5.0	15.5	54.30**	7.2	6.7		BQ10410	792	0.32	6.74*	5.1	204.1	20.36**	8.5	12.0
BARCLv2_1020	837	0.32	0.00			14.52	1.1	298.6	putative autotransporters[c]	BQ10280	799	0.25	0	9.6	1.0	14.65*	1.5	43.5
BARCLv2_0145	608	0.25	3.24			7.43	3.3	3.7		BQ10290	1753	0.26	282.17**	21.9	41.5	345.48**	23.3	30.4
BARCLv2_0416	244	0.53	0.37			0.53			TonB protein	BQ04170	248	0.41	10.25*	4.5	7.0	15.14**	6.1	5.5
BARCLv2_1236	377	0.28	0.20			0.78			TolA protein	BQ11830	381	0.25	0			5.1		
BARCLv2_0171	160	0.28	2.03			2.25			Alanine acetyltransferase	BQ02150	160	0.29	0			0.0		
BARCLv2_1171	328	0.25	20.56**	4.1	8.3	23.36**	4.6	7.5	putative phage protein	BQ11060	260	0.26	0			0.0		
BARCLv2_0847	293	0.30	0.00			0.00			Lysophospholipase l2	BQ05220	300	0.28	0			1.5		
BARCLv2_0125	50	0.50	0.99			1.06			Heme exporter protein CcmD	BQ01050	83	0.50	0			0.0		
BARCLv2_0115	157	0.35	2.27			3.00			FxsA cytoplasmic membrane protein	BQ00960	163	0.26	1.63			4.9		
BARCLv2_1107	391	0.27	0.62			2.17			Filament-A percursor	BQ01520	424	0.28	0			0.7		
BARCLv2_1274	194	0.30	2.17			3.83			ATP synthase delta subunit	BQ12260	194	0.30	0.13			1.2		
BARCLv2_1270	138	0.38	1.00			1.30			ATP synthase epsilon chain	BQ12220	138	0.26	0			0.7		
BARCLv2_1329	143	0.37	-9.41			6.16			Antioxidant, AhpC/TSA family	BQ13100	226	0.30	0.13			0.7		
BARCLv2_1038	144	0.34	2.25			2.88			twin arginine translocase TatB	BQ04750	134	0.84	4.26			5.2		
BARCLv2_0226	473	0.27	0.00			0.51			conserved protein of unknown function	BQ02810	447	0.25	0.16			3.0		
BARCLv2_0279	218	0.41	0.00			0.00			conserved protein of unknown function	BQ09660	228	0.49	1.72			3.5		
BARCLv2_0327	84	0.30	0.20			0.27			conserved protein of unknown function	BQ03350	85	0.27	3.11			4.2		
BARCLv2_0670	813	0.32	4.85			6.04*	0.7	12.8	conserved protein of unknown function	BQ07710	819	0.35	11.5*	1.1	11.5	16.37**	1.5	9.5
BARCLv2_0949	181	0.27	0.00			0.00			conserved protein of unknown function	BQ08360	184	0.28	5.59			7.19*		
BARCLv2_0978	40	0.41	0.92			1.20			conserved protein of unknown function	BQ08690	60	0.32	5.19			5.2		
BARCLv2_1013	158	0.43	2.71			3.46			conserved protein of unknown function	BQ09010	176	0.35	2.67			3.5		
BARCLv2_1059	141	0.42	5.04			5.44			conserved protein of unknown function	BQ03120	135	0.57	2.13			3.6		
BARCLv2_1163	97	0.65	2.11			2.27			conserved protein of unknown function	BQ11040	114	0.31	0			0.0		
BARCLv2_1164	139	0.41	3.14			3.26			conserved protein of unknown function	BQ11050	148	0.35	0.17			0.6		
BARCLv2_1199	63	0.33	0.00			0.00			conserved protein of unknown function	BQ11430	62	0.46	2.47			3.9		

Lineage 3:

			PAML results							**Lineage 4:**			PAML results					
			Model 1 vs Model 2[a]			Model 7 vs Model 8[b]							Model 1 vs Model 2[a]			Model 7 vs Model 8[b]		
Gene	**Length**	**ω**	**LRT**	**% sites > 1**	**dn/ds**	**LRT**	**% sites > 1**	**dn/ds**	**Protein product**	**Gene**	**Length**	**ω**	**LRT**	**% sites >1**	**dn/ds**	**LRT**	**% sites >1**	**dn/ds**
BARCLv2_1200	126	0.33	1.08			1.44			conserved protein of unknown function	BQ11450	129	0.26	0			2.3		
BARCLv2_1261	95	0.30	0.00			0.00			conserved protein of unknown function	BQ12190	93	0.31	0			0.6		
BARCLv2_1321	85	0.82	1.36			1.77			conserved protein of unknown function	BQ12930	110	0.40	2.55			3.3		
BARCLv2_1327	387	0.42	5.28			5.85			conserved protein of unknown function	BQ13000	389	0.31	1.57			4.8		
BARCLv2_0768	77	16.69	0.00			0.00			conserved exported protein	BQ06010	103	0.35	1.27			0.0		

[a] Likelihood ratio test (LRT) statistics for model 1 versus model 2a, percentage of sites under positive selection and estimated *d*n/*d*s value are shown, if model 2a (positive selection) fitted data significantly better than model 1 (* <0.05, **<0.001); [b] same analysis for of model 7 versus model 8; [c] no clear ortholog assignment possible between the genes of lineage 3 and lineage 4

Supplementary Table 5. Prediction of tyrosine-phosphorylation sites for Beps of Bep-clade 9 using NetPhos2.0 (*38*), ScanSite (*41*), and KinasePhos (*42*).

			NetPhos2.0 prediction[a]		ScanSite prediction[b]			KinasePhos prediction[c]	
Locus_tag	**Species**	**Pos.**	**Motif**	**Score**	**High**	**Medium**	**predicted kinase**	**100% specificity**	**predicted kinase**
B11Cv2_60018	*B.*sp. 1-1C	255	SSGIYTNYN	0.966	no	yes	Grb2 SH2, Lck kinase	yes	INSR kinase
B11Cv2_60018	*B.*sp. 1-1C	278	SSGIYTNYN	0.966	no	yes	Grb2 SH2, Lck kinase	yes	INSR kinase
B11Cv2_60018	*B.*sp. 1-1C	297	SEEEYSAIY	0.989	yes	yes	Shc SH2, **INSR kinase**	yes	Src kinase
B11Cv2_60018	*B.*sp. 1-1C	317	NEEEYSGIY	0.979	yes	yes	**Fgr**, Lck, INSR kinases, **Shc SH2**	yes	Src, INSR kinases
B11Cv2_60018	*B.*sp. 1-1C	321	YSGIYESYD	0.899	yes	yes	**Src**, Lck kinases, Lck SH2	yes	Syk, INSR kinases
B11Cv2_60018	*B.*sp. 1-1C	340	NEEEYSGIY	0.979	yes	yes	**Fgr**, Lck, INSR kinases, **Shc SH2**	yes	Syk, INSR kinases
B11Cv2_60018	*B.*sp. 1-1C	344	YSGIYANYD	0.923	yes	yes	Fgr, Lck, Src, **Grb2 SH**	yes	INSR kinase
B11Cv2_60018	*B.*sp. 1-1C	363	NEEEYSGIY	0.979	yes	yes	**Fgr**, Lck, INSR kinases, **Shc SH2**	yes	Src, INSR kinases
B11Cv2_60018	*B.*sp. 1-1C	367	YSGIYANCD	0.898	no	yes	Itk SH2	yes	INSR kinase
B11Cv2_60018	*B.*sp. 1-1C	396	NDDIYDNKD	0.987	no	yes	Src kinase	yes	Syk, INSR kinases
B11Cv2_60018	*B.*sp. 1-1C	402	NKDIYDSAN	0.983	no	no	-	yes	EGFR, INSR kinases
B11Cv2_60018	*B.*sp. 1-1C	414	NDDIYDNKD	0.987	no	yes	Src kinase	yes	Syk, INSR kinases
B11Cv2_60018	*B.*sp. 1-1C	420	NKDIYDSAN	0.983	no	no	-	yes	EGFR, INSR kinases
B11Cv2_60018	*B.*sp. 1-1C	432	NDDIYDNKD	0.987	no	yes	Src kinase	yes	Syk, INSR kinases
B11Cv2_60018	*B.*sp. 1-1C	438	NKDIYDNPN	0.99	yes	yes	**Abl**, **Nck**, Crk, Itk SH2	yes	Syk, EGFR, INSR kinases
B11Cv2_60018	*B.*sp. 1-1C	660	KEDDYQTLA	0.991	no	no	-	yes	Syk, EGFR, INSR kinases
BARROv2_70017	*B. rochalimae*	4	-MPNYVLVP	0.969	no	yes	PDFGR kinase, PLCg SH2	no	-
BARROv2_70017	*B. rochalimae*	253	QEGIYANYN	0.979	yes	yes	Fgr, **Lck**, Abl, Src kinases	yes	INSR kinase
BARROv2_70017	*B. rochalimae*	277	NEDIYDTTD	0.986	yes	yes	Lck, **Src**, Abl kinases	yes	Syk, Jak, EGFR, INSR kinases
BARROv2_70017	*B. rochalimae*	283	TTDIYDNPD	0.964	yes	yes	**Abl**, Crk, **Nck**, Itk SH2	yes	Syk, INSR kinases
BARROv2_70017	*B. rochalimae*	301	EEDIYANYN	0.993	yes	yes	**Fgr**, **Lck**, **Src**, Abl kinases, **Grb2, Itk SH2**	yes	Src, Syk, INSR kinasess
BARROv2_70017	*B. rochalimae*	320	NEGEYSDTY	0.986	no	no	-	yes	Srk, EGFR, INSR kinases
BARROv2_70017	*B. rochalimae*	324	YSDTYNTTG	0.938	no	no	-	yes	Syk kinases
BARROv2_70017	*B. rochalimae*	330	TTGIYENPD	0.927	yes	yes	**Abl**, Lck, Itk SH2	yes	INSR kinase

			NetPhos2.0 prediction[a]		ScanSite prediction[b]			KinasePhos prediction[c]	
Locus_tag	**Species**	**Pos.**	**Motif**	**Score**	**High**	**Medium**	**predicted kinase**	**100% specificity**	**predicted kinase**
BARROv2_70017	*B. rochalimae*	350	FSDIYDTTD	0.954	no	yes	Nck SH2	no	-
BARROv2_70017	*B. rochalimae*	356	TTDIYDNPD	0.964	yes	yes	**Abl**, Crk, **Nck**, Itk SH2	yes	Syk, INSR kinases
BARROv2_70017	*B. rochalimae*	374	EEDIYANYN	0.993	yes	yes	**Fgr, Lck, Src**, Abl kinases, **Itk, Grb2 SH2**	yes	Src, Syk, INSR kinasess
BARROv2_70017	*B. rochalimae*	393	NKGEYSDTY	0.973	no	no	-	yes	INSR kinase
BARROv2_70017	*B. rochalimae*	397	YSDTYNTTG	0.938	no	no	-	yes	Syk kinase
BARROv2_70017	*B. rochalimae*	403	TTGIYENPD	0.927	yes	yes	**Abl**, Lck, Itk SH2	yes	INSR kinase
BARROv2_70017	*B. rochalimae*	421	EEDIYANYN	0.993	yes	yes	**Fgr, Lck, Src**, Abl kinases, **Grb2, Itk SH2**	yes	Src, Syk, INSR kinasess
BARROv2_70017	*B. rochalimae*	440	NKGEYSDTY	0.973	no	no	-	yes	INSR kinase
BARROv2_70017	*B. rochalimae*	444	YSDTYNTTG	0.938	no	no	-	yes	Syk kinase
BARROv2_70017	*B. rochalimae*	450	TTGIYENPD	0.927	yes	yes	**Abl**, Lck, Itk SH2	yes	INSR kinase
BARROv2_70017	*B. rochalimae*	468	EENIYENYN	0.99	yes	yes	**Src, Lck**, Fgr kinases, **Fgr**, Lck, **Grb2, Itk**, Shc SH2	yes	Src, Syk, INSR kinasess
BARROv2_70017	*B. rochalimae*	675	KENDYQTLA	0.98	no	no	-	yes	Syk
BAR15v2_80031	*B.* sp. AR 15-3	255	DQNIYESYD	0.866	yes	yes	**Src kinase**	yes	Syk, INSR kinases
BAR15v2_80031	*B.* sp. AR 15-3	285	IDSIYDNPS	0.84	yes	yes	**Abl**, Crk, **Nck SH2**	yes	Syk, INSR kinases
BAR15v2_80031	*B.* sp. AR 15-3	296	VDPVYANYT	0.929	yes	yes	Abl, Lck kinases, **Grb2**, Itk SH2	no	-
BAR15v2_80031	*B.* sp. AR 15-3	323	VDPVYDDPS	0.928	yes	yes	Abl kinase, **Nck**, Abl, Crk SH2	yes	Syk kinases
BAR15v2_80031	*B.* sp. AR 15-3	340	IDSIYDNPS	0.84	yes	yes	**Abl**, Crk, **Nck SH2**	yes	Syk, INSR kinases
BAR15v2_80031	*B.* sp. AR 15-3	351	VDPVYANYT	0.929	yes	yes	Abl, Lck kinases, **Grb2**, Itk SH2	no	-
BAR15v2_80031	*B.* sp. AR 15-3	378	IDPVYDDPS	0.908	yes	yes	Abl kinase, **Nck**, Abl, Crk SH2	yes	Syk kinases
BAR15v2_80031	*B.* sp. AR 15-3	395	IDSIYDNPS	0.84	yes	yes	**Abl**, Crk, **Nck SH2**	yes	Syk, INSR kinases
BAR15v2_80031	*B.* sp. AR 15-3	406	VDPVYANYT	0.929	yes	yes	Abl, Lck kinases, **Grb2**, Itk SH2	no	-
BAR15v2_80031	*B.* sp. AR 15-3	433	VDPVYDDPS	0.928	yes	yes	Abl kinase, **Nck**, Abl, Crk SH2	yes	Syk kinase
BAR15v2_80031	*B.* sp. AR 15-3	450	IDSIYDNPS	0.84	yes	yes	**Abl**, Crk, **Nck SH2**	yes	Syk, INSR kinases
BAR15v2_80031	*B.* sp. AR 15-3	461	VDPVYANYT	0.929	yes	yes	Abl, Lck kinases, **Grb2**, Itk SH2	no	-
BAR15v2_80031	*B.* sp. AR 15-3	488	IDPVYDDPS	0.908	yes	yes	Abl kinase, **Nck**, Abl, Crk SH2	yes	Syk kinases
BAR15v2_80031	*B.* sp. AR 15-3	608	KNPRYKQAK	0.948	no	no	-	no	-
BARCLv2_1034	*B. clarridgeiae*	14	NSEIYENSE	0.981	no	yes	Src kinase, Fgr SH2	yes	Syk, EGFR, INSR kinases
BARCLv2_1034	*B. clarridgeiae*	20	NSEIYDNPA	0.983	yes	yes	**Abl, Nck SH2**	yes	Syk, INSR kinasess

			NetPhos2.0 prediction[a]		ScanSite prediction[b]			KinasePhos prediction[c]	
Locus_tag	**Species**	**Pos.**	**Motif**	**Score**	**High**	**Medium**	**predicted kinase**	**100% specificity**	**predicted kinase**
BARCLv2_1034	*B. clarridgeiae*	30	DSGIYDTPA	0.891	yes	yes	**Abl kinase, Nck, Crk**, Abl SH2	yes	Syk, INSR kinases
BARCLv2_1034	*B. clarridgeiae*	40	NSEIYDNPA	0.983	yes	yes	Abl kinase, **Abl, Nck**, Crk, Itk SH2	yes	Syk, INSR kinases
BARCLv2_1034	*B. clarridgeiae*	50	DSGIYDTPA	0.891	yes	yes	**Abl kinase, Nck, Crk**, Abl SH2	yes	Syk, INSR kinases
BARCLv2_1034	*B. clarridgeiae*	60	NSEIYGNPA	0.978	no	yes	Abl, Srk kinases, Abl, Itk SH2	yes	INSR kinase
BARCLv2_1034	*B. clarridgeiae*	70	DSGIYDTPA	0.891	yes	yes	**Abl kinase, Nck, Crk,** Abl SH2	yes	Syk, INSR kinase
BARCLv2_1034	*B. clarridgeiae*	80	NSEIYENSE	0.981	no	yes	Src kinase	yes	Syk, EGFR, INSR kinases
BARCLv2_1034	*B. clarridgeiae*	86	NSEIYGNPA	0.978	no	yes	Abl kinase, Abl, Nck, Crk, Itk SH2	yes	INSR kinase
BARCLv2_1034	*B. clarridgeiae*	96	DSGIYDTPA	0.891	yes	yes	**Abl kinase, Nck, Crk,** Abl SH2	yes	Syk, INSR kinases
BARCLv2_1034	*B. clarridgeiae*	106	NSEIYENSE	0.981	no	yes	Src kinase	yes	Syk, EGFR, INSR kinases
BARCLv2_1034	*B. clarridgeiae*	112	NSEIYGNSA	0.973	no	yes	Src kinase	yes	INSR kinase
BARCLv2_1034	*B. clarridgeiae*	122	DSEIYENYD	0.983	yes	yes	**Src, Lck, Fgr kinases, Grb2**, Fgr SH2	yes	Syk, INSR kinases
BARCLv2_1034	*B. clarridgeiae*	128	NYDTYKKNK	0.844	no	no	-	no	-
BARCLv2_1034	*B. clarridgeiae*	318	RNNEYKLLA	0.84	no	no	-	no	-
BARCLv2_1032	*B. clarridgeiae*	17	QEVEYTEH	0.949	no	yes	Fyn, Lck, Src SH2	no	-
BARCLv2_1032	*B. clarridgeiae*	42	QEVEYAEIR	0.962	no	yes	Lck kinase	no	-
BARCLv2_1035	*B. clarridgeiae*	16	EEVEYAEVF	0.971	no	yes	Src, Lck, INSR, PDFGR kinase, PLCg SH2	yes	Syk, INSR kinases
BARCLv2_1035	*B. clarridgeiae*	197	SKPVYMLSR	0.89	no	yes	PCLg SH2	no	-

[a] For NetPhos2.0 predictions, a threshold of 0.8 was used.

[b] Scansite was used with the high stringency filter and the medium stringency filter. Kinases depicted in bold type were predicted with high stringency filter.

[a] For KinasePhos predictions, the 100% specificity filter was used.

Supplementary Table S6. 454-sequencing and newbler-assemby data for genomes sequenced in this study.

	B. clarridgeiae	***B. rochalimae***	***B.* sp. 1-1C**	***B.* sp. AR15-3**	***B. schoenbuchensis***
Final length (bp)	1,522,743	1,552,290	1,584,739	1,607,248	1,684,713
Number of 454-reads	212,905	287,512	253,303	254,535	173,358
Average read length (bp)	247	211	241	236	285
Number of Newbler contigs (> one reads)	43	78	74	93	220
454-sequence coverage	35x	39x	39x	37x	29x

References

1. Z. Yang, R. Nielsen, *Mol Biol Evol* **17**, 32 (Jan, 2000).
2. Z. Yang, R. Nielsen, *Mol Biol Evol* **19**, 908 (Jun, 2002).
3. Z. Yang, *Mol Biol Evol* **24**, 1586 (Aug, 2007).
4. M. Anisimova, J. P. Bielawski, Z. Yang, *Mol Biol Evol* **18**, 1585 (Aug, 2001).
5. H. L. Saenz *et al.*, *Nat Genet* **39**, 1469 (Dec, 2007).
6. M. C. Schmid *et al.*, *Mol Microbiol* **52**, 81 (Apr, 2004).
7. R. Schulein, C. Dehio, *Mol Microbiol* **46**, 1053 (Nov, 2002).
8. R. Schulein *et al.*, *Proc Natl Acad Sci U S A* **102**, 856 (Jan 18, 2005).
9. J. A. Carroll, S. A. Coleman, L. S. Smitherman, M. F. Minnick, *Infect Immun* **68**, 6750 (Dec, 2000).
10. M. F. Minnick *et al.*, *Infect Immun* **71**, 814 (Feb, 2003).
11. R. Fronzes *et al.*, *Science* **323**, 266 (Jan 9, 2009).
12. T. A. Rhomberg, M. C. Truttmann, P. Guye, Y. Ellner, C. Dehio, *Cell Microbiol* **11**, 927 (Jun, 2009).
13. F. Scheidegger *et al.*, *Cell Microbiol* **11**, 1088 (Jul, 2009).
14. M. C. Schmid *et al.*, *PLoS Pathog* **2**, e115 (Nov, 2006).
15. C. Dehio, *Cell Microbiol* **10**, 1591 (Aug, 2008).
16. R. Fronzes, P. J. Christie, G. Waksman, *Nat Rev Microbiol* **7**, 703 (Oct, 2009).
17. A. Seubert, R. Hiestand, F. de la Cruz, C. Dehio, *Mol Microbiol* **49**, 1253 (Sep, 2003).
18. B. Nystedt, A. C. Frank, M. Thollesson, S. G. Andersson, *Mol Biol Evol* **25**, 287 (Feb, 2008).
19. K. Orth, *Curr Opin Microbiol* **5**, 38 (Feb, 2002).

20. A. C. Frank, C. M. Alsmark, M. Thollesson, S. G. Andersson, *Mol Biol Evol* **22**, 1325 (May, 2005).

21. L. Chen, Y. Chen, D. W. Wood, E. W. Nester, *J Bacteriol* **184**, 4838 (Sep, 2002).

22. T. M. Finan *et al.*, *Proc Natl Acad Sci U S A* **98**, 9889 (Aug 14, 2001).

23. E. C. Berglund *et al.*, *PLoS Genet* **5**, e1000546 (Jul, 2009).

24. R. Heller *et al.*, *J Clin Microbiol* **35**, 1327 (Jun, 1997).

25. J. W. Lin, C. Y. Chen, W. C. Chen, B. B. Chomel, C. C. Chang, *J Med Microbiol* **57**, 1496 (Dec, 2008).

26. M. E. Eremeeva *et al.*, *N Engl J Med* **356**, 2381 (Jun 7, 2007).

27. K. Inoue *et al.*, *Emerg Infect Dis* **15**, 526 (Apr, 2009).

28. C. Dehio *et al.*, *Int J Syst Evol Microbiol* **51**, 1557 (Jul, 2001).

29. M. Margulies *et al.*, *Nature* **437**, 376 (Sep 15, 2005).

30. S. Bocs, S. Cruveiller, D. Vallenet, G. Nuel, C. Medigue, *Nucleic Acids Res* **31**, 3723 (Jul 1, 2003).

31. D. Vallenet *et al.*, *Nucleic Acids Res* **34**, 53 (2006).

32. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res* **22**, 4673 (Nov 11, 1994).

33. J. C. Wilgenbusch, D. Swofford, *Curr Protoc Bioinformatics* **Chapter 6**, Unit 6 4 (Feb, 2003).

34. J. P. Huelsenbeck, F. Ronquist, *Bioinformatics* **17**, 754 (Aug, 2001).

35. D. Posada, *Curr Protoc Bioinformatics* **Chapter 6**, Unit 6 5 (Feb, 2003).

36. J. A. A. Nylander. (Program distributed by the author, 2004).

37. S. L. Pond, S. D. Frost, S. V. Muse, *Bioinformatics* **21**, 676 (Mar 1, 2005).

38. B. B. Chomel *et al.*, *Vet Res* **40**, 27 (Mar 10, 2009).

39. N. Blom, S. Gammeltoft, S. Brunak, *J Mol Biol* **294**, 1351 (Dec 17, 1999).

40. P. Engel, C. Dehio, *Genome Dyn* **6**, 158 (2009).

41. J. C. Obenauer, L. C. Cantley, M. B. Yaffe, *Nucleic Acids Res* **31**, 3635 (Jul 1, 2003).

42. H. D. Huang, T. Y. Lee, S. W. Tzeng, J. T. Horng, *Nucleic Acids Res* **33**, W226 (Jul 1, 2005).

3.3 Research article III

The Trw type IV secretion system of *Bartonella* mediates host-specific infection of erythrocytes

Muriel Vayssier-Taussat, Danielle Le Rhun, Hongkuan Deng, Francis Biville, Sandra Cescau, Antoine Danchin, Geneviève Marignac, Evelyne Lenaour, Henri-Jean Boulouis, Maria Mavris, Le Menach Arnaud, Huan-Ming Yang, Jing Wang, Philipp Engel, Henri Saenz, and Christoph Dehio

Manuscript submitted.

Format: Research article for PLoS Pathogens

3.3.1 Summary

As *Bartonella* species can only cause intraerythrocytic infection in one or a few related mammalian reservoir hosts, the genus *Bartonella* displays a great model to investigate the molecular basis of host specificity in bacterial pathogens.

In this study, our aim was to identify bacterial factors which may confer host-specificity for the colonization of erythrocytes. An *in vitro* erythrocyte adhesion and invasion assay was established and different *Bartonella* species were tested for their ability to colonize erythrocytes of different hosts. Quantification by gentamicin protection assays and flow cytometry showed that most *Bartonella* species can only adhere to and invade erythrocytes isolated from their respective reservoir hosts. These findings fully reflected the host specificity found *in vivo*.

Using the *Bartonella birtlesii*-mouse model, a signature-tagged mutagenesis was performed. This resulted in the identification of 45 abacteremic mutants of *B. birtlesii*. By testing them in the *in vitro* assay, nine of these abacteremic mutants were found to be impaired in erythrocyte colonization. In seven out of nine mutants the transposon insertion was found to be located in genes of the Trw type IV secretion system (T4SS). These results confirmed the previously suggested role of this system in erythrocyte colonization. To test whether the Trw T4SS is directly responsible for the *in vitro* and *in vivo* observed host-specific colonization of erythrocytes, the entire Trw T4SS of *B. tribocorum* (rat-specific) was expressed in *B. henselae* (cat-specific). Erythrocyte infection experiments showed that *B. henselae* harboring the Trw T4SS of *B. tribocorum* was able to colonize erythrocytes isolated from cats and from rats implicating a direct role of the Trw system in host specificity.

We further analyzed the molecular evolution of different *trw* genes to identify components of this T4SS which appear likely to mediate the observed host specificity. To this end, we applied phylogenetic and natural selection analyses, and compared the *trw* loci from the available genomes of *B. tribocorum*, *B. henselae*, *B. quintana*, and *B. birtlesii*. The latter genome was sequenced in relation to this project. In conjunction with previously published data (Nystedt et al. 2008), our computational analysis revealed that *trwL* and *trwJ* genes have been amplified and diversified in the different *Bartonella* species. Therefore, these two surface-exposed pilus subunits of the Trw T4SS represent primary candidates for mediating the observed host-specificity.

Statement of the own participation

I contributed to this publication by annotating the Trw T4SS sequence of *B. birtlesii* and performing molecular evolution analyses of different Trw T4SS genes (Figure 7). I designed Figure 7 and was involved in writing parts of the results chapter and producing Supporting Table 1.

References

Nystedt, B., A. C. Frank, et al. (2008). "Diversifying selection and concerted evolution of a type IV secretion system in Bartonella." Mol Biol Evol 25(2): 287-300.

3.3.2 Manuscript

The Trw type IV secretion system of *Bartonella* mediates host-specific infection of erythrocytes

M. Vayssier-Taussat[1,2,*], D. Le Rhun[1], H.K. Deng[1], F. Biville[3,4], S. Cescau[5], A. Danchin[5], G. Marignac[1], E. Lenaour[1], H.J. Boulouis[4], M. Mavris[1], L. Arnaud[6], H.M. Yang[7], J. Wang[8], M. Quebatte[2], P. Engel[2], H. Saenz[2], and C. Dehio[2,*]

[1]Unité de Biologie Moléculaire et Immunologie Parasitaire et Fongique, ENVA/AFSSA/INRA, 94700 Maisons-Alfort, France; [2]Focal Area Infection Biology, Biozentrum, University of Basel, 4056 Basel, Switzerland; [3]Département de Microbiologie, CNRS URA 2172, Institut Pasteur, 75724 Paris Cedex 15, France; [4]Unité des Bactéries Zoonotiques Hémotropes, ENVA/AFSSA, 94700 Maisons-Alfort, France; [5]Unité des Génomes Bactériens, Institut Pasteur, CNRS URA 2171, 75724 Paris Cedex 15, France; [6]Institut de Transfusion Sanguine, 75015 Paris, France, [7]Beijing Institute for Genomics, Chinese Academy of Science, Beijing 101300, China; [8]Behavioral Genetics Center, Institute of Psychology, Chinese Academy of Science, Beijing 100101, China

[*]Corresponding authors:

Prof. Christoph Dehio
Focal Area Infection Biology
Biozentrum, University of Basel
Klingelbergstrasse 70
CH-4056 Basel, Switzerland
Tel.: +41-61-267 2140
Fax.: +41-61-267 2118
e-mail: christoph.dehio@unibas.ch

Dr. Muriel Vayssier-Taussat (Ph.D.)
UMR 956 BIPAR
23 avenue du General de Gaulle
AFSSA
F-94700 Maisons-Alfort, France
Tel.: +33-1-4977 4655
Fax: +33-1-4396 7332
e-mail: mvayssier@vet-alfort.fr

Running title: *Bartonella* host specificity

ABSTRACT

Bacterial pathogens typically infect only a limited range of hosts; however, the genetic mechanisms governing this host-specificity are poorly understood. The α-proteobacterial genus *Bartonella* comprises 21 species that cause host-specific intraerythrocytic bacteremia as hallmark of infection in their respective mammalian reservoirs, including the human-specific pathogens *Bartonella quintana* and *Bartonella bacilliformis* that cause trench fever and Oroya fever, respectively. Here, we have identified bacterial factors that mediate host-specific erythrocyte colonization in the mammalian reservoirs. Using mouse-specific *Bartonella birtlesii*, human-specific *Bartonella quintana*, cat-specific *Bartonella henselae* and rat-specific *Bartonella tribocorum*, we established *in vitro* adhesion and invasion assays with isolated erythrocytes that fully reproduce the host-specificity of erythrocyte infection as observed *in vivo*. By signature-tagged mutagenesis of *B. birtlesii* and mutant selection in a mouse infection model we identified mutants impaired in establishing intraerythrocytic bacteremia. Among 45 abacteremic mutants, five failed completely to infect mouse erythrocytes *in vitro*. The corresponding genes encode components of the type IV secretion system (T4SS) Trw, demonstrating that this virulence factor laterally acquired by the *Bartonella* lineage is directly involved in erythrocyte infection. Strikingly, ectopic expression of Trw of rat-specific *B. tribocorum* in cat-specific *B. henselae* expanded the host range for erythrocyte infection to rat, indicating that Trw mediates host-specific erythrocyte infection. A molecular evolutionary analysis of the *trw* locus further indicated that the variable, surface-located TrwL and TrwJ represent the T4SS components that determine host-specificity of erythrocyte parasitism. In conclusion, we show that the laterally

acquired Trw T4SS diversified in the *Bartonella* lineage to facilitate the infection of erythrocytes in a host-restricted manner.

AUTHOR SUMMARY

Pathogens are - as the result of adaptive evolution in their principle host(s) - typically limited in the range of hosts that they can infect successfully. However, infrequently such host-restricted pathogens may undergo a spontaneous host switch, which can lead to the evolution of pathogens with altered host specificity. Most human pathogens evolved this way, and animal-specific pathogens have thus to be considered as an important reservoir for the emergence of novel human pathogens. Despite of host-specificity representing a common feature of pathogens the underlying molecular mechanisms are largely unknown. In this study we have used bacterial pathogens of the genus *Bartonella* to identify bacterial factors involved in the determination of host specificity. The bartonellae represent an excellent model to study host specificity as each species is adapted to cause an intracellular infection of erythrocytes exclusively in its respective reservoir host(s). Using a genetic approach in combination with erythrocyte infection models *in vitro* and *in vivo* we demonstrate that a surface-located bacterial nanomachine – a socalled type IV secretion system - determines host specificity of erythrocyte infection. Our work sheds light on the molecular basis of host specificity and established an experimental model for studying the evolutionary processes facilitating sudden host shifts.

INTRODUCTION

The successful infection of a mammalian host by a bacterial pathogen typically involves a series of intimate host-pathogen interactions. On the molecular level this is reflected by specific receptor-ligand interactions between bacterial virulence factors and their targeted host factors (Finlay and Cossart 1997). Adaptation of a bacterial virulence factor to a host factor that displays variability within the host population can restrict the host range that is susceptible to infection. The resulting host-specificity is an inherent feature of most bacterial pathogens of humans, including *Helicobacter pylori*, *Listeria monocytogenes, Neisseria gonorrhoae*, *Salmonella typhi*, *Streptococcus pyogenes* and *Staphylococcus aureus*. However, remarkably little is known about the molecular determinants of host specificity in bacterial infections, with the only exception of *L. monocytogenes* for which the conjugated action of two distinct host-specific invasion proteins was shown to be critical for fetoplacental listeriosis (Lecuit et al. 1999; Khelef et al. 2006; Disson et al. 2008).

Bartonellae represent an interesting but largely unexplored model for host specificity. These facultative intracellular bacteria use arthropod transmission and hemotropism as mammalian parasitism strategy (Dehio 2004). As the result of an adaptive radiation each of the 21 species infects only one or a few closely related mammalian reservoir host(s), which is highlighted by their capacity to cause a long-lasting intraerythrocytic bacteremia (Saenz et al. 2007). Non-reservoir hosts may get incidentally infected without resulting in an intraerythrocytic infection (Dehio 2005). Two *Bartonella* species are human-specific: *Bartonella bacilliformis* causes the biphasic Carrion's disease, with acute Oroya fever followed by the chronic verruga peruana, and *Bartonella quintana* causing trench fever. The life-threatening Oroya

fever and the much milder course of trench fever represent the characteristic intraerythrocytic stages of these pathogens. The other 19 species cause intraerythrocytic infections in various non-primate mammalian reservoirs. At least seven of them are recognized as zoonotic pathogens which incidentally infect humans. Most prominently, *B. henselae* is associated with cat scratch disease (Dehio 2005).

The life cycle of *Bartonella* in the reservoir host has been analyzed in detail in rats experimentally infected with *B. tribocorum* (Schulein et al. 2001). Following intravenous inoculation, bacteria initially infect a primary niche outside of circulating blood, which is considered to comprise the vascular endothelium and possibly other cell types. Approximately on day five of infection, large numbers of bacteria are released into the bloodstream where they invade mature erythrocytes. Bacteria then replicate in a membrane-bound compartment until reaching a critical number. For the remaining life span of the erythrocytes the intracellular bacteria remain in a non-dividing state. Monitoring of bacteremia in other animal models, such as the *B. birtlesii*/mouse (Boulouis et al. 2001) and *B. henselae*/cat models (Yamamoto et al. 2002), or in captive naturally infected animals has yielded results that match those observed in the *B. tribocorum*/rat model, suggesting a common mode of infection of the different species in their respective animal reservoirs (Chomel et al. 2009a). The only exception is *B. bacilliformis*, which causes lysis of the infected human erythrocytes, eventually resulting in a severe hemolytic anemia.

The *B. tribocorum*/rat model was further explored to identify bacterial pathogenicity factors that are required for colonization of the mammalian reservoir host. A signature-tagged mutagenesis (STM) screen identified 98 essential bacterial loci (Saenz and Dehio 2005), including genes encoding components of two distinct

type IV secretion systems (T4SS), VirB/VirD4 and Trw, the invasion-locus B (IalB) protein, the trimeric autotransporter adhesin BadA, as well as other members of the autotransporter family (Saenz et al. 2007). Whether any of the identified genes is critical for host-specificity is unknown, although it is conceivable to assume that host-specificity loci are essential for infection and may thus be represented among the hits of the performed STM screen.

Experimental infections of different mammalian hosts by a given *Bartonella* strain have reproduced the species-specificity of erythrocyte invasion as observed in natural infections (Kosoy et al. 2000; Kabeya et al. 2003; Chomel et al. 2009a; Chomel et al. 2009b). However, despite their availability, *in vitro* erythrocyte infection assays (Scherer et al. 1993; Mehock et al. 1998) have not been investigated for the study of host specificity. Here, we demonstrate for the first time that host specificity is reflected by the exclusive capacity of *Bartonella* species to infect erythrocytes isolated from their natural host(s). Second, by performing STM in *Bartonella birtlesii* followed by screening in mice *in vivo* and in isolated erythrocytes *in vitro* we identified the T4SS Trw as the molecular determinant of host-specific erythrocyte infection.

RESULTS

An *in vitro* erythrocyte colonization assay to study host-restricted infection

Based on described *in vitro* models of human and feline erythrocyte infection by *B. bacilliformis* and *B. henselae*, respectively (Scherer et al. 1993; Mehock et al. 1998), we established for *B. birtlesii* an *in vitro* infection model for erythrocytes isolated form the murine reservoir host. Balb/C mice were used as the source of erythrocytes as they are known to develop a long lasting intraerythrocytic infection upon experimental infection with *B. birtlesii* (Boulouis et al. 2001). The intraerythrocytic presence of bacteria was evaluated over a period of three days using the gentamicin protection assay (Fig. 1A). Bacterial entry into erythrocytes was dependent on the number of bacteria per erythrocyte (multiplicity of infection, MOI; tested MOI range: 0.01 to 10) and time of infection (days post infection, DPI; tested time range: 1 to 3 DPI). The highest intraerythrocytic bacterial content over time was obtained for MOI=0.1 and 1, with approximately $2x10^5$ colony forming units (CFU) per 10^{10} erythrocytes (≈0.002% infected erythrocytes) at 3 DPI. Given that mouse blood contains approximately 10^{10} erythrocytes/ml, this value corresponds well to the bacteremia reported for experimentally infected Balb/C mice (≈$1x10^5$ CFU/ml; 0.001% infected erythrocytes) (Boulouis et al. 2001). For MOI=10, erythrocytes were highly infected at 1 DPI, but lysed entirely until 3 DPI. At MOI=0.01, only low numbers of intraerythrocytic bacteria were detected over time. Based on these data, MOI=1 was used for all subsequent erythrocyte infection assays. To evaluate whether the increase of intraerythrocytic bacteria over time was mainly due to continued bacterial invasion, or to intraerythrocytic bacterial multiplication, or to a combination of both, erythrocytes were infected with *B. birtlesii* for one day in the absence of gentamicin, followed by

incubation in the continuous presence of gentamicin to kill extracellular bacteria. Fig. 1B shows that the number of intracellular bacteria increased over time in the presence of gentamicin, albeit to a lesser extent than in the untreated control. Bacteria thus appear to enter erythrocytes beyond 1 DPI and, moreover, to replicate in an intra-erythrocytic location.

Invasion of erythrocytes by *Bartonella* is preceded by bacterial adhesion to the erythrocyte surface (Schulein et al. 2005). To quantify erythrocytes infected by adherent extracellular and/or intracellular bacteria, we used GFP-expressing bacteria in combination with flow cytometry (Fig. 1.C, D). Similar as described for intraerythrocytic bacteria in the gentamicin protection assay, erythrocyte colonization revealed by flow cytometry was dependent on time (Fig. 1.C) and MOI (Fig. 1.D). However, the rate of erythrocyte colonization evaluated by flow cytometry (55% for MOI=1 at 3 DPI) was approximately 20’000-fold higher than erythrocytes invasion determined by the gentamicin protection assay (compare Fig. 1.A), indicating that the vast majority of bacteria detected by flow cytometry were associated extracellularly with erythrocytes. Confocal microscopy confirmed the predominant extracellular localization of erythrocyte-associated bacteria (Fig. 1.E).

Next we investigated whether *Bartonella* species differ in their capacity to interact *in vitro* with erythrocytes of different mammalian origin, and whether this capacity may reflect the host-restriction displayed during natural infection. First, mouse erythrocytes were infected with either *B. birtlesii*, *B. vinsonii arupensis* (both mouse-specific), *B. alsatica* (rabbit-specific), *B. vinsonii berkhoffii* (dog-specific), *B. henselae* (cat-specific), *B. quintana* (human-specific), *B. chomelii* (cattle-specific), or *B. tribocorum* (rat-specific). Erythrocyte invasion was quantified by the gentamicin protection assay (Fig. 2). *B. vinsonii arupensis* displayed invasion rates similar to *B.*

birtlesii, while none of the other strains tested resulted in significant erythrocyte invasion. These findings indicate that specificity for a mouse reservoir *in vivo* correlates with efficient interaction with mouse erythrocytes *in vitro*.

Next, we tested whether - similarly as observed for mouse erythrocytes and *B. birtlesii* – the capacity of *B. henselae, B. quintana* and *B. tribocorum* to colonize erythrocytes *in vitro* is also restricted to erythrocytes from their natural reservoir host, i.e. cat, human and rat, respectively. GFP-expressing bacteria were used for erythrocyte infection, and colonization was quantified by flow cytometric analysis. Fig. 3 and Table 1 illustrate that all tested *Bartonella* species were able to efficiently interact with erythrocytes isolated from their respective reservoir hosts, while they essentially did not interact with erythrocytes from non-reservoir hosts. The only exception is *B. quintana*, which further to erythrocytes from the human reservoir also colonized cat erythrocytes. However, isolation of *B. quintana* from cat blood has been described, suggesting that this organism has indeed a limited capacity to infect cat erythrocytes (Table 1) (Breitschwerdt et al. 2007). Together, these data indicate that the established *in vitro* model of erythrocyte colonization reflects host restriction as observed during natural infection.

B. birtlesii genes required for intra-erythrocytic bacteremia in mice

As a basis for identifying genetic factors involved in host-restricted erythrocyte colonization, we identified a comprehensive set of *B. birtlesii* genes required for establishing intraerythrocytic bacteremia in mice. To this end, an STM library of *B. birtlesii* was constructed as previously described for *B. tribocorum* (Mavris et al. 2005; Saenz and Dehio 2005; Saenz et al. 2007). From each conjugation assay, we selected 96 single kanamycin-resistant colonies and assembled an STM mutant

library of 3456 mutants. We then identified mutants that have lost the capacity to cause intraerythrocytic bacteremia by screening the library in the mouse infection model (Boulouis et al. 2001). Of 1456 mutants tested in the input pools, 98 were not detected in the output pools from mice at days 7 and 14 post infection and were thus classified as abacteremic mutant candidates. All 98 abacteremic mutant candidates were retested by reassembling them into 49 pools of 9 mutants, each pool containing two abacteremic mutant candidates and an invariable set of seven mutants displaying wild-type behavior (bacteremic mutants). The rescreen confirmed an abacteremic phenotype for 48 of the initial 98 abacteremic mutant candidates, corresponding to 3.3 % of the total number of mutants screened. Growth of all of the 48 confirmed abacteremic mutants on solid media was similar to the parental wild-type strain (data not shown).

We determined the transposon insertion sites for all 48 abacteremic mutants by direct sequencing out of the transposon into the flanking chromosomal region and mapping of the derived sequenced onto the draft genome sequence of *B. birtlesii* (S. Cescau, H.M. Yang, J. Wang, M. Vayssier-Taussat, A. Danchin, and F. Biville, unpublished data). Three mutants harboring two separate transposon insertions were not considered further in the analysis. Table S1 lists the loci inactivated by single transposon insertion in the remaining 45 abacteremic mutants. Five mutants carried the transposon insertion in an intergenic region: one (83D04) was near a gene encoding a tRNA; three of them (04A01, 86C05, 69B07) were upstream of genes encoding proteins of unknown function and one (69C09) was in proximity to a putative transcriptional regulator gene. In these mutants, the transposon may have thus disrupted a promoter or other regulatory sequence. 40 transposon insertions were mapped to the coding region of 38 different protein-encoding genes. In 8

mutants the insertions were found in genes encoding a conserved protein of unknown function, among them three putative surface proteins. Sixteen mutants carried insertions in genes previously implicated in bacterial pathogenicity, either in *Bartonella* (*virB/D4*, *trw*, *ialA/B*, *badA*, *omp43*, *iba*) or other pathogenic bacteria (encoding heat shock proteins) (Dehio 2004). Moreover, mutant genes encoding proteins involved in transport and metabolism, as well as phage-related function were also identified.

B. birtlesii genes required for erythrocytic infection *in vitro*

The 45 confirmed abacteremic mutants with single transposon insertion were individually tested for their capacity to invade murine erythrocytes using the gentamicin protection assay (Fig. 4 and Table S1). Nine mutants were found to be impaired in murine erythrocyte colonization. Seven of them harbored a mutation in the same operon encoding the T4SS Trw (two in *trwD*, *trwE*, *trwF, trwJ2, trwL1, trwL2)* which was previously shown to be important for establishing an intraerythrocytic bacteremia in *B. tribocorum* (Seubert et al. 2003). Compared to wild-type, both *trwD* mutants (04B03 and 41C12) showed a five-fold decrease in invasion efficiency. All other *trw* mutants failed to invade erythrocytes. We confirmed the direct role of the *trw* operon in erythrocyte invasion in the *B. tribocorum*/rat model. Δ*trwE* mutants (Saenz et al. 2007) failed to invade rat erythrocytes, while complementation of the Δ*trwE* mutant restored invasiveness (Fig. 5). The mutant harboring an insertion in the invasion-associated locus *ialA/B* locus also showed an impaired erythrocyte invasion phenotype (10-fold reduced, $p<0.01$), confirming the previously suggested role of this locus in erythrocyte infection (Mitchell and Minnick 1995; Coleman and Minnick 2001). One clone (25A02) mutated in *livG* (encoding an amino acid ABC-

transporter) showed a slight decrease of invasion efficiency (4-fold, $p<0.05$) compared to wild-type. None of the other abacteremic mutants appeared to be involved in erythrocytes invasion indicating that they probably are required for an earlier step of infection, i.e. for colonization of the primary niche.

Role of Trw T4SS in host-specific infection of erythrocytes

Next we tested whether the Trw system shown here to be essential for erythrocyte infection *in vitro* and *in vivo* may also determine host-specificity. To this end we introduced pAB2, a plasmid expressing the entire *trw* locus of *B. tribocorum* (Seubert et al. 2003), in *B. henselae* and tested the capacity of this strain [*B. henselae* (pAB2)] to infect rat erythrocytes in comparison to the parental *B. henselae* strain and *B. tribocorum* (+/- pAB2). As shown in Fig. 6, expression of the *trw* locus of *B. tribocorum* rendered *B. henselae* permissive to the infection of rat erythrocytes, thereby supporting a major role of the Trw system in host-specific erythrocyte recognition.

To further assess which components of the Trw T4SS may mediate host specificity we analyzed the molecular evolution of different *trw* genes of *B. birtlesii* and related species. Candidate genes for mediating host specificity are surface exposed components, i.e. the T4SS pilus components TrwL and TrwJ. As shown by Nystedt *et al.* (Nystedt et al. 2008) for other *Bartonella* species, *trwL* and *trwJ* genes have been amplified and diversified several times during evolution. The *trw* locus of *B. birtlesii* also displays amplification of *trwL* (five copies) and co-amplification of *trwJ* together with *trwH and trwI* (two copies) (Fig. 7, panel A). Phylogenetic analyses and calculation of the non-synonymous (*d*N) and synonymous (*d*S) substitution frequencies of different *trw* genes further showed that *trwJ* and *trwL* homologs have

diversified to much higher degree than other components of the Trw T4SS, within and among different species [Fig. 7, panel B-G, and (Nystedt et al. 2008)].

DISCUSSION

Host-specificity is a prominent feature of pathogenic bacteria that reflects the host range susceptible to infection. Subtle changes in the molecular mechanisms that govern host-specificity may result in sudden host shifts, which represent a major risk for the emergence of novel human pathogens from animal reservoirs. Striking examples for this evolutionary scenario are the bartonellae, which cause host-restricted intra-erythrocytic infections in their mammalian reservoirs. In conjunction with repeated host shifts, the large number of *Bartonella* species evolved by adaptive radiation (Saenz et al. 2007), including the human-specific pathogen *B. quintana* that evolved from cat-specific *B. henselae* (Alsmark et al. 2004). Here we explored the bacterial genetic basis for host-restricted infection of erythrocytes. The establishment of an *in vitro* model of erythrocyte adherence and invasion allowed us to demonstrate for the first time a direct correlation of host-restricted erythrocyte infection *in vivo* and *in vitro*, demonstrating that host-specificity is determined by direct interaction of bacteria with erythrocytes. In order to identify the bacterial factors critical for host-restricted erythrocyte infection we have used a two-step experimental protocol. First, we performed an STM screen for *B. birtlesii* in mice which allowed us to identify 45 abacteremic mutants defective in establishing intra-erythrocytic infection. Among the corresponding set of 38 protein-encoding genes, 13 loci were also indentified in a similar STM screen performed in the *B. tribocorum*/rat model (Saenz et al. 2007), indicating extensive similarities in the repertoire of pathogenesis factors in these closely related organisms as well as robustness of the performed genetic screens. Second, rescreening of the entire set of 45 abacteremic *B. birtlesii* mutants in the *in vitro* mouse erythrocyte infection model resulted in the identification of nine mutants

impaired in this assay. The other mutants (36 of 45 = 80%) displaying a wild-type phenotype in this assay are therefore not directly involved in erythrocyte infection, but rather may contribute to the establishment of infection in the primary niche. Prominent examples are the *virB/virD4* genes encoding the VirB/VirD4 T4SS, which is known to be required for primary niche infection in the *B. tribocorum*/rat model (Schulein, 2002). Moreover, a recent study inferred the VirB/VirD4 T4SS as major bacterial factor facilitating bacterial adaptation to novel hosts (Saenz et al. 2007). The nine mutants impaired in erythrocyte infection *in vitro* include transposon insertions in the invasion locus (*ialA/B*) previously implicated in erythrocyte invasion (Mitchell and Minnick 1995; Coleman and Minnick 2001), *livG* encoding an amino acid ABC-transporter, and most prominently six genes encoding different components of the T4SS Trw. Trw is known to be required for establishing intra-erythrocytic infection in the *B. tribocorum/*rat model (Seubert et al. 2003; Dehio 2004, 2008), however, evidence for a direct role of the Trw system in erythrocyte infection as provided here was lacking so far. Based on the presumable surface location of components of Trw (Seubert et al. 2003) this T4SS may directly interact with the erythrocyte surface and thus may restrict the host range of erythrocyte infection. To test the hypothesis that Trw determines host range we have expressed Trw of rat-specific *B. tribocorum* in cat-specific *B. henselae*. Strikingly, this genetic manipulation resulted in an extension of the host range for *in vitro* erythrocyte infection towards rats, demonstrating that Trw indeed represents a major determinant of host-specificity of erythrocyte infection. Further to the molecular paradigm of host-specificity exemplified by the interaction of two surface proteins of *L. monocytogenes*, InlA and InlB, with their respective host receptors (Lecuit et al. 1999; Khelef et al. 2006; Disson et al. 2008), this finding

establishes a new experimental model to study the molecular mechanisms governing host restriction.

The Trw locus was laterally acquired during evolution of the bartonellae. It is present in the largest sub-branch of the genus tree, comprising 13 species that are adapted to diverse mammalian reservoir hosts, while it is absent from human-specific *B. bacillifformis*, cat-specific *Bartonella clarridgeiae* and the species of the ruminent-specific sub-branch, which diverted early during evolution of the bartonellae (Saenz et al. 2007). Interestingly, the acquisition of Trw by the modern lineage correlates with the loss of flagella, which are know to represent a major pathogencity factor for the invasion of erythroctes by *B. bacilliformis* and probably other flagellated bartonellae (Dehio 2008). The Trw system of *Bartonella* represents an interesting example of a pathogenesis-related T4SS that evolved rather recently by functional diversification of a laterally acquired bacterial conjugation system. Its locus displays characteristic features of a pathogenicity island and shares extensive similarity with the *trw* locus of IncW broad-host range plasmid R388 encoding a genuine conjugation system. The *trw* loci of *Bartonella* and R388 are colinear, except for multiple tandem gene duplications of *trwL* and *trwJ-trwH* in *Bartonella*. Complementation of R388 derivatives carrying mutations in different *trw* genes with their *Bartonella* homologues allowed to demonstrate functional interchangability for some T4SS components (Seubert et al. 2003; de Paz et al. 2005), underscoring the structural and functional conservation of individual subunits of these functionally diversified T4SSs. However, a major difference between these homologous systems is the lack of the coupling protein TrwB in *Bartonella*, which in R388 is required for export of T4SS substrates. The lack of TrwB in *Bartonella* thus indicates that its Trw system does not translocate substrates. However, the multiple copies of *trwL* and

trwJ in the *Bartonella trw* locus encode variant forms of surface-exposed pilus components, which probably are all co-expressed (Seubert et al. 2003), indicating that the primary function of the *Bartonella* Trw system may be the formation of variant pilus forms (Dehio 2008). Based on the essential role of the Trw system for erythrocyte invasion and its role in determining host range it is conceivable to assume that these variant pili may facilitate the specific interaction with polymorphic erythrocyte receptors, either within the reservoir host population (e.g. different blood group antigens), or among different reservoir hosts. Phylogenetic analyses and calculation of the non-synonymous (*d*N) and synonymous (*d*S) substitution frequencies of different *trw* genes indeed demonstrated that *trwJ* and *trwL* homologs have diversified to much higher degree than other components of the Trw T4SS, within and among different species (Nystedt et al. 2008). Together with the notion that the number of tandem repeats of *trwL* and *trwIJIH* are variable among different *Bartonella* species these findings indicate that *trwL* and *trwJ* genes have been amplified and diversified several times during evolution. Horizontal transfer of such genes from a different bartonellae – similarly as we have demonstrated here for the entire *trw* operon of *B. tribocorum* resulting in an extension of the host range of *B. birtlesii* – or alternatively pre-adaption of superfluous copies of *trwL* and *trwJ* may represent realistic molecular evolutionary scenarios for host-shifts and thereby the evolution of pathogens with an altered host-specificity as it has happened repeatedly during the evolution of the bartonellae. Future studies should identify the nature of the erythrocyte receptors targeted by the Trw system and their specific interaction that facilitate host-specific erythrocyte infection.

MATERIAL AND METHODS

Ethics Statement. Animals were handled in strict accordance with good animal practice as defined by the relevant European (European standards of welfare for animals in research), national (Information and guidelines for animal experiments and alternative methods, Federal Veterinary Office of Switzerland) and/or local animal welfare bodies. Animal work performed at the Biozentrum of the University of Basel was approved by the Veterinary Office of the Canton Basel City on June 2003 (licence no. 1741), and animal work performed at the Ecole Nationale Vétérinaire d'Alfort (ENVA/AFSSA) was approved by the institute's ethics committee on September 2005.

Bacterial strains and growth conditions. *B. alsatica* (IBS 382^T, CIP 105477^T) (Heller et al. 1999), *B. birtlesii* (IBS 135^T, CIP 106691^T) (Bermond et al. 2000), *B. chomelii* (A 828^T, CIP 107869^T) (Maillard et al. 2004), *B. henselae* (Houston-1, ATCC 49882^T), *B. quintana* (Fuller^T, ATCC VR-358^T), *B. tribocorum* (IBS 506^T, CIP 105476^T) (Heller et al. 1998), *B. vinsonii subsp. berkhoffii* (ATCC 51672^T), *B. vinsonii subsp arupensis* (ATCC 700727) (Welch et al. 1999) were grown for 5 days on Columbia agar containing 5% defibrinated sheep blood (CBA) in a humidified atmosphere with 5% $C0_2$ at 35°C.

Construction of bacterial strains. *B. tribocorum-gfp* containing a chromosomally-integrated *gfp*-expression cassette (Schulein et al. 2001) was used as GFP-expressing *B. tribocorum* strain. GFP-expressing bacteria of other *Bartonella* species were obtained by electroporation with plasmid pBBR1-MCS2 as previously described

(Kovach et al. 1995; Fournier et al. 2001). This plasmid was extracted and purified from *B. quintana* using a Midi Prep Kit (Qiagen). The electroporation procedures was described previously (Fournier et al. 2001). Transformed bacteria were selected by plating on CBA-Km. A signature-tagged mutant library of *B. birtlesii* IBS135^T was constructed as described for *B. tribocorum* (Saenz and Dehio 2005; Saenz et al. 2007). *B. tribocorum* Δ*trwE* mutant and complemented Δ*trwE* (Δ*trwE*-comp.) have been described (Seubert et al. 2003). Cosmid pAB2 encoding the entire *trw* locus of *B. tribocorum* (Dehio et al. 1998; Seubert et al. 2003; de Paz et al. 2005) was introduced into *B. henselae* by three parental mating (Dehio and Meyer 1997; Dehio et al. 1998).

***In vitro* infection of erythrocytes.** Erythrocytes from peripheral blood of mice (Balb/C), cats, rats (Wistar) and humans were isolated and purified by Ficoll gradient centrifugation. After washing in PBS, they were maintained in F12 modified medium [supplemented with 10% fetal calf serum, 2 mM glutamine, 1 mM sodium pyruvate, 0.1 mM Hepes, 257 mM histidine, 0.1 mg/ml hematin/histidine, non-essential amino acid (Gibco, FRANCE)] at $2x10^8$/ml. For *in vitro* infection experiments, *Bartonella* species were grown on CBA or CBA-km (*Bartonella-gfp* and STM mutants) plates. After 5 days of culture (10 days for GFP-expressing *Bartonella*), bacteria were harvested, washed, suspended in PBS, and added to erythrocytes at a multiplicity of infection (MOI) varying from 0.01 to 10 and incubated at 35°C in 5% $C0_2$ for various periods of time (from 1 to 3 days).

Detection of erythrocyte-associated bacteria. Colonization of erythrocytes by *Bartonella* was assessed and quantified by the gentamicin protection assay, flow

cytometry and confocal microscopy. For the quantification of intracellular bacteria by gentamicin protection, 100 µl were withdrawn from the invasion mixtures after 1, 2 or 3 days of *in vitro* infection. Mouse erythrocytes were separated from non-associated bacteria by washing with PBS and centrifuged at 500 g for 5 min. Erythrocytes were then incubated for 2 h at 35°C with gentamicin sulfate (250 µg/ml) to kill residual extracellular bacteria. Erythrocytes were then washed three times in PBS to remove the antibiotic and intracellular bacteria were released from erythrocytes by hypotonic lyses of erythrocytes in 10 µl of sterile water by freezing at -20°C for 15 min. After thawing, serial dilutions of bacteria in PBS were inoculated onto CBA plates and incubated at 35°C for 5 days before being counted. For data presentation, all measurements were expressed as the number of CFU/10^{10} erythrocytes (corresponding to ≈1 ml of blood).

For flow cytometric detection of erythrocyte-associated bacteria, measurements were performed at day 1, 2, 3 after *in vitro* infection of erythrocytes with a ten days old culture of GFP-expressing *Bartonella sp.* 100 µl of the infection mixtures was washed 3 times in PBS and fixed for 10 min with 0.8% paraformaldehyde and 0.025% glutaraldehyde. After fixing, erythrocytes were analyzed by flow cytometry (FACScan, Becton Dickinson Bioscience, France). Data were analyzed using the CellQuestPro software, version 4.0.2. Data for 10'000 gated erythrocytes were collected and analyzed.

For confocal microscopy, 100 µl of the infection mixtures was washed three times in PBS and the erythrocytes cell surface was stained using PE-labeled anti-GPA antibodies. Samples were viewed with a Nikon Eclipse C1 Plus confocal laser scanning microscope (Nikon, Amstelveen, Netherlands) with detection in channel 1 (GFP fluorescence) and channel 2 (PE fluorescence) at original magnification x100.

STM library. The transposon vectors pHS006-Tag-001 to pHS006-Tag-036 are containing each: an *oriT* for conjugative transfer, the *Himar1* transposon, a kanamycin resistant marker, a hyperactive transposase and one of 36 distinct signature-tags (Saenz et al. 2007). These 36 signature-tagged mariner transposon vectors were separately transferred from *E. coli* β2155 to *B. birtlesii* by two-parental mating as previously described (Dehio and Meyer 1997). From each mating, 96 single kanamycin-resistant *B. birtlesii* transconjugants were transferred to a 96-well plate with cryo-medium and stored at -80°C.

Mouse infections. Eight weeks old female Balb/C mice from Charles River Laboratories were housed in an animal facility (2 animals/cage) and allowed to acclimate to the facility and the diet for at least 5 days prior infection. Food and water were provided *ad libitum*. 36 differently signature-tagged mutants were grown separately from the transposon library for each input pool. They were pooled in PBS immediately before infection, and used to infect two mice with a total inoculum of $5x10^7$ colony forming units (10 µl of OD_{595}=1) in the ear dermis of Balb/C mice. The remainder of the input pools was heated at 100°C for 10 min and used as template for PCR detection. Fifty µl of blood were taken from the tail vein of the infected mice when bacteremia is peaking (days 7 and 14 post-infection) (Boulouis et al. 2001). Bacteria released from erythrocytes by a freeze/thaw cycle were plated on CBA-km. After 10 days, bacterial colonies (output pool) were counted, harvested in PBS, suspended to OD_{595}=1 and heated at 100°C for 10 min to be used as template for PCR detection. The rescreen was done following the same protocol using pools of nine mutants (two abacteremic mutants and seven mutants displaying a wild-type phenotype)

PCR detection of abacteremic mutants. For signature-tag identification, the generic primer Srev01 corresponding to a sequence in the transposon and a set of tag-specific primer were used for amplification of a fragment of approximately 600 bp (Saenz et al. 2007). The conditions for the PCR were as follows: a first denaturation step at 95°C for 5 min, followed by 30 cycles of PCR with denaturation at 95°C for 1 min, annealing for 30 s at 52°C, and extension at 72°C for 1 min. The program was completed by an extension step at 72°C for 5 min. The amplified fragments were displayed on a 1% agarose gel. Mutants that were detected in the input pools and absent from the out put pools (days 7 and 14) in both mice were considered as abacteremic mutants.

Identification and analysis of transposon insertion sites. Genomic DNA from abacteremic mutants, regrown from the library, was prepared with the ROCHE Genomic DNA Isolation Kit. Genomic DNA was sent to QIAGEN for sequencing with primers Tn_{start} and Tn_{end} (Saenz et al. 2007). The sequences obtained by the genomic sequencing were compared by BlastN to the nr data base of NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The exact transposon insertion sites were found by comparing the genomic sequences to contigs of the ongoing *B. birtlesii* genome sequencing project by BlastN.

Screening of abacteremic mutants for their capacity to infect murine erythrocytes. Mutants displaying an abacteremic phenotype were tested for their capacity to invade murine erythrocytes using the gentamicin protection assay. Each mutant was tested at MOI=1 in at least two independent experiments performed in

triplicate samples. For mutants displaying an impaired erythrocyte invasion phenotype, invasion assays were performed at least three times in triplicate samples.

Statistical analysis. Numerical data are reported as the mean of at least 3 replicate samples +/- standard errors of the means. Statistical significance of the data was measured by use of Student's t test. A p-value <0.05 was considered significant.

Phylogenetic and evolutionary analysis. The sequence of the *B. birtlesii trw* locus was deposited under the EMBL-EBI accession no. FN555106. Sequence alignments were calculated with ClustalW as implemented in MEGA4. Phylogenetic trees were inferred by maximum likelihood methods with Paup 4.0 (Wilgenbusch and Swofford 2003) and 100 bootstrap replicates were calculated. To select an appropriate substitution model the Akaike information criterion of Modeltest 3.7 was used (Posada and Crandall 1998). The models obtained were general time reversible (GTR) + I for *trwFED* and *trwN*, transversion model (TVM) + I for *trwI*, and TVM + I + G for *trwJ* and *trwL*. Nonsynonymous (dN) and synonymous (dS) substitution frequencies were calculated using the method of Yang and Nielson (Yang and Nielsen 2000) as implemented in the PAML package (Yang 1997, 2007).

ACKNOWLEDGEMENTS:

We thank Dr. Arto Pulliainen for critically reading of the manuscript and Drs Didier Raoult and Pierre Edouard Fournier for providing the pBBR1-MCS2 plasmid.

REFERENCES:

Alsmark CM, Frank AC, Karlberg EO, Legault BA, Ardell DH et al. (2004) The louse-borne human pathogen Bartonella quintana is a genomic derivative of the zoonotic agent Bartonella henselae. Proc Natl Acad Sci U S A 101(26): 9716-9721.

Bermond D, Heller R, Barrat F, Delacour G, Dehio C et al. (2000) *Bartonella birtlesii* sp. nov., isolated from small mammals (*Apodemus* spp.). Int J Syst Evol Microbiol 50: 1973-1979.

Boulouis HJ, Barrat F, Bermond D, Bernex F, Thibault D et al. (2001) Kinetics of *Bartonella birtlesii* infection in experimentally infected mice and pathogenic effect on reproductive functions. Infect Immun 69(9): 5313-5317.

Breitschwerdt EB, Maggi RG, Sigmon B, Nicholson WL (2007) Isolation of *Bartonella quintana* from a woman and a cat following putative bite transmission. J Clin Microbiol 45(1): 270-272.

Chomel BB, Boulouis HJ, Breitschwerdt EB, Kasten RW, Vayssier-Taussat M et al. (2009a) Ecological fitness and strategies of adaptation of *Bartonella* species to their hosts and vectors. Vet Res 40(2): 29.

Chomel BB, Henn JB, Kasten RW, Nieto NC, Foley J et al. (2009b) Dogs are more permissive than cats or guinea pigs to experimental infection with a human isolate of *Bartonella rochalimae*. Vet Res 40(4): 27.

Coleman SA, Minnick MF (2001) Establishing a direct role for the *Bartonella bacilliformis* invasion- associated locus B (*ialB*) protein in human erythrocyte parasitism. Infect Immun 69(7): 4373-4381.

de Paz HD, Sangari FJ, Bolland S, Garcia-Lobo JM, Dehio C et al. (2005) Functional interactions between type IV secretion systems involved in DNA transfer and virulence. Microbiology 151(Pt 11): 3505-3516.

Dehio C (2004) Molecular and cellular basis of *Bartonella* pathogenesis. Annu Rev Microbiol 58: 365-390.

Dehio C (2005) *Bartonella*-host-cell interactions and vascular tumour formation. Nat Rev Microbiol 3(8): 621-631.

Dehio C (2008) Infection-associated type IV secretion systems of *Bartonella* and their diverse roles in host cell interaction. Cell Microbiol 10(8): 1591-1598.

Dehio C, Meyer M (1997) Maintenance of broad-host-range incompatibility group P and group Q plasmids and transposition of Tn5 in *Bartonella henselae* following conjugal plasmid transfer from *Escherichia coli*. J Bacteriol 179(2): 538-540.

Dehio M, Knorre A, Lanz C, Dehio C (1998) Construction of versatile high-level expression vectors for *Bartonella henselae* and the use of green fluorescent protein as a new expression marker. Gene 215(2): 223-229.

Disson O, Grayo S, Huillet E, Nikitas G, Langa-Vives F et al. (2008) Conjugated action of two species-specific invasion proteins for fetoplacental listeriosis. Nature 455(7216): 1114-1118.

Finlay BB, Cossart P (1997) Exploitation of mammalian host cell functions by bacterial pathogens. Science 276(5313): 718-725.

Foucault C, Rolain JM, Raoult D, Brouqui P (2004) Detection of *Bartonella quintana* by direct immunofluorescence examination of blood smears of a patient with acute trench fever. J Clin Microbiol 42(10): 4904-4906.

Fournier PE, Minnick MF, Lepidi H, Salvo E, Raoult D (2001) Experimental model of human body louse infection using green fluorescent protein-expressing *Bartonella quintana*. Infect Immun 69(3): 1876-1879.

Heller R, Riegel P, Hansmann Y, Delacour G, Bermond D et al. (1998) *Bartonella tribocorum* sp. nov., a new *Bartonella* species isolated from the blood of wild rats. Int J Syst Bacteriol 48: 1333-1339.

Heller R, Kubina M, Mariet P, Riegel P, Delacour G et al. (1999) *Bartonella alsatica* sp. nov., a new *Bartonella* species isolated from the blood of wild rabbits. Int J Syst Bacteriol 49: 283-288.

Kabeya H, Tsunoda E, Maruyama S, Mikami T (2003) Immune responses of immunocompetent and immunocompromised mice experimentally infected with *Bartonella henselae*. J Vet Med Sci 65(4): 479-484.

Khelef N, Lecuit M, Bierne H, Cossart P (2006) Species specificity of the *Listeria monocytogenes* InlB protein. Cell Microbiol 8(3): 457-470.

Kosoy MY, Saito EK, Green D, Marston EL, Jones DC et al. (2000) Experimental evidence of host specificity of *Bartonella* infection in rodents. Comp Immunol Microbiol Infect Dis 23(4): 221-238.

Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA et al. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene 166(1): 175-176.

Lecuit M, Dramsi S, Gottardi C, Fedor-Chaiken M, Gumbiner B et al. (1999) A single amino acid in E-cadherin responsible for host specificity towards the human pathogen *Listeria monocytogenes*. Embo J 18(14): 3956-3963.

Maillard R, Riegel P, Barrat F, Bouillin C, Thibault D et al. (2004) *Bartonella chomelii* sp. nov., isolated from French domestic cattle (Bos taurus). Int J Syst Evol Microbiol 54(Pt 1): 215-220.

Mavris M, Saenz H, Monteil M, Boulouis HJ, Dehio C et al. (2005) Characterization of genes involved in long-term bacteremia in mice by *Bartonella birtlesii*. Ann N Y Acad Sci 1063: 312-314.

Mehock JR, Greene CE, Gherardini FC, Hahn TW, Krause DC (1998) *Bartonella henselae* invasion of feline erythrocytes *in vitro*. Infect Immun 66(7): 3462-3466.

Mitchell SJ, Minnick MF (1995) Characterization of a two-gene locus from *Bartonella bacilliformis* associated with the ability to invade human erythrocytes. Infect Immun 63(4): 1552-1562.

Nystedt B, Frank AC, Thollesson M, Andersson SG (2008) Diversifying selection and concerted evolution of a type IV secretion system in *Bartonella*. Mol Biol Evol 25(2): 287-300.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14(9): 817-818.

Rolain JM, La Scola B, Liang Z, Davoust B, Raoult D (2001) Immunofluorescent detection of intraerythrocytic *Bartonella henselae* in naturally infected cats. J Clin Microbiol 39(8): 2978-2980.

Rolain JM, Locatelli C, Chabanne L, Davoust B, Raoult D (2004) Prevalence of *Bartonella clarridgeiae* and *Bartonella henselae* in domestic cats from France and detection of the organisms in erythrocytes by immunofluorescence. Clin Diagn Lab Immunol 11(2): 423-425.

Rolain JM, Foucault C, Guieu R, La Scola B, Brouqui P et al. (2002) *Bartonella quintana* in human erythrocytes. Lancet 360(9328): 226-228.

Saenz HL, Dehio C (2005) Signature-tagged mutagenesis: technical advances in a negative selection method for virulence gene identification. Curr Opin Microbiol 8(5): 612-619.

Saenz HL, Engel P, Stoeckli MC, Lanz C, Raddatz G et al. (2007) Genomic analysis of *Bartonella* identifies type IV secretion systems as host adaptability factors. Nat Genet 39(12): 1469-1476.

Scherer DC, DeBuron-Connors I, Minnick MF (1993) Characterization of *Bartonella bacilliformis* flagella and effect of antiflagellin antibodies on invasion of human erythrocytes. Infect Immun 61(12): 4962-4971.

Schulein R, Seubert A, Gille C, Lanz C, Hansmann Y et al. (2001) Invasion and persistent intracellular colonization of erythrocytes. A unique parasitic strategy of the emerging pathogen *Bartonella*. J Exp Med 193(9): 1077-1086.

Schulein R, Guye P, Rhomberg TA, Schmid MC, Schroder G et al. (2005) A bipartite signal mediates the transfer of type IV secretion substrates of *Bartonella henselae* into human cells. Proc Natl Acad Sci U S A 102(3): 856-861.

Seubert A, Hiestand R, de la Cruz F, Dehio C (2003) A bacterial conjugation machinery recruited for pathogenesis. Mol Microbiol 49(5): 1253-1266.

Welch DF, Carroll KC, Hofmeister EK, Persing DH, Robison DA et al. (1999) Isolation of a new subspecies, *Bartonella vinsonii* subsp. arupensis, from a cattle rancher: identity with isolates found in conjunction with *Borrelia burgdorferi* and *Babesia microti* among naturally infected mice. J Clin Microbiol 37(8): 2598-2601.

Wilgenbusch JC, Swofford D (2003) Inferring evolutionary trees with PAUP*. Curr Protoc Bioinformatics Chapter 6: Unit 6 4.

Yamamoto K, Chomel BB, Kasten RW, Hew CM, Weber DK et al. (2002) Experimental infection of specific pathogen free (SPF) cats with two different strains of *Bartonella henselae* type I: a comparative study. Vet Res 33(6): 669-684.

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13(5): 555-556.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24(8): 1586-1591.

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17(1): 32-43.

FIGURES

Figure 1: *B. birtlesii* invades murine erythrocytes *in vitro*. (A, B) Time- and bacterial number-dependency of *B. birtlesii* invasion of murine erythrocytes determined by the gentamicin protection assay. **(A)** Freshly isolated murine erythrocytes were infected with *B. birtlesii* at the indicated multiplicity of infection (MOI) and the numbers of intra-erythrocytic bacteria (colony forming units, CFU) was determined by the gentamicin protection assay at 1, 2 and 3 days post infection (DPI); n=6, mean +/-SD; *, **: significant difference of data compared to 1 DPI. **(B)** Freshly isolated murine erythrocytes were infected with *B. birtlesii* at MOI=1. At 1 DPI, gentamicin was added to half of the samples, while as control the other half was not treated. For both untreated and gentamicin treated samples, numbers of intra-erythrocytic bacteria were determined by the gentamicin protection assay at 1, 2 and 3 DPI (n=6; mean +/-SD, *,**: significant difference in gentamicin treated samples compared to one DPI). **(C)** Time- and **(D)** bacterial number-dependency of *B. birtlesii* associated to murine erythrocytes determined by flow cytometry. Freshly isolated murine erythrocytes were infected with *B. birtlesii-gfp* (MOI=1, detection at two and three DPI in C and MOI=0.1 or 1, detection at three DPI in D). The percentage of erythrocytes associated with bacteria were quantified by flow cytometric analysis at 2 and 3 DPI. Representative data for the fluorescence (FL-1) of 10’000 erythrocytes are shown as histogram plots. **(E)** Confocal microscopic analysis of murine erythrocytes infected for 2 days with GFP-expressing *B. birtlesii*. Arrows point to bacteria found in close association with erythrocytes.

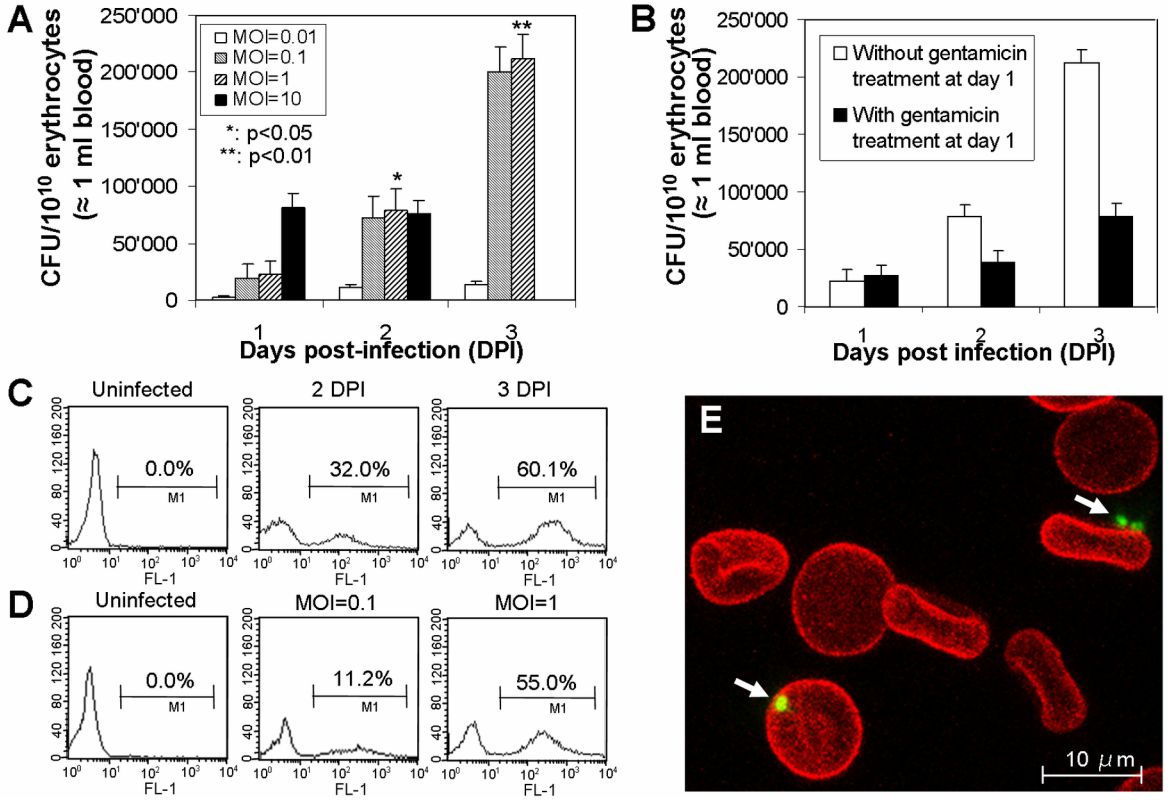

A
CFU/10^10 erythrocytes (≈ 1 ml blood)
250'000
200'000
150'000
100'000
50'000
0
MOI=0.01
MOI=0.1
MOI=1
MOI=10
*: p<0.05
**: p<0.01
*
**
1
2
3
Days post-infection (DPI)
B
CFU/10^10 erythrocytes (≈ 1 ml blood)
Without gentamicin treatment at day 1
With gentamicin treatment at day 1
Days post infection (DPI)
C
Uninfected
2 DPI
3 DPI
0.0%
32.0%
60.1%
M1
FL-1
D
Uninfected
MOI=0.1
MOI=1
0.0%
11.2%
55.0%
E
10 µm

Figure 1

Figure 2: Efficiency of *in vitro* invasion of murine erythrocytes by different *Bartonella* species. Freshly isolated murine erythrocytes were infected with the indicated *Bartonella* species with a MOI=1. The numbers of intra-erythrocytic bacteria (colony forming units, CFU) was determined by the gentamicin protection assay at 1 and 2 days post infection (DPI); mean +/-SD of triplicate samples.

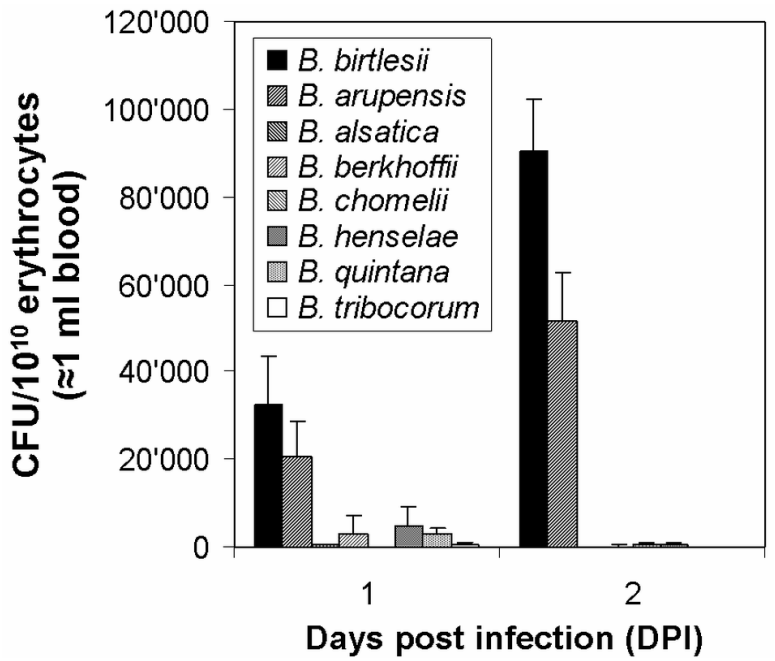

120'000
100'000
80'000
60'000
40'000
20'000
0
CFU/10¹⁰ erythrocytes
(≈1 ml blood)
B. birtlesii
B. arupensis
B. alsatica
B. berkhoffii
B. chomelii
B. henselae
B. quintana
B. tribocorum
1
2
Days post infection (DPI)

Figure 2

Figure 3. Efficiency of interaction between erythrocyte and *Bartonella* sp. according to host origin and *Bartonella* species. Freshly isolated erythrocytes from mouse, cat, human or rat were infected with *gfp*-expressing bacteria of the indicated *Bartonella* species (MOI=1). The percentages of infected erythrocytes were determined by flow cytometry at two DPI. Representative histogram plots for GFP-fluorescence (FL-1) of 10’000 erythrocytes are shown.

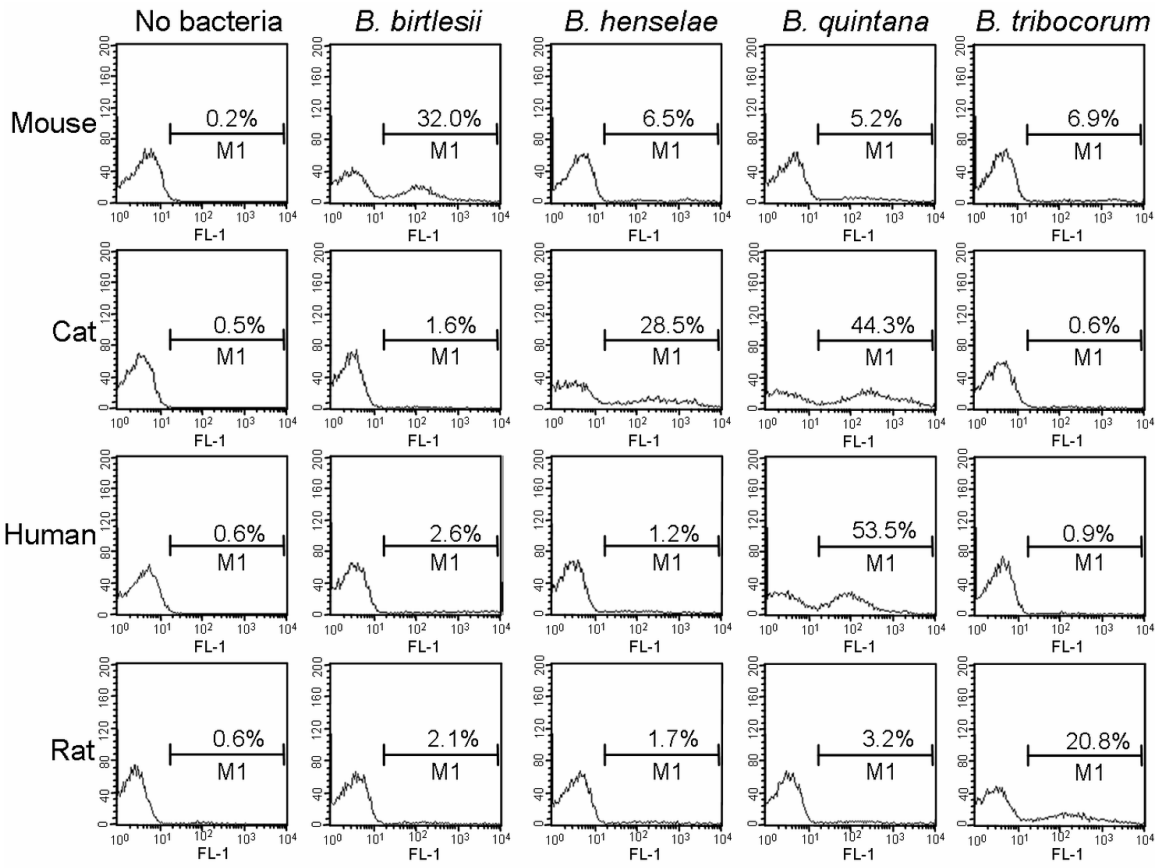

No bacteria
B. birtlesii
B. henselae
B. quintana
B. tribocorum
Mouse
Cat
Human
Rat
0.2%
32.0%
6.5%
5.2%
6.9%
0.5%
1.6%
28.5%
44.3%
0.6%
0.6%
2.6%
1.2%
53.5%
0.9%
0.6%
2.1%
1.7%
3.2%
20.8%
M1
FL-1

Figure 3

Figure 4. Role of Trw in erythrocyte infection. Efficiency of *in vitro* invasion of murine erythrocyte by abacteremic mutants of *B. birtlesii*. The *in vitro* erythrocyte invasion phenotype of abacteremic mutants identified in the STM screen was evaluated by the gentamicin protection assay at 1 DPI. The efficiency of erythrocyte invasion of each tested mutant is expressed as the percentage of erythrocyte invasion of the isogenic wild-type strain (mean +/- SD of triplicate samples; n=1 for *badA* and *virD4*, n=3 for other mutants). All mutants listed in Table S1 that do not appear in this figure display wild-type phenotype in regard of *in vitro* erythrocyte invasion.

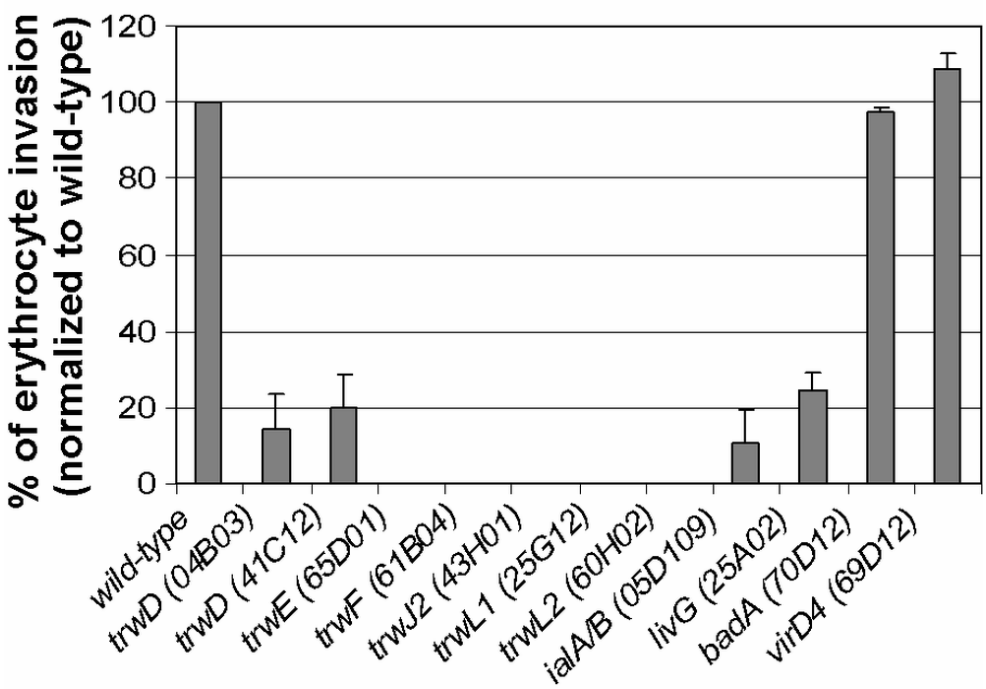

% of erythrocyte invasion (normalized to wild-type)
120
100
80
60
40
20
0
wild-type
trwD (04B03)
trwD (41C12)
trwE (65D01)
trwF (61B04)
trwJ2 (43H01)
trwL1 (25G12)
trwL2 (60H02)
ialA/B (05D109)
livG (25A02)
badA (70D12)
virD4 (69D12)

Figure 4

Figure 5: Role of *B. tribocorum trwE* in rat erythrocyte infection. Efficiency of rat erythrocyte invasion by the Δ*trwE* mutant of *B. tribocorum*, the isogenic wild-type strain (wild-type), and the complemented Δ*trwE* mutant (Δ*trwE* comp.) at 1 DPI (n=3, mean +/- SD).

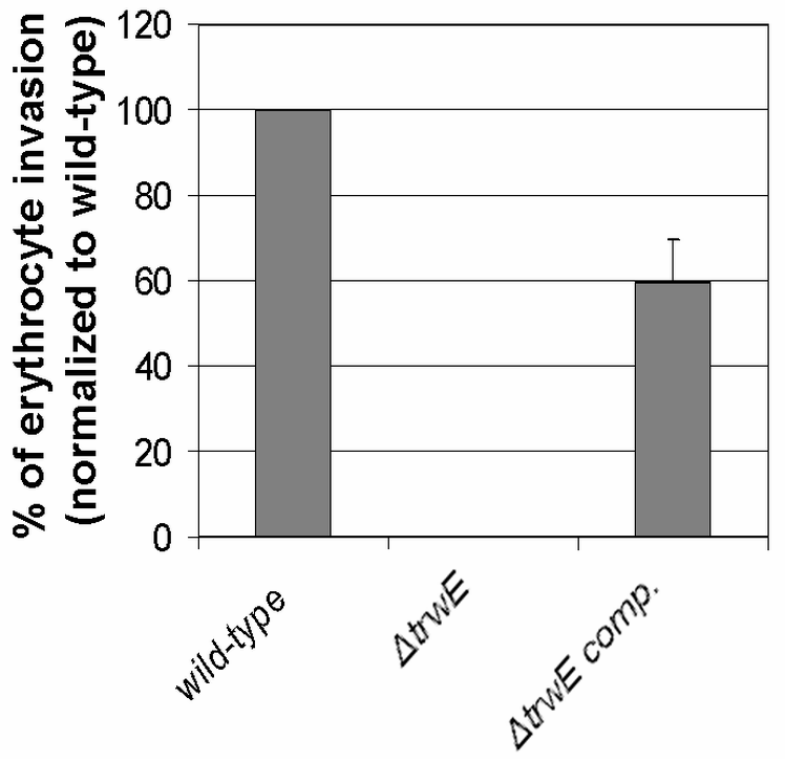

% of erythrocyte invasion
(normalized to wild-type)
120
100
80
60
40
20
0
wild-type
ΔtrwE
ΔtrwE comp.

Figure 5

Figure 6. Role of Trw T4SS in mediating host-specific erythrocyte invasion.

Freshly isolated rat erythrocytes were infected with *B. tribocorum, B. tribocorum* (pAB2), *B. henselae* and *B. henselae* (pAB2) at a MOI=1. Intra-erythrocytic bacteria were enumerated at 1, 2 and 3 days post infection (DPI) by the gentamicin protection assay (n=3; mean +/-SD).

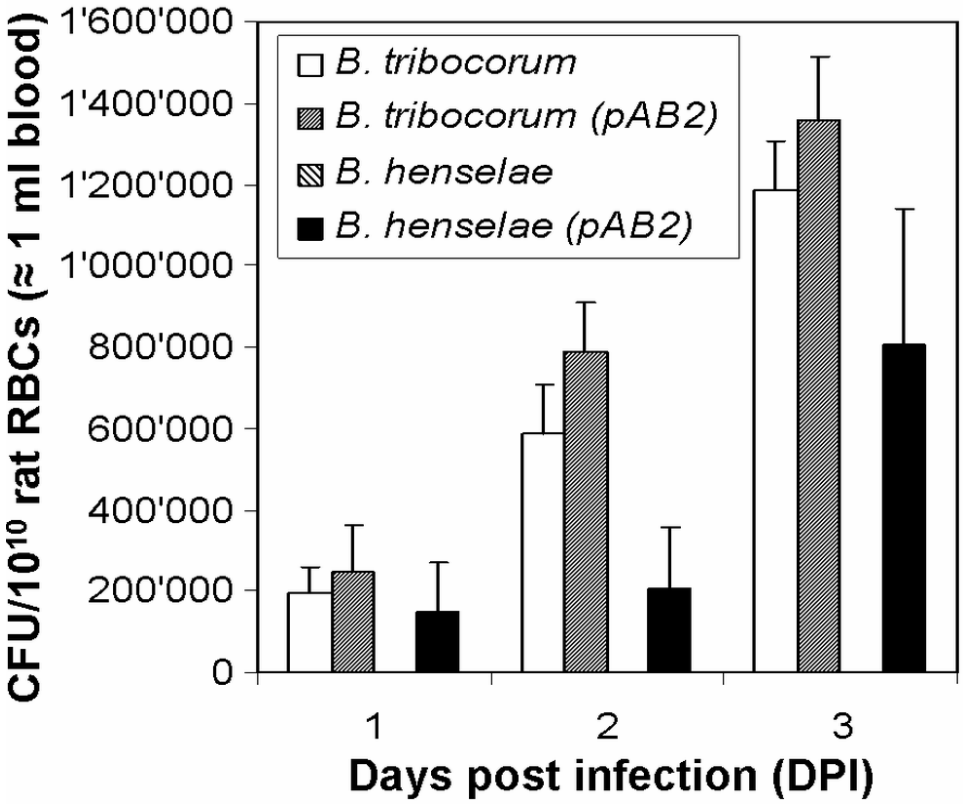

1'600'000
1'400'000
1'200'000
1'000'000
800'000
600'000
400'000
200'000
0
CFU/10^10 rat RBCs (≈ 1 ml blood)
B. tribocorum
B. tribocorum (pAB2)
B. henselae
B. henselae (pAB2)
1
2
3
Days post infection (DPI)

Figure 6

Figure 7: Genetic organization of the *Bartonella trw* locus, and phylogenies and synonymous (*d*S) vs. nonsynonymous (*d*N) substitution frequencies of the encoded *trw* genes. (A) Gene order structure of the *trw* locus of *B. birtlesii* and comparison to other *Bartonella* species. The copy number of amplified genes or segments in other *Bartonella* species is indicated within brackets. Maximum Likelihood phylogenies of **(B)** the concatenated nucleotide alignments of *trwF*, *trwE*, and *trwD*, the nucleotide alignments of **(C)** *trwJ* copies, **(D)** *trwI*, **(E)** *trwL* copies, and **(F)** *trwN* of *B. birtlesii* (*Bb*), *B. grahamii* (*Bg*), *B. henselae* (*Bh*), *B. quintana* (Bq) , and *B. tribocorum* (*Bt*). For *trwJ* **(C)** and *trwL* **(E)**, the range of pairwise *d*N/*d*S ratios of different phylogenetic subclusters (shaded areas) are indicated at the upper right of each cluster. For *trwL1*, the range of pairwise *d*N/*d*S ratios is indicated as well, although they do not cluster. **(G)** The pairwise *d*N/*d*S ratios of orthologous *trw* genes and the two adjacent genes *ubiH* and *sdhA* of *B. birtlesii* and *B. grahamii*, *B. henselae*, *B. quintana*, or *B. tribocorum* are plotted according to their gene order. For the tandem repeated genes *trwL, trwJ*, *trwI*, and *trwH* only *trwL5*, *trwJ1*, *trwI1*, and *trwH1* are shown, since ortholog assignment is difficult for the others due to copy number variation and the occurrence of recombination among different species (Nystedt et al. 2008).

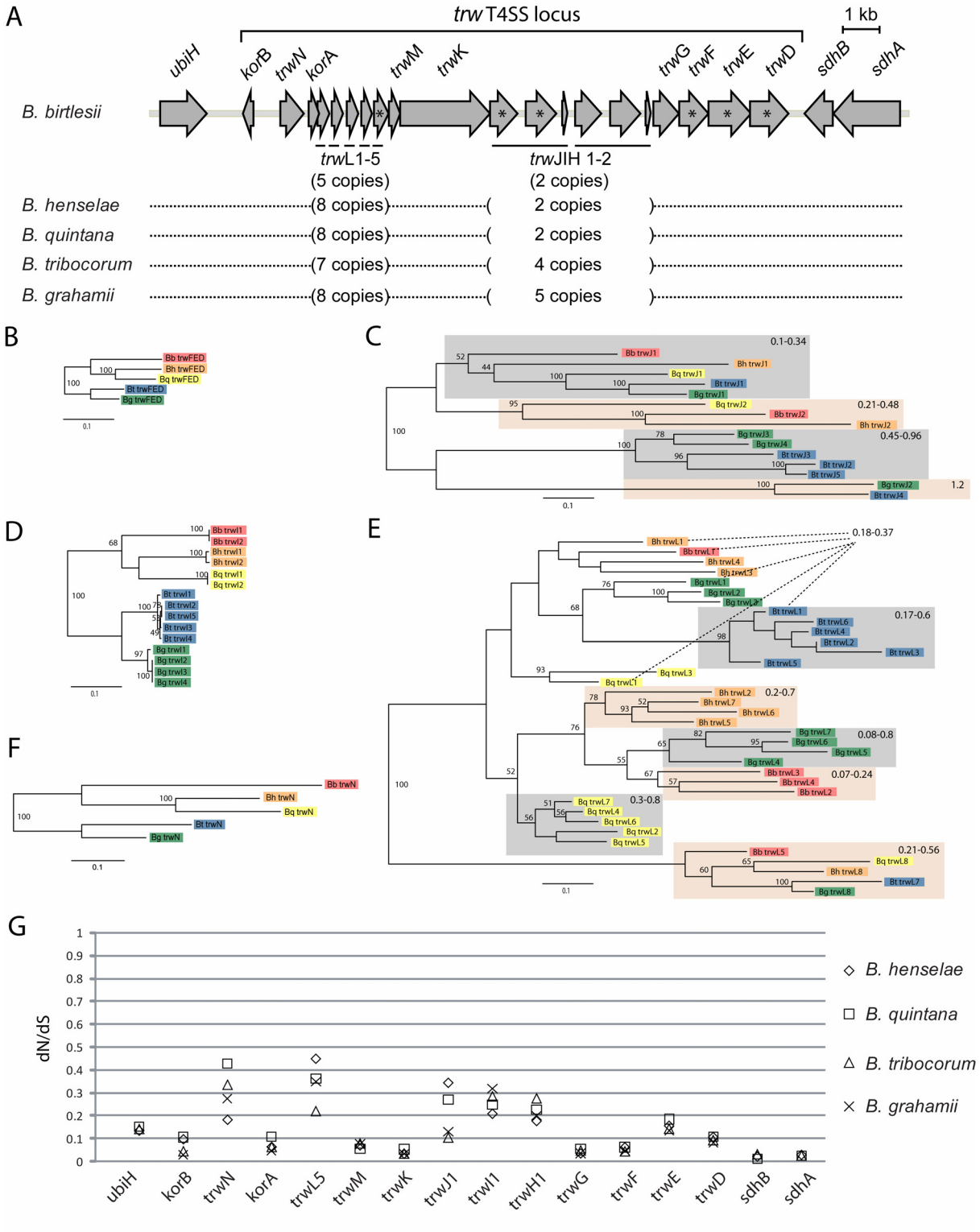

A
trw T4SS locus
1 kb
ubiH
korB
trwN
korA
trwM
trwK
trwG
trwF
trwE
trwD
sdhB
sdhA
B. birtlesii
trwL1-5
(5 copies)
trwJIH 1-2
(2 copies)
B. henselae (8 copies) (2 copies)
B. quintana (8 copies) (2 copies)
B. tribocorum (7 copies) (4 copies)
B. grahamii (8 copies) (5 copies)
B
C
D
E
F
G
dN/dS
B. henselae
B. quintana
B. tribocorum
B. grahamii

Figure 7

TABLES

Table 1: Efficiency of erythrocyte colonization according to host origin and *Bartonella* species

	B. birtlesii		***B. henselae***		***B. quintana***		***B. tribocorum***	
	in vivo	*in vitro* (%)*	*in vivo*	*in vitro* (%)*	*in vivo*	*in vitro* (%)*	*in vivo*	*in vitro* (%)*
Mouse	+[a]	26.3 +/- 2.2	-[b]	1.3 +/- 0.5	n.r.	0.8 +/- 0.3	n.r.	1.6 +/- 0.4
Cat	n.r.	1.6 +/- 0.8	+[c]	28.5 +/- 4.1	+[d]	42.2 +/- 3.1	n.r.	0.5 +/- 0.3
Human	n.r.	2.7 +/- 1.0	n.r.	1.3 +/- 0.3	+[e]	58.4 +/- 1.2	n.r.	0.9 +/- 0.2
Rat	n.r.	2.0 +/- 1.3	n.r.	1.8 +/- 0.7	n.r.	3.5 +/- 1.1	+[f]	20.7 +/- 2.8

[a](Boulouis et al. 2001), [b](Kabeya et al. 2003), [c](Rolain et al. 2001; Rolain et al. 2004), [d](Breitschwerdt et al. 2007), [e](Schulein et al. 2001; Rolain et al. 2002; Foucault et al. 2004), [f](Schulein et al. 2001)

*Freshly isolated erythrocytes from mouse, cat, human or rat were infected with *gfp*-expressing bacteria of the indicated *Bartonella* species (MOI=1) . The percentages of colonized erythrocytes were determined by flow cytometry at day two post infection. Data for 10'000 erythrocytes per time-point were analyzed (n=6 for tests with homologous species and n=3 for tests with heterologous species, mean +/- SD). For previously described infections of the respective mammalian hosts with the indicated *Bartonella* species the presence (+) or absence (-) of intraerythrocytic bacteremia is indicated (n.r. = not reported).

Supporting Table S1: Genotypic characterization of the abacteremic mutants. The columns BARBAKC, BH, BQ and BT list the extensions of systematic names of orthologous genes from the published genomes of *B. bacilliformis* (accession no. CP000524), *B. henselae* (accession no. BX897699), *B. quintana* (accession no. 897700) and *B. tribocorum* (accession no. AM260525) respectively. * The *in vitro* erythrocate invasion phenotype of each mutant was determined by the gentamicin protection assay after 1 day of infection (triplicate samples) and categorized as normal (>70% of wild-type), reduced (<70% of wild-type but >1% of wild-type) or none (<1% of wild-type). Mutants with reduced or none *in vitro* invasion were tested again (n=3) and the resulting mean and SD of all three experiments are represented in Figure 3.

Gene Name	Mutant	Putative function	BARBAKC	BH	BQ	BT	*In vitro* infection*
Adhesion/Invasion							
badA	05A04	adhesin	583_0314	01490 01510	01390 01400 01410	0168	normal
badA	70D12	adhesin	583_0314	01490 01510	01390 01400 01410	0168	normal
ialA/ialB	05D10	invasion associated gene B	583_0326	01650	01550	0181	reduced
ibaA	44H12	putative inducible autotransporter	583_1132	13160	10410	1655	normal
omp43	43G09	outer membrane protein, adhesin	583_0447	12500	09890	1902	normal
trwD	04B03	T4SS component, VirB11 homolog	-	15760	12680	2533	reduced
trwD	41C12	T4SS component, VirB11 homolog	-	15760	12680	2533	reduced
trwE	65D01	T4SS component, VirB10 homolog	-	15750	12670	2532	none
trwF	61B04	T4SS component, VirB9 homolog	-	15740	12660	2531	none
trwJ2	43H01	T4SS component, VirB5 homolog	-	15670 15700	12590 12620	2519 2522 2524 2526 2528	none
trwL1	25G12	T4SS component, VirB2 homolog	-	15570 15580 15590 12600 12610 12620 12630 12640	12490 12500 12510 12520 12530 12540 12550 12560	2511 2512 2513 2514 2515 2516 2516a	none
trwL2	60H02	T4SS component, VirB2 homolog	-	15570 15580 15590 12600 12610 12620	12490 12500 12510 12520 12530 12540	2511 2512 2513 2514 2515 2516	none

Gene Name	Mutant	Putative function	BARBAKC	BH	BQ	BT	*In vitro* infection*
				12630 12640	12550 12560	2516a	
virB4	61H02	T4SS component, VirB4 homolog	-	13280	10550	1691	normal
virD4	69D12	T4SS component, VirD4 homolog	-	13380	10640	1701	normal
virD4	79C12	T4SS component, VirD4 homolog	-	13380	10640	1701	normal
	61C01	autotransporter protein	-	13030	10290	1796	normal
Iron uptake							
hutA	5B10	outer membrane heme receptor	583_0460	04970	04160	0774	normal
Transport function							
ilvE	45B03	amino acid transporter	583_0747	10010	07730	1376	normal
livF	04A08	amino acid transporter	-	08250	06330	1144	normal
livG	25A02	amino acid transporter	-	08260	06320	1145	reduced
livH	45C07	ABC transporter	-	08280	06300	1147	normal
phaA	05G09	K+/H+ transmembrane protein	583_0030	16460	13360	2670	normal
Cell stress response							
ibpA (*hsp20*)	86B07	chaperon	583_0614	07300	05230	1333	normal
hslO (*hsp33*)	41A03	chaperon	583_1292	01080	00990	0118	normal
Metabolism / cell integrity							
carD	41C07	transcriptional regulator factor	583_0123	15240	12150	2444	normal
glnE	70D02	glutamate ammonialigase adenyl transferase	-	4800	04000	0707	normal
ftsK	15G10	cell division transmembrane protein	583_0291	03840	02850	0572	normal
cobS	44G10	cobalamin biosynthesis	583_0080	15880	12800	2554	normal
lpcC	69H08	lipopolysaccharide core biosynthesis mannosyltransferase	583_0983	11690	09300	0746	normal
mfd	41B10	transcription repair coupling factor	583_0798	08750	05840	1197	normal
Unknown function							
	43H05	unknown function	-	03150	-	0332 0505 1229 1275 1394 1812 2614	normal
BA0981	65D04	putative exported protein	-	02590	02450	0286	normal
BA1484	15A08	putativemembrane protein	583_1009	11960	09380	0713	normal
BA1559	35D02	helicase/methyltransferase	-	15450	-	0164 0455 0541 1006 1021 1035 1053 1080 1105 2491	normal

Gene Name	Mutant	Putative function	BARBAKC	BH	BQ	BT	*In vitro* infection*
BA1819	41C02	unknown function	-	09350	-	0466 1089 1090 2281 2282	normal
BA1566	61D04	unknown function	-	-	-	1926	normal
	05H01	conserved/putative (Tm helices) membrane protein	583_0492	05300	04480	0812	normal
	44G12	putative efflux transport protein	-	12560	-	1909	normal
Phage origin							
BA1301	86C10	putative anti-repressor protein	583_1070	06900 02990 03430 03240 03670 03690	-	0325 0355 0372 0373 0431 0470 0486 0494 0556 0557 0954 0976 2290 2301	normal
BA1052	60B07	Putative anti-repressor protein	-	02890 03020 03250 03440 0345 03460 03470	-	0475	normal
Intergenic region							
	83D04	intergenic	583_1301/ 583_1302	00970/ 00960	00900/ 0089	tRNA- BT000 1/BT0 082	normal
	69C09	intergenic, close to putative transcriptional regulator	583_1248	13590 14160 14370 14380 14970	02090	2389 2390 2397 2399 2400	normal
	04A01	intergenic	583_1132	13140 or 13160	10380	1660 or 1661	normal
	69B07	intergenic	583_0094/ 583_0093	15500/ 15510	12420/ 12430	2497/ 2504	normal
	86C05	intergenic	583_1019/ 583_1020	12050/ 12060	09460/ 09470	1641/ 1642	normal

3.4 Unpublished results

The VirB-homologous T4SS of *B. schoenbuchensis:* An evolutionary link between conjugation machineries and T4SSs adopted for host interaction

Introduction

The horizontal acquisition of the VirB type IV secretion system (T4SS) was shown to play a central role for the host adaptation of *Bartonella* species (see 3.1. Research article I). By translocating a versatile cocktail of different effector proteins, bartonellae are able to subtly modulate their cellular host niches which results in a highly specific adaptation (Schmid et al. 2004; Schulein et al. 2005). The ease by which evolution can alter these effector genes seemed to facilitate *Bartonella* species of different lineages to adapt to a wide range of different hosts (see 3.2. Research article II).

Interestingly, phylogenetic trees revealed two lineages of *Bartonella* which seem not to harbor the VirB T4SS. The deeply-branching lineage of *B. bacilliformis* has presumably diverged before the VirB T4SS was acquired by horizontal gene transfer. In a second phylogenetic lineage, strictly comprising ruminant-infecting species, PCR screening for conserved VirB T4SS genes, as *virB4* or v*irB11*, did neither detect any VirB T4SS. However, in all analyzed ruminant-infecting species, the screening revealed the presence of two genes with homology to *virB4* and *virB11*. By sequence analysis, these genes turned out to constitute orthologs of the VirB-homologous (Vbh) T4SS identified in *B. tribocorum* (see 3.1. Research article I). As the *virB4*-homologous gene carried a frame shift mutation, and any effector- or coupling protein-encoding genes were absent, the Vbh system of *B. tribocorum* was assumed not to be functional. However, the presence of the Vbh T4SS throughout the ruminant-infecting lineage, plus the fact that the highly related VirB T4SS is absent from these species, suggested the Vbh T4SS to functionally substitute the VirB T4SS in this lineage (Saenz et al. 2007).

Here, we assessed the coding content and function of the Vbh T4SS of the ruminant-infecting species *B. schoenbuchensis*.

Results

Sequence analysis of the *vbh* T4SS locus of *B. schoenbuchensis* indicates a bifunctional role for the Vbh T4SS

Starting from the gene sequences of *vbh4* and *vbh11*, the adjacently located chromosomal regions of *B. schoenbuchensis* were sequenced by primer walking on genomic DNA. This approach revealed the presence of a continuous *vbh* gene cluster in the genome of *B. schoenbuchensis.* Whole genome-shotgun sequencing of *B. schoenbuchensis* using 454-pyrosequencing confirmed the results from the primer walking approach (Liesch 2008).

In contrast to the *vbh* locus of *B. tribocorum*, manual sequence analysis and gene annotation of the *vbh* locus of *B. schoenbuchensis* did not reveal any frame shift mutations. In addition to the machinery-encoding genes *vbh2*-*vbh11*, a putative effector gene, a relaxase-encoding gene, and a gene with homology to the coupling protein-encoding *virD4* gene were identified (Fig. 1). These components are not present in the v*bh* locus of *B. tribocorum*. The putative effector gene of *B. schoenbuchensis* was identified by comparison to known *Bartonella* effector protein (*bep*) genes of the VirB T4SS. The analysis revealed that the effector gene of *B. schoenbuchensis* consists of an N-terminal FIC (filamentation-induced by cAMP) domain and a C-terminal BID (*Bartonella* intracellular delivery) domain, a domain structure typical found in Beps of the VirB T4SS (Fig.1). Interestingly, the BID domains of the putative effector gene and the relaxase gene of *B. schoenbuchensis* revealed a high degree of similarity.

Comparative analyses of the Vbh T4SS of *B. schoenbuchensis* and the closely related VirB T4SS of e.g. *B. tribocorum* revealed the genomic organization of the machinery-encoding genes (*vbh2*-*vbh11* and *virB2*-*virB11*) to be conserved (Fig. 1). However, the gene content of the downstream located effector region displays marked differences. A relaxase-encoding gene is absent from the *virB* T4SS loci of *Bartonella*. Furthermore, the single putative effector protein of the Vbh T4SS of *B. schoenbuchensis* is encoded on the reverse strand, whereas effector protein-encoding genes of the VirB T4SS are located on the forward strand (Fig. 1).

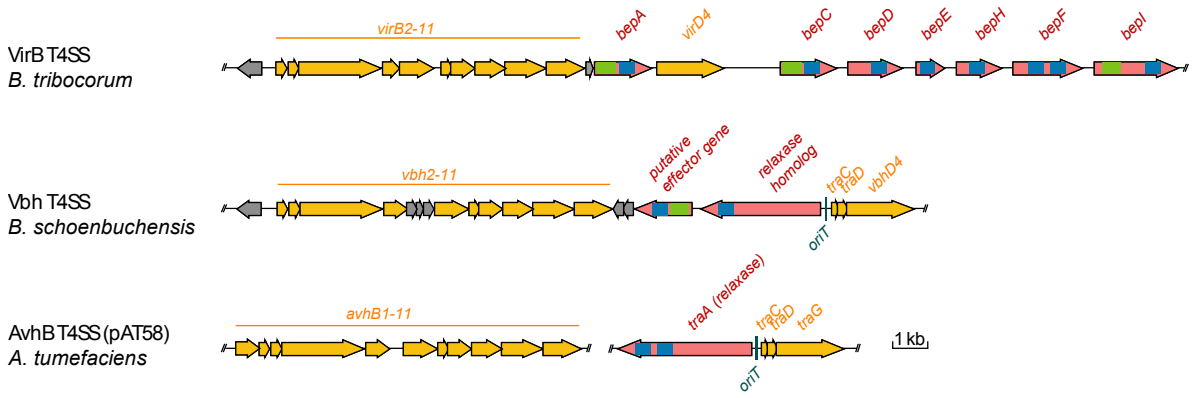


Figure 1: Comparison of the *virB* T4SS locus of *B. tribocorum*, the *vbh* T4SS locus of *B. schoenbuchensis*, and the *avhB* T4SS locus of *A. tumefaciens*. T4SS genes are shown in yellow and genes coding for translocated effector genes are shown in light red. Other genes are shown in gray color. Boxes in blue and green color indicate BID and FIC domains, respectively. In case of the conjugation system AvhB and the Vbh T4SS, the coupling protein-encoding gene (*virD4*) is named *traG* and *vbhD4*.

Comparison between the *vbh* T4SS locus of *B. schoenbuchensis* and gene clusters encoding related conjugation machineries, as for example the AvhB T4SS of *A. tumefaciens*, revealed striking similarities (Fig. 1). Beside the conservation of the machinery-encoding genes, as well as the relaxase and the coupling protein genes, additional genetic features essential for conjugation machineries are present in the *vbh* locus of *B. schoenbuchensis*. This included an intergenic sequence stretch resembling the *oriT* site, the origin of transfer, and two close-by encoded genes, *traC* and *traD* involved in the processing of the *oriT* during conjugative transfer.

The BID domain-containing C-terminus of the putative effector protein of *B. schoenbuchensis* is transferred by the VirB T4SS of *B. henselae* into Ea.hy926 cells

Here, it was tested whether the BID domain-containing C-terminus of the putative effector protein of *B. schoenbuchensis* can be transferred into eukaryotic cells, either directly by *B. schoenbuchensis* or via the VirB T4SS of *B. henselae*. To this end, we used the CRAfT reporter assay, which allows the quantification of protein transfer from bacteria into eukaryotic cells of the Cre-tester cell line Ea.hy926/pRS56-c#B1 (Schulein et al. 2005). A reporter plasmid pPE1001 was constructed harboring a

fusion protein of the BID domain-containing C-terminus of the putative effector gene to a nuclear localization signal (NLS)-Cre recombinase.

First, it was tested whether *B. schoenbuchensis* is capable of transferring this reporter construct into eukaryotic cells. When using standard conditions (5 days of infection, multiplicity of infection [MOI] of 200), no translocation of the reporter construct encoded on plasmid pPE1001 was detectable. Neither the variation of the infection dose (MOI of 50, 100, and 300) nor the elongation of the infection time (7 days of infection) resulted in detectable translocation events (data not shown).

To test translocation via the VirB T4SS of *B. henselae*, cells were infected with different *B. henselae* strains (wild-type and Δ*virB4)* carrying the plasmid pPE1001. These experiments demonstrated that the BID domain-containing C-terminus of the putative effector protein of *B. schoenbuchensis* is translocated by *B. henselae* in a VirB T4SS-dependent manner into eukaryotic cells (Fig 2).

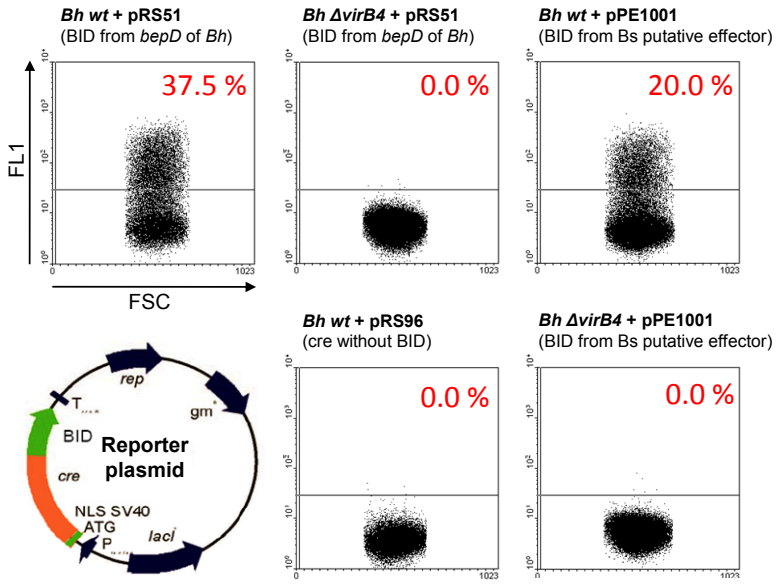


Figure 2: Protein transfer was determined by CRAfT. The Cre-tester cell line Ea.hy926/pRS56-c#B1 was infected with the indicated *B. henselae* (*Bh*) strains (wt or Δ*virB4*) expressing different NLS-Cre fusion proteins from plasmids (5 days of infection, MOI of 200). A schematic representation of the reporter construct plasmid depicts the gene encoding the NLS-Cre-BID fusion protein. GFP-positive cells (implicating the transfer of the NLS-Cre-BID fusion protein) were detected by FACS analysis. Dot plots of forward scatter (FSC) and GFP fluorescence (FL1) are shown for the indicated *Bh* strains. Percentages of GFP-positive cells are indicated. As positive control, plasmid pRS51 containing the BID domain-containing C-terminus of *B. henselae bepD* was used (Schulein et al. 2005). Plasmid pRS96 expressing the NLS-Cre protein without C-terminal BID domain served as negative control.
Bs, *B. schoenbuchensis.*

Further Results (generated by Marius Liesch)

The Vbh T4SS is plasmid-encoded and mediates conjugative transfer

Marius Liesch, a Master student under my supervision, continued to investigate the above-described aspects. The assembly of the *B. schoenbuchensis* genome indicated that the Vbh T4SS is encoded on a plasmid. The existence of this plasmid harboring the *vbh* T4SS locus could be demonstrated experimentally (Liesch 2008). By tagging the plasmid with a gentamicin resistance cassette, Marius Liesch further showed that the plasmid of *B.schoenbuchensis* is horizontally transferred between different *Bartonella* species. The disruption of the relaxase gene of the *vbh* locus by single cross-over recombination abolished the conjugative transfer of the plasmid between *B. schoenbuchensis* and recipient *Bartonella* strains (Liesch 2008). This implicated that the Vbh T4SS is directly involved in the conjugative transfer of the extrachromosomal replicon of *B. schoenbuchensis*.

B. schoenbuchensis is not able to translocate effector proteins into Ea.hy926 cells

Marius Liesch also investigated the role of the Vbh T4SS in host interaction. He extended the above-described CRAfT assays by changing various experimental parameters for the infections with *B. schoenbuchensis*. However, in none of the tested conditions, he could detect effector gene translocation from *B. schoenbuchensis* into Ea.hy926 cells. The fact that an increase of the MOI to more than 2000 did not lead to any cytotoxic effects suggested that there might be only poor interaction between *B. schoenbuchensis* and human endothelial cells (Liesch 2008). This implication was bolstered by the fact that *B. schoenbuchensis* bacteria expressing plasmid-encoded GFP did not show any obvious interaction with Ea.hy926 cells. Typical cellular phenotypes resulting from infections with *B. henselae*, like invasome formation, peri-nuclear localization, or bacterial aggregation, were not observed by microscopy (Liesch 2008).

Perspective

To show translocation by the Vbh T4SS, a new approach which was established by Marius Liesch before he left the laboratory was based on a β-lactamase reporter assay (Marketon et al. 2005). As this method can be used with any cell line, it has clear advantage over the CRAfT assay. Preliminary experiments infecting bovine cells with *B. schoenbuchensis* expressing the corresponding reporter construct did not show any significant levels of effector translocation. Still, this approach will be followed in the future by modifying the experimental conditions and testing different cell lines. To exclude that the Vbh T4SS is inactive, the gene expression of this T4SS in conditions of *in vitro* cell infections will be assessed.

Conclusions

The Vbh T4SS detected in ruminant-infecting *Bartonella* species was suggested to functionally replace the missing VirB T4SS (Saenz et al. 2007). Sequencing of the *vbh* T4SS locus of *B. schoenbuchensis* revealed that homologous genes of all VirB T4SS subunits are conserved in this ruminant-adapted species including the coupling protein-encoding gene. Furthermore, the detection of an adjacently located putative effector protein harboring an N-terminal FIC domain and a C-terminal BID domain, as found for many effector proteins of the VirB T4SS (Schulein et al. 2005), corroborated the assumption that the Vbh T4SS might display a host-interacting T4SS with similar function as the VirB. Although, this putative effector protein was translocated into host cells by *B. henselae* in a VirB T4SS-dependent manner, the direct translocation via the Vbh T4SS of *B. schoenbuchensis* could not be demonstrated yet. Remarkably, several features of the *vbh* T4SS locus of *B. schoenbuchensis* were found to be reminiscent of ancestrally related conjugation machineries like the presence of a relaxase gene or a putative origin of transfer. Genomic and functional investigation showed that the Vbh T4SS of *B. schoenbuchensis* is not only encoded on a plasmid, but also mediates the conjugative transfer of this plasmid between different *Bartonella* species.

Although conclusive experimental data for a bi-functional role is not yet available, the Vbh T4SS appears to display an evolutionary link between the VirB T4SS of *Bartonella,* solely dedicated to host interaction, and ancestrally related conjugation

machineries mediating horizontal transfer of plasmids among bacteria. Hence, the Vbh T4SS reveals insights into the evolutionary mechanisms resulting in the functional transformation of these nanomachines. The putative effector gene of the Vbh T4SS, for example, provides compelling evidence for the evolutionary origin of the effector genes of the VirB T4SS. It was already suggested by Schulein *et al*. (2005) that *bep* genes might have originated from relaxases by adopting their C-terminal translocation signal. The high degree of sequence identities shared by the BID domains of the putative effector gene and the adjacently located relaxase of *B. schoenbuchensis* provides the proof for the evolutionary origin of VirB T4SS effector genes. As relaxase genes frequently harbour two BID domains, one could have been split off by integration of foreign DNA sequence. By such a mechanism, the FIC domain was possibly introduced into the T4SS locus and coupled to the BID domain. The coupling of domains to an export signal was also found to be a prominent mechanisms in the evolutionary emergence of novel type III secretion effector proteins (Stavrinides et al. 2006).

Whether the effector gene of the Vbh T4SS of *B. schoenbuchensis* is already adapted for host interaction, or plays an unknown role in the process of conjugation will be revealed in the future.

Experimental procedures

Bacterial strains and Growth conditions. *Bartonella* spp. were grown on Columbia agar (CBA) plates containing 5% defibrinated sheep blood in a humified atmosphere at 35°C and 5% CO2 for 2-3 days (*B. henselae*) or 4-5 days (*B. schoenbuchensis*). Strain RSE247, a spontaneous streptomycin-resistant strain of ATCC 49882T (Schmid et al., 2004) served as *B. henselae* wild-type in this study. The typing strain R1 (CHDE252) was used as wild-type for *B. schoenbuchensis*. Media were supplemented with 30 µg/ml kanamycin, 100 µg/ml streptomycin, 12.5 µg/ml gentamicin, and/or 500 µM isopropyl β-D-thiogalactosidase (IPTG) if needed.

Sequencing on genomic DNA by primer walking. To isolate genomic DNA the QIAGEN Genomic-tip was used as recommended by the supplier (QIAGEN, Hilden, Germany). Bacteria were harvested from one to two plates, resuspended, and centrifuged for 5 minutes at 4000 rpm. DNA was frozen at -20°C or directly used for DNA extraction. For sequencing, 3 µg of genomic DNA was mixed with 0.5 µl of primer (10 pmol), 4 µl of BigDye1.1 reaction mix, 2 µl of BigDye1.1 5x Buffer, 1 µl Betaine, and ddH_2O added to a final volume of 18 µl. The thermocycler program used for amplification was as follows: 95°C 5min, 95°C 30 sec, 50°C 20 sec, 60°C 4 min, repeat step 2 to 4 99 times, store at 4°C. After linear DNA amplification, 2.2% SDS was added to each sequencing sample and incubated for 5 min at 95°C. The purification and precipitation of the DNA was done by adding 40 µl H_2O, 4 µl sodium-acetate (3 M) pH5.2, and 100 µl ethanol (95-100%). The samples were put into the -70°C freezer for 15 min to facilitate DNA precipitation. Thereafter, the DNA was pelleted in a 30 min centrifugation step at 13'000 rpm and the supernatant discarded. The pellet was washed twice with 70% EtOH and centrifuged for 5 min after each washing step. The precipitated and purified DNA was dried by a 10 min centrifugation in a Speedvac or by air-drying for 30 min in the dark. Thereafter, the samples were stored at 4°C in the dark until use. The sequencing was performed with the in-house facility (ABI Prism 7000 Sequence Detector and ABI 7000 Sequence Detection System 1.1 software from Applied Biosystem). The generated sequences were assembled with existing sequence information using VectorNTI software (Invitrogen).

Cloning of expression plasmid encoding the NLS-Cre-BID fusion protein. To construct plasmid pPE1001, the BID domain-containing C-terminus was amplified from genomic DNA of *B. schoenbuchensis* with primer pairs containing flanking *Xma*I/*Sal*I sites: prPE199 (ATGGTGTCGACAAACACAACAGGCACCATATC) and prPE201 (TCCCCCCCGGGTTACCTTGTAATTCCCTTTGAAG). PCR products were digested with enzymes *Xma*I and *Sal*I and ligated into the *Xma*I/*Sal*I sites of the backbone of plasmid pRS51 (Schulein et *al.* 2005). All plasmid DNA isolations and PCR purifications were performed with Macherey-Nagel and Promega columns according to manufacturer's instruction

Cell Culture. Ea.hy926/pRS56-c#B1 cells were grown in Dulbecco's Modified Eagle Media (DMEM,Gibco) with 10% fetal calf serum (FCS). Cells were grown in a humified atmosphere at 37°C and 5% CO2.

Experimental setup for CRAfT. One day prior infection, 20'000 Ea.hy926/pRS56-c#B1 cells (per ml and per well) were seeded into a 24-well plate. The next morning, the cells were washed once with pre-warmed (37°C) M199/10%. Then, 1 ml M199/10% FCS containing 500 μM IPTG was added. The cells were infected with variable MOIs of either *B. henselae* or *B. schoenbuchensis* strains carrying different reporter plasmids. After adding the bacteria to the cells, the 24-well plate was centrifuged for 5 min at 1'200 rpm (Biofuge stratos, Heraeus) and incubated at 35°C, 5% CO2. After 5-7 days of infection, the cell medium was removed and the cells were washed with 400 μl trypsine. Afterwards, the cells were incubated with 150 μl trypsine for 3 min at 35°C to detach them from the surface. The reaction was stopped with 450 μl M199 / 10% FCS. The suspension was mixed thoroughly by pipetting up and down and transferred into a FACS tube. 3×10^4 Ea.hy926/pRS56-c#B1 cells were analyzed by FACS-CaliburTM and data were exported for analysis to WinMDI or FlowJo software.

References

Liesch, M. (2008). Bartonella schoenbuchensis genome assembly and functional characterization of the VirB-homologous (Vbh) type IV secretion system. Biozentrum Basel. Basel, University of Basel. Master Thesis.

Marketon, M. M., R. W. DePaolo, et al. (2005). "Plague bacteria target immune cells during infection." Science 309(5741): 1739-1741.

Nystedt, B., A. C. Frank, et al. (2008). "Diversifying selection and concerted evolution of a type IV secretion system in Bartonella." Mol Biol Evol 25(2): 287-300.

Saenz, H. L., P. Engel, et al. (2007). "Genomic analysis of Bartonella identifies type IV secretion systems as host adaptability factors." Nat Genet 39(12): 1469-1476.

Schmid, M. C., R. Schulein, et al. (2004). "The VirB type IV secretion system of Bartonella henselae mediates invasion, proinflammatory activation and antiapoptotic protection of endothelial cells." Mol Microbiol 52(1): 81-92.

Schulein, R., P. Guye, et al. (2005). "A bipartite signal mediates the transfer of type IV secretion substrates of Bartonella henselae into human cells." Proc Natl Acad Sci U S A 102(3): 856-861.

Stavrinides, J., W. Ma, et al. (2006). "Terminal reassortment drives the quantum evolution of type III effectors in bacterial pathogens." PLoS Pathog 2(10): e104.

3.5 Review article

Genomics of host-restricted pathogens of the genus *Bartonella*

Philipp Engel and Christoph Dehio

Microbial Pathogenomics, Genome Dynamics, Karger, 2009, volume 6, 158-159.

3.5.1 Summary

The published genomes of *B. bacilliformis*, *B. henselae*, *B. quintana*, and *B. tribocorum* revealed interesting insights into evolution, host interaction, and ecology of the alphaproteobacterial pathogen *Bartonella*. In this review, we summarized published findings concerning general genomic features, genome dynamics on the inter- and intra-species level, as well as functional genomics of *Bartonella*. Beside the published data, we provided a comprehensive comparative analysis of the four *Bartonella* genome sequences highlighting aspects not yet covered by the primary research articles.

We used a combination of synteny plots (dot plots) and circular genome maps to display size differences among the four *Bartonella* genomes (1.44 – 2.62 Mb) and to indicate the conservation of a genomic backbone in the genus *Bartonella*. Their small syntenic core genome (~ 1000 genes) is reflecting the common facultative intracellular life-style of *Bartonella*. Still, the variation in genome size indicated expansion and reduction of genomic content in different lineages of *Bartonella*. We mapped the core genome genes on these circular genome representations showing that the accessory genome of *Bartonella* is organized in genomic islands of different dimensions. We further assembled a comprehensive list of all genomic islands identified in each of the four genomes and provided information about their gene content, dimensions and localization, as well as their presence or absence in other *Bartonella* genomes. This analysis depicted the highly dynamic genome evolution in the genus *Bartonella* reflecting their exceptional adaptability.

Statement of the own participation

This book chapter about the current state of *Bartonella* genomics was written by Prof. Christoph Dehio and me. Besides reviewing published data, I analyzed the four publicly available genomes of *Bartonella* for presence, absence, and structure of genomic islands and assembled the information in a comprehensive table. Additionally, I generated circular genome maps and dot plot representations of the four *Bartonella* genomes to allow direct comparison of genomic features. All figures of this review article were produced by me.

de Reuse H, Bereswill S (eds): Microbial Pathogenomics.
Genome Dyn. Basel, Karger, 2009, vol 6, pp 158–169

Genomics of Host-Restricted Pathogens of the Genus *Bartonella*

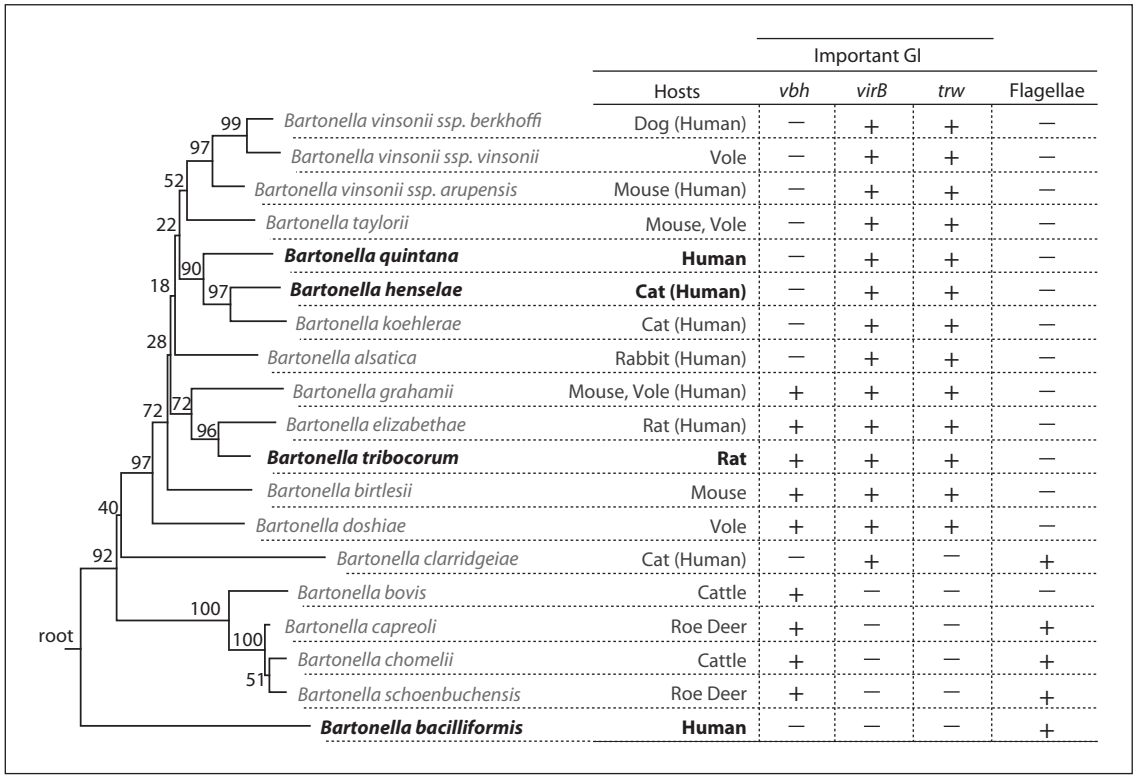
P. Engel · C. Dehio

Biozentrum, University of Basel, Basel, Switzerland

Abstract

The α-proteobacterial genus *Bartonella* comprises numerous arthropod-borne pathogens that share a common host-restricted life-style, which is characterized by long-lasting intraerythrocytic infections in their specific mammalian reservoirs and transmission by blood-sucking arthropods. Infection of an incidental host (e.g. humans by a zoonotic species) may cause disease in the absence of intraerythrocytic infection. The genome sequences of four *Bartonella* species are known, i.e. those of the human-specific pathogens *Bartonella bacilliformis* and *Bartonella quintana*, the feline-specific *Bartonella henselae* also causing incidental human infections, and the rat-specific species *Bartonella tribocorum*. The circular chromosomes of these bartonellae range in size from 1.44 Mb (encoding 1,283 genes) to 2.62 Mb (encoding 2,136 genes). They share a mostly synthenic core genome of 959 genes that features characteristics of a host-integrated metabolism. The diverse accessory genomes highlight dynamic genome evolution at the species level, ranging from significant genome expansion in *B. tribocorum* due to gene duplication and lateral acquisition of prophages and genomic islands (such as type IV secretion systems that adopted prominent roles in host adaptation and specificity) to massive secondary genome reduction in *B. quintana*. Moreover, analysis of natural populations of *B. henselae* revealed genomic rearrangements, deletions and amplifications, evidencing marked genome dynamics at the strain level.

Until the early 1990s, the genus *Bartonella* comprised a single species, *B. bacilliformis*. Since then, the reclassification of previously described bacteria based on 16S rRNA sequences (i.e., *Grahamella* and *Rochalimea*) and the description of novel *Bartonella* species isolated from various animal reservoirs resulted in a major expansion of the genus to currently 19 approved species, one of which *(Bartonella vinsonii)* is split into 3 subspecies. Among those, nine have been associated with human diseases (fig. 1) [1, 2]. The arthropod-borne bartonellae are widespread pathogens that colonize mammalian endothelial cells and erythrocytes as major target cells [3]. While endothelial cells and potentially other nucleated cells may get infected in both reservoir and incidental hosts, erythrocyte invasion takes place exclusively in the reservoir host,



		Important GI			
	Hosts	*vbh*	*virB*	*trw*	Flagellae
Bartonella vinsonii ssp. berkhoffi	Dog (Human)	−	+	+	−
Bartonella vinsonii ssp. vinsonii	Vole	−	+	+	−
Bartonella vinsonii ssp. arupensis	Mouse (Human)	−	+	+	−
Bartonella taylorii	Mouse, Vole	−	+	+	−
Bartonella quintana	**Human**	−	+	+	−
Bartonella henselae	**Cat (Human)**	−	+	+	−
Bartonella koehlerae	Cat (Human)	−	+	+	−
Bartonella alsatica	Rabbit (Human)	−	+	+	−
Bartonella grahamii	Mouse, Vole (Human)	+	+	+	−
Bartonella elizabethae	Rat (Human)	+	+	+	−
Bartonella tribocorum	**Rat**	+	+	+	−
Bartonella birtlesii	Mouse	+	+	+	−
Bartonella doshiae	Vole	+	+	+	−
Bartonella clarridgeiae	Cat (Human)	−	+	−	+
Bartonella bovis	Cattle	+	−	−	−
Bartonella capreoli	Roe Deer	+	−	−	+
Bartonella chomelii	Cattle	+	−	−	+
Bartonella schoenbuchensis	Roe Deer	+	−	−	+
Bartonella bacilliformis	**Human**	−	−	−	+

Fig. 1. Phylogeny and epidemiology of the genus *Bartonella*, distribution of important genomic islands (GI) encoding virulence factors, and presence/absence of flagella. For zoonotic species, man as an incidental host is indicated in brackets. Species with known genome sequences are highlighted in bold. The phylogenetic tree was calculated on the basis of protein sequences of *rpoB*, *groEL*, *ribC*, and *gltA* as described by [9]. Numbers at the nodes of the tree indicate bootstrap values for 1,000 replicates. Except for *Bartonella talpae* and *Bartonella peromysci,* for which no type strains exist, all approved species are included in the tree.

resulting in the establishment of a long-lasting intraerythrocytic bacteremia. Despite the fact that most *Bartonella* species are restricted to one reservoir host, there is an increasing body of evidence that some species can infect several different mammalian hosts [4–8]. The bartonellae represent an interesting model to study the evolution of host adaptation/host restriction as most mammals infested by blood-sucking arthropods serve as a reservoir host for at least one *Bartonella* species [9].

The highly virulent human-specific pathogen *B. bacilliformis* (causing life-threatening Oroya fever and verruga peruana) holds an isolated position in the *Bartonella* phylogeny as sole representative of an ancestral lineage. All other species evolved in a separate 'modern' lineage by radial speciation. These modern species represent host-adapted pathogens of rather limited virulence potential within their diverse mammalian reservoirs. Examples are the human-specific species *B. quintana* causing trench

Table 1. General features of *Bartonella* genome sequences. PCG, protein-coding genes; n.d., not determined. The coding content of *B. bacilliformis* and *B. tribocorum* were (re-)calculated by dividing the total length of all protein-coding genes and tRNA/rRNA coding regions by the chromosome length. In addition, the average length of PCG was calculated for *B. bacilliformis* by dividing the total length of all PCG by the number of PCG.

	B. bacilliformis	*B. tribocorum*	*B. henselae*	*B. quintana*
Chromosome size	1,445,021 bp	2,619,061 bp	1,931,047 bp	1,581,384 bp
G+C content	38.2%	38.8% (35.0%)[a]	38.2%	38.8%
Total number of PCG	1,283	2,136 (18)[a]	1,488	1,142
Average length of PCG	909 bp	906 bp	942 bp	999 bp
Integrase remnants	n.d.	47 (0)[a]	43	4
Number of rRNA operons	2	2 (0)[a]	2	2
Number of tRNA genes	44	42 (0)[a]	44	44
Percentage coding	81.6%	74.6% (69.8%)[a]	72.3%	72.7%
Plasmid	0	1 (23,343 bp)[a]	0	0

[a] Numbers in brackets refer to the plasmid

fever, the cat-adapted zoonotic pathogen *B. henselae* causing cat-scratch-disease and various other disease manifestations in the incidental human host, and the rat-specific pathogen *B. tribocorum* not yet associated with human infection (fig. 1). Over the last decade, the availability of animal and cell culture infection models in combination with powerful bacterial genetics has facilitated research aiming at understanding the cellular and molecular interactions that contribute to the complex relationship between *Bartonella* and its mammalian hosts [1–3]. More recently, *Bartonella* has entered the post-genomic era by the release of several complete genome sequences. Here, we summarize the comparative and functional genomic studies on *Bartonella* that have been reported to date.

General Features of *Bartonella* Genomes

Complete genome sequences are presently available for four *Bartonella* species, i.e., *B. henselae* and *B. quintana* [10], *B. tribocorum* [9], and *B. bacilliformis* (GenBank accession no. CP000525). Additionally, the genome composition of *Bartonella koehlerae* has been analyzed by comparative genomic hybridization profiling (CGH) based on the genome sequence of the closely related species *B. henselae* [11]. The four available *Bartonella* genomes are composed of single circular chromosomes (plus one plasmid in *B. tribocorum*), which display a uniformly low G+C content of 38.2% to 38.8%, and a noteworthy low coding density of 72.3% to 81.6% (table 1). The chromosome sizes range from 1,445 kb (encoding 1,283 genes) for *B. bacilliformis* to 2,619 kb (encoding

2,136 genes) for *B. tribocorum* (table 1, fig. 2). Orthologous gene assignments resulted in the identification of a core genome of 959 genes [9], which is encoded by a rather well conserved chromosomal backbone in a largely synthenic manner (fig. 2, see dotplots).

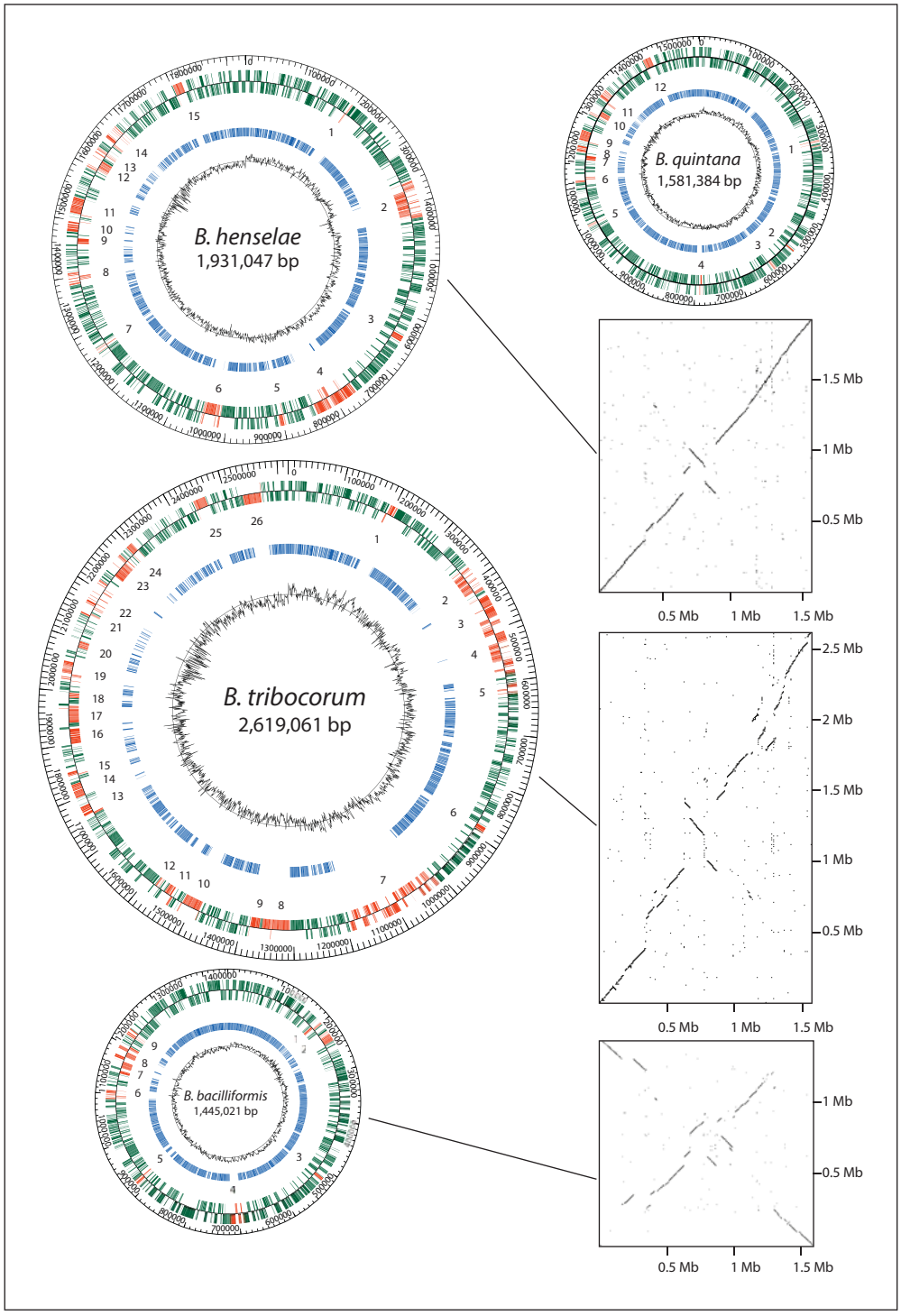
The relatively small core genome of the bartonellae reflects specific adaptations to the genus-specific lifestyle. For instance, a striking example of host-integrated metabolism is represented by hemin. This important source for iron and porphyrin is particularly abundant in the host niches colonized by bartonellae, i.e. the intracellular space of erythrocytes and the midgut lumen of blood-sucking arthropods. The strict hemin requirement for growth of *B. quintana* (and probably other bartonellae) in vitro correlates with the presence of multiple genes encoding hemin binding and hemin uptake proteins, while no hemin biosynthesis enzyme is encoded by this organism [10]. A large-scale mutagenesis screen in the *B. tribocorum*-rat model identified several of the hemin-uptake genes as essential for establishing intraerythrocytic infection. Moreover, this screen revealed that the majority of pathogenicity factors required for establishing intraerythrocytic bacteremia is encoded by the core genome inferred from the four available *Bartonella* genome sequences (66 of 97 pathogenicity genes) [9], indicating that this genus-specific infection strategy is to a large extent dependent on a conserved set of core genome-encoded pathogenicity factors.

Genome Dynamics by Lineage-Specific Expansion and Reduction

Despite of a largely synthenic core genome, the known *Bartonella* genomes are diversified by the variable size and composition of their accessory genomes. These were shaped in evolution by massive expansions (due to lateral gene transfer and gene duplication) and reductions (due to gene decay and deletion), which mostly occurred in a lineage-specific manner.

A marked example for genome reduction is *B. quintana*, which shares 1,106 orthologous genes with *B. henselae* as its closest relative (fig. 1). *B. henselae* codes for 382 genes without orthologs in *B. quintana*, while only 36 genes are unique to *B. quintana* [9, 10]. Interestingly, *Rickettsia prowazekii* representing another pathogen transmitted by the human body louse has also undergone recent genome reduction, suggesting that the extensive genome decay in the *B. quintana* lineage may be related to the biology of this arthropod vector [10]. However, *B. bacilliformis*, a pathogen vectored by the sandfly *Lutzomyia verrucarum*, displays also a remarkably small genome sequence, indicating that adaptation to humans could be accompanied by reductive genome evolution. Consistently, several of the more recently evolved human-specific pathogens display marked genome decay, e.g. *Salmonella typhi* and *Mycobacterium leprae* [12].

With an accessory genome exceeding the size of the core genome (1,195 vs. 959 genes), *B. tribocorum* represents a remarkable example of lineage-specific genome


B. henselae
1,931,047 bp
B. quintana
1,581,384 bp
B. tribocorum
2,619,061 bp
B. bacilliformis
1,445,021 bp
0.5 Mb
1 Mb
1.5 Mb
2 Mb
2.5 Mb

expansion, and to a lesser extent genome expansion is also evident in *B. henselae* (accessory genome of 529 genes). The primary source for these genome expansions are prophages and other laterally acquired genomic islands (GIs, table 2 and fig. 2). One phage-related GI is conserved in all known *Bartonella* genomes (table 2; BB-GI2 and homologs). *B. tribocorum* and *B. henselae* encode in addition large (>50 kb) prophage regions (table 2; BH-GI2, BT-GI2/4) that are homologous but highly plastic in their genetic organization [10]. These mosaic prophage regions and the related GIs encoding homologous phage genes were probably shaped during evolution by a consecutive acquisition of different prophages, followed by duplication, excision, reintegration, and reduction of prophage segments of different size and origin. Exclusively *B. tribocorum* encodes another large prophage (>30 kb) that, moreover, is present in multiple copies (table 2; BT-GI8/10/17/26). The different copies of this prophage display a strictly conserved gene order (fig. 3a) and a marked similarity to the genetic organization and sequence of P2- and Mu-like prophages described in other bacterial taxa. GIs encoding two-partner secretion systems, which often also carry phage genes, have also contributed to the large accessory genomes of *B. tribocorum* and *B. henselae* (table 2; BH-GI4/6, and BT-GI3/7/9/11). Remnants of these GIs are found in the reduced genome of *B. quintana*, while they are absent from the ancestral *B. bacilliformis* lineage and closely related α-proteobacterial taxa. A prototype of these GIs was thus likely acquired by the common ancestor of the modern *Bartonella* lineage, followed by lineage-specific expansions and reductions. At present it is unknown whether the prophages, phage-related GIs and GIs encoding two-partner secretion systems, that contributed to the remarkable genome expansion exemplified by *B. tribocorum* and *B. henselae*, have any beneficial role in host interaction, or whether these two species are just not under the selective pressure that resulted in massive genome reduction in *B. quintana*.

Some other GIs constituting the accessory genomes of the bartonellae are well established pathogenicity factors with important roles in the process of host colonization. Unlike *B. bacilliformis*, all species of the modern lineage encode at least one of the closely related type IV secretion systems (T4SSs) VirB/VirD4 or Vbh (VirB homolog) (fig. 1), which likely emanated from an ancestral duplication event and which are redundant in function. These VirB-like T4SSs are considered to represent major host adaptability factors that contributed to the remarkable evolutionary success of the modern lineage [9]. T4SSs are transporters ancestrally related to bacterial conjugation systems that mediate the vectorial translocation of virulence factors across the

Fig. 2. Circular genome maps of the four *Bartonella* genome sequences and Dot-plot representation of genome colinearity (micro-syntheny). The genome maps indicate (outside circles to inside circles) the genes on the + and – strands (genes located on genomic islands which are >5 kb or encoding more than five CDS are colored in red, all other genes in green), the genes belonging to the core genome (in blue), and the GC skew (black). Dot-plots were plotted for the *B. quintana* genome against any other genome for a sliding window of 20 nucleotides. Numbers in the genome circles refer to the different genomic islands (see also table 2).

Table 2. List of genomic islands (GIs) >10 kb of the four known *Bartonella* genomes. The first and last gene of each island is indicated by its locus tag (only the number of each locus tag is shown). The length refers to the start and end of the first and last gene of the island, respectively.

GI#	Similar genomic Islands	Description	tRNA	Begin	End	Length
B. bacilliformis						
BB-GI2	BT-GI20/23, BH-GI6/12, BQ-GI10	*Bartonella*-specific island encoding phage genes	yes	0217	0240	22115
BB-GI4		duplicated genomic region encoding housekeeping genes	no	0679	0710	26295
BB-GI5		conserved exported protein and transporter encoding genes	yes	0883	0894	10151
BB-GI6		conserved exported protein and phage genes	yes	1055	1080	17466
BB-GI8	BT-GI13, BH-GI10, BQ-GI8	flagella genes and inducible *Bartonella* autotransporter (*iba*) genes	yes	1116	1160	46499
BB-GI9		conserved exported protein and phage-related genes	no	1180	1190	12068
B. tribocorum						
BT-GI1		BT-specific helicase and phage-related genes	yes	0156	0167	15612
BT-GI2	BH-GI2, BQ-GI1	phage island	yes	0303	0377	51254
BT-GI3	BH-GI4/6	type II secretion system island	yes	0387	0422	44997
BT-GI4	BH-GI2/6	phage island	yes	0423	0564	110682
BT-GI5		BT-specific island encoding predicted membrane proteins	yes	0577	0596	17292
BT-GI6	BH-GI3, BQ-GI2, BB-GI3	putative membrane proteins not present in other alphaproteobacteria	no	0832	0834	11826
BT-GI7	BH-GI2/4/5/6	phage genes, type II secretion systems and helicase genes	yes	0941	1122	181527
BT-GI8		BT-specific phage island I	yes	1218	1283	53256
BT-GI9		BT-specific type II secretion systems and hypothetical genes	yes	1292	1301	18348
BT-GI10		BT-specific phage island II	yes	1382	1429	37682
BT-GI11	BH-GI4	type II secretion system island	no	1446	1464a	18888
BT-GI13	BH-GI10, BQ-GI8, BB-GI8	inducible *Bartonella* autotransporter (*iba*) genes	no	1650	1663	21879

Table 2. Continued

GI#	Similar genomic Islands	Description	tRNA	Begin	End	Length
BT-GI14	BH-GI11, BQ-GI9	VirB T4SS and *Bartonella* effector protein (Bep) genes	no	1689	1710	25598
BT-GI16	BH-GI9, BQ-GI7, BB-GI7	conserved *Bartonella*-specific autotransporter encoding genes	no	1785	1796	28492
BT-GI17		BT-specific phage island III	yes	1810	1849	32182
BT-GI19	BH-GI8, BQ-GI6	transporter-associated genes, and restriction system specific to BT	yes	1897	1930	35415
BT-GI20	BH-GI6/12, BQ-GI10, BB-GI2	*Bartonella*-specific island encoding phage genes	yes	1965	1983	12384
BT-GI22	BH-GI14, BQ-GI11, BB-GI1	*Bartonella*-specific island encoding *yopP* gene(s) in BQ and BT	yes	2113	2225	53002
BT-GI23	BH-GI6/12, BQ-GI10, BB-GI2	*Bartonella*-specific island encoding phage genes	yes	2263	2306	37989
BT-GI24		VirB-homologous (Vbh) T4SS	no	2331	2351	13874
BT-GI25	BH-GI15, BQ-GI12	Trw T4SS	no	2507	2533	22519
BT-GI26		BT-specific phage island IV	yes	2603	2646	35567
B. henselae						
BH-GI2	BT-GI2/4/7, BQ-GI1	phage island	yes	02730	03760	65723
BH-GI4	BT-GI3/7/11	type II secretion system island	yes	06500	07260	75441
BH-GI6	BT-GI3/4/7/20/23, BQ-GI10, BB-GI2	phage genes and type II secretion	yes	08980	09500	33315
BH-GI8	BT-GI19, BQ-GI6	transporter-associated genes	yes	12470	12600	20850
BH-GI10	BT-GI13, BQ-GI8, BB-GI8	inducible *Bartonella* autotransporter (*iba*) genes	no	13120	13190	19100
BH-GI11	BT-GI14, BQ-GI9	VirB T4SS and *Bartonella* effector protein (Bep) genes	no	13250	13440	28575
BH-GI12	BT-GI20/23, BQ-GI10, BB-GI2	*Bartonella*-specific island encoding phage genes	yes	13900	14090	21639
BH-GI14	BT-GI22, BQ-11, BB-GI1	*Bartonella*-specific island	yes	14450	14630	29125
BH-GI15	BT-GI25, BQ-GI12	Trw T4SS	no	15530	15760	16156
B. quintana						
BQ-GI1	BT-GI2/4, BH-GI2	Remnants of phage island present in BH and BT	yes	02600	02760	12764
BQ-GI6	BT-GI19, BH-GI8	Transporter-associated genes	yes	09850	09930	10161
BQ-GI8	BT-GI13, BH-GI10, BB-GI8	inducible *Bartonella* autotransporter (*iba*) genes	no	10360	10410	12121

Table 2. Continued

GI#	Similar genomic Islands	Description	tRNA	Begin	End	Length
BQ-GI9	BT-GI14, BH-GI11	VirB T4SS and *Bartonella* effector protein (Bep) genes	no	10510	10680	22110
BQ-GI10	BT-GI20/23, BH-GI6/12, BB-GI2	*Bartonella*-specific island encoding phage genes	yes	11020	11160	17399
BQ-GI11	BT-GI22, BH-GI14, BB-GI1	*Bartonella*-specific island encoding *yopP* gene(s) in BQ and BT	yes	11400	11630	20809
BQ-GI12	BT-GI25, BH-GI15	Trw T4SS	no	12450	12680	16587

two Gram-negative bacterial membranes and the host cell plasma membrane directly into the host cell cytoplasm [1]. The VirB/VirD4 T4SS of *B. henselae* was shown to translocate several effector proteins, termed Beps, into endothelial cells that subvert cellular functions, such as apoptosis and the inflammatory response, that are considered critical for establishing chronic infection [13–15]. The molecular mechanism by which VirB-like T4SSs mediate host adaptability is probably also dependent on the translocated Beps. Comparison of the *virB/virD4/bep* T4SS loci of *B. henselae*, *B. quintana* and *B. tribocorum* revealed that the *virB/virD4* genes encoding the 11 essential T4SS components are highly conserved, while the *bep* genes encoding the translocated Beps displayed a higher degree of sequence variation (fig. 3b), suggesting an increased rate of evolution as the result of positive selection for adaptive functions in the infected host [9].

A third T4SS, Trw, is present in a sub-branch of the modern lineage (fig. 1) and essential for the process of erythrocyte invasion [16]. Interestingly, the presence of Trw by the modern lineage correlates with the loss of flagella (fig. 1), which are required for the invasion of erythrocytes by *B. bacilliformis* and probably also the flagellated bacteria of the modern lineage [1]. Trw does not translocate any known effectors, but produces multiple variant pilus subunits due to tandem gene duplication and diversification (by combinatorial sequence shuffling and point mutations) of *trwL* (encoding the major pilus subunit TrwL) and *trwJ* (encoding the minor pilus-associated subunit TrwJ) (fig. 3c) [17]. The variant pilus subunits exposed on the bacterial surface are thought to facilitate the interaction with different erythrocyte receptors or blood group antigens, and may thus represent major determinants of host specificity [1].

Genome Dynamics on the Strain Level

Evidence for genome dynamics on the intra-species level is accumulating for different *Bartonella* species. To access the natural variation in gene content and genome

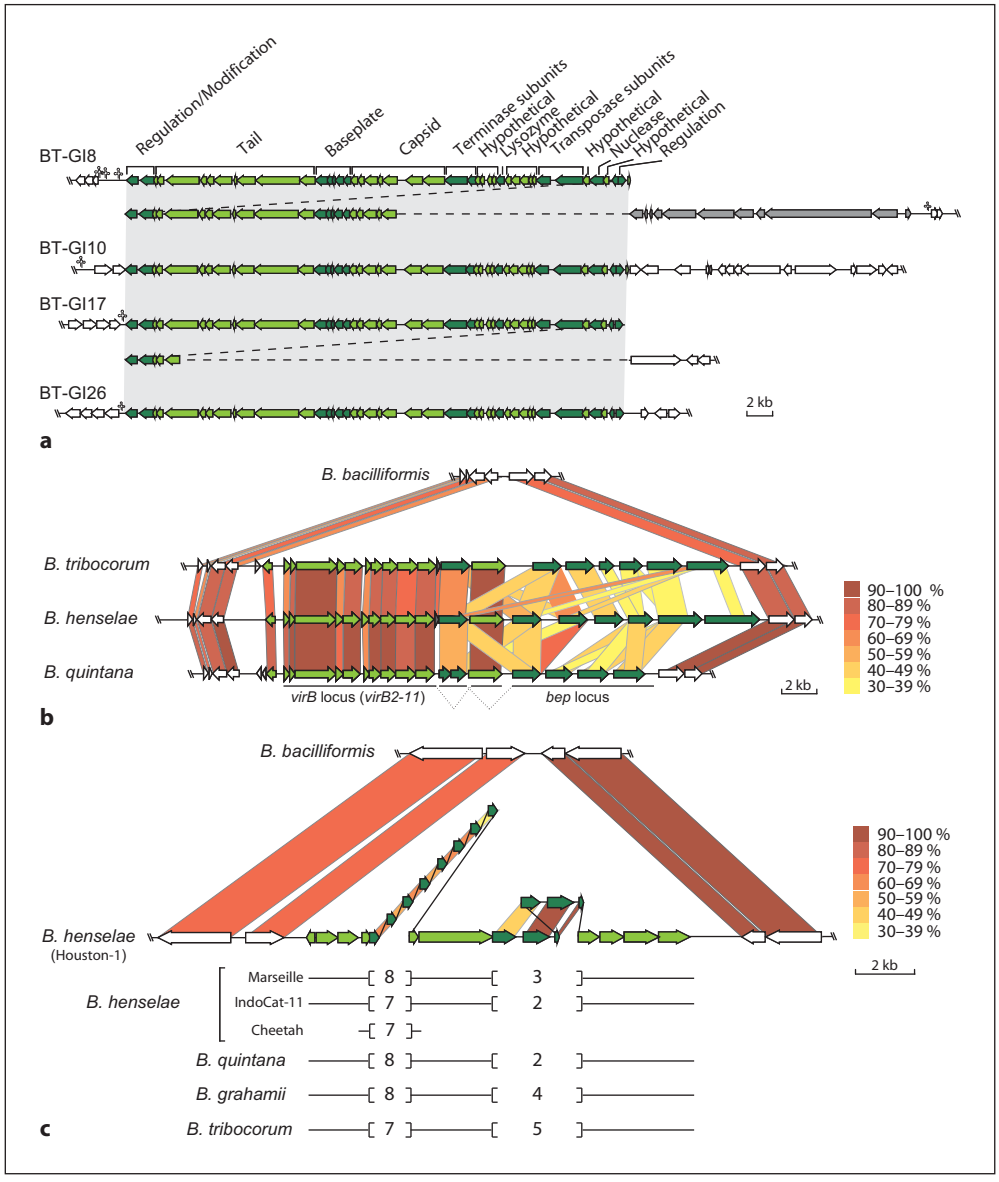


Fig. 3. Representation of selected GIs encoded in *Bartonella* genomes. Genes belonging to the GIs are shown in green, flanking genes are shown in white. (**a**) Alignment of the GIs encoding a *B. tribocorum*-specific prophage. Genes belonging to the prophage are located within the gray area. Noteworthy, BT-GI8 is flanked on one side by another island (gray gene symbols); (**b**) Alignment of the GI encoding the conserved T4SS VirB/VirD4 (*virB2–11* and *virD4* genes, colored in light green) and the highly variable translocated effectors (*bep* genes, colored in dark green); (**c**) Alignment of the GI (and flanking genes) encoding the T4SS-locus *trw*. The number of tandem repeats of *trwL* and *trwIJH* is indicated by gene symbols (colored in dark green) for the sequenced Houston-1 strain of *B. henselae* and by numbers in brackets for further *B. henselae* strains and the other species with known gene sequences. For (**b**) and (**c**), sequence similarity is shown with the percent identity indicated according to the color scales.

structure of *B. henselae*, a set of 38 strains isolated from cats and humans was analyzed by comparative genome hybridization [18]. The variation in gene content was modest and confined to the mosaic prophage region and other GIs, whereas extensive rearrangements were detected across the terminus of replication with breakpoints frequently locating to GIs. Moreover, in some strains a growth-phase dependent DNA-amplification was detected that centered at a putative phage replication initiation site located in a large plasticity region exemplified by a particularly low coding density [18]. Another study suggested that *B. henselae* exists as a mosaic of different genetic variants in the infected host [19]. Finally, genomic rearrangements due to gene deletions were elegantly demonstrated in serial isolates of *B. quintana* from an experimentally infected macaque [20]. Together, these data strongly suggest that various mechanisms contribute to a dynamic genome variation on the strain level.

Conclusions

Comparative and functional analysis of the four available complete genome sequences of species belonging to the genus *Bartonella* yielded first insights into the evolution, ecology and host interaction of this largely understudied group of bacterial pathogens. The small core genome reflects a host-integrated metabolism and codes for the majority of genes involved in the genus-specific infection strategy characterized by long-lasting intraerythrocytic infections in specific mammalian reservoir hosts. However, it is also evident that the accessory genomes contribute significantly to this infection strategy, e.g. flagella serving in the process of erythrocyte invasion by more ancestral species are considered to be functionally replaced by a laterally-acquired T4SS in more recently evolved species. Other laterally-acquired T4SSs were associated with the remarkable host adaptability exemplified by the radiating modern lineage. Genome expansion by lateral gene transfer in combination with secondary genome reduction has shaped the variable accessory genomes of the known *Bartonella* genomes. Additional *Bartonella* genome sequences expected to get available in the near future should result in a better understanding of the evolutionary processes that facilitated the emergence of a radiating group of host-restricted pathogens adapted to colonize a large variety of mammalian species that is infested by blood-sucking arthropods.

Acknowledgements

We are grateful to Arto Pulliainen for critically reading of the manuscript. The work was supported by grant 3100A0–109925/1 from the Swiss National Science Foundation (SNF), and grant 55005501 from the Howard Hughes Medical Institute (HHMI).

References

1 Dehio C: Infection-associated type IV secretion systems of *Bartonella* and their diverse roles in host cell interaction. Cell Microbiol 2008;10:1591–1598.
2 Dehio C: Molecular and cellular basis of *Bartonella* pathogenesis. Annu Rev Microbiol 2004;58:365–390.
3 Dehio C: *Bartonella*-host-cell interactions and vascular tumour formation. Nat Rev Microbiol 2005; 3:621–631.
4 Harms C, Maggi RG, Breitschwerdt EB, Clemons-Chevis CL, Solangi M, et al: *Bartonella* species detection in captive, stranded and free-ranging cetaceans. Vet Res 2008;39:59.
5 Jones SL, Maggi R, Shuler J, Alward A, Breitschwerdt EB: Detection of *Bartonella henselae* in the blood of 2 adult horses. J Vet Intern Med 2008;22:495–498.
6 Maggi RG, Harms CA, Hohn AA, Pabst DA, McLellan WA, et al: *Bartonella henselae* in porpoise blood. Emerg Infect Dis 2005;11:1894–1898.
7 Bown KJ, Bennet M, Begon M: Flea-borne *Bartonella grahamii* and *Bartonella taylorii* in bank voles. Emerg Infect Dis 2004;10:684–687.
8 Engbaek K, Lawson PA: Identification of *Bartonella* species in rodents, shrews and cats in Denmark: detection of two *B. henselae* variants, one in cats and the other in the long-tailed field mouse. Apmis 2004;112:336–341.
9 Saenz HL, Engel P, Stoeckli MC, Lanz C, Raddatz G, et al: Genomic analysis of *Bartonella* identifies type IV secretion systems as host adaptability factors. Nat Genet 2007;39:1469–1476.
10 Alsmark CM, Frank AC, Karlberg EO, Legault BA, Ardell DH, et al: The louse-borne human pathogen *Bartonella quintana* is a genomic derivative of the zoonotic agent *Bartonella henselae*. Proc Natl Acad Sci USA 2004;101:9716–9721.
11 Lindroos HL, Mira A, Repsilber D, Vinnere O, Naslund K, et al: Characterization of the genome composition of *Bartonella koehlerae* by microarray comparative genomic hybridization profiling. J Bacteriol 2005;187:6155–6165.
12 Pallen MJ, Wren BW: Bacterial pathogenomics. Nature 2007;449:835–842.
13 Schmid MC, Scheidegger F, Dehio M, Balmelle-Devaux N, Schulein R, et al: A translocated bacterial protein protects vascular endothelial cells from apoptosis. PLoS Pathog 2006;2:e115.
14 Schulein R, Guye P, Rhomberg TA, Schmid MC, Schroder G, et al: A bipartite signal mediates the transfer of type IV secretion substrates of *Bartonella henselae* into human cells. Proc Natl Acad Sci USA 2005;102:856–861.
15 Schmid MC, Schulein R, Dehio M, Denecker G, Carena I, Dehio C: The VirB type IV secretion system of *Bartonella henselae* mediates invasion, proinflammatory activation and antiapoptotic protection of endothelial cells. Mol Microbiol 2004;52:81–92.
16 Seubert A, Hiestand R, de la Cruz F, Dehio C: A bacterial conjugation machinery recruited for pathogenesis. Mol Microbiol 2003;49:1253–1266.
17 Nystedt B, Frank AC, Thollesson M, Andersson SG: Diversifying selection and concerted evolution of a type IV secretion system in *Bartonella*. Mol Biol Evol 2008;25:287–300.
18 Lindroos H, Vinnere O, Mira A, Repsilber D, Naslund K, Andersson SG: Genome rearrangements, deletions, and amplifications in the natural population of *Bartonella henselae*. J Bacteriol 2006;188: 7426–7439.
19 Berghoff J, Viezens J, Guptill L, Fabbi M, Arvand M: *Bartonella henselae* exists as a mosaic of different genetic variants in the infected host. Microbiology 2007;153:2045–2051.
20 Zhang P, Chomel BB, Schau MK, Goo JS, Droz S, et al: A family of variably expressed outer-membrane proteins (Vomp) mediates adhesion and autoaggregation in *Bartonella quintana*. Proc Natl Acad Sci USA 2004;101:13630–13635.

Christoph Dehio
Biozentrum, University of Basel
Klingelbergstrasse 70
CH–4056 Basel (Switzerland)
Tel. +41 61 267 2140, Fax +41 61 267 2118, E-Mail christoph.dehio@unibas.ch

4. Summary

4. Summary

Species of the alphaproteobacterial genus *Bartonella* present some stunning characteristics. They have adopted a unique infection strategy by persistently colonizing the bloodstream of various mammalian hosts. Whereas the presence of bacteria in the blood normally leads to severe disease manifestations, most *bartonellae* cause only mild or even asymptomatic infections, which indicates their perfect adaptation to the hosts. This remarkable infection strategy of *Bartonella* seems to be generally applicable, since the bloodstream of a wide range of different mammalian hosts can be colonized by *Bartonella*. Despite their adaptability, bartonellae exhibit a strong specificity, as a given species can only cause intraerythrocytic infection in one or few closely related mammalian hosts. By using genomic approaches, this work describes molecular features crucial for the adaptation of this unique infection strategy of *Bartonella* and assesses the underlying evolutionary processes.

Research article I presents a combinatorial study using functional and comparative genomics to identify colonization factors (i.e. pathogenicity genes) which could have contributed to the remarkable host adaptability of *Bartonella*. Our results show that most of the comprehensively identified colonization factors of *B. tribocorum* are conserved among different *Bartonella* species. This reflects the common infection strategy adopted by bartonellae to colonize their different mammalian hosts. Interestingly, we identified the VirB and Trw T4SS to be restricted to rapidly multiplying lineages comprising species causing generally benign infections in their various reservoir hosts. No T4SS was found in the genome of the highly virulent species *B. bacilliformis*. Comparative and phylogenetic analyses revealed that the VirB and Trw T4SSs were acquired by horizontal transfer after the divergence from the ancestral lineage of *B. bacilliformis*. By constituting molecular devices dedicated to the interaction with cellular host niches, T4SSs of *Bartonella* seem to be involved in the fine-tuning of the infection strategy resulting in increased host adaptability. This is reflected by the high degree of plasticity found among the genes encoding the VirB T4SS effector proteins known to subvert various cellular functions of the host. In this context, we propose that T4SSs display host adaptability factors facilitating virulence adjustment and the colonization of new hosts.

In **Research article II**, our aim was to further understand the evolutionary process resulting in the adaptation to different hosts. Thus, we investigated the parallel diversifying evolution of two *Bartonella* lineages which both harbor the VirB T4SS. Phylogenetic analysis revealed that the diversification of these two lineages resembled an evolutionary process known as adaptive radiation – i.e. the evolution of ecological and phenotypic diversity within a rapidly multiplying lineage. Based on our previous findings (3.1. Research article I) and a genome-wide natural selection analysis, the VirB T4SS seems to display a horizontally acquired key innovation which has independently triggered the adaptive radiation of the two lineages. Our phylogenetic and comparative analyses revealed that the VirB T4SS of the two lineages has been chromosomally integrated by two separate evolutionary events. Subsequently, two distinct VirB T4SS effector gene sets with functional similar characteristics have evolved in the two lineages by independent amplification and diversification processes. As evolutionary parallelism displays stark evidence for comparable, yet independent responses to similar selective pressures, our results imply that the lineage-specific evolution of the VirB T4SS and its effector proteins was driving the radiation of the two lineages. This is the first study demonstrating adaptive radiations for bacteria in their natural habitats, and moreover, assesses the underlying molecular mechanisms.

Research article III is shifting the focus on another aspect of host adaptation of *Bartonella* by addressing the specificity on the level of erythrocytes exhibited by this bacterial pathogen. With our *in vitro* model for erythrocyte adhesion and invasion, we demonstrated that the interaction with erythrocytes underlies a strict host-specificity in the genus *Bartonella*. When testing abacteremic transposon mutants of *B. birtlesii* in the *in vitro* model, the Trw T4SS was identified to be essential for erythrocyte colonization. The expression of the entire Trw T4SS of the rat-specific species *B. tribocorum* in *B. henselae* enabled this cat-adapted strain to colonize erythrocytes isolated from cats as well as from rats. This implicates a direct role for the Trw system in host specificity. Genomic inspection and evolutionary analysis of the *trw* loci showed that several pilus-components of the Trw T4SS are under diversifying selection in different *Bartonella* species. Hence, these components represent good candidates to confer host specificity of the Trw T4SS.

Complementary to the previously mentioned study on adaptive radiations, the chapter of **unpublished results** addresses the functional role of the VirB-homologous (Vbh) T4SS in a phylogenetic lineage of *Bartonella* strictly comprising of ruminant-infecting species. The Vbh T4SS was suggested to functionally replace the VirB T4SS which is missing in this sub-lineage of *Bartonella*. Sequencing of the entire *vbh* locus of *B. schoenbuchensis*, a representative of that lineage, indicated a bi-functional role for this T4SS. Besides the presence of an effector-encoding gene similar to the ones of the VirB T4SS, the *vbh* locus of *B. schoenbuchensis* also harbors genetic features reminiscent of a conjugation machinery. Genomic and functional investigation showed that the Vbh T4SS of *B. schoenbuchensis* is encoded on a plasmid and mediates the conjugative transfer of this plasmid between different *Bartonella* species. Attempts to demonstrate Vbh T4SS-mediated translocation of the identified effector gene into eukaryotic cells did so far not lead to any conclusive results. However, with its ambiguous characteristics, the Vbh T4SS appears to constitute an evolutionary snapshot of the adaptive transformation of a conjugation machinery into a host-interacting protein secretion system.

Finally, to provide an overview of *Bartonella* genomics, our **Review article** summarizes commonalities and differences of the publicly available genomes of *B. bacilliformis, B. henselae, B. quintana,* and *B. tribocorum*. It highlights the variation of genome size ranging from 1.44 Mb to 2.62 Mb and addresses the accessory genome content of *Bartonella*. Dot plot analyses and comparison of the coding content of the four genomes show that bartonellae share a small, syntenic core genome (~ 1000 genes). The variation in gene content can mostly be attributed to lineage-specific acquisition and reduction of various genomic islands of different appearance. The highly dynamic genome evolution seems to reflect the great adaptability of the genus *Bartonella*.

5. Discussion

5. Discussion

In this study, I used genomic approaches to understand the molecular basis of the infection strategy of *Bartonella* and to infer the underlying evolutionary process of host adaptation. Since my thesis included sequencing and analysis of several *Bartonella* genomes, I will first discuss some global findings in regard to *Bartonella* genomics and reflect on the life-style of this bacterial pathogen. Further, I will focus on more specific results of my PhD and integrate different aspects of the research articles and unpublished results to obtain a comprehensive picture for the molecular evolution of *Bartonella* by adaptive radiations.

5.1 Genomic overview of *Bartonella* and its lifestyle

In 2004, *Bartonella* entered the post-genomic era when the genomes of the zoonotic species *B. henselae* and the human pathogen *B. quintana* had been released *(Alsmark et al. 2004)*. During my thesis, we sequenced the genomes of six additional *Bartonella* strains: *B. tribocorum, B. clarridgeiae, B. rochalimae, B. sp. 1-1C, B. sp. AR 15-3, and B. schoenbuchensis* (3.1 and 3.2 Research article I and II). Four of these six genomes consist of draft sequences only (*B. rochalimae, B. sp. 1-1C, B. sp. AR 15-3, and B. schoenbuchensis*). However, the advanced state of the assemblies and the automated gene annotations allowed comparative analyses of these draft genomes with the completed sequences. Together with genomes published by other research groups (*B. henselae*, *B. quintana* (Alsmark et al. 2004), *B. grahamii* (Berglund et al. 2009)*, and B. bacilliformis* (CP000524)), the existing sequence data covers species from all major lineages of *Bartonella* and thus, provides a comprehensive overview of the genome dynamics in the genus *Bartonella.*

In general, bartonellae harbor small genomes with a size of 1.4-2.6 Mb, a GC content of 35% - 39% and a low coding content (70 - 80 %). These are typical characteristics also found among genomes of other vector-borne intracellular

alphaproteobacteria, as *Rickettsia*, *Wolbachia*, or *Anaplasma* (Boussau et al. 2004). Hence, genomic evolution in the genus *Bartonella* consistently follows the common trend towards small genomes of intracellular bacteria confirming the strong correlation between environmental selection pressures and the evolutionary processes shaping genomic structures (Sallstrom and Andersson 2005; Casadevall 2008).

As *Bartonella* species share a common lifestyle in protected host niches, it is not surprising that their small core genome encodes most of the pathogenicity factors identified by the signature-tagged mutagenesis of *B. tribocorum* (3.1 Research article I). However, it is rather remarkable that some *Bartonella* species, despite their common infection strategy and the general trend towards gene loss, harbor large accessory gene pools resulting in the observed variability in genome size (3.5 Review article). These differences could reflect the adaptation to different blood-sucking arthropods used by *Bartonella* to be transmitted between hosts. As arthropods are colonized by heterogeneous bacterial populations (Dillon and Dillon 2004; Tonetti et al. 2009), they display a potential source for HGT most likely responsible for the lineage-specific expansions of the genetic content in *Bartonella*. A substantial fraction of this accessory gene content might be specifically dedicated to the survival and replication within the distinct arthropod hosts. However, as the adaptation to narrow niches results in gene loss, it needs to be considered that not only the number of different arthropod vectors, but also the host range of these vectors, as well as the bacterial population size might influence the evolution of genomic structures in *Bartonella* (Felsenstein 1974; Moran 1996; Moran and Plague 2004). In this line, it was proposed that *B. quintana* might have reduced its genomic content to the core genome as a consequence of the adaptation to a single host-restricted vector, the human body louse (Page et al. 1998; Alsmark et al. 2004).

For a substantial fraction of the accessory gene pool of *Bartonella* no function could yet be assigned. This might implicate the selfish nature of these DNA segments, which litter the genomes of *Bartonella*. Several prophage regions identified in the genome of *B. tribocorum* might display such elements, as they encode typical phage components needed for a selfish lifestyle in a bacterial population. Also, toxin-antitoxin-like genes and restriction-modification systems found in the genomes of *Bartonella* are evidence for the selfish nature of a substantial

fraction of the accessory genetic content (Kobayashi 2001; Van Melderen and Saavedra De Bast 2009). However, for *B. grahamii,* a bacteriophage was identified which not only packaged phage DNA into viral particles but also large fractions of genomic regions harboring genes important for host colonization. It was proposed that this so called gene transfer agent might allow an increased genetic exchange of genes involved in infectious processes (Berglund et al. 2009). These findings go in line with the fact that the packaged region displays the highest degree of gene content variability indicating that this mechanism might substantially shape the size of the accessory genomes of different *Bartonella* lineages.

Despite the significant differences in genome size and the high abundance of repeats in the accessory gene pools, the organization of the genetic backbone encoding the core genome appears to be rather conserved across different *Bartonella* lineages. This is astonishing, as in other bacterial species the presence of repeats resulted in massive rearrangements of their genomic backbone (Mira et al. 2002). The high degree of conservation implies that the chromosomal organization of *Bartonella* is under strong selection pressure, and the genomic plasticity is restricted to genomic hot spots. This compatibility of genomic conservation and plasticity might provide the framework of the evolutionary success of this bacterial pathogen.

5.2 Adaptive radiations in *Bartonella*

Since species of the genus *Bartonella* are specifically adapted to colonize the bloodstream of a wide array of different mammals, they display a great model to study molecular and evolutionary aspects of host adaptation. The availability of genome sequences offered new approaches for assessing the genetic basis and the molecular evolution of host adaptation in this bacterial pathogen. The genomic approach of this study inferred that type IV secretion systems (T4SSs), acquired at different time points of evolution, enabled *Bartonella* to successfully adapt its characteristic infection strategy for the colonization of a wide array of different mammals. In particular, these are the VirB, the Trw, and most likely also the VirB-homologous (Vbh) T4SS (3.1 Research article I). At least two lineages of *Bartonella* (here referred to as lineage 3 and lineage 4, see 3.2 Research article II) harbor the VirB T4SS and appear to be exceptionally successful in adapting to various hosts, as they comprise most of the extant *Bartonella* species. In both lineages species have independently adapted to the same reservoir hosts. In research article II, we showed that evolution of these two lineages follows an evolutionary process known as adaptive radiation. This process describes the rapid evolution of a multitude of species from a single ancestor by the adaptation to divergent niches (Schluter 2000) which in case of *Bartonella* represent the mammalian hosts. This process can either be driven by a key innovation, the availability of unoccupied niches (i.e. ecological opportunities), or a combination of these two factors. Hallmark examples in the Metazoan kingdom are Darwin's finches (Grant and Grant 2008) on the Galapagos Islands, cichlid fishes in the Great East African lakes (Salzburger 2009), or the Carribean Anolis lizards (Butler et al. 2007). Although adaptive radiations of rapidly multiplying bacteria have been directly observed in evolution experiments (Rainey and Travisano 1998; MacLean 2005; Blount et al. 2008; Kassen 2009), no compelling example of adaptive radiation has been described for bacteria in their natural habitats. Reasons for this may be the difficulty to consistently define bacterial species boundaries with respect to phenotypic and ecotypic characteristics, as well as the varying degree of genetic diversity (Fraser et al. 2009) across different bacterial genera. Herein, the ecology of *Bartonella* seems to provide the appropriate framework which allows us to understand mechanisms of adaptive radiations of bacteria in their natural habitats for the first time.

5.2.1 The VirB system: a key innovation driving parallel adaptive radiations

Our data provides strong evidence that the VirB T4SS displays a key innovation for the adaptive radiations of *Bartonella*. We have found the VirB T4SS to be exclusively present in the two radiating lineages 3 and 4 of *Bartonella*, where it has presumably been acquired in a common ancestor. T4SS are ancestrally related to plasmid-encoded conjugation machineries mediating the horizontal transfer of these extra-chromosomal replicons among bacteria (Schulein and Dehio 2002; Frank et al. 2005). Thus, it is tempting to speculate that the common ancestor of the two radiating lineages acquired the VirB T4SS as a conjugation system encoded on a conjugative plasmid. This scenario becomes even more likely by the fact that the highly related VirB-homologous (Vbh) T4SS of *B. schoenbuchensis* and *B. grahamii* is encoded on a plasmid and at least in *B. schoenbuchensis* is capable of mediating conjugative transfer (Berglund et al. 2009). Further, our comparative and phylogenetic analyses comparing the VirB T4SS of the two radiating lineages revealed a separate but highly similar evolution of the VirB T4SS. This included the independent chromosomal integration of the T4SS genes and the lineage-specific emergence of translocated effector alleles. These independently evolved similar functional features in the two lineages, as for example tyrosine-phosphorylation motifs. These parallel evolutionary processes preceded the adaptive radiations in both lineages implicating that the functional versatility of the VirB T4SS had been readily evolved before the lineages started to diversify. Evolutionary parallelisms, as identified on the ecological as well as the molecular level for *Bartonella,* are characteristic for adaptive radiations. They display compelling evidence for comparable, but independent, responses to similar selective conditions in different lineages (Salzburger 2009). Therefore, the correlation between the diversification of the VirB T4SS and the occurrence of adaptive radiation in two independent lineages provides strong evidence that this T4SS displays a key innovation for the radiation of the two lineages.

The more recent onset of the radiation in lineage 3 compared to lineage 4 might suggest that the integration of the VirB T4SS and the diversification of the effector genes happened at different time-points of evolution in these two lineages (3.2 research article II). The genetic differences which we have found between the VirB T4SS loci of the two lineages seem to reflect this. In the genomes of lineage 3,

the T4SS gene cluster is present in two or even three copies. Further, we have found more duplicated effector genes in lineage 3 compared to lineage 4. It was argued that gene amplification processes display the initial adaptive responses to selection pressures resulting in an increase of the gene dosage of important factors (Sandegren and Andersson 2009). This is explained by the much higher frequency of gene duplications over specific point mutations (as for example promoter up mutations) in a bacterial population (Bergthorsson et al. 2007). The multiple copies of the VirB T4SS and effector genes present in the chromosomes of lineage 3 might confer such a gene dosage effect. However, gene amplifications are highly instable mutations and display only transient solutions to a selective problem until more stable adaptive mutations occur (Andersson and Hughes 2009). This might explain why one of the three VirB loci was partially deleted again in a common ancestor of *B. rochalimae* and *B. sp*. 1-1C. The same seems to hold true for some effector genes of lineage 3, as gene remnants at given genomic integration sites provide evidence for their secondary deletion in some of the analyzed species.

By harboring only one locus encoding the T4SS and effector genes, the genomic organization of the VirB system in lineage 4 seems to be more stream-lined likely reflecting its advanced evolutionary state in the species of this older radiation. This is also supported by the finding that the VirB T4SS of *B. henselae*, a species of lineage 4, was put under the regulatory control of the vertically inherited two-component system BatR/BatS which belongs to the alphaproteobacterial core genome. Apparently, after the HGT and chromosomal integration, appropriate promoter sequences had evolved upstream of the VirB T4SS and effector genes (unpublished results, Maxime Quebatte). The control of the expression of the VirB T4SS of lineage 3 is unknown and would be an interesting aspect to study in the future.

While gene dosage effects are supposed to constitute the initial force driving gene amplification, secondary adaptive mutations often lead to the emergence of functionally new variants (Francino 2005; Bergthorsson et al. 2007). This is exactly what we observe for the effector genes of lineage 3 and 4. The duplicated effector alleles have diversified by the accumulation of adaptive point mutations, as measured by positive selection analysis, as well as by recombination resulting in effector variants with altered domain structures. Particularly in lineage 3, where many effector

genes harbor the ancestral domain architecture, a substantial fraction of the adaptive mutations might also reflect a co-evolutionary arms-race between the pathogen and the host immune-system. In the plant pathogenic species complex *Pseudomonas syringae*, it was shown that the T3SS effector gene YopJ appears in various allelic forms which have evolved under strong positive selection (Ma et al. 2006; McCann and Guttman 2008). These variants have retained the same biochemical function but show very different patterns of recognition in different hosts, thereby also conferring host-specificity.

For *Bartonella*, however, experimental studies have demonstrated that many of the analyzed effector variants exhibit distinct phenotypic properties in host cells indicative for their diverse roles in host subversion (Schmid et al. 2006; Rhomberg et al. 2009; Scheidegger et al. 2009; Selbach et al. 2009). But how can a single ancestral effector gene give rise to such a variety of distinct functions? The explanation seems to lie in the functional adaptability of the different domains and motifs which these effector genes are built on. The tandem-repeated phosphorylation motifs as well as the FIC domains encoded by many effector genes are supposed to mediate post-translational modifications of a diversity of host cell proteins (Roy and Mukherjee 2009; Selbach et al. 2009; Yarbrough et al. 2009). The adaptive evolution seems to allow these different effector variants to target distinct host cell functions resulting in a highly specific modulation of the cellular niches. Remarkably, also the BID domain, being part of the translocation signal of effector genes, seems to be capable of adopting functions in the host cell (Schulein et al. 2005). In case of BepA, it was shown that the BID domain is sufficient to mediate the anti-apoptotic property of this effector gene (Schmid et al. 2006). BID domains of other effectors with the same domain constitution as BepA do not exhibit this phenotype indicating the specific adaptive modulation of the translocation signal in BepA. The large reservoirs of biologically distinct functions, which apparently can be invented from a single ancestor-gene by amplification-diversification processes, provide the framework for the adaptive evolution of *Bartonella* in different ecological niches and ultimately result in the radiation of lineages harboring the VirB T4SS.

Although the VirB T4SS and its effector genes are essential for colonization of the intraerythrocytic niche, the exact cell types targeted by this system *in vivo* remain unknown. However, there is evidence that the VirB T4SS fulfills its function at an

early stage of the infection process before the bacteria get seeded into the bloodstream (Schulein and Dehio 2002). Particularly, endothelial cells are likely to be targeted *in vivo* by *Bartonella* and the VirB T4SS, because these cells line the interior surface of blood vessels where the bacterial seeding could take place from (Dehio 2005). Several *Bartonella* species show a strong tropism for endothelial cells (Brouqui and Raoult 1996; Verma et al. 2000; Schmid et al. 2004). *In vitro* studies of *B. henselae* infecting endothelial cells revealed various phenotypes dependent on the T4SS and its effector genes which get translocated into these cells (Schmid et al. 2004; Schulein et al. 2005; Schmid et al. 2006; Rhomberg et al. 2009; Scheidegger et al. 2009). However, it is very likely that also other cell types, particularly those of the immune system, are targeted by the VirB T4SS, as the modulation of the host adaptive immune response represents the key towards a persistent infection (Batut et al. 2004); and also, the innate immune response displays one of the first barrier towards the colonization of new hosts (Albiger et al. 2007). Cell types of the immune system known to be targeted by *Bartonella* are, for instance, macrophages (Musso et al. 2001; Schweyer and Fayyazi 2002; Kyme et al. 2005).

Interestingly, recent data shows that *B. henselae* is translocating effector proteins into dendritic cells implicating these tissue-located phagocytes as potential first targets of *Bartonella* at the site of vector-infection (unpublished results, Rusudan Okujava). The action of the VirB T4SS at an early stage of infection and the ability to modulate the host-specific immune response would go in line with the adaptive capabilities of this nanomachinery. Future studies aiming at the identification of the primary cellular niches *in vivo* would help to understand the molecular action of the VirB T4SS allowing *Bartonella* to reach the bloodstream. In this context, it would be of particular interest to study commonalities and differences in the way how *Bartonella* species of the two radiating lineages with their different effector gene sets are subverting the cellular niches of their hosts.

5.2.2 The Trw system: a lineage-specific key innovation

The fact that the two lineages 3 and 4 have undergone separate adaptive radiations raises the question whether lineage-specific factors could have contributed to these independent evolutionary processes. The Trw T4SS might display such a lineage-specific factor. It is exclusively found in lineage 4, where it was acquired by HGT (Seubert et al. 2003). Together with the VirB, the Trw T4SS belongs to the few essential host colonization factors absent in the highly virulent species *B. bacilliformis*. Thus, the Trw T4SS was suggested to contribute to the observed host adaptability in the corresponding lineage of *Bartonella*. Further, our natural selection analysis showed that different components of the Trw T4SS have substantially evolved by positive selection suggesting a role in the adaptive process of *Bartonella*.

On the basis of *in vivo* data (Seubert et al. 2003), the Trw T4SS was hypothesized to be involved in erythrocyte binding and invasion; however, the direct interaction with erythrocytes was never shown. By using *in vitro* cell assays, we have demonstrated for the first time the direct role of the Trw T4SS in erythrocyte adherence and invasion (3.3 Research article III). Further, our *in vitro* assay fully reproduced the host specificity of *Bartonella* observed *in vivo*. The ectopic expression of the Trw T4SS of rat-specific *B. tribocorum* in cat-specific *B. henselae* expanded the host range for erythrocyte infection from cat to rat, thus indicating the Trw T4SS to mediate host-specific erythrocyte infection. Since niche specificity is a direct consequence of adaptive evolution to divergent niches (Cohan and Koeppel 2008), these finding are strongly supporting the adaptive potential of the Trw T4SS. In case of plant pathogenic bacteria, as *Pseudomonas syringae* or *Pantoea agglomerans,* Type III secretion effector genes were shown to have evolved by adaptive evolution, thereby conferring the host-specificity of different strains (Nissan et al. 2006; McCann and Guttman 2008; Almeida et al. 2009). Also, adaptive evolution detected in Type III secretion effector genes of *Salmonella* was supposed to mediate host specificity (Eswarappa et al. 2008; Eswarappa et al. 2009). Most *trw* T4SS genes, which were found to be under positive selection, are encoding different pilus subunits. Interestingly, these genes have been amplified and diversified several times during evolution resulting in the presence of up to eight paralogous gene copies in a given *Bartonella* species. It was hypothesized that these diverse pilus variants might confer

the binding to surface molecules of erythrocytes, as e.g. glycoproteins. Mammalian erythrocytes are known to have highly variable glycoprotein surface structures, both within and between individual host species (Gagneux and Varki 1999). Further, they have been suggested to evolve by diversifying selection (Baum et al. 2002; Altheide et al. 2006). Therefore, the selection for a diverse arsenal of bacterial binding components appears likely to explain the amplification and diversification of the pilus components of the Trw T4SS. Additionally, the selection pressure imposed by the host immune system may contribute to the observed rapid evolution of these pilus proteins (Nystedt et al. 2008). Although the Trw T4SS confers host specificity, the rapid diversifying evolution of the pilus subunits might facilitate spontaneous host switches, which eventually lead to the evolution of pathogens with altered host specificity. Hence, the Trw T4SS appears likely to contribute to the adaptive radiation of lineage 4.

The adoption of improved molecular traits to colonize a given niche often results in the loss of formerly essential factors. This can be due to the incompatibility of these factors with the novel lifestyle of the bacterium or to the replacing function of the novel trait (Forterre 1999; Maurelli 2007). Intriguingly, we find a marked correlation between the presence of the Trw T4SS and the deletion of the flagella loci in different lineages of *Bartonella*. Flagella are supposed to have been vertically inherited in the alphaproteobacterial subdivision (Boussau et al. 2004). For the ancestral species *B. bacilliformis*, flagella are required for erythrocyte invasion (Scherer et al. 1993). In other *Bartonella* species lacking the Trw T4SS, genomic loci encoding the flagella genes were detected and corresponding surface structures identified by electron microscopy (Kordick et al. 1997; Dehio et al. 2001). Thus, it is tempting to speculate that the horizontally acquired Trw T4SS has functionally replaced the vertically inherited flagellae in the radiating lineage 4. This makes sense as flagellin, the major structural component of flagella, is a potent activator of the innate and adaptive immune response (Honko and Mizel 2005; Salazar-Gonzalez and McSorley 2005).

So far it is not known, whether flagellae directly bind to the erythrocyte surface and confer host specificity, or whether they only support the adherence and invasion process by mobilizing the bacteria to the erythrocytes. However, the availability of an erythrocyte invasion and adherence assay allows addressing these questions in the

future. Particularly, the comparison of species adapted to the same host but either harboring the Trw T4SS or flagellae could broaden our understanding of the functional commonalities and differences of these two traits.

The step-wise acquisition of two different T4SS, Trw and VirB, both substantially fine-tuning the molecular strategy of *Bartonella* to colonize the host bloodstream, might explain why the adaptive radiation in lineage 4 constitutes the most species-rich clade of *Bartonella*.

5.2.3 Ecological opportunities

Besides the acquisition of key innovations, the driving forces for an adaptive radiation also include ecological opportunities, which is the availability of unoccupied niches where competition with other organisms is absent (Schluter 2000). For the adaptive radiations of the Darwin's finches and the cichlid fishes, the Galapagos Islands and the Great Lakes of East Africa display ecological opportunities (Grant and Grant 2008; Salzburger 2009).

In case of *Bartonella*, the blood stream of different mammalian hosts obviously represents the equivalents of such unoccupied niches as it is seldom colonized by other microorganisms (Brooks et al. 1991). However, an important aspect to consider is to what extent different host niches have been available for *Bartonella* during evolution. If we assume that the ancestors of different *Bartonella* lineages were already adapted to arthropod-borne transmission routes, the ecological opportunities are strictly dependent on the host-range of such vectors. Whereas some tick species can feed on a variety of different hosts (Sonenshine and Mather 1994) other arthropods, known to be vectors of *Bartonella,* exhibit a rather strict host-specificity (Krasnov et al. 2006). Although our knowledge about the nature of *Bartonella* vectors is currently limited, ecological opportunities to colonize new niches might not have been readily available for *Bartonella* during evolution. This appears to be reflected by the phylogenetic branching pattern of the radiating lineages 3 and 4 of *Bartonella*. The internal branches at the onset of the radiations are not exclusively short compared to the tip branches (i.e. branches of contemporary taxa) indicating that the

acquisition and diversification of the VirB T4SS did not result in an immediate but rather continuous branching of the two lineages. Short internal branches and long tip branches are typically found, when an ancestor arrives in an environment with plenty of unoccupied niches available resulting in a burst of new species over a rather short period of time until all niches are occupied (Gavrilets and Vose 2005). In the adaptive radiation of the cichlid fishes, the evolutionary young Great Lakes of East Africa presented such “still-to-be-filled” niches, as they were offering diverse habitats along their shores (Salzburger 2009). Over a short period of time, the cichlid fishes arriving in these lakes diversified by adaptation to the distinct ecological niches resulting in the evolution of enormous species flocks (Verheyen et al. 2003). Thus, these readily available niches result in the rapid multiplication of lineages over a limited time of evolution which slows down again, when new niches become rare.

As shown by our phylogenetic analysis, the process of radiation in the *Bartonella* lineages seems not to have been limited to a certain time window in evolution implicating that the process is still ongoing. This is supported by a body of evidence. First, several lineage splits have occurred recently in evolution, as for example the adaptive diversification of *Bartonella rochalimae* and *B.sp*. 1-1C, whereas others appear to be rather old. Second, host niches can be colonized more than once, since several species of different lineages have adapted to same reservoir hosts. Apparently, niche competition does not exist, which is also substantiated by several reports of co-infections with two different *Bartonella* species, in particular with *B. henselae* and *B. clarridgeiae* in cats (Gurfield et al. 1997; Kosoy et al. 1997; Gurfield et al. 2001; Maruyama et al. 2001). Third, although *Bartonella* exhibits host-specificity, some species are able to explore potentially new hosts. This becomes evident by (i) the zoonotic potential of *Bartonella* species, like *B. henselae or B. rochalimae*, and (ii) the capability of particularly rodent-adapted species to colonize several related hosts.

Taken extrinsic factors into account, the adaptive radiations of *Bartonella* may not only be driven by the acquisition of given key innovations, but also by the availability of ecological opportunities. As shown for other radiations, the combination of the availability of unoccupied niches and the readiness of a particular species to occupy them seems to be the key for the occurrence of adaptive radiations (Schluter 2000; Salzburger 2009).

5.2.4 The Vbh system in the ruminant-infecting lineage

Phylogenetic analysis of the genus *Bartonella* consistently revealed the clustering of four closely related ruminant-infecting species. They constituted a sister-clade of the two radiating lineages 3 and 4 and the ancestral lineage of *B. bacilliformis* (see 3.2 Research article II). Further, our analyses revealed that they evolved from a deep-branching lineage. The close relationship among these species might indicate a recent adaptive radiation to different hosts of the ruminant suborder. However, as the different species of this lineage are known to have an overlapping host range (Chang et al. 2000; Dehio et al. 2001; Rolain et al. 2003), their genetic diversification does not display an adaptive process to distinct ruminant niches.

The fact that this lineage is deep-branching and strictly comprises ruminants-infecting species suggests that the adaptation to this suborder of mammals might have prevented the colonization of other potential hosts. This is supported by our findings that neither of the two adaptive T4SS, VirB or Trw, was identified in this sub-lineage of *Bartonella*. However, in Saenz et *al.* (2007), we argued that the presence of a VirB-homologous T4SS likely substitutes for the function of the lacking VirB system in the ruminant-infecting species (3.1 Research article I). These two T4SS share a high degree of sequence similarity indicating close evolutionary relatedness. Further, the *vbh* locus was found to encode a hypothetical effector protein consisting of the same two functional domains as many effector genes of the VirB T4SS. However, subsequent shotgun sequencing of *B. schoenbuchensis* revealed this gene to be the only effector identified in the genome of this ruminant-adapted species. While the adaptive function of the VirB T4SS is conferred by an arsenal of divergent effector proteins, the existence of a sole effector might indicate a limited potential of the Vbh T4SS to distinctively modulate cellular host niches. On one side, this could explain the restriction to colonize other hosts than ruminants. On the other side, it suggests that virulence attenuation and adaptation to this specific suborder of mammals might have included other factors than T4SSs.

Closer inspection of the *vbh* locus of *B. schoenbuchensis* by genomic sequencing revealed characteristics typical for related conjugation machineries (3.4 Unpublished results): (i) in the genome of *B. schoenbuchensis,* the *vbh* T4SS was found to be located on an extrachomosomal replicon and (ii) the downstream region of the *vbh*

locus encoded conserved genetic determinants essential for conjugative transfer of plasmids (e.g. relaxase gene and *oriT*). Our two-parental mating experiments confirmed the Vbh T4SS to mediate conjugative plasmid transfer between different *Bartonella* species. However, the likely bi-functional role of this T4SS remains hypothetical as our experimental approaches did so far not succeed to detect direct effector protein translocation into mammalian cells via the Vbh T4SS. Still, the genomic locus of the Vbh T4SS seems to offer a snapshot of evolution depicting the transformation of a conjugation machinery into a host-interacting T4SS, as the VirB system is constituting. The sequence similarity between the BID domains of the Vbh effector gene and the adjacently located relaxase provides the proof for the previously suggested evolutionary origin of the translocation signal of the VirB effector proteins (Schulein et al. 2005). The coupling of an enzymatic function to this translocation signal in the form of the FIC domain seems to have armed the T4SS with a prototypical effector protein. As FIC domains are evolutionary conserved from bacteria to eukaryotes and mediate post-translational modifications in cell signaling, they display suitable functional modules to couple to the secretion signal (Kinch et al. 2009; Yarbrough and Orth 2009). This is not only the case for *Bartonella*. Many other pathogenic bacteria seem to use FIC domains to interfere with their hosts cellular functions, though there are using different secretion systems (Roy and Mukherjee 2009; Worby et al. 2009; Yarbrough et al. 2009).

Further, the Vbh T4SS reveals insights into the chromosomal integration mechanism of the VirB T4SS. The recently published genome of *B. grahamii* seems to harbor two Vbh T4SS copies, one integrated into the chromosome and another one located on the plasmid. Interestingly, a site-specific recombinase gene is present on the plasmid of *B. grahamii.* This is a typical characteristic of integrative and conjugative elements (Burrus and Waldor 2004) and might display the mechanism by which the T4SS has been inserted into the chromosome. A similar mechanism appears likely to have resulted in the independent chromosomal fixation of the VirB T4SS in lineage 3 and 4. The conjugative function of a plasmid-encoded T4SS is supposedly under strong selection, as it confers the maintenance of the extra chromosomal replicon in the bacterial population (Frank et al. 2005). This functional constraint is released as soon as the T4SS locus is transferred into the chromosome. Hence, the chromosomal fixation might display the crucial step towards the complete adoption of a new functionality, as observed for the VirB T4SS in lineage 3 and 4.

5.3 References

Albiger, B., S. Dahlberg, et al. (2007). "Role of the innate immune system in host defence against bacterial infections: focus on the Toll-like receptors." J Intern Med 261(6): 511-528.

Almeida, N. F., S. Yan, et al. (2009). "A draft genome sequence of Pseudomonas syringae pv. tomato T1 reveals a type III effector repertoire significantly divergent from that of Pseudomonas syringae pv. tomato DC3000." Mol Plant Microbe Interact 22(1): 52-62.

Alsmark, C. M., A. C. Frank, et al. (2004). "The louse-borne human pathogen Bartonella quintana is a genomic derivative of the zoonotic agent Bartonella henselae." Proc Natl Acad Sci U S A 101(26): 9716-9721.

Altheide, T. K., T. Hayakawa, et al. (2006). "System-wide genomic and biochemical comparisons of sialic acid biology among primates and rodents: Evidence for two modes of rapid evolution." J Biol Chem 281(35): 25689-25702.

Andersson, D. I. and D. Hughes (2009). "Gene Amplification and Adaptive Evolution in Bacteria." Annu Rev Genet.

Batut, J., S. G. Andersson, et al. (2004). "The evolution of chronic infection strategies in the alpha-proteobacteria." Nat Rev Microbiol 2(12): 933-945.

Baum, J., R. H. Ward, et al. (2002). "Natural selection on the erythrocyte surface." Mol Biol Evol 19(3): 223-229.

Berglund, E. C., A. C. Frank, 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(7): e1000546.

Bergthorsson, U., D. I. Andersson, et al. (2007). "Ohno's dilemma: evolution of new genes under continuous selection." Proc Natl Acad Sci U S A 104(43): 17004-17009.

Blount, Z. D., C. Z. Borland, et al. (2008). "Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli." Proc Natl Acad Sci U S A 105(23): 7899-7906.

Boussau, B., E. O. Karlberg, et al. (2004). "Computational inference of scenarios for alpha-proteobacterial genome evolution." Proc Natl Acad Sci U S A 101(26): 9722-9727.

Brooks, G. F., J. S. Butel, et al. (1991). Jawetz, Melnick & Adelberg's medical microbiology. Norwalk, Conn., Appleton & Lange: v.

Brouqui, P. and D. Raoult (1996). "Bartonella quintana invades and multiplies within endothelial cells in vitro and in vivo and forms intracellular blebs." Res Microbiol 147(9): 719-731.

Burrus, V. and M. K. Waldor (2004). "Shaping bacterial genomes with integrative and conjugative elements." Res Microbiol 155(5): 376-386.

Butler, M. A., S. A. Sawyer, et al. (2007). "Sexual dimorphism and adaptive radiation in Anolis lizards." Nature 447(7141): 202-205.

Casadevall, A. (2008). "Evolution of intracellular pathogens." Annu Rev Microbiol 62: 19-33.

Chang, C. C., B. B. Chomel, et al. (2000). "Bartonella spp. isolated from wild and domestic ruminants in North America." Emerg Infect Dis 6(3): 306-311.

Cohan, F. M. and A. F. Koeppel (2008). "The origins of ecological diversity in prokaryotes." Curr Biol 18(21): R1024-1034.

Dehio, C. (2005). "Bartonella-host-cell interactions and vascular tumour formation." Nat Rev Microbiol 3(8): 621-631.

Dehio, C., C. Lanz, et al. (2001). "Bartonella schoenbuchii sp. nov., isolated from the blood of wild roe deer." Int J Syst Evol Microbiol 51(Pt 4): 1557-1565.

Dillon, R. J. and V. M. Dillon (2004). "The gut bacteria of insects: nonpathogenic interactions." Annu Rev Entomol 49: 71-92.

Eswarappa, S. M., J. Janice, et al. (2009). "Host-specificity of Salmonella enterica serovar Gallinarum: insights from comparative genomics." Infect Genet Evol 9(4): 468-473.

Eswarappa, S. M., J. Janice, et al. (2008). "Differentially evolved genes of Salmonella pathogenicity islands: insights into the mechanism of host specificity in Salmonella." PLoS One 3(12): e3829.

Felsenstein, J. (1974). "The evolutionary advantage of recombination." Genetics 78(2): 737-756.

Forterre, P. (1999). "Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins." Mol Microbiol 33(3): 457-465.

Francino, M. P. (2005). "An adaptive radiation model for the origin of new gene functions." Nat Genet 37(6): 573-577.

Frank, A. C., C. M. Alsmark, et al. (2005). "Functional divergence and horizontal transfer of type IV secretion systems." Mol Biol Evol 22(5): 1325-1336.

Fraser, C., E. J. Alm, et al. (2009). "The bacterial species challenge: making sense of genetic and ecological diversity." Science 323(5915): 741-746.

Gagneux, P. and A. Varki (1999). "Evolutionary considerations in relating oligosaccharide diversity to biological function." Glycobiology 9(8): 747-755.

Gavrilets, S. and A. Vose (2005). "Dynamic patterns of adaptive radiation." Proc Natl Acad Sci U S A 102(50): 18040-18045.

Grant, P. R. and B. R. Grant (2008). How and why species multiply : the radiation of Darwin's finches. Princeton, Princeton University Press.

Gurfield, A. N., H. J. Boulouis, et al. (1997). "Coinfection with Bartonella clarridgeiae and Bartonella henselae and with different Bartonella henselae strains in domestic cats." J Clin Microbiol 35(8): 2120-2123.

Gurfield, A. N., H. J. Boulouis, et al. (2001). "Epidemiology of Bartonella infection in domestic cats in France." Vet Microbiol 80(2): 185-198.

Honko, A. N. and S. B. Mizel (2005). "Effects of flagellin on innate and adaptive immunity." Immunol Res 33(1): 83-101.

Kassen, R. (2009). "Toward a general theory of adaptive radiation: insights from microbial experimental evolution." Ann N Y Acad Sci 1168: 3-22.

Kinch, L. N., M. L. Yarbrough, et al. (2009). "Fido, a novel AMPylation domain common to fic, doc, and AvrB." PLoS One 4(6): e5818.

Kobayashi, I. (2001). "Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution." Nucleic Acids Res 29(18): 3742-3756.

Kordick, D. L., E. J. Hilyard, et al. (1997). "Bartonella clarridgeiae, a newly recognized zoonotic pathogen causing inoculation papules, fever, and lymphadenopathy (cat scratch disease)." J Clin Microbiol 35(7): 1813-1818.

Kosoy, M. Y., R. L. Regnery, et al. (1997). "Distribution, diversity, and host specificity of Bartonella in rodents from the Southeastern United States." Am J Trop Med Hyg 57(5): 578-588.

Krasnov, B. R., Morand, S., et al. (2006). "Resource predictability and host specificity in fleas: the effect of host body mass." Parasitology 133(1): 81-86.

Kyme, P. A., A. Haas, et al. (2005). "Unusual trafficking pattern of Bartonella henselae -containing vacuoles in macrophages and endothelial cells." Cell Microbiol 7(7): 1019-1034.

Ma, W., F. F. Dong, et al. (2006). "Type III effector diversification via both pathoadaptation and horizontal transfer in response to a coevolutionary arms race." PLoS Genet 2(12): e209.

MacLean, R. C. (2005). "Adaptive radiation in microbial microcosms." J Evol Biol 18(6): 1376-1386.

Maruyama, S., T. Sakai, et al. (2001). "Prevalence of Bartonella species and 16s rRNA gene types of Bartonella henselae from domestic cats in Thailand." Am J Trop Med Hyg 65(6): 783-787.

Maurelli, A. T. (2007). "Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens." FEMS Microbiol Lett 267(1): 1-8.

McCann, H. C. and D. S. Guttman (2008). "Evolution of the type III secretion system and its effectors in plant-microbe interactions." New Phytol 177(1): 33-47.

Mira, A., L. Klasson, et al. (2002). "Microbial genome evolution: sources of variability." Curr Opin Microbiol 5(5): 506-512.

Moran, N. A. (1996). "Accelerated evolution and Muller's rachet in endosymbiotic bacteria." Proc Natl Acad Sci U S A 93(7): 2873-2878.

Moran, N. A. and G. R. Plague (2004). "Genomic changes following host restriction in bacteria." Curr Opin Genet Dev 14(6): 627-633.

Musso, T., R. Badolato, et al. (2001). "Interaction of Bartonella henselae with the murine macrophage cell line J774: infection and proinflammatory response." Infect Immun 69(10): 5974-5980.

Nissan, G., S. Manulis-Sasson, et al. (2006). "The type III effectors HsvG and HsvB of gall-forming Pantoea agglomerans determine host specificity and function as transcriptional activators." Mol Microbiol 61(5): 1118-1131.

Nystedt, B., A. C. Frank, et al. (2008). "Diversifying selection and concerted evolution of a type IV secretion system in Bartonella." Mol Biol Evol 25(2): 287-300.

Page, R. D., P. L. Lee, et al. (1998). "A different tempo of mitochondrial DNA evolution in birds and their parasitic lice." Mol Phylogenet Evol 9(2): 276-293.

Rainey, P. B. and M. Travisano (1998). "Adaptive radiation in a heterogeneous environment." Nature 394(6688): 69-72.

Rhomberg, T. A., M. C. Truttmann, et al. (2009). "A translocated protein of Bartonella henselae interferes with endocytic uptake of individual bacteria and triggers uptake of large bacterial aggregates via the invasome." Cell Microbiol 11(6): 927-945.

Rolain, J. M., E. Rousset, et al. (2003). "Bartonella schoenbuchensis isolated from the blood of a French cow." Ann N Y Acad Sci 990: 236-238.

Roy, C. R. and S. Mukherjee (2009). "Bacterial FIC Proteins AMP Up Infection." Sci Signal 2(62): pe14.

Salazar-Gonzalez, R. M. and S. J. McSorley (2005). "Salmonella flagellin, a microbial target of the innate and adaptive immune system." Immunol Lett 101(2): 117-122.

Sallstrom, B. and S. G. Andersson (2005). "Genome reduction in the alpha-Proteobacteria." Curr Opin Microbiol 8(5): 579-585.

Salzburger, W. (2009). "The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes." Mol Ecol 18(2): 169-185.

Sandegren, L. and D. I. Andersson (2009). "Bacterial gene amplification: implications for the evolution of antibiotic resistance." Nat Rev Microbiol 7(8): 578-588.

Scheidegger, F., Y. Ellner, et al. (2009). "Distinct activities of Bartonella henselae type IV secretion effector proteins modulate capillary-like sprout formation." Cell Microbiol 11(7): 1088-1101.

Scherer, D. C., I. DeBuron-Connors, et al. (1993). "Characterization of Bartonella bacilliformis flagella and effect of antiflagellin antibodies on invasion of human erythrocytes." Infect Immun 61(12): 4962-4971.

Schluter, D. (2000). The ecology of adaptive radiation. Oxford, Oxford University Press.

Schmid, M. C., F. Scheidegger, et al. (2006). "A translocated bacterial protein protects vascular endothelial cells from apoptosis." PLoS Pathog 2(11): e115.

Schmid, M. C., R. Schulein, et al. (2004). "The VirB type IV secretion system of Bartonella henselae mediates invasion, proinflammatory activation and antiapoptotic protection of endothelial cells." Mol Microbiol 52(1): 81-92.

Schulein, R. and C. Dehio (2002). "The VirB/VirD4 type IV secretion system of Bartonella is essential for establishing intraerythrocytic infection." Mol Microbiol 46(4): 1053-1067.

Schulein, R., P. Guye, et al. (2005). "A bipartite signal mediates the transfer of type IV secretion substrates of Bartonella henselae into human cells." Proc Natl Acad Sci U S A 102(3): 856-861.

Schweyer, S. and A. Fayyazi (2002). "Activation and apoptosis of macrophages in cat scratch disease." J Pathol 198(4): 534-540.

Selbach, M., F. E. Paul, et al. (2009). "Host cell interactome of tyrosine-phosphorylated bacterial proteins." Cell Host Microbe 5(4): 397-403.

Seubert, A., R. Hiestand, et al. (2003). "A bacterial conjugation machinery recruited for pathogenesis." Mol Microbiol 49(5): 1253-1266.

Sonenshine, D. E. and T. N. Mather (1994). Ecological dynamics of tick-borne zoonoses. New York ; Oxford, Oxford University Press.

Tonetti, N., M. Berggoetz, et al. (2009). "Ticks and tick-borne pathogens from wildlife in the Free State Province, South Africa." J Wildl Dis 45(2): 437-446.

Van Melderen, L. and M. Saavedra De Bast (2009). "Bacterial toxin-antitoxin systems: more than selfish entities?" PLoS Genet 5(3): e1000437.

Verheyen, E., W. Salzburger, et al. (2003). "Origin of the superflock of cichlid fishes from Lake Victoria, East Africa." Science 300(5617): 325-329.

Verma, A., G. E. Davis, et al. (2000). "Infection of human endothelial cells with Bartonella bacilliformis is dependent on Rho and results in activation of Rho." Infect Immun 68(10): 5960-5969.

Worby, C. A., S. Mattoo, et al. (2009). "The fic domain: regulation of cell signaling by adenylylation." Mol Cell 34(1): 93-103.

Yarbrough, M. L., Y. Li, et al. (2009). "AMPylation of Rho GTPases by Vibrio VopS disrupts effector binding and downstream signaling." Science 323(5911): 269-272.

Yarbrough, M. L. and K. Orth (2009). "AMPylation is a new post-translational modiFICation." Nat Chem Biol 5(6): 378-379.

6. Acknowledgments

6. Acknowledgments

I carried out the work for this thesis as a member of the group of Prof. Christoph Dehio in the Focal area Infection Biology at the Biozentrum of the University of Basel.

First, I would like to thank Prof. Christoph Dehio for his guidance and support, as well as for the freedom he has given me to explore genomic and evolutionary aspects of *Bartonella*. His valuable suggestions, analytical criticism, and positive attitude were an immense contribution to the success of this work.

I am very grateful to Prof. Walter Salzburger who spontaneously agreed to join my PhD committee and introduced me to the field of molecular evolution. I appreciate not only his support and assistance, but also the time he spent discussing current research with me, helping me write my paper, or proof-reading my manuscripts.

I also want to thank the other member of my thesis committee, Prof. Urs Jenal and Prof. Guy Cornelis, for accompanying me with my work, giving me guidelines, good advice, and constructive criticism.

I would like to express my special gratitude to Dr. Henri Saenz for introducing me to the field of science, for his valuable advice, and for the great fun we had while carrying out mutual projects.

I am very grateful to Maxime Quebatte for always having an open ear for me and an open mind in discussing problems, strategies, or even some foolish ideas about *Bartonella* evolution. I highly appreciated the peaceful atmosphere in our shared office/lab space 485.

I have had the great fortune of being able to introduce two very dedicated and enthusiastic Master students, Marius Liesch and Alexander Harms, to the world of *Bartonella*. I thank Marius Liesch very much for his cooperation with our mutual projects. In the future, I look forward to explore new aspects of *Bartonella* evolution together with Alexander Harms.

I also want to thank Rusudan Okujava and Claudia Mistl for their assistance with our rats in the basement.

In addition to those mentioned above, I wish to express my thanks to all present and past members of the Dehio lab for all the help, discussions, ideas, and the extra-lab activities. I thank Andrea Basler, Dr. Arto Pulliainen, Dr. Michaela Dehio, Nadège Devaux, Yvonne Ellner, Marco Faustmann, Jérémie Gay-Fraret, Arnaud Goepfert, Dr. Patrick Guye, Barbara Hauert, Sonja Huser, Dr. Thomas Rhomberg, Dr. Florine Scheidegger, Hermine Schein, Dr. Michael Schmid, Dr. Gunnar Schröder, Isabella Toller, Dr. David Topel, and Matthias Truttmann.

For the collaboration on the genome sequencing projects, I want to thank Dr. Stephan Schuster, at Penn State University, Pennsylvania, USA, as well as Dr. Christa Lanz from the Max-Planck Institut für Entwicklungsbiologie and Günter Raddatz from the Max-Planck Institut für biologische Kybernetik in Tübingen, Germany.

For the collaboration on the genome annotation projects, I am very grateful to Alexander Calteau, Claudine Médigue and the other members of the Comparative Genomics Atelier at Genoscope in Paris, France.

Also, I would like to acknowledge the invaluable support from the many people working on the technical and administrative staff of the 4th floor. Thank you all very much for all the things running in the back of the everyday business of science.

For the critical and careful reading of my thesis I am very grateful to Maryalice Wichmann.

Finally, I want to thank my whole family for their support throughout my studies and their interest in my work.

Thank you, Fabienne, for your loving, support, constant encouragement, and, above all, your boundless patience with me. You have been my greatest source of equilibrium during all these years!

7. Curriculum Vitae

7. Curriculum Vitae

Name:	Philipp Engel
Date of birth:	27.07.1981
Place of birth:	Essen, Germany
Nationality:	German
Family status:	Single

Address	**Office:**	**Private:**
	Biozentrum University of Basel Infection Biology Klingelbergstrasse 50/70 CH-4056 Basel Switzerland Phone: +41-61-267 2137 Fax: +41-61-267 2118 Email: philipp.engel@unibas.ch	Elsternweg 23 CH-4103 Bottmingen Switzerland Phone: +41-61-411 01 14 +41-61-261 49 38

Education and Research experience:

2006 – 2010	Ph.D. thesis Focal area Infection Biology, Biozentrum, University of Basel CH Supervisor, Prof. Christoph Dehio Committee Members, Prof. Urs Jenal, Prof. Walter Salzburger, Prof. Guy Cornelis Title: “A Genomic Approach to the Role of Type IV Secretion Systems in *Bartonella* Host Adaptation”
2004 - 2005	Graduate studies Awarded M.Sc. in Molecular Biology Focal area Infection Biology, Biozentrum, University of Basel CH Supervisor, Prof. Christoph Dehio Title: “The Genome Sequence of *Bartonella tribocorum* and Comparative Genomics of the Genus *Bartonella*”
2001 - 2004	Undergraduate studies Major studies in Molecular Biology Minor studies in Computer Science University of Basel CH
1997 - 2000	Matura Type B (university-entrance diploma) Gymnasium Münchenstein, Münchenstein BL, CH

Teaching experience:

2009 – 2010	Tutorage of a master student Focal area Infection Biology, Biozentrum, University of Basel CH
2004 – 2009	Assistant and tutor for undergraduate students in practical courses “Microbiology” once a year Focal area Infection Biology, Biozentrum, University of Basel CH
2007 – 2008	Tutorage of a master student Focal area Infection Biology, Biozentrum, University of Basel CH

Oral presentations:

2009	Short presentation 4th European Conference on Prokaryotic Genomes” (ProkaGenomics2009), Göttingen, Germany Title: “Lineage-specific diversification of VirB Type IV secretion systems confers host adaptability in the genus *Bartonella*”
2008	Presentation at the Biozentrum Symposium in Basel, CH Title: “Genomic analysis of *Bartonella* identifies a role for type IV secretion systems in host adaptability”
2007	Presentation at the Swiss Molecular Microbiology Conference (SWIMM07), Villars-sur-Ollon, CH Title: “Inferring pathogenicity factors critical for host adaptation in the genus *Bartonella*”

Poster presentations:

2008	Poster at the Genomes 2008 Conference, Institut Pasteur, Paris, France Title: “The genome sequence of *Bartonella clarridgeiae* and evolution of VirB Type IV secretion systems”
2007	Poster at the European Conference on Synthetic Biology (ECSB), Sant Feliu de Guixols, Barcelona, Spain Title: “From a devastating to a low virulence pathogen - Evolution of the modern bartonellae”

2007	Poster at the 6th BioValley Life Science Week, Basel, CH Title: “From a devastating to a low virulence pathogen - Evolution of the modern bartonellae”
2006	Poster at the Biozentrum Symposium in Basel, CH Title: “From a devastating to a low virulence pathogen - Evolution of the modern bartonellae”
2005	Poster at the 2nd European Conference on Prokaryotic Genomes (Prokagen2005), Göttingen, Germany Title: “The Complete Genome Sequence of *Bartonella tribocorum* and Prophage Reduction during Evolution of *Bartonella*”

Publications:

1. Henri L Saenz, Philipp Engel, Michèle C Stoeckli, Christa Lanz, Günter Raddatz, Muriel Vayssier-Taussat, Richard Birtles, Stephan C Schuster, Christoph Dehio (2007) Genomic analysis of *Bartonella* identifies type IV secretion systems as host adaptability factors. Nature Genet. 39 (12):1469-1476

2. Philipp Engel, Christoph Dehio (2009) Genomics of host-restricted pathogens of the genus *Bartonella.* Microbial Pathogenomics, Genome Dynamics, Karger. Volume 6: 158-159.

3. Muriel Vayssier-Taussat, Danielle Le Rhun, Hongkuan Deng, Francis Biville, Sandra Cescau, Antoine Danchin, Geneviève Marignac, Evelyne Lenaour, Henri-Jean Boulouis, Maria Mavris, Le Menach Arnaud, Huan-Ming Yang, Jing Wang, Philipp Engel, Henri Saenz and Christoph Dehio. The Trw type IV secretion system of *Bartonella* mediates host-specific infection of erythrocytes. Submitted to PLoS Pathogens.

4. Philipp Engel, Walter Salzburger, Marius Liesch, Chao-Chin Chang, Soichi Maruyama, Christa Lanz, Alexandra Clateau, Aurélie Lajus, Claudine Médigue, Stephan C. Schuster, Christoph Dehio. Parallel adaptive radiations in a bacterial pathogen. In preparation.

About the cover

The graphic shows part of the genome of the bacterium *Bartonella tribocorum* depicting annotated genes, location of genomic islands, the core genome of *Bartonella*, and the GC skew. (Chapter 3.5.2, p. 201.)