Characterisation of the fibronectin binding domains and genomic variation of the *Bartonella* adhesin A of *Bartonella henselae*

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biowissenschaften
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Arno Thibau
aus Lokeren, Belgien

Frankfurt am Main, 2022
D30

vom Fachbereich 15 der Biowissenschaften der Johann Wolfgang Goethe-Universität als Dissertation angenommen

Dekan: Prof. Dr. Sven Klimpel

Gutachter: Prof. Dr. Volker Müller

                   Prof. Dr. Volkhard A. J. Kempf

Datum der Disputation:

Table of contents

1. Introduction

1.1. The species *Bartonella henselae*

1.1.1. The genus *Bartonella* as emerging pathogens

Bartonella species have long since been marked as emerging pathogens, yet *B. quintana* has been detected in human remains dating back to 2,000 BC [1–3]. The genus *Bartonella* belongs to the alpha-2 subgroup of the class proteobacterium, currently consists of more than 40 identified species, and is categorised into three clades [4–7]. The first two clades consist solely of the species *B. apis* and *B. tamiae*, respectively, while the third clade, designated as the eubartonellae [8], is further classified into four phylogenetic lineages based on several core genes and six house-keeping genes (*rpoB*, *groEL*, *ribC*, *ftsZ*, *nuoG*, and *gltA*) [9]. Lineage 1 consists of the human pathogenic *B. bacilliformis* and the phylogenetically related *B. ancashensis* [10–12]. The other three lineages include both zoonotic and human pathogens with a wide array of mammals as reservoir hosts [13–17]. *Bartonella* species are hematophagous-arthropod-borne, rod-shaped facultative intracellular bacteria that are characterised by their 'stealthy' course of infection [13, 18]. From a clinical point of view, long-lasting host reservoir infections are commonly asymptomatic, while incidental host infections show a higher morbidity, despite often being self-limiting. *Bartonella* species are the causative agent of a wide variety of human diseases with a remarkable range of symptoms, examples are Carrion's disease (*B. bacilliformis*), trench fever (*B. quintana*), and cat scratch disease (CSD; *B. henselae*) [19, 20].

1.1.2. Genomic adaptation and genotypes of *B. henselae*

Bartonella species are evolutionary driven by adaptive radiation where host-specific adaptation is suggested to be essential for their survival and pathogenicity [21]. Homologous recombination, for example by phase variation, and horizontal gene transfer mediate the origination of different strains within one species, while simultaneously maintaining genome integrity [16, 22, 23]. In addition, reductive genome evolution is commonly observed within the genus *Bartonella* and is in line with the overall intracellular lifestyle and hematophagous-arthropod-borne transmission [24–26].

B. henselae strains demonstrate variable genomic and phenotypic differences [27, 28] and thus far, two *B. henselae* genotypes were identified based on 16S ribosomal (r)RNA gene sequences. Genotype I is represented by *B. henselae* ATCC49882^{T} Houston-I and genotype II by *B. henselae* Marseille [29–32]. Classification and correct bacterial genotyping of

B. henselae strains is a challenging task and has been performed using different techniques such as pulse field gel electrophoresis, restriction fragment length polymorphism, or various polymerase chain reaction (PCR)-based techniques [33–36]. More recently, genomic DNA (gDNA) sequencing-based approaches, for example 16S rRNA gene sequencing, multilocus sequence typing, or whole genome sequencing techniques have been implemented [24, 37–41].

Differing correlations of both genotypes regarding their infection strategy of feline and human endothelial cell lines have been described [42–44]. As such, some strains are associated with certain host related niches and thus display distinct characteristics, in which outer membrane protein variation is suggested to be mediated by phase variation [27, 28, 35, 45]. Phenotypic differences are among others related to the production of two important pathogenicity factors: the *Bartonella* adhesin A (BadA) and the VirB/D4 type IV secretion system [35, 46].

1.1.3. *B. henselae* is the causative agent of cat scratch disease

B. henselae is an endemic, fastidious, and slow-growing bacteria with cats (*Felis catus*) serving as the primary reservoir host. Infection of cats usually results in a long-lasting asymptomatic bacteraemia [47, 48]. Transmission among cats occurs through the faeces or bites of infected cat fleas (*Ctenocephalides felis*) and possibly via ticks (for example *Ixodes ricinus*) [49–52]. Other reservoir hosts such as dogs and guinea pigs have also been suggested [53–56]. Incidental transmission of *B. henselae* to humans occurs indirectly by contaminated flea faeces via cat scratches or bites. Thus far, no direct *B. henselae* transmission from flea to human has been reported [47, 49].

B. henselae is the causative agent of CSD, a self-limiting zoonotic disease characterised by localised lymphadenopathy appearing 2-3 weeks after inoculation. Other common symptoms are skin lesions at the site of the cat scratch, fatigue, and fever. Infection of immunocompromised patients might induce more severe diseases such as 'culture-negative' endocarditis, bacillary angiomatosis, or other related vasculoproliferative disorders [19, 57–60] and are often treated with antibiotics such as rifampicin, azithromycin, and/or doxycycline [61, 62]. The majority of CSD-cases are reported in the United States of America where for instance a yearly average of ca. 12,000 diagnoses and ca. 500 hospitalisations have been observed during the period 2005-2013 [63, 64]. Diagnosis of CSD is most often established by histological examination, bacterial cultivation, serologic testing, and/or PCR-based methods, combined with a history of exposure to cats [65].

1.1.4. Pathogenicity of *B. henselae*

The interaction of *B. henselae* with host cells has been studied, however, conclusive knowledge on the primary niche of *B. henselae* or on the means of invasion and colonisation of host cells remains unclear. After inoculation in cats, *B. henselae* establishes a long-lasting and relapsing infection of erythrocytes [66, 67]. Thus far, no invasion of human erythrocytes by *B. henselae* has been observed [68]. Other identified niches in cats are the liver, heart, and lymph nodes [69]. Inoculation of human cells by *B. henselae* occurs usually through infected flea faeces that is superficially scratched into the skin. Human umbilical vein endothelial cells (HUVEC) have often been associated with intracellular or attached *B. henselae* and endothelial cells were therefore suggested as a primary niche [70]. Moreover, several outer membrane proteins have been identified to adhere to HUVECs *in vitro* [71]. Accordingly, cultured HUVECs have been proven an efficient *in vitro* model system to study infection experiments with *B. henselae* [72, 73]. Interactions between endothelial cells and *B. henselae* are characterised by internalisation via *Bartonella*-containing vacuoles or via an actin-dependent invasome-mediated uptake, suggested to facilitate long-lasting host cell colonisation [18, 74, 75]. Subsequent activation of the key transcription factor hypoxia-inducible factor-1, secretion of the vascular endothelial growth factor, and inhibition of apoptosis results in the formation of vascular tumours [76–79].

Two major pathogenicity factors of *B. henselae* have been described more extensively and have been demonstrated to participate actively in the interaction with host cells: the VirB/D4 type IV secretion system and the trimeric autotransporter adhesin (TAA) BadA. The VirB/D4 type IV secretion system is known to translocate a mixture of *Bartonella* effector proteins into the host cell and is associated with invasome formation via actin cytoskeleton reorganisation, inhibition of apoptosis, activation of proinflammatory signalling, and stimulation of vasoproliferation [18, 78, 80, 81]. Activation of the VirB/D4 type IV secretion system is regulated by the BatR/BatS two-component system in combination with the alternative sigma factor RpoH1 [82]. The BatR/BatS two-component system functions as a pH sensor and controls the adaptive response of *B. henselae* during infection of human host cells and during changes in the host cell environment [83, 84]. Furthermore, it has been demonstrated that the presence of BadA on the outer membrane of *B. henselae* negatively effects the translocation of VirB/D4-dependent *Bartonella* effector proteins and subsequent host cell invasion via invasome formation [46]. Other identified pathogenicity factors of *B. henselae* are the Trw type IV secretion system [85, 86], the CAMP-like factor autotransporter [87], the invasion associated locus B (IalB) [88], the lipopolysaccharide [89], the hemin binding protein A (HbpA/Pap31) [90–93], and few outer membrane proteins such as Omp43 and Omp89 [71, 94].

1.2. *Bartonella* adhesin A

1.2.1. Characteristics of trimeric autotransporter adhesins

B. henselae is characterised by the enormous BadA protein present abundantly on the bacteria's surface. BadA belongs to the class of TAAs, also designated as the type Vc secretion system (Figure 1). TAAs are widely present in human pathogenic Gram-negative bacteria and in the genus *Bartonella* and display a common modular architecture consisting of a long N-terminal passenger domain and a C-terminal anchor domain [95, 96]. The *Yersinia* adhesin A (YadA) of *Yersinia enterocolitica* is considered the prototypical TAA while other well-known and clinically relevant examples are the *Acinetobacter* trimeric autotransporter (Ata) of *Acinetobacter baumannii*, the *Neisseria* adhesin A (NadA) of *Neisseria meningitidis*, and the *Salmonella* adhesin A (SadA) of *Salmonella enterica* [97–100].
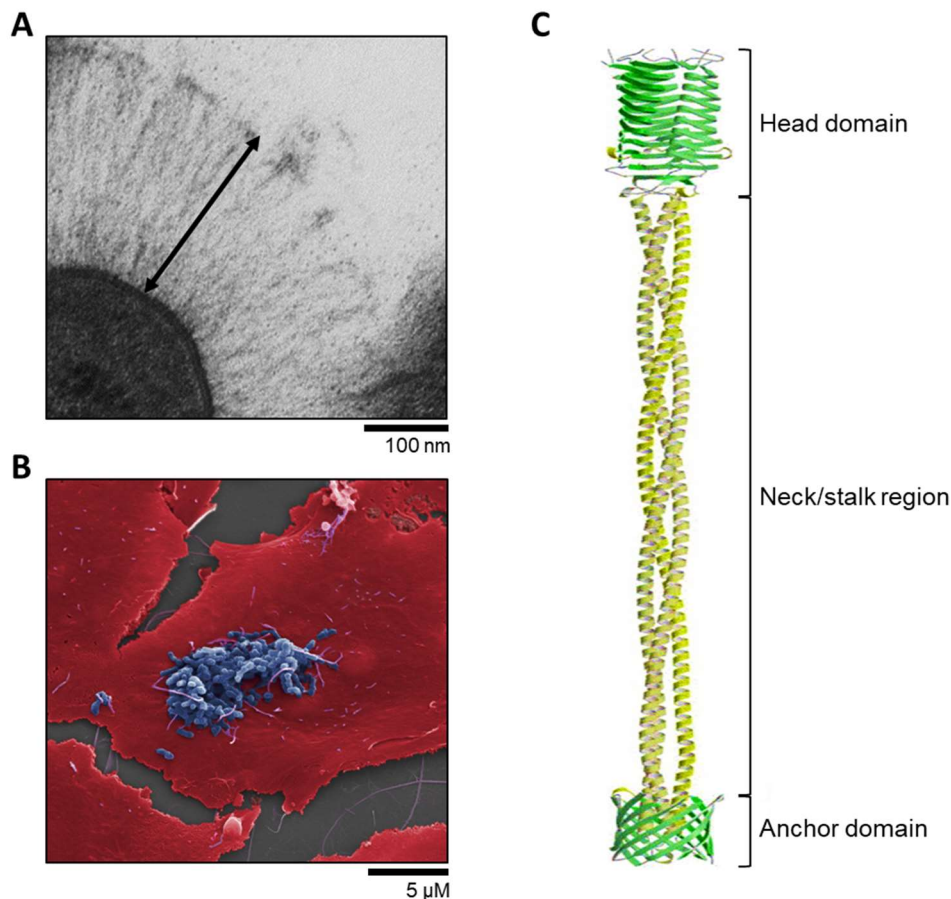
Figure 1. Overview of BadA from *B. henselae* as representative of trimeric autotransporter adhesins (adapted from [95, 101]). **(A)** Transmission electron microscopy image of *B. henselae* Marseille with BadA (black arrow) present as long fibres on the bacterial outer membrane. Scale bar: 100 nm. **(B)** Scanning electron microscopy image of *B. henselae* Marseille (blue) adhering and aggregating on the surface of HUVECs (red) 30 minutes after infection. Scale bare: 5 µm. **(C)** Structural simplified model representing the common architecture of TAAs consisting of an N-terminal head domain, a long neck/stalk region, and a C-terminal anchor domain.

The passenger domain of BadA consists of a head domain and a long neck/stalk region and has been organised into shorter sequence motifs using the 'domain dictionary' approach of the domain annotation of TAA (daTAA)-server [96, 102–104]. Moreover, the crystal structure of the N-terminal head domain has been solved and contains so-called YadA-like head repeats, a head insert motif (HIM)1, a tryptophan (Trp) ring motif, a GIN motif, and a neck motif [105]. The neck/stalk region was shown to consist of a recurring pattern including FGG motifs, coiled-coil motifs, and DALL-neck tandem connectors [106]. FGG motifs come as different subtypes but are generally described as an insertion of a 3-stranded β-meander into a coiled-coil region causing a 120 ° twist of the subunit chains around the trimer axis and usually support the long BadA trimer against vertical shear stress [97, 104]. In addition, the presence of multiple glycine residues in the FGG motif sequence facilitate tight turns and the formation of short loops [107]. Coiled-coil regions are characterised by a super-helical α-helix structure and generally show a 'knobs-into-holes' packing where hydrophobic residues from one strand are situated in the trimeric core and fit into a space enveloped by the other two strands [108]. Most coiled-coil motif sequences consist of a heptad repeat where the first and fourth amino acid (aa) are hydrophobic residues buried in the 'knob' and are separated by polar residues present on the trimer's exterior [104]. Transitions from α-helices to β-stranded structures, and back to α-helices are facilitated via DALL and neck motif sequences, respectively. DALL motifs consist of two β-strands forming a hairpin perpendicular to the fibre axis connected to the β-strand of the following neck motif. Neck motifs are distributed abundantly among the BadA neck/stalk region, are considered the most conserved part of the passenger domain in TAAs, and consist of N-terminal β-sheets forming a hairpin making it a conformationally flexible region [104, 106, 109]. The C-terminal anchor domain is homologous among TAAs and was shown to consists of a 12-stranded β-barrel embedded in the bacterial outer membrane [95, 104].

1.2.2. Characteristics and biological function of BadA

BadA was originally identified as a type IV-like pilus and was first characterised in *B. henselae* Marseille [76, 110]. It was initially demonstrated that *B. henselae* Marseille expresses a 9.3 kb *badA* gene encoding for a BadA protein (3,082 aa) with an enormous fibre length of ca. 240 nm and mass of ca. 328-340 kDa per monomer [76]. The corresponding BadA neck/stalk region was depicted to consist of 22 neck/stalk domains, defined by their respective neck sequence [106]. However, while analysing several *B. henselae* strains, variations in length of the respective *badA* genes were observed and were attributed to differences in the repetitive neck/stalk region. In addition, some of the analysed *B. henselae* strains lacked the ability to express *badA* because of the presence of a premature stop codon [111].

Thus far, regulation of *badA* expression remains rather inconclusive but has been linked to both the general stress response system making use of an alternate sigma factor [112] and to the *Bartonella* regulatory transcript 1 (Brt1) in association with the transcriptional regulatory protein 1 (Trp1) [113, 114]. Accordingly, the alternate sigma factor RpoE and the anti-anti-sigma factor PhyR are positive regulators of the general stress response system that were shown to repress *badA* expression and be activated under conditions mimicking the cat flea host environment. On the other hand, Trp1 was shown to increase *badA* transcription through binding with the suggested *badA* promoter region. Transcription of *trp1* is controlled by *brt1* that was shown to be upregulated under human host cell conditions. In addition, cultivating *B. henselae* under human host cell conditions *in vitro* with a neutral pH of 7.4 and a temperature of 37 °C has proven to stimulate *badA* transcription compared to environments mimicking cat flea host conditions [112, 115, 116].

BadA is a major pathogenicity factor of *B. henselae* and expression of *badA* has been associated directly with angiogenic reprograming of infected host cells via the activation of the hypoxia-inducible factor-1 and secretion of vasculoproliferative cytokines [76, 77]. Furthermore, *badA*-expressing strains have been shown to aggregate in large agglomerates mimicking biofilm formation *in vitro* [113, 115]. In addition, BadA antibodies were detectable in ca. 75 % of sera samples from patients with a suspected *B. henselae* infection, suggesting an immunodominant role for BadA and making it an interesting candidate for the development of a vaccine [76, 101, 117]. Moreover, BadA was demonstrated to be crucial for adhesion to various extracellular matrix (ECM) proteins and angiogenic reprogramming of host cells [72].

1.2.3. The adhesion of BadA to fibronectin

Adhesion to host cells is the first and most crucial step in infections with *B. henselae* and with pathogenic Gram-negative bacteria more generally. It has been demonstrated that *B. henselae* binds ECM proteins using its long and membrane exposed BadA fibres to attach the bacteria to the host cell surface [72, 76, 106, 111, 118]. Consequently, *B. henselae* strains lacking *badA* expression showed only minimal binding to various ECM proteins such as fibronectin, laminin, vitronectin, and several collagens, under both static and dynamic flow conditions [76, 118]. Moreover, via binding assays using modified BadA proteins it was demonstrated that exclusively the BadA neck/stalk region, and not the head domain, is responsible for the adhesion of *B. henselae* to fibronectin [106]. As such, *B. henselae* strains expressing modified *badA* genes that do not include a head domain and only contain 4 neck/stalk domains were still able to bind fibronectin, while *B. henselae* strains expressing modified *badA* genes that

include a head domain and only 1 neck/stalk domain were not. Consequently, it was demonstrated that a minimal number of 4 neck/stalk domains is necessary to bind fibronectin.

Fibronectin was proven to be a key first binding partner of *B. henselae* during infection of human endothelial host cells, therefore, *B. henselae* Marseille showed a significant lower binding to HUVECs deficient in producing fibronectin. It was shown via cross-linking mass spectrometry (XL-MS) that BadA-fibronectin interactions occur via the fibronectin heparin-binding domains [73]. Fibronectin is a multi-domain (type I, II, and III repeats) and heterodimeric glycoprotein that is present abundantly on the cell surface of endothelial cells as a fibrillary-type matrix (cellular fibronectin) or in blood, saliva, and other fluids (plasma fibronectin) making it an excellent first binding partner in infections of blood vessels, heart valves, or in the case of a cat scratch in the human skin [119, 120].

1.3. Aim of the thesis

TAAs are important virulence factors of various pathogenic Gram-negative bacteria and TAA-targeted anti-adhesive strategies might represent a universal strategy to counteract many bacterial infections. BadA is one of the best characterised TAAs and because of its high number of neck/stalk domains it serves as an attractive adhesin to study the domain-function relationship of TAAs in the infection process. Moreover, BadA neck/stalk domains consist of specific and repeated sequence motifs with characteristic and predicted structural conformations. The first and major binding partner during infection is the ECM glycoprotein fibronectin. However, it remains unknown which exact domains or motifs of the BadA neck/stalk region are responsible for fibronectin binding.

Nonetheless, to analyse such BadA-fibronectin binding motifs in detail, it must first be verified that such motifs are conserved within the species of *B. henselae*. It is known that early-passage *B. henselae* strains are able to express *badA*, resulting in long trimeric protein fibres present on the bacterial outer membrane mediating adhesion to the ECM and host cells. The *badA* gene of *B. henselae* stands out because of its enormous length, its highly modular and repetitive structure, and its variability in length. Underlying repeats on the *badA* sequence level suggest easy gene rearrangements via recombination, making correct species typing or phylogenetic analysis challenging. Moreover, the lifestyle of *B. henselae* with frequent transitions from the cat flea's gut to cats and to accidental human hosts might require efficient and quick adaptation strategies. Differences in the *badA* gene length were already observed and are presumably driven by genetic adaptation to frequently alternating host environments. However, it was never possible to correctly sequence this genomic region because of its highly repetitive nature. In the scope of this thesis, genomes of eight *B. henselae* strains were

sequenced using long-read sequencing techniques with the aim of analysing differences in the highly variable *badA* gene and study *badA* expression and functional binding to ECM proteins. Compared to short-read sequencing technologies, long-read sequencing facilitates to differentiate close variants and to cover highly repetitive stretches without major assembly problems.

To identify specific repeated motif sequences that are involved in the adherence to fibronectin, several modified *badA* mutants were constructed in a *B. henselae* strain deficient in expressing *badA* and used in whole-cell binding ELISA experiments. Expression of the truncated and modified *badA* mutants was verified by Western blotting, immunofluorescence microscopy, confocal laser scanning microscopy, and transmission electron microscopy. The identification of common binding motifs between BadA and the major binding partner fibronectin might provide a basis towards the design and development of novel 'anti-adhesive' compounds preventing initial adherence of *B. henselae* to endothelial cells during infection and might also function as a model for studies involving other TAA-producing pathogenic bacteria.

2. Materials and Methods

2.1. Bacterial strains and culture media

Bacterial strains used in this work are listed in Table 1. The different *B. henselae* isolates were collected from human patients and domestic cats and were preserved in the strain collection of the Institute of Medical Microbiology and Infection Control at the University Hospital Frankfurt am Main. The passage number of the different *B. henselae* isolates is < 10, however, the exact passage number before arriving in Frankfurt am Main is unknown. Three of the *B. henselae* strains are variants of the type-strain ATCC49882^T Houston-I [29]. *B. henselae* Marseille ΔBadA-T/pS27, /pHN2S27, /pS28, /pS29, /pS30, and /pHNS30 were designed and/or generated previously by Patrick Kaiser, PhD.

Chemically competent *Escherichia coli* DH5α from New England BioLabs® (NEB) are suitable for high efficiency transformation and were used for all cloning procedures.

Table 1. Overview of *B. henselae* strains used in this work.

B. henselae strain	Characteristics	Source
Marseille	*badA*-expressing isolate from a human patient diagnosed with cat scratch disease	[30]
ATCC49882^T var-1	*badA*-deficient laboratory streptomycin-resistant variant of *B. henselae* ATCC49882^T Houston-I	[29, 80]
ATCC49882^T var-2	*badA*-expressing laboratory variant of *B. henselae* ATCC49882^T Houston-I	[29, 46, 111]
Berlin-I	*badA*-deficient skin biopsy specimen-isolate from a human patient diagnosed with bacillary angiomatosis	[121]
G-5436	*badA*-expressing isolate from a human patient and variant of *B. henselae* ATCC49882^T Houston-I	[29, 122, 123]
88-64 Oklahoma	*badA*-expressing blood isolate from a human patient diagnosed with human immunodeficiency viruses (HIV)	[124]
FR96/BK38	*badA*-expressing blood isolate from a domestic cat	[125]
FR96/BK3	*badA*-expressing blood isolate from a domestic cat	[125]
Marseille		
ΔBadA-T	*B. henselae* Marseille *badA*-deficient mutant with a TN <KAN-2> transposon integrated in *badA*; Kan (30 µg/ml) resistant	[76, 126]
ΔBadA-D	*B. henselae* Marseille *badA*-deficient mutant with *badA* deleted via homologous recombination	This work, [127]

Table 1. Overview of *B. henselae* strains used in this work. *Continued.*

∆BadA-T/pS27	*B. henselae* Marseille ∆BadA-T transformed with pS27; previously referred to as *B. henselae* badA-/pF12; Kan (30 µg/ml) and Gen (10 µg/ml) resistant	[106]
∆BadA-T/pS28	*B. henselae* Marseille ∆BadA-T transformed with pS28; Kan (30 µg/ml) and Gen (10 µg/ml) resistant	This work, [128]
∆BadA-T/pS29	*B. henselae* Marseille ∆BadA-T transformed with pS29; Kan (30 µg/ml) and Gen (10 µg/ml) resistant	This work, [128]
∆BadA-T/pS30	*B. henselae* Marseille ∆BadA-T transformed with pS30; previously referred to as *B. henselae* badA-/pN23; Kan (30 µg/ml) and Gen (10 µg/ml) resistant	[106]
∆BadAT/pHN2S27	*B. henselae* Marseille ∆BadA-T transformed with pHN2S27; previously referred to as *B. henselae* badA-/pHN2F12; Kan (30 µg/ml) and Gen (10 µg/ml) resistant	[106]
∆BadA-T/pHNS30	*B. henselae* Marseille ∆BadA-T transformed with pHNS30; previously referred to as *B. henselae* badA-/pHN23; Kan (30 µg/ml) and Gen (10 µg/ml) resistant	[72, 106]
∆BadA-T/pD16S28	*B. henselae* Marseille ∆BadA-T transformed with pD16S28; Kan (30 µg/ml) and Gen (10 µg/ml) resistant	This work, [128]
∆BadA-T/pD19S28	*B. henselae* Marseille ∆BadA-T transformed with pD19S28; Kan (30 µg/ml) and Gen (10 µg/ml) resistant	This work, [128]
∆BadA-T/pD25S28	*B. henselae* Marseille ∆BadA-T transformed with pD25S28; Kan (30 µg/ml) and Gen (10 µg/ml) resistant	This work, [128]
∆BadA-T/pD27S29	*B. henselae* Marseille ∆BadA-T transformed with pD27S29; Kan (30 µg/ml) and Gen (10 µg/ml) resistant	This work, [128]

2.2. Vectors and primers

Table 2. Overview of vectors used in this work.

Vector	Characteristics	Source
pBIISK_*sacB*/*kanR*	Suicide cloning vector including a *sacB*/*kanR*-cassette; Kan (50 µg/ml) resistance	[129]
pBIISK_*sacB*/*kan_*UpDownBadA	pBIISK_*sacB*/*kanR* including a ca. 2 kb up- and downstream region of *badA* (strain Marseille)	This work, [127]
pMK-RQ	GeneArt® cloning vector; Kan (50 µg/ml) resistance	GeneArt®
pMA	GeneArt® cloning vector; Amp (50 µg/ml) resistance	GeneArt®
pMS-RQ	GeneArt® cloning vector; Spec (30 µg/ml) resistance	GeneArt®
pBBR1MCS-5	Broad-host range vector; Gen (10 µg/ml) resistance	[130]

Table 2. Overview of vectors used in this work. *Continued.*

pS27	pBBR1MCS-5 including a ca. 3.0 kb *badA* fragment (*badA* S27); previously referred to as *badA* F12	[106]
pS28	pBBR1MCS-5 including a ca. 2.6 kb *badA* fragment (*badA* S28)	This work, [128]
pS29	pBBR1MCS-5 including a ca. 2.2 kb *badA* fragment (*badA* S29)	This work, [128]
pS30	pBBR1MCS-5 including a ca. 1.9 kb *badA* fragment (*badA* S30); previously referred to as *badA* N23	[106]
pHN2S27	pBBR1MCS-5 including a ca. 4.3 kb *badA* fragment (*badA* HN2S27); previously referred to as *badA* HN2F12	[106]
pHNS30	pBBR1MCS-5 including a ca. 3.1 kb *badA* fragment (*badA* HNS30); previously referred to as *badA* HN23	[72, 106]
pD16S28	pBBR1MCS-5 including a ca. 3.0 kb *badA* fragment (*badA* D16S28)	This work, [128]
pD19S28	pBBR1MCS-5 including a ca. 2.8 kb *badA* fragment (*badA* D19S28)	This work, [128]
pD25S28	pBBR1MCS-5 including a ca. 2.8 kb *badA* fragment (*badA* D28S28)	This work, [128]
pD27S29	pBBR1MCS-5 including a ca. 2.6 kb *badA* fragment (*badA* D27S29)	This work, [128]
pCR™2.1-TOPO®_*glyA*	Standard cloning vector including a 120 bp fragment of *glyA*; Kan (50 µg/ml) and Amp (50 µg/ml) resistance	Invitrogen

Table 3. Overview of primers used in this work (Metabion, Munich, Germany).

Primer	5'-3' sequence*	Application
*badA*FrUp_Fw	<u>AGCCCGGGGGATCC</u>ATTGGTTTGGCAGTGTCCAACT	PCR; upstream flanking region of *badA* (*B. henselae* Marseille)
*badA*FrUp_Rv	<u>CCTGCGCC</u>TCTCTTTGATGTGACAGATAATTTTTTTCATAATAAATTCTCCAA	PCR; upstream flanking region of *badA* (*B. henselae* Marseille)
*badA*FrDown_Fw	<u>AAAGAGA</u>GGCGCAGGGATTACTTTGAG	PCR; downstream flanking region of *badA* (*B. henselae* Marseille)
*badA*FrDown_Rv	<u>GGCCGCTCTAGAACTAG</u>ACTTCAAATAATATCTCTTACTATTGAATAATATTTTCCTAAAACTACG	PCR; downstream flanking region of *badA* (*B. henselae* Marseille)
pBIISK_Fw	CTAGTTCTAGAGCGGCCGC	PCR; vector pBIISK_*sacB*/*kanR*
pBIISK_Rv	TGGATCCCCCGGGCTG	PCR; vector pBIISK_*sacB*/*kanR*
pBIISK_seq_Fw	ATTGGTTGTAACACTGGCAGAG	PCR and sequencing; vector pBIISK_*sacB*/*kanR*_UpDownBadA
pBIISK_seq_Fv	GCGCAATTAACCCTCACTAAAG	PCR and sequencing; vector pBIISK_*sacB*/*kanR*_UpDownBadA
IntegrationA_Fw	GGATCCATTGGTTTGGCAGTG	PCR; gDNA *B. henselae* Marseille and vector pBIISK_*sacB*/*kanR*_UpDownBadA

*Underlined primer protein sequences do not anneal to the template DNA

Table 3. Overview of primers used in this work (Metabion, Munich, Germany). *Continued*.

IntegrationA_Rv	GTTTTTCCATGGTAGCATCACTC	PCR; gDNA *B. henselae* Marseille and vector pBIISK_*sacB/kanR*_UpDownBadA
IntegrationB_Fw	TTTAAGCTACGCGGTTGAGGATG	PCR and sequencing; gDNA *B. henselae* Marseille and vector pBIISK_*sacB/kanR*_UpDownBadA
IntegrationB_Rv	CGCTCTAGAACTAGACTTCAAATAATATCTCTTAC	PCR; gDNA *B. henselae* Marseille and vector pBIISK_*sacB/kanR*_UpDownBadA
Segregation_Fw	TTAAAGGGTTAAGTGCAGC	Sequencing; gDNA *B. henselae* Marseille ΔBadA-D
Repeat_Fw	GAGATTGCTAATGCGAATGG	PCR and sequencing; gDNA *B. henselae* strains
Repeat_Rv	ATTACCAACACCATTGCC	PCR and sequencing; gDNA *B. henselae* strains
S28domains_Fw	<u>CTCTAGAACTAGTGGATCCC</u>GAATTCCTGAATTTAGAGAGTG	PCR and sequencing; *badA* mutants in vector pBBR1MCS-5
S28domains_Rv	<u>GTATCGATAAGCTTGATATC</u>GAATTCTTTTTCGTAGAAACAAGAG	PCR and sequencing; *badA* mutants in vector pBBR1MCS-5
pBBR1MCS-5_Fw	<u>GTTTCTACGAAAAA</u>GAATTCGATATCAAGCTTATCGATACC	PCR; vector pBBR1MCS-5
pBBR1MCS-5_Rv	<u>CTCTCTAAATTCAGGAATTC</u>GGGATCCACTAGTTCTAGAGC	PCR; vector pBBR1MCS-5
pBBR1MCS-5_GA_Fw	CTCACTATAGGGCGAATTGG	PCR; *badA* mutants in vector pBBR1MCS-5
pBBR1MCS-5_GA_Rv	GAGTTAGCTCACTCATTAGGC	PCR; *badA* mutants in vector pBBR1MCS-5
BadA1_Fw	CGTTCCCAATTTGACCAC	PCR and sequencing; gDNA *B. henselae* Marseille ΔBadA-D and *badA* mutants in vector pBBR1MCS-5
BadA2_Fw	TATTCAATGAGCAATATGCTTGCGACC	Sequencing; *badA* mutants in vector pBBR1MCS-5
BadA3_Fw	AGTTGGTACGTCTTTCACC	PCR and sequencing; *badA* mutants in vector pBBR1MCS-5
BadA4_Fw	ACGACTGTTACGCAACAGG	PCR and sequencing; *badA* mutants in vector pBBR1MCS-5
BadA5_Fw	TGATAAGGACGGTTCAGC	PCR and sequencing; *badA* mutants in vector pBBR1MCS-5
BadA_Rv	ATGATTCGACGTGTTTCACC	PCR and sequencing; *badA* mutants in vector pBBR1MCS-5
glyA_Fw	GACAGGAAAATGTGCCGAAT	qPCR; *glyA* in vector pCR™2.1-TOPO®_*glyA*
glyA_Rv	GCAGGTGAACCAAGACGAAT	qPCR; *glyA* in vector pCR™2.1-TOPO®_*glyA*

*Underlined primer protein sequences do not anneal to the template DNA

2.3. Antibodies

Table 4. Overview of antibodies used in this work.

Antibody	Dilution	Application	Source
Rabbit anti-BadA	1:4,000 (WB); 1:400 (IFM)	Primary antibody; WB and IFM	This work
Rabbit anti-BadA-DALL	1:4,000 (WB); 1:2,000 (ELISA); 1:100 (IFM)	Primary antibody and inhibiting compound; WB, ELISA, and IFM	This work
Rabbit anti-*B. henselae*	1:1,000	Primary antibody; WB and ELISA	[131]
Mouse anti-fibronectin	1:1,000	Primary antibody; ELISA	Becton-Dickinson (610077)
HRP-conjugated swine anti-rabbit	1:2,000	Secondary antibody; WB and ELISA	Agilent-Dako (P0217)
Alexa 488-conjugated goat anti-rabbit	1:200	Secondary antibody; IFM	Dianova (111-545-045)
HRP-conjugated goat anti-mouse	1:1,000	Secondary antibody; ELISA	Agilent-Dako (P0260)

2.4. Proteins and markers

Table 5. Overview of proteins and markers used in this work.

Protein	Application	Producer
Fibronectin (human plasma)	ELISA	Sigma-Aldrich (F2006)
Collagen-I (human)	ELISA	Merck (CC050)
Bovine serum albumin (BSA)	ELISA	Sigma-Aldrich (A7030)
TypeOne™ Restriction Inhibitor	Transformation by electroporation	Lucigen (192000)
PCRBIO HiFi Polymerase	PCR	PCR Biosystems (PB10.41)
Marker		
PageRuler™ Plus Prestained Protein Ladder	SDS-PAGE and WB	Thermo Scientific (26619)
Quick-Load® 1 kb DNA Ladder	Agarose gel electrophoresis	NEB (N0468)
Quick-Load® 100 bp DNA Ladder	Agarose gel electrophoresis	NEB (N0551)

2.5. Reagent kits

Table 6. Overview of reagent kits used in this work.

Kit	Application	Producer
NucleoSpin® Plasmid	Isolation of plasmid DNA	Macherey-Nagel (740588)
NucleoBond® Xtra Maxi	Isolation of high yield plasmid DNA	Macherey-Nagel (740414)
MagAttract® HMW DNA Kit	Isolation of HMW gDNA	Qiagen (67563)
QIAquick® PCR Purification Kit	Purification of amplified DNA fragments	Qiagen (28106)
MinElute Gel Extraction Kit	Purification of amplified DNA from gels	Qiagen (28604)
Quant-iT™ PicoGreen™ dsDNA Assay-Kit	Quantification of dsDNA	Thermo Scientific (P11496)
Luna® Universal qPCR Master Mix	Detection and amplification of DNA	NEB (M3003)
Gibson Assembly® Master Mix	Assembly of DNA fragments	NEB (E2611)

2.6. Chemicals and substrates

Table 7. Overview of chemicals and substrates used in this work.

Chemical or substrate	Producer
Acetic acid	Sigma-Aldrich
Acrylamid/Bisacrylamid (37.5:1)	Roth
Agarose	Roth
Ammonium Peroxydisulphate (APS)	AppliChem
Ampicillin (Amp)	Roth
Calcium chloride dihydrate ($CaCl_2.2H_2O$)	Sigma-Aldrich
Coomassie Brilliant Blue R-250 Destaining	Bio-Rad
Coomassie Brilliant Blue R-250 Staining	Bio-Rad
4′,6-diamidino-2-phenylindole (DAPI)	Sigma-Aldrich
Disodium phosphate heptahydrate ($Na_2HPO_4.7H_2O$)	Sigma-Aldrich
Dulbecco's phosphate buffered saline (PBS)	Pan Biotech
Ethanol (≥ 99.5 %)	Roth
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich
Fluorescence mounting media	Agilent-Dako
Foetal calf serum (FCS)	Sigma-Aldrich
Gelatine sepharose-4B	GE Healtcare
Gentamycin (Gen)	Gibco
Glutaraldehyde	Electron Microscopy Sciences

Table 7. Overview of chemicals and substrates used in this work. *Continued.*

Glycerol	Roth
Glycine	Sigma-Aldrich
HEPES	Sigma-Aldrich
Hydrogen chloride (HCl)	Sigma-Aldrich
Kanamycin (Kan)	MP Biomedicals
Laemmli sample buffer (2X)	Sigma-Aldrich
Luria/Miller (LB) Agar	Roth
LB medium	Roth
Methanol	Roth
Midori Green Direct	Nippon Genetics
Milk powder	Heirler Frema Reform
Nuclease-free H_2O	Thermo Scientific
Paraformaldehyde (PFA)	Electron Microscopy Sciences
Polyethylene glycol (PEG) 6000	Roth
2-propanol	Sigma-Aldrich
Rabbit blood	Acile AG
RPMI-1640	PAN-Biotech
Schneider's Drosophila Medium	Serva
SOC Outgrowth Medium	NEB
Sodium bicarbonate ($NaHCO_3$)	Sigma-Aldrich
Sodium chloride (NaCl)	Sigma-Aldrich
Sodium dihydrogen phosphate monohydrate ($NaH_2PO_4.H_2O$)	Sigma-Aldrich
Sodium dodecyl sulphate (SDS)	Roth
Sodium hydroxide (NaOH)	VWR
Sodium pyruvate	Sigma-Aldrich
Spectinomycin (Spec)	Sigma-Aldrich
Sucrose	Sigma-Aldrich
SuperSignal West Pico PLUS Luminol/Enhancer	Thermo Scientific
SuperSignal West Pico PLUS Stable Peroxide	Thermo Scientific
N,N,N',N'-Tetramethylethane-1,2-diamine (TEMED)	Roth
3,3′,5,5′-tetramethylbenzidine liquid substrate (TMB)	Sigma-Aldrich
Tris(hydroxymethyl)-aminomethane (Tris)	Applichem
Tween 20	Roth

2.7. Equipment and consumables

Table 8. Overview of equipment used in this work.

Agarose gel electrophoresis	Producer
Gel casters	Bio-Rad
Horizontal electrophoresis chambers	Bio-Rad
Power supply EV265	Consort
Gel imaging system FastGene FAS V	Nippon Genetics
SDS-PAGE	
Vertical electrophoresis chambers Mini-PROTEAN System	Bio-Rad
Power supply PowerPac™ Basic	Bio-Rad
Combs, gel casters, short and spacer glass plates	Bio-Rad
Western Blotting	
Trans-Blot® Turbo™ Transfer System	Bio-Rad
Gel imaging ChemiDOC XRS+ system	Bio-Rad
Pipettes	
Single Channel pipettes	Gilson
Multichannel Research Plus pipette	Eppendorf
Pipetus®	Hirschmann-Laborgeräte
Analytic scales	
Kern 572	Kern
Kern PNJ	Kern
Centrifuges	
Heraeus Biofuge™ Stratos™ centrifuge	Thermo Scientific
Centrifuge 5430 R	Eppendorf
MiniSpin®	Eppendorf
Mini centrifuge ROTILABO®	Roth
Microscopes	
Axio Imager 2 microscope with Spot RT3 camera	Zeiss and Diagnostic Instruments Inc.
Eclipse Ti microscope	Nikon
Incubators	
CH-4103 Bottmingen Incubator Shaker	Infors HT
Minitron CO_2 Incubator Shaker	Infors HT
Heraeus BBD 6220 CO_2 Incubator	Thermo Scientific
Shaking devices	
Unimax 2010	Heidolph
Rocking platform Rocky®	LTF Labortechnik
Rotator roller mixer SRT9D	Stuart

Table 8. Overview of equipment used in this work. *Continued.*

Other equipment	
Thermomixer Comfort	Eppendorf
Water bath WNB45	Memmert
pH-meter 654	Metrohm
Gene Pulser® II and Pulse Controller Plus	Bio-Rad
Microplate Sunrise-Basic™ reader	Tecan
Microplate Infinite® 200	Tecan
HERAsafe KS 18 Class II Safety Cabinet	Thermo Scientific
T3000 Thermocycler	Biometra
NanoPhotometer® P300 with Pearl Submicroliter Cell	Implen
LightCycler® 480 II	Roche

Table 9. Overview of consumables used in this work.

Consumable	Producer
Nunc™ EasYFlask™ 25 cm^2	Thermo Scientific
Bio-One Round Bottom Polypropylene Culture Tube	Greiner
Centrifuge tubes (15 and 50 ml)	Greiner
Nunc Maxisorp flat-bottom 96-wells (ELISA)	Thermo Scientific
Inoculation loops	Greiner
SAPPHIRE microplate 96-wells (qPCR)	Greiner
CELLSTAR® serological pipettes (5, 10, 25, and 50 ml)	Greiner
Cotton swabs	Süsse Labortechnik
Whatman™ Gel Blot Paper	GE Healthcare
Amersham™ Protran® 0.2 µm NC nitrocellulose membrane	Cytiva
Cryovials (2 ml)	Biozym
SafeSeal reaction tubes (1.5 and 2 ml)	SARSTEDT
PCR SingleCap 8er-SoftStrips	Biozym
Nalgene Syringe filter Nylon (0.2 µm)	Thermo Scientific
Bottle-top vacuum filter (0.22 µm)	Corning
Coverslips (IFM)	Roth
Microscopy slides StarFrost®	KNITTEL
Electroporation cuvettes (0.2 cm-gap)	Bio-Rad
Cuvettes (polystyrol)	SARSTEDT
Petri dishes	Greiner
Poly-Prep® Chromatography Columns	Bio-Rad

2.8. Software and bioinformatic tools

Table 10. Overview of software and bioinformatic tools used in this work.

Software/tool	Application*	Source
NCBI Prokaryotic Genome Annotation Pipeline (PGAP)	Annotation of genome sequences	[132]
Rapid Annotation Subsystem Technology Tool kit (RASTtk)	Annotation of genome sequences	[133, 134]
Minimap2	Alignment of PacBio reads to genome sequences	[135]
Geneious Prime v2020.0.5	Alignment of PacBio reads to genome sequences using Minimap2	Dotmatics
progressiveMAUVE	Alignment of homologous regions in genome sequences	[136, 137]
JSpecies Web Server	Genome comparison as per average nucleotide identity (ANI)	[138, 139]
Phage Search Tool Enhanced Release (PHASTER)	Prediction of prophage sequences in genome sequences	[140, 141]
VisiView v2.0.5	Analysis of IFM images	Visitron System
NIS-Elements BR v4.30.02	Analysis of DAPI-stained images	Nikon
Magellan v6	Measurement of ELISA colorimetric absorbance	Tecan
LightCycler® 480 v1.5.0	Analysis of qPCR measurements	Roche
ImageLab V6.0.1.	Analysis of WB images	Bio-Rad
HHpred	Prediction of remote homologues using PDB and pfam-A databases	[142–145]
Quick2D	Prediction of structural motifs in the BadA protein sequence	[146–148]
ClustalΩ	Alignment of BadA domain protein sequences	[144]
Expasy molecular mass (MW) tool	Prediction of the MW for protein sequences	Expasy (SIB)
NEB Tm calculator	Prediction of PCR annealing temperature	NEB
daTAA-server	Prediction of BadA domain protein structures	[102]
Cluster analysis of sequences (CLANS)	Prediction of pairwise BadA domain similarity	[149]
SnapGene	General analysis of genomes, genes, proteins, and *in silico* cloning steps	GSL Biotech
Photoshop CS6	Measurement of BadA fibre lengths	Adobe
GraphPad Prism v7.04	Statistical analysis	Dotmatics
Mendeley	Reference management	Elsevier

*All software and tools were run on default parameters unless noted otherwise.

2.9. Buffers and bacterial growth media

Table 11. Overview of buffers and solutions used in this work.

Buffer/solution	Components
SDS-PAGE running buffer (5X)	50 mM Tris
	384 mM glycine
	0.1 % (w/v) SDS
	in dH$_2$O
SDS-PAGE resolving gel (8 %)	1.3 ml Acrylamid/Bisacrylamid (37.5:1)
	1.3 ml 1.5 M Tris-HCl (pH 8.8)
	50 µl 10 % (w/v) SDS
	50 µl 10 % (w/v) APS
	3 µl TEMED
	2.3 ml dH$_2$O
SDS-PAGE stacking gel	335 µl Acrylamid/Bisacrylamid (37.5:1)
	250 µl 0.5 M Tris-HCl (pH 6.8)
	20 µl 10 % (w/v) SDS
	20 µl 10 % (w/v) APS
	2 µl TEMED
	1.4 ml dH$_2$O
Towbin transfer buffer (10X)	25 mM Tris
	192 mM glycine
	in dH$_2$O
Towbin transfer buffer (1X)	100 ml Towbin transfer buffer (10X)
	200 ml cold methanol
	700 ml dH$_2$O
WB washing buffer (10X)	10 mM Tris-HCl (pH 7.4)
	150 mM NaCl
	0.5 % (v/v) Tween 20
	in dH$_2$O
WB blocking buffer	5 % (w/v) milk powder
	in WB washing buffer (1X with dH$_2$O)
TAE buffer (50X)	2 M Tris (pH 8)
	50 mM EDTA
	1 M acetic acid
	in dH$_2$O
ELISA washing buffer	0.05 % (v/v) Tween 20
	in PBS (1X)
ELISA blocking buffer	2 % (w/v) BSA
	in ELISA washing buffer
Phosphate Buffer (200 mM)	2.022 g Na$_2$HPO$_4$.7H$_2$O
	0.339 g NaH$_2$PO$_4$.H$_2$O
	in 50 ml dH$_2$O (pH 7.4) and filter sterilised (0.2 µm)

Table 12. Overview of bacterial growth media used in this work.

Growth medium	Components
LB agar medium	40 g of premade mixture in 1 L dH$_2$O (autoclaved)
LB medium	25 g of premade mixture in 1 L dH$_2$O (autoclaved)
BALI medium	24.45 g Schneider's Drosophila Medium
	50 g sucrose
	0.79 g CaCl$_2$.2H$_2$O
	0.4 g NaHCO$_3$
	in 1 L dH$_2$O (pH 6.5 NaOH)
	filter sterilised (0.2 μm)
	100 ml heat-inactivated (fibronectin-depleted) FCS
Recovery medium (1 ml)	10 μl HEPES buffer
	10 μl sodium pyruvate
	50 μl FCS
	50 μl rabbit blood lysate
	in 880 μl RPMI-1640 and filter sterilised (0.2 μm)
Columbia agar with 5 % (v/v) sheep blood)	Becton-Dickinson

Fibronectin-depleted FCS was prepared by incubating heat-inactivated (30 min at 56 °C) FCS with PBS-washed gelatine sepharose-4B overnight at 4°C while gently shaking. Fibronectin-bound gelatine sepharose-4B was subsequently removed using Poly-Prep® Chromatography columns and filter-sterilised fibronectin-depleted FCS aliquots were stored at -20 °C. Rabbit blood lysate was prepared by incubating rabbit blood in cold dH$_2$O (1:4) for 10 min at room temperature (RT) while gently shaking. After centrifugation at 3500 x *g* for 10 min at 4 °C, supernatant was filter sterilised and aliquots were stored at -20 °C.

2.10. Bacterial growth conditions

2.10.1. Cultivation and handling of *B. henselae*

B. henselae strains were cultivated in *Bartonella* liquid (BALI) medium while shaking at 120 rotations per minute (rpm) for 3 days or on Columbia blood agar (CBA) plates for 5 days, both in a humidified atmosphere at 37 °C with 5 % CO$_2$. Liquid medium cultures were inoculated with *B. henselae* (ca. 1.5 x 10^7 bacteria/ml) from cryostocks or with cultivated *B. henselae* from CBA agar plates. Fibronectin-depleted FCS was used in BALI medium for the cultivation of *B. henselae* strains that were implemented in fibronectin binding experiments. In case of transformed *B. henselae* strains (Table 1 and Table 2), required selection markers (Kan and Gen) were equally distributed on the CBA plates or added to the BALI medium before

inoculation. Single colonies of CBA-cultivated *B. henselae* strains were collected after 14 days of incubation.

Cultivated *B. henselae* strains used in experiments were collected either from CBA plates using cotton swabs or from BALI medium by centrifugation at 3,500 x *g* for 10 minutes (min) at 4 °C. Bacteria are washed three times by centrifugation in PBS and the approximate bacterial number is determined via optical density (OD) measurements at a wavelength of 600 nm (OD_{600}) in disposable cuvettes on a NanoPhotometer® P300. An OD_{600} of 1 corresponds with a bacterial number of ca. 5×10^8 per ml.

To prepare cryostocks, bacteria were cultivated, collected, washed in LB medium, and stored at -80 °C in LB medium supplemented with 20 % glycerol. To determine the bacterial number, serial dilutions of cryostock *B. henselae* strains were cultivated on CBA plates and resulting colony forming units were counted.

2.10.2. Cultivation of *E. coli*

E. coli cultures were inoculated with strains from cryostocks or with cultivated single colonies from LB agar plates. *E. coli* strains were cultivated overnight at 37 °C in LB medium while shaking (180 rpm) or on LB agar plates. In case of transformed *E. coli* strains (Table 1 and Table 2), required selection markers (Kan, Gen, Spec, and Amp) were equally distributed on the LB agar plates or added to the LB medium before inoculation. Cryostocks were prepared in LB medium supplemented with 20 % glycerol and stored at -80 °C.

2.11. Molecular biological methods

2.11.1. Isolation of vector DNA

Vector DNA was isolated from LB medium-cultivated *E. coli* (10 ml) using a NucleoSpin® plasmid kit following the manufacturer's guidelines for low-copy plasmids. The elution step was performed at 70 °C and with an elution volume of 30 µl.

To obtain a higher yield and concentration for transformations steps, vector DNA was isolated from LB medium-cultivated *E. coli* (600 ml) using a NucleoBond® Xtra Maxikit following the manufacturer's guidelines for low-copy plasmids. The overnight *E. coli* culture (600 ml) was inoculated by adding 600 µl of a small starter culture (5 ml). The volumes of the resuspension buffer, lysis buffer, and neutralisation buffer were increased (40-45 ml) according to the OD_{600} of the overnight *E. coli* culture. Eluted vector DNA was precipitated by the addition of room-

temperature 2-propanol and subsequent centrifugation at 10,000 x g for 45 min at 4 °C. Precipitated vector DNA was washed with room-temperature ethanol (70 %) at 10,000 x g for 10 min at RT. Air-dried vector DNA was resuspended in 150 µl nuclease-free H_2O and stored at 4 °C. Final concentrations of isolated vector DNA (1 µl) were measured on a NanoPhotometer® P300 using a Pearl Submicroliter Cell.

2.11.2. Isolation of genomic DNA

gDNA was isolated from BALI medium-cultivated *B. henselae* strains by boiling at 95 °C for 15 min in nuclease-free H_2O. After centrifugation at 10,000 x g for 5 min at 4 °C, gDNA was resuspended in 250 µl nuclease-free H_2O. High molecular weight (HMW) gDNA was isolated from BALI medium-cultivated *B. henselae* strains with an OD_{600} of 4 using a MagAttract® HMW DNA Kit following the manufacturer's guidelines. To obtain higher concentrations, isolated HMW gDNA was eluted with 150 µl of elution buffer. To obtain higher yields, a second elution step was performed with a volume of 100 µl and pooled with the first elution. Amplified DNA fragments were purified from PCR reaction mixtures using a QIAquick® PCR Purification Kit following the manufacturer's guidelines. Amplified DNA fragments were purified from sliced agarose gel pieces using a MinElute Gel Extraction Kit via centrifuge processing following the manufacturer's guidelines. Final concentrations of isolated gDNA were measured on a NanoPhotometer® P300 using a Pearl Submicroliter Cell. The concentration of isolated HMW gDNA was measured using a Quant-iT™ PicoGreen™ dsDNA Assay Kit following the manufacturer's guidelines.

2.11.3. Amplification of DNA by polymerase chain reaction

DNA fragments from both gDNA and vector DNA were amplified *in vitro* by PCR using a HiFi Polymerase following the manufacturer's guidelines. The general PCR reaction mixture and PCR cycle protocol are given in Table 13 and Table 14, respectively. PCR reaction mixtures with a total volume of 50 µl were prepared on ice and centrifuged shortly to avoid bubbles. The annealing temperature of the PCR cycle protocol depends on the used primers and was calculated using the NEB Tm calculator software. The extension time depends on type of polymerase and the size (bp) of the amplified DNA fragment. In case of colony PCR, LB agar-cultivated *E. coli* single colonies were collected with an inoculation loop and added directly to the PCR mixture as template DNA.

Table 13. Overview of the general PCR reaction mixture.

PCR mixture component	Volume per PCR-reaction (µl)
PCRBIO reaction buffer (5X)	10
Forward primer (10 µM)	2
Reverse primer (10 µM)	2
Template DNA (< 100 ng)	2
PCRBIO HiFi polymerase (2 U/µl)	0.5
Nuclease-free H_2O	33.5

Table 14. Overview of the general PCR cycle protocol.

PCR step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	10 min	1
Denaturation	95	15 sec	
Annealing	55-65	30 sec	30
Extension	72	30 sec per kb	
Final extension	72	10 min	1
Cooling	12	∞	1

To amplify DNA fragments with primers that create overhangs, a two-step PCR cycle protocol was used in which the annealing temperature in the first 10 cycles was calculated using the part of the primer protein sequence that anneals to the template DNA. The annealing temperature in the subsequent 20 cycles was adjusted according to the complete primer protein sequence. To amplify the *B. henselae* 18-bp repeat regions, only 22 cycles for denaturation, annealing, and extension were used (Figure 6).

2.11.4. Amplification of DNA by real-time polymerase chain reaction

To quantify the *B. henselae* bacterial number, quantitative real-time PCR (qPCR) was performed using a Luna Universal qPCR Master Mix following the manufacturer's guidelines. The qPCR reaction mixture and qPCR cycle protocol are given in Table 15 and Table 16, respectively. qPCR reaction mixtures with a total volume of 20 µl were prepared on ice and centrifuged shortly to avoid bubbles. *B. henselae* strains were washed by centrifugation in nuclease-free water and added directly to the PCR mixture as template DNA. Logarithmic numbers of gene copy equivalents were calculated via an internal standard by amplifying a fragment (120 bp) of the housekeeping gene *glyA* (serine hydroxymethyltransferase) cloned in the pCR™2.1-TOPO® vector [150].

Table 15. Overview of the qPCR reaction mixture.

PCR mixture component	Volume per PCR-reaction (µl)
Luna Universal qPCR Master Mix (2X)	10
Forward primer (10 µM)	0.5
Reverse primer (10 µM)	0.5
Template DNA (< 100 ng)	5
Nuclease-free H_2O	4

Table 16. Overview of the qPCR cycle protocol including the melt curve.

PCR step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	15 min	1
Denaturation	95	15 sec	
Annealing	57	20 sec	40
Extension	60	30 sec	
Melt curve	60 to 90	various	1

2.11.5. Agarose gel electrophoresis

Amplified DNA fragments were size-separated by electrophoresis on a 1 % (w/v) agarose gel in TAE buffer (1X) for ca. 40 min at 85 V. Agarose was dissolved in TAE buffer (1X) at high temperatures, cooled down for 10 min at RT, and poured in gel casters to polymerise. Amplified DNA fragments and Quick-Load® 1 kb or 100 bp DNA ladders were mixed (1:10) with Midori Green Direct DNA dyes and pipetted in the pockets of the agarose gel. Size-separated DNA fragments were imaged by fluorescent lightning at a wavelength of 470-520 nm on a FastGene FAS V system.

2.11.6. Ligation of DNA fragments

Multiple amplified DNA fragments with overlapping ends were assembled via Gibson Assembly® technology for 30 min at 50 °C using three enzymes (5'-exonuclease, DNA polymerase, and DNA ligase) included in the Gibson Assembly® Master Mix. The required amount of pmols of each amplified DNA fragment was calculated based on DNA length and DNA mass. For each cloning step, 75 ng of amplified vector DNA was mixed with a 3- or 5-fold molar excess of amplified insert DNA, nuclease-free H_2O, and the Gibson Assembly® Master Mix (20 µl).

2.11.7. Heat-shock transformation of *E. coli*

Chemically competent *Escherichia coli* DH5α were transformed by heat-shock transformation following the manufacturer's guidelines. Bacteria were thawed on ice and 2 µl of assembled DNA product was added. After incubating on ice for 30 min, the mixture was heat shocked for 30 seconds (sec) at 42 °C in a warm water bath and immediately placed back on ice for 2 min. Heat-shocked *E. coli* DH5α were incubated in 1 ml of SOC medium for 1 h at 37 °C while shaking (250 rpm) and subsequently cultivated on LB agar supplemented with the required selection markers.

2.11.8. Electroporation of *B. henselae*

B. henselae strains were transformed by electroporation following a previously described protocol [126, 128]. *B. henselae* strains were cultivated on CBA plates and collected via cotton swabs in BALI medium. Electrocompetent *B. henselae* were prepared by washing three times with ice-cold 10 % glycerol in dH_2O by centrifugation. Ca. 1×10^{10} of electrocompetent *B. henselae* cells were mixed with 10 µg of purified vector DNA and 1 µl of TypeOne™ restriction inhibitor in pre-cooled electroporation cuvettes and incubated on ice for 15 min. After electroporation at 2.5 kV, 200 Ω, and 25 µF, samples were incubated in 1 ml of RT-recovery medium for 4 h at 37 °C in a humidified atmosphere with 5 % CO_2 while shaking (120 rpm). Transformed bacteria were subsequently incubated on CBA plates supplemented with the required selection markers.

2.11.9. Sanger sequencing and Pacbio sequencing

Short-read sequencing of isolated and purified DNA fragments was done via Sanger sequencing and performed by Microsynth Seqlab (Göttingen, Germany). Long-read sequencing of isolated and purified HMW gDNA was done via Pacific Biosciences (PacBio) single-molecule real-time (SMRT) sequencing and performed by the Norwegian Sequencing Centre, a national technology platform hosted by the University of Oslo, Norway. In short, HMW gDNA was sheared to ca. 10-12 kb fragments with g-TUBEs (Covaris) and the sequencing library was prepared following the PacBio protocol for SMRTbell™ libraries using PacBio® Barcoded Adapters for multiplex SMR® Sequencing. Library samples were size selected using 0.45x AMPure PB beads and sequenced in a single run with a movie and pre-extension time of 20 h and 4 h, respectively, on a PacBio Sequel instrument using v3.0 sequencing chemistry, Sequel polymerase v3.0, and an SMRT cell v3 LR tray. Reads were demultiplexed using a Barcoding pipeline on SMRT Link Analysis Services (v5.1.0.26412 and GUI v5.1.0.26411) with

a barcode score of ≥ 26. De novo genome assembly was performed using the HGAP 4 pipeline via SMRT Link Analysis Services (v6.0.0.47841) with an expected genome size of 2 Mbp, resulting in single contigs. Circular consensus sequences (CCS) reads were computed for the demultiplexed dataset with ≤ 1 as a number of passes and ≤ 0.9 as predicted accuracy

2.12. Protein biochemical methods

2.12.1. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

Bacterial proteins were separated based on MW via electrophoresis on an 8 % sodium dodecyl sulphate (SDS)-polyacrylamide gel. The resolving and stacking gel solutions were prepared as described in Table 11 and consecutively casted between a short glass plate and a spacer glass plate. Immediately after casting the resolving gel, 2-propanol was added on top of the resolving gel to avoid bubbles and was removed again after solidification of the gel. The stacking gel was casted on top of the solidified resolving gel and a comb was placed afterwards. Casted gels were placed in a vertical electrophoresis chamber and filled with SDS-polyacrylamide gel electrophoresis (PAGE) running buffer (1X).

Cultivated bacteria were washed three times by centrifugation in PBS, diluted to an OD_{600} of 1, incubated for 10 min at 98 °C in Laemmli sample buffer (1:1), and pipetted in the pockets of the SDS-PAGE gel. Bacterial proteins were first concentrated in the stacking gel for ca. 10 min by a constant current of 15 mA/gel after which the current was increased to 30 mA/gel until the bromophenol blue buffer front had completely migrated through the gel. A pre-stained protein ladder (10-250 kDa) was used as marker. After electrophoresis, certain SDS-PAGE gels were stained overnight in a Coomassie Brilliant Blue R-250 staining solution, destained by several washes in Coomassie Brilliant Blue R-250 destaining solution, and imaged on a ChemiDOC XRS+ system.

2.12.2. Western blotting

Bacterial proteins that were separated by SDS-PAGE were transferred to a nitrocellulose membrane (0.2 µm) by semi-dry Western blotting (WB) for 30 min at 25 V in Towbin transfer buffer (1X) between three layers of Whatman™ blot paper. Free binding places on the blotted nitrocellulose membranes were blocked with WB blocking buffer for 2 h at RT while gently shaking. After blocking, nitrocellulose membranes were incubated with primary antibodies diluted in WB blocking buffer overnight at 4 °C while gently shaking. Nitrocellulose membranes were washed three times for 5 min with WB washing buffer (1X) and incubated with secondary

antibody diluted in WB blocking buffer for 2 h at RT while gently shaking. After three washes in WB washing buffer (1X), nitrocellulose membranes were incubated with SuperSignal West Pico PLUS Chemiluminescent Substrates for 5 min at RT and subsequently imaged on a ChemiDOC XRS+ system with ImageLab software using different exposure times.

2.12.3. Isolation of BadA proteins

To isolate BadA proteins from the bacterial surface, BALI medium-cultivated *B. henselae* Marseille were vortexed thoroughly for 5 min and separated by centrifugation at 6,000 x *g* for 15 min at 4 °C. BadA and other large proteins were precipitated by incubating the supernatant with 20 % (w/v) PEG 6000 in dH$_2$O overnight at 4 °C while gently shaking and subsequently collected by centrifugation at 10,000 x *g* for 1 h at 4 °C. Isolated BadA and other large proteins were further separated by SDS-PAGE in a large single-pocket without preparative heating (see 2.12.1). Coomassie Brilliant Blue R-250 stained BadA proteins were precisely sliced from the top of the SDS-PAGE stacking gel and used as antigen (ca. 75 μg/injection) for the generation of rabbit anti-BadA antibodies (performed by Eurogentec).

Excised gel fragments were analysed by mass spectrometry (MS) to check for isolated BadA proteins. MS and preparative steps were performed by Sounak Chowdhury, PhD, from Lund University, Sweden (Suppl. Figure 1). In short, Coomassie Brilliant Blue R-250 stained SDS-PAGE gel fragments were first cut into small pieces and proteins were denatured by incubation in urea (8 M) and ammonium (100 mM) bicarbonate. Tris(2-carboxyethyl)phosphine hydrochloride (5 mM) was added to reduce the amount of disulphide bonds and samples were incubated in iodoacetamide (10 mM) for alkylation in a dark room. Samples were further diluted in ammonium bicarbonate (100 mM) after which sequencing-grade trypsin (0.5 mg/ml) was added to digest the proteins. Resulting peptides were analysed on a Q Exactive HF-X Quadrupole-Orbitrap Mass Spectrometer connected to an Easy-nLC 1200 System.

2.13. Immunological methods

2.13.1. Purification of antibodies by pre-adsorption

Generated rabbit anti-BadA antibodies were further purified by pre-adsorption to reduce unspecific background reactions in WB and microscopy imaging. In short, BALI medium-cultivated *B. henselae* Marseille ∆BadA-T were washed in PBS by centrifugation and ca. 5×10^8 pelleted bacteria were resuspended in and incubated with 100 µl of anti-BadA antibodies for 2 h at RT while shaking (900 rpm). Anti-BadA antibodies were subsequently isolated in the supernatant by centrifugation at 10,000 x *g* for 30 min.

2.13.2. ELISA

The ability of *B. henselae* to bind immobilised human plasma fibronectin or human collagen-I was assessed via enzyme-linked immunosorbent assays (ELISA) using anti-*B. henselae* antibodies. Nunc Maxisorp flat-bottom 96-wells were coated with 1 µg of fibronectin or collagen-I in 100 µl PBS overnight at 4 °C and subsequently blocked with 200 µl ELISA blocking buffer for 2 h at 37 °C.

BALI medium-cultivated *B. henselae* were washed three times with PBS by centrifugation and ca. 2.5×10^7 bacterial cells/well were added and incubated for 2 h at 37 °C. Bound bacteria were first incubated with 100 µl primary anti-*B. henselae* antibodies (1:1,000) and afterwards with 100 µl secondary HRP-conjugated swine anti-rabbit antibodies (1:2,000), both for 1 h at RT. Between each incubation step, unbound bacteria and antibodies were removed by three consecutive washes with 300 µl ELISA washing buffer. During each incubation step a protective seal was used to avoid evaporation. Assays were developed by incubation with 100 µl TMB liquid substrate for 1 min at RT and the reaction was stopped by the addition of 100 µl HCl (1 M). Colorimetric absorbance was measured on a microplate Sunrise-Basic™ reader at a wavelength of 450 nm. Assays were done in triplicate and negative controls include samples without the addition of bacteria or without prior fibronectin coating.

2.13.3. Antibody inhibition ELISA

The ability of anti-BadA-DALL antibodies to inhibit the binding of *B. henselae* S27 to human plasma fibronectin was assessed via ELISA using anti-fibronectin antibodies. BALI medium-cultivated *B. henselae* Marseille ∆BadA-T and S27 were washed three times with PBS by centrifugation and Nunc Maxisorp flat-bottom 96-wells were coated with ca. 5×10^7 bacteria in PBS overnight at 4 °C. Coated wells were blocked with 200 µl ELISA blocking buffer for 2 h at

37 °C and incubated with 100 µl of a specific dilution of anti-BadA-DALL antibodies in ELISA blocking buffer (0 µg/ml, 1 µg/ml, 10 µg/ml, or 100 µg/ml) for 1.5 h at 37 °C. Anti-BadA antibodies were similarly diluted and used as positive control. Afterwards, coated wells were incubated with 1 µg of human plasma fibronectin in 100 µl PBS for 1.5 h at 37 °C. Bound fibronectin was first incubated with 100 µl primary anti-fibronectin antibodies (1:1,000) and afterwards with 100 µl secondary HRP-conjugated goat anti-mouse antibodies (1:1,000), both for 1 h at RT. Assays were developed by incubation with 100 µl TMB liquid substrate for 1 min at RT and the reaction was stopped by the addition of 100 µl HCl (1 M). Colorimetric absorbance was measured on a microplate Sunrise-Basic™ reader at a wavelength of 450 nm. Assays were done in quintuplicates and negative controls include samples without the addition of fibronectin or without prior bacterial coating.

2.14. Microscopy

2.14.1. Immunofluorescence microscopy

The presence of BadA on the surface of *B. henselae* strains was assessed via immunofluorescence microscopy (IFM) using anti-BadA antibodies. BALI medium-cultivated *B. henselae* were washed three times by centrifugation in PBS, diluted to an OD_{600} of 1, air-dried on glass microscopy slides, and fixed with 40 µl 3.75 % (w/v) PFA for 10 min at 4 °C. Fixed bacteria were first incubated with 40 µl of primary anti-BadA antibodies in PBS (1:400) and afterwards with 40 µl secondary Alexa 488-conjugated goat anti-rabbit antibodies in PBS (1:200), both for 1 h at RT. Bacterial DNA was stained with a 40 µl DAPI solution in PBS (1 µg/ml) for 10 min at 4 °C. All incubation steps were performed in a humid chamber and followed by three washes with 25 ml PBS using serological pipettes. Microscopy slides were mounted by coverslips with fluorescence mounting media, air-dried overnight in dark chamber at RT, and sealed with nail polish. Prepared cover slips were imaged on an Axio Imager 2 microscope with a 63x objective (glycerol) and an exposure time of 50 ms and analysed using VisiView software.

In addition, the ability of the different *B. henselae* Marseille ΔBadA-T mutant strains, each expressing a truncated and modified *badA* gene, to bind immobilised human plasma fibronectin was assessed via IFM. Bacteria were processed as described in section 2.13.2 and fibronectin-bound bacteria were subsequently incubated with a 100 µl DAPI solution in PBS (1 µg/ml) for 15 min at RT and washed with ELISA washing buffer. Fibronectin-bound DAPI-stained bacteria (in PBS) were analysed via IFM on a Nikon Eclipse Ti microscope with a 20x objective and an exposure time of 4 sec using NIS-Elements BR software.

2.14.2. Confocal laser scanning microscopy

The presence of BadA on the surface of *B. henselae* Marseille ∆BadA-T mutant strains was assessed via confocal laser scanning microscopy (CLSM) using anti-BadA antibodies. Bacterial samples were processed similarly as described in section 2.14.1. Prepared cover slips were imaged on a Stellaris 8 System confocal microscope (Leica) using Las X software (v4.4.0). Samples were captured (65 % gain) with a 93x objective (glycerol) at an excitation and emission wavelength of 499 nm and 530-575 nm, respectively. CLSM images were adjusted *in silico* to a brightness of 50 % and a contrast of 10 %. CLSM imaging was performed by Daniela Bender, PhD, from the Paul-Ehrlich-Institut (Langen, Germany).

2.14.3. Transmission electron microscopy

The presence of BadA on the surface of *B. henselae* was assessed via transmission electron microscopy (TEM). BALI medium-cultivated *B. henselae*, inoculated by CBA plate-cultivated single colonies, were washed three times by centrifugation in PBS and fixed with 4 % PFA and 2.5 % glutaraldehyde in phosphate buffer for 90 min at RT and stored at 4 °C. Subsequent sample preparation and TEM imaging was performed by Katharina Hipp, PhD, from the Max Planck Institute for Developmental Biology (Tübingen, Germany).

Two different processing methods were used to obtain optimal visualisation of the BadA fibres. Accordingly, fixed samples were processed either by progressive lowering of temperature (PLT) in dimethylformamide (DMF) and embedding in Lowicryl K4M, or by high-pressure freezing (HPF) by freeze substitution (FS) and embedding in Epon.

In case of PLT in DMF, fixed bacterial samples were washed twice in phosphate buffer by centrifugation (1,000 x *g*), embedded in 12 % (w/v) melted (37 °C) gelatine (Merck), and subsequently cut into small cubes of ca. 1 mm^3. After a second fixation in 1 % (v/v) glutaraldehyde for 5 min at 4 °C, cubes were dehydrated by gradually increasing DMF concentration from 30 % (v/v) in dH$_2$O for 30 min at 0 °C, to 100 % (v/v) for 1.5 h at -35 °C. Afterwards, Lowicryl K4M was infiltrated at -35 °C and samples were polymerised by UV. In case of HPF/FS, fixed bacteria were cryo-fixed in cellulose capillaries in planchettes filled with 1-hexadecene in a high-pressure freezer (Compact 03, Wohlwend), freeze-substituted in a 2 % (w/v) osmium tetroxide/0.4 % (w/v) uranyl acetate solution in acetone, and subsequently embedded in Epon. Prepared samples were cut into ultrathin sections, stained with uranyl acetate and lead citrate, and analysed with a Tecnai Spirit electron microscope (Thermo Fisher Scientific) operated at 120 kV.

2.15. Statistics

Statistical analyses were performed on Prism V7.04 software using one-way ANOVA and assuming parametric data distribution.

3. Results

3.1. Analysis of the *B. henselae* genome

For the first part of this study and to characterise the difference in *badA* expression among various *B. henselae* strains, eight isolates with a passage number lower than 10 were selected from the laboratory's strain collection (Table 1) and prepared for sequencing. The exact passage number of all strains before arriving in Frankfurt am Main, Germany, is unknown. *B. henselae* Marseille, ATCC49882^T var-1, ATCC49882^T var-2, Berlin-I, G-5436, and 88-64 Oklahoma are isolates from human patients diagnosed with either CSD, HIV, or bacillary angiomatosis [29, 30, 80, 111, 121–123]. *B. henselae* FR96/BK38 and FR96/BK3 were originally collected from domestic cats [124, 125].

3.1.1. Overview of the *B. henselae* genome sequencing parameters

HMW input gDNA was isolated using magnetic beads and quality of the samples was verified via agarose gel electrophoresis and fluorescence absorption measurements (data not shown). The whole genome of eight *B. henselae* strains was determined via next-generation long-read PacBio SMRT sequencing and resulted in eight complete and single circular chromosomes. Library preparation, sequencing, initial quality control, and genome assembly was performed by the Norwegian Sequencing Centre.

The Phred quality (Q) score of circular consensus sequences (CCS) reads reflects the probability of an incorrect base call. For example, a Q score of 20 (Q20) is equivalent to the probability of an incorrect base call of 1 in 100 times or a base call accuracy of 99 %. CCS reads consist of aligned subreads that are generated during a sequencing pass around a circular template. CCS reads of all strains, except for Marseille and ATCC49882^T var-1, were filtered to have a score higher than Q20 and showed a Q score between Q32 (99.94 % accuracy) and Q34 (99.96 % accuracy). The unfiltered CCS reads of strains Marseille and ATCC49882^T var-1 showed an average accuracy of 98.7 % and 98.9 %, respectively. In conclusion, the overall high base call accuracy provides the quality necessary to confidently analyse genomic differences between the studied *B. henselae* strains. More detailed information and additional sequencing data is given in Table 17.

Table 17. Long-read sequencing parameters of the *B. henselae* genomes (adapted from [127]).

B. henselae strain	Genome size (bp)	Mean coverage[a]	Number subreads	Mean subread length (bp)	N50 subread length (bp)[b]	CCS reads (≥Q20)[c]	CCS mean read length (≥Q20[c], bp)	Number predicted CDSs
Marseille	1,906,759	802x	335,562	7,486	9,778	18,542	7,486	1,566
ATCC49882[T] var-1	1,955,459	1,113x	631,218	5,801	9,778	28,343	5,801	1,574
ATCC49882[T] var-2	1,931,585	651x	306,776	8,307	10,446	10,695	6,051	1,579
Berlin-I	1,931,655	652x	238,737	9,695	10,446	9,490	7,440	1,584
G-5436	1,963,796	629x	255,359	9,561	10,446	9,597	7,025	1,577
88-64 Oklahoma	1,969,298	566x	215,349	9,663	10,446	8,300	7,157	1,610
FR96/BK38	1,944,393	762x	371,120	8,161	10,446	12,787	8,792	1,561
FR96/BK3	1,935,288	922x	356,346	9,460	10,446	13,834	7,089	1,571

[a]Total #bases/genome size, [b]for entire SMRT cell, [c]no Q-filtered reads for strains Marseille and ATCC49882[T] var-1

3.1.2. Overview of the general *B. henselae* genome organisation

All analysed *B. henselae* genomes were oriented (sense) to start with the housekeeping gene *gltA* (citrate synthase) and show a consistent low GC-content of 38 %. *B. henselae* 88-64 Oklahoma and Marseille demonstrate the largest (1.97 Mbp) and smallest (1.91 Mbp) genome size, respectively. Whole-genome comparison as per average nucleotide identity (ANI) resulted in a high pairwise sequence identity ranging from 98.57 % between FR96/BK38 and FR96/BK3, to 99.99 % between Berlin-I and ATCC49882[T] var-2 (Table 18). Moreover, strains ATCC49882[T] var-1, ATCC49882[T] var-2, Berlin-I, G-5436, and 88-64 Oklahoma demonstrate a pairwise sequence identity of 99.92 % or higher. The GenBank accession numbers of the *B. henselae* strains are given in Table 1 and in the section *Data availability*.

Table 18. Overview of the pairwise *B. henselae* genome sequence identity. The upper number shows the whole-genome comparison as per average nucleotide identity (%), while the number in parentheses shows the proportion of aligned nucleotides (%). A grey scale gradient from white to dark grey (low to high) visualizes the percentage of pairwise genome sequence identity (adapted from [127]).

B. henselae strain	Marseille	ATCC49882^T var-1	ATCC49882^T var-2	Berlin-I	G-5436	88-64 Oklahoma	FR96/BK38	FR96/BK3
Marseille	/	98.83 (97.88)	98.84 (97.61)	98.84 (97.61)	98.84 (97.89)	98.83 (97.89)	98.82 (98.40)	98.91 (99.06)
ATCC49882^T var-1	98.84 (96.17)	/	99.99 (99.53)	99.98 (99.53)	99.98 (100.00)	99.92 (98.89)	99.37 (98.19)	98.60 (97.36)
ATCC49882^T var-2	98.85 (96.95)	99.99 (99.99)	/	99.99 (100.00)	99.99 (100.00)	99.92 (99.70)	99.37 (98.17)	98.61 (98.14)
Berlin-I	98.85 (96.94)	99.98 (99.99)	99.99 (100.00)	/	99.99 (100.00)	99.92 (99.70)	99.37 (98.17)	98.62 (98.14)
G-5436	98.84 (96.17)	99.99 (99.99)	99.99 (99.53)	99.99 (99.53)	/	99.92 (98.90)	99.39 (98.21)	98.61 (97.26)
88-64 Oklahoma	98.83 (96.31)	99.92 (99.96)	99.92 (99.72)	99.92 (99.72)	99.92 (99.97)	/	99.39 (97.23)	98.60 (98.38)
FR96/BK38	98.82 (96.99)	99.38 (98.45)	99.38 (97.98)	99.38 (97.98)	99.39 (98.45)	99.40 (97.45)	/	98.58 (97.46)
FR96/BK3	98.89 (97.75)	98.60 (97.74)	98.61 (97.40)	98.61 (97.40)	98.60 (97.74)	98.60 (97.75)	98.57 (97.58)	/

In more detail, most differentiations in genomic organisation among the different strains are present in the first ca. 400,000 bp as is visualised via a comparative genome alignment depicting homologous regions as similarly coloured blocks (Figure 2). Depending on the strain, the first ca. 400,000 bp region contains one or two potential prophage sequences and has been designated as a 'type II secretion system island' and predicted to consist of several phage genes [24, 32]. An additional ca. 60,000 bp large prophage region was identified in all strains (via PHASTER) and is located in a more conserved region (Marseille; nt start position 1,530,422).

From approximately nt position 400,000 on, the genomic backbone is highly collinear among all analysed *B. henselae* strains with the exception of a major inversion of two large adjacent regions (green and purple in Figure 2) in case of strains G-5436, 88-64 Oklahoma, and FR96/BK38. Both inversion breakpoints include a copy of the highly conserved *tuf* gene (elongation factor Tu; EF-Tu), flanked by either *fusA* (elongation factor G; EF-G) or genes encoding for ribosomal and transcription-related proteins. Both breakpoint regions showed a high sequencing coverage (via Minimap2) lowering the possibility of a faulty assembly. In conclusion, all strains show a similar genomic organisation except for a variable region of ca. 400,000 bp. Interestingly, a low similarity profile is observed at the position of *badA*, indicating a lower sequence similarity among the aligned *B. henselae* genomes.
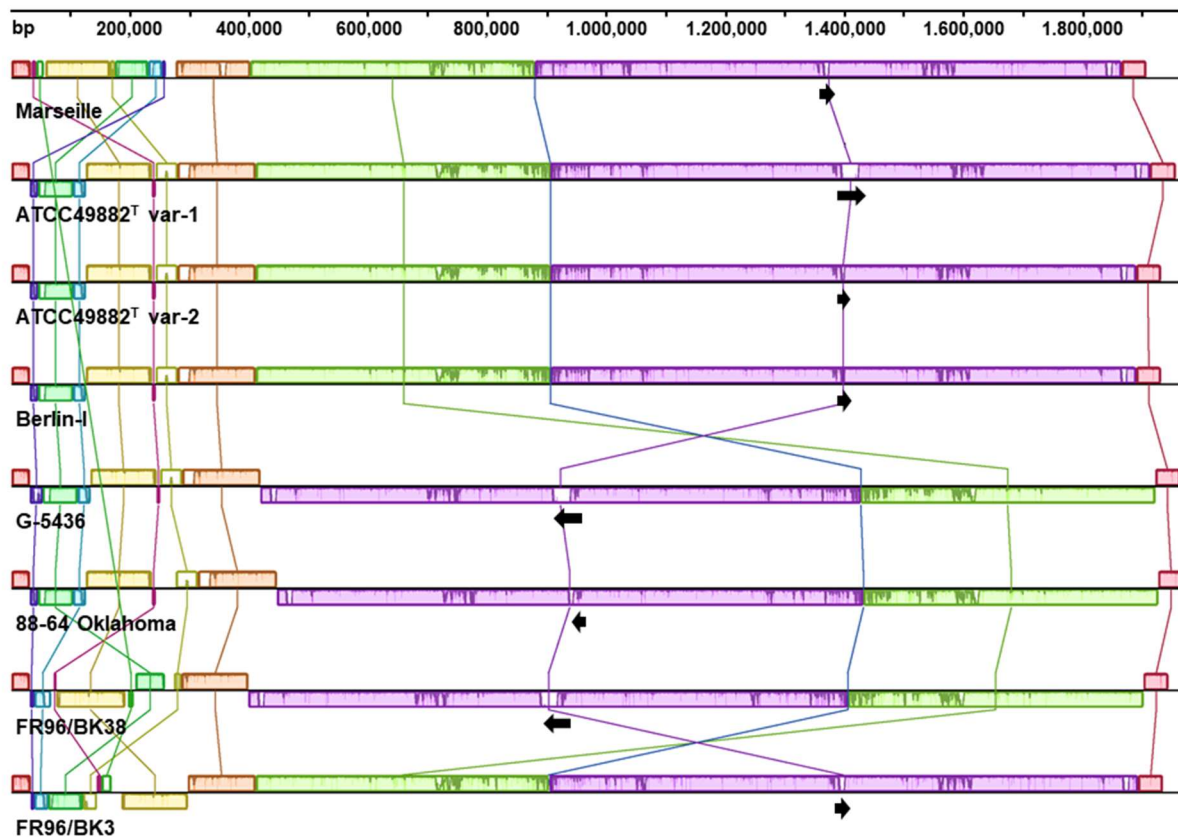
Figure 2. Multiple genome alignment of the *B. henselae* strains (adapted from [127]). The alignment shows a conserved *B. henselae* genome with few differences mainly located in the first ca. 400,000 bp. Genomes are shown as a black line. Coloured blocks and vertical lines depict localised collinear regions of the genome that align to part of another genome, that are homologous, and that are internally free from major genomic rearrangements. Blocks above or below the horizontal line are in the same or reverse complement orientation, respectively, compared to the reference genome of *B. henselae* Marseille. Inside each block, a similarity profile of the genome sequence is drawn. The height of this profile corresponds to the average level of conservation in that particular region. Regions outside blocks or white areas lack detectable similarity and might contain sequence elements unique for that strain. Black arrows indicate the length, orientation, and position of the respective *badA* island. The upper scale depicts the genome size (bp).

3.1.3. Overview of the *B. alsatica* IBS 382 (CIP 105477) genome sequencing parameters

Together with the eight *B. henselae* strains, an additional *Bartonella alsatica* strain IBS 382 (CIP 105477) was analysed via next-generation long-read PacBio SMRT sequencing and resulted in a single circular genome with a size of 1.66 kb and a GC-content of 36.85 % [151]. A total of 14,153 CCS reads were filtered ($\geq$ Q20) resulting in an average base call accuracy of 99.95% (Q33) with an average read length of 6,559 bp, a raw N50 read length of 7,814 bp, and a mean coverage of 1,503 times. The genome was likewise annotated via the NCBI PGAP and oriented (sense) to start with *glyA*.

The genome of *B. alsatica* contains various pathogenicity factors also found in other *Bartonella* species, such as a *badA*-homologue (11,850 bp), the VirB/D4-locus, the Trw-locus, and an *hbpA/pap31* gene. Moreover, the observed open reading frames (ORFs) flanking the *badA*-homologue are highly similar to those identified in *B. henselae*, including an upstream *badA* pseudogene (4,383 bp). These observations match the former classification of *B. alsatica* in the *Bartonella* evolution lineage 4 [13].

3.2. Analysis of the *badA* island and flanking regions

The *badA* island of *B. henselae* is located within a conserved genomic region (purple in Figure 2), however a low similarity profile is observed at the location of the *badA* island (black arrow in Figure 2) indicating a higher sequence variability among the studied *B. henselae* strains. The observed sequence variability of the *badA* island might be due to the influence of flanking genes and/or their corresponding proteins. In addition, the functional role of the *badA* genomic neighbourhood on the regulation of *badA* expression remains unexplored. In that regard, a selection of interesting flanking ORFs of the *badA* island is given in the following two chapters.

The *badA* island itself consists of the *badA* gene, four or five smaller ORFs, and one or two flanking *badA* pseudogenes depending on the *B. henselae* strain (Figure 3). Directly upstream of the *badA* gene, four or five shorter ORFs ($\leq$ 462 bp) were identified, in which the ORF directly upstream of *badA* is predicted to produce a surface protein sharing homology with BadA. The genome annotation of the *badA* island and flanking regions was manually checked and modified *in silico* (Table 19).
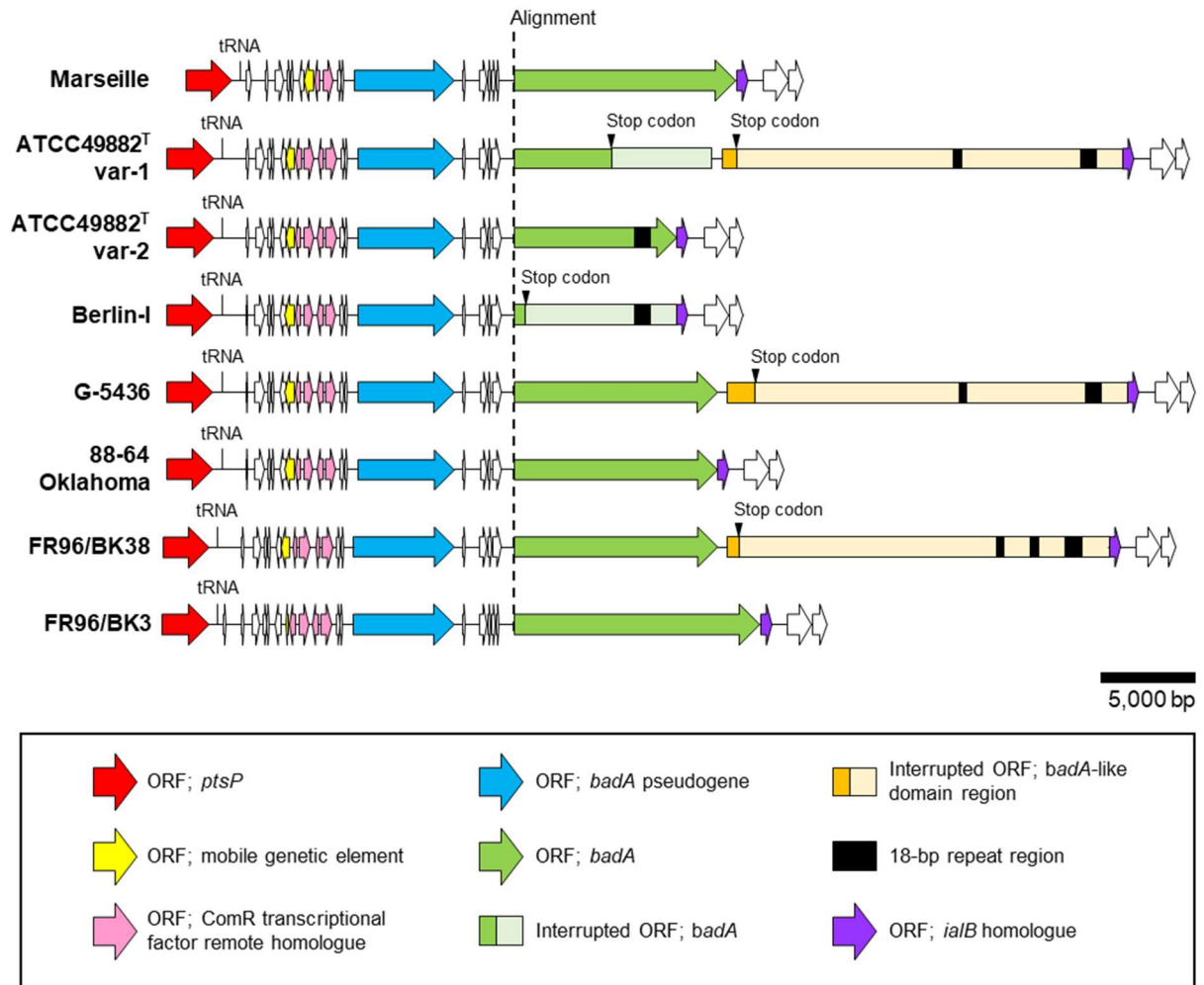
Figure 3. Comparative genomic organisation of the *badA* island and flanking regions (adapted from [127])**.** All genomic islands are shown in the same orientation. The *badA* island is flanked up- and downstream by a conserved region in which some ORFs were identified to produce for example a phosphoenolpyruvate-protein phosphotransferase (red), a mobile genetic element (yellow), a remote homologue of the ComR transcriptional factor (pink), or an IalB homologue (purple). The *badA* pseudogene (blue) is highly conserved and the corresponding protein sequence displays the typical TAA organisation. A higher pairwise sequence variability is found among the different *badA* genes (green), where a premature stop codon is observed in strains ATCC49882^T var-1 and Berlin-I. Strains ATCC49882^T var-1, G-5436, and FR96/BK38 contain an exceptionally long interrupted ORF, designated as the *badA*-like domain region (orange). The corresponding protein sequence contains the characteristic C-terminal anchor domain and numerous repeated neck/stalk domains. Five strains include one or more unique 18-bp repeat regions (black) in their respective *badA* gene or in the *badA*-like domain region.

3.2.1. Identification of the major genes upstream of the *badA* island

The *badA* island is flanked upstream by a conserved region of ORFs in which remote homologues from the corresponding protein sequences were predicted using HHpred software (Suppl. Table 1). For example, a conserved transcriptional repressor gene (504 bp) is located 23.5 kb upstream of the *badA* island and the corresponding protein sequence was predicted to be a remote homologue of a ferric uptake regulation protein (Fur) from *Francisella tularensis*, a Gram-negative gammaproteobacterium, or *Rhizobium leguminosarum bv. viciae*, a Gram-negative alphaproteobacterium. Other conserved ORFs upstream of the *badA* island were identified as (i) an efflux resistance-nodulation-cell division transporter permease subunit (3,135 bp), (ii) a glucose-6-phosphate isomerase (1,665 bp), (iii) two serine transfer (t)RNAs (90 bp), (iv) two large subunit ribosomal proteins (477 and 270 bp, respectively), and (v) a phosphoenolpyruvate-protein phosphotransferase (*ptsP;* 2,505 bp). Among the studied strains, only *B. henselae* Marseille, FR96/BK38, and FR96/BK3 show a limited number (< 29) of point mutations in any of the above-mentioned ORFs compared to the consensus sequence.

Closer to the *badA* island, an ORF (480 bp) was annotated as mobile genetic element (Figure 3; yellow ORF). Strain FR96/BK3 displays a shorter version of this ORF (201 bp) due to a point mutation causing a premature stop codon. The corresponding aa-sequence was predicted as a protein of the AAA+ ATPase (ATPase associated with diverse cellular activities) superfamily involved in DNA transposition, recombination-dependent replication, and/or DNA damage repair. Identified remote homologues of this mobile genetic element that showed the highest probability scores were (i) the MuB transposition protein from the bacteriophage Mu [152], (ii) the transposon Tn7 transposition protein (*tnsC*) from *E. coli* [153], (iii) the RuvB-like protein from *homo sapiens* [154, 155], and (iv) the replication-associated recombination protein A from *E. coli* [156].

Directly upstream of the *badA* island, four ORFs were identified whose corresponding protein sequences were all predicted to be remote homologous of a competence regulator (ComR) transcriptional factor of *Streptococcus vestibularis* (Figure 3; pink ORFs). However, strain Marseille is characterised by a 1.2 kb deletion immediately downstream of the mobile genetic element and therefore contains only two ORFs that correspond to the first two ComR-like transcriptional factors identified in the other *B. henselae* strains. The ComR transcriptional factor is involved in the adaptive response contributing to gene expression and genome plasticity [157]. Additional corresponding aa-sequences of shorter ORFs present in the upstream region (ca. 15 kb) of the *badA* island were identified as proteins with unknown function or as remote homologues of (trans)membrane proteins and phage-related transcriptional regulators.

3.2.2. Identification of the major genes downstream of the *badA* island

The region downstream of the *badA* island (ca. 15 kb) is likewise highly conserved among all studied *B. henselae* strains (Suppl. Table 2). Compared to the other *B. henselae* strains, minor genomic differences were observed in strains Marseille, FR96/BK3, and FR96/BK38 due to point mutations and the deletion and/or insertion of short DNA fragments (< 8 bp). Several corresponding proteins of annotated ORFs in that downstream region were identified as enzymes, for instance a guanosine pentaphosphate phosphohydrolase (1,311 bp), an RNA methyltransferase (744 bp), and a glutamate 5-kinase (1,140 bp). Other major ORFs in that same region were predicted as a guanosinetrifosfaat (GTP) binding protein (1,023 bp), a ribosomal silencing factor (444 bp), and a murein hydrolase activator (1,281 bp) [158–161].

The corresponding protein sequence of the ORF (573 bp; Marseille) located immediately downstream of the *badA* island is a homologue of the IalB family protein found among various *Bartonella* spp. [88, 162]. Due to a single bp deletion, strains Marseille and FR96/BK3 display a shorter ORF (573 bp) compared to the ORF (594 bp) identified in the remaining *B. henselae* strains (Figure 3; purple ORF). Interestingly, the originally identified *ialB* gene (561 bp) is located ca. 13 kb further downstream. IalB is an outer membrane protein involved in erythrocyte invasion [67, 88, 163, 164].

3.2.3. Identification of the *badA* pseudogenes

Among all analysed *B. henselae* strains, an ORF with a sequence length ranging from 5,181 to 5,430 bp is located upstream of the *badA* gene, is annotated as 'Bartonella autotransporter adhesin', and is herein designated as the *badA* pseudogene (Figure 3; blue ORF). The *badA* pseudogene shows a relatively high pairwise sequence identity (≥ 79.68 %), in which strains ATCC49882[T] var-1, ATCC49882[T] var-2, Berlin-I, G-5436, and 88-64 Oklahoma display an identical *badA* pseudogene (Table 20). Moreover, the final ca. 1.4 kb-region is identical among all studied strains. The corresponding aa-sequence displays the characteristic modular organisation of TAAs, including a head domain (375 aa; Marseille), a neck/stalk region (1,323 aa and 11 domains; Marseille), and an anchor domain (89 aa) [102]. However, only few regions of the *badA* pseudogenes align partly to regions in their respective *badA* genes, with the highest similarities observed in the anchor domain. Thus far, expression of the *badA* pseudogene is not yet examined nor demonstrated.

Table 19. Location and size of the major ORFs of the *badA* island in the *B. henselae* genomes (adapted from [127]).

B. henselae strain	*badA* pseudogene			*badA*			*badA*-like domain region		
	NCBI locus tag	Start position	Size (bp)	NCBI locus tag	Start position	Size (bp)	NCBI locus tag	Start position	Size (bp)
Marseille	KAE73_05700	1,353,514	5,361	KAE73_05715	1,362,098	11,922	/	/	/
ATCC49882[T] var-1	KAE76_05665	1,383,089	5,181	bapA; KAE76_05685	1,391,491; 1,396,757	5,266; 5,398	KAE76_05690 to KAE76_05770	1,402,734	21,532
ATCC49882[T] var-2	KAE74_05750	1,383,227	5,181	KAE74_05770; KAE74_05780	1,391,629	8,763	/	/	/
Berlin-I	KAE72_05775	1,383,299	5,181	KAE72_05790; KAE72_05800	1,391,701; 1,392,330	627; 8,133	/	/	/
G-5436	KAE71_04240	945,048	5,181	KAE71_04225	941,826	10,926	KAE71_04220 to KAE71_04150	930,320	21,532
88-64 Oklahoma	KAE77_04380	949,628	5,181	KAE77_04365	935,475	10,932	/	/	/
FR96/BK38	KAE75_04135	923,247	5,430	KAE75_04120	909,103	10,926	KAE75_04115 to KAE75_04045	908,530	20,510
FR96/BK3	KAE70_05705	1,379,217	5,430	KAE70_05720	1,387,870	13,224	/	/	/

Only three *B. henselae* strains (ATCC49882[T] var-1, G-5436, and FR96/BK38), contain an enormous region directly downstream of the *badA* gene, designated as the *badA*-like domain region (Figure 3). In case of strains ATCC49882[T] var-1 and G-5436, two distinct and single bp insertions are causing a frameshift mutation and subsequent premature stop codon in the otherwise perfect ORFs (21,532 bp) of their respective *badA*-like domain region. Similarly, a premature stop codon was observed in the otherwise perfect ORF (20,510 bp) of the *badA*-like domain region in strain FR96/BK38, due to a single bp deletion. The *badA*-like domain region of strain FR96/BK38 includes multiple mutations compared to the *badA*-like domain region of the other two strains showing a pairwise sequence identity of 88.60 %. The putative aa-sequence of all three *badA*-like domain regions contains a typical C-terminal anchor domain as well as a neck/stalk region that consists of 55 domains (for strains G-5436 and ATCC49882[T] var-1), however, no distinguishable *badA*-like head domain is observed. Expression of the *badA*-like domain region is not yet examined nor demonstrated.

Table 20. Pairwise sequence identity of the *B. henselae badA* pseudogene. The identity (%) is determined using a local alignment (Smith-Waterman) and is visualised with a grey scale gradient ranging from white to dark grey (lower to higher % identity).

B. henselae strain	Marseille	ATCC49882^T var-1	ATCC49882^T var-2	Berlin-I	G-5436	88-64 Oklahoma	FR96/BK38	FR96/BK3
Marseille	/	91.61	91.61	91.61	91.61	91.61	85.23	91.18
ATCC49882^T var-1	91.61	/	100.00	100.00	100.00	100.00	79.68	84.83
ATCC49882^T var-2	91.61	100.00	/	100.00	100.00	100.00	79.68	84.83
Berlin-I	91.61	100.00	100.00	/	100.00	100.00	79.68	84.83
G-5436	91.61	100.00	100.00	100.00	/	100.00	79.68	84.83
88-64 Oklahoma	91.61	100.00	100.00	100.00	100.00	/	79.68	84.83
FR96/BK38	85.23	79.68	79.68	79.68	79.68	79.68	/	89.70
FR96/BK3	91.18	84.83	84.83	84.83	84.83	84.83	89.70	/

3.2.4. Characterisation of the *badA* gene and corresponding BadA protein

Six out of eight analysed *B. henselae* genomes contain an uninterrupted *badA* ORF with a sequence length ranging from 8,763 bp (ATCC49882^T var-2) to 13,224 bp (FR96/BK3). Strains ATCC49882^T var-1 and Berlin-I include a premature stop codon in their putative *badA* gene due to a frameshift mutation caused by the deletion of a 262 bp region and a single bp, respectively (Figure 3). The *badA* promoter has been suggested to be located in a region ca. 250 bp upstream of the *badA* start codon site [111, 114].

The corresponding BadA protein sequences demonstrate the characteristic TAA domain organisation consisting of a passenger domain, including an N-terminal head domain and a long and repetitive neck/stalk region, and a C-terminal anchor domain [102]. While the head domain (≥ 97 % pairwise protein sequence similarity) and anchor domain (≥ 91 % pairwise protein sequence similarity) are highly conserved, numerous differences in the neck/stalk region were observed in regard to protein sequence length, similarity, and domain composition (Figure 4). The neck/stalk region is modularly built and consists of various repeated domains with a total number ranging from 18 domains in strains ATCC49882^T var-2 and Berlin-I to 34 domains in strain FR96/BK3. Certain neck/stalk domains show high pairwise similarities and are visualised by identical colours or patterns, with the exception of domains 14, 15, and 16 in strains ATCC49882^T var-2 and Berlin-I. Overall, the size of neck/stalk domains varies between 68 aa and 147 aa and are defined by their respective neck motif sequence.
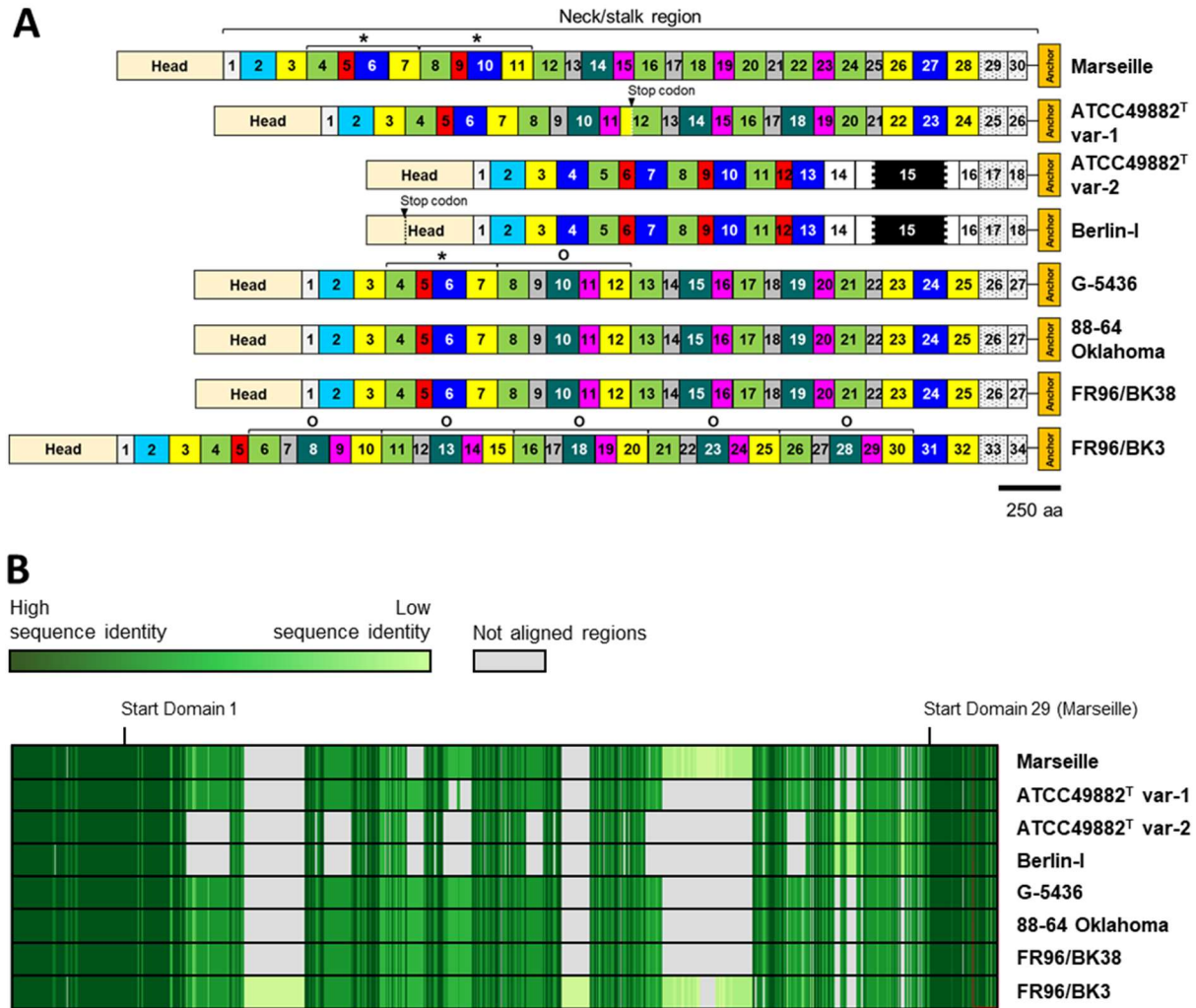
Figure 4. Schematic organisation and protein sequence alignment of the corresponding (and putative) BadA proteins (adapted from [127]). **(A)** The neck/stalk region of BadA is divided in different domains based on their respective neck sequence motif, in which identical coloured neck/stalk domains show a high pairwise protein sequence similarity. The BadA neck/stalk region is highly diverse among the studied strains. The presence of several repeated domain series within a single BadA protein as well as among the different BadA proteins suggests the occurrence of frequent recombination events. Two examples thereof are indicated by a star (*) and a circle (O). Due to a frameshift mutation, a premature stop codon is observed in the *badA* gene of strains ATCC49882[T] var-1 and Berlin-I. The black region within domain 15 of strains ATCC49882[T] var-1 and Berlin-I represents the 18-bp repeat region. **(B)** A multiple pairwise BadA protein sequence alignment (via MUSCLE) shows a conserved N-terminal region (first 1,000 aa) and C-terminal region (last 330 aa) with a higher diversity observed in the BadA neck/stalk region.

Three types of *badA* anchor domain sequences were identified among the studied *B. henselae* strains (Figure 5). Type 1 differs 37 and 44 bp with type 2 and 3, respectively, and is found in strains ATCC49882[T] var-1, G-5436, and FR96/BK38 that all contain a downstream *badA*-like domain region. Type 2 is observed in strains Berlin-I, ATCC49882[T] var-2, and 88-64 Oklahoma and is homologous to the anchor domain of the *badA*-like domain region in the strains that include a *badA* type 1 anchor domain. Type 3 differs 16 bp with type 2 and is observed in strains Marseille and FR96/BK3.

B. henselae	BadA anchor domain protein sequence	Type
	******:**:***********************:*:**:**::*:*********.*************:*::**.***:**:*:	
Marseille	FEALSYTVEEVRKEARQAAAIGLAVSNLRYYDIPGSLSLSFGTGIWRSQSAFAIGAGYTSEDGNIRSNLSITSSGGQWGVGAGITLRLK	3
ATCC49882[T] var-1	FEALSYTVEDVRKEARQAAAIGLAVSNLRYNDLPGKLSVALGSGIWRSQSAFAFGAGYTSEDGNIRSNLSVTNAGGSWGVGVGLSLTLN	1
ATCC49882[T] var-2	FEALSYTVAVEDVRKEARQAAAIGLAVSNLRYYDIPGSLSLSFGTGIWRSQSAFAVGAGYTSEDGNIRSNLSITNAGGHWGVGAGITLRLK	2
Berlin-I	FEALSYTVAVEDVRKEARQAAAIGLAVSNLRYYDIPGSLSLSFGTGIWRSQSAFAVGAGYTSEDGNIRSNLSITNAGGHWGVGAGITLRLK	2
G-5436	FEALSYTVEDVRKEARQAAAIGLAVSNLRYNDLPGKLSVALGSGIWRSQSAFAFGAGYTSEDGNIRSNLSVTNAGGSWGVGVGLSLTLN	1
88-64 Oklahoma	FEALSYTVAVEDVRKEARQAAAIGLAVSNLRYYDIPGSLSLSFGTGIWRSQSAFAVGAGYTSEDGNIRSNLSITNAGGHWGVGAGITLRLK	2
FR96/BK38	FEALSYTVEDVRKEARQAAAIGLAVSNLRYNDLPGKLSVALGSGIWRSQSAFAFGAGYTSEDGNIRSNLSVTNAGGSWGVGVGLSLTLN	1
FR96/BK3	FEALSYTVEEVRKEARQAAAIGLAVSNLRYYDIPGSLSLSFGTGIWRSQSAFAIGAGYTSEDGNIRSNLSITSSGGQWGVGAGITLRLK	3
BadA-like domain region	FEALSYTVAVEDVRKEARQAAAIGLAVSNLRYYDIPGSLSLSFGTGIWRSQSAFAVGAGYTSEDGNIRSNLSITNAGGHWGVGAGITLRLK	2

Figure 5. Comparative overview of the different BadA anchor domain types (adapted from [127]). A multiple protein sequence alignment (via MUSCLE) of the identified anchor domains in the putative BadA proteins and *badA*-like domain regions reveals the presence of three different anchor types (1, 2, and 3). Highlighted aa (grey) show differences among aligned sequences.

3.2.5. Analysis of an 18-bp repeat region

Five out of eight analysed *B. henselae* strains (ATCC49882[T] var-1, ATCC49882[T] var-2, Berlin-I, G-5436, and FR96/BK38) contain one or more regions of either 468 bp or 918 bp that strictly consist of a peculiar 18-bp long sequence [5'-GGA AG(C/T) AA(C/T) GG(C/T) A(G/A)T GGC-3']. Strains ATCC49882[T] var-1, G-5436, and FR96/BK38 show two or three 18-bp repeat regions in their respective *badA*-like domain region, while strains ATCC49882[T] var-2 and Berlin-I contain only one 18-bp repeat region in their respective *badA* gene (Figure 3 and Table 21). The authenticity of this sequence motif was verified via Sanger sequencing and PCR (Figure 6) using the primers Repeat_Fw and Repeat_Rv. Via PCR, a second 18-bp repeat region of ca. 500 bp was identified in strain ATCC49882[T] var-2, while only one 18-bp repeat region of 918 bp was initially observed via long-read sequencing technology.

Table 21. Overview of *B. henselae* strains that contain one or more 18-bp repeat regions.

B. henselae strain	18-bp repeat region size (bp)			Location
ATCC49882[T] var-1	468		918	*badA*-like repeat region
ATCC49882[T] var-2		918		*badA*
Berlin-I		918		*badA*
G-5436	468		918	*badA*-like repeat region
FR96/BK38	504	522	918	*badA*-like repeat region

The corresponding protein structure of the 18-bp aa-sequence [GSNG(N/S)G] might mimic collagen-like helixes which would fit the common TAA structure. Nonetheless, only strain ATCC49882[T] var-2 includes this unique repeat region in a membrane localised BadA protein. Future research is necessary to clarify the exact function and origin of this 18-bp repeat region.
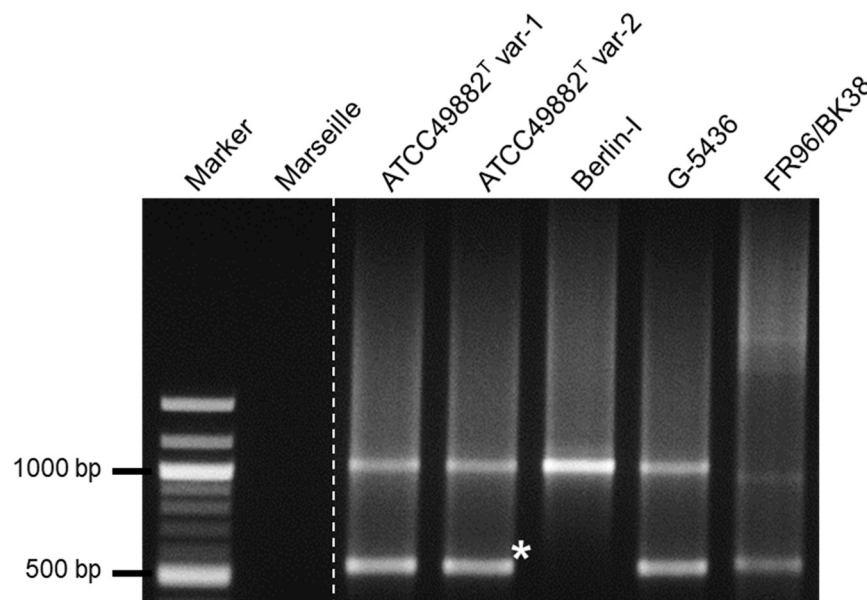
Figure 6. Analysis of the different 18-bp repeat regions. The presence and size (bp) of all 18-bp repeat regions is verified via PCR and agarose gel electrophoresis. However, a second and smaller (ca. 500 bp) 18-bp repeat region is identified in strain ATCC49882[T] var-2 (*) that is not observed in the long-read sequenced genome. Negative control strain Marseille does not contain an 18-bp repeat region. All strains were analysed on the same agarose gel in which the image was adjusted *in silico* (dotted line).

3.3. Generation of a *B. henselae* Marseille *badA* deletion mutant

To assess the role of BadA in binding ECM proteins and to verify the functionality of the *B. henselae* Marseille *badA*-deficient transposon mutant (∆BadA-T), an additional markerless *B. henselae* Marseille *badA*-deficient deletion mutant (∆BadA-D) was constructed via the principle of homologous recombination following a previously described two-step selection process [129, 150]. Vectors and primers are listed in Table 2 and Table 3.

Two flanking regions up- and downstream of the *badA* gene (from *B. henselae* Marseille) were amplified and ligated into the suicide vector pBIISK_*sacB*/*kanR* using Gibson Assembly®. The first fragment consists of the upstream region of *badA* (ca. 1 kb) including the first 30 bp of 3'-*badA* and was amplified using the primers *badA*FrUp_Fw and *badA*FrUp_Rv. The second fragment consists of the region downstream of *badA* (ca. 1 kb) including the last 30 bp of 5'-*badA* and was amplified using the primers *badA*FrDown_Fw and *badA*FrDown_Rv. Vector pBIISK_*sacB*/*kanR* was amplified using the primers pBIISK_Fw and pBIISK_Rv. The resulting vector pBIISK_*sacB*/*kanR*_UpDownBadA was multiplied in heat-shock transformed *E. coli* DH5α, selected on kanamycin-supplemented LB-agar plates, assessed via colony PCR and Sanger sequencing using the primers pBIISK_seq_Fw and pBIISK_seq_Rv, and finally used to transform electrocompetent *B. henselae* Marseille.

Electroporated *B. henselae* Marseille were selected on kanamycin-supplemented CBA plates and checked for successful integration of pBIISK_*sacB*/*kanR*_UpDownBadA in the gDNA (Figure 7A). It was demonstrated via colony PCR using the primers pBIISK_Fw, pBIISK_Rv, IntegrationA_Fw, IntegrationA_Rv, IntegrationB_Fw, and IntegrationB_Rv that the vector was integrated upstream of *badA*. Resulting colonies were subsequently transferred onto CBA plates supplemented with 10 % sucrose to select for segregation of the gDNA-integrated vector. Vector pBIISK_*sacB*/*kanR* carries a levansucrase gene (*sacB*) converting sucrose into levan, a toxin that destabilises the membrane.
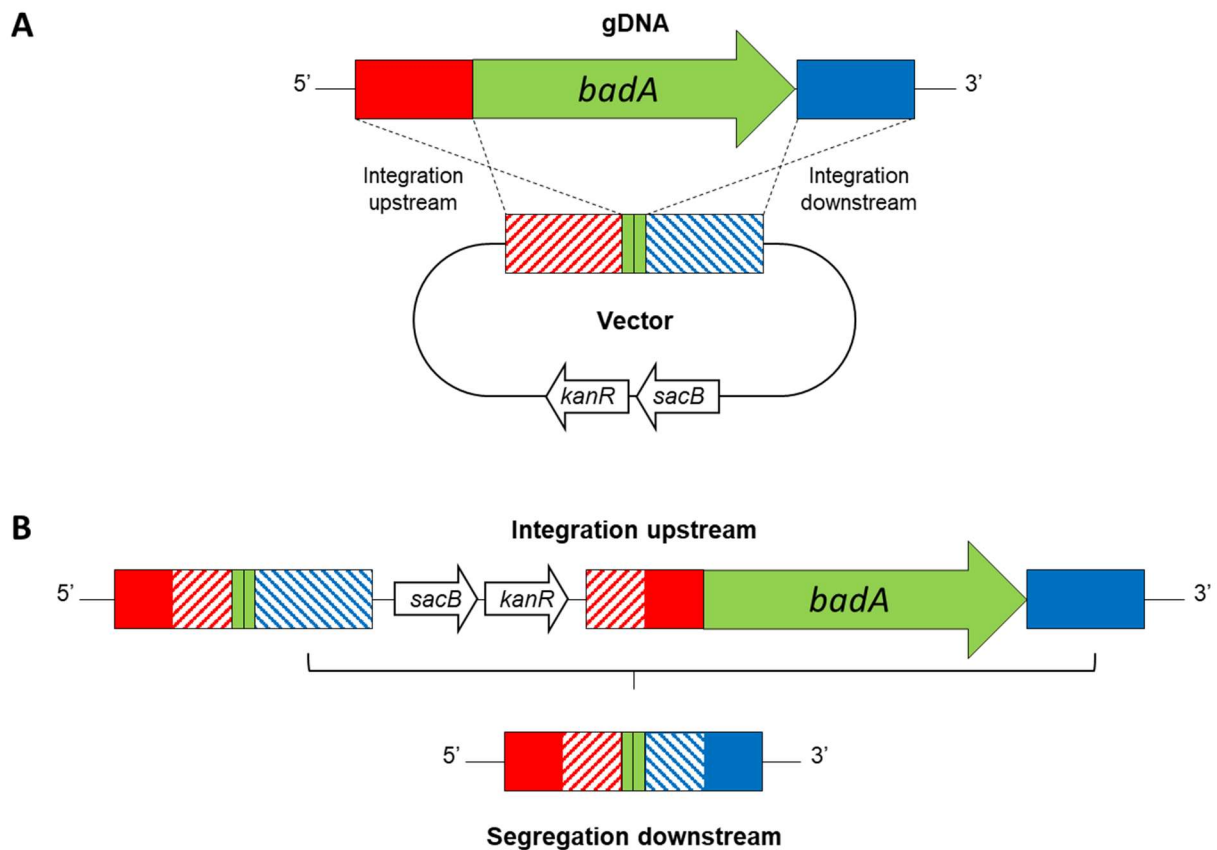
Figure 7. Schematic overview of the deletion process of *badA* in *B. henselae* Marseille via homologous recombination. (A) *B. henselae* Marseille is transformed with a suicide vector carrying the upstream (red) and downstream (blue) flanking regions of *badA*, 30 bp of the up- and downstream *badA* gene (green), and two selection markers *kanR* and *sacB*. Colonies with correct integration of the vector in the gDNA via the up- or downstream homologous region were selected using kanamycin. **(B)** It was demonstrated via colony PCR that the vector was integrated upstream of *badA* in the gDNA. Single colonies were subsequently selected using sucrose for downstream segregation of the vector together with the *badA* gene. The resulting gDNA of *B. henselae* Marseille ∆BadA-D contains a truncated (60 bp) non-functional *badA* gene.

Proper vector segregation is characterised by colonies that do not grow on kanamycin-supplemented CBA plates and was analysed via selection on CBA plates with and without supplemented kanamycin. Segregation of the integrated vector results either in the recovery of the original gDNA sequence including an intact *badA* gene, or in a truncated (60 bp) non-functional *badA* gene (Figure 7B). The successful deletion of the *badA* gene was assessed via Sanger sequencing using the primers IntegrationB_Fw, Segregation_Fw, and BadA1_Fw, via colony PCR using the primers IntegrationB_Fw and IntegrationA_Rv, and via Western blotting (WB) using anti-BadA antibodies (Figure 8).
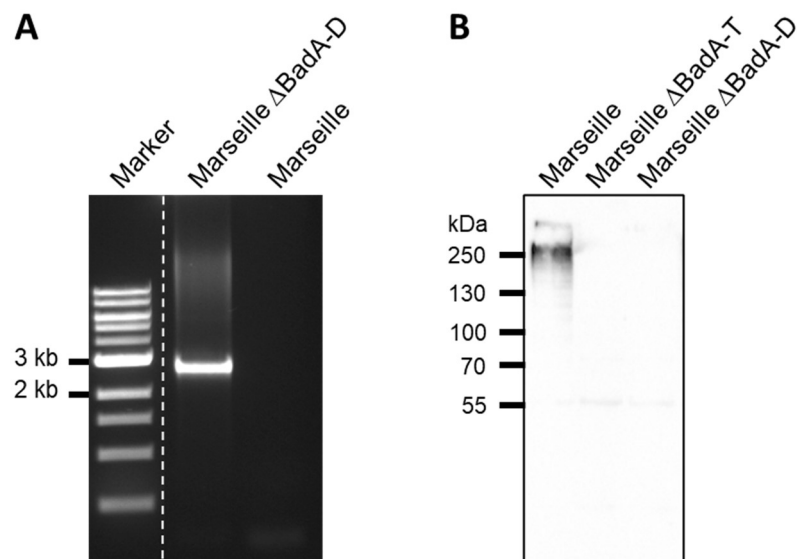
Figure 8. Analysis of the deletion of *badA* in *B. henselae* Marseille ∆BadA-D via colony PCR and Western blotting. (A) Deletion of *badA* was assessed via colony PCR using the primers IntegrationB_Fw and IntegrationA_Rv showing a ca. 2.5 kb band including the 1 kb up- and downstream region of *badA* and the truncated (60 bp) non-functional *badA* gene. Negative control *B. henselae* Marseille does not show a similar band. **(B)** It was demonstrated via WB using anti-BadA antibodies that *B. henselae* Marseille ∆BadA-D does not express *badA*. Likewise, negative control *B. henselae* Marseille ∆BadA-T does not express *badA,* as opposed to the wild type *B. henselae* Marseille.

3.4. Generation of anti-BadA antibodies and verification of *badA* expression via Western blotting

It has been demonstrated that the passage number of *B. henselae* influences the expression status of *badA* [76]. Therefore, anti-BadA antibodies were generated to check for *badA* expression and to localise BadA on the bacterial surface. Torn off BadA fibres were precipitated from the supernatant of *B. henselae* Marseille-incubated growth medium and subsequently isolated via gel electrophoresis (Figure 9). Excised gel fragments were verified to contain BadA proteins using MS (Suppl. Figure 1; performed by Sounak Chowdhury, PhD, from Lund University, Sweden) and were used as antigen for rabbit immunisation (performed by Eurogentec). To reduce unspecific binding, anti-BadA antibodies were further purified via pre-adsorption with *B. henselae* Marseille ∆BadA-T.
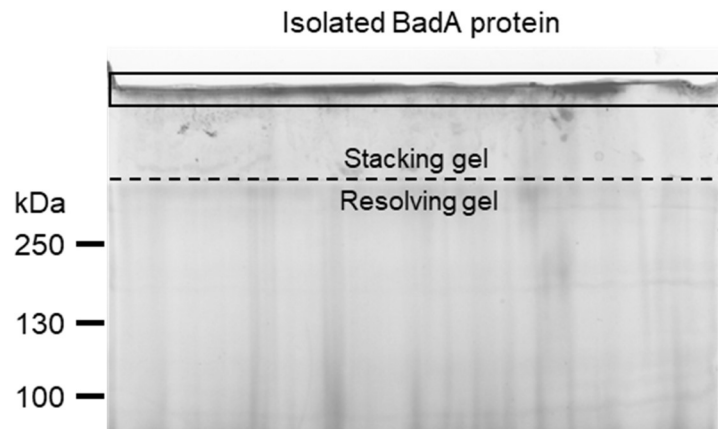
Figure 9. Isolation of BadA proteins via gel electrophoresis. Precipitated BadA fibres were collected and isolated via SDS-PAGE (8 % gel). Because of its large predicted trimeric protein mass (1,251 kDa), BadA was sliced from the top of the stacking gel (black frame), verified via MS, and used as antigen for rabbit immunisation. A pre-stained protein ladder (10-250 kDa) was used as marker.

Expression of *badA* and production of the corresponding BadA protein was analysed via WB using anti-BadA antibodies (Figure 10). Results are in accordance with the long-read sequencing data in which strains ATCC49882[T] var-2, Berlin-I, and both *badA*-deficient mutants (strains Marseille ΔBadA-T and Marseille ΔBadA-D) do not express *badA*. Predictions of the MW of the different monomeric BadA proteins range from 296 kDa (for strain ATCC49882[T] var-2) to 464 kDa (for strain FR96/BK3). Accordingly, strains Marseille and FR96/BK3 show a slightly larger monomeric BadA protein compared to the other strains. The largest detected protein, at the top of the stacking gel, is presumed to be trimeric BadA that is unable to travel down because of its enormous size, its heat stability, and its incomplete denaturation [165, 166]. Therefore, the limited resolution of SDS-PAGE analysis does not allow for a precise MW quantification of the trimeric BadA protein.
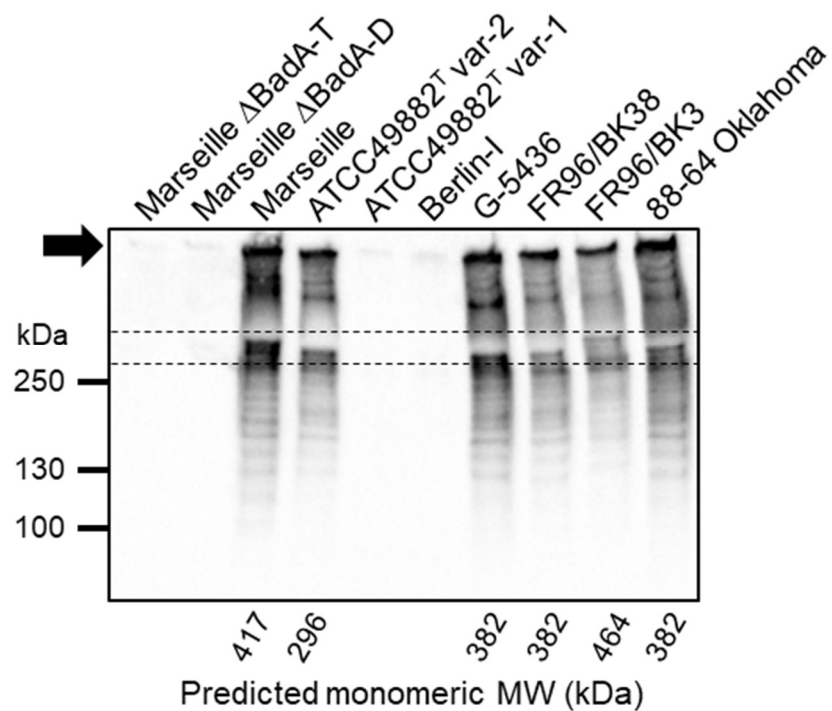
Figure 10. Analysis of *badA* expression via Western blotting using anti-BadA antibodies (adapted from [127])**.** The band between the dashed lines is considered monomeric BadA protein, while the uppermost band is presumably trimeric BadA protein stuck at the top of the stacking gel (black arrow). Strains Marseille and FR96/BK3 display a higher band between the dashed lines, corresponding to the predicted MW of the respective monomeric BadA protein (indicated underneath the blotted membrane). The numerous lower MW-bands, as well as the bands in between the presumed trimeric and monomeric proteins, are considered degradation product of the high MW BadA protein. Strains ATCC49882^T var-2 and Berlin-I contain a premature stop codon in their respective *badA* gene and do not produce BadA accordingly. Likewise, no BadA is detected for both *badA*-deficient mutant strains Marseille ΔBadA-T and Marseille ΔBadA-D.

3.5. Analysis of BadA on the bacterial surface via immunofluorescence and transmission electron microscopy

The presence of BadA on the bacterial outer membrane is essential for efficient host cell adherence in the course of infection and was therefore assessed via IFM using anti-BadA antibodies and additionally visualised via TEM. BadA proteins on the bacterial surface are characterised by a green halo surrounding the DAPI-stained intracellular DNA and was observed for strains Marseille, ATCC49882^T var-2, G-5436, 88-64 Oklahoma, FR96/BK38, and FR96/BK3. Strains ATCC49882^T var-1 and Berlin-I did not show such green halo because of a premature stop codon in their respective *badA* gene, preventing correct translation. Negative control strains Marseille ΔBadA-T and Marseille ΔBadA-D did not show any *badA* expression (Figure 11).
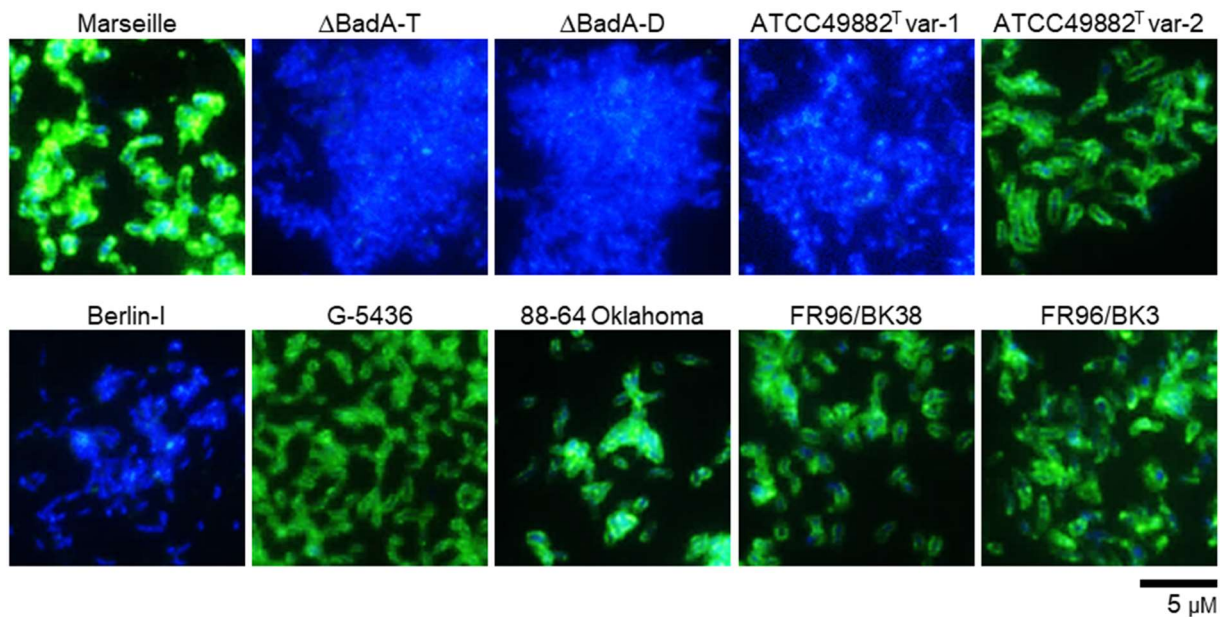
Figure 11. Analysis of surface exposed BadA in various *B. henselae* strains via immunofluorescence microscopy (adapted from [127])**.** The presence of BadA on the bacterial membrane was analysed via IFM using anti-BadA antibodies (green) in which bacterial DNA was stained with DAPI (blue). BadA is present in strains Marseille, ATCC49882[T] var-2, G-5436, 88-64 Oklahoma, FR96/BK38, and FR96/BK3, illustrated by a green halo. Strains ATCC49882[T] var-1, Berlin-I, and both negative control strains Marseille ΔBadA-T and Marseille ΔBadA-D do not express *badA* and do not show a green halo. Scale bar: 5 µm.

IFM results were supported by TEM imaging in which two different processing methods (PLT in DMF and HPF/FS) were used to obtain optimal visualisation of the BadA fibres. TEM was performed by Katharina Hipp, PhD, from the Max Plank Institute for Developmental Biology (Tübingen, Germany). All *B. henselae* strains with an intact *badA* ORF show a dense layer of BadA fibres along the entire bacterial surface. In contrast, strains ATCC49882[T] var-1, Berlin-I, and both negative control strains Marseille ΔBadA-T and Marseille ΔBadA-D show a smooth outer membrane (Figure 12). In conclusion, both microscopy approaches confirm the genomic data obtained via long-read sequencing in which the presence of BadA on the bacterial surface of the various *B. henselae* strains corresponds to their respective *badA* ORF.
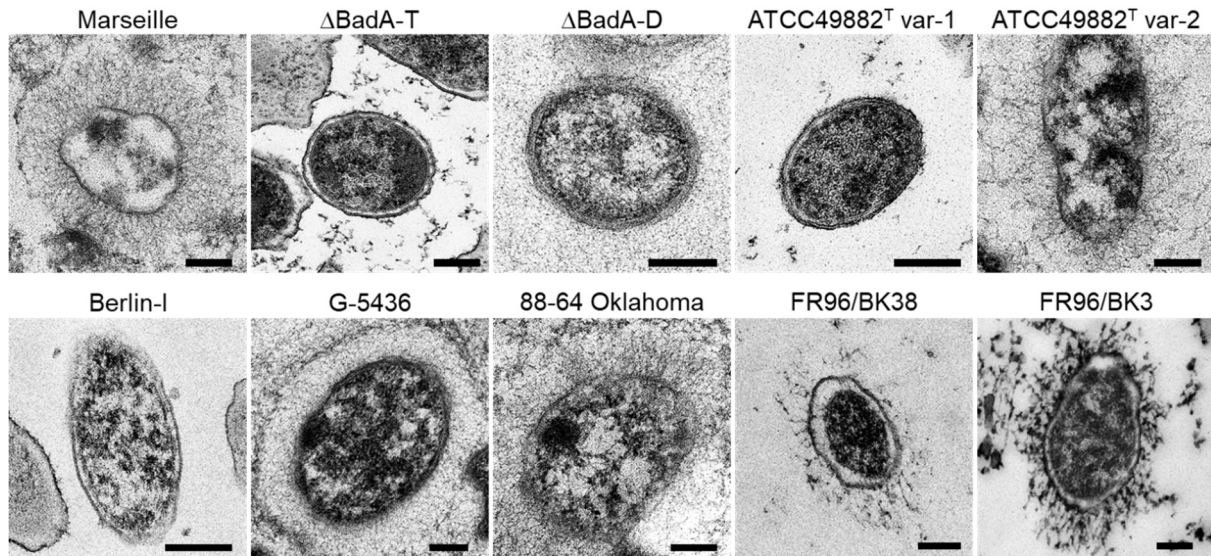
Figure 12. Analysis of BadA on the bacterial surface of various *B. henselae* strains via transmission electron microscopy (adapted from [127])**.** The presence of BadA on the bacterial surface was demonstrated via TEM. Surface-exposed BadA fibres are observed in the representative images for strains Marseille, ATCC49882^T var-2, G-5436, 88-64 Oklahoma, FR96/BK38, and FR96/BK3, but not for strains ATCC49882^T var-1 and Berlin-I, nor for the negative control strains Marseille ∆BadA-T and Marseille ∆BadA-D. For technical reasons, samples were prepared by PLT in DMF and K4M embedding (for strains Marseille, Marseille ∆BadA-D, ATCC49882^T var-2, G-5436, and 88-64 Oklahoma) or by HPF/FS and Epon embedding (for strains Marseille ∆BadA-T, ATCC49882^T var-1, Berlin-I, FR96/BK38, and FR96/BK3). Scale bars: 200 nm.

In addition, the average BadA fibre length from each *B. henselae* strain was measured using ca. 30 to 50 TEM images of bacterial cells per strain (Figure 13). Calculated BadA fibre lengths correspond with their respective BadA protein sequence lengths. For instance, strains Marseille and FR96/BK3 display the longest BadA fibres measuring 243 nm (3,973 aa) and 238 nm (4,407 aa) on average, respectively, while strain ATCC49882^T var-2 displays the shortest BadA fibres measuring 155 nm (2,920 aa) on average.
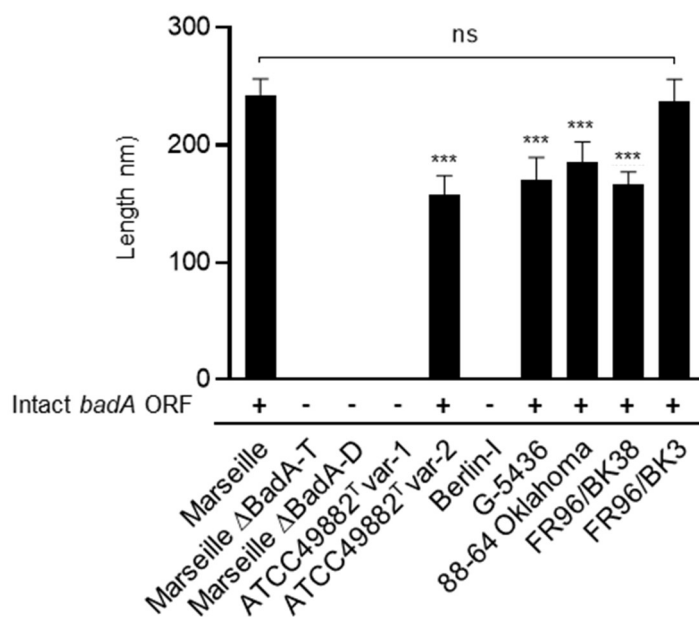
Figure 13. Overview of the average BadA fibre lengths via transmission electron microscopy images (adapted from [127]). To determine the phenotype and to measure the average length of bacterial surface-exposed BadA fibres of the various *B. henselae* strains, 30-50 TEM images of bacterial cells per *B. henselae* strain were used. The average BadA fibre lengths of strains ATCC49882[T] var-2 (157 nm), G-5436 (171 nm), 88-64 Oklahoma (186 nm) and FR96/BK38 (166 nm) are significantly shorter compared to those of strains Marseille (243 nm) and FR96/BK3 (238 nm). The BadA fibre lengths of the latter two strains do not show a significant difference (ns). Strains ATCC49882[T] var-1, Berlin-I, Marseille ∆BadA-T, and Marseille ∆BadA-D do not display BadA fibres on the bacterial surface. *B. henselae* strains are indicated to contain an intact *badA* gene (+) or not (-) based on long-read sequencing data. Statistical significance was determined using one-way ANOVA (***; $p < 0.001$).

3.6. Functional analysis of the fibronectin and collagen binding of *B. henselae*

The importance of *badA* expression on the binding ability of *B. henselae* to ECM proteins was evaluated via whole-cell ELISA using anti-*B. henselae* antibodies (Figure 14). In short, multiwell plates were coated with human plasma fibronectin or human collagen-I in which attached bacteria were identified using anti-*B. henselae* antibodies. Unbound bacteria and antibodies were removed by intermediate washes. Assays were done in triplicate and negative controls include samples without the addition of bacteria or without prior fibronectin coating. *B. henselae* strains with an intact *badA* ORF (Figure 3) and with BadA fibres on their respective bacterial surfaces (Figure 12) show a significantly higher binding to both fibronectin and collagen-I compared to the *badA*-deficient mutant strains ATCC49882[T] var-1, Berlin-I, and both negative control strains Marseille ∆BadA-T and Marseille ∆BadA-D. The presence of BadA on the bacterial surface is essential for binding to ECM proteins and is suggested to be independent of the strain-specific BadA length and neck/stalk domain organisation.
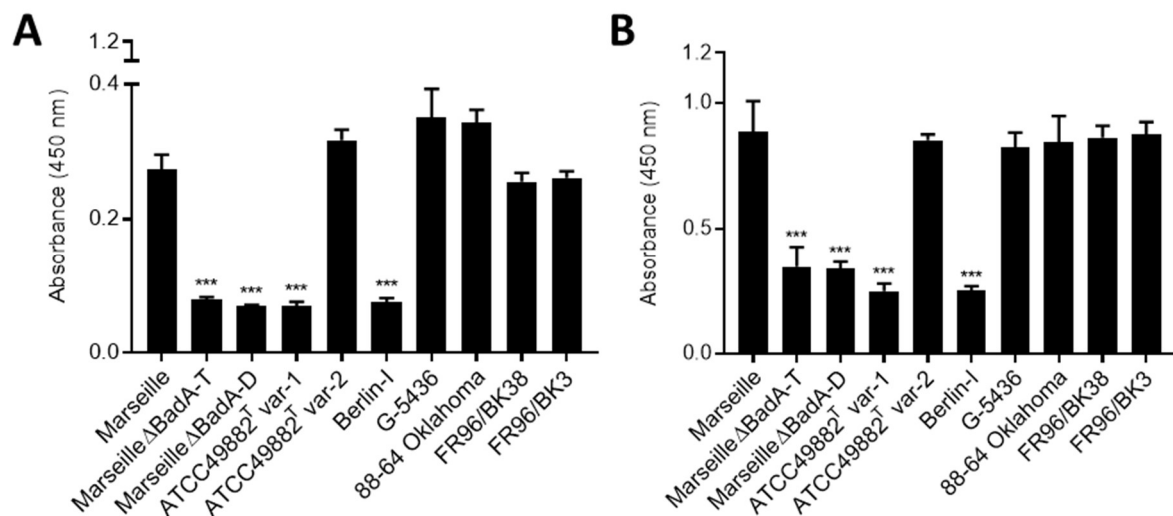
Figure 14. Analysis of the binding ability of *B. henselae* to ECM proteins via ELISA (adapted from [127])**.** Multiwell plates were coated with either **(A)** human collagen-I or **(B)** human plasma fibronectin in which attached bacteria were quantified (colorimetric absorbance measurements at 450 nm) via whole-cell ELISA using anti-*B. henselae* antibodies. Strains expressing *badA* show a significantly higher binding to fibronectin and collagen-I compared to strains lacking *badA* expression. Statistical significance was determined using one-way ANOVA testing (***; $p < 0.001$).

3.7. Characterisation and schematic domain organisation of the repetitive *Bartonella* adhesin A of *B. henselae* Marseille

For the second part of this study and to characterise the major fibronectin binding region(s) of BadA, a more detailed analysis of BadA was performed using the *badA*-expressing and laboratory-model strain *B. henselae* Marseille. The BadA neck/stalk region of all studied *B. henselae* strains, including strain Marseille, shows a modular structure consisting of various repeated domains defined by their respective neck motif sequence. The 30 BadA neck/stalk domains from strain Marseille were clustered into groups within a three-dimensional plot using the application CLANS to visualise their high pairwise protein sequence similarity (Figure 15A).
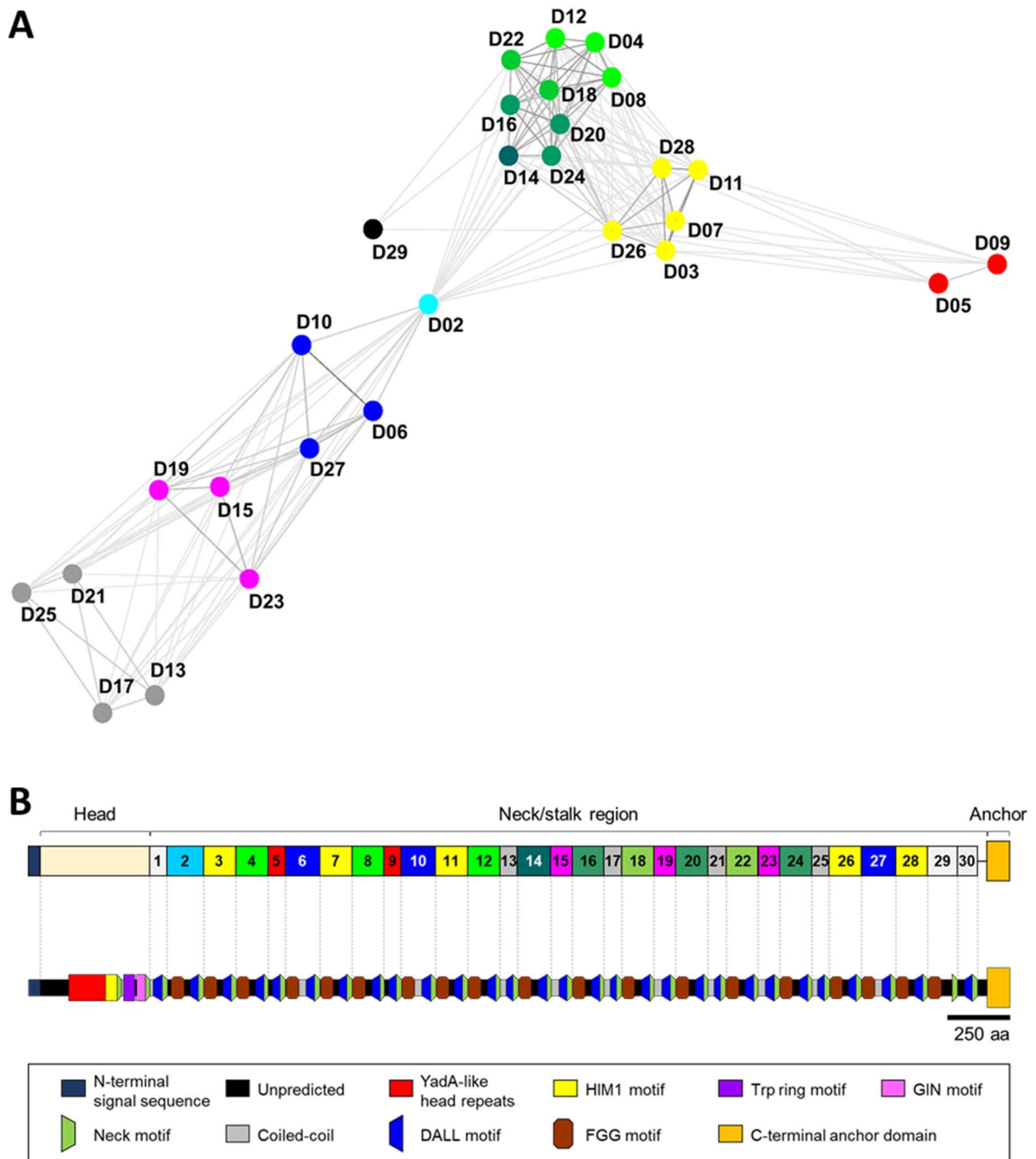
Figure 15. Analysis of the domain organisation and pairwise sequence similarity of BadA from _B. henselae_ Marseille (adapted from [128]). **(A)** BadA is modularly structured and is organised in 30 repetitive neck/stalk domains. The high pairwise domain similarity is visualised via a protein sequence similarity plot created in a three-dimensional space using the application CLANS. Neck/stalk domains with a high protein sequence similarity are displayed by identically coloured dots, are clustered together, and are connected by a darker line. Neck/stalk domains 1 and 30 are not included because of their higher divergent domain sequence. **(B)** BadA domains can be further organised into shorter sequence motifs using the daTAA server. The signal sequence is cleaved off during transport into the periplasm. The head domain includes YadA-like head repeats, a HIM1 motif, a Trp ring motif, and a GIN motif. The long and repetitive neck/stalk region is organised in a recurring pattern including FGG motifs, coil-coiled motifs, and DALL-neck tandem connectors. BadA images are drawn to scale according to aa-sequence length. Scale bar: 250 aa.

Repetitions of certain motif sequences are observed throughout BadA in which domains can be further organised into shorter sequence motifs using the 'domain dictionary' approach of the daTAA server (Figure 15B) [102, 104]. The prototypical neck/stalk domain consists of an FGG motif sequence, a coiled-coil motif sequence, a DALL motif sequence, and a final neck motif sequence. FGG motifs are defined by their characteristic 'FGG' (type 1) or 'LGG' (type 2) aa-sequence. A total of 19 FGG motifs were observed in the BadA protein of strain Marseille and only domains 6, 10, and 27 include a type 2 FGG motif. FGG and neck motifs are consistently followed by a coiled-coil motif that mostly consists of a heptad repeat in which the seven aa positions are designated 'abcdefg', with a hydrophobic residue on positions 'a' and 'd' and hydrophilic residues on positions 'b', 'c', 'e', 'f', and 'g'. A previously resolved crystal structure of the BadA head domain including part of the neck/stalk region has confirmed the presence of a coiled-coil segment in domain 1 [105]. All neck/stalk domains include a DALL-neck tandem connector at their C-terminal end. Three DALL motif variants are identified with signature sequences 'DSAV', 'DALL', and 'DSLV'. Neck motifs appear as either long (22 aa) or short (19 aa) variants. Long neck motifs are consistently preceded by a signature motif sequence 'DSAV' or 'DALL', while the signature sequence 'DSLV' is always followed by a short neck motif. The repetitive architecture of the BadA neck/stalk region is illustrated by aligning the domains according to their motif sequences (Figure 16)

Figure 16. Protein sequence alignment of BadA neck/stalk domains of *B. henselae* Marseille (adapted from [128])**. (A)** The repetitive and modular architecture of the BadA neck/stalk region is illustrated by aligning the neck/stalk domains according to their predicted motif sequences [102, 104]. Previously identified fibronectin-BadA interaction sites [73] are highlighted in light blue. Domains that were used to generate the *badA* mutants D16S28, D19S28, D25S28, and D27S29 are highlighted in red. **(B)** Two FGG motif subtypes are observed in 19 of 30 neck/stalk domains and are characterised by a signature '(F/L)GG' sequence (purple). **(C)** Motif sequences above the dotted line are predicted *in silico* to encode a coiled-coil and to comprise a heptad repeat with residue positions labelled as 'abcdefg'. Hydrophobic residues on the positions 'a' and 'd' are highlighted green and are consistently separated by two or three hydrophilic residues. Occasional hydrophilic residues on positions 'a' and 'd' are highlighted in grey. Sequence motifs below the dotted line are predicted *in silico* to encode for α-helix structures, yet show a similar heptad repeat. **(D)** The DALL-neck tandem connector is present at the end of each neck/stalk domain. Three variants of the DALL motif exist and are characterised by signature sequences 'DSAV', 'DALL', and 'DSLV' (purple). Neck motif sequences appear as either long (22 aa) or short (19 aa) variants and show a common signature motif (purple). **(B-D)** Framed sequences show the consensus sequence of each motif. Residues in upper case are present in all observed motif sequences, while residues in lower case represent the relatively most frequent.

A

Domain	FGG motif region	Coiled-coil	DALL motif region	Neck motif

```
1    --------------------------------------------------------------------------KKIETEVK--------EQVAA-SGFVK-QDSDT-KY-----L-TIGKDTDGDTINIANNKS-DKRTLTGIKEGDISKDSSEAITGSQL
29   ------YSLNEQ-------LL-TYFGGDAGYKDGQWIAPKFHVLQFKSDGSSGEKESYDNVAAAFEGV-----N-KSLA------GMNE--R-I-NNVTAGQNVSSSSLNWNE-TE--GG--YDA-R-HNGVDSK------LTHVENGDVSEKSKEAVNGSQL
30   --------------------------WNTN--------------EK----V-EAVE------KDVKNI----------E--KKVQD-IAT-V-ADSAV-KYEK--DSTGKKTNVIKLVG-GSESEPVL---IDNVADGKIEADSKQAVNGGQL

5    --------------------------------------------------------------------------DKG---L-KHLSDSL-QS-D-DSAVVHYDKKTDETGGINYTSVTLG-GKDKTPVA--LHNVADGSISKDSHDAINGGQI
9    --------------------------------------------------------------------------DKG---L-KHLSDSL-QS-D-DSAVVHYDKKTDETGGINYTSVTLG-GKDKTPVA---LHNVADGSISKDSHDAINGGQI
3    ------YSLGDK-------IA-SYLGGNAKYEDGEWTAPTFKVKTVKEDGKE-EEKTYQNVAEALTGVGTSFTNVKN----------EITKQI-NHL-QS-D-DSAVVHYDKNKDETGGINYASVTLGKGKDSAAV---TLHNVADGSISKDSRDAINGSQI
7    ------YSLGDK-------IA-SYLGGNAKYENGEWTAPTFKVKTVKEDGKE-EEKTYQNVAEALTGVGASFTNVKN----------EITKQI-NHL-QS-D-DSAVVHYDKNKDETGGINYASVTLGKGKDSAAV---TLHNVADGSISKDSRDAINGSQI
11   ------YSLGDK-------IA-SYLGGNAKYENGEWTAPTFKVKTVKEDGKE-EEKTYQNVAEALTGVGASFTNVKN----------EITKQI-NHL-QS-D-DSAVVHYDKNKDETGGINYASVTLGKGKDSAAV---TLHNVADGSISKDSRDAINGSQI
26   ------YSLNEQ-------LA-TYFGGGAKYENGQWTAPTFKVKTVNGEGKE-EEKTYQNVAAAFEGVGTSFTHVKN----------EITKQI-NHL-QS-D-DSAVVHYDK-DDKNGSINYASVTLGKGKDSAAVA--LHNVADGSISKDSHDAINGGQI
28   ------YSMSNM-------LA-TYLGGNAKYENGQWTAPTFKVKTVKEDGEE-EEKTYQNVAEALTGVGTSFTNIKS----------EIAKQI-NHL-QS-D-DSAVIHYDKNKDETGTINYASVTLGKGKDSAAVA--LHNVAAGNIAKDSRDAINGSQL
4    ------YSLNEQ-------LA-TYFGGGAKYENGQWTAPIFKVKTVKEDGEE-EEKTYQNVAEALTGVGTSFTNIKS----------EITKQIANEI-SSVTGDSLV-KKDLATNLI-TIGKEVAGTEINIASVSKADRTLS-GVK-EAVKDN-EAVNKGQL
8    ------YSLNEQ-------LA-TYFGGGAKYENGQWTAPIFKVKTVKEDGEE-EEKTYQNVAEALTGVGTSFTNIKS----------EITKQIANEI-SSVTGDSLV-KKDLATNLI-TIGKEVAGTEINIASVSKADRTLS-GVK-EAVKDN-EAVNKGQL
12   ------YSLNEQ-------LA-TYFGGGAKYENGQWTAPIFKVKTVKEDGEE-EEKTYQNVAEALTGVGTSFTNIKS----------EITKQIANEI-SSVTGDSLV-KKDLATNLI-TIGKEVAGTEINIASVSKADRTLS-GVK-EAVKDN-EAVNKGQL
18   ------YSLNEQ-------LA-TYFGGGAKYENGQWTAPTFKVKTVNGEGKE-EEKTYQNVAAAFEGVGTSFTNIKS----------EITKQIANEI-SNVTGDSLV-KKDLDTNLI-TIGKEIAGTEINIASVSKADRTLS-GVK-EAVNDN-EAVNKGQL
22   ------YSLNEQ-------LA-TYFGGGAKYENGQWTAPTFKVKTVNGEGKE-EEKTYQNVAAAFEGVGTSFTNIKS----------EITKQIANEI-SNVTGDSLV-KKDLDTNLI-TIGKEIAGTEINIASVSKADRTLS-GVK-EAVNDN-EAVNKGQL
24   ------YSLNEQ-------LA-TYFGGGAKYENGQWTAPTFKVKTVNGEGKE-EEQTYQNVAAAFEGVGTSFTNIKS----------EITKQINNEII-NVKGDSLV-KRDLATNLI-TIGKEIEGSVINIANKSGEARTIS-GVK-EAVKDN-EAVNKGQL
20   ------YSLNEQ-------LA-TYFGGGAKYENGQWTAPTFKVKTVNGEGKE-EEQTYQNVAAAFEGVGTSFTNIKS----------EITKQINNEII-NVKGDSLV-KRDLATNLI-TIGKEIEGSVINIANKSGEARTIS-GVK-EAVKDN-EAVNKGQL
16   ------YSLNEQ-------LA-TYFGGGAKYENGQWTAPTFKVKTVNGEGKE-EEQTYQNVAAAFEGVGTSFTNIKS----------EITKQINNEII-NVKGDSLV-KRDLATNLI-TIGKEIEGSVINIANKSGEARTIS-GVK-EAVKDN-EAVNKGQL
14   ------YSLNEQ-------LA-TYFGGGAKYENGQWTAPTFKVKTVNGEGKE-EEQTYQNVAEALTGVGASFMN-------VQNKITNEITNQVNN-AITKVEGDSLV-KQDNLG-II-TLGKERGGLKVDFANRDGLDRTLS-GVK-EAVNDN-EAVNKGQL
2    FTTNQNVKTVSDNLQTAATNIAKT-FGGGAKYEDGEWIAPAFKVKTVTGEGKE-EEKRYQNVADALAGVGSSITN-------VQNKVT-E---QVNN-AITKVEGDALL-WSDEANAFVARHEKS--LGKGASKATQENSKITYLLDGDVSKDSTDAITGKQL
6    -------HTIGED-------VAK-FLGGAASFNNGAFTGPTYKLSNIDAKG-DVQQSEFKDIGSAFAGL--D-TNIKNVNNNVTNKFN-ELTQNITN-VTQQVKGDALL-WSDEANAFVARHEKS--LGKGASKATQENSKITYLLDGDVSKDSTDAITGKQL
10   -------HTIGED-------VAK-FLGGAASFNNGAFTGPTYKLSNIDAKG-DVQQSEFKDIGSAFAGL--D-TNIKNVNNNVTNKFN-ELTQNITN-VTQQVKGDALL-WSDEANAFVARHEKS--LGKGASKATQENSKITYLLDGDVSKDSTDAITGKQL
27   -------HTIGED-------VAK-FLGGDAAFKDGAFTGPTYKLSNIDAKG-DVQQSEFKDIGSAFAGL--D-TNIKNVNNNVTNKLS-ELTQNITT-VTQQVKGNALL-WSDEANAFVARHEKS--LEKGASKAIQENSKITYLLDGDVSKGSTDAVTGGQL
15   --------------------------------------------------------DA-DISKVNNNVTNKFN-ELTQNITN-VTQQVKGDALL-WSDEANAFVARHEKS--LEKGVSKATQENSKITYLLDGDISKGSTDAVTGGQL
19   --------------------------------------------------------DA-NISKVNNNVTNKFN-ELTQSITN-VTQQVKGDALL-WSDEANAFVARHEKS--LEKGVSKATQENSKITYLLDGDISKGSTDAVTGGQL
23   --------------------------------------------------------DA-NISKVNNNVTNKFN-ELTQSITN-VTQQVKGDALL-WSDEANAFVARHEKS--LEKGVSKATQENSKITYLLDGDISKGSTDAVTGGQL
13   --------------------------------------------------------D-TNIKKV----EDKLT-EAV--G--KVTQQVKGDALL-WSNEDNAFVADHGKD--------S-A-KTKSKITHLLDGNIASGSTDAVTGGQL
17   --------------------------------------------------------D-TNIKKV----EDKLT-EAV--G--KVTQQVKGDALL-WSNEDNAFVADHGKD--------S-A-KTKSKITHLLDGNIASGSTDAVTGGQL
21   --------------------------------------------------------D-TNIKKV----EDKLT-EAV--G--KVTQQVKGDALL-WSNEDNAFVADHGKD--------S-A-KTKSKITHLLDGNIASGSTDAVTGGQL
25   --------------------------------------------------------D-TNIKKV----EDKLT-EAV--G--KVTQQVKGDALL-WSNEDNAFVADHGKD--------S-A-KTKSKITHLLDGNIASGSTDAVTGGQL
```

B

Type 1 FGG	YSlneqlAtYfFGGgAKYEnGqWTAPtFKVKTVngeGkEEEkTYQNVAeAltGV
Type 2 FGG	HTIGEDVAKFLGCaAsFnnGAFTGPTYKLSNIDAKGDVQQSEFKDIGSAFAGL

C

Domain	Coiled coil – heptad repeat
1	`      d   a  d` `KKTETEVKEQVAA`
30	`  a d   a d   a d   a d   a d` `EKVEAVEKDVKNIEKKVQDIATVADSAVKY`
6, 10, 27, 15, 19, 23	`   a d   a -a d   a -a d   a` `DtnIkkVNNNVTNKFnELTQnITnVTQQVKG`
13, 17, 21, 25	`   a d   a -a d   a` `DTNIKKVEDKLTEAVGKVTQQVKG`
	- - - - - - - - - - - - - - - - -
5, 9	`   a  d  a  d` `DKGLKHLSDSLQSD`
3, 7, 11, 26, 28	`  a d  a  d  a` `GtSETnvKnEItKQINHLQSD`
4, 8, 12, 18, 22, 24, 20, 16, 14	`  a  d  a -a d  a` `GTSETNIKSEITKQIaNEIsnVtG`

D

DALL 1	DSAVvHYDKnkDetGgINYaSVTLGkGkDsaaVa
DALL 2	dALLWSdEaNAFVArHeksklgKgaSkAtqenSK
DALL 3	DSLVKkDLatnlITiGKEiaGteiniAsvskadRT

Short neck	lS-GVK-EAVkDN-EAVNKGQI
Long neck	itnlldGdiskdStdAitGgQI

Figure 16. Protein sequence alignment of BadA neck/stalk domains of *B. henselae* Marseille.

3.8. Generation of modified *badA* genes expressed in a *B. henselae* Marseille *badA*-deficient transposon mutant

3.8.1. Overview of the design and construction of modified *badA* genes

To further investigate the role of certain individual BadA neck/stalk domains in their ability to bind human plasma fibronectin, *B. henselae* Marseille ΔBadA-T was transformed with various truncated and modified *badA* mutants (Figure 17A and Table 1). *B. henselae* Marseille ΔBadA-T mutant strains are consistently mentioned by their vector name, for instance strain *B. henselae* Marseille ΔBadA-T/pS27 is indicated as strain S27. Furthermore, 'S' refers to 'stalk', 'H' refers to 'head', 'N' refers to 'neck', 'D' refers to 'domain', and subsequent numbers denote the first observed N-terminal neck/stalk domain number.

Mutant strains S27, HN2S27, S28, S29, S30, and HNS30 were designed previously by Patrick Kaiser, PhD [72, 106] and their respective *badA* mutant sequences (Suppl. Figure 2-6) were assessed via colony PCR and Sanger sequencing using the primers BadA1_Fw, BadA3_Fw, BadA4_Fw, BadA5_Fw, and BadA_Rv. Additional *badA* mutant sequences D16S28, D19S28, D25S28, and D27S29 were designed *in silico* according to the *badA* sequence of *B. henselae* Marseille including the native promotor region (ca. 250 bp) and signal sequence (141 bp) and were synthesised by Invitrogen GeneArt® Gene Synthesis Services (Suppl. Figure 7-10). Synthesised *badA* mutants were amplified from their respective GeneArt® vectors (Table 2) using the primers S28domains_Fw and S28domains_Rv and subsequently cloned into the broad-host range vector pBBR1MCS-5 via Gibson Assembly®. Vector pBBR1MCS-5 was amplified using primers pBBR1MCS-5_Fw and pBBR1MCS-5_Rv. Resulting vectors were multiplied in heat-shock transformed *E. coli* DH5α and selected on kanamycin- and gentamycin-supplemented LB-agar plates. Bacterial vector integration was assessed via colony PCR using the primers pBBR1MCS-5_GA_Fw and pBBR1MCS-5_GA_Rv. Insert sequences were verified to be error-free via Sanger sequencing using the primers S28domains_Fw, S28domains_Rv, BadA1_Fw, BadA2_Fw, and BadA3_Fw. Electrocompetent *B. henselae* Marseille ΔBadA-T were transformed with the resulting vectors and further selected on kanamycin- and gentamycin-supplemented CBA plates.

Expression of all truncated and modified *badA* mutant genes was verified via WB using anti-BadA antibodies and illustrates the differences in protein size corresponding to their predicted trimeric MW (Figure 17B). The predicted trimeric MW of truncated and modified BadA constructs ranges from 77 kDa (for strain S30) to 327 kDa (for strain HN2S27). Strain *B. henselae* Marseille shows the highest MW band for BadA (predicted to be 1,252 kDa). Strains *B. henselae* Marseille ΔBadA-T and ΔBadA-D function as negative control and do not show any *badA* expression.
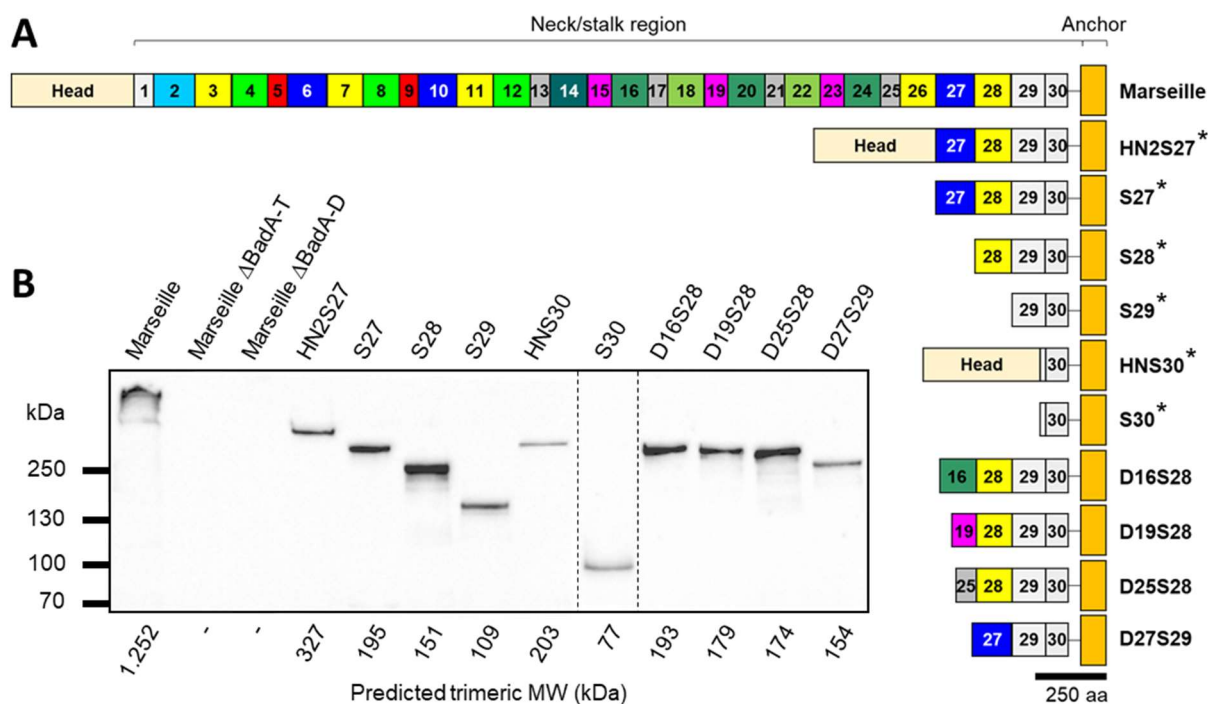
Figure 17. Overview of the truncated and modified BadA constructs and analysis of the MW via Western blotting (adapted from [128])**. (A)** Various truncated and modified BadA proteins are constructed by removing or rearranging specific passenger domains. BadA constructs that are indicated by a star (*) were designed previously [72, 106]. BadA constructs D16S28, D19S28, D25S28, and D27S29 were generated by combining a single neck/stalk domain with BadA S28 or S29. Images are drawn to scale according to aa-sequence length. Scale bar: 250 aa. **(B)** The variety in MW of the generated BadA constructs is demonstrated via WB using anti-BadA antibodies. The predicted MW of the trimeric BadA constructs is indicated underneath the blotted membranes. *B. henselae* strain HN2S27 displays the largest modified BadA construct (predicted to be 327 kDa), while strain S30 displays the smallest truncated BadA construct (predicted to be 77 kDa). The trimeric BadA protein of strain *B. henselae* Marseille is predicted to be 1,252 kDa, while negative control strains ΔBadA-T and ΔBadA-D lack the ability to express *badA*. Bacteria were analysed on two separate nitrocellulose membranes in which the order of columns has been rearranged *in silico* (dotted line).

Modified *badA* mutants D16S28, D19S28, D25S28, and D27S29 were created by combining a single *badA* neck/stalk domain with *badA* S28 or S29, the latter two functioning as a scaffold, exploiting the modular architecture of TAAs. Domains 16, 19, 25, and 27 (highlighted red in Figure 16A) were selected to represent different variants of BadA neck/stalk domains. Domain 16 (131 aa) is a representative for all nine domains present within the cluster of green-coloured domains (Figure 15A). Domain 19 is shorter (87 aa) and represents the three domains present within the cluster of purple-coloured domains. Domain 25 is likewise short (72 aa) and represents the three domains present within the cluster of grey-coloured domains.

3.8.2. Analysis of modified BadA fibres on the bacterial surface via confocal laser scanning and transmission electron microscopy

In order to be used in binding experiments, the presence of truncated and modified BadA constructs on the outer membrane of *B. henselae* Marseille ΔBadA-T was assessed via CLSM using anti-BadA antibodies and was additionally visualised via TEM. Truncated BadA constructs were identified on the bacterial surface via CLSM by a green halo surrounding the DAPI-stained intracellular DNA (Figure 18). A weaker CLSM-fluorescence signal is observed for strains S29 and S30 expressing the shortest *badA* mutants. Negative control strain Marseille ΔBadA-T does not express *badA* and therefore does not show a green halo. CLSM was performed by Daniela Bender, PhD, from the Paul-Ehrlich-Institut (Langen, Germany).
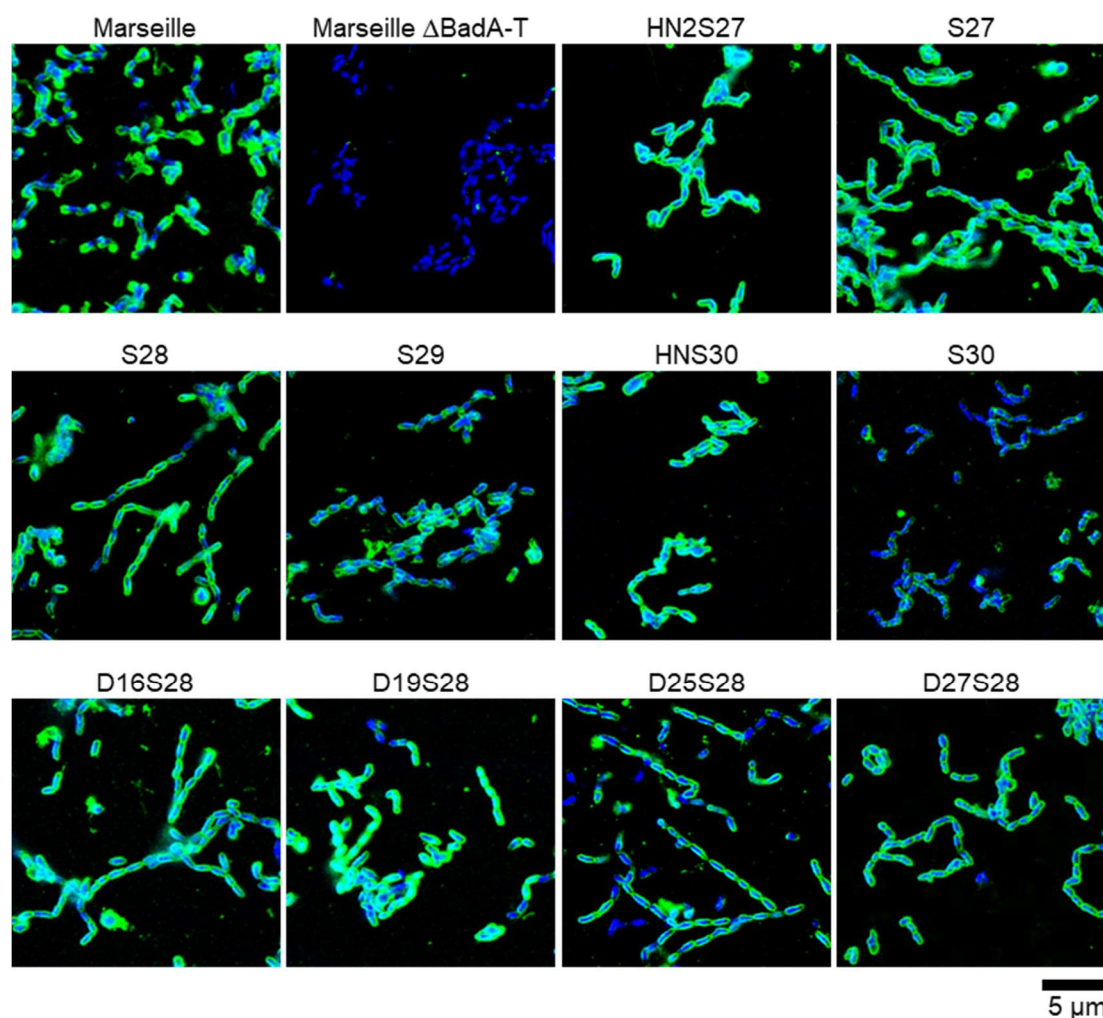
Figure 18. Analysis of truncated and modified BadA constructs on the bacterial surface via confocal laser scanning microscopy (adapted from [128]). The presence of truncated BadA constructs on the bacterial membrane was analysed via CLSM using anti-BadA antibodies (green halo) in which bacterial DNA was stained with DAPI (blue). Negative control strain *B. henselae* Marseille ΔBadA-T does not express *badA* and does not show a green halo. Depicted CLSM images are representative of at least four images from different areas on the same microscopy slide and were selected from over twenty representative images using conventional IFM. Scale bar: 5 µm.

CLSM results are supported by TEM in which the *B. henselae* ΔBadA-T mutant strains display a strongly truncated BadA protein, visible as short fibres on the bacterial surface (Figure 19). Observed BadA fibres vary in length from ca. 17 nm (for strain S30) to ca. 45 nm (for strain HN2S27), corresponding to their respective aa-sequence length. For *B. henselae* Marseille, a dense layer of long BadA fibres is observed (ca. 240 nm), while the negative control strain *B. henselae* Marseille ΔBadA-T is characterised by a smooth outer membrane. TEM imaging was performed by Katharina Hipp, PhD, from the Max Planck Institute for Developmental Biology (Tübingen, Germany). In conclusion, both microscopy approaches confirm the correct production of truncated and modified BadA fibres on the bacterial surface of *B. henselae* Marseille ΔBadA-T in which the fibre length corresponds to their respective BadA sequence length.
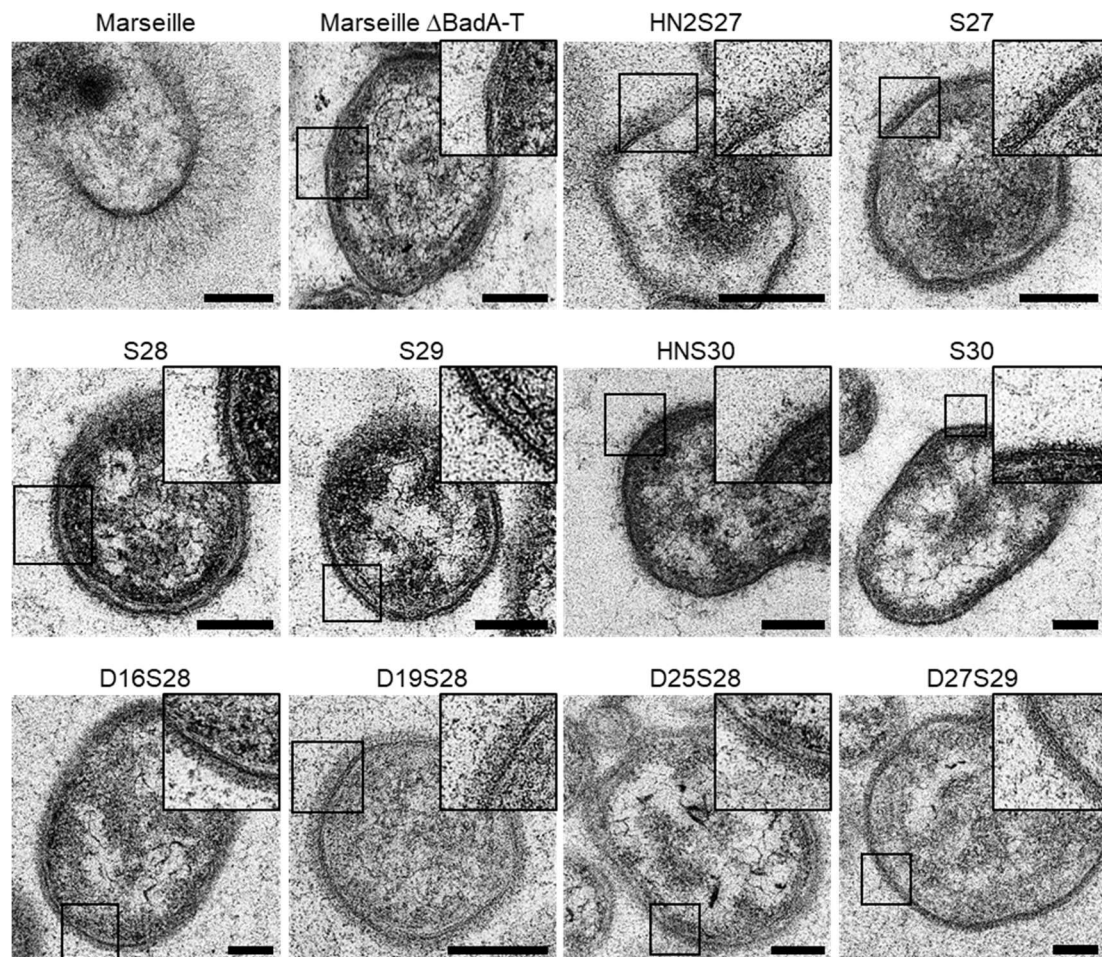
Figure 19. Analysis of truncated and modified BadA constructs on the bacterial surface via transmission electron microscopy (adapted from [128])**.** TEM images of all *badA*-expressing strains depict a dense layer of fibres protruding from the outer membrane. By contrast, negative control strain *B. henselae* Marseille ΔBadA-T is characterised by a smooth outer membrane. BadA fibre lengths vary from ca. 20 nm (for strain S30) to ca. 240 nm (for strain Marseille). Enlarged images of the BadA fibre structures are given in the framed boxes, respectively. Scale bars: 200 nm.

3.9. Functional analysis of the fibronectin binding of modified BadA constructs via ELISA and fluorescence microscopy

3.9.1. Analysis of fibronectin binding via ELISA

The ability of the different *B. henselae* Marseille ΔBadA-T mutant strains, each expressing a truncated and modified *badA* gene, to bind human plasma fibronectin was evaluated via whole-cell ELISA (Figure 20). In short, multiwell plates were coated with fibronectin in which attached bacteria were identified using anti-*B. henselae* antibodies and quantified via colorimetric absorbance measurements. All *B. henselae* Marseille ΔBadA-T mutant strains showed a significantly lower fibronectin binding compared to strain Marseille (100 %). Likewise, both negative control strains Marseille ΔBadA-T (16 %) and Marseille ΔBadA-D (17 %) showed a strongly reduced fibronectin binding.

Strains HN2S27 (78 %), S27 (82 %), and D19S28 (79 %) showed the highest fibronectin binding despite the deletion of ca. 85 % of the Marseille *badA* gene. Strain S28 (37 %) showed a significantly lower fibronectin binding compared to S27, although only missing domain 27. Strains S29 (23 %) and S30 (21 %) contain an even more truncated *badA* mutant and showed an even lower fibronectin binding. Strains D16S28, D19S28, and D25S28 express a *badA* mutant consisting of domain 16, 19, and 25 merged to *badA* S28, respectively. Compared to strain S28, strains D16S28 (60 %) and D19S28 showed a significant increase in fibronectin binding, while D25S28 (42 %) did not. Similarly, strain D27S29 (60 %) expresses a *badA* mutant consisting of domain 27 merged to *badA* S29 and showed a significantly higher fibronectin binding compared to strain S29. In addition, no significant difference in fibronectin binding was observed between strains HNS30 (31 %) and HN2S27, and their headless variants S30 and S27, respectively. In conclusion, BadA neck/stalk domains 19 and 27 (and highly similar domains) are suggested to play a major role in fibronectin binding.
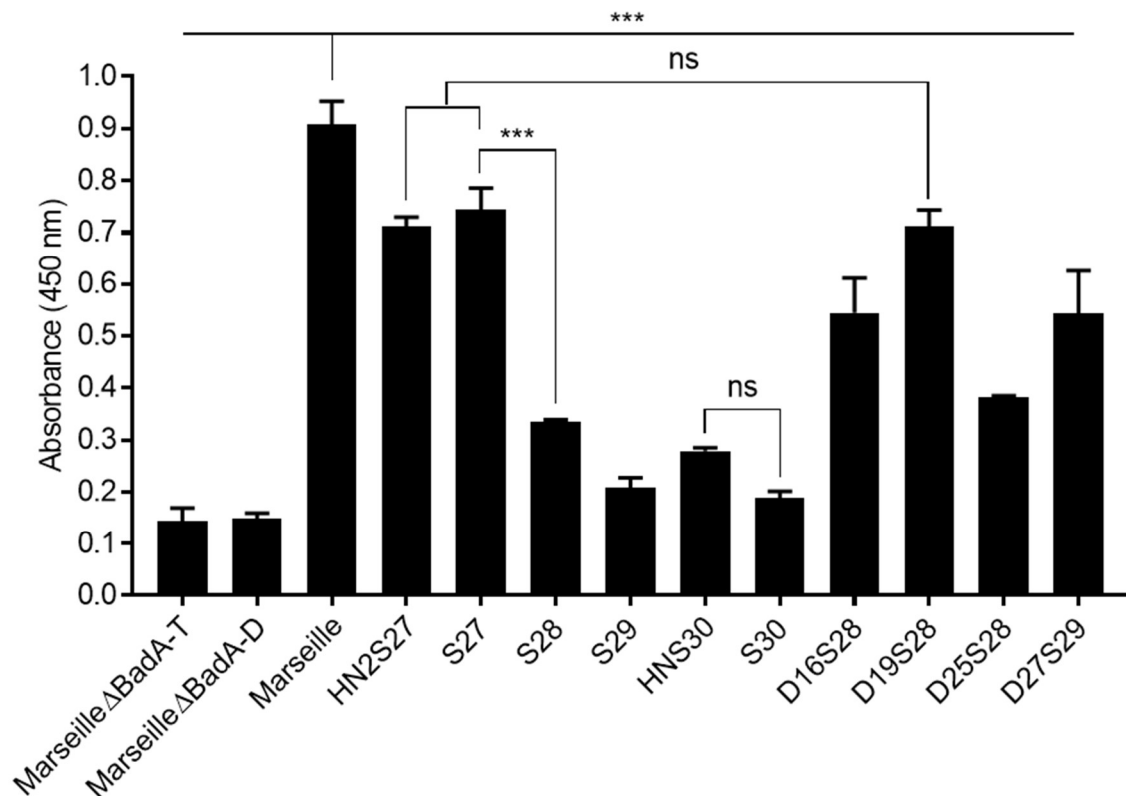
Figure 20. Analysis of the fibronectin binding of *B. henselae* Marseille ΔBadA-T mutant strains via ELISA (adapted from [128])**.** The ability of the different *B. henselae* Marseille ΔBadA-T mutant strains to bind human plasma fibronectin was quantified (colorimetric absorbance measurements at 450 nm) via whole-cell ELISA using anti-*B. henselae* antibodies. All *B. henselae* Marseille ΔBadA-T mutant strains, including both negative control strains Marseille ΔBadA-T (0.144) and Marseille ΔBadA-D (0.150), show a significant lower fibronectin binding than strain Marseille (0.908). Strains S28 (0.334), S29 (0.209), HNS30 (0.279), S30 (0.189), and D25S28 (0.383) show a reduced fibronectin binding. Strains D16S28 (0.545), D19S28 (0.713), and D27S29 (0.544) show a significant higher fibronectin binding than strains S28 and S29. No significant (ns) difference in fibronectin binding is observed between strains D19S28 (0.713) and S27 (0.745). Moreover, strains HN2S27 (0.711) and HNS30 do not show a significant higher fibronectin binding than their headless variants S27 and S30, respectively. Statistical significance was determined using one-way ANOVA testing (***; $p < 0.001$).

3.9.2. Analysis of fibronectin binding via fluorescence microscopy

Findings obtained via quantitative whole-cell ELISA analysis were verified by a fluorescence microscopy analysis (Figure 21). In short, bacteria were processed on the same multiwell plate but identified using DAPI staining. Accordingly, a higher number of fibronectin-bound bacteria were identified for strains Marseille, HN2S27, S27, D16S28, D19S28, and D27S29. No fibronectin-bound bacteria were detected for negative control strain Marseille ΔBadA-T.
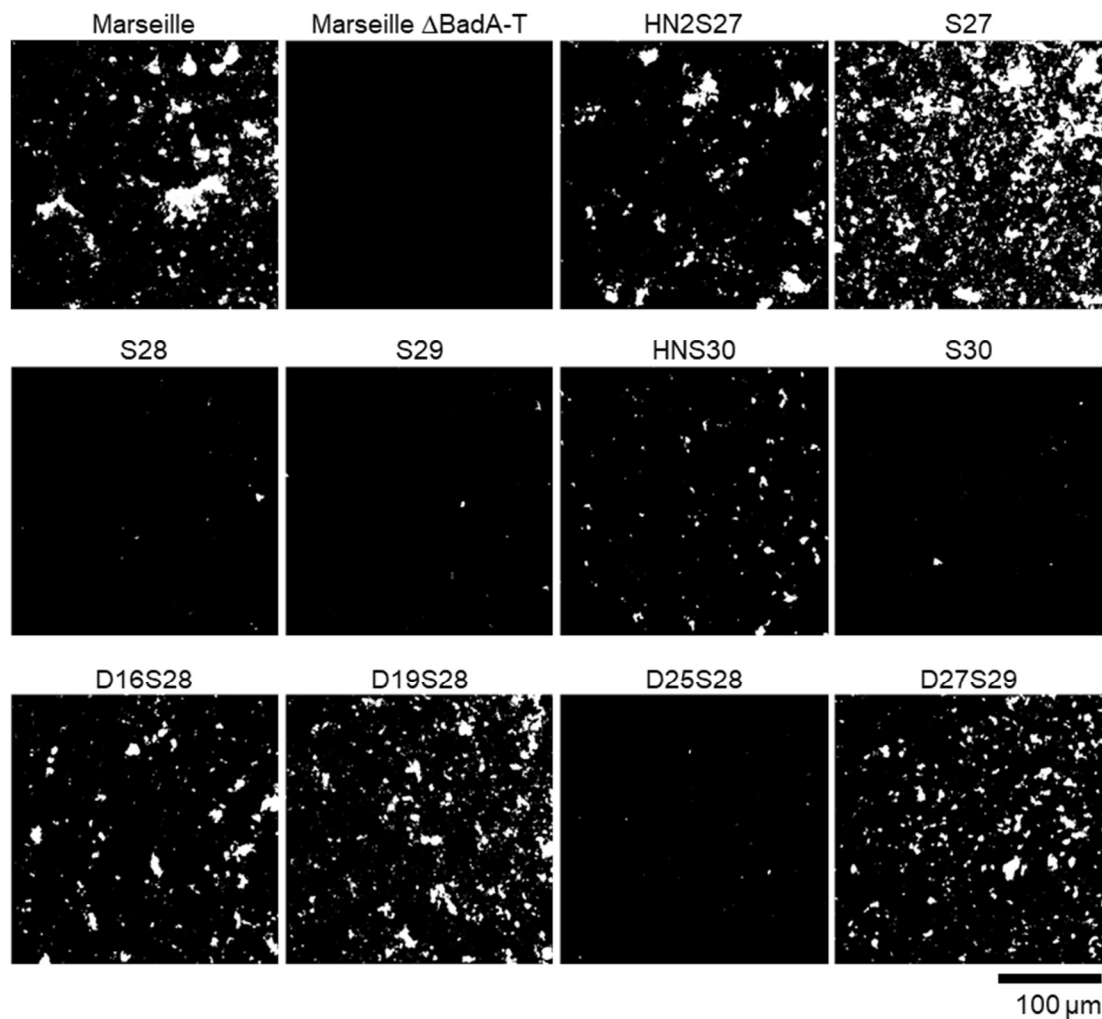
Figure 21. Analysis of the fibronectin binding of *B. henselae* Marseille ΔBadA-T mutant strains via fluorescence microscopy (adapted from [128]). The ability of the different *B. henselae* Marseille ΔBadA-T mutant strains to bind human plasma fibronectin was assessed via fluorescence microscopy using DAPI staining (white). Strains Marseille, HN2S27, S27, D16S28, D19S28, and D27S29 show a high number of fibronectin-bound bacteria, while negative control strain Marseille ΔBadA-T does not. Strains S28, S29, HNS30, S30, and D25S28 show a low number of fibronectin-bound bacteria. Scale bar: 100 µm.

3.9.3. Analysis of the bacterial seeding number via quantitative real-time polymerase chain reaction

To assess the addition of approximate equal amounts of *B. henselae* Marseille mutant strain cells per well for both the whole-cell ELISA analysis and the fluorescence microscopy analysis, a qPCR was performed (Figure 22). In short, bacterial numbers for a bacterial solution with an OD_{600} of 0.5 were calculated via an internal standard by amplifying a fragment (120 bp) of the housekeeping gene *glyA* from the vector pCR™2.1-TOPO®_*glyA* using the primers glyA_Fw and glyA_Rv. While strains Marseille ΔBadA-T and Marseille ΔBadA-D show a significant lower number of gene copy equivalents compared to the wild type strain Marseille, strains D16S28, D19S28, and D25S28 show a significant higher number of gene copy equivalents. Nonetheless, all strains depict a number of gene copy equivalents within the same order of magnitude confirming the approximate equal addition of bacterial cells per well.
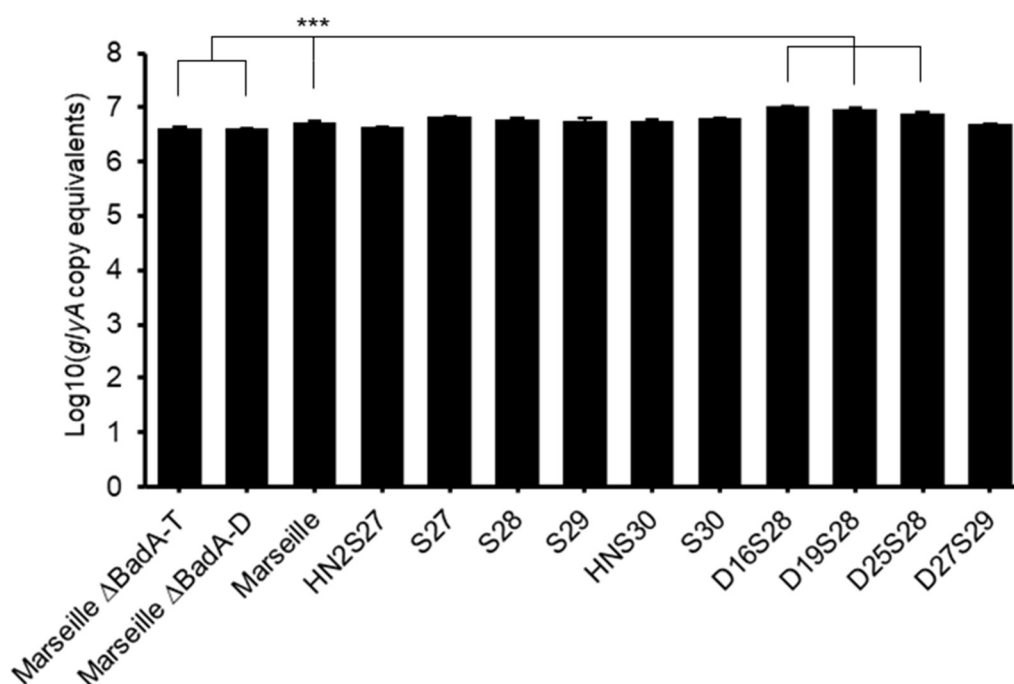
Figure 22. Analysis of the bacterial seeding number via quantitative real-time PCR (adapted from [128])**.** The observed logarithmic numbers of *glyA* copy equivalents correspond to bacterial solutions (in PBS) with an OD_{600} of 0.5. The addition of approximate equal amounts of *B. henselae* Marseille mutant strain cells in both the whole-cell ELISA analysis and fluorescence microscopy analysis was confirmed via qPCR using logarithmic numbers of *glyA* copy equivalents as an internal standard. Statistical significance was determined using one-way ANOVA testing (***; *p* < 0.001).

3.10. Characterisation of a BadA-fibronectin binding site using specific anti-BadA-DALL antibodies

3.10.1. Generation and specificity of anti-BadA-DALL antibodies

BadA neck/stalk domain 27 was, together with homologous domains, suggested to be important for binding human plasma fibronectin. The common DALL motif is predicted to consist of β-sheets forming a hairpin structure and might act as an optimal fibronectin interaction site. To assess the role of the DALL motif of domain 27 in binding fibronectin, a 15-mer peptide sequence (RHEKSKLEKGASKAI) from the DALL motif was synthesised and used as antigen for rabbit immunisation (performed by Eurogentec) to generate anti-BadA-DALL antibodies.

Specificity of the anti-BadA-DALL antibodies was assessed via WB, ELISA, and IFM using anti-BadA-DALL antibodies (Figure 23). Strains Marseille, S27, and D27S29 produce a modified BadA protein including the targeted DALL motif sequence and were identified via WB (Figure 23A), ELISA (Figure 23B), and IFM (Figure 23C) using the anti-BadA-DALL antibodies. Negative control strains Marseille ∆BadA-T, S28, S29, S30, D16S28, and D25S28 do not express a modified *badA* gene that contains the targeted DALL motif sequence and were thus not identified. The DALL motif of domain 19 includes a DALL motif sequence (RHEKSKLEKGVSKAT) that differs only by two aa with the targeted DALL motif sequence of domain 27. Hence, strain D19S28 was also identified via ELISA using the anti-BadA-DALL antibodies. In conclusion, the generated anti-BadA-DALL antibodies specifically identify the 15-mer peptide sequence from the DALL motif in domains 27 and 19.
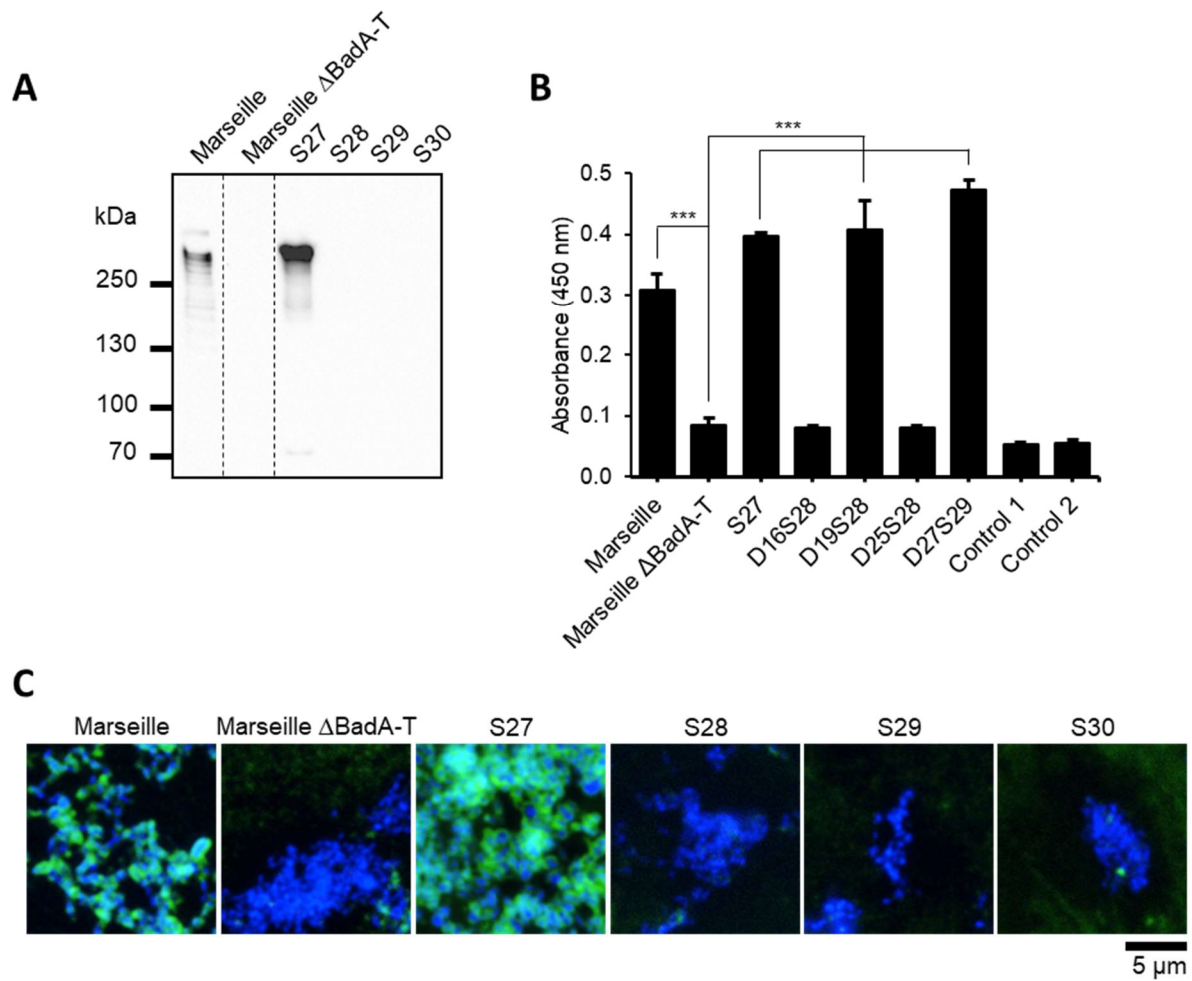
Figure 23. Analysis of the specificity of anti-BadA-DALL antibodies via Western blotting, ELISA, and immunofluorescence microscopy (adapted from [128]). **(A-C)** Only strains Marseille, S27, D19S28, and D27S29 expressing *badA* including the targeted (or highly similar) DALL motif sequence were identified via WB, ELISA, and/or IFM analysis using anti-BadA-DALL antibodies. Negative control strains Marseille ΔBadA-T, S28, S29, and S30 lack the targeted DALL motif sequence and are thus not detected. **(A)** WB: Bacteria were analysed on a single nitrocellulose membrane in which the order of columns has been rearranged *in silico* (dotted line). **(B)** ELISA: Multiwell plates were coated with human plasma fibronectin and attached bacteria were quantified via colorimetric absorbance measurements at 450 nm. For negative control 1 no bacteria were added, for negative control 2 strain Marseille was added without prior fibronectin coating. Statistical significance was determined using one-way ANOVA testing (***; $p < 0.001$). **(C)** IFM: Identification of the targeted DALL motif sequence is indicated by a green halo surrounding the DAPI-stained intracellular DNA. Scale bar: 5 µm.

3.10.2. Analysis of a decreased BadA-fibronectin binding via anti-BadA-DALL antibodies

To assess the role of the DALL motif of domain 27 in binding fibronectin, specific anti-BadA-DALL antibodies were generated, and evaluated for BadA-fibronectin inhibiting features via ELISA testing (Figure 24A). In short, multiwell plates were coated with bacteria (strains S27 or Marseille ΔBadA-T) and successively incubated with a dilution of anti-BadA-DALL antibodies (1 μg/ml, 10 μg/ml, or 100 μg/ml) and human plasma fibronectin. Anti-BadA antibodies were similarly diluted and used as positive control. Bound fibronectin was identified using anti-fibronectin antibodies.

The stepwise increase of anti-BadA-DALL antibodies resulted in a gradual decrease of bound fibronectin to strain S27, indicating that the 15-mer protein sequence of the DALL motif of domain 27 is involved in fibronectin binding (Figure 24B). A similar observation was made for the anti-BadA antibodies targeting various unknown sites of BadA.
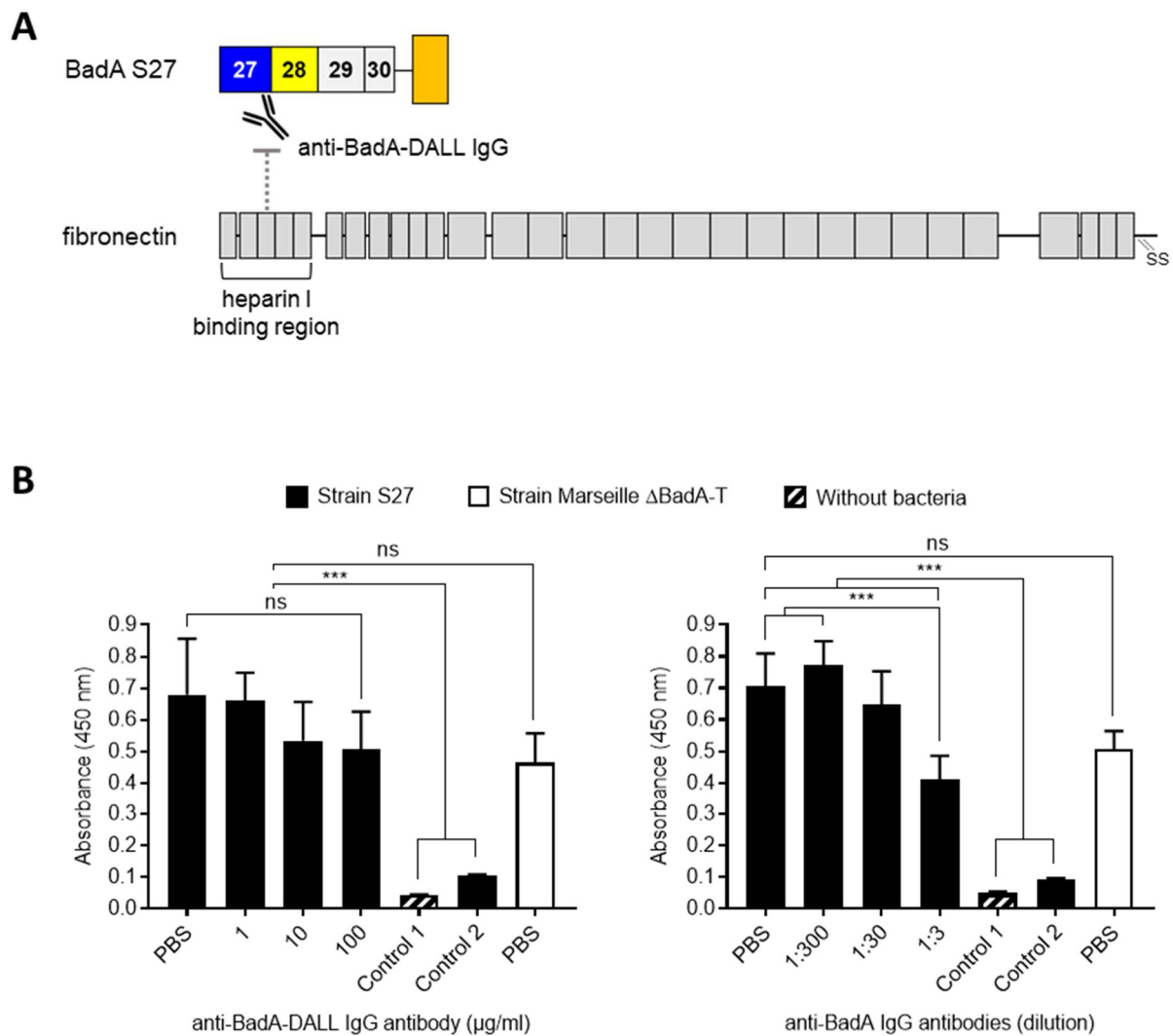
Figure 24. Analysis of the inhibiting effect of anti-BadA-DALL antibodies on BadA-fibronectin binding (adapted from [128]). **(A)** Overview of principle where anti-BadA-DALL antibodies inhibit the binding of fibronectin to BadA S27 by occupying specifically the DALL motif sequence in domain 27. The heparin binding region I of fibronectin was suggested previously to be a BadA interaction site [73]. **(B)** Multiwell plates were coated with strains S27 or Marseille ΔBadA-T and successively incubated with a dilution of anti-BadA-DALL antibodies (1 µg/ml, 10 µg/ml, or 100 µg/ml) and human plasma fibronectin. Anti-BadA antibodies were similarly diluted and used as positive control. Bound fibronectin was quantified via colorimetric absorbance measurements at 450 nm. Increasing concentrations of anti-BadA-DALL antibodies resulted in a decrease of fibronectin binding of strain S27, back to the level of strain Marseille ΔBadA-T (negative control). Similar observations were made for the anti-BadA antibodies (positive control). For negative control 1, no bacteria were coated, for negative control 2, no fibronectin was added. Statistical significance was determined using one-way ANOVA testing (***; $p < 0.001$).

4. Discussion

B. henselae is an endemic, fastidious, and slow-growing bacterial species. Cats serve as the major reservoir host of *B. henselae* and transmission among cats occurs through the faeces or bites of infected cat fleas [47–49]. The incidental inoculation of humans usually occurs indirectly through infected flea faeces that is superficially scratched into the skin. *B. henselae* is the causative agent of CSD, a self-limiting zoonotic disease characterised by localised lymphadenopathy. Other common symptoms are skin lesions at the site of the cat scratch, fatigue, and fever. Immunocompromised patients might suffer from life-threating endocarditis [5, 60] or vasculoproliferative disorders such as bacillary angiomatosis [59, 61].

Adhesion to host cells is the first and foremost step during infection with pathogens. In Gram-negative bacteria, TAAs represent a major class of pathogenicity factors and TAA-producing bacteria cause a wide spectrum of human diseases such as CSD (*B. henselae*), enterocolitis (*Y. enterocolitica*), meningitis (*N. meningitis*), and blood stream infections (multi-drug resistant *A. baumannii*). TAA-targeted anti-adhesive strategies might represent a universal strategy to counteract such bacterial infections. YadA of *Y. enterocolitica* is considered the prototypical TAA despite its relatively short passenger domain and fibre length of ca. 23 nm [98, 167]. Examples of other well-studied TAAs are Ata of *A. baumannii* [168], NadA of *N. meningitidis* [100], and SadA of *S. enterica* [99]. One of the best characterised TAAs is BadA of *B. henselae* and is associated with angiogenic reprograming of infected host cells [76–79]. Expression of *badA* has been suggested to be upregulated under conditions mimicking the human host cell environment and downregulated under conditions mimicking the cat flea host environment [169]. BadA is primarily responsible for bacterial binding to ECM proteins such as collagen, laminin, and fibronectin [118]. Moreover, it has been described that the interaction of BadA with fibronectin represents the basis for adhesion of *B. henselae* to endothelial host cells [73].

TAAs are characterised by a homologous C-terminal membrane anchor and share a modular and repetitive passenger domain consisting of head, neck, and stalk domains [95, 102]. Neck/stalk domains share specific sequence motifs with characteristic conformations, annotated by the 'domain dictionary' approach of the daTAA server [102, 104]. While it was shown that the BadA head domain is crucial for adhesion to various ECM proteins and angiogenic reprogramming of host cells [72, 117], adhesion of *B. henselae* to fibronectin was demonstrated to be mediated via the BadA neck/stalk region [106]. Thus far, the exact BadA neck/stalk domains or motifs that are responsible for fibronectin binding remain unknown. However, to determine specific BadA binding motifs, it must be first verified that such motifs are conserved within the species of *B. henselae*.

4.1. Long-read sequencing reveals genetic adaptation of the *badA* island among different *B. henselae* strains

4.1.1. Long-read sequencing covers the highly repetitive *badA* island

Genomes of eight *B. henselae* strains were sequenced using long-read sequencing techniques with the aim of analysing differences in the highly variable *badA* gene, studying *badA* expression, and investigating the role of BadA in binding ECM proteins. Because of the long and repetitive *badA* gene and flanking regions, assembly of short-read sequencing data often fell short or was considered questionable [170]. In contrast, long-read PacBio sequencing yields 20-25 kb reads that cover highly repetitive stretches bypassing few genome assembly challenges related to short-read sequencing data [171].

For instance, it was demonstrated that the *badA* gene from strain Marseille consists of 11,922 bp instead of the earlier established 9,249 bp [76, 127]. It can be reasoned that two repeated regions of 1.4 kb and 1.3 kb, respectively, were omitted during the assembly of the short-read Sanger sequencing data. Furthermore, the PCR-identified variations in the repetitive neck/stalk region of the *badA* gene of different *B. henselae* strains [111] were validated via long-read sequencing data analyses (Figure 3). The average BadA fibre length from each *B. henselae* strain was determined via TEM-image analysis and correlates with their respective *badA* gene length (Figure 12). Small variations between the measured BadA fibre lengths of *B. henselae* G-5436, 88-64 Oklahoma, and FR96/BK38, despite showing similar *badA* gene lengths, might derive from structural changes due to the TEM-processing.

4.1.2. Classification of the *B. henselae* species

All sequenced *B. henselae* strains show a conserved genome with a high pairwise genome sequence similarity (≥ 98.57 %). An ANI-score of ≥ 95 % is suggested as a cut-off value for defining prokaryotic species [172]. Divergences are primarily observed in a ca. 400,000 bp region, previously designated as a type II secretion system island [32], and are associated with the presence of prophage sequences (Figure 2). The remaining genome sequence is highly conserved among the species. Nonetheless, a major inversion of two adjacent collinear regions (ca. 1.5 Mbp) is observed in strains G-5436, 88-64 Oklahoma, and FR96/BK38. Both inversion breakpoints are characterised by a copy of the highly conserved *tuf* gene and genes encoding for ribosomal and transcription-related proteins [173, 174].

Thus far, only two *B. henselae* genotypes have been described. Genotype I (represented by *B. henselae* ATCC49882[T] Houston-I) is suggested to be more associated with human infections, while genotype II (represented by *B. henselae* Marseille) might favour the cat host

environment as they appear to outcompete genotype I during bloodstream infections in cats [31, 44]. *B. henselae* Marseille, ATCC49882^T var-1, ATCC49882^T var-2, Berlin-I, G-5436, and 88-64 Oklahoma are isolates from human patients diagnosed with either CSD, HIV, or bacillary angiomatosis, while *B. henselae* FR96/BK38 and FR96/BK3 are isolates from domestic cats. The genomes of *B. henselae* ATCC49882^T var-1, ATCC49882^T var-2, Berlin-I, G-5436, and 88-64 Oklahoma can be classified as genotype I strains because of their high pairwise genome sequence identity (≥ 99.9 %) and homologous copy of two 16S rRNA genes [41]. In addition, strains ATCC49882^T var-1, ATCC49882^T var-2, Berlin-I, and G-5436 have been described as variants of the genotype I type-strain ATCC49882^T Houston-I. Similarly, *B. henselae* Marseille, FR96/BK38, and FR96/BK3 share a homologous copy of two 16S rRNA genes and can thus be classified as genotype II strains. However, the comparatively low pairwise genome sequence identity (≥ 98.57 %) of the genotype II strains might suggest otherwise (Table 18). Moreover, strain FR96/BK38 shows a higher pairwise genome sequence identity (≥ 99.37 %) to strains classified as genotype I. The lowest pairwise genome sequence identities are observed for strains Marseille (≥ 98.82 %) and FR96/BK3 (≥ 98.57 %) that appear to be more genetically distinct. The frequent emergence of new *B. henselae* genetic variants *in vivo* has been described before and demonstrates that classifying *B. henselae* species into particular genotypes cannot be solely attributed to the source of isolation or 16S rRNA gene sequences [37, 175].

In addition, numerous variations within the *badA* island among the studied *B. henselae* strains are observed that can be attributed to a series of recombination events. As a result, none of the analysed *badA* sequences are identical, only three *B. henselae* strains contain a downstream *badA*-like domain region, and three different types of *badA* anchor domain sequences are identified. Based on the intact *badA* gene, the type of *badA* anchor domain sequence, and the long downstream badA-like domain region, *B. henselae* G-5436 is suggested to be the evolutionary ancestor of the strains analysed in this work. In conclusion, future classification and genotyping of *B. henselae* strains should be supported by either long-read whole genome sequencing techniques or by a detailed analysis of the *badA* island.

4.1.3. Potential influence of flanking genes on *badA* expression

Expression of *badA* is influenced by environmental signals such as pH, iron availability, and temperature and is correlated with the bacterial lifestyle of frequently alternating host environments [112, 114, 116]. The functional role of the *badA* genomic neighbourhood on the regulation of *badA* expression and modification remains unexplored [176]. The variable *badA* island sequence is flanked up- and downstream by a highly conserved region including genes

with predicted proteins that are involved in various functions such as transcriptional regulation (for example Fur-like proteins, ComR-like proteins, and two ribosomal silencing factors), post-translational modification (for example RNA methyltransferase), protein synthesis (for example the GTP binding protein, two serine tRNAs, and two 50S ribosomal proteins), membrane transport (for example an efflux resistance-nodulation-cell division transporter permease subunit), metabolism (for example phosphoenolpyruvate-protein phosphotransferase, glucose-6-phosphate isomerase, and glutamate 5-kinase), membrane integrity (for example a murein hydrolase activator), pathogenicity (for example IalB), or DNA damage repair, transposition, and recombination (for example ComR-like proteins, and a protein of the AAA+ ATPase superfamily).

The rather close vicinity (ca. 23.5 kb) of a predicted Fur-like protein to the *badA* island might suggest a repressed *badA* transcription in *B. henselae* upon residing within the iron-saturated flea gut environment. Conversely, within mammal hosts where free heme as an iron source is rare, upregulation of *badA* might facilitate initial adhesion of *B. henselae* to the ECM during infection [177]. Furthermore, the identified *ptsP* gene upstream of *badA* is predicted to be a component of a phosphoenolpyruvate phosphotransferase system that has been associated to virulence gene expression related to nutrient availability [178]. Lastly, an *ialB* pseudogene and *ialB* gene are located downstream of the *badA* island in which the corresponding IalB protein is involved in erythrocyte invasion [88, 162, 164]. In contrast to *badA* expression, *ialB* is shown to be upregulated *in vitro* under conditions with a lower pH (< 7.2) or lower temperatures (< 37 °C) suggesting the bacteria to be primed for erythrocyte invasion under conditions mimicking the cat flea host [162]. TAA genes are generally monocistronic, however operon configurations including a TAA gene and an additional downstream gene associated with TAA biogenesis have been observed in *Brucella abortus*, *Salmonella* spp., *Burkholderia* spp., and *Acinetobacter* sp. Tol 5 [179–182]. Nonetheless, the influence of any of the abovementioned genes or predicted proteins on the regulation of *badA* expression remains to be explored.

4.1.4. The *badA* island is a recombination hotspot

The pathogenic life cycle of *B. henselae* with frequent host transitions from the cat flea to the cat flea's gut, to cats, and potentially to humans suggests the need for an efficient and quick adaptation strategy to the differing host environments. In contrast to other variable regions, the *badA* island has not yet been predicted nor described as a prophage or defined genomic island [24, 32]. Prophage sequence-containing regions have been described to stimulate

diversification and dispersion of specific host-adaptability genes throughout the *B. henselae* species [24, 27, 183].

The observed diversity between the *badA* islands of the analysed *B. henselae* strains might be the result of evolutionary selection for few beneficial traits such as host-specific colonisation or immune evasion. Moreover, the numerous versions of the *badA* island and its extensive number of repeats make this region and excellent 'toolbox' for TAA adaptation and might function as a back-up system against reductive genome evolution, emphasizing its evolutionary significance. Expression of the *badA* pseudogene and the *badA*-like domain region has not yet been examined nor demonstrated. Furthermore, the potential role of several flanking genes involved in recombination-dependent replication, DNA damage repair, and/or genome plasticity (for example the mobile genetic element and the ComR-like transcriptional factors) remains unclear. Variations in the expression status and the length of *badA* from different *B. henselae* strains have been identified via WB using anti-*B. henselae* antibodies and via long-distance PCR approaches, respectively [111]. It is suggested that a shuffling mechanism comparable to phase variation via recombination might mediate the occurrence of new *badA* variants [35, 104, 184]. The highly repetitive regions in the sequences of *badA*, the *badA* pseudogene, and the *badA*-like domain region might stimulate such site-specific recombination or slipped-strand mispairing, resulting in new variations of the *badA* island.

In particular, the BadA proteins of *B. henselae* Berlin-I and ATCC49882[T] var-2 differ considerably from the BadA proteins observed in other *B. henselae* strains regarding their size, number of repetitive domains, and overall domain organisation (Figure 4A). The *badA* gene of strains Berlin-I and ATCC49882[T] var-2 might be the result of a recombination event deleting a large portion of the *badA* gene and a former *badA*-like domain region, combining the first part of the *badA* gene (up to 69 bp before the start of domain 4) with the latter part of a former *badA*-like domain region (last 6,360 bp). Hence, the presence of an 18-bp repeat region and a type 2 BadA anchor domain in their respective *badA* genes. A similar recombination event might have happened in the *badA* island of strain 88-64 Oklahoma connecting a large part of *badA* (up to domain 26) with the last 697 bp of a former *badA*-like domain region, resulting in a *badA* gene with a type 2 BadA anchor domain. Moreover, the current *badA* gene of strains Marseille and FR96/BK3 might derive from a common *badA* gene (for example the *badA* gene of strain G-5436) that has been through several recombination events. Indications thereof can be found through the presence of multiple *badA* neck/stalk domain repeats present only one time in the *badA* gene of strain G-5436 (Figure 4).

Low-passage *B. henselae* isolates are usually characterised by the abundant expression of *badA*, while extensive passaging on CBA plates has resulted in the occurrence of a *badA*-deficient *B. henselae* strain [185], strengthening the hypothesis of existing phase-on and

phase-off phenotypes [23, 35]. Accordingly, it is possible that the observed frameshift mutations in *badA* in *B. henselae* strains ATCC49882[T] var-1 and Berlin-I have occurred because of cultivating the bacteria under laboratory conditions. Attenuated *B. henselae* strains, deficient in expressing *badA*, might even stay restricted to the cat flea host or might survive only under artificial laboratory conditions where the metabolic burden of *badA* expression selects for faster growing colonies that have silenced the *badA* gene. Loss of biologically important genetic information might also happen under *in vivo* conditions. Thus far, regaining the ability to express *badA* has never described. However, a suitable animal infection model mimicking human infections that could confirm this hypothesis does not exist.

4.1.5. The 18-bp repeat region follows a periodic glycine-x-x motif

Five analysed *B. henselae* strains include at least one region that solely consists of a strict 18-bp repeat motif sequence. The predicted protein structure [GSNG(N/S)G] follows a periodic glycine-x-x (pGxx) motif mimicking the structure of a poly-proline collagen helix. Similar pGxx motifs have been identified in other TAAs or YadA-like proteins from different pathogens (for example in *A. johnsonii*, *S. enterica,* and *Ewingella americana*). Thus far, and within the *B. henselae* species, only *B. henselae* ATCC49882[T] var-2 contains a pGxx motif region in a membrane exposed BadA protein. It is possible that the identified pGxx motifs are simply present as junk DNA [186], however, DNA tandem repeats forming so-called minisatellites are known to be hypermutable regions that act as engines for genetic variability and bacterial adaptation to changing environments [187]. In conclusion, follow-up research is necessary to clarify the function, if existing, and origin of this peculiar 18-bp repeat motif.

4.2. Adhesion of *B. henselae* to fibronectin is mediated via repetitive motifs present in the stalk of BadA

4.2.1. BadA is crucial for adhesion in the initial phase of infection

BadA production is crucial for adhesion to ECM proteins, endothelial cell infection, and induction of a proangiogenic response [19, 76, 118]. It has been demonstrated that *B. henselae* binds ECM proteins using its long and membrane exposed BadA fibres to connect the bacteria to the host cell surface. The binding of *B. henselae* to fibronectin has been attributed to the BadA neck/stalk region, while the BadA head domain was found to be responsible for binding collagen [72, 106].

Consequently, *B. henselae* strains deficient in expressing *badA* showed only minimal binding to the ECM proteins fibronectin and collagen-I (Figure 14). While collagen-I is abundantly present in the human body, especially in the dermis where it represents a major binding partner for TAAs [188–190], fibronectin was proven to be a key first binding partner of *B. henselae* during infection of human endothelial host cells in blood vessels or heart valves [73, 120]. Moreover, the binding ability of the different *badA*-expressing *B. henselae* strains seemed to be independent of the BadA fibre length, domain composition, or number of neck/stalk domains (Figure 14). Though rearranged, certain neck/stalk domains remain conserved among all analysed *badA* sequences. To be able to inhibit the initial binding of *B. henselae* in the course of infection, it must first be determined which conserved domains or motifs of the BadA neck/stalk region are involved in binding fibronectin.

4.2.2. The role of specific BadA neck/stalk domains in binding fibronectin

To specify the role of certain domains and motifs in binding fibronectin, numerous truncated and modified BadA constructs were generated (Figure 17A) and their ability to bind fibronectin was analysed under static binding conditions (Figure 20). For this work, the laboratory model strain *B. henselae* Marseille was used.

In general, all analysed *B. henselae* Marseille ΔBadA-T mutant strains showed a significantly lower fibronectin binding compared to strain Marseille. The addition of approximate equal amounts of *B. henselae* Marseille ΔBadA-T mutant strain cells per well for both the whole-cell ELISA and the fluorescence microscopy analysis was confirmed. The low but measurable fibronectin binding of negative control strains *B. henselae* Marseille ΔBadA-T and Marseille ΔBadA-D might be attributed to other *B. henselae* adhesion proteins such as HbpA/Pap31, Omp89, and Omp43 [71, 90, 94]. No significant difference in fibronectin binding was observed between strains HNS30 and HN2S27 on the one hand, and strains producing their respective headless counterparts S30 and S27 on the other hand, confirming prior findings that the head domain is not directly involved in binding fibronectin [106]. Strains S29 and S30 showed a relatively low fibronectin binding, approximating the level of both negative control strains Marseille ΔBadA-T and Marseille ΔBadA-D. The short-chain lipopolysaccharide of *B. henselae* [89] and other outer membrane proteins should not obstruct binding in case of the long wild-type BadA fibre, however, it is unknown whether the BadA fibres of strains S29 (ca. 23 nm) and S30 (ca. 17 nm) are large enough to stick out from the glycolipid layer. Future analyses using an LPS-deficient *B. henselae* strain could clarify this matter. However, the YadA fibre of *Y. enterocolitica* likewise measures only 23 nm and is able to bind various ECM proteins, though to a lower degree compared to *B. henselae* Marseille [96, 118].

Despite the deletion of 26 *badA* neck/stalk domains (ca. 10 kb), strains S27 and D19S28 still showed a relatively high fibronectin binding (ca. 80 %) compared to the fibronectin binding of the wild type strain Marseille. In contrast, strain S28 showed a significantly lower fibronectin binding compared to strains S27 or D19S28, although only missing one *badA* neck/stalk domain. Therefore, BadA neck/stalk domains 19 and 27 are considered important regions for the initial attachment of *B. henselae* to fibronectin. BadA neck/stalk domains 6 and 10 are homologous to domain 27, while domain 15 and 23 are homologous to domain 19. Moreover, strain D27S29 showed a significantly higher fibronectin binding compared to strain S29, although only including the extra *badA* neck/stalk domain 27, which confirms the important role of domain 27 in the adherence of BadA to fibronectin.

4.2.3. BadA-fibronectin binding is mediated via repetitive motif sequences

Overall, BadA neck/stalk domains demonstrate a recurring pattern including structural FGG motifs, coiled-coil motifs, and DALL-neck tandem connectors [102, 104]. BadA neck/stalk domains 19 and 27 share a similar coiled-coil motif and DALL-neck tandem connector sequence (Figure 16). A complex network showing close interactions between BadA and fibronectin has been resolved using XL-MS, of which six interaction sites were verified *in vitro* to be directly involved in fibronectin binding (Figure 16A) [73]. Four of the identified interaction sites are located exclusively in domain 19 and domain 27 (and homologous domains) and emphasise the role of domains 19 and 27 in the initial attachment of *B. henselae* to fibronectin. The first interaction site (VNNNVTNKFNELTQSITNVTQQVK) is part of the coiled-coil motif, whereas the other three identified interaction sites are all part of the DALL-neck tandem connector motif (LEKGASKATQENSKITYLLDGDVSK).

The DALL motif is predicted to consist of two β-strands forming a hairpin structure and is considered an optimal interaction site for adherence to the numerous unpaired β-strands present in fibronectin [105, 109, 191]. For instance, *Staphylococcus aureus* and *Streptococcus pyogenes* have been observed to bind fibronectin via extended tandem β-zippers [192]. To assess the role of DALL motifs in binding fibronectin, an anti-BadA-DALL antibody was generated targeting a 15-mer peptide sequence (RHEKSKLEKGASKAI) from the DALL motif of domain 27 that is observed partly in one of the BadA-fibronectin interaction sites identified via XL-MS [73]. The stepwise increase of anti-BadA-DALL antibodies resulted in a gradual decrease of fibronectin attached to strain S27, indicating that the 15-mer sequence of the DALL motif of domain 27 is involved in fibronectin binding. A stronger decrease of fibronectin binding was observed when using anti-BadA antibodies as inhibiting component, however, the exact antibody concentration and BadA targeting sites are unknown.

The observed increase in fibronectin binding of strain D16S28 compared to strain S28, might be the result of an additional FGG motif present in domain 16. FGG motifs are distributed abundantly in the BadA neck/stalk region, are usually characterised by the insertion of a 3-stranded β-meander into a coiled-coil region, and have been described as a potential fibronectin binding region [97, 104, 106, 107]. Moreover, domain 16 and domain 28 each contain a BadA-fibronectin interaction site interaction site in their FGG motif, likewise identified via XL-MS [73]. Nonetheless, the contribution of a single FGG motif in binding fibronectin must be low as a similar motif sequence present in domain 28 and domain 29 do not result in a high fibronectin binding of strain S28 and strain S29.

In conclusion, BadA is suggested to bind fibronectin in a cumulative fashion with quick saturation making use of unpaired β-sheet hairpins present in the DALL motifs of BadA neck/stalk domains 19 and 27 (and homologous domains). Moreover, this type of DALL motif sequence is present at least once in every *badA* sequence analysed in this work. Because of the highly repetitive nature of the BadA neck/stalk region, it might be hypothesised that the cumulation of DALL motifs would lead to a linear increase of the overall fibronectin binding capacity. However, an increasing number of DALL motifs only slightly enhanced the fibronectin binding capacity, as was demonstrated by the minimal difference in fibronectin binding between the wild type strain (30 BadA neck/stalk domains) and strain S27 (4 BadA neck/stalk domains). Therefore, the enormous size of BadA and the extensive number of repeats in the BadA neck/stalk region might primarily function as an evolutionary 'toolbox' for TAA adaptation or as a long 'grab' to facilitate adhesion of *B. henselae* to fibronectin.

4.3. Outlook

Additional infection experiments using HUVECs or binding assays using dynamic flow conditions (lab-on-chip) could be pursued to mimic bacterial adhesion under *in vivo*-like conditions and might even partly replace animal infection models. Follow-up research via evolutionary modelling or in-depth genomic analyses might clarify the function and origin of the 18-bp repeat motif. Moreover, the functional role of the *badA* island's genomic neighbourhood on the regulation of *badA* expression and genomic modification remains unexplored. Future research should also be focused on the production of recombinant BadA mutant proteins and single structural motifs with the aim of elucidating BadA-fibronectin binding sites more in detail, for instance via atomic force microscopy. Further identification of common binding motifs between BadA and fibronectin will provide a basis towards the design of novel 'anti-adhesive' compounds that might prevent the initial adherence of *B. henselae* and other TAA-producing pathogens during infection of host cells.

Summary

Adhesion to host cells is the first and most crucial step in infections with pathogenic Gram-negative bacteria and is often mediated by trimeric autotransporter adhesins (TAAs). TAA-producing bacteria are the causative agent of many human diseases and TAA-targeted anti-adhesive compounds might counteract such bacterial infections. The modularly structured *Bartonella* adhesin A (BadA) is one of the best characterised TAAs and serves as an attractive adhesin to study the domain-function relationship of TAAs during infection. BadA is a major virulence factor of *B. henselae* and is essential for the initial attachment to host cells via adhesion to extracellular matrix proteins. *B. henselae* is the causative agent of cat scratch disease and adheres to fibronectin using its long BadA fibres. The life cycle of this pathogen, with alternating host conditions, drives evolutionary and host-specific adaptations.

Human, feline, and laboratory adapted *B. henselae* isolates display genomic and phenotypic differences. By analysing the genomes of eight *B. henselae* strains using long-read sequencing, a variable genomic *badA* island with a diversified and highly repetitive *badA* gene flanked by *badA* pseudogenes was identified. Moreover, numerous conserved flanking genes were characterised, however, their influence on the regulation of *badA* expression and modification remains to be explored. It seems that *B. henselae* G-5436 is the evolutionary ancestor of the other *B. henselae* strains analysed in this work. The diversity of the *badA* island among the *B. henselae* strains indicates that the downstream *badA*-like domain region might be used as a 'toolbox' for rearrangements in the *badA* gene. Overall, it is suggested that *badA*-domain duplications, insertions, and/or deletions are the result of active phase variation via site-specific recombination and contribute to rapid host adaptation in the scope of pathogenicity, immune evasion, and/or enhanced long-term colonisation.

The model strain *B. henselae* Marseille expresses a *badA* gene that includes 30 repetitive neck/stalk domains, each consisting of several predicted structural motifs. To further elucidate the motif sequences that mediate fibronectin binding, various modified *badA* constructs were generated. Their ability to bind fibronectin was assessed via whole-cell ELISA and fluorescence microscopy. In conclusion, it is suggested that BadA adheres to fibronectin in a cumulative fashion with quick saturation via unpaired β-strands appearing in structural motifs present in BadA neck/stalk domains 19, 27, and other homologous domains. Furthermore, antibodies targeting a 15-mer amino acid sequence in the DALL motif of BadA neck/stalk domain 27 were able to reduce fibronectin binding of the *B. henselae* mutant strain S27. Moreover, this DALL motif sequence is conserved in the genome of all analysed *B. henselae* strains. The identification of common binding motifs between BadA and fibronectin supports the development of new anti-adhesive compounds that might inhibit the initial adherence of *B. henselae* and other TAA-producing pathogens during infection.

Zusammenfassung

Einleitung

Die Adhäsion von Infektionserregern an Wirtszellen ist der erste und wichtigste Schritt bei Infektionen und wird bei Infektionen mit pathogenen gramnegativen Bakterien häufig durch trimere Autotransporter-Adhäsine (TAAs) vermittelt. TAA-exprimierende Bakterien sind die Verursacher vieler menschlicher Krankheiten, wie Katzenkratzkrankheit (hervorgerufen durch *Bartonella henselae*), Enterokolitis (hervorgerufen durch z.B. *Yersinia enterocolitica*), Meningitis (hervorgerufen durch z.B. *Neisseria meningitis*) und Blutstrominfektionen (hervorgerufen durch z.B. multiresistente *Acinetobacter baumannii*). Dementsprechend könnten auf TAA ausgerichtete Antiadhäsionsstrategien eine universelle Strategie in der Therapie vieler bakterieller Infektionen darstellen.

TAAs weisen eine gemeinsame modulare Architektur auf, die eine lange N-terminale *passenger* Domäne und eine C-terminale Ankerdomäne beinhaltet. Das *Yersinia*-Adhäsin A (YadA) von *Y. enterocolitica* gilt als prototypisches TAA, während z.B. der *Acinetobacter* trimere Autotransporter (Ata) von *A. baumannii*, das *Neisseria*-Adhäsin A (NadA) von *N. meningitidis* und das *Salmonella*-Adhäsin A (SadA) von *S. enterica* andere bekannte Beispiele sind. Das modular aufgebaute *Bartonella*-Adhäsin A (BadA) ist eines der am besten charakterisierten TAAs und ist zur Untersuchung der Domänen-Funktions-Beziehung von TAAs in Infektionen sehr gut geeignet. Die *passenger* Domäne von BadA besteht aus einer Kopfdomäne und einer langen Hals-/Stielregion. Domänen aus der Hals-/Stielregion teilen sich spezifische Sequenzmotive mit charakteristischen Konformationen, die durch den *domain dictionary*-Ansatz des daTAA-Servers annotiert wurden, darunter FGG-Motive, *coiled-coil*-Motive und DALL-Neck-Tandemkonnektoren.

BadA vermittelt die Adhäsion von *B. henselae* an Wirtszellen und extrazelluläre Matrixproteine. *B. henselae*, der Erreger der Katzenkratzkrankheit, adhäriert mit seinen ca. 150-250 nm langen BadA-Adhäsinen an Fibronektin. Darüber hinaus wurde nachgewiesen, dass ausschließlich die BadA Hals-/Stielregion und nicht die Kopfdomäne für die Adhäsion von *B. henselae* an Fibronektin verantwortlich ist. Fibronektin ist nachweislich ein wichtiger erster Bindungspartner von *B. henselae* während der Infektion menschlicher Endothelzellen, und die BadA-Fibronektin-Interaktion erfolgt über die Heparin-bindenden Domänen. Fibronektin ist ein heterodimeres Glykoprotein, das auf der Zelloberfläche von Endothelzellen als fibrilläre Matrix (zelluläres Fibronektin) oder in Blut, Speichel und anderen Flüssigkeiten (Plasma Fibronektin) reichlich vorhanden ist, was es zu einem ausgezeichneten ersten Bindungspartner bei Infektionen von Blutgefäßen, Herzklappen oder im Falle eines Katzenkratzers in der menschlichen Haut macht.

Zielsetzung

Bislang sind die genauen BadA Hals-/Stieldomänen oder Motive, die für die Fibronektinbindung verantwortlich sind, noch unbekannt. Außerdem muss zur Bestimmung spezifischer BadA-Bindungsmotive zunächst überprüft werden, ob solche Motive innerhalb der Spezies von *B. henselae* konserviert sind. Die zugrundeliegenden Wiederholungen auf der *badA*-Sequenzebene deuten auf häufige Genumlagerungen durch Rekombination hin, was eine korrekte Speziestypisierung oder phylogenetische Analyse schwierig macht. Darüber hinaus könnte die Lebensweise von *B. henselae* mit häufigen Übergängen vom Katzenfloh zur Katze und zu zufälligen menschlichen Wirten effiziente und schnelle Anpassungsstrategien erfordern. Außerdem wurde festgestellt, dass die Expression von *badA* unter Bedingungen, die der menschlichen Wirtszelle entsprechen hochreguliert und unter Bedingungen die denen des Katzenflohs entsprechen herunterreguliert wird.

Ergebnisse und Diskussion

Im Rahmen dieser Arbeit wurden die Genome mehrerer *B. henselae*-Stämme mit *long-read*-Sequenzierungstechniken sequenziert, um Unterschiede im hochvariablen *badA*-Gen zu analysieren und die *badA*-Expression sowie die funktionelle Bindung an extrazelluläre Matrixproteine zu untersuchen. Die *long-read*-Sequenzierung mit PacBio liefert 20 bis 25 kb lange *reads*, die stark repetitive Abschnitte abdecken und einige Probleme bei der Genomassemblierung im Zusammenhang mit *short-read*-Sequenzierungsdaten umgehen. So wurde beispielsweise deutlich, dass *B. henselae* Marseille ein *badA*-Gen exprimiert, das tatsächlich aus 11.922 bp und 30 Hals-/Stieldomänen besteht, anstatt der früher ermittelten Länge von 9.249 bp und 22 Hals-/Stieldomänen. Es ist zu vermuten, dass bei der Assemblierung der *short-read*-Sanger-Sequenzierungsdaten zwei sich wiederholende Regionen von 1,4 kb bzw. 1,3 kb ausgelassen wurden.

Humane, katzenartige, und laboradaptierte *B. henselae*-Isolate weisen also genomische und phänotypische Unterschiede auf. Durch die Analyse der Genome von acht *B. henselae*-Stämmen mittels *next-generation long-read*-Sequenzierung wurde eine variable genomische *badA*-Insel mit einem diversifizierten und stark repetitiven *badA*-Gen identifiziert, das von *badA*-Pseudogenen flankiert wird. Darüber hinaus wurden zahlreiche flankierende Gene identifiziert, deren Einfluss auf die Regulierung der *badA*-Expression und -Modifikation jedoch noch erforscht werden muss. Die Vielfalt der *badA*-Insel deutet darauf hin, dass die nachgelagerte *badA*-ähnliche Domänenregion als Rekombinationspool für Umlagerungen in der Zusammensetzung des *badA*-Gens verwendet werden könnte. Insgesamt dürften Duplikationen, Insertionen und/oder Löschungen der *badA*-Domäne das Ergebnis einer

aktiven Phasenvariation durch Rekombinationen darstellen, die zu einer schnellen Wirtsanpassung beigetragen hat (z.B. in Bezug auf Pathogenität, Immunevasion, und/oder Kolonisierung).

Insbesondere die BadA-Proteine von *B. henselae* Berlin-I und ATCC49882^T var-2 unterscheiden sich erheblich von den BadA-Proteinen anderer *B. henselae*-Stämme hinsichtlich ihrer Größe, der Anzahl der repetitiven Domänen und der gesamten Domänenorganisation. Das *badA*-Gen der Stämme Berlin-I und ATCC49882^T var-2 könnte das Ergebnis eines Rekombinationsereignisses sein, bei dem ein großer Teil des *badA*-Gens und eine frühere *badA*-ähnliche Domänenregion deletiert und der erste Teil des *badA*-Gens mit dem letzten Teil einer früheren *badA*-ähnlichen Domänenregion kombiniert wurde. Ein ähnliches Rekombinationsereignis könnte in der *badA*-Insel von Stamm 88-64 Oklahoma stattgefunden haben, die einen großen Teil von *badA* mit dem letzten Teil einer früheren *badA*-ähnlichen Domänenregion verbindet. Darüber hinaus könnte das aktuelle *badA*-Gen der Stämme Marseille und FR96/BK3 von einem gemeinsamen *badA*-Gen (z.B. dem *badA*-Gen von Stamm G-5436) abstammen, das mehrere Rekombinationen durchlaufen hat. Hinweise darauf liefert das Vorhandensein mehrerer *badA*-Hals-/Stieldomänen-Wiederholungen, die nur einmal im *badA*-Gen von Stamm G-5436 vorhanden sind. Es scheint, dass *B. henselae* G-5436 der evolutionäre Vorfahre der in dieser Arbeit analysierten Stämme ist.

Trotz der beobachteten Variationen in der Zusammensetzung von *badA* sind bestimmte Hals-/Stielbereiche in allen untersuchten *badA*-Gensequenzen konserviert. Zur Klärung der Motivsequenzen, die die Fibronektinbindung vermitteln, wurden verschiedene verkürzte und modifizierte *badA*-Konstrukte erzeugt und in *B. henselae* Marseille ΔBadA-T exprimiert. Die Expression aller verkürzten und modifizierten *badA*-Mutanten wurde durch *Western blotting*, Immunfluoreszenzmikroskopie, konfokale *laser-scanning*-Mikroskopie und Transmissions Elektronenmikroskopie überprüft. Die Fähigkeit entsprechender *badA*-Mutanten-exprimierender *B. henselae* Stämme, humanes Fibronektin zu binden, wurde mittels Ganzzell-ELISA untersucht und durch Fluoreszenzmikroskopie verifiziert. Generell zeigten alle untersuchten *B. henselae* Marseille ΔBadA-T-Mutantenstämme eine deutlich geringere Fibronektinbindung im Vergleich zum Wildtyp-Stamms Marseille.

Trotz der Deletion von 26 *badA* Hals-/Stieldomänen zeigten die Stämme S27 und D19S28 eine relativ hohe Fibronektinbindung (ca. 80 %) im Vergleich zur Fibronektinbindung des Wildtyp-Stamms Marseille. Im Gegensatz dazu zeigte der Stamm S28 im Vergleich zu den Stämmen S27 und D19S28 eine deutlich geringere Fibronektinbindung, obwohl ihm nur eine *badA* Hals-/Stieldomäne fehlt. Daher kann angenommen werden, dass die BadA Hals-/Stieldomänen 19 und 27 wichtige Regionen für die anfängliche Bindung von *B. henselae* an Fibronektin sind. Die BadA Hals-/Stieldomänen 6 und 10 sind homolog zu Domäne 27,

während die Domänen 15 und 23 homolog zu Domäne 19 sind. Darüber hinaus zeigte der Stamm D27S29 im Vergleich zum Stamm S29 eine deutlich höhere Fibronektinbindung, obwohl er nur die zusätzliche BadA Hals-/Stieldomäne 27 enthielt, was die wichtige Rolle der Domäne 27 bei der Adhäsion von BadA an Fibronektin bestätigt.

Die Domänen 19 und 27 haben eine sehr ähnliche DALL-Hals-Tandemverbindung. Darüber hinaus wird vorhergesagt, dass das DALL-Motiv aus zwei β-Strängen besteht, die eine Haarnadelstruktur bilden. Dies gilt als optimale Interaktionsstelle für die Adhäsion an die zahlreichen ungepaarten β-Stränge, die in Fibronektin vorhanden sind. Um die Rolle der DALL-Motive bei der Bindung von Fibronektin zu analysieren, wurden anti-BadA-DALL-Antikörpern erzeugt, die eine 15-mer Peptidsequenz (RHEKSKLEKGASKAI) aus dem DALL-Motiv der Domäne 27 erkennen. Die schrittweise Erhöhung der Menge von anti-BadA-DALL-Antikörpern führte zu einer korrelierenden Abnahme des an den Stamm S27 gebundenen Fibronektin, was darauf hindeutet, dass die 15-mer-Sequenz des DALL-Motivs der Domäne 27 an der Fibronektinbindung wesentlich beteiligt ist und als *druggable target* im Sinne der Verhinderung der Adhäsion an Wirtszellen in Frage kommen könnte. Eine stärkere Abnahme der Fibronektinbindung wurde bei der Verwendung von anti-BadA-Antikörpern beobachtet, die gegen das ganze BadA gerichtet waren. Die genaue Antikörperkonzentration und die BadA-Zielorte sind jedoch unbekannt.

Es kann vermutet werden, dass BadA an Fibronektin auf kumulative Weise mit schneller Sättigung über ungepaarte β-Stränge adhäriert, die in repetitiven strukturellen Motiven in den BadA Hals-/Stieldomänen 19, 27 und anderen homologen Domänen vorhanden sind. Antikörper, die eine 15-mer Aminosäuresequenz aus dem DALL-Motiv der BadA Hals-/Stieldomäne 27 erkennen, reduzierten die Fibronektinbindung. Diese DALL-Motivsequenz liegt in den *badA*-Sequenzen aller in dieser Arbeit analysierten *B. henselae* Stämme konserviert vor. Die Identifizierung gemeinsamer Bindungsmotive zwischen BadA und Fibronektin unterstützt die Entwicklung neuer antiadhäsiver Therapiestrategien, die die anfängliche Adhäsion von *B. henselae* und anderen TAA-exprimierenden Humanpathogenen während der initialen Phase einer Infektion hemmen könnten.

References

1. Drancourt M, Tran-Hung L, Courtin J, Lumley H de, Raoult D (2005) *Bartonella quintana* in a 4000-year-old human tooth. J Infect Dis 191:607–11. https://doi.org/10.1086/427041

2. Vorou RM, Papavassiliou VG, Tsiodras S (2007) Emerging zoonoses and vector-borne infections affecting humans in Europe. Epidemiol Infect 135:1231–1247. https://doi.org/10.1017/S0950268807008527

3. Wormser GP (2007) Discovery of new infectious diseases - *bartonella* species. N Engl J Med 356:2346–2347. https://doi.org/10.1056/NEJMP078069

4. Breitschwerdt EB (2017) Bartonellosis, one health and all creatures great and small. Vet Dermatol 28:96-e21. https://doi.org/10.1111/vde.12413

5. Okaro U, Addisu A, Casanas B, Anderson B (2017) *Bartonella* species, an emerging cause of blood-culture-negative endocarditis. Clin Microbiol Rev 30:709–746. https://doi.org/10.1128/CMR.00013-17

6. Buffet J, Kosoy M, Vayssier-Taussat M (2013) Natural history of *Bartonella*-infecting rodents in light of new knowledge on genomics, diversity, and evolution. Future Microbiol 8:1117–1128. https://doi.org/10.2217/fmb.13.77

7. Kešnerová L, Moritz R, Engel P (2016) *Bartonella apis* sp. nov., a honey bee gut symbiont of the class Alphaproteobacteria. Int J Syst Evol Microbiol 66:414–421. https://doi.org/10.1099/ijsem.0.000736

8. Zhu Q, Kosoy M, Olival KJ, Dittmar K (2014) Horizontal transfers and gene losses in the phospholipid pathway of *Bartonella* reveal clues about early ecological niches. Genome Biol Evol 6:2156–2169. https://doi.org/10.1093/gbe/evu169

9. Gutiérrez R, Vayssier-Taussat M, Buffet JP, Harrus S (2017) Guidelines for the isolation, molecular detection, and characterization of *Bartonella* species. Vector-Borne Zoonotic Dis 17:42–50. https://doi.org/10.1089/vbz.2016.1956

10. Mullins KE, Hang J, Clifford RJ, Onmus-Leone F, Yang Y, Jiang J, Leguia M, Kasper MR, Maguina C, Lesho EP, Jarman RG, Richards A, Blazes D (2017) Whole-genome analysis of *Bartonella ancashensis* , a novel pathogen causing verruga peruana, rural ancash region, Peru. Emerg Infect Dis 23:430–438. https://doi.org/10.3201/eid2303.161476

11. Hang J, Mullins KE, Clifford RJ, Onmus-Leone F, Yang Y, Jiang J, Leguia M, Kasper MR, Maguiña C, Lesho EP, Jarman RG, Richards AL, Blazes D (2015) Complete genome sequence of *Bartonella ancashensis* strain 20.00, isolated from the blood of a patient with verruga peruana. Genome Announc 3:. https://doi.org/10.1128/genomeA.01217-15

12. Dichter AA, Schultze TG, Becker SA, Tsukayama P, Kempf VAJ (2020) Complete genome sequence of *Bartonella bacilliformis* strain KC584 (ATCC 35686). Microbiol Resour Announc 9:. https://doi.org/10.1128/MRA.01377-19

13. Engel P, Salzburger W, Liesch M, Chang C-C, Maruyama S, Lanz C, Calteau A, Lajus A, Médigue C, Schuster SC, Dehio C (2011) Parallel evolution of a type IV secretion system in radiating lineages of the host-restricted bacterial pathogen *Bartonella*. PLoS Genet 7:e1001296. https://doi.org/10.1371/journal.pgen.1001296

14. Buffet JP, Pisanu B, Brisse S, Roussel S, Félix B, Halos L, Chapuis JL, Vayssier-Taussat M (2013) Deciphering *Bartonella* diversity, recombination, and host specificity in a rodent community. PLoS One 8:. https://doi.org/10.1371/journal.pone.0068956

15. Wagner A, Dehio C (2019) Role of distinct type-IV-secretion systems and secreted effector sets in host adaptation by pathogenic *Bartonella* species. Cell Microbiol 21:1–9. https://doi.org/10.1111/cmi.13004

16. Québatte M, Dehio C (2019) *Bartonella* gene transfer agent: evolution, function, and proposed role in host adaptation. Cell Microbiol 21:1–9. https://doi.org/10.1111/cmi.13068

17. Pulliainen AT, Dehio C (2012) Persistence of *Bartonella* spp. stealth pathogens: from subclinical infections to vasoproliferative tumor formation. FEMS Microbiol Rev 36:563–599. https://doi.org/10.1111/j.1574-6976.2012.00324.x

18. Harms A, Dehio C (2012) Intruders below the radar: molecular pathogenesis of *Bartonella* spp. Clin Microbiol Rev 25:42–78. https://doi.org/10.1128/CMR.05009-11

19. Kaiser PO, Riess T, O'Rourke F, Linke D, Kempf VAJ (2011) *Bartonella* spp.: throwing light on uncommon human infections. Int J Med Microbiol 301:7–15. https://doi.org/10.1016/j.ijmm.2010.06.004

20. Garcia-Quintanilla M, Dichter AA, Guerra H, Kempf VAJ (2019) Carrion's disease: more than a neglected disease. Parasites and Vectors 12:141. htpps://doi.org/10.1186/s13071-019-3390-2

21. Harms A, Segers FHID, Quebatte M, Mistl C, Manfredi P, Körner J, Chomel BB, Kosoy M, Maruyama S, Engel P, Dehio C (2017) Evolutionary dynamics of pathoadaptation revealed by three independent acquisitions of the VirB/D4 type IV secretion system in *Bartonella*. Genome Biol Evol 9:761–776. https://doi.org/10.1093/gbe/evx042

22. Guy L, Nystedt B, Toft C, Zaremba-Niedzwiedzka K, Berglund EC, Granberg F, Näslund K, Eriksson A-S, Andersson SGE (2013) A gene transfer agent and a dynamic repertoire of secretion systems hold the keys to the explosive radiation of the emerging pathogen *Bartonella*. PLoS Genet 9:e1003393. https://doi.org/10.1371/journal.pgen.1003393

23. van der Woude MW (2011) Phase variation: how to create and coordinate population diversity. Curr Opin Microbiol 14:205–211. https://doi.org/10.1016/j.mib.2011.01.002

24. Alsmark CM, Frank AC, Karlberg EO, Legault BA, Ardell DH, Canbäck B, Eriksson AS, Näslund AK, Handley SA, Huvet M, La Scola B, Holmberg M, Andersson SGE (2004) 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. https://doi.org/10.1073/pnas.0305659101

25. Segers FH, Kešnerová L, Kosoy M, Engel P (2017) Genomic changes associated with the evolutionary transition of an insect gut symbiont into a blood-borne pathogen. ISME J 11:1232-1244. https://doi.org/10.1038/ismej.2016.201

26. Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SGE (2004) Computational inference of scenarios for α-proteobacterial genome evolution. Proc Natl Acad Sci USA 101:9722–9727. https://doi.org/10.1073/pnas.0400975101

27. Lindroos H, Vinnere O, Mira A, Repsilber D, Näslund K, Andersson SGE (2006) Genome rearrangements, deletions, and amplifications in the natural population of *Bartonella henselae*. J Bacteriol 188:7426–7439. https://doi.org/10.1128/JB.00472-06

28. Kosoy M, Hayman DTS, Chan KS (2012) *Bartonella* bacteria in nature: where does population variability end and a species start? Infect Genet Evol 12:894–904. https://doi.org/10.1016/j.meegid.2012.03.005

29. Regnery RL, Anderson BE, Clarridge JE, Rodriguez-Barradas MC, Jones DC, Carr JH (1992) Characterization of a novel *Rochalimaea* species, *R. henselae* sp. nov., isolated from blood of a febrile, human immunodeficiency virus-positive patient. J Clin Microbiol 30:265–274. https://doi.org/10.1128/jcm.30.2.265-274.1992

30. Drancourt M, Birtles R, Raoult D, Chaumentin G, Vandenesch F, Etienne J (1996) New serotype of *Bartonella henselae* in endocarditis and cat-scratch disease. Lancet 347:441–443. https://doi.org/10.1016/S0140-6736(96)90012-4

31. Bouchouicha R, Durand B, Monteil M, Chomel BB, Berrich M, Arvand M, Birtles RJ, Breitschwerdt EB, Koehler JE, Maggi R, Maruyama S, Kasten R, Petit E, Boulouis HJ, Haddad N (2009) Molecular epidemiology of feline and human *bartonella henselae* isolates. Emerg Infect Dis 15:813–816. https://doi.org/10.3201/eid1505.080995

32. Engel P, Dehio C (2009) Genomics of host-restricted pathogens of the genus *Bartonella*. Microb Pathog 6:158–169. https://doi.org/10.1159/000235769

33. Sander A, Ruess M, Bereswill S, Schuppler M, Steinbrueckner B (1998) Comparison of different DNA fingerprinting techniques for molecular typing of *Bartonella henselae* isolates. J Clin Microbiol 36:2973–2981. https://doi.org/10.1128/jcm.36.10.2973-2981.1998

34. Handley SA, Regnery RL (2000) Differentiation of pathogenic *Bartonella* species by infrequent restriction site PCR. J Clin Microbiol 38:3010–3015. https://doi.org/10.1128/jcm.38.8.3010-3015.2000

35. Kyme P, Dillon B, Iredell J (2003) Phase variation in *Bartonella henselae*. Microbiology 149:621-629. https://doi.org/10.1099/mic.0.26014-0

36. Dillon B, Iredell J (2004) *Dde*I RFLP for 16S rDNA typing in *Bartonella henselae*. J Med Microbiol 53:1263–1265. https://doi.org/10.1099/jmm.0.45606-0

37. Iredell J, Blanckenberg D, Arvand M, Grauling S, Feil EJ, Birtles RJ (2003) Characterization of the natural population of *Bartonella henselae* by multilocus sequence typing. J Clin Microbiol 41:5071–5079. https://doi.org/10.1128/JCM.41.11.5071-5079.2003

38. Li W, Raoult D, Fournier PE (2007) Genetic diversity of *Bartonella henselae* in human infection detected with multispacer typing. Emerg Infect Dis 13:1178–1183. https://doi.org/10.3201/eid1308.070085

39. Omasits U, Varadarajan AR, Schmid M, Goetze S, Melidis D, Bourqui M, Nikolayeva O, Québatte M, Patrignani A, Dehio C, Frey JE, Robinson MD, Wollscheid B, Ahrens CH (2017) An integrative strategy to identify the entire protein coding potential of prokaryotic genomes by proteogenomics. Genome Res 27:2083–2095. https://doi.org/10.1101/gr.218255.116

40. Boulouis HJ, Chomel BB, Guillaume G, Benoît D, Chang C chin, Monteil M, Kasten RW, Jack A, Nadia H (2020) Multiple locus variable number tandem repeat analysis for the characterization of wild feline *Bartonella* species and subspecies. Vet Microbiol 247:108788. https://doi.org/10.1016/j.vetmic.2020.108788

41. Viezens J, Arvand M (2008) Simultaneous presence of two different copies of the 16S rRNA gene in *Bartonella henselae*. Microbiology 154:2881–2886. https://doi.org/10.1099/mic.0.2008/018630-0

42. Berrich M, Kieda C, Grillon C, Monteil M, Lamerant N, Gavard J, Boulouis HJ, Haddad N (2011) differential effects of *Bartonella henselae* on human and feline macro- and micro-vascular endothelial cells. PLoS One 6:e20204. https://doi.org/10.1371/journal.pone.0020204

43. Chang C-C, Chen Y-J, Tseng C-S, Lai W-L, Hsu K-Y, Chang C-L, Lu C-C, Hsu Y-M (2011) A comparative study of the interaction of *Bartonella henselae* strains with human endothelial cells. Vet Microbiol 149:147–156. https://doi.org/10.1016/j.vetmic.2010.09.033

44. Huwyler C, Heiniger N, Chomel BB, Kim M, Kasten RW, Koehler JE (2017) dynamics of co-infection with *Bartonella henselae* genotypes I and II in naturally infected cats: implications for feline vaccine development. Microb Ecol 74:474–484. https://doi.org/10.1007/s00248-017-0936-8

45. Chomel BB, Boulouis H-J, Breitschwerdt EB, Kasten RW, Vayssier-Taussat M, Birtles RJ, Koehler JE, Dehio C (2009) Ecological fitness and strategies of adaptation of *Bartonella* species to their hosts and vectors I. Vet Res 40:29. https://doi.org/10.1051/vetres/2009011

46. Lu YY, Franz B, Truttmann MC, Riess T, Gay-Fraret J, Faustmann M, Kempf VAJ, Dehio C (2013) *Bartonella henselae* trimeric autotransporter adhesin BadA expression interferes with effector translocation by the VirB/D4 type IV secretion system. Cell Microbiol 15:759–778. https://doi.org/10.1111/cmi.12070

47. Koehler JE, Glaser CA, Tappero JW (1994) *Rochalimaea henselae* infection: a new zoonosis with the domestic cat as reservoir. JAMA 271:531–535. https://doi.org/10.1001/jama.1994.03510310061039

48. Yamamoto K, Chomel BB, Kasten RW, Hew CM, Weber DK, Lee WI, Koehler JE, Pedersen NC (2003) Infection and re-infection of domestic cats with various *Bartonella* species or types: *B. henselae* type I is protective against heterologous challenge with *B. henselae* type II. Vet Microbiol 92:73–86. https://doi.org/10.1016/S0378-1135(02)00347-4

49. Chomel BB, Kasten RW, Floyd-Hawkins K, Chi B, Yamamoto K, Roberts-Wilson J, Gurfield AN, Abbott RC, Pedersen NC, Koehler JE (1996) Experimental transmission of *Bartonella henselae* by the cat flea. J Clin Microbiol 34:1952–1956. https://doi.org/10.1128/jcm.34.8.1952-1956.1996

50. Finkelstein JL, Brown TP, O'Reilly KL, Wedincamp J, Foil LD (2002) Studies on the growth of *Bartonella henselae* in the cat flea (Siphonaptera: Pulicidae). J Med Entomol 39:915–919. https://doi.org/10.1603/0022-2585-39.6.915

51. Sanogo YO, Zeaiter Z, Caruso G, Merola F, Shpynov S, Brouqui P, Raoult D (2003) *Bartonella henselae* in *Ixodes ricinus* ticks (Acari: Ixodida) removed from humans, Belluno Province, Italy. Emerg Infect Dis 9:329–332. https://doi.org/10.3201/eid0903.020133

52. Hercík K, Hášová V, Janeček J, Branny P (2007) Molecular evidence of *Bartonella* DNA in ixodid ticks in Czechia. Folia Microbiol (Praha) 52:503–509. https://doi.org/10.1007/BF02932111

53. Brenner EC, Chomel BB, Singhasivanon OU, Namekata DY, Kasten RW, Kass PH, Cortés-Vecino JA, Gennari SM, Rajapakse RP, Huong LT, Dubey JP (2013) *Bartonella* infection in urban and rural dogs from the tropics: Brazil, Colombia, Sri Lanka and Vietnam. Epidemiol Infect 141:54–61. https://doi.org/10.1017/S0950268812000519

54. Pérez C, Maggi RG, Diniz PPVP, Breitschwerdt EB (2011) molecular and serological diagnosis of *Bartonella* infection in 61 dogs from the United States. J Vet Intern Med 25:805–810. https://doi.org/10.1111/J.1939-1676.2011.0736.X

55. Godet C, Roblot F, Le Moal G, Roblot P, Frat JP, Becq-Giraudon B (2004) Cat-scratch disease presenting as a breast mass. Scand J Infect Dis 36:493–494. https://doi.org/10.1080/00365540410020235

56. Rizzo MF, Osikowicz L, Cáceres AG, Luna-Caipo VD, Suarez-Puyen SM, Bai Y, Kosoy M (2019) Identification of *Bartonella rochalimae* in guinea pigs (*cavia porcellus*) and fleas collected from rural Peruvian households. Am J Trop Med Hyg 101:1276–1281. https://doi.org/10.4269/AJTMH.19-0517

57. Lamps LW, Scott MA (2004) Cat-scratch disease: historic, clinical, and pathologic perspectives. Am J Clin Pathol 121 Suppl:71–80. https://doi.org/10.1309/JC8YM53L4E0L6PT5

58. Kempf VAJ, Schairer A, Neumann D, Grassl GA, Lauber K, Lebiedziejewski M, Schaller M, Kyme P, Wesselborg S, Autenrieth IB (2005) *Bartonella henselae* inhibits apoptosis in Mono Mac 6 cells. Cell Microbiol 7:91–104. https://doi.org/10.1111/j.1462-5822.2004.00440.x

59. Relman DA, Loutit JS, Schmidt TM, Falkow S, Tompkins LS (1990) The agent of bacillary angiomatosis. N Engl J Med 323:1573–1580. https://doi.org/10.1056/NEJM199012063232301

60. Hadfield TL, Warren R, Kass M, Brun E, Levy C (1993) Endocarditis caused by *Rochalimaea henselae*. Hum Pathol 24:1140–1141. https://doi.org/10.1016/0046-8177(93)90196-N

61. Anderson BE, Neuman MA (1997) *Bartonella* spp. as emerging human pathogens. Clin Microbiol Rev 10:203–219. https://doi.org/10.1128/cmr.10.2.203-219.1997

62. Kempf V, Petzold H, Autenrieth I (2014) Cat scratch disease due to *Bartonella henselae* infection mimicking parotid malignancy. Eur J Clin Microbiol Infect Dis 2001 2010 20:732–733. https://doi.org/10.1007/S100960100605

63. Nelson CA, Saha S, Mead PS (2016) Cat-scratch disease in the United States, 2005–2013. Emerg Infect Dis 22:1741–1746. https://doi.org/10.3201/eid2210.160115

64. Jackson LA, Perkins BA, Wenger JD (1993) Cat scratch disease in the United States: an analysis of three national databases. Am J Public Health 83:1707–1711. https://doi.org/10.2105/AJPH.83.12.1707

65. Jost M, Latz A, Ballhorn W, Kempf VAJ (2018) Development of a specific and sensitive enzyme-linked immunosorbent assay as an *in vitro* diagnostic tool for detection of *Bartonella henselae* antibodies in human serum. J Clin Microbiol 56. https://doi.org/10.1128/JCM.01329-18

66. Mehock JR, Greene CE, Gherardini FC, Hahn T-W, Krause DC (1998) *Bartonella henselae* invasion of feline erythrocytes *in vitro*. Infect Immun 66:3462–3466. https://doi.org/10.1128/IAI.66.7.3462-3466.1998

67. Chenoweth MR, Greene CE, Krause DC, Gherardini FC (2004) Predominant outer membrane antigens of *Bartonella henselae*. Infect Immun 72:3097–3105. https://doi.org/10.1128/IAI.72.6.3097-3105.2004

68. Mändle T, Einsele H, Schaller M, Neumann D, Vogel W, Autenrieth IB, Kempf VAJ (2005) Infection of human CD34+ progenitor cells with *Bartonella henselae* results in intraerythrocytic presence of *B. henselae*. Blood 106:1215–1222. https://doi.org/10.1182/BLOOD-2004-12-4670

69. Kordick DL, Brown TT, Shin K, Breitschwerdt EB (1999) Clinical and pathologic evaluation of chronic *Bartonella henselae* or *Bartonella clarridgeiae* infection in cats. J Clin Microbiol 37:1536–1547. https://doi.org/10.1128/JCM.37.5.1536-1547.1999

70. Dehio C (2001) *Bartonella* interactions with endothelial cells and erythrocytes. Trends Microbiol 9:279–285. https://doi.org/10.1016/S0966-842X(01)02047-9

71. Burgess AWO, Anderson BE (1998) Outer membrane proteins of *Bartonella henselae* and their interaction with human endothelial cells. Microb Pathog 25:157–164. https://doi.org/10.1006/MPAT.1998.0223

72. Kaiser PO, Riess T, Wagner CL, Linke D, Lupas AN, Schwarz H, Raddatz G, Schäfer A, Kempf VAJ (2008) The head of *Bartonella* adhesin A is crucial for host cell interaction of *Bartonella henselae*. Cell Microbiol 10:2223–2234. https://doi.org/10.1111/j.1462-5822.2008.01201.x

73. Vaca DJ, Thibau A, Leisegang MS, Malmström J, Linke D, Eble JA, Ballhorn W, Schaller M, Happonen L, Kempf VAJ (2022) Interaction of *Bartonella henselae* with fibronectin represents the molecular basis for adhesion to host cells. Microbiol Spectr. 10. https://doi.org/10.1128/spectrum.00598-22

74. Dehio C, Meyer M, Berger J, Schwarz H, Lanz C (1997) Interaction of *Bartonella henselae* with endothelial cells results in bacterial aggregation on the cell surface and the subsequent engulfment and internalisation of the bacterial aggregate by a unique structure, the invasome. J Cell Sci 110:2141–2154. https://doi.org/10.1242/JCS.110.18.2141

75. Truttmann MC, Rhomberg TA, Dehio C (2011) Combined action of the type IV secretion effector proteins BepC and BepF promotes invasome formation of *Bartonella henselae* on endothelial and epithelial cells. Cell Microbiol 13:284–299. https://doi.org/10.1111/J.1462-5822.2010.01535.X

76. Riess T, Andersson SGE, Lupas A, Schaller M, Schäfer A, Kyme P, Martin J, Wälzlein J-H, Ehehalt U, Lindroos H, Schirle M, Nordheim A, Autenrieth IB, Kempf VAJ (2004) *Bartonella* adhesin A mediates a proangiogenic host cell response. J Exp Med 200:1267–1278. https://doi.org/10.1084/jem.20040500

77. Kempf VAJ, Lebiedziejewski M, Alitalo K, Wälzlein JH, Ehehalt U, Fiebig J, Huber S, Schütt B, Sander CA, Müller S, Grassl G, Yazdi AS, Brehm B, Autenrieth IB (2005) Activation of hypoxia-inducible factor-1 in bacillary angiomatosis. Circulation 111:1054–1062. https://doi.org/10.1161/01.CIR.0000155608.07691.B7

78. Schmid MC, Scheidegger F, Dehio M, Balmelle-Devaux N, Schulein R, Guye P, Chennakesava CS, Biedermann B, Dehio C (2006) A translocated bacterial protein protects vascular endothelial cells from apoptosis. PLOS Pathog 2:e115. https://doi.org/10.1371/JOURNAL.PPAT.0020115

79. Dehio C (2005) *Bartonella*–host-cell interactions and vascular tumour formation. Nat Rev Microbiol 2005 38 3:621–631. https://doi.org/10.1038/nrmicro1209

80. Schmid MC, Schulein R, Dehio M, Denecker G, Carena I, Dehio C (2004) The VirB type IV secretion system of *Bartonella henselae* mediates invasion, proinflammatory activation and antiapoptotic protection of endothelial cells. Mol Microbiol 52:81–92. https://doi.org/10.1111/j.1365-2958.2003.03964.x

81. Scheidegger F, Ellner Y, Guye P, Rhomberg TA, Weber H, Augustin HG, Dehio C (2009) Distinct activities of *Bartonella henselae* type IV secretion effector proteins modulate capillary-like sprout formation. Cell Microbiol 11:1088–1101. https://doi.org/10.1111/J.1462-5822.2009.01313.X

82. Québatte M, Dick MS, Kaever V, Schmidt A, Dehio C (2013) Dual input control: activation of the *Bartonella henselae* VirB/D4 type IV secretion system by the stringent sigma factor RpoH1 and the BatR/BatS two-component system. Mol Microbiol 90:756–775. https://doi.org/10.1111/mmi.12396

83. Quebatte M, Dehio M, Tropel D, Basler A, Toller I, Raddatz G, Engel P, Huser S, Schein H, Lindroos HL, Andersson SGE, Dehio C (2010) The BatR/BatS two-component regulatory system controls the adaptive response of *Bartonella henselae* during human endothelial cell infection. J Bacteriol 192:3352–3367. https://doi.org/10.1128/JB.01676-09

84. Saenz HL, Engel P, Stoeckli MC, Lanz C, Raddatz G, Vayssier-Taussat M, Birtles R, Schuster SC, Dehio C (2007) Genomic analysis of *Bartonella* identifies type IV secretion systems as host adaptability factors. Nat Genet 39:1469–1476. https://doi.org/10.1038/NG.2007.38

85. Vayssier-Taussat M, Rhun D Le, Deng HK, Biville F, Cescau S, Danchin A, Marignac G, Lenaour E, Boulouis HJ, Mavris M, Arnaud L, Yang H, Wang J, Quebatte M, Engel P, Saenz H, Dehio C (2010) The Trw type IV secretion system of *Bartonella* mediates host-specific adhesion to erythrocytes. PLoS Pathog 6:. https://doi.org/10.1371/journal.ppat.1000946

86. Seubert A, Hiestand R, De La Cruz F, Dehio C (2003) A bacterial conjugation machinery recruited for pathogenesis. Mol Microbiol 49:1253–1266. https://doi.org/10.1046/J.1365-2958.2003.03650.X

87. Siewert LK, Korotaev A, Sedzicki J, Fromm K, Pinschewer DD, Dehio C (2022) Identification of the *Bartonella* autotransporter CFA as a protective antigen and hypervariable target of neutralizing antibodies in mice. Proc Natl Acad Sci U S A 119:e2202059119. https://doi.org/10.1073/pnas.2202059119

88. Deng H, Pang Q, Xia H, Le Rhun D, Le Naour E, Yang C, Vayssier-Taussat M, Zhao B (2016) Identification and functional analysis of invasion associated locus B (IalB) in *Bartonella* species. Microb Pathog 98:171–177. https://doi.org/10.1016/j.micpath.2016.05.007

89. Zähringer U, Lindner B, Knirel YA, van den Akker WMR, Hiestand R, Heine H, Dehio C (2004) structure and biological activity of the short-chain lipopolysaccharide from *Bartonella henselae* ATCC49882^T. J Biol Chem 279:21046–21054. https://doi.org/10.1074/jbc.M313370200

90. Dabo SM, Confer AW, Anderson BE, Gupta S (2006) *Bartonella henselae* Pap31, an extracellular matrix adhesin, binds the fibronectin repeat III 13 module. Infect Immun 74:2513-2521. https://doi.org/10.1128/IAI.74.5.2513-2521.2006

91. Sander A (2000) Hemin-dependent growth and hemin binding of *Bartonella henselae*. FEMS Microbiol Lett 189:55–59. https://doi.org/10.1016/S0378-1097(00)00243-3

92. Berglund EC, Frank AC, Calteau A, Pettersson OV, Granberg F, Eriksson AS, Näslund K, Holmberg M, Lindroos H, Andersson SGE (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:. https://doi.org/10.1371/journal.pgen.1000546

93. Bowers TJ, Sweger D, Jue D, Anderson B (1998) Isolation, sequencing and expression of the gene encoding a major protein from the bacteriophage associated with *Bartonella henselae*. Gene 206:49–52. https://doi.org/10.1016/S0378-1119(97)00580-5

94. Dabo SM, Confer AW, Saliki JT, Anderson BE (2006) Binding of *Bartonella henselae* to extracellular molecules: identification of potential adhesins. Microb Pathog 41:10–20. https://doi.org/10.1016/j.micpath.2006.04.003

95. Linke D, Riess T, Autenrieth IB, Lupas A, Kempf VAJ (2006) Trimeric autotransporter adhesins: variable structure, common function. Trends Microbiol 14:264–270. https://doi.org/10.1016/j.tim.2006.04.005

96. Hoiczyk E (2000) Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. EMBO J 19:5989–5999. https://doi.org/10.1093/emboj/19.22.5989

97. Koiwai K, Hartmann MD, Linke D, Lupas AN, Hori K (2016) Structural basis for toughness and flexibility in the C-terminal passenger domain of an *Acinetobacter* trimeric autotransporter adhesin. J Biol Chem 291:3705–3724. https://doi.org/10.1074/jbc.M115.701698

98. Mühlenkamp M, Oberhettinger P, Leo JC, Linke D, Schütz MS (2015) *Yersinia* adhesin A (YadA) – beauty & beast. Int J Med Microbiol 305:252–258. https://doi.org/10.1016/j.ijmm.2014.12.008

99. Raghunathan D, Wells TJ, Morris FC, Shaw RK, Bobat S, Peters SE, Paterson GK, Jensen KT, Leyton DL, Blair JMA, Browning DF, Pravin J, Flores-Langarica A, Hitchcock JR, Moraes CTP, Piazza RMF, Maskell DJ, Webber MA, May RC, MacLennan CA, Piddock LJ, Cunningham AF, Henderson IR (2011) SadA, a trimeric autotransporter from *Salmonella enterica* serovar Typhimurium, can promote biofilm formation and provides limited protection against infection. Infect Immun 79:4342 LP – 4352. https://doi.org/10.1128/IAI.05592-11

100. Comanducci M, Bambini S, Brunelli B, Adu-Bobie J, Aricò B, Capecchi B, Giuliani MM, Masignani V, Santini L, Savino S, Granoff DM, Caugant DA, Pizza M, Rappuoli R, Mora M (2002) NadA, a novel vaccine candidate of *Neisseria meningitidis*. J Exp Med 195:1445–1454. https://doi.org/10.1084/jem.20020407

101.   Thibau A, Dichter AA, Vaca DJ, Linke D, Goldman A, Kempf VAJ (2019) Immunogenicity of trimeric autotransporter adhesins and their potential as vaccine targets. Med Microbiol Immunol 209:243–263. https://doi.org/10.1007/s00430-019-00649-y

102.   Szczesny P, Lupas A (2008) Domain annotation of trimeric autotransporter adhesins - daTAA. Bioinformatics 24:1251–1256. https://doi.org/10.1093/bioinformatics/btn118

103.   Leo JC, Grin I, Linke D (2012) Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. Philos Trans R Soc B Biol Sci 367:1088–1101. https://doi.org/10.1098/RSTB.2011.0208

104.   Bassler J, Hernandez Alvarez B, Hartmann MD, Lupas AN (2015) A domain dictionary of trimeric autotransporter adhesins. Int J Med Microbiol 305:265–275. https://doi.org/10.1016/j.ijmm.2014.12.010

105.   Szczesny P, Linke D, Ursinus A, Bär K, Schwarz H, Riess TM, Kempf VAJ, Lupas AN, Martin J, Zeth K (2008) Structure of the head of the *Bartonella* adhesin BadA. PLOS Pathog 4:e1000119. https://doi.org/10.1371/journal.ppat.1000119

106.   Kaiser PO, Linke D, Schwarz H, Leo JC, Kempf VAJ (2012) Analysis of the BadA stalk from *Bartonella henselae* reveals domain-specific and domain-overlapping functions in the host cell infection process. Cell Microbiol 14:198–209. https://doi.org/10.1111/j.1462-5822.2011.01711.x

107.   Krieger F, Möglich A, Kiefhaber T (2005) Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. J Am Chem Soc 127:3346–3352. https://doi.org/10.1021/ja042798i

108.   Lupas AN, Gruber M (2005) The structure of α-helical coiled coils. Adv Protein Chem. 70:37–38. https://doi.org/10.1016/S0065-3233(05)70003-6

109.   Hartmann MD, Grin I, Dunin-Horkawicz S, Deiss S, Linke D, Lupas AN, Hernandez Alvarez B (2012) Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria. Proc Natl Acad Sci USA 109:20907–20912. https://doi.org/10.1073/pnas.1211872110

110.   Batterman HJ, Peek JA, Loutit JS, Falkow S, Tompkins LS (1995) *Bartonella henselae* and *Bartonella quintana* adherence to and entry into cultured human epithelial cells. Infect Immun 63:4553–4556. https://doi.org/10.1128/IAI.63.11.4553-4556.1995

111.   Riess T, Raddatz G, Linke D, Schäfer A, Kempf VAJ (2007) Analysis of *Bartonella* adhesin A expression reveals differences between various *B. henselae* strains. Infect Immun 75:35–43. https://doi.org/10.1128/IAI.00963-06

112.   Tu N, Lima A, Bandeali Z, Anderson B (2016) Characterization of the general stress response in *Bartonella henselae*. Microb Pathog 92:1–10. https://doi.org/10.1016/j.micpath.2015.12.010

113.   Tu N, Carroll RK, Weiss A, Shaw LN, Nicolas G, Thomas S, Lima A, Okaro U, Anderson B (2017) A family of genus-specific RNAs in tandem with DNA-binding proteins control expression of the *badA* major virulence factor gene in *Bartonella henselae*. Microbiologyopen 6:e00420. https://doi.org/10.1002/mbo3.420

114. Okaro U, George S, Valdes S, Macaluso K, Anderson B (2020) A non-coding RNA controls transcription of a gene encoding a DNA binding protein that modulates biofilm development in *Bartonella henselae*. Microb Pathog 147:104272. https://doi.org/https://doi.org/10.1016/j.micpath.2020.104272

115. Okaro U, Green R, Mohapatra S, Anderson B (2019) The trimeric autotransporter adhesin BadA is required for in vitro biofilm formation by *Bartonella henselae*. npj Biofilms Microbiomes 5:10. https://doi.org/10.1038/s41522-019-0083-8

116. Riess T, Dietrich F, Schmidt K V., Kaiser PO, Schwarz H, Schäfer A, Kempf VAJ (2008) Analysis of a novel insect cell culture medium-based growth medium for *Bartonella* species. Appl Environ Microbiol 74:5224–5227. https://doi.org/10.1128/AEM.00621-08

117. Wagner CL, Riess T, Linke D, Eberhardt C, Schäfer A, Reutter S, Maggi RG, Kempf VAJ (2008) Use of *Bartonella* adhesin A (BadA) immunoblotting in the serodiagnosis of *Bartonella henselae* infections. Int J Med Microbiol 298:579–590. https://doi.org/https://doi.org/10.1016/j.ijmm.2008.01.013

118. Müller NF, Kaiser PO, Linke D, Schwarz H, Riess T, Schäfer A, Eble JA, Kempf VAJ (2011) Trimeric autotransporter adhesin-dependent adherence of *Bartonella henselae*, *Bartonella quintana*, and *Yersinia enterocolitica* to matrix components and endothelial cells under static and dynamic flow conditions. Infect Immun 79:2544 LP – 2553. https://doi.org/10.1128/IAI.01309-10

119. Cho J, Mosher DF (2006) Role of fibronectin assembly in platelet thrombus formation. J Thromb Haemost 4:1461–1469. https://doi.org/10.1111/j.1538-7836.2006.01943.x

120. Mao Y, Schwarzbauer JE (2005) Fibronectin fibrillogenesis, a cell-mediated matrix assembly process. Matrix Biol 24:389–399. https://doi.org/10.1016/J.MATBIO.2005.06.008

121. Arvand M, Wendt C, Regnath T, Ullrich R, Hahn H (1998) Characterization of *Bartonella henselae* isolated from bacillary angiomatosis lesions in a human immunodeficiency virus–infected patient in Germany. Clin Infect Dis 26:1296–1299. https://doi.org/10.1086/516348

122. Zbinden R, Höchli M, Nadal D (1995) Intracellular location of *Bartonella henselae* cocultivated with Vero cells and used for an indirect fluorescent-antibody test. Clin Diagn Lab Immunol 2:693 LP – 695. https://doi.org/10.1128/cdli.2.6.693-695.1995

123. Zbinden R, Michael N, Sekulovski M, Von Graevenitz A, Nadal D (1997) Evaluation of commercial slides for detection of immunoglobulin G against *Bartonella henselae* by indirect immunofluorescence. Eur J Clin Microbiol Infect Dis 16:648–652. https://doi.org/10.1007/BF01708554

124. Welch DF, Pickett DA, Slater LN, Steigerwalt AG, Brenner DJ (1992) *Rochalimaea henselae* sp. nov., a cause of septicemia, bacillary angiomatosis, and parenchymal bacillary peliosis. J Clin Microbiol 30:275–280. https://doi.org/10.1128/JCM.30.2.275-280.1992

125. Sander A, Bühler C, Pelz K, von Cramm E, Bredt W (1997) Detection and identification of two *Bartonella henselae* variants in domestic cats in Germany. J Clin Microbiol 35:584–587. https://doi.org/10.1128/JCM.35.3.584-587.1997

126. Riess T, Anderson B, Fackelmayer A, Autenrieth IB, Kempf VAJ (2003) Rapid and efficient transposon mutagenesis of *Bartonella henselae* by transposome technology. Gene 313:103–109. https://doi.org/10.1016/S0378-1119(03)00636-X

127. Thibau A, Hipp K, Vaca DJ, Chowdhury S, Malmström J, Saragliadis A, Ballhorn W, Linke D, Kempf VAJ (2022) Long-read sequencing reveals genetic adaptation of *Bartonella* adhesin A among different *Bartonella henselae* isolates. Front Microbiol 13:1–17. https://doi.org/10.3389/fmicb.2022.838267

128. Thibau A, Vaca DJ, Bagowski M, Hipp K, Bender D, Ballhorn W, Linke D, Kempf VAJ (2022) Adhesion of *Bartonella henselae* to fibronectin is mediated via repetitive motifs present in the stalk of *Bartonella* adhesin A. Microbiol Spectr 10:. https://doi.org/10.1128/spectrum.02117-22

129. Stahl J, Bergmann H, Göttig S, Ebersberger I, Averhoff B (2015) *Acinetobacter baumannii* virulence is mediated by the concerted action of three phospholipases D. PLoS One 10:1–19. https://doi.org/10.1371/journal.pone.0138360

130. Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM, Peterson KM (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene 166:175–176 https://doi.org/10.1016/0378-1119(95)00584-1

131. Kempf VAJ, Schaller M, Behrendt S, Volkmann B, Aepfelbacher M, Cakman I, Autenrieth IB (2000) Interaction of *Bartonella henselae* with endothelial cells results in rapid bacterial rRNA synthesis and replication. Cell Microbiol 2:431–441. https://doi.org/10.1046/j.1462-5822.2000.00072.x

132. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J (2016) NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569

133. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75

134. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F (2015) RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365

135. Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191

136. Darling AE, Mau B, Perna NT (2010) Progressivemauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:. https://doi.org/10.1371/journal.pone.0011147

137. Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704

138. Kim M, Oh H-S, Park S-C, Chun J (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 64:346–351. https://doi.org/10.1099/ijs.0.059774-0

139. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J (2016) JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32:929–931. https://doi.org/10.1093/bioinformatics/btv681

140. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS (2011) PHAST: a fast phage search tool. Nucleic Acids Res 39:W347–W352. https://doi.org/10.1093/nar/gkr485

141. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387

142. Söding J (2005) Protein homology detection by HMM-HMM comparison. Bioinformatics 21:951–960. https://doi.org/10.1093/bioinformatics/bti125

143. Hildebrand A, Remmert M, Biegert A, Söding J (2009) Fast and accurate automatic structure prediction with HHpred. Proteins Struct Funct Bioinforma 77:128–132. https://doi.org/10.1002/prot.22499

144. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V (2018) A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol 430:2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007

145. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J (2019) HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics 20:473. https://doi.org/10.1186/s12859-019-3019-7

146. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V (2020) Protein sequence analysis using the MPI bioinformatics toolkit. Curr Protoc Bioinforma 72:1–30. https://doi.org/10.1002/cpbi.108

147. Klausen MS, Jespersen MC, Nielsen H, Jensen KK, Jurtz VI, Sønderby CK, Sommer MOA, Winther O, Nielsen M, Petersen B, Marcatili P (2019) NetSurfP-2.0: improved prediction of protein structural features by integrated deep learning. Proteins Struct Funct Bioinforma 87:520–527. https://doi.org/10.1002/prot.25674

148. Delorenzi M, Speed T (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 18:617–625. https://doi.org/10.1093/bioinformatics/18.4.617

149. Frickey T, Lupas A (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics 20:3702–3704. https://doi.org/10.1093/bioinformatics/bth444

150. Weidensdorfer M, Chae JI, Makobe C, Stahl J, Averhoff B, Müller V, Schürmann C, Brandes RP, Wilharm G, Ballhorn W, Christ S, Linke D, Fischer D, Göttig S, Kempf VAJ (2016) Analysis of endothelial adherence of *Bartonella henselae* and *Acinetobacter baumannii* using a dynamic human *ex vivo* infection model. Infect Immun 84:711–722. https://doi.org/10.1128/IAI.01502-15

151. Thibau A, Schultze TG, Ballhorn W, Kempf VAJ (2020) complete genome sequence of *Bartonella alsatica* strain IBS 382 (CIP 105477). Microbiol Resour Announc 9:1–2. https://doi.org/10.1128/MRA.00769-20

152.   Hung L-H (2000) The solution structure of the C-terminal domain of the Mu B transposition
       protein. EMBO J 19:5625–5634. https://doi.org/10.1093/emboj/19.21.5625

153.   Shen Y, Gomez-Blanco J, Petassi MT, Peters JE, Ortega J, Guarné A (2022) Structural basis
       for DNA targeting by the Tn7 transposon. Nat Struct Mol Biol 29:143–151.
       https://doi.org/10.1038/s41594-022-00724-8

154.   Matias PM, Gorynia S, Donner P, Carrondo MA (2006) Crystal structure of the human AAA+
       protein RuvBL1. J Biol Chem 281:38918–38929. https://doi.org/10.1074/JBC.M605625200

155.   Silva STN, Brito JA, Arranz R, Sorzano CÓS, Ebel C, Doutch J, Tully MD, Carazo J-M,
       Carrascosa JL, Matias PM, Bandeiras TM (2018) X-ray structure of full-length human RuvB-
       Like 2 – mechanistic insights into coupling between ATP binding and mechanical action.
       Sci Rep 8:13726. https://doi.org/10.1038/s41598-018-31997-z

156.   Page AN, George NP, Marceau AH, Cox MM, Keck JL (2011) Structure and biochemical
       activities of *Escherichia coli* MgsA. J Biol Chem 286:12075–12085.
       https://doi.org/10.1074/JBC.M110.210187

157.   Ledesma-Garcia L, Thuillier J, Guzman-Espinola A, Ensinck I, de la Sierra-Gallay IL, Lazar N,
       Aumont-Nicaise M, Mignolet J, Soumillion P, Nessler S, Hols P (2020) Molecular dissection of
       pheromone selectivity in the competence signaling system ComRS of streptococci. Proc Natl
       Acad Sci USA 117:7745–7754. https://doi.org/10.1073/PNAS.1916085117

158.   Gkekas S, Singh RK, Shkumatov A V., Messens J, Fauvart M, Verstraeten N, Michiels J,
       Versées W (2017) Structural and biochemical analysis of *Escherichia coli* ObgE, a central
       regulator of bacterial persistence. J Biol Chem 292:5871–5883.
       https://doi.org/10.1074/JBC.M116.761809

159.   Hillen HS, Lavdovskaia E, Nadler F, Hanitsch E, Linden A, Bohnsack KE, Urlaub H, Richter-
       Dennerlein R (2021) Structural basis of GTPase-mediated mitochondrial ribosome biogenesis
       and recycling. Nat Commun 2021 121 12:1–10. https://doi.org/10.1038/s41467-021-23702-y

160.   Khusainov I, Fatkhullin B, Pellegrino S, Bikmullin A, Liu W ti, Gabdulkhakov A, Shebel A Al,
       Golubev A, Zeyer D, Trachtmann N, Sprenger GA, Validov S, Usachev K, Yusupova G,
       Yusupov M (2020) Mechanism of ribosome shutdown by RsfS in *Staphylococcus aureus*
       revealed by integrative structural biology approach. Nat Commun 2020 111 11:1–10.
       https://doi.org/10.1038/s41467-020-15517-0

161.   Cook J, Baverstock TC, McAndrew MBL, Stansfeld PJ, Roper DI, Crow A (2020) Insights into
       bacterial cell division from a structure of EnvC bound to the FtsX periplasmic domain. Proc Natl
       Acad Sci USA 117:28355–28365. https://doi.org/10.1073/PNAS.2017134117

162.   Coleman SA, Minnick MF (2003) Differential expression of the invasion-associated locus B
       (*ialB*) gene of *Bartonella bacilliformis* in response to environmental cues. Microb Pathog
       34:179–186. https://doi.org/10.1016/S0882-4010(03)00005-6

163.   Mitchell SJ, Minnick MF (1995) Characterization of a two-gene locus from *Bartonella
       bacilliformis* associated with the ability to invade human erythrocytes. Infect Immun 63:1552–
       1562. https://doi.org/10.1128/iai.63.4.1552-1562.1995

164.   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 69:4373–4381. https://doi.org/10.1128/IAI.69.7.4373-4381.2001

165.   Grosskinsky U, Schütz M, Fritz M, Schmid Y, Lamparter MC, Szczesny P, Lupas AN, Autenrieth IB, Linke D (2007) a conserved glycine residue of trimeric autotransporter domains plays a key role in *Yersinia* adhesin A autotransport. J Bacteriol 189:9011–9019. https://doi.org/10.1128/JB.00985-07

166.   Wollmann P, Zeth K, Lupas AN, Linke D (2006) Purification of the YadA membrane anchor for secondary structure analysis and crystallization. Int J Biol Macromol 39:3–9. https://doi.org/10.1016/j.ijbiomac.2005.11.009

167.   El Tahir Y, Skurnik M (2001) YadA, the multifaceted *Yersinia* adhesin. Int J Med Microbiol 291:209–218. https://doi.org/10.1078/1438-4221-00119

168.   Weidensdorfer M, Ishikawa M, Hori K, Linke D, Djahanschiri B, Iruegas R, Ebersberger I, Riedel-Christ S, Enders G, Leukert L, Kraiczy P, Rothweiler F, Cinatl J, Berger J, Hipp K, Kempf VAJ, Göttig S (2019) The *Acinetobacter* trimeric autotransporter adhesin Ata controls key virulence traits of *Acinetobacter baumannii*. Virulence 10:68–81. https://doi.org/10.1080/21505594.2018.1558693

169.   Okaro U, George S, Anderson B (2021) What is in a cat scratch? growth of *Bartonella* henselae in a biofilm. Microorganisms 9:835. https://doi.org/10.3390/microorganisms9040835

170.   Tørresen OK, Star B, Mier P, Andrade-Navarro MA, Bateman A, Jarnot P, Gruca A, Grynberg M, Kajava A V., Promponas VJ, Anisimova M, Jakobsen KS, Linke D (2019) Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. Nucleic Acids Res 47:10994–11006. https://doi.org/10.1093/nar/gkz841

171.   Koren S, Harhay GP, Smith TPL, Bono JL, Harhay DM, Mcvey SD, Radune D, Bergman NH, Phillippy AM (2013) Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome Biol 14:. https://doi.org/10.1186/gb-2013-14-9-r101

172.   Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106:19126–19131. https://doi.org/10.1073/PNAS.0906412106

173.   Harvey KL, Jarocki VM, Charles IG, Djordjevic SP (2019) The diverse functional roles of elongation factor Tu (EF-Tu) in microbial pathogenesis. Front Microbiol 10:2351. https://doi.org/10.3389/fmicb.2019.02351

174.   Lathe WC, Bork P (2001) Evolution of *tuf* genes: ancient duplication, differential loss and gene conversion. FEBS Lett 502:113–116. https://doi.org/10.1016/S0014-5793(01)02639-4

175.   Berghoff J, Viezens J, Guptill L, Fabbi M, Arvand M (2007) *Bartonella henselae* exists as a mosaic of different genetic variants in the infected host. Microbiology 153:2045–2051. https://doi.org/10.1099/mic.0.2007/006379-0

176.   De S, Babu MM (2010) Genomic neighbourhood and the regulation of gene expression. Curr Opin Cell Biol 22:326–333. https://doi.org/10.1016/j.ceb.2010.04.004

177. Battisti JM, Smitherman LS, Sappington KN, Parrow NL, Raghavan R, Minnick MF (2007) Transcriptional regulation of the heme binding protein gene family of *Bartonella* quintana is accomplished by a novel promoter element and iron response regulator. Infect Immun 75:4373–4385. https://doi.org/10.1128/IAI.00497-07

178. Bier N, Hammerstrom TG, Koehler TM (2020) Influence of the phosphoenolpyruvate:carbohydrate phosphotransferase system on toxin gene expression and virulence in *Bacillus anthracis*. Mol Microbiol 113:237–252. https://doi.org/10.1111/mmi.14413

179. Rahbar MR, Zarei M, Jahangiri A, Khalili S, Nezafat N, Negahdaripour M, Fattahian Y, Savardashtaki A, Ghasemi Y (2020) Non-adaptive evolution of trimeric autotransporters in *Brucellaceae*. Front Microbiol 11:2664. https://doi.org/10.3389/fmicb.2020.560667

180. Ishikawa M, Yoshimoto S, Hayashi A, Kanie J, Hori K (2016) Discovery of a novel periplasmic protein that forms a complex with a trimeric autotransporter adhesin and peptidoglycan. Mol Microbiol 101:394–410. https://doi.org/10.1111/mmi.13398

181. Grin I, Hartmann MD, Sauer G, Hernandez Alvarez B, Schütz M, Wagner S, Madlung J, Macek B, Felipe-Lopez A, Hensel M, Lupas A, Linke D (2014) A trimeric lipoprotein assists in trimeric autotransporter biogenesis in enterobacteria. J Biol Chem 289:7388–7398. https://doi.org/10.1074/jbc.M113.513275

182. Adler NRL, Stevens JM, Stevens MP, Galyov EE (2011) Autotransporters and their role in the virulence of *Burkholderia pseudomallei* and *Burkholderia mallei*. Front Microbiol 2:151. https://doi.org/10.3389/fmicb.2011.00151

183. Anderson B, Goldsmith C, Johnson A, Padmalayam I, Baumstark B (1994) Bacteriophage-like particle of *Rochalimaea henselae*. Mol Microbiol 13:67–73. https://doi.org/10.1111/j.1365-2958.1994.tb00402.x

184. Bentley SD, Vernikos GS, Snyder LAS, Churcher C, Arrowsmith C, Chillingworth T, Cronin A, Davis PH, Holroyd NE, Jagels K, Maddison M, Moule S, Rabbinowitsch E, Sharp S, Unwin L, Whitehead S, Quail MA, Achtman M, Barrell B, Saunders NJ, Parkhill J (2007) Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. PLoS Genet 3:0230–0240. https://doi.org/10.1371/journal.pgen.0030023

185. Kempf VAJ, Volkmann B, Schaller M, Sander CA, Alitalo K, Rieß T, Autenrieth IB (2001) Evidence of a leading role for VEGF in *Bartonella henselae* -induced endothelial cell proliferations. Cell Microbiol 3:623–632. https://doi.org/10.1046/j.1462-5822.2001.00144.x

186. Gemmell NJ (2021) Repetitive DNA: genomic dark matter matters. Nat Rev Genet 2021 226 22:342–342. https://doi.org/10.1038/s41576-021-00354-8

187. Zhou K, Aertsen A, Michiels CW (2014) The role of variable DNA tandem repeats in bacterial adaptation. FEMS Microbiol Rev 38:119–141. https://doi.org/10.1111/1574-6976.12036

188. Vaca DJ, Thibau A, Schütz M, Kraiczy P, Happonen L, Malmström J, Kempf VAJ (2019) Interaction with the host: the role of fibronectin and extracellular matrix proteins in the adhesion of Gram-negative bacteria. Med Microbiol Immunol 209:277–299. https://doi.org/10.1007/s00430-019-00644-3

189. Łyskowski A, Leo JC, Goldman A (2011) Structure and biology of trimeric autotransporter adhesins. Adv Exp Med Biol 715:143–158. https://doi.org/10.1007/978-94-007-0940-9_9

190. Singh B, Fleury C, Jalalvand F, Riesbeck K (2012) Human pathogens utilize host extracellular matrix proteins laminin and collagen for adhesion and invasion of the host. FEMS Microbiol Rev 36:1122–1180. https://doi.org/10.1111/j.1574-6976.2012.00340.x

191. Pankov R, Yamada KM (2002) Fibronectin at a glance. J Cell Sci 115:3861–3863. https://doi.org/10.1242/jcs.00059

192. Schwarz-Linek U, Werner JM, Pickford AR, Gurusiddappa S, Ewa JHK, Pilka S, Briggs JAG, Gough TS, Höök M, Campbell ID (2003) Pathogenic bacteria attach to human fibronectin through a tandem β-zipper. Nature 423:177–181. https://doi.org/10.1038/nature01589

List of figures

List of tables

Abbreviations

AAA+ ATPase	ATPase associated with diverse cellular activities
aa	amino acid(s)
Amp	ampicillin
ANI	average nucleotide identity
ATCC	American type culture collection
ATP	adenosine triphosphate
BadA	*Bartonella* adhesin A
BALI	*Bartonella* liquid
bp	base pair(s)
BSA	bovine serum albumin
ca.	circa
CBA	Columbia blood agar
CCS	circular consensus sequences
CDS	coding sequences
CLANS	cluster analysis of sequences
CLSM	confocal laser scanning microscopy
ComR	competence regulator
CSD	cat scratch disease
D	domain
daTAA	domain annotation of TAAs
ΔBadA-D	*badA*-deficient deletion mutant
ΔBadA-T	*badA*-deficient transposon mutant
dH$_2$O	distilled water
DNA	deoxyribonucleic acid
ECM	extracellular matrix
ELISA	enzyme-linked immunosorbent assay
FCS	foetal calf serum
FS	freeze substitution
Fur	ferric uptake regulator
Fw	forward
g	grams
gDNA	genomic DNA

Gen	gentamycin
GTP	guanosine triphosphate
h	hours
H	head
HbpA	hemin binding protein A
HIM	head insert motif
HMW	high molecular weight
HPF	high-pressure freezing
HRP	horseradish peroxidase
HUVEC	human umbilical vein endothelial cell
IalB	invasion associated locus B
IFM	immunofluorescence microscopy
Kan	kanamycin
kb	kilobase pairs
kDa	kilo-Dalton
LB	Luria/Miller
LPS	lipopolysaccharides
µl	microliter
µm	micrometer
min	minutes
mA	milliampere
ml	millilitre
MW	molecular mass
ms	milliseconds
MS	mass spectrometry
N	neck
nm	nanometer
nt	nucleotide(s)
OD	optical density
OD_{600}	optical density measured at a wavelength of 600 nm
ORF	open reading frame
PacBio	Pacific Biosciences
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline

PCR	polymerase chain reaction
PGAP	prokaryotic genome annotation pipeline
pGxx	periodic glycine-x-x
PHASTER	phage search tool enhanced release
PLT	progressive lowering of temperature
Q	Phred quality
qPCR	quantitative real-time PCR
RASTtk	rapid annotation using subsystem technology tool kit
RNA	ribonucleic acid
rRNA	ribosomal RNA
rpm	rotations per minute
RT	room temperature
Rv	reverse
S	stalk
SDS	sodium dodecyl sulphate
sec	seconds
SMRT	single-molecule real-time
Spec	spectinomycin
suppl.	supplementary
TAA	trimeric autotransporter adhesin
TEM	transmission electron microscopy
tRNA	transfer RNA
v/v	volume/volume
w/v	weight/volume
WB	Western blotting
XL-MS	cross-linking mass spectrometry

Data availability

The genome sequences of all sequenced *B. henselae* strains, together with their corresponding Sequence Read Archive data, have been deposited in the NCBI GenBank database under BioProject PRJNA720375 with the following genome accession numbers: CP072904 (Marseille), CP072903 (ATCC49882^T var-1), CP072902 (ATCC49882^T var-2), CP072901 (Berlin-I), CP072900 (G-5436), CP072899 (88-64 Oklahoma), CP072898 (FR96/BK38), and CP072897 (FR96/BK3).

The *badA* gene sequence from *B. henselae* Marseille was deposited separately under the GenBank accession number MK993576. The genome sequence of *B. alsatica* strain IBS 382 (CIP 105477), together with the corresponding Sequence Read Archive data, was deposited in the NCBI GenBank database under BioProject PRJNA641327 with the GenBank accession number CP058235.

Supplementary figures

```
MKKLSVTSKRQYNLYASPISRRLSLLMKLSLETVTVMFLLGASPVLASNLALTGAKNLSQNSPGVNYSKGSHGSIVLSGDDDFCGADYVLG
RGGNSTVRNGIPISVEEEYERFVKQKLMNNATSPYSQSSEQQVWTGDGLTSKGSGYMGGKSTDGDKNILPEAYGIYSFATGCGSSAQGNYS
VAFGANATALTGGSQAFGVAALASGRVSVAIGVGSEATGEAGVSLGGLSKAAGARSVAIGTRAKAQGEESIAIGSSVKNGDKDGSAVAQGA
KAIAIGSNSISFQHYAVAVGAKAHALLSKTVALGYDSVADVDAGIRGYDPVEDEPSKDVSFVWKSSLGAVSVGNRKEGLTRQIIGVAAGTE
DTDAVNVAQLKALRGMISEKGGWNLTVNNDNNTVVSSGGALDLSSGSKNLKIVKDGKKNNVTFDVARDLTLKSIKLDGVTLNETGLFIANG
PQITASGINAGSQKITGVAEGTDANDAVNFGQLKKIETEVKEQVAASGFVKQDSDTKYLTIGKDTDGDTINIANNKSDKRTLTGIKEGDIS
KDSSEAITGSQLFTTNQNVKTVSDNLQTAATNIAKTFGGGAKYEDGEWIAPAFKVKTVTGEGKEEEKRYQNVADALAGVGSSITNVQNKVT
EQVNNAITKVEGDALLWSDEANAFVARHEKSKLGKGASKATQENSKITYLLDGDVSKDSTDAITGKQLYSLGDKIASYLGGNAKYEDGEWT
APTFKVKTVKEDGKEEEKTYQNVAEALTGVGTSFTNVKNEITKQINHLQSDDSAVVHYDKNKDETGGINYASVTLGKGKDSAAVTLHNVAD
GSISKDSRDAINGSQIYSLNEQLATYFGGGAKYENGQWTAPIFKVKTVKEDGEEEEKTYQNVAEALTGVGTSFTNIKSEITKQIANEISSV
TGDSLVKKDLATNLITIGKEVAGTEINIASVSKADRTLSGVKEAVKDNEAVNKGQLDKGLKHLSDSLQSDDSAVVHYDKKTDETGGINYTS
VTLGGKDKTPVALHNVADGSISKDSHDAINGGQIHTIGEDVAKFLGGAASFNNGAFTGPTYKLSNIDAKGDVQQSEFKDIGSAFAGLDTNI
KNVNNNVTNKFNELTQNITNVTQQVKGDALLWSDEANAFVARHEKSKLGKGASKATQENSKITYLLDGDVSKDSTDAITGKQLYSLGDKIA
SYLGGNAKYENGEWTAPTFKVKTVKEDGKEEEKTYQNVAEALTGVGASFTNVKNEITKQINHLQSDDSAVVHYDKNKDETGGINYASVTLG
KGKDSAAVTLHNVADGSISKDSRDAINGSQIYSLNEQLATYFGGGAKYENGQWTAPIFKVKTVKEDGEEEEKTYQNVAEALTGVGTSFTNI
KSEITKQIANEISSVTGDSLVKKDLATNLITIGKEVAGTEINIASVSKADRTLSGVKEAVKDNEAVNKGQLDKGLKHLSDSLQSDDSAVVH
YDKKTDETGGINYTSVTLGGKDKTPVALHNVADGSISKDSHDAINGGQIHTIGEDVAKFLGGAASFNNGAFTGPTYKLSNIDAKGDVQQSE
FKDIGSAFAGLDTNIKNVNNNVTNKFNELTQNITNVTQQVKGDALLWSDEANAFVARHEKSKLGKGASKATQENSKITYLLDGDVSKDSTD
AITGKQLYSLGDKIASYLGGNAKYENGEWTAPTFKVKTVKEDGKEEEKTYQNVAEALTGVGASFTNVKNEITKQINHLQSDDSAVVHYDKN
KDETGGINYASVTLGKGKDSAAVTLHNVADGSISKDSRDAINGSQIYSLNEQLATYFGGGAKYENGQWTAPIFKVKTVKEDGEEEEKTYQN
VAEALTGVGTSFTNIKSEITKQIANEISSVTGDSLVKKDLATNLITIGKEVAGTEINIASVSKADRTLSGVKEAVKDNEAVNKGQLDTNIK
KVEDKLTEAVGKVTQQVKGDALLWSNEDNAFVADHGKDSAKTKSKITHLLDGNIASGSTDAVTGGQLYSLNEQLATYFGGGAKYENGQWTA
PTFKVKTVNGEGKEEEQTYQNVAEALTGVGASFMNVQNKITNEITNQVNNAITKVEGDSLVKQDNLGIITLGKERGGLKVDFANRDGLDRT
LSGVKEAVNDNEAVNKGQLDADISKVNNNVTNKFNELTQNITNVTQQVKGDALLWSDEANAFVARHEKSKLEKGVSKATQENSKITYLLDG
DISKGSTDAVTGGQLYSLNEQLATYFGGGAKYENGQWTAPTFKVKTVNGEGKEEEQTYQNVAAAFEGVGTSFTNIKSEITKQINNEIINVK
GDSLVKRDLATNLITIGKEIEGSVINIANKSGEARTISGVKEAVKDNEAVNKGQLDTNIKKVEDKLTEAVGKVTQQVKGDALLWSNEDNAF
VADHGKDSAKTKSKITHLLDGNIASGSTDAVTGGQLYSLNEQLATYFGGGAKYENGQWTAPTFKVKTVNGEGKEEEKTYQNVAAAFEGVGT
SFTNIKSEITKQIANEISNVTGDSLVKKDLDTNLITIGKEIAGTEINIASVSKADRTLSGVKEAVNDNEAVNKGQLDANISKVNNNVTNKF
NELTQSITNVTQQVKGDALLWSDEANAFVARHEKSKLEKGVSKATQENSKITYLLDGDISKGSTDAVTGGQLYSLNEQLATYFGGGAKYEN
GQWTAPTFKVKTVNGEGKEEEQTYQNVAAAFEGVGTSFTNIKSEITKQINNEIINVKGDSLVKRDLATNLITIGKEIEGSVINIANKSGEA
RTISGVKEAVKDNEAVNKGQLDTNIKKVEDKLTEAVGKVTQQVKGDALLWSNEDNAFVADHGKDSAKTKSKITHLLDGNIASGSTDAVTGG
QLYSLNEQLATYFGGGAKYENGQWTAPTFKVKTVNGEGKEEEKTYQNVAAAFEGVGTSFTNIKSEITKQIANEISNVTGDSLVKKDLDTNL
ITIGKEIAGTEINIASVSKADRTLSGVKEAVNDNEAVNKGQLDANISKVNNNVTNKFNELTQSITNVTQQVKGDALLWSDEANAFVARHEK
SKLEKGVSKATQENSKITYLLDGDISKGSTDAVTGGQLYSLNEQLATYFGGGAKYENGQWTAPTFKVKTVNGEGKEEEQTYQNVAAAFEGV
GTSFTNIKSEITKQINNEIINVKGDSLVKRDLATNLITIGKEIEGSVINIANKSGEARTISGVKEAVKDNEAVNKGQLDTNIKKVEDKLTE
AVGKVTQQVKGDALLWSNEDNAFVADHGKDSAKTKSKITHLLDGNIASGSTDAVTGGQLYSLNEQLATYFGGGAKYENGQWTAPTFKVKTV
NGEGKEEEKTYQNVAAAFEGVGTSFTHVKNEITKQINHLQSDDSAVVHYDKDDKNGSINYASVTLGKGKDSAAVALHNVADGSISKDSHDA
INGGQIHTIGEDVAKFLGGDAAFKDGAFTGPTYKLSNIDAKGDVQQSEFKDIGSAFAGLDTNIKNVNNNVTNKLSELTQNITTVTQQVKGN
ALLWSDEANAFVARHEKSKLEKGASKAIQENSKITYLLDGDVSKGSTDAVTGGQLYSMSNMLATYLGGNAKYENGEWTAPTFKVKTVNGEG
KEEEQTYQNVAEALTGVGTSFTNIKSEIAKQINHLQSDDSAVIHYDKNKDETGTINYASVTLGKGEDSAAVALHNVAAGNIAKDSRDAING
SQLYSLNEQLLTYFGGDAGYKDGQWIAPKFHVLQFKSDGSSGEKESYDNVAAAFEGVNKSLAGMNERINNVTAGQNVSSSSLNWNETEGGY
DARHNGVDSKLTHVENGDVSEKSKEAVNGSQLWNTNEKVEAVEKDVKNIEKKVQDIATVADSAVKYEKDSTGKKTNVIKLVGGSESEPVLI
DNVADGKIEADSKQAVNGGQLRDYTEKQMKIVLDDAKKYTDERFNDVVNNGINEAKAYTDVKFEALSYTVEEVRKEARQAAAIGLAVSNLR
YYDIPGSLSLSFGTGIWRSQSAFAIGAGYTSEDGNIRSNLSITSSGGQWGVGAGITLRLK
```

Suppl. Figure 1. Mass spectrometry analysis of isolated BadA fibres from *B. henselae* Marseille. For immunisation of rabbits, high MW BadA fibres from *B. henselae* Marseille were isolated and identified as BadA proteins using MS analysis. Grey highlighted sequences are MS-identified peptides, with a total of 154 peptide hits of which 84 are unique peptides (MS and preparative steps were performed by Sounak Chowdhury, PhD, from Lund University, Sweden).

badA HN2S27 gene sequence

CTGAATTTAGAGAGTGTAAGCTTTTATAGAAGCGTGCTGTTCTCTTTGAAAAGGAATGGTATTGTTCACAAAAAGTACTGTTTTTATTATG
AACTAAAAAAATTTATTTTTAGCTTGCTATTTTACTCAATAGAGGATAGTGATACAGAAGGTATATCAGTATACTCATTTTAATTATAACT
TCAAAAGGGGAGGAAGTAATGCGTAAAAGACGAAACGCCACTCTAAAAGCAAATTTACATACCGCATCAATATAAAGAAACACTC
GTAACAGAAATCAACTAAGCATACAGATTTCTTTTAAATATTCTTCAAATTCTCTTATTATTAAGAAAAGATGCTCCTTAATGAAAAAATT
TTTTAATAAAACAGATAGCAATAAAAGAATGATTGAAATATTATTTAAACAACACCACCCTAACGTAAAACGTCTTAATATTTAAAACAGA
AAAATTCTTTTTTAAGTACACAACAAAAACAACCGCTCAACCCCTATTACAATCCAAATGCGCTATTTACACGCTTCCTACCAAGCTTTCG
CATTCAGATTTCATTACAGAAAGTACACACAAAATAAAAATAAAGACTCAAAACGTTCCCAATTTGACCACCCTCCTTATTTTAATCCTCA
TTACAAGGGAGTAGGTAATACTAAAATGTGTCTTTTTTATGTTTTGGATGTGCTTTGTAATTTTTTTCATTGGAGAATTTATT**ATG**AAAAA
ATTATCTGTCACATCAAAGAGACAATATAATTTATATGCTTCGCCTATTTCTCGACGTTTATCTTTGTTAATGAAGCTCTCATTGGAAACT
GTAACAGTTATGTTCTTATTGGGTGCATCTCCTGTATTGGCTTCGAATCTTGCGCTTACAGGAGCAAAGAATCTGAGTCAAAACTCTCCAG
GTGTAAATTACTCTAAAGGTAGCCATGGTAGTATTGTTCTCTCTGGTGATGATGATTTTTGCGGTGCGGATTATGTTCTTGGTCGTGGAGG
CAATTCTACTGTACGTAATGGGATTCCAATAAGTGTAGAAGAAGAATATGAGAGATTTGTCAAACAAAAATTAATGAATAATGCTACTTCT
CCTTATAGTCAGAGTTCAGAGCAACAAGTTTGGACTGGTGATGGGCTAACAAGCAAAGGTTCGGGTTATATGGGAGGGAAGTCGACTGACG
GTGATAAAAATATCTTGCCTGAGGCTTATGGTATATATTCTTTTGCAACTGGTTGTGGTTCTTCTGCGCAGGGGAATTATTCAGTTGCATT
TGGTGCAAATGCAACTGCACTTACTGGGGGGTCGCAAGCTTTTGGTGTTGCTGCACTTGCAAGTGGAAGGGTAAGTGTTGCTATTGGTGTA
GGGTCAGAAGCGACGGGAGAGGCTGGAGTTTCTTTGGGTGGACTCTCAAAGGCAGCTGGTGCTCGTAGTGTTGCTATAGGGACGCGGGCTA
AAGCTCAGGGTGAAGAATCTATTGCGATAGGGTAGTAGCGTAAAGAATGGTGATAAGGACGGTTCAGCTGTAGCGCAGGGTGCAAAAGCGAT
TGCTATAGGTTCTAATTCTATTAGTTTTCAGCACTATGCAGTTGCGGTTGGTGCTAAAGCCCATGCTCTTCTCTCGAAAACTGTTGCCTTG
GGTTATGATTCTGTTGCTGATGTTGATGCTGGCATTAGAGGTTATGATCCTGTGGAGGATGAGCCATCGAAAGACGTTAGTTTTGTATGGA
AAAGCTCTCTAGGTGCTGTTAGTGTTGGTAATCGTAAAGAAGGCTTAACGCGACAAATTATAGGAGTTGCAGCTGGTACTGAAGACACTGA
TGCAGTAAATGTTGCACAGCTAAAAGCATTAAGGGGAATGATATCAGAAAAAGGAGGTTGGAATCTTACTGTTAATAATGACAATAATACA
GTTGTTAGCTCAGGTGGTGCATTAGATTTGTCATCTGGAAGTAAAAATCTCAAAATTGTAAAGATGGAAAAAAGAATAATGTAACCTTTG
ATGTCGCTAGGGATCTCACGTTAAAGAGCATAAAATTAGACGGTGTTACTTTAAATGAAACAGGTTTATTTATTGCAAACGGACCACAAAT
CACCGCTTCAGGTATTAATGCTGGTAGTCAAAAAATTACAGGCGTAGCAGAGGGTACTGATGCGAACGATGCAGTAAACTTTGGACAACT<mark>G</mark>
<mark>C</mark>ATACAATCGGTGAGGATGTTGCAAAATTCTTGGGTGGAGATGCAGCTTTTAAAGATGGCGCTTTTACCGGCCCAACTTATAAGTTGTCGA
ATATTGATGCAAAGGGTGATGTACAACAGATGAGTTTAAAGATATAGGTTCAGCCTTTGCGGGTCTTGATACGAACATCAAGAATGTCAA
TAATAATGTAACGAATAAGCTCAGTGAACTTACTCAAAACATAACGACTGTTACGCAACAGGTAAAAGGCAATGCCTTATTATGGAGCGAT
GAAGCTAATGCCTTTGTGGCGCGTCATGAAAGAGCAAGTTAGAAAAAGGTGCATCTAAAGCGATACAAGAAAACAGCAAGATTACGTATC
TGTTAGATGGTGATGTTTCGAAAGGTTCCACGGATGCCGTTACTGGTGGTCAGCTTTATTCAATGAGCAATATGCTTGCGACCTATTTGGG
TGGTAACGCTAAATATGAGAATGGTGAATGGACCGCACCTACCTTTAAGGTTAAAACAGTTAACGGTGAAGGCAAGGAAGAAGAGCAAACT
TATCAGAATGTAGCGGAAGCTTTGACTGGAGTTGGTACGTCTTTCACCAATATAAAAAGTGAGATTGCCAAACAGATTAATCATCTCCAGT
CTGATGATTCAGCGGTTATTCATTATGATAAGAATAAGATGAAACTGGCACCATTAATTATGCGAGTGTAACTTTGGGTAAAGGTGAAGA
TTCTGCAGCTGTTGCCCTTCATAATGTCGCTGCAGGTAATATTGCTAAGGATTTCACGTGATGCAATCAATGGTTCTTCAGCTTTATTCTTTG
AACGAGCAGTTATTGACCTATTTTGGCGGTGATGCTGGCTATAAAGATGGGCAATGGATAGCTCCCAAGTTCCATGTTTTGCAGTTCAAGA
GTGATGGTAGTTCTGGTGAGAAGGAGAGCTATGATAATGTAGCGGCTGCGTTTGAAGGAGTTAACAAAAGTCTTGCAGGTATGAACGAGCG
TATTAATAATGTTACTGCTGGCCAGAATGTTTCGTCGAGCAGTTTAAATTGGAATGAGACGGAGGGAGGTTATGACGCTCGTCATAATGGT
GTGGACAGTAAGCTTACGCATGTAGAGAATGGTGACGTATCCGAAAAATCGAAAGAAGCCGTTAATGGAAGTCAACTATGGAATACGAATG
AGAAAGTTGAAGCGGTTGAGAAGGATGTAAAGAATATTGAGAAGAAGGTACAAGATATTGCTACAGTAGCAGATAGTGCTGTTAAGTATGA
GAAAGATAGTACTGGCAAGAAAACGAATGTAATCAAATTAGTTGGTGGGAGTGAAAGTGAGCCAGTATTGATAGACAATGTAGCGGATGGT
AAAATTGAAGCAGACTCTAAGCAGGCAGTCAATGGAGGTCAGTTGCGTGATTATACTGAGAAACAGATGAAGATAGTGCTTGATGATGCGA
AGAAATATACGGATGAACGCTTCAATGATGTCGTCAATAATGGTATTAATGAGGCTAAAGCTTATACAGATGTGAAGTTTGAGGCTTTAAG
TTACACTGTTGAGGAAGTCCGGAAAGAAGCAAGACAAGCAGCGGCTATTGGTTTAGCAGTATCTAACTTACGTTACTATGATATACCAGGA
TCTTTAAGTCTTTCATTTGGTACGGGTATATGGCGTAGTCAGTCTGCATTTGCTATTGGTGCTGGTTATACATCTGAAGATGGCAATATTC
GTTCTAATTTATCTATCACGAGTTCTGGTGGTCAGTGGGGAGTAGGCGCAGGGATTACTTTGAGACTGAAA**TGA**TAAAAAAACTAATATTA
TGATAGAAAAACGAAGTATTTTGATAAATATTCTGTTCTTCCTTGCCTTATTAGGCAAGGGAGAAAGTTTTGCTGATGAAACGATAGTGT
TTATACGGTGCATCCACCGCATTTATCTATTCCTAATGGGGTAGCGGGTGAAACACGTCGAATCATCATGCAGTTTTATTATTGGACTTTA
ATTTGTGATGAAAACAAAGCTTAGGCAAGGCATATGTAATGTGACGCAAACTGTCCATGATAAGGAAGGCAATACTATTTTCAGTTGGT
CTCTTGTTTCTACGAAAAA

BadA HN2S27 protein sequence

MKKLSVTSKRQYNLYASPISRRLSLLMKLSLETVTVMFLLGASPVLASNLALTGAKNLSQNSPGVNYSKGSHGSIVLSGDDDFCGADYVLG
RGGNSTVRNGIPISVEEEYERFVKQKLMNNATSPYSQSSEQQVWTGDGLTSKGSGYMGGKSTDGDKNILPEAYGIYSFATGCGSSAQGNYS
VAFGANATALTGGSQAFGVAALASGRVSVAIGVGSEATGEAVSLGGLSKAAGARSVAIGTRAKAQGEESIAIGSSVKNGDKDGSAVAQGA
KAIAIGSNSISFQHYAVAVGAKAHALLSKTVALGYDSVADVDAGIRGYDPVEDEPSKDVSFVWKSSLGAVSVGNRKEGLTRQIIGVAAGTE
DTDAVNVAQLKALRGMISEKGGWNLTVNNDNNTVVSSGGALDLSSGSKNLKIVKDGKKNNVTFDVARDLTLKSIKLDGVTLNETGLFIANG
PQITASGINAGSQKITGVAEGTDANDAVNFGQ<mark>LH</mark>TIGEDVAKFLGGDAAFKDGAFTGPTYKLSNIDAKGDVQQSEFKDIGSAFAGLDTNIK
NVNNNVTNKLSELTQNITTVTQQVKGNALLWSDEANAFVARHEKSKLEKGASKAIQENSKITYLLDGDVSKGSTDAVTGGQLYSMSNMLAT
YLGGNAKYENGEWTAPTFKVKTVNGEGKEEEQTYQNVAEALTGVGTSFTNIKSEIAKQINHLQSDDSAVIHYDKNKDETGTINYASVTLGK
GEDSAAVALHNVAAGNIAKDSRDAINGSQLYSLNEQLLTYFGGDAGYKDGQWIAPKFHVLQFKSDGSSGEKESYDNVAAAFEGVNKSLAGM
NERINNVTAGQNVSSSSLNWNETEGGYDARHNGVDSKLTHVENGDVSEKSKEAVNGSQLWNTNEKVEAVEKDVKNIEKKVQDIATVADSAV
KYEKDSTGKKTNVIKLVGGSESEPVLIDNVADGKIEADSKQAVNGGQLRDYTEKQMKIVLDDAKKYTDERFNDVVNNGINEAKAYTDVKFE
ALSYTVEEVRKEARQAAAIGLAVSNLRYYDIPGSLSLSFGTGIWRSQSAFAIGAGYTSEDGNIRSNLSITSSGGQWGVGAGITLRLK–

Suppl. Figure 2. Synthesised and truncated *badA* gene and BadA protein HN2S27 sequences.
The grey highlighted region translates to the truncated BadA HN2S27 protein sequence shown directly
underneath. The underlined sequence represents the cleaved-off signal sequence during transport into
the periplasm. Yellow highlighted nt and aa depict the transition site of combined *badA* regions.

badA S27 gene sequence

```
CTGAATTTAGAGAGTGTAAGCTTTTATAGAAGCGTGCTGTTCTCTTTGAAAAGGAATGGTATTGTTCACAAAAAGTACTGTTTTTATTATG
AACTAAAAAAATTTATTTTTAGCTTGCTATTTTACTCAATAGAGGATAGTGATACAGAAGGTATATCAGTATACTCATTTTAATTATAACT
TCAAAAGGGGAGGAAGTAATGCGTAAAAGACGAAACGCCACTCTAAAAGCAAATTTACATACCGCATCACACTCAATATAAAGAAACACTC
GTAACAGAAATCAACTAAGCATACAGATTTCTTTTAAATATTCTTCAAATTCTCTTATTATTAAGAAAAGATGCTCCTTAATGAAAAAATT
TTTTAATAAAACAGATAGCAATAAAAGAATGATTGAAATATTATTTAAACAACACCACCCTAACGTAAAACGTCTTAATATTTAAAACAGA
AAAATTCTTTTTTAAGTACACAACAAAAACAACCGCTCAACCCCTATTACAATCCAAATGCGCTATTTACACGCTTCCTACCAAGCTTTCG
CATTCAGATTTCATTACAGAAAGTACACACAAAATAAAAATAAAGACTCAAAACGTTCCCAATTTGACCACCCTCCTTATTTTAATCCTCA
TTACAAGGGAGTAGGTAATACTAAAATGTGTCTTTTTTATGTTTTGGATGTGCTTTGTAATTTTTTTCATTGGAGAATTTATT**ATG**AAAAA
ATTATCTGTCACATCAAAGAGACAATATAATTTATATGCTTCGCCTATTTCTCGACGTTTATCTTTGTTAATGAAGCTCTCATTGGAAACT
GTAACAGTTATGTTCTTATTGGGTGCATCTCCTGTATTGGCTTCGAATCTTGC**GG**ACAGATTCATCAATCGGTGAGGATGTTGCAAAAT
TCTTGGGTGGAGATGCAGCTTTTAAAGATGGCGCTTTTACCGGCCCAACTTATAAGTTGTCGAATATTGATGCAAAGGGTGATGTACAACA
GAGTGAGTTTAAAGATATAGGTTCAGCCTTTGCGGGTCTTGATACGAACATCAAGAATGTCAATAATAATGTAACGAATAAGCTCAGTGAA
CTTACTCAAAACATAACGACTGTTACGCAACAGGTAAAAGGCAATGCCTTATTATGGAGCGATGAAGCTAATGCCTTTGTGGCGCGTCATG
AAAAGAGCAAGTTAGAAAAAGGTGCATCTAAAGCGATACAAGAAAACAGCAAGATTACGTATCTGTTAGATGGTGATGTTCGAAAGGTTC
CACGGATGCCGTTACTGGTGGTCAGCTTTATTCAATGAGCAATATGCTTGCGACCTATTTGGGTGGTAACGCTAAATATGAGAATGGTGAA
TGGACCGCACCTACCTTTAAGGTTAAAACAGTTAACGGTGAAGGCAAGGAAGAAGAGCAAACTTATCAGAATGTAGCGGAAGCTTTGACTG
GAGTTGGTACGTCTTTCACCAATATAAAAAGTGAGATTGCCAAACAGATTAATCATCTCCAGTCTGATGATTCAGCGGTTATTCATTATGA
TAAGAATAAAGATGAAACTGGCACCATTAATTATGCGAGTGTAACTTTGGGTAAAGGTGAAGATTCTGCAGCTGTTGCCCTTCATAATGTC
GCTGCAGGTAATATTGCTAAGGATTCACGTGATGCAATCAATGGTTCTCAGCTTTATTCTTTGAACGAGCAGTTATTGACCTATTTTGGCG
GTGATGCTGGCTATAAAGATGGGCAATGGATAGCTCCCAAGTTCCATGTTTTGCAGTTCAAGAGTGATGGTAGTTCTGGTGAGAAGGAGAG
CTATGATAATGTAGCGGCTGCGTTTGAAGGAGTTAACAAAAGTCTTGCAGGTATGAACGAGCGTATTAATAATGTTACTGCTGGCCAGAAT
GTTTCGTCGAGCAGTTTAAATTGGAATGAGACGGAGGGAGGTTATGACGCTCGTCATAATGGTGTGGACAGTAAGCTTACGCATGTAGAGA
ATGGTGACGTATCCGAAAAATCGAAAGAAGCCGTTAATGGAAGTCAACTATGGAATACGAATGAGAAAGTTGAAGCGGTTGAGAAGGATGT
AAAGAATATTGAGAAGAAGGTACAAGATATTGCTACAGTAGCAGATAGTGCTGTTAAGTATGAGAAGATAGTACTGGCAAGAAAACGAAT
GTAATCAAATTAGTTGGTGGGAGTGAAAGTGAGCCAGTATTGATAGACAATGTAGCGGATGGTAAAATTGAAGCAGACTCTAAGCAGGCAG
TCAATGGAGGTCAGTTGCGTGATTATACTGTGAAAACGATAGTGAAGATAGTGCTTGATGATGCGAAGAAAATATACGGATGAACGCTTCAATGA
TGTCGTCAATAATGGTATTAATGAGGCTAAAGCTTATACAGATGTGAAGTTTGAGGCTTTAAGTTACACTGTTGAGGAAGTCCGGAAAGAA
GCAAGACAAGCAGCGGCTATTGGTTTAGCAGTATCTAACTTACGTTACTATGATATACCAGGATCTTTAAGTCTTTCATTGGTACGGGTA
TATGGCGTAGTCAGTCTGCATTTGCTATTGGTGCTGGTTATACATCTGAAGATGGCAATATTCGTTCTAATTTATCTATCACGAGTTCTGG
TGGTCAGTGGGGAGTAGGCGCAGGGATTACTTTGAGACTGAAA**TGA**TAAAAAAACTAATATTATGATAGAAAAACGAAGTATTTTGATAAA
TATTCTGTTCTTCCTTGCCTTATTAGGCAAGGGAGAAAGTTTTGCTGATGAAACGATAGTGTTTATACGGTGCATCCACCGCATTTATCT
ATTCCTAATGGGGTAGCGGGTGAAACACGTCGAATCATCATGCAGTTTTATTATTGGACTTTAATTTGTGATGAAAACAAAAGCTTAGGC
AAGGCATATGTAATGTGACGCAAACTGTCCATGATAAGGAAGGCAATACTATTTTCAGTTGGTCTCTTGTTTCTACGAAAAA
```

BadA S27 protein sequence

```
MKKLSVTSKRQYNLYASPISRRLSLLMKLSLETVTVMFLLGASPVLASNLAGQIHTIGEDVAKFLGGDAAFKDGAFTGPTYKLSNIDAKGD
VQQSEFKDIGSAFAGLDTNIKNVNNNVTNKLSELTQNITTVTQQVKGNALLWSDEANAFVARHEKSKLEKGASKAIQENSKITYLLDGDVS
KGSTDAVTGGQLYSMSNMLATYLGGNAKYENGEWTAPTFKVKTVNGEGKEEEQTYQNVAEALTGVGTSFTNIKSEIAKQINHLQSDDSAVI
HYDKNKDETGTINYASVTLGKGEDSAAVALHNVAAGNIAKDSRDAINGSQLYSLNEQLLTYFGGDAGYKDGQWIAPKFHVLQFKSDGSSGE
KESYDNVAAAFEGVNKSLAGMNERINNVTAGQNVSSSSLNWNETEGGYDARHNGVDSKLTHVENGDVSEKSKEAVNGSQLWNTNEKVEAVE
KDVKNIEKKVQDIATVADSAVKYEKDSTGKKTNVIKLVGGSESEPVLIDNVADGKIEADSKQAVNGGQLRDYTEKQMKIVLDDAKKYTDER
FNDVVNNGINEAKAYTDVKFEALSYTVEEVRKEARQAAAIGLAVSNLRYYDIPGSLSLSFGTGIWRSQSAFAIGAGYTSEDGNIRSNLSIT
SSGGQWGVGAGITLRLK-
```

Suppl. Figure 3. Synthesised and truncated badA gene and BadA protein S27 sequences. The grey highlighted region translates to the truncated BadA S27 protein sequence shown directly underneath. The underlined sequence represents the cleaved-off signal sequence during transport into the periplasm. Yellow highlighted nt and aa depict the transition site of combined badA regions.

badA S28 gene sequence

```
CTGAATTTAGAGAGTGTAAGCTTTTATAGAAGCGTGCTGTTCTCTTTGAAAAGGAATGGTATTGTTCACAAAAAGTACTGTTTTTATTATG
AACTAAAAAAATTTATTTTTAGCTTGCTATTTTACTCAATAGAGGATAGTGATACAGAAGGTATATCAGTATACTCATTTTAATTATAACT
TCAAAAGGGGAGGAAGTAATGCGTAAAAGACGAAACGCCACTCTAAAAGCAAATTTACATACCGCATCACACTCAATATAAAGAAACACTC
GTAACAGAAATCAACTAAGCATACAGATTTCTTTTAAATATTCTTCAAATTCTCTTATTATTAAGAAAAGATGCTCCTTAATGAAAAAATT
TTTTAATAAAACAGATAGCAATAAAAGAATGATTGAAATATTATTTAAACAACACCACCCTAACGTAAAACGTCTTAATATTTAAAACAGA
AAAATTCTTTTTTAAGTACACAACAAAAACAACCGCTCAACCCCTATTACAATCCAAATGCGCTATTTACACGCTTCCTACCAAGCTTTCG
CATTCAGATTTCATTACAGAAAGTACACACAAAATAAAAATAAAGACTCAAAACGTTCCCAATTTGACCACCCTCCTTATTTTAATCCTCA
TTACAAGGGAGTAGGTAATACTAAAATGTGTCTTTTTTATGTTTTGGATGTGCTTTGTAATTTTTTTCATTGGAGAATTTATT**ATG**AAAAA
ATTATCTGTCACATCAAAGAGACAATATAATTTATATGCTTCGCCTATTTCTCGACGTTTATCTTTGTTAATGAAGCTCTCATTGGAAACT
GTAACAGTTATGTTCTTATTGGGTGCATCTCCTGTATTGGCTTCGAATCTTGC<mark>GG</mark>GTCAGCTTTATTCAATGAGCAATATGCTTGCGACCT
ATTTGGGTGGTAACGCTAAATATGAGAATGGTGAATGGACCGCACCTACCTTTAAGGTTAAAACAGTTAACGGTGAAGGCAAGGAAGAAGA
GCAAACTTATCAGAATGTAGCGGAAGCTTTGACTGGAGTTGGTACGTCTTTCACCAATATAAAAAGTGAGATTGCCAAACAGATTAATCAT
CTCCAGTCTGATGATTCAGCGGTTATTCATTATGATAAGAATAAAGATGAAACTGGCACCATTAATTATGCGAGTGTAACTTTGGGTAAAG
GTGAAGATTCTGCAGCTGTTGCCCTTCATAATGTCGCTGCAGGTAATATTGCTAAGGATTCACGTGATGCAATCAATGGTTCTCAGCTTTA
TTCTTTGAACGAGCAGTTATTGACCTATTTTGGCGGTGATGCTGGCTATAAAGATGGGCAATGGATAGCTCCCAAGTTCCATGTTTTGCAG
TTCAAGAGTGATGGTAGTTCTGGTGAGAAGGAGAGCTATGATAATGTAGCGGCTGCGTTTGAAGGAGTTAACAAAAGTCTTGCAGGTATGA
ACGAGCGTATTAATAATGTTACTGCTGGCCAGAATGTTTCGTCGAGCAGTTTAAATTGGAATGAGACGGAGGGAGGTTATGACGCTCGTCA
TAATGGTGTGGACAGTAAGCTTACGCATGTAGAGAATGGTGACGTATCCGAAAAATCGAAAGAAGCCGTTAATGGAAGTCAACTATGGAAT
ACGAATGAGAAAGTTGAAGCGGTTGAGAAGGATGTAAAGAATATTGAGAAGAAGGTACAAGATATTGCTACAGTAGCAGATAGTGCTGTTA
AGTATGAGAAAGATAGTACTGGCAAGAAAACGAATGTAATCAAATTAGTTGGTGGGAGTGAAAGTGAGCCAGTATTGATAGACAATGTAGC
GGATGGTAAAATTGAAGCAGACTCTAAGCAGGCAGTCAATGGAGGTCAGTTGCGTGATTATACTGAGAAACAGATGAAGATAGTGCTTGAT
GATGCGAAGAAATATACGGATGACAACGCTTCAATGATGTCGTCAATAATGGTATTAATGAGGCTAAAGCTTATACAGATGTGAAGTTTGAGG
CTTTAAGTTACACTGTTGAGGAAGTCCGGAAAGAAGCAAGACAAGCAGCGGCTATTGGTTTAGCAGTATCTAACTTACGTTACTATGATAT
ACCAGGATCTTTAAGTCTTTCATTTGGTACGGGTATATGGCGTAGTCAGTCTGCATTTGCTATTGGTGCTGGTTATACATCTGAAGATGGC
AATATTCGTTCTAATTTATCTATCACGAGTTCTGGTGGTCAGTGGGGAGTAGGCGCAGGGATTACTTTGAGACTGAAA**TGA**TAAAAAAACT
AATATTATGATAGAAAAACGAAGTATTTTGATAAATATTCTGTTCTTCCTTGCCTTATTAGGCAAGGGAGAAAGTTTTGCTGATGAAAACG
ATAGTGTTTATACGGTGCATCCACCGCATTTATCTATTCCTAATGGGGTAGCGGGTGAAACACGTCGAATCATCATGCAGTTTTATTATTG
GACTTTAATTTGTGATGAAAAACAAAAGCTTAGGCAAGGCATATGTAATGTGACGCAAACTGTCCATGATAAGGAAGGCAATACTATTTTC
AGTTGGTCTCTTGTTTCTACGAAAAA
```

BadA S28 protein sequence

```
MKKLSVTSKRQYNLYASPISRRLSLLMKLSLETVTVMFLLGASPVLASNL<mark>AG</mark>QLYSMSNMLATYLGGNAKYENGEWTAPTFKVKTVNGEGK
EEEQTYQNVAEALTGVGTSFTNIKSEIAKQINHLQSDDSAVIHYDKNKDETGTINYASVTLGKGEDSAAVALHNVAAGNIAKDSRDAINGS
QLYSLNEQLLTYFGGDAGYKDGQWIAPKFHVLQFKSDGSSGEKESYDNVAAAFEGVNKSLAGMNERINNVTAGQNVSSSSLNWNETEGGYD
ARHNGVDSKLTHVENGDVSEKSKEAVNGSQLWNTNEKVEAVEKDVKNIEKKVQDIATVADSAVKYEKDSTGKKTNVIKLVGGSESEPVLID
NVADGKIEADSKQAVNGGQLRDYTEKQMKIVLDDAKKYTDERFNDVVNNGINEAKAYTDVKFEALSYTVEEVRKEARQAAAIGLAVSNLRY
YDIPGSLSLSFGTGIWRSQSAFAIGAGYTSEDGNIRSNLSITSSGGQWGVGAGITLRLK−
```

Suppl. Figure 4. Synthesised and truncated *badA* gene and BadA protein S28 sequences. The grey highlighted region translates to the truncated BadA S28 protein sequence shown directly underneath. The underlined sequence represents the cleaved-off signal sequence during transport into the periplasm. Yellow highlighted nt and aa depict the transition site of combined *badA* regions.

badA S29 gene sequence

```
CTGAATTTAGAGAGTGTAAGCTTTTATAGAAGCGTGCTGTTCTCTTTGAAAAGGAATGGTATTGTTCACAAAAAGTACTGTTTTTATTATG
AACTAAAAAAATTTATTTTTAGCTTGCTATTTTACTCAATAGAGGATAGTGATACAGAAGGTATATCAGTATACTCATTTTAATTATAACT
TCAAAAGGGGAGGAAGTAATGCGTAAAAGACGAAACGCCACTCTAAAAGCAAATTTACATACCGCATCACACTCAATATAAAGAAACACTC
GTAACAGAAATCAACTAAGCATACAGATTTCTTTTAAATATTCTTCAAATTCTCTTATTATTAAGAAAAGATGCTCCTTAATGAAAAAATT
TTTTAATAAAACAGATAGCAATAAAAGAATGATTGAAATATTATTTAAACAACACCACCCTAACGTAAAACGTCTTAATATTTAAAACAGA
AAAATTCTTTTTTAAGTACACAACAAAAACAACCGCTCAACCCCTATTACAATCCAAATGCGCTATTTACACGCTTCCTACCAAGCTTTCG
CATTCAGATTTCATTACAGAAAGTACACACAAAATAAAAATAAAGACTCAAAACGTTCCCAATTTGACCACCCTCCTTATTTTAATCCTCA
TTACAAGGGAGTAGGTAATACTAAAATGTGTCTTTTTTATGTTTTGGATGTGCTTTGTAATTTTTTTCATTGGAGAATTTATT**ATG**AAAAA
ATTATCTGTCACATCAAAGAGACAATATAATTTATATGCTTCGCCTATTTCTCGACGTTTATCTTTGTTAATGAAGCTCTCATTGGAAACT
GTAACAGTTATGTTCTTATTGGGTGCATCTCCTGTATTGGCTTCGAATCTTGC GT CTCAGCTTTATTCTTTGAACGAGCAGTTATTGACCT
ATTTTGGCGGTGATGCTGGCTATAAAGATGGGCAATGGATAGCTCCCAAGTTCCATGTTTTGCAGTTCAAGAGTGATGGTAGTTCTGGTGA
GAAGGAGAGCTATGATAATGTAGCGGCTGCGTTTGAAGGAGTTAACAAAAGTCTTGCAGGTATGAACGAGCGTATTAATAATGTTACTGCT
GGCCAGAATGTTTCGTCGAGCAGTTTAAATTGGAATGAGACGGAGGGAGGTTATGACGCTCGTCATAATGGTGTGGACAGTAAGCTTACGC
ATGTAGAGAATGGTGACGTATCCGAAAAATCGAAAGAAGCCGTTAATGGAAGTCAACTATGGAATACGAATGAGAAGTTGAAGCGGTTGA
GAAGGATGTAAAGAATATTGAGAAGAAGGTACAAGATATTGCTACAGTAGCAGATAGTGCTGTTAAGTATGAGAAAGATAGTACTGGCAAG
AAAACGAATGTAATCAAATTAGTTGGTGGGAGTGAAAGTGAGCCAGTATTGATAGACAATGTAGCGGATGGTAAAATTGAAGCAGACTCTA
AGCAGGCAGTCAATGGAGGTCAGTTGCGTGATTATACTGAGAAACAGATGAAGATAGTGCTTGATGATGCGAAGAAAATATACGGATGAACG
CTTCAATGATGTCGTCAATAATGGTATTAATGAGGCTAAAGCTTATACAGATGTGAAGTTTGAGGCTTTAAGTTACACTGTTGAGGAAGTC
CGGAAAGAAGCAAGACAAGCAGCGGCTATTGGTTTAGCAGTATCTAACTTACGTTACTATGATATACCAGGATCTTTAAGTCTTTCATTTG
GTACGGGTATATGGCGTAGTCAGTCTGCATTTGCTATTGGTGCTGGTTATACATCTGAAGATGGCAATATTCGTTCTAATTTATCTATCAC
GAGTTCTGGTGGTCAGTGGGGAGTAGGCGCAGGGATTACTTTGAGACTGAAA**TGA**TAAAAAACTAATATTATGATAGAAAAACGAAGTAT
TTTGATAAATATTCTGTTCTTCCTTGCCTTATTAGGCAAGGGAGAAAGTTTTGCTGATGAAAACGATAGTGTTTATACGGTGCATCCACCG
CATTTATCTATTCCTAATGGGGTAGCGGGTGAAACACGTCGAATCATCATGCAGTTTTATTATTGGACTTTAATTTGTGATGAAAAACAAA
AGCTTAGGCAAGGCATATGTAATGTGACGCAAACTGTCCATGATAAGGAAGGCAATACTATTTTCAGTTGGTCTCTTGTTTCTACGAAAAA
```

BadA S29 protein sequence

```
MKKLSVTSKRQYNLYASPISRRLSLLMKLSLETVTVMFLLGASPVLASNL AS QLYSLNEQLLTYFGGDAGYKDGQWIAPKFHVLQFKSDGS
SGEKESYDNVAAAFEGVNKSLAGMNERINNVTAGQNVSSSSLNWNETEGGYDARHNGVDSKLTHVENGDVSEKSKEAVNGSQLWNTNEKVE
AVEKDVKNIEKKVQDIATVADSAVKYEKDSTGKKTNVIKLVGGSESEPVLIDNVADGKIEADSKQAVNGGQLRDYTEKQMKIVLDDAKKYT
DERFNDVVNNGINEAKAYTDVKFEALSYTVEEVRKEARQAAAIGLAVSNLRYYDIPGSLSLSFGTGIWRSQSAFAIGAGYTSEDGNIRSNL
SITSSGGQWGVGAGITLRLK-
```

Suppl. Figure 5. Synthesised and truncated *badA* gene and BadA protein S29 sequences. The grey highlighted region translates to the truncated BadA S29 protein sequence shown directly underneath. The underlined sequence represents the cleaved-off signal sequence during transport into the periplasm. Yellow highlighted nt and aa depict the transition site of combined *badA* regions.

badA S30 gene sequence

```
CTGAATTTAGAGAGTGTAAGCTTTTATAGAAGCGTGCTGTTCTCTTTGAAAAGGAATGGTATTGTTCACAAAAAGTACTGTTTTTATTATG
AACTAAAAAAATTTATTTTTAGCTTGCTATTTTACTCAATAGAGGATAGTGATACAGAAGGTATATCAGTATACTCATTTTAATTATAACT
TCAAAAGGGGAGGAAGTAATGCGTAAAAGACGAAACGCCACTCTAAAAGCAAATTTACATACCGCATCACACTCAATATAAAGAAACACTC
GTAACAGAAATCAACTAAGCATACAGATTTCTTTTAAATATTCTTCAAATTCTCTTATTATTAAGAAAAGATGCTCCTTAATGAAAAAATT
TTTTAATAAAACAGATAGCAATAAAAGAATGATTGAAATATTATTTAAACAACACCACCCTAACGTAAAACGTCTTAATATTTAAAACAGA
AAAATTCTTTTTTAAGTACACAACAAAAACAACCGCTCAACCCCTATTACAATCCAAATGCGCTATTTACACGCTTCCTACCAAGCTTTCG
CATTCAGATTTCATTACAGAAAGTACACACAAAATAAAAATAAAGACTCAAAACGTTCCCAATTTGACCACCCTCCTTATTTTAATCCTCA
TTACAAGGGAGTAGGTAATACTAAAATGTGTCTTTTTTATGTTTTGGATGTGCTTTGTAATTTTTTTCATTGGAGAATTTATT**ATG**AAAAA
ATTATCTGTCACATCAAAGAGACAATATAATTTATATGCTTCGCCTATTTCTCGACGTTTATCTTTGTTAATGAAGCTCTCATTGGAAACT
GTAACAGTTATGTTCTTATTGGGTGCATCTCCTGTATTGGCTTCGAATCTTGC**GC**TTACGCATGTAGAGAATGGTGACGTATCCGAAAAAT
CGAAAGAAGCCGTTAATGGAAGTCAACTATGGAATACGAATGAGAAAGTTGAAGCGGTTGAGAAGGATGTAAAGAATATTGAGAAGAAGGT
ACAAGATATTGCTACAGTAGCAGATAGTGCTGTTAAGTATGAGAAAGATAGTACTGGCAAGAAAACGAATGTAATCAAATTAGTTGGTGGG
AGTGAAAGTGAGCCAGTATTGATAGACAATGTAGCGGATGGTAAAATTGAAGCAGACTCTAAGCAGGCAGTCAATGGAGGTCAGTTGCGTG
ATTATACTGAGAAACAGATGAAGATAGTGCTTGATGATGCGAAGAAATATACGGATGAACGCTTCAATGATGTCGTCAATAATGGTATTAA
TGAGGCTAAAGCTTATACAGATGTGAAGTTTGAGGCTTTAAGTTACACTGTTGAGGAAGTCCGGAAAGAAGCAAGACAAGCAGCGGCTATT
GGTTTAGCAGTATCTAACTTACGTTACTATGATATACCAGGATCTTTAAGTCTTTCATTTGGTACGGGTATATGGCGTAGTCAGTCTGCAT
TTGCTATTGGTGCTGGTTATACATCTGAAGATGGCAATATTCGTTCTAATTTATCTATCACGAGTTCTGGTGGTCAGTGGGGAGTAGGCGC
AGGGATTACTTTGAGACTGAAA**TGA**TAAAAAAACTAATATTATGATAGAAAACGAAGTATTTTGATAAATATTCTGTTCTTCCTTGCCTTT
ATTAGGCAAGGGAGAAAGTTTTGCTGATGAAAACGATAGTGTTTATACGGTGCATCCACCGCATTTATCTATTCCTAATGGGGTAGCGGGT
GAAACACGTCGAATCATCATGCAGTTTTATTATTGGACTTTAATTTGTGATGAAAAACAAAAGCTTAGGCAAGGCATATGTAATGTGACGC
AAACTGTCCATGATAAGGAAGGCAATACTATTTTCAGTTGGTCTCTTGTTTCTACGAAAAA
```

BadA S30 protein sequence

```
MKKLSVTSKRQYNLYASPISRRLSLLMKLSLETVTVMFLLGASPVLASNLALTHVENGDVSEKSKEAVNGSQLWNTNEKVEAVEKDVKNIE
KKVQDIATVADSAVKYEKDSTGKKTNVIKLVGGSESEPVLIDNVADGKIEADSKQAVNGGQLRDYTEKQMKIVLDDAKKYTDERFNDVVNN
GINEAKAYTDVKFEALSYTVEEVRKEARQAAAIGLAVSNLRYYDIPGSLSLSFGTGIWRSQSAFAIGAGYTSEDGNIRSNLSITSSGGQWG
VGAGITLRLK-
```

Suppl. Figure 6. Synthesised and truncated badA gene and BadA protein S30 sequences. The grey highlighted region translates to the truncated BadA S30 protein sequence shown directly underneath. The underlined sequence represents the cleaved-off signal sequence during transport into the periplasm. Yellow highlighted nt and aa depict the transition site of combined badA regions.

badA D16S28 gene sequence

```
GAATTCCTGAATTTAGAGAGTGTAAGCTTTTATAGAAGCGTGCTGTTCTCTTTGAAAAGGAATGGTATTGTTCACAAAAAGTACTGTTTTT
ATTATGAACTAAAAAAATTTATTTTTAGCTTGCTATTTTACTCAATAGAGGATAGTGATACAGAAGGTATATCAGTATACTCATTTTAATT
ATAACTTCAAAAGGGGAGGAAGTAATGCGTAAAAGACGAAACGCCACTCTAAAAGCAAATTTACATACCGCATCACACTCAATATAAAGAA
ACACTCGTAACAGAAATCAACTAAGCATACAGATTTCTTTTAAATATTCTTCAAATTCTCTTATTATTAAGAAAAGATGCTCCTTAATGAA
AAAATTTTTTAATAAAACAGATAGCAATAAAAGAATGATTGAAATATTATTTAAACAACACCACCCTAACGTAAAACGTCTTAATATTTAA
AACAGAAAAATTCTTTTTTAAGTACACAACAAAAACAACCGCTCAACCCCTATTACAATCCAAATGCGCTATTTACACGCTTCCTACCAAG
CTTTCGCATTCAGATTTCATTACAGAAAGTACACACAAAATAAAAATAAAGACTCAAAACGTTCCCAATTTGACCACCCTCCTTATTTTAA
TCCTCATTACAAGGGAGTAGGTAATACTAAAATGTGTCTTTTTTATGTTTTGGATGTGCTTTGTAATTTTTTTCATTGGAGAATTTATT**AT**
**G**AAAAAATTATCTGTCACATCAAAGAGACAATATAATTTATATGCTTCGCCTATTTCTCGACGTTTATCTTTGTTAATGAAGCTCTCATTG
GAAACTGTAACAGTTATGTTCTTATTGGGTGCATCTCCTGTATTGGCTTCGAATCTTGC<span style="background:yellow">GG</span>GTCAGCTTTATTCTCTGAACGAGCAACTTG
CGACCTATTTTGGCGGCGGTGCTAAGTACGAGAATGGCCAATGGACTGCACCTACCTTTAAGGTTAAAACAGTTAACGGTGAAGGCAAGGA
AGAAGAGCAGACTTATCAGAATGTAGCAGCAGCTTTTGAAGGAGTTGGTACGTCTTTCACCAATATAAAAAGTGAGATTACTAAACAGATT
AATAATGAGATTATCAATGTAAAAGGTGATAGTCTTGTTAAGAGAGATCTCGCTACGAATCTCATCACCATTGGTAAAGAAATAGAAGGCA
GTGTAATCAATATTGCTAATAAGAGTGGTGAAGCTCGGACCATTTCTGGTGTTAAGGAAGCAGTAAAAGATAATGAAGCTGTTAACAAAGG
GCAGCT<span style="background:yellow">TT</span>ATTCAATGAGCAATATGCTTGCGACCTATTTGGGTGGTAACGCTAAATATGAGAATGGTGAATGGACCGCACCTACCTTTAAG
GTTAAAACAGTTAACGGTGAAGGCAAGGAAGAAGAGCAAACTTATCAGAATGTAGCGGAAGCTTTGACTGGAGTTGGTACGTCTTTCACCA
ATATAAAAAGTGAGATTGCCAAACAGATTAATCATCTCCAGTCTGATGATTCAGCGGTTATTCATTATGATAAGAATAAAGATGAAACTGG
CACCATTAATTATGCGAGTGTAACTTTGGGTAAAGGTGAAGATTCTGCAGCTGTTGCCCTTCATAATGTCGCTGCAGGTAATATTGCTAAG
GATTCACGTGATGCAATCAATGGTTCTCAGCTTTATTCTTTGAACGAGCAGTTATTGACCTATTTTGGCGGTGATGCTGGCTATAAAGATG
GGCAATGGATAGCTCCCAAGTTCCATGTTTTGCAGTTCAAGAGTGATGGTAGTTCTGGTGAGAAGGAGAGCTATGATAATGTAGCGGCTGC
GTTTGAAGGAGTTAACAAAAGTCTTGCAGGTATGAACGAGCGTATTAATAATGTTACTGCTGGCCAGAATGTTTCGTCGAGCAGTTTAAAT
TGGAATGAGACGGAGGGAGGTTATGACGCTCGTCATAATGGTGTGGACAGTAAGCTTACGCATGTAGAGAATGGTGACGTATCCGAAAAAT
CGAAAGAAGCCGTTAATGGAAGTCAACTATGGAATACGAATGAGAAAGTTGAAGCGGTTGAGAAGGATGTAAAGAATATTGAGAAGAAGGT
ACAAGATATTGCTACAGTAGCAGATAGTGCTGTTAAGTATGAGAAAGATAGTACTGGCAAGAAAACGAATGTAATCAAATTAGTTGGTGGG
AGTGAAAGTGAGCCAGTATTGATAGACAATGTAGCGGATGGTAAAATTGAAGCAGACTCTAAGCAGGCAGTCAATGGAGGTCAGTTGCGTG
ATTATACTGAGAAACAGATGAAGATAGTGCTTGATGATGCGAAGAAAATATACGGATGAACGCTTCAATGATGTCGTCAATAATGGTATTAA
TGAGGCTAAAGCTTATACAGATGTGAAGTTTGAGGCTTTAAGTTACACTGTTGAGGAAGTCCGGAAAGAAGCAAGACAAGCAGCGGCTATT
GGTTTAGCAGTATCTAACTTACGTTACTATGATATACCAGGATCTTTAAGTCTTTCATTTGGTACGGGTATATGGCGTAGTCAGTCTGCAT
TTGCTATTGGTGCTGGTTATACATCTGAAGATGGCAATATTCGTTCTAATTTATCTATCACGAGTTCTGGTGGTCAGTGGGGAGTAGGCGC
AGGGATTACTTTGAGACTGAAA**TGA**TAAAAAAACTAATATTATGATAGAAAACGAAGTATTTTGATAAATATTCTGTTCTTCCTTGCCTTT
ATTAGGCAAGGGAGAAAGTTTTGCTGATGAAAACGATAGTGTTTATACGGTGCATCCACCGCATTTATCTATTCCTAATGGGGTAGCGGGT
GAAACACGTCGAATCATCATGCAGTTTTATTATTGGACTTTAATTTGTGATGAAAACAAAAGCTTAGGCAAGGCATATGTAATGTGACGC
AAACTGTCCATGATAAGGAAGGCAATACTATTTTCAGTTGGTCTCTTGTTTCTACGAAAAAGAATTC
```

BadA D16S28 protein sequence

```
MKKLSVTSKRQYNLYASPISRRLSLLMKLSLETVTVMFLLGASPVLASNL<span style="background:yellow">AG</span>QLYSLNEQLATYFGGGAKYENGQWTAPTFKVKTVNGEGK
EEEQTYQNVAAAFEGVGTSFTNIKSEITKQINNEIINVKGDSLVKRDLATNLITIGKEIEGSVINIANKSGEARTISGVKEAVKDNEAVNK
GQ<span style="background:yellow">LY</span>SMSNMLATYLGGNAKYENGEWTAPTFKVKTVNGEGKEEEQTYQNVAEALTGVGTSFTNIKSEIAKQINHLQSDDSAVIHYDKNKDET
GTINYASVTLGKGEDSAAVALHNVAAGNIAKDSRDAINGSQLYSLNEQLLTYFGGDAGYKDGQWIAPKFHVLQFKSDGSSGEKESYDNVAA
AFEGVNKSLAGMNERINNVTAGQNVSSSSLNWNETEGGYDARHNGVDSKLTHVENGDVSEKSKEAVNGSQLWNTNEKVEAVEKDVKNIEKK
VQDIATVADSAVKYEKDSTGKKTNVIKLVGGSESEPVLIDNVADGKIEADSKQAVNGGQLRDYTEKQMKIVLDDAKKYTDERFNDVVNNGI
NEAKAYTDVKFEALSYTVEEVRKEARQAAAIGLAVSNLRYYDIPGSLSLSFGTGIWRSQSAFAIGAGYTSEDGNIRSNLSITSSGGQWGVG
AGITLRLK-
```

Suppl. Figure 7. Synthesised and truncated _badA_ gene and BadA protein D16S28 sequences.
The grey highlighted region translates to the truncated BadA D16S28 protein sequence shown directly underneath. The underlined sequence represents the cleaved-off signal sequence during transport into the periplasm. Yellow highlighted nt and aa depict the transition site of combined _badA_ regions.

badA D19S28 gene sequence

```
GAATTCCTGAATTTAGAGAGTGTAAGCTTTTATAGAAGCGTGCTGTTCTCTTTGAAAAGGAATGGTATTGTTCACAAAAAGTACTGTTTTT
ATTATGAACTAAAAAAATTTATTTTTAGCTTGCTATTTTACTCAATAGAGGATAGTGATACAGAAGGTATATCAGTATACTCATTTTAATT
ATAACTTCAAAAGGGGAGGAAGTAATGCGTAAAAGACGAAACGCCACTCTAAAAGCAAATTTACATACCGCATCACACTCAATATAAAGAA
ACACTCGTAACAGAAATCAACTAAGCATACAGATTTCTTTTAAATATTCTTCAAATTCTCTTATTATTAAGAAAAGATGCTCCTTAATGAA
AAAATTTTTTAATAAAACAGATAGCAATAAAAGAATGATTGAAATATTATTTAAACAACACCACCCTAACGTAAAACGTCTTAATATTTAA
AACAGAAAAATTCTTTTTTAAGTACACAACAAAAACAACCGCTCAACCCCTATTACAATCCAAATGCGCTATTTACACGCTTCCTACCAAG
CTTTCGCATTCAGATTTCATTACAGAAAGTACACACAAAATAAAAATAAAGACTCAAAACGTTCCCAATTTGACCACCCTCCTTATTTTAA
TCCTCATTACAAGGGAGTAGGTAATACTAAAATGTGTCTTTTTTATGTTTTGGATGTGCTTTGTAATTTTTTTCATTGGAGAATTTATT**AT**
**G**AAAAAATTATCTGTCACATCAAAGAGACAATATAATTTATATGCTTCGCCTATTTCTCGACGTTTATCTTTGTTAATGAAGCTCTCATTG
GAAACTGTAACAGTTATGTTCTTATTGGGTGCATCTCCTGTATTGGCTTCGAATCTTGC<mark>GG</mark>GCCAGCTTGATGCCAATATCAGTAAAGTTA
ATAATAATGTAACGAATAAGTTTAATGAACTTACTCAAAGCATAACGAATGTTACGCAACAGGTAAAAGGCGATGCCTTATTATGGAGCGA
TGAAGCCAATGCTTTTGTGGCGCGTCATGAAAGAGTAAGTTAGAAAAAGGCGTATCTAAAGCGACACAAGAAAATAGCAAGATTACGTAT
CTGTTAGATGGTGATATTTCGAAAGGTTCCACGGATGCCGTTACCGGTGGTCAGCT<mark>TT</mark>ATTCAATGAGCAATATGCTTGCGACCTATTTGG
GTGGTAACGCTAAATATGAGAATGGTGAATGGACCGCACCTACCTTTAAGGTTAAAACAGTTAACGGTGAAGGCAAGGAAGAAGAGCAAAC
TTATCAGAATGTAGCGGAAGCTTTGACTGGAGTTGGTACGTCTTTCACCAATATAAAAAGTGAGATTGCCAAACAGATTAATCATCTCCAG
TCTGATGATTCAGCGGTTATTCATTATGATAAGAATAAAGATGAAACTGGCACCATTAATTATGCGAGTGTAACTTTGGGTAAAGGTGAAG
ATTCTGCAGCTGTTGCCCTTCATAATGTCGCTGCAGGTAATATTGCTAAGGATTCACGTGATGCAATCAATGGTTCTCAGCTTTATTCTTT
GAACGAGCAGTTATTGACCTATTTTGGCGGTGATGCTGGCTATAAAGATGGGCAATGGATAGCTCCCAAGTTCCATGTTTGCAGTTCAAG
AGTGATGGTAGTTCTGGTGAGAAGGAGAGCTATGATAATGTAGCGGCTGCGTTTGAAGGAGTTAACAAAAGTCTTGCAGGTATGAACGAGC
GTATTAATAATGTTACTGCTGGCCAGAATGTTTCGTCGAGCAGTTTAAATTGGAATGAGACGGAGGGAGGTTATGACGCTCGTCATAATGG
TGTGGACAGTAAGCTTACGCATGTAGAGAATGGTGACGTATCCGAAAAATCGAAAGAAGCCGTTAATGGAAGTCAACTATGGAATACGAAT
GAGAAAGTTGAAGCGGTTGAGAAGGATGTAAAGAATATTGAGAAGAAGGTACAAGATATTGCTACAGTAGCAGATAGTGCTGTTAAGTATG
AGAAAGATAGTACTGGCAAGAAAACGAATGTAATCAAATTAGTTGGTGGGAGTGAAAGTGAGCCAGTATTGATAGACAATGTAGCGGATGG
TAAAATTGAAGCGAGACTCTAAGCAGGCAGTCAATGGAGGTCAGTTGCGTGATTATACTGAGAAACAGATGAAGATAGTGCTTGATGATGCG
AAGAAATATACGGATGAACGCTTCAATGATGTCGTCAATAATGGTATTAATGAGGCTAAAGCTTATACAGATGTGAAGTTTGAGGCTTTAA
GTTACACTGTTGAGGAAGTCCGGAAAGAAGCAAGCAAGCAGCGGCTATTGGTTTAGCAGTATCTAACTTACGTTACTATGATATACCAGG
ATCTTTAAGTCTTTCATTTGGTACGGGTATATGGCGTAGTCAGTCTGCATTTGCTATTGGTGCTGGTTATACATCTGAAGATGGCAATATT
CGTTCTAATTTATCTATCACGAGTTCTGGTGGTCAGTGGGGAGTAGGCGCAGGGATTACTTTGAGACTGAAA**TGA**TAAAAAACTAATATT
ATGATAGAAAAACGAAGTATTTTGATAAATATTCTGTTCTTCCTTGCCTTATTAGGCAAGGGAGAAAGTTTTGCTGATGAAAACGATAGTG
TTTATACGGTGCATCCACCGCATTTATCTATTCCTAATGGGGTAGCGGGTGAAACACGTCGAATCATCATGCAGTTTTATTATTGGACTTT
AATTTGTGATGAAAACAAAAGCTTAGGCAAGGCATATGTAATGTGACGCAAACTGTCCATGATAAGGAAGGCAATACTATTTTCAGTTGG
TCTCTTGTTTCTACGAAAAAGAATTC
```

BadA D19S28 protein sequence

```
MKKLSVTSKRQYNLYASPISRRLSLLMKLSLETVTVMFLLGASPVLASNL<mark>AG</mark>QLDANISKVNNNVTNKFNELTQSITNVTQQVKGDALLWS
DEANAFVARHEKSKLEKGVSKATQENSKITYLLDGDISKGSTDAVTGGQ<mark>LY</mark>SMSNMLATYLGGNAKYENGEWTAPTFKVKTVNGEGKEEEQ
TYQNVAEALTGVGTSFTNIKSEIAKQINHLQSDDSAVIHYDKNKDETGTINYASVTLGKGEDSAAVALHNVAAGNIAKDSRDAINGSQLYS
LNEQLLTYFGGDAGYKDGQWIAPKFHVLQFKSDGSSGEKESYDNVAAAFEGVNKSLAGMNERINNVTAGQNVSSSSLNWNETEGGYDARHN
GVDSKLTHVENGDVSEKSKEAVNGSQLWNTNEKVEAVEKDVKNIEKKVQDIATVADSAVKYEKDSTGKKTNVIKLVGGSESEPVLIDNVAD
GKIEADSKQAVNGGQLRDYTEKQMKIVLDDAKKYTDERFNDVVNNGINEAKAYTDVKFEALSYTVEEVRKEARQAAAIGLAVSNLRYYDIP
GSLSLSFGTGIWRSQSAFAIGAGYTSEDGNIRSNLSITSSGGQWGVGAGITLRLK-
```

Suppl. Figure 8. Synthesised and truncated *badA* gene and BadA protein D19S28 sequences. The grey highlighted region translates to the truncated BadA D19S28 protein sequence shown directly underneath. The underlined sequence represents the cleaved-off signal sequence during transport into the periplasm. Yellow highlighted nt and aa depict the transition site of combined *badA* regions.

badA D25S28 gene sequence

```
GAATTCCTGAATTTAGAGAGTGTAAGCTTTTATAGAAGCGTGCTGTTCTCTTTGAAAAGGAATGGTATTGTTCACAAAAAGTACTGTTTTT
ATTATGAACTAAAAAAATTTATTTTTAGCTTGCTATTTTACTCAATAGAGGATAGTGATACAGAAGGTATATCAGTATACTCATTTTAATT
ATAACTTCAAAAGGGGAGGAAGTAATGCGTAAAAGACGAAACGCCACTCTAAAAGCAAATTTACATACCGCATCACACTCAATATAAAGAA
ACACTCGTAACAGAAATCAACTAAGCATACAGATTTCTTTTAAATATTCTTCAAATTCTCTTATTATTAAGAAAAGATGCTCCTTAATGAA
AAAATTTTTTAATAAAACAGATAGCAATAAAAGAATGATTGAAATATTATTTAAACAACACCACCCTAACGTAAAACGTCTTAATATTTAA
AACAGAAAAATTCTTTTTTAAGTACACAACAAAAACAACCGCTCAACCCCTATTACAATCCAAATGCGCTATTTACACGCTTCCTACCAAG
CTTTCGCATTCAGATTTCATTACAGAAAGTACACACAAAATAAAAATAAAGACTCAAAACGTTCCCAATTTGACCACCCTCCTTATTTTAA
TCCTCATTACAAGGGAGTAGGTAATACTAAAATGTGTCTTTTTTATGTTTTGGATGTGCTTTGTAATTTTTTTCATTGGAGAATTTATT̲A̲T̲
G̲̲AAAAAATTATCTGTCACATCAAAGAGACAATATAATTTATATGCTTCGCCTATTTCTCGACGTTTATCTTTGTTAATGAAGCTCTCATTG
GAAACTGTAACAGTTATGTTCTTATTGGGTGCATCTCCTGTATTGGCTTCGAATCTTGC GG GGCAGCTTGATACCAATATCAAGAAAGTAG
AAGATAAATTAACAGAAGCAGTCGGTAAAGTTACGCAACAGGTAAAAGGTGATGCTTTATTGTGGAGCAATGAAGATAACGCGTTTGTTGC
TGATCATGGTAAAGATAGCGCAAAGACAAAGAGCAAGATTACACATTTATTAGATGGAAATATTGCGTCTGGCTCAACCGATGCCGTTACC
GGTGGTCAACT CT ATTCAATGAGCAATATGCTTGCGACCTATTTGGGTGGTAACGCTAAATATGAGAATGGTGAATGGACCGCACCTACCT
TTAAGGTTAAAACAGTTAACGGTGAAGGCAAGGAAGAAGAGCAAACTTATCAGAATGTAGCGGAAGCTTTGACTGGAGTTGGTACGTCTTT
CACCAATATAAAAAGTGAGATTGCCAAACAGATTAATCATCTCCAGTCTGATGATTCAGCGGTTATTCATTATGATAAGAATAAAGATGAA
ACTGGCACCATTAATTATGCGAGTGTAACTTTGGGTAAAGGTGAAGATTCTGCAGCTGTTGCCCTTCATAATGTCGCTGCAGGTAATATTG
CTAAGGATTCACGTGATGCAATCAATGGTTCTCAGCTTTATTCTTTCTTTGAACGAGCAGTTATTGACCTATTTTGGCGGTGATGCTGGCTATAA
AGATGGGCAATGGATAGCTCCCAAGTTCCATGTTTTGCAGTTCAAGAGTGATGGTAGTTCTGGTGAGAAGGAGAGCTATGATAATGTAGCG
GCTGCGTTTGAAGGAGTTAACAAAAGTCTTGCAGGTATGAACGAGCGTATTAATAATGTTACTGCTGGCCAGAATGTTTCGTCGAGCAGTT
TAAATTGGAATGAGACGGAGGGAGGTTATGACGCTCGTCATAATGGTGTGGACAGTAAGCTTACGCATGTAGAGAATGGTGACGTATCCGA
AAAATCGAAAGAAGCCGTTAATGGAAGTCAACTATGGAATACGAATGAGAAAGTTGAAGCGGTTGAGAAGGATGTAAAGAATATTGAGAAG
AAGGTACAAGATATTGCTACAGTAGCAGATAGTGCTGTTAAGTATGAGAAAGATAGTACTGGCAAGAAAACGAATGTAATCAAATTAGTTG
GTGGGAGTGAAAGTGAGCCAGTATTGATAGACAATGTAGCGGATGGTAAAATTGAAGCAGACTCTAAGCAGGCAGTCAATGGAGGTCAGTT
GCGTGATTATACTGAGAAACAGATGAAGATAGTGCTTGATGATGCGAAGAAATATACGGATGAACGCTTCAATGATGTCGTCAATAATGGT
ATTAATGAGGCTAAAGCTTATACAGATGTGAAGTTTGAGGCTTTAAGTTACACTGTTGAGGAAGTCCGGAAAGAAGCAAGACAAGCAGCGG
CTATTGGTTTAGCAGTATCTAACTTACGTTACTATGATATACCAGGATCTTTAAGTCTTTCATTTGGTACGGGTATATGGCGTAGTCAGTC
TGCATTTGCTATTGGTGCTGGTTATACATCTGAAGATGGCAATATTCGTTCTAATTTATCTATCACGAGTTCTGGTGGTCAGTGGGGAGTA
GGCGCAGGGATTACTTTGAGACTGAAA**TGA**TAAAAAAACTAATATTATGATAGAAAACGAAGTATTTTGATAAATATTCTGTTCTTCCTT
GCCTTATTAGGCAAGGGAGAAAGTTTTGCTGATGAAACGATAGTGTTTATACGGTGCATCCACCGCATTTATCTATTCCTAATGGGGTAG
CGGGTGAAACACGTCGAATCATCATGCAGTTTTATTATTGGACTTTAATTTGTGATGAAAACAAAAGCTTAGGCAAGGCATATGTAATGT
GACGCAAACTGTCCATGATAAGGAAGGCAATACTATTTTCAGTTGGTCTCTTGTTTCTACGAAAAAGAATTC
```

BadA D25S28 protein sequence

```
M̲K̲K̲L̲S̲V̲T̲S̲K̲R̲Q̲Y̲N̲L̲Y̲A̲S̲P̲I̲S̲R̲R̲L̲S̲L̲L̲M̲K̲L̲S̲L̲E̲T̲V̲T̲V̲M̲F̲L̲L̲G̲A̲S̲P̲V̲L̲A̲S̲N̲L AG QLDTNIKKVEDKLTEAVGKVTQQVKGDALLWSNEDNAFV
ADHGKDSAKTKSKITHLLDGNIASGSTDAVTGGQ LY SMSNMLATYLGGNAKYENGEWTAPTFKVKTVNGEGKEEEQTYQNVAEALTGVGTS
FTNIKSEIAKQINHLQSDDSAVIHYDKNKDETGTINYASVTLGKGEDSAAVALHNVAAGNIAKDSRDAINGSQLYSLNEQLLTYFGGDAGY
KDGQWIAPKFHVLQFKSDGSSGEKESYDNVAAAFEGVNKSLAGMNERINNVTAGQNVSSSSLNWNETEGGYDARHNGVDSKLTHVENGDVS
EKSKEAVNGSQLWNTNEKVEAVEKDVKNIEKKVQDIATVADSAVKYEKDSTGKKTNVIKLVGGSESEPVLIDNVADGKIEADSKQAVNGGQ
LRDYTEKQMKIVLDDAKKYTDERFNDVVNNGINEAKAYTDVKFEALSYTVEEVRKEARQAAAIGLAVSNLRYYDIPGSLSLSFGTGIWRSQ
SAFAIGAGYTSEDGNIRSNLSITSSGGQWGVGAGITLRLK–
```

Suppl. Figure 9. Synthesised and truncated *badA* gene and BadA protein D25S28 sequences.
The grey highlighted region translates to the truncated BadA D25S28 protein sequence shown directly
underneath. The underlined sequence represents the cleaved-off signal sequence during transport into
the periplasm. Yellow highlighted nt and aa depict the transition site of combined *badA* regions.

badA D27S29 gene sequence

```
GAATTCCTGAATTTAGAGAGTGTAAGCTTTTATAGAAGCGTGCTGTTCTCTTTGAAAAGGAATGGTATTGTTCACAAAAAGTACTGTTTTT
ATTATGAACTAAAAAAATTTATTTTTAGCTTGCTATTTTACTCAATAGAGGATAGTGATACAGAAGGTATATCAGTATACTCATTTTAATT
ATAACTTCAAAAGGGGAGGAAGTAATGCGTAAAAGACGAAACGCCACTCTAAAAGCAAATTTACATACCGCATCACACTCAATATAAAGAA
ACACTCGTAACAGAAATCAACTAAGCATACAGATTTCTTTTAAATATTCTTCAAATTCTCTTATTATTAAGAAAAGATGCTCCTTAATGAA
AAAATTTTTTAATAAAACAGATAGCAATAAAAGAATGATTGAAATATTATTTAAACAACACCACCCTAACGTAAAACGTCTTAATATTTAA
AACAGAAAAATTCTTTTTTAAGTACACAACAAAAACAACCGCTCAACCCCTATTACAATCCAAATGCGCTATTTACACGCTTCCTACCAAG
CTTTCGCATTCAGATTTCATTACAGAAAGTACACACAAAATAAAAATAAAGACTCAAAACGTTCCCAATTTGACCACCCTCCTTATTTTAA
TCCTCATTACAAGGGAGTAGGTAATACTAAAATGTGTCTTTTTTATGTTTTGGATGTGCTTTGTAATTTTTTTCATTGGAGAATTTATT<u>**AT**</u>
<u>**G**</u>AAAAAATTATCTGTCACATCAAAGAGACAATATAATTTATATGCTTCGCCTATTTCTCGACGTTTATCTTTGTTAATGAAGCTCTCATTG
GAAACTGTAACAGTTATGTTCTTATTGGGTGCATCTCCTGTATTGGCTTCGAATCTTGC<mark>GG</mark>GACAGATTCATACAATCGGTGAGGATGTTG
CAAAATTCTTGGGTGGAGATGCAGCTTTTAAAGATGGCGCTTTTACCGGCCCAACTTATAAGTTGTCGAATATTGATGCAAAGGGTGATGT
ACAACAGAGTGAGTTTAAAGATATAGGTTCAGCCTTTGCGGGTCTTGATACGAACATCAAGAATGTCAATAATAATGTAACGAATAAGCTC
AGTGAACTTACTCAAAACATAACGACTGTTACGCAACAGGTAAAAGGCAATGCCTTATTATGGAGCGATGAAGCTAATGCCTTTGTGGCGC
GTCATGAAAAGAGCAAGTTAGAAAAAGGTGCATCTAAAGCGATACAAGAAACAGCAAGATTACGTATCTGTTAGATGGTGATGTTTCGAA
AGGTTCCACGGATGCCGTTACTGGTGGTCAGCT<mark>TT</mark>ATTCTTTGAACGAGCAGTTATTGACCTATTTTGGCGGTGATGCTGGCTATAAAGAT
GGGCAATGGATAGCTCCCAAGTTCCATGTTTTGCAGTTCAAGAGTGATGGTAGTTCTGGTGAGAAGGAGAGCTATGATAATGTAGCGGCTG
CGTTTGAAGGAGTTAACAAAAGTCTTGCAGGTATGAACGAGCGTATTAATAATGTTACTGCTGGCCAGAATGTTTCGTCGAGCAGTTTAAA
TTGGAATGAGACGGAGGGAGGTTATGACGCTCGTCATAATGGTGTGGACAGTAAGCTTACGCATGTAGAGAATGGTGACGTATCCGAAAAA
TCGAAAGAAGCCGTTAATGGAAGTCAACTATGGAATACGAATGAGAAGTTGAAGCGGTTGAGAAGGATGTAAAGAATATTGAGAAGAAGG
TACAAGATATTGCTACAGTAGCAGATAGTGCTGTTAAGTATGAGAAGATAGTACTGGCAAGAAAACGAATGTAATCAAATTAGTTGGTGG
GAGTGAAAGTGAGCCAGTATTGATAGACAATGTAGCGGATGGTAAAATTGAAGCAGACTCTAAGCAGGCAGTCAATGGAGGTCAGTTGCGT
GATTATACTGAGAAACAGATGAAGATAGTGCTTGATGATGCGAAGAAATATACGGATGAACGCTTCAATGATGTCGTCAATAATGGTATTA
ATGAGGCTAAAGCTTATACAGATGTGAAGTTTGAGGCTTTAAGTTACACTGTTGAGGAAGTCCGGAAAGAAGCAAGACAAGCAGCGGCTAT
TGGTTTAGCAGTATCTAACTTACGTTACTATGATATACCAGGATCTTTAAGTCTTTCATTTGGTACGGGTATATGGCGTAGTCAGTCTGCA
TTTGCTATTGGTGCTGGTTATACATCTGAAGATGGCAATATTCGTTCTAATTTATCTATCACGAGTTCTGGTGGTCAGTGGGGAGTAGGCG
CAGGGATTACTTTGAGACTGAAA**TGA**TAAAAAAACTAATATTATGATAGAAAACGAAGTATTTTGATAAATATTCTGTTCTTCCTTGCCT
TATTAGGCAAGGGAGAAAGTTTTGCTGATGAAAACGATAGTGTTTATACGGTGCATCCACCGCATTTATCTATTCCTAATGGGGTAGCGGG
TGAAACACGTCGAATCATCATGCAGTTTTATTATTGGACTTTAATTTGTGATGAAAACAAAAGCTTAGGCAAGGCATATGTAATGTGACG
CAAACTGTCCATGATAAGGAAGGCAATACTATTTTCAGTTGGTCTCTTGTTTCTACGAAAAAGAATTC
```

BadA D27S29 protein sequence

```
<u>MKKLSVTSKRQYNLYASPISRRLSLLMKLSLETVTVMFLLGASPVLASNL</u><mark>AG</mark>QIHTIGEDVAKFLGGDAAFKDGAFTGPTYKLSNIDAKGD
VQQSEFKDIGSAFAGLDTNIKNVNNNVTNKLSELTQNITTVTQQVKGNALLWSDEANAFVARHEKSKLEKGASKAIQENSKITYLLDGDVS
KGSTDAVTGGQ<mark>LY</mark>SLNEQLLTYFGGDAGYKDGQWIAPKFHVLQFKSDGSSGEKESYDNVAAAFEGVNKSLAGMNERINNVTAGQNVSSSSL
NWNETEGGYDARHNGVDSKLTHVENGDVSEKSKEAVNGSQLWNTNEKVEAVEKDVKNIEKKVQDIATVADSAVKYEKDSTGKKTNVIKLVG
GSESEPVLIDNVADGKIEADSKQAVNGGQLRDYTEKQMKIVLDDAKKYTDERFNDVVNNGINEAKAYTDVKFEALSYTVEEVRKEARQAAA
IGLAVSNLRYYDIPGSLSLSFGTGIWRSQSAFAIGAGYTSEDGNIRSNLSITSSGGQWGVGAGITLRLK–
```

Suppl. Figure 10. Synthesised and truncated _badA_ gene and BadA protein D27S29 sequences.
The grey highlighted region translates to the truncated BadA D27S29 protein sequence shown directly
underneath. The underlined sequence represents the cleaved-off signal sequence during transport into
the periplasm. Yellow highlighted nt and aa depict the transition site of combined _badA_ regions.

Supplementary tables

Suppl. Table 1. Overview of the major flanking genes upstream of the *badA* island. ORFs are identified in a region ca. 15 kb upstream of the *badA* island. Length and start position of the listed ORFs refer to the genome of *B. henselae* Marseille, unless indicated otherwise.

Identified ORF	Length (bp)	Start position (nt)	Predicted remote homologue[b]	*B. henselae* genomic conservation[c]
Transcriptional repressor/ iron response regulator	504	1,330,292	Ferric uptake regulation (Fur) protein	1 bp mutation in strain FR96/BK3
Efflux resistance-nodulation-cell division transporter permease	3,135	1,336,864	/	≤ 9 bp mutations in strains Marseille, FR96/BK38, and FR96/BK3
Glucose-6-phosphate isomerase	1,665	1,340,196	/	≤ 8 bp mutations in strains Marseille, FR96/BK38, and FR96/BK3
50S ribosomal protein L21	477	1,342,273	/	≤ 3 bp mutations in strains Marseille, FR96/BK38, and FR96/BK3
50S ribosomal protein L27	270	1,342,764	/	≤ 3 bp mutations in strains Marseille and FR96/BK3
Phosphoenolpyruvate-protein phosphotransferase	2,505	1,344,453	/	≤ 22 bp mutations in strains Marseille, FR96/BK38, and FR96/BK3
Mobile genetic element	480	1,350,853	AAA+ ATPase superfamily proteins: 1) MuB transposition protein 2) transposon Tn7 transposition protein 3) RuvB-like protein 4) replication-associated recombination protein A	≤ 29 bp mutations in strains Marseille and FR96/BK3 (201 bp)
Helix-turn-helix domain-containing protein	294	1,351,401	Competence regulator (ComR) transcriptional factor	≤ 14 bp mutations in strains Marseille, FR96/BK38 (204 bp), and FR96/BK3 (351 bp); all other strains show an ORF of 204 bp
Helix-turn-helix transcriptional regulator	540	1,351,840	ComR transcriptional factor	≤ 15 bp mutations in strains Marseille, FR96/BK38 (504 bp), and FR96/BK3 (528 bp); all other strains show an ORF of 504 bp
Helix-turn-helix domain-containing protein[a]	300[a]	1,380,964[a]	ComR transcriptional factor	≤ 4 bp mutations in strains FR96/BK38 and FR96/BK3; not present in strain Marseille
Helix-turn-helix transcriptional regulator[a]	543[a]	1,381,410[a]	ComR transcriptional factor	≤ 6 bp mutations in strains FR96/BK38 (174 bp) and FR96/BK3; not present in strain Marseille

[a]Refer to the genome of strain ATCC49882[T] var-1, [b]via HHpred software, [c]compared to the consensus sequence

Suppl. Table 2. Overview of the major flanking genes downstream of the *badA* island. ORFs are identified in a region ca. 15 kb downstream of the *badA* island. Length and start position of the listed ORFs refer to the genome of *B. henselae* Marseille.

Identified ORF	Length (bp)	Start position (nt)	Predicted remote homologue(s)[a]	*B. henselae* genomic conservation[b]
Invasion associated locus B (IalB) family protein	573	1,374,036	IalB family protein	≤ 14 bp mutations in strains Marseille and FR96/BK3 (573 bp); all other strains show an ORF of 594 bp
Guanosine triphosphate (GTP) binding protein	1,023	1,378,580	1) GTP-binding protein (ObgE) 2) Mitochondrial ribosome-associated GTPase 2	≤ 5 bp mutations in strains Marseille, FR96/BK38, and FR96/BK3
Ribosome silencing factor	444	1,383,010	Ribosomal silencing factor (RsfS)	≤ 9 bp mutations in strains Marseille, FR96/BK38, and FR96/BK3
Peptidoglycan DD-metalloendopeptidase family protein	1,281	1,384,033	Murein hydrolase activator (EnvC)	≤ 2 bp mutations in strains Marseille, FR96/BK38, and FR96/BK3
IalB family protein	561	1,387,316	/	1 bp in strain FR96/BK38
YggT family protein	288	1,388,206	/	1 bp in strain FR96/BK38

[a]via HHpred, [b]compared to the consensus sequence

Acknowledgments

I would first like to thank Professor Volkhard A. J. Kempf for the opportunity to be part of an international EU-funded MSCA ITN-consortium (ViBrANT) and to be able to conduct research in the *Bartonella* laboratory group at the University Hospital Frankfurt. I am mostly thankful for his continuous support, feedback, and guidance during my 4-years period as a PhD-researcher. I would also like to thank Professor Volker Müller for acting as my main supervisor and for the possibility of pursuing a PhD at the Johann Wolfgang Goethe University and the Faculty of Biological Sciences.

Furthermore, I would like to thank the ViBrANT-associated supervisory committee members Professor Sally A. Peyman (University of Leeds, United Kingdom) and Carina Almeida, PhD, (National Institute of Agrarian and Veterinary Research, Portugal) for their support and insightful advice during our biannual progress report meetings. Special thanks go to Professor Dirk Linke (University of Oslo, Norway) who was always available for a constructive discussion and for answering my endless stream of research-related questions throughout the full period of my PhD and especially during my 2-months research stay at his laboratory group. I would also like to thank my ViBrANT-associated peers for the great times and mental support during our many seminars, summer schools, and conferences.

In addition, I would like to thank everyone from the institute, the office, and the *Bartonella* laboratory group, especially Diana J. Vaca and Alexander Dichter, for the constant support and helpful discussions, but mostly for all the enjoyable and unforgettable moments we shared. Special thanks go to Wibke Ballhorn, who single-handedly kept the laboratory running and who was always an enormous help in all laboratory-related matters.

I would also like to thank my parents for their help and encouragement. Lastly and most importantly, I would like to thank Paulien Van de Velde, who has motivated me repeatedly and who was my biggest support throughout the 4-years of my PhD-project, she continuously believed in my capabilities and always encouraged me to give my best in every situation. A big thank you.

Curriculum Vitae

━━━━━━━━ **Personal data**

THIBAU Arno

Birthday and -place: 29 June 1995, Lokeren, Belgium   **Phone number**: 0032 471 62 17 69

Nationality: Belgian   **E-mail address**: arno.thibau@hotmail.com

━━━━━━━━ **Education**

PhD-candidate in natural sciences   Goethe University Frankfurt, Institute
(2018 - 2022)   for Medical Microbiology, Germany

Master of Science in Biochemical Engineering Technology   Ghent University, Belgium
(2016 - 2017)   ~ *Magna Cum Laude*

Bachelor of Science in Bioscience Engineering Technology   Ghent University, Belgium
(2013 - 2016)   ~ *Cum Laude*

Secondary education in Latin-Sciences   Sint-Lodewijkscollege, Belgium
(2007 - 2013)

━━━━━━━━ **Professional experience**

Early Stage Researcher   Institute for Medical Microbiology - University Hospital Frankfurt
PhD-candidate   Part of the European MSCA-ITN called 'ViBrANT'
(2018 - 2022)   Project: *Characterisation of the fibronectin binding domains and
                               genomic variation of the Bartonella adhesin A in Bartonella henselae*

ViBrANT2021 conference   Organisation of 2-day online conference on bacterial and viral
(2019 - 2021)   adhesion with 100 registered participants (MSCA-ITN project)

Tutoring medical student   University Hospital Frankfurt - Institute for Medical Microbiology
(2019-2021)   Project: *Immunodominant domains within the Bartonella adhesin A*

Scientific secondment   University of Oslo - Bacterial Cell Surface Group
(May - June 2019)   in the context of the MSCA-ITN project

Trainee Innovation Manager   Ghent University - Faculty of Bioscience Engineering
(April - August 2018)   Strategic Relation Office & International Training Centre

FWO-SB PhD application   Ghent University - Laboratory for Applied Biochemistry
(August - December 2017)   Project: *Cell wall-deficient Escherichia coli L-forms as innovative
                               microbial cell factories*

Master thesis   Ghent University - Laboratory for Applied Biochemistry
(2016 - 2017)   Project: *Construction of designer cellulosomes towards the
                               degradation of galacto- and glucomannan in bioethanol production*

Short internship   Flanders Research institute for Agriculture, Fisheries and Food
(August 2016)   (ILVO)

━━━━━━━━ **Languages**

Dutch   C2 - Native proficiency

English   C2 - Full professional proficiency

German   B2 - Professional working proficiency

French   B1 - Limited working proficiency

Publications

Thibau A, Vaca DJ, Bagowski M, Hipp K, Bender D, Ballhorn W, Linke D, and Kempf VAJ (2022). Adhesion of *Bartonella henselae* to fibronectin is mediated via repetitive motifs present in the stalk of *Bartonella* adhesin A. Microbiol. Spectr. https://doi.org/10.1128/spectrum.02117-22

Thibau A, Hipp K, Vaca DJ, Chowdhury S, Malmström J, Saragliadis A, Ballhorn W, Linke D, and Kempf VAJ (2022). Long-read sequencing reveals genetic adaptation of *Bartonella* adhesin A among different *Bartonella henselae* isolates. Front. in Microbiol. https://doi.org/10.3389/fmicb.2022.838267

Vaca DJ, **Thibau A**, Leisegang MS, Malmström J, Linke D, Eble JA, Ballhorn W, Schaller M, Happonen L, and Kempf VAJ (2022). Interaction of *Bartonella henselae* with fibronectin represents the molecular basis for adhesion to host cells. Microbiol. Spectr. https://doi.org/10.1128/spectrum.00598-22

Van Belkum A, Almeida B, Bardiaux B, Barrass SV, Butcher SJ, Çaykara T, Chowdhury S, Datar R, Eastwood I, Goldman A, Goyal M, Happonen L, Izadi-Pruneyre N, Jacobsen T, Johnson PH, Kempf VAJ, Kiessling A, Bueno JL, Malik A, Malmström J, Meuskens I, Milner PA, Nilges M, Pamme N, Peyman SA, Rodrigues LR, Rodriguez-Mateos P, Sande MG, Silva CJ, Stasiak AC, Stehle T, **Thibau A**, Vaca DJ, and Linke D (2021). Host-pathogen adhesion as the basis of innovative diagnostics for emerging pathogens. Diagnostics. https://doi.org/10.3390/diagnostics11071259

Thibau A, Schultze TG, Ballhorn W, and Kempf VAJ (2020). Complete genome sequence of *Bartonella alsatica* strain IBS 382 (CIP 105477). Microbiol Resour Announc. https://doi.org/10.1128/MRA.00769-20

Vaca DJ, **Thibau A**, Schütz M, Kraiczy P, Happonen L, Malmström J, and Kempf VAJ, et al (2019). Interaction with the host: the role of fibronectin and extracellular matrix proteins in the adhesion of Gram-negative bacteria. Med Microbiol Immunol. https://doi.org/10.1007/s00430-019-00644-3

Thibau A, Dichter AA, Vaca DJ, Linke D, Goldman A, and Kempf VAJ, et al (2019). Immunogenicity of trimeric autotransporter adhesins and their potential as vaccine targets. Med Microbiol Immunol. https://doi.org/10.1007/s00430-019-00649-y

Conference participations

Participation in the 71st Annual Conference of German Society for Hygiene and Microbiology (DGHM) e.V. (Göttingen, Germany, 25-27 February 2019).

Poster presentation titled 'Analysing the fibronectin binding domain of the *Bartonella henselae* adhesin A' at the 8th Congress of European Microbiologists FEMS2019 (Glasgow, Scotland, 7-11 July 2019).

Oral presentation titled 'Analysing the fibronectin binding domain of the *Bartonella* adhesin A' at the 9th International Congress on *Bartonella* spp. as an Emerging Pathogen (Paris, France, 18-20 September 2019).

Poster presentation titled 'Analysing the fibronectin binding properties of *Bartonella henselae* adhesin A' at the 6th Joint Conference of the DGHM & VAAM – 72nd Annual Meeting of DGHM & Annual Meeting of VAAM (Leipzig, Germany, 8-11 March 2020).

Poster presentation titled 'Analysing the fibronectin binding properties and genomic variation of *Bartonella henselae* adhesin A' at the World Microbe Forum 2021 (online, 20-24 June 2021).

Poster presentation titled 'Analysis of fibronectin binding properties and genomic variation of *Bartonella henselae* adhesin A' at the 31st European Congress of Clinical Microbiology and Infectious Diseases (online, 9-12 July 2021).

Poster presentation titled 'Analysis of fibronectin binding properties and genomic variation of *Bartonella henselae* adhesin A' and organisation of the ViBrANT2021 conference (online, 15-16 July 2021).

Oral presentation titled 'Analysing the fibronectin binding properties and genomic variation of *Bartonella henselae* adhesin A' at the 73rd Annual Meeting of DGHM (online, 12-14 September 2021).

Oral presentation titled 'Molecular and functional analysis of *Bartonella henselae* host cell adhesion' at the International intracellular bacteria meeting 2022, a Joint ESCCAR International congress on Rickettsia and 9th Meeting of the European Society for Chlamydia Research (ESCR) (Lausanne, Switzerland, 23-26 August 2022).

Poster presentation titled 'Identifying the fibronectin binding regions and genomic variation of *Bartonella* adhesin A' at the 74th Annual Meeting of DGHM (Berlin, Germany, 5-7 September 2022).