Studies of Genome Diversity in \textit{Bartonella} Populations

\textit{A journey through cats, mice, men and lice}

HILLEVI LINA LINDROOS
Dissertation presented at Uppsala University to be publicly examined in Lindahlsalen, Evolutionsbiologiskt centrum, Norbyvägen 18, Uppsala, Friday, June 1, 2007 at 13:00 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

Bacteria of the genus Bartonella inhabit the red blood cells of many mammals, including humans, and are transmitted by blood-sucking arthropod vectors. Different species of Bartonella are associated with different mammalian host species, to which they have adapted and normally do not cause any symptoms. Incidental infection of other hosts is however often followed by various disease symptoms, and several Bartonella species are considered as emerging human pathogens.

In this work, I have studied the genomic diversity within and between different Bartonella species, with focus on the feline-associated human pathogen B. henselae and its close relatives, the similarly feline-associated B. koehlerae and the trench-fever agent B. quintana which is restricted to humans.

In B. henselae, the overall variability in sequence and genome content was modest and well correlated, suggesting low levels of intra-species recombination in the core genome. The variably present genes were located in the prophage and the genomic islands, which are also absent from B. quintana and B. koehlerae, indicating multiple independent excision events. In contrast, diversity of genome structures was immense and probably associated with rearrangements between the repeated genomic islands located around the terminus of replication, possibly to avoid the host’s immune system. In both B. henselae and the mouse-associated species B. grahamii a large portion of the chromosome was manifold amplified in long-time cultures and packaged into phage particles, allowing for different recombination rates for different chromosomal regions.

In B. quintana, diversity was studied by sequencing non-coding spacers. The low variability might be due to the recent emergence of this species. Surprisingly, also this species displayed high variability in genome structures, despite its lack of repeated sequences.

The results indicate that genome rearrangements and gain or loss of mobile elements are major mechanisms of evolution in Bartonella.

Keywords: Bartonella, genome content, genome rearrangement, gene expression, phage, microarray, PFGE, MLST

Hillevi Lina Lindroos, Department of Evolution, Genomics and Systematics, Norbyv. 18C, Uppsala University, SE-75236 Uppsala, Sweden

© Hillevi Lina Lindroos 2007

ISSN 1651-6214
ISBN 978-91-554-6917-7
urn:nbn:se:uu:diva-7927 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-7927)
List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I  **Lindroos, H. L.,** Mira, A., Repsilber, D., Vinnere, O., Näslund, K., Dehio, M., Dehio, C. and S. G. E. Andersson
Characterization of the genome composition of *Bartonella koehlerae* by microarray comparative genomic hybridization profiling.

Genome rearrangements, deletions, and amplifications in the natural population of *Bartonella henselae*.
*Journal of Bacteriology*, 2006, 188:7426-39

III  Foucault, C., La Scola, B., **Lindroos, H.,** Andersson, S. G. E. and D. Raoult
Multi-Spacer-Typing (MST) for sequence-based typing of *Bartonella quintana*.
*Journal of Clinical Microbiology*, 2005, 43:41-8

IV  **Lindroos, H. L.,** Berglund, E.*, Vinnere Pettersson, O.* and S. G. E. Andersson
Escape Replication and Packaging of the Chromosome II-like Segment into *Bartonella* Phage Particles in Long-time Cultures.
*Manuscript.*

V  Dehio, M.*, Quebatte, M.*, Basler, A., Toller, I., Raddatz, G., Tropel, D., **Lindroos, H. L.,** Andersson, S. G. E. and C. Dehio
The Expression Profile of *Bartonella* during Human Cell Infections Reveals the Adaptability of Two-Component System Regulators.
*Manuscript.*

* These authors contributed equally to this work.

Reprints of papers I-III were made with permissions from the publishers.
1. **Lindroos, H.** and S. G. E. Andersson  
   Visualizing Metabolic Pathways: Comparative Genomics and Expression Analysis.  

   Bartonella adhesin a mediates a proangiogenic host cell response.  

3. Repsilber, D., Mira, A., **Lindroos, H.**, Andersson, S. and A. Ziegler  
   Data rotation improves genomotyping efficiency.  
   *Biometrical Journal*, 2005, 47:585-98
Contents

1. Introduction ................................................................................................................................. 9
   Bacterial genomes .................................................................................................................. 9
       Genome size and content ................................................................................................. 9
       Genome structures ......................................................................................................... 11
   Genome evolution in bacteria ............................................................................................. 12
       Rearrangements – genome structure evolution ........................................................... 12
       Insertions, deletions and duplications – gene content evolution ................................ 14
       Point mutations – sequence evolution ........................................................................... 15
   The genus *Bartonella* ......................................................................................................... 15
       *Bartonella henselae* ...................................................................................................... 16
       *Bartonella quintana* .................................................................................................. 20
       *Bartonella koehlerae* ................................................................................................. 20
       *Bartonella grahamii* .................................................................................................. 21
       *Bartonella bacilliformis* ............................................................................................ 21
   The *Bartonella* genomes ..................................................................................................... 21
   The *Bartonella* phage(s) ..................................................................................................... 22

2. Aims ........................................................................................................................................ 24

3. Methodology ............................................................................................................................... 25
   Bacterial culture ................................................................................................................... 25
   Multi Locus Sequence Typing ............................................................................................. 25
   Multi Spacer Typing ............................................................................................................. 26
   Microarray Comparative Genomic Hybridizations .......................................................... 26
       Microarray design and construction ............................................................................ 27
       Microarray CGH hybridizations ................................................................................... 27
       Data analysis and visualization ...................................................................................... 27
       Gene-content based phylogeny ....................................................................................... 28
   Microarray Gene Expression Analysis .............................................................................. 29
   Pulsed Field Gel Electrophoresis and Southern Blotting ................................................. 29
       Experimental procedure ............................................................................................... 29
       Data analysis – hypothesis testing ................................................................................ 30

4. Results .................................................................................................................................... 32
   Validity of the microarray CGH results (Paper I) ............................................................ 32
   Genome content of *Bartonella koehlerae* (Paper I) ...................................................... 33
Genomic diversity within *Bartonella henselae* (Paper II) ..................34
  Sequence diversity ...........................................................................34
  Genome content diversity .................................................................35
  Genome structure diversity ...............................................................35
  Amplification – escape replication ....................................................36
Spacer typing of *Bartonella quintana* (Paper III) .............................37
Phage DNA content and growth-time related gene expression (Paper IV) ..................................................................................................................38
  *B. henselae* gene expression during infection (Paper V) .....................39
5. Summary and discussion ....................................................................41
Sammanfattning på svenska ..................................................................44
Acknowledgements ..................................................................................47
References .............................................................................................50
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-value</td>
<td>average of the base 2-logarithm of the Cy3 and Cy5 intensities for a microarray spot</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>CGH</td>
<td>Comparative Genomic Hybridization</td>
</tr>
<tr>
<td>indel</td>
<td>Insertion or deletion</td>
</tr>
<tr>
<td>kb</td>
<td>kilo base pairs ($10^3$ bp)</td>
</tr>
<tr>
<td>Mb</td>
<td>mega base pairs ($10^6$ bp)</td>
</tr>
<tr>
<td>MLST</td>
<td>Multi Locus Sequence Typing</td>
</tr>
<tr>
<td>MST</td>
<td>Multi Spacer Typing</td>
</tr>
<tr>
<td>M-value</td>
<td>base 2-logarithm of the ratio of the intensities for a microarray spot, probe or gene</td>
</tr>
<tr>
<td>PFGE</td>
<td>Pulsed Field Gel Electrophoresis</td>
</tr>
<tr>
<td>RFLP</td>
<td>Restriction Fragment Length Polymorphism</td>
</tr>
<tr>
<td>ST</td>
<td>Sequence Type</td>
</tr>
</tbody>
</table>
1. Introduction

“A subtle chain of countless rings;
the next unto the farthest brings...”

Ralph Waldo Emerson, *Nature*

This thesis describes studies of genome diversity and evolution in three pathogenic species of *Bartonella*, intracellular bacteria of the alpha subgroup of proteobacteria. *Bartonella* bacteria infect the red blood cells and the endothelial cells lining the blood vessels of a wide range of mammals, including humans. I have studied genomic diversity within and between several *Bartonella* species at different levels (genome structure, genome content and gene sequence) and tried to assess the relative importance of different evolutionary mechanisms. In addition, I have analyzed bacterial gene expression during various conditions.

**Bacterial genomes**

The last decade, several hundreds bacterial genomes have been sequenced, and more are in the pipeline (Binnewies et al. 2006; Ochman and Davalos 2006; Ward and Fraser 2005) (see [http://www.genomesonline.org](http://www.genomesonline.org) and [http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi](http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi)). The large amount of sequence data has given unprecedented insights into bacterial genome variability and evolution, showing some general trends but also great diversity, and still yielding new surprises with every new sequenced genome.

**Genome size and content**

Genome size varies a lot among different bacterial species; the smallest known to date (*Carsonella ruddi*) is only 160 kb (Nakabachi et al. 2006), which is smaller than some viruses, while the largest, *Sorangium cellulosum* is over 13 Mb (Liolios et al. 2006; Pradella et al. 2002), i.e. almost 100 times larger. The average bacterial genome size (of those sequenced to date) is 3.5 Mb.

Bacteria that live in close association with and are dependent on other organisms often have small genomes, since they can rely on the host to provide many metabolic functions. Obligate intracellular bacteria are also prone to
genome reduction since deletions are more common than duplications (see below). In contrast, free-living bacteria that encounter fluctuating environments often have large genomes and plasmids, that can easily be gained or lost (Moreno 1998).

Genome size varies widely even between closely related strains of the same species, for example natural isolates of the common model organism *Escherichia coli* differ by 1 Mb (Bergthorsson and Ochman 1998; Liolios et al. 2006), and those of *Ochrobactrum* by as much as 3 Mb (Teyssier et al. 2005). The differences in genome size within a species can often be attributed to the gain or loss of mobile elements, as discussed below.

**Base composition**

In addition to size, bacterial genomes also vary widely in their base composition, i.e. the relative amounts of the nucleotides A and T versus G and C, with the amount of G+C ranging from 17 % in the smallest genome *C. ruddi* to 75 % in the free-living gamma-proteobacterium *Anaeromyxobacter dehalogenans*. The average G+C content of the sequenced bacterial genomes is 48 %. There is a correlation between genome size and GC- content, such that small genomes of host-associated bacteria tend to have low content of G+C. The reason for the high variability in G+C-content is however still unclear, and has been suggested to be caused by differences in metabolism (Rocha and Danchin 2002), environment (Foerstner et al. 2005) or the composition of the DNA polymerase involved in genome replication (Zhao et al. 2007). The base composition also varies within individual genomes (GC-skew); see below.

**Mobile elements: Plasmids, phages and genomic islands**

In addition to, or inserted in, the core genome of many bacteria are sequences encoded by mobile elements, which may act as selfish parasites and/or contribute vital fitness factors to the bacterial host.

Plasmids are autonomously replicating circular sequences, which can be transferred between neighboring bacteria by conjugation, where the donor cell builds a channel through which the plasmid is transferred. Plasmids may encode sequences that are only needed under special circumstances, and which are therefore not necessary for all cells to harbor. Antibiotic resistance genes are often located on plasmids, facilitating their spread in the bacterial community (Alanis 2005). However, plasmids can also be seen as selfish elements using the host cell to sustain and be transmitted, and many plasmids even express stable toxins and rapidly degrading antitoxins to ascertain that the host dies if the plasmid is lost.

Phages are bacterial viruses and as such naturally parasitic. However, some phages occasionallypackage bacterial DNA in addition to or instead of their own, and can thus contribute to gene flux between species, strains or individual cells. Temperate phages may insert themselves in the bacterial
chromosome rather than killing the cell, thus hitchhiking (and possibly increasing in number) along with the bacteria, a process known as lysogeny. The inserted phage is known as a prophage and many of the sequenced genomes contain one or several prophages, and even more remains of degraded prophages. Prophages may contain genes beneficial to the host, picked up during some earlier infection, and possibly kept as a means for the phage to be kept by the bacterial host. Virulence genes of pathogenic bacteria are sometimes encoded by prophages. When resources get scarce, the prophage is excised from the genome, replicated and packaged into phage particles, which are released to search novel hosts as the original host is lysed. When several phages and/or prophages inhabit the same host, their genomes may recombine, accounting for the mosaic nature of phage genomes and the difficulty in establishing evolutionary relationships between phages. Phages are believed to be the most abundant organisms on earth, with an estimated number of $10^{31}$, which is ten times more than their bacterial hosts. They play important roles both in bacterial evolution and in metabolic cycling (by killing the bacteria and releasing their metabolites), particularly in the oceans, where each milliliter contains millions of bacteria (Breitbart and Rohwer 2005).

Genomic islands are stretches of the bacterial chromosome that can be distinguished from the core genome by for example different G+C content, suggesting that they have been horizontally transferred rather than vertically inherited. Genomic islands often encode fitness factors that enhance bacterial survival in certain environments (which could be a mammalian host – since the first sequenced genomes were human pathogens, genomic islands were first known as pathogenicity islands), and like prophages they are often inserted at a bacterial tRNA-gene (Dobrindt et al. 2004). They also commonly contain phage- and plasmid-related genes for transfer and integration.

Genome structures

Most bacteria have a single circular chromosome, which replicates bi-directionally from a single origin. To prove this general rule, there are however exceptions which have several and/or linear chromosomes, especially among the alpha-proteobacteria. The terminus of replication is generally located approximately at the opposite site of the chromosome circle, making the two replicohores equally sized (Achaz et al. 2003).

Gene order in bacteria is in general not conserved except at short time-scales (Mushegian and Koonin 1996; Suyama and Bork 2001) and short distances (genes in operons). However, highly expressed genes involved in transcription or translation tend to be located closer to the origin of replication – at least in bacteria with short generation times and hence multiple copies of genes near the replication origin (Rocha and Danchin 2003a).
Although the order of genes is not conserved, the chromosome sequence typically displays a characteristic pattern (GC-skew) with an overabundance of the nucleotide G relative to C on the leading strand (Frank and Lobry 1999; Lobry 1996). This bias is believed to be caused by different mutational biases on the leading and lagging strands; since the lagging strand is synthesized discontinuously, it is single-stranded for longer time and thus more vulnerable to deamination of C to T. The orientation of this skew can be used to map the origin and terminus of replication, and its presence in most sequenced genomes indicates that the genomes are relatively stable at least on a short time scale and that mutations quickly restore the base composition pattern if it is disrupted by rearrangements, or that such rearrangements decrease fitness and thus are counter-selected.

Many bacterial genomes also display gradients in the abundance of various short sequence patterns across the replication axis. These may be involved in determining the location of the terminus of replication and in segregation of the daughter chromosomes (Hendrickson and Lawrence 2006; Lobry and Louarn 2003).

Genes, especially those that are essential, are preferentially located on the leading strand, probably because collisions between the replication and transcription polymerases would otherwise result in truncated transcripts, which decreases gene expression and may even be toxic (Rocha and Danchin 2003a; Rocha and Danchin 2003b).

Genome evolution in bacteria
Bacterial genomes have been shown to be constantly changing. The short generation time of bacteria (compared to multicellular organisms) enable rapid sequence evolution by mutations, and despite their lack of sex they have several ways of exchanging genes with each other, even between distantly related species. In addition, the order of genes on the chromosome can be scrambled, potentially generating new genes, combinations of genes or regulatory mechanisms. In the constant fight for survival in a world of limited resources, novelty can at times be crucial – although the majority of changes are detrimental or neutral. For pathogenic or otherwise host-associated bacteria, genomic change may also be essential in order to avoid recognition and elimination by the immune system of the host.

Rearrangements – genome structure evolution

Genome rearrangements – inversions of part of the chromosome or translocation to another chromosomal location - can be caused by homologous recombination between similar, repeated sequences, via the enzyme RecA.
Closely located repeated sequences can also contribute to rearrangements by illegitimate recombination, without RecA activity (Rocha 2003a).

Comparing the gene order of pairs of closely related species, an X-shaped pattern is often observed, indicating that inversions or translocations that are symmetric across the origin and/or terminus of replication are more common than those within a replichore (Eisen et al. 2000; Mira et al. 2002; Suyama and Bork 2001; Tillier and Collins 2000). The reasons for the preference for symmetric inversions are however unclear. They could be caused by an interaction between the recombination and replication machineries, with strand switching between the two replication forks, which are equidistant from the origin of replication (Tillier and Collins 2000); i.e. a mutational bias. The same pattern could however also be explained by selection against unbalanced chromosomes (with replichores of different sizes), and against inversions within the same replichore, which move genes from the leading to the lagging strand and also change the base composition pattern. Decrease in fitness after such inversions has been implied by some experimental studies (Campo et al. 2004; Mackiewicz et al. 2001; Rocha 2006). Selection against rearrangement is also apparent from a comparison of *Escherichia coli* and *Salmonella* (Rocha 2006).

The extent of genome rearrangements between pairs of bacteria is correlated with the number of repeats (Rocha 2003b), supporting the hypothesis that rearrangements are often caused by homologous recombination. However, inversions across the terminus appear not to involve homologous recombination between repeated sequences (Tillier and Collins 2000).

A consequence of genome rearrangements could be generation of novel genes (if they occur within coding regions) or changes in the expression of genes. For host-associated bacteria changes in gene sequence or expression (antigenic variation) might be vital to avoid elimination by the immune system. Bacteria that infect eukaryotic hosts seem to undergo genome rearrangements more frequently than free-living bacteria (Hacker et al. 2003; Hughes 2000). This conclusion has however been questioned by E. Rocha (Rocha 2006), who found that pathogens have average conservation of gene order and suggested that the increased rearrangement rate in certain species might be related to recent change of ecological niche (i.e. switch to pathogenesis) rather than with the pathogenic lifestyle *per se*.

Rearrangements between similar prophage sequences also contribute to the generation of mosaic, chimeric phages.

Genome architecture is also used in strain typing by Pulsed-Field Gel Electrophoresis (PFGE), where the genome is cleaved by restriction enzymes and the fragments separated and visualized by gel electrophoresis. Similarities in the fragment pattern suggest similar genome structure and thus close evolutionary relationship.
Rearrangements appear to be more common in laboratory than natural strains (Achaz et al. 2003; Rocha 2006), suggesting that they are counter-selected under natural conditions.

Insertions, deletions and duplications – gene content evolution

Variation in genome content can be caused by insertions, deletions and duplications of genetic material. Deletions and duplications are often mediated by homologous recombination or slipped-strand mispairing during replication between directly repeated sequences.

Deletions are more common than duplications, explaining why genomes are not infinitely large. Sequences that don’t contribute important functions thus tend to be inactivated and subsequently degraded and lost (Andersson and Andersson 1999; Mira et al. 2001).

Deletional bias also explains why the genomes of obligate intracellular bacteria tend to be small, as access to novel sequences is limited in the intracellular niche. The small population size of intracellular bacteria also contributes to fixation of deleterious mutations (including deletions), a process known as Muller’s ratchet (Andersson and Hughes 1996; Andersson and Kurland 1998; Muller 1964). Laboratory experiments have shown that genome reduction can happen quickly (Nilsson et al. 2005).

Tandem duplications are prone to deletion via recombination of the newly generated repeat, and are rare in sequenced genomes (Achaz et al. 2003) and account for few differences in genome composition (Lawrence and Hendrickson 2005). Duplications are however believed to play an important role in the evolution of novel genes, since one of the copies will be relieved from negative selection to perform the original function (Hooper and Berg 2003; Ohta 1987; Ohta 1989).

Insertion of novel genetic material is common among bacteria and can be achieved in several ways. Horizontal gene transfer (HGT) can be mediated by phages, which incidentally package and spread bacterial DNA instead of or in addition to their own genome, a process called transduction. Intact phages are also commonly inserted into the bacterial chromosome. Some bacteria are also capable of taking up free DNA from the environment (transformation), or of actively transferring and taking up DNA from another cell (conjugation). The extent to which these processes take place varies between different bacterial species, and influences the spread of genes (e.g. for antibiotic resistance) in the population, and the structure of the population. In species that frequently exchange and replace genes – i.e. recombine – the evolutionary linkage between genes on the chromosome is broken such that different genes will have different evolutionary histories.

Since it is difficult to distinguish between insertion and deletion events when comparing different bacteria (has the sequence been gained in one or
lost from the other?), insertion and deletion events are often collectively called indels.

Point mutations – sequence evolution
Genome rearrangements and horizontal gene transfer can alter the chromosome on a large scale. Meanwhile, the individual gene sequences evolve by the accumulation of point mutations, where a single base pair is substituted, inserted or deleted, due to errors during replication or to DNA damage. Provided that the rate of recombination between strains or species is sufficiently low, differences in sequences can be used to trace the evolutionary history of organisms by constructing phylogenetic trees.

The genus Bartonella
*Bartonella* are gram-negative bacteria, belonging to the order alpha-proteobacteria. Like many other members of this order, such as the close relative *Brucella*, *Bartonella* live in close association with eukaryotic hosts. 19 species of *Bartonella* have as yet been identified, and are associated with a wide span of mammalian hosts, such as mice, cows, deer, cats and humans. At least seven species have been associated with human disease (Rolain et al. 2004), and several of these are considered emerging or re-emerging pathogens, i.e. the diseases have appeared recently and/or are rapidly increasing in incidence. The bacteria are transferred between hosts by blood-sucking arthropod vectors such as fleas, lice and ticks. Some species of *Bartonella* are specific to a certain host, but host-specificity is not absolute and a single host species may often be infected by several *Bartonella* species, and vice versa (Houpikian and Raoult 2001). There is some correlation between host specificity of different *Bartonella* species and their phylogenetic (evolutionary) relationships, although several host switches must have occurred (Houpikian and Raoult 2001).

In the natural host, *Bartonella* first infects a primary niche, believed to be the vascular endothelial cells. From there, they are periodically seeded into the blood stream, where they invade the erythrocytes (red blood cells), a niche that is shielded from the immune system and which enables transmission to another host by blood-sucking arthropods. The ability to invade erythrocytes – and to multiply and persist throughout the life of the infected cell – is unique to *Bartonella* (Jacomo et al. 2002; Schulein et al. 2001; Seubert et al. 2002). Also unique to *Bartonella* is their ability to form tumors consisting of proliferating infected endothelial cells, and induce angiogenesis (i.e. the creation of new blood vessels) to support them.

Infection of the natural host is normally asymptomatic (with exception for the human specialists *B. quintana* and *B. bacilliformis*), allowing for long-
term persistence and increased transmission of the bacterium and indicating that the bacterium has adapted to its natural host. In contrast, incidental infection of other host species frequently results in a wide variety of disease symptoms, the severity of which depend on the immune status of the host, and the bacteria are usually unable to invade the erythrocytes.

Figure 1. Phylogeny of the *Bartonella* species, based on maximum likelihood-analysis of six concatenated genes (courtesy of Carolin Frank), with animal host indicated.

*Bartonella henselae*

**Ecology and prevalence**

*Bartonella henselae*, the main focus of this work, was first isolated from the blood of an HIV-infected man with fever in 1990 (Regnery et al. 1992). It is naturally associated with cats, where it infects the erythrocytes and normally
does not cause any disease symptoms, although rare cases have been reported (Shaw et al. 2001). The bacterium is transmitted between cats by the cat flea, *Ctenocephalides felis* (Chomel et al. 1996), by intradermal inoculation of infected flea feces (figure 2). The bacterium multiplies in the digestive system of the flea, and survives in the flea feces for several days (Finkelstein et al. 2002). *C. felis* is the most common flea species on cats, and has a worldwide distribution. It is however not restricted to cats but has been detected on more than 50 mammalian species (Rust and Dryden 1997; Szabo et al. 2000).

*B. henselae* DNA has also been detected in ticks (*Ixodes pacificus* and *I. ricinus*) (Chang et al. 2001; Holden et al. 2006; Sanogo et al. 2003), but it is as yet unknown whether ticks can act as vectors and spread the bacteria (Stricker et al. 2006). However, there have been cases of human infection not associated with cat contact, suggesting other infection routes. *B. henselae* and other *Bartonella* species have also been implied in sudden cardiac death among Swedish orienteers, which are exposed to insect bites (Wesslen et al. 2001).

The prevalence of *B. henselae* in cats varies widely between different geographic regions, from 0 % in cold areas such as Norway up to 93 % in North Carolina, US (Boulouis et al. 2005). Prevalence can be measured either by seropositivity, antibodies that show current or past infection, or by bacteremia, culture of the bacterium from blood, and the techniques yield different results. Bacteremic animals or humans do not always have antibodies. In the US, prevalence of *B. henselae* was found to increase with increasing climatic warmth and annual precipitation, ranging from 4 to 55 % in different regions (Jameson et al. 1995). The low prevalence of *B. henselae* in regions with a cold climate is probably due to the low prevalence of the cat flea. In Sweden, only 1-2 % of domestic cats have antibodies against *B. henselae*, and the infection is higher in the southern part of the country (Boulouis et al. 2005; Hjelm et al. 2002). In Denmark, 23 % of cats were bacteremic, while 46 % had antibodies indicating previous or ongoing infection (Chomel et al. 2002). In most studies, stray cats are more often infected than domestic cats, and outdoor cats more often than indoor cats.

*B. henselae* has also been found in wild felids, such as cheetahs, panthers and lions (Kelly et al. 1998; Molia et al. 2004; Pretorius et al. 2004; Rotstein et al. 2000; Yamamoto et al. 1998). It has also been detected in dogs (which can also be infected by the cat flea) and in field mouse (Engbaek and Lawson 2004). Mice can also be experimentally infected with *B. henselae*.

**Human pathogenicity**

Humans can be infected by *B. henselae* by a scratch or bite from an infected cat - probably by infected flea feces on the claws or teeth of the cat - and possibly also by direct transmission from insect vectors. Infection can give rise to a range of different disease symptoms, depending on immune status
of the infected person (Resto-Ruiz et al. 2003). Judging from seropositivity screens (i.e. presence of antibodies), many infections are un-noticed.

Figure 2. Life cycle of *Bartonella henselae*. The vector (the cat flea *C. felis*) transmits the bacteria between cats, which are the natural host. Humans are incidentally infected by a scratch from a cat, or possibly directly from an insect bite.

In immuno-competent persons, the bacterium cannot infect the erythrocytes but instead targets the endothelial cells. The most common disease caused by *B. henselae* is cat scratch disease (CSD). CSD is characterized by swollen lymph nodes and mild fever. It is usually self-eliminated within a few months, but can be cured by antibiotics. In the US, more than 22,000 cases of CSD are estimated to have occurred in 1992, giving an incidence rate of 9.3 per 100,000 per year (Jackson et al. 1993).

In immuno-compromised people, such as AIDS patients, *B. henselae* infections can cause more severe symptoms. Most common is bacillary angiomatosis, in which the bacterium induces the formation of tumor-like lesions of the skin or other organs, thus creating their own habitat. Bacillary peliosis and endocarditis are also common in immuno-compromised patients.

*B. henselae* may also cause a wide range of less common symptoms, including arthropathy (joint disease) and encephalopathy (brain disease), which may possibly be due to the ammonia produced by the amino acid catabolism which the bacterium uses instead of glycolysis to derive energy (Stricker et al. 2006).
Molecular mechanisms of infection

Research on the molecular mechanisms of *B. henselae* infection has revealed several sets of genes involved in adhesion and entry into endothelial cells and/or erythrocytes, prevention of phagocytosis (ingestion by immune defense cells) or host-cell apoptosis (suicide) and stimulation of angiogenesis.

The protein BadA (*Bartonella* adhesin A) is an adhesin that prevents phagocytosis, enables attachment and stimulates host cell expression of the vascular endothelial growth factor, a protein plays a key role in endothelial cell proliferation and angiogenesis (Kempf et al. 2001; Riess et al. 2004). The protein forms pilus-like structures extruding from bacterial cell, which have been shown to be highly immunogenic and also to undergo phase variation (Kyme et al. 2003) – possibly to escape immune recognition after the initial function has been performed. The gene encoding this protein is extremely long (ca 10 kb in *B. henselae*) and highly repetitive, and is present in multiple, different copies in the sequenced *Bartonella* genomes. The homologues in *B. quintana* are called Vomps (variably expressed outer membrane proteins), and the different proteins are expressed at different stages of infection, associated with rearrangements facilitated by repeats (Zhang et al. 2004). However, the *B. quintana* Vomps did not mediate adherence to host cells but only induced vascular endothelial growth factor production (Schulte et al. 2006).

Two different type IV secretion systems, which export proteins or DNA to the eukaryotic host, have been identified in *B. henselae*; the trw operon (Seubert et al. 2003) and the virB operon (Schmid et al. 2004; Schulein and Dehio 2002). The Trw proteins are required for invasion of erythrocytes, although the precise mechanism is unknown, as is the translocated molecule, and they may be involved in attachment rather than secretion (Seubert et al. 2003). The VirB proteins are required for invasion of the primary niche, i.e. the endothelial cells, and for their manipulation in terms of inhibition of apoptosis, cytoskeletal rearrangements and inflammation. These effects are mediated by the Bep proteins, which are secreted through the VirB complex (Schmid et al. 2006; Schulein et al. 2005). Both the trw and the virB genes originate from plasmid conjugation systems, and have been horizontally transferred into *Bartonella* (Frank et al. 2005).

Diversity and typing

Based on sequences of the gene coding for 16S ribosomal RNA, isolates of *B. henselae* can be divided into two types (Bergmans et al. 1996), which have also been found to correspond to different serotypes (La Scola et al. 2002), i.e. elicit different antibodies in infected animals. It has been suggested that type I might be over-represented in human samples and thus more apt to infect humans, although results from different studies are contradictive and frequently based on small samples. Other genes have also
been used for typing, leading to a set of 9 genes used in a multi-locus sequence typing scheme, separating different isolates into 7 sequence types (Iredell et al. 2003).

**Bartonella quintana**

*B. quintana* is a close relative to *B. henselae* but has humans as its natural host, and is transmitted by the human body louse, *Pediculus humanus corporis*.

*B. quintana* often causes chronic bacteremia with no or few disease symptoms, but may also cause severe disease. During World War I and II, millions of soldiers were affected by trench fever (also called quintan fever), which recurs with intervals of approximately five days, corresponding to the periodic seeding of the bacterium from the primary niche. After louse control was introduced, the disease was rare, but is now re-emerging among homeless people with poor hygienic conditions, causing “urban trench fever” (Brouqui et al. 2005).

Like *B. henselae*, *B. quintana* can cause severe symptoms such as bacillary angiomatosis and endocarditis in immuno-compromised people.

*B. quintana* DNA has also been detected in cat fleas (Rolain et al. 2003b), in rodent fleas (Marie et al. 2006), in a human flea taken from a pet monkey (Rolain et al. 2005), as well as in a captive-bred *Cynomolgus* monkey (O’Rourke et al. 2005) and in the dental pulp of a cat (La et al. 2005). Humans however seem to be the main reservoir and natural host of *B. quintana*, although it is likely that cats were the original host and that the bacterium switched host in connection with the domestication of cats (Frank et al. 2006; Lindroos et al. 2005) and with the use of clothing, which induced the emergence of the human body louse approximately 70,000 years ago (Alsmark et al. 2004; Kittler et al. 2003).

**Bartonella koehlerae**

*Bartonella koehlerae* is a recently discovered member of the genus *Bartonella*. Phylogenetically, it is closely related to *B. henselae* and *B. quintana* (Houpikian and Raoult 2001). Like *B. henselae*, it seems to have the cat as the natural host (Avidor et al. 2004; Droz et al. 1999; Rolain et al. 2003a) and it has been detected in cat fleas (Rolain et al. 2003b), but also in rodent fleas (Marie et al. 2006). It has been associated with one case of human endocarditis (Avidor et al. 2004).

Since *B. koehlerae* does not grow under conditions ordinarily used to isolate *Bartonella* from blood, its prevalence may be highly underestimated.
**Bartonella grahamii**

*B. grahamii* is a common inhabitant of various small rodent species, such as rats, mice and voles (Bown et al. 2004; Engbaek and Lawson 2004; Holmberg et al. 2003; Knap et al. 2007; Mediannikov et al. 2005). Thus far, two cases of human disease (ocular neuroretinitis) have been associated with *B. grahamii* (Kerkhoff et al. 1999; Serratrice et al. 2003). It is evolutionarily quite distantly related to the other three species used in this study (*B. henselae*, *B. quintana* and *B. koehlerae*).

**Bartonella bacilliformis**

*B. bacilliformis* was one of the first *Bartonella* species to be recognized, due to its high pathogenicity and mortality in humans, where it causes Carrións disease (with a acute phase known as Oroya fever, said to have killed 7,000 out of 10,000 railroad builders in Peru in 1871, and a chronic phase known as Verruga peruana). Like *B. quintana* it has humans as its only and natural host, and is transmitted by the sand fly *Lutzomyia verrucarum* (Ihler 1996). However, it is phylogenetically distantly related to *B. quintana* and the other *Bartonella* species here discussed. It also differs from other *Bartonella* species in that it frequently causes haemolysis, i.e. destruction of the erythrocytes (up to 90 % of the red blood cells are lysed) and subsequent anaemia. It is geographically limited to a region in the Andes in South America, probably due to the limited distribution of the vector. In the endemic area, 5-15 % of the population are bacteremic, many of which remain asymptomatic (Weinman 1944 and Ricketts 1947, as cited on [http://www.interamericaninstitute.org/bartonellosis.htm](http://www.interamericaninstitute.org/bartonellosis.htm)), and 45 % have antibodies to *B. bacilliformis* (Chamberlin et al. 2000).

The **Bartonella genomes**

The genomes of *B. henselae*, *B. quintana* (Alsmark et al. 2004) and *B. bacilliformis* (The Institute for Genomic Research, unpublished) have been sequenced, and the *B. grahamii* sequence is in the last stage of completion (Frank et al., unpublished).

The 1.58 Mb genome of *B. quintana*, the human specialist, was found to be a subset of the larger 1.93 Mb genome of *B. henselae*, which has a broader host distribution. The difference lies predominantly in the presence of a prophage and three genomic islands in the genome of *B. henselae*. These elements have probably been horizontally transferred, since they have an atypical nucleotide distribution and are flanked by tRNA genes, which are frequently used as targets for phages. Since *B. quintana* contains some remains of these sequences and the corresponding sites are associated with rearrangements between the genomes, the transfer probably took place be-
fore the divergence of *B. henselae* and *B. quintana*, and the sequences were subsequently deleted in *B. quintana*. Both genomes are balanced, i.e. the two replicores are of approximately equal size.

Like many prophages, the 57 kb prophage of *B. henselae* appears to be of mosaic origin and contains genes with homologues in different phages, for examples phages of entrobacteria and from the insect endosymbiont *Wolbachia*. The genomic islands contain genes believed to be involved in infection and virulence, such as filamentous hemagglutinin and the associated hemolysin activation protein, interspersed with phage-related genes and *Bartonella*-specific genes of unknown function. The genomic islands and the prophage contain highly repeated sequences, some of which are present in several regions, suggesting recombination between the islands and prophage.

Both genomes also contain a large region with low coding content and many repeated sequences. Many of the genes in this region are specific to *Bartonella*, and of the rest many show homology to the second replicon (ChrII) of *Brucella suis*, a closely related species which harbors two chromosomes, whereas the rest of the *Bartonella* genomes is similar to ChrI of *B. suis*. This region (spanning 282 kb in *B. henselae* and 200 kb in *B. quintana*) is therefore believed to be the remains of a second chromosome that has been integrated into the main chromosome, and is thus called the ChrII-like region.

Recently, the genome of *B. bacilliformis*, a species distantly related to *B. henselae* and *B. quintana*, was completed by The Institute for Genomic Research (TIGR) and submitted to the public database of the National Center for Biotechnology Information (NCBI). Like the other human specialist pathogen *B. quintana*, *B. bacilliformis* has a small genome of 1.45 Mb, and lacks the prophage and genomic islands that are present in *B. henselae*.

The genome of *B. grahamii* is currently being sequenced by the department of Molecular Evolution at Uppsala University. It has a larger genome than the other sequenced *Bartonella* genomes, with a genome size over 2.3 Mb, containing virtually all of the genes present in *B. henselae* and *B. quintana*, with extra copies of some (e.g. the prophage) plus additional unique sequences. In contrast to the other studied *Bartonella* species, *B. grahamii* also harbors a plasmid, containing some of the genes that are also present on the main chromosome. The analysis of the *B. grahamii* genome has also led to the identification of previously undetected genomic islands and degraded islands in *B. henselae* and *B. quintana*.

As expected for host-associated bacteria, the sizes (and G+C content of approximately 38 %) of the *Bartonella* genomes are far below average.

**The Bartonella phage(s)**

Several studies have shown the production of phage particles in cultures of *B. henselae* (Anderson et al. 1994; Chenoweth et al. 2004), as well as in *B.
bacilliformis (Barbian and Minnick 2000; Umemori et al. 1992) and B. vinsonii (Maggi and Breitschwerdt 2005). The particles have round to icosahedral shape and contain double-stranded, linear DNA fragments of 14 kb. The capsid seems to be composed by three different proteins, which have similar sizes in B. henselae and B. bacilliformis. However, the nature of the packaged DNA is unclear and appears to be heterogenic, possibly representing random samples of the bacterial chromosome (Anderson et al. 1994; Barbian and Minnick 2000). The heterogenous or random DNA content suggests that the phage is defective (i.e. unable to propagate itself), and might instead be involved in horizontal gene transfer.

Defective phages that package host DNA have previously been identified in a few other bacterial species, for example Bacillus subtilis (Wood et al. 1990), Bacillus natto (Tsutsumi et al. 1990), Rhodobacter capsulatus (Lang and Beatty 2000), Brachyspira hyodysenteriae (Humphrey et al. 1997) and Desulfovibrio desulfuricans (Rapp and Wall 1987). However, it is not yet clear if the phenomenon is an accident caused by a defective prophage unable to package its own genome, or if it has some adaptive value for the bacterial community by increasing the horizontal gene transfer during difficult times.
2. Aims

The aims of this thesis were to study the diversity and evolution of bacteria on a genomic scale, using the genus *Bartonella* as a model system.

In particular, I wanted to study the frequency of different mechanisms of evolution – single nucleotide mutations, deletions and duplications, and genome rearrangements – within the species *B. henselae*. It was also of interest to find out which genes were dispensable (i.e. could be deleted). Another question was whether specific genes are required for human infection, or whether any strain can cause human infection and disease.

The question of host-specificity was also addressed by studying the genome content of *B. koehlerae*, which is closely related to *B. henselae* and has a similar host range, but a genome size closer to the human specialist *B. quintana*.

The sequence diversity within *B. henselae* was also compared to that within *B. quintana*.

Since microarrays were a central technology in the study of genome content diversity, I first verified the accuracy of microarrays for comparative genomics.

Microarrays were also used to study the DNA content of the *B. henselae* and *B. grahamii* bacteriophage particles, a subject that has been under some debate. To identify genes involved in the formation of phage particles, *B. grahamii* gene expression was monitored after different culture times.

Gene expression was also studied in *B. henselae* during intracellular infection of human cells, to detect genes involved in infection and their regulation.
3. Methodology

Bacterial culture

*Bartonella* are fastidious, slow-growing bacteria that are dependent on heme and CO$_2$ for growth. The bacteria were cultured on blood or chocolate agar plates at 35 °C with 5 % CO$_2$ for 5-15 days.

Multi Locus Sequence Typing

Several methods have been developed in order to subtype bacterial isolates from a given species more closely. This is of interest both for epidemiology (tracking the spread of diseases) and for studying bacterial evolution. If recombination between organisms is rare or nonexistent, and if mutations are infrequent enough to rule out homoplasies (the same mutation happening independently in several lineages), it would be enough to study a single character, such as the sequence of a specific gene or the properties of a single protein. Combining several characters gives better resolution, and also information on the extent of recombination in the population.

Multi-locus sequence typing (MLST) groups bacterial isolates into different sequence types and clonal complexes, based on partial sequences of 7-9 genes (Cooper and Feil 2004; Maiden 2006; Maiden et al. 1998; Urwin and Maiden 2003). Housekeeping genes (i.e. genes involved in basic metabolism) are usually used for MLST typing, although for highly similar bacteria a larger number of or more variable genes may be required. The advantage of using housekeeping genes is that those can be assumed to be under negative rather than positive selection, so that the risk of homoplasies is minimized. They can also be expected to be universally present, and be less prone to horizontal gene transfer. Approximately 400-700 nucleotides are sequenced from each gene – the sequence length that can be obtained from a single read on current sequencing machines. For each gene, different allele versions are given different numbers (in order of their discovery), and the combination of different allele types gives the sequence type (ST) of a given isolate. Sequence types sharing most alleles can be organized into clonal complexes. Recombination in the population can be detected from allelic profiles which would require that the same mutations (allele type) has developed more than once, which is considered unlikely in genes that are not under positive selection (i.e. selection for change).
Iredell *et al.* described a MLST typing scheme for *B. henselae* (Iredell *et al.* 2003), and we used the same primers in our study.

**Multi Spacer Typing**

Since closely related isolates display few differences in coding sequences, MLST has been extended to intergenic sequences, spacers (Fournier *et al.* 2004). Spacers are assumed to be under no selection and thus accumulate more mutations. Primers are preferably designed to be located within the flanking genes, which can be assumed to be present and more conserved in sequence in various isolates.

**Microarray Comparative Genomic Hybridizations**

In order to study the genome content of different isolates without performing whole genome sequencing (which is still expensive and time-consuming), comparative genomic hybridizations (CGH) to microarrays were performed. A microarray is a glass chip with spots of DNA, each corresponding to different genes. By labeling genomic DNA from the test isolate and a reference with different fluorescent dyes and hybridizing them simultaneously to the microarray, the absence or presence of a specific sequence in the test isolate can be deduced from the relative fluorescent intensities of the two dyes in a certain spot (figure 3). The intensity of the test strain must be compared to that of the reference strain for each spot, since the absolute fluorescent intensity depends not only on the amount of hybridized DNA, but also on factors such as the concentration of the probe in the spot and the length and hybridization strength of the probe.

A drawback with microarray technology for comparative genomics is that only the status of the genes spotted on the slide can be tested; the tested strains may harbor additional genes, which will not be detected. Also, absence of a gene may be confounded by sequence divergence, which also results in reduced hybridization. For genes present in multiple copies in the reference genome, interpretation is more difficult since a lower signal in the test strain may be caused either by its complete absence or by a lower copy number. Conversely, absence of a gene may not be detected due to cross-hybridization with similar sequences. For single-copy genes in closely related species, however, microarrays offer a quick and accurate view of their distribution.
Microarray design and construction

A *B. henselae* microarray was constructed with 1650 PCR-products as probes. For *B. grahamii*, a microarray with 4438 60-mer oligo-nucleotides was constructed.

Microarray CGH hybridizations

For each studied *B. henselae* strain, several separate experiments (technical replicates) were made with the same genomic DNA. Genomic DNA from the studied strain was labeled with one of the fluorescent dyes Cy3 or Cy5, and genomic DNA from the reference strain was labeled with the other dye. The labeled DNA was then mixed and applied to the microarray, allowing it to bind to the complementary sequences (spots). Using a laser scanner, the fluorescent intensity for each dye was measured in pixels of 10x10 μm over the microarray. The image is then analyzed and spots detected. This step often requires manual inspection and correction. For each fluorescent dye, the average intensity and its standard deviation are computed for each spot, and for the local background of the spot.

Data analysis and visualization

Before deciding on the status (absent, present or amplified) of a gene in the test strain, the raw data has to be corrected (normalized) to avoid bias resulting from differences in concentrations and labeling efficiencies of the samples, and in detection sensitivity and saturation of the fluorescent dyes. Normalization for different amounts of material or different labeling efficiencies is often done by global total-intensity normalization, i.e. the signals are scaled to give the same total or average signal for both dyes. Sometimes, there is also intensity-dependent bias, i.e. the dye bias is different for spots with high or low average intensities. This can be caused by saturation effects or by a non-linear relationship between the amount of fluorescent molecules and the measured signal (quenching). Intensity-dependent dye bias can be
detected by plotting the 2-logarithm of the ratio of the intensities (M-value) against the 2-logarithm of the average of the intensities (A-value) for every spot, and is corrected for by fitting a smooth line (lowess or loess) to the data and making that the new baseline.

For the *B. quintana* and *B. koehlerae* hybridizations, the common method of lowess normalization of M vs A was found to be flawed by the large number of absent genes causing natural correlation between the two; absent genes have both low M-values and low A-values. Instead, intensity-dependent dye bias was investigated and corrected for by comparing M and the 2-logarithm of the reference intensity. Also, the large number of absent genes reduced the average M-value significantly as compared to the M-value of present genes. Therefore, global intensity normalization was made to the peak of the distribution of all M-values, rather than the average.

After normalization, median M-values for all replicate spots were computed and used to decide the status of the probe. For the *B. quintana* hybridizations, a cutoff of -2 (corresponding to a 4-fold difference in intensity) was found to suit the sequence similarity data. For the *B. koehlerae* hybridizations, the threshold for absence had to be adjusted since the average sequence similarity to the reference species *B. henselae* is much higher for *B. koehlerae* than for *B. quintana*. Since the average probes have M=0, highly conserved probes (with > 96 % sequence similarity between *B. quintana* and *B. henselae*) had a median M-value of 1.12 in the *B. quintana* CGH experiments, while the same probes had a median M-value of 0.41 in *B. koehlerae*. Thus, the -2.0 cutoff used for *B. quintana* corresponds to -2.7 for *B. koehlerae*. For the *B. henselae* hybridizations, the cutoff for absence was set to -2, since there was less risk for confusion of deleted or diverged genes.

Normalization and analysis was done in the statistics language R with my own scripts.

To visualize and explore the results, I wrote several interactive programs in Perl/Tk that display the results in the context of the sequenced *B. henselae* and/or *B. quintana* genome, on a linear or circular scale (figure 4).

Gene-content based phylogeny

The gene content data generated by CGH was used to cluster the strains on the basis of gene content. Neighboring probes were sorted into contiguous regions displaying the same pattern of absence or presence. The absence (0) or presence (1) of the regions was then used to create a maximum parsimony tree, which was compared to a phylogenetic tree based on the concatenated MLST sequences.
Microarray Gene Expression Analysis

Microarrays were also used to monitor gene expression after different growth times, on agar plates or during intracellular infection. RNA from each time point was compared to that from bacteria grown on plates for a certain time. After reverse transcription of RNA to DNA, the procedure is similar to the CGH analysis.

Pulsed Field Gel Electrophoresis and Southern Blotting

Experimental procedure

Restriction enzyme length polymorphism (RFLP) followed by pulsed field gel electrophoresis (PFGE) is a common method for estimating genome sizes and for strain typing. The chromosome is first cleaved by restriction enzymes, which recognize and cut the DNA at specific sequences, and the resulting fragments are then separated by size by gel electrophoresis (circular chromosomes do not move proportionally to their size). Since the fragments are large, they must be forced through the gel by switching the polarity of the electric field in pulses. By comparing migration distance of the separated fragments to that of known standards, the sizes of the fragments can be estimated. If there are no double bands, i.e. multiple fragments of similar size, and if the chromosome was completely digested so that no un-
cleaved fragments remain, the total genome size can be computed as the sum of all fragments. When studying different isolates of the same species, the fragment sizes are often found to differ, due to rearrangements and indels. Differences in fragment patterns are often used in epidemiology to study the relatedness of different isolates.

We initially did RFLP-PFGE with two different enzymes (NotI and AscI) in order to study the genome sizes of the isolates, to see if they had additional sequences absent from the sequenced strain and hence not detected by CGH. A previous study had suggested that genome size varied a lot between isolates of *B. henselae*. When the banding pattern was found to differ a lot, we wanted to investigate the underlying rearrangement events by hybridizing different probes to blots of the PFGE gels (figure 5). It turned out to be trickier than expected due to the complexity of the rearrangement events, and to noise and uncertainties in the data. In the end, we hybridized 24 different probes, and two additional restrictions; Sgfl and a double digest combination of NotI and AscI.

**Figure 5.** PFGE gel (left) with several isolates cleaved by different restriction enzymes, and hybridization of a gene-specific probe by Southern blotting (right).

**Data analysis – hypothesis testing**

Due to the complex banding patterns, the loss and gain of restriction sites by mutation (in addition to rearrangements), presence of double bands (i.e. two or more fragments of approximately the same size, appearing as a single band on the gel), heterogeneity within the population of bacterial cells such that some had and some lacked a certain restriction site, and possibly rearrangements occurring during cultivation, it was difficult to determine the rearrangement events from the PFGE band patterns precisely.

In order to visualize the results of the different PFGE and Southern blotting experiments along with the genome data, and help to create and test different hypotheses about the rearrangement events and genome structures, I wrote an interactive program (XPulSee) in Perl/Tk, where the expected
PFGE patterns and probe hybridizations after hypothetical rearrangement events could be compared to those observed (figure 6).

The rearrangement events chosen were those that required the smallest number of events to produce PFGE and Southern blot profiles similar to the observed ones.

Figure 6. Screenshot of the XPulSee program for visualization of PFGE and Southern blot data, and inference of genome rearrangements. One inversion and one new restriction site are inferred to account for the observed data for the isolate CA-1. The expected profiles after rearrangements are labelled “Exp”, the observed profiles “Obs”, and those expected for the sequenced reference strain “Ref”.

31
4. Results

In this work, I have studied the genome content of B. koehlerae and of different B. henselae isolates by comparative genomic hybridizations to microarrays. I have also studied the diversity of genome sizes and structures and of gene sequences within B. henselae, and of intergenic sequences in B. quintana. Finally, I have studied the content of the Bartonella phage particles and growth time-related gene expression, and analyzed gene expression during intracellular growth.

Validity of the microarray CGH results (Paper I)

To test the correctness of the microarray for CGH experiments, we hybridized genomic DNA from the closely related species B. quintana, the genome of which has been sequenced (Alsmark et al. 2004), against that of B. henselae. The B. quintana genome is smaller than that of B. henselae, and the genes that are shared between the two species are on average 87% identical. The sequence similarity between each probe on the microarray and the most similar sequence of the B. quintana genome was determined. For genes that are absent in B. quintana, the sequence similarity ranged from 35% to 78%, and we decided to use 75% as the limit for absence or presence of a sequence.

The CGH experiments produce a so-called M-value for each spot, which is the 2-logarithm of the ratio of the signal from B. quintana relative to that of B. henselae. The M-values were adjusted to be centered around 0 for the majority of probes. Thus, probes corresponding to genes with average (ca 87%) sequence similarity to B. henselae will have M-values close to 0, whereas genes with higher sequence similarity will have M-values above 0. Probes that have lower sequence similarity or are absent from B. quintana will have negative M-values. Some probes are present in several identical or similar copies in B. henselae. For those, it can be difficult to judge whether they are absent or present in fewer copies in another strain/species. Thus, we tested the accuracy of the CGH results separately for single-copy probes and multi-copy probes.

Of the 1428 single-copy probes, 1147 were classified as present in B. quintana based on the sequence similarity cutoff. 1136 of these were correctly classified as present by the CGH, i.e. they had an M-value above -2, giving a specificity of 99%. Conversely, of the 278 single-copy probes ab-
sent in *B. quintana* based on sequence similarity, 242 had an M-value below -2 and were thus correctly reported as absent, giving a sensitivity of 87% (see figure 1 of paper I). As expected, the specificity was lower for multi-copy probes (80%; 52 of 65 probes correctly classified as present), while the sensitivity was somewhat higher (90%; 138 of 153).

The CGH study also showed that the *B. henselae* Houston-1 used for the CGH experiments differed from that of the sequenced isolate of the same strain in a 10 kb deletion and a 10 kb duplication, as was verified by PCR and sequencing. Both events were associated with flanking repeats. Intermediate forms of the *B. henselae* Houston-1 strain were retrieved, supporting that it had undergone those changes during laboratory cultivation.

**Genome content of Bartonella koehlerae (Paper I)**

*B. koehlerae* is more closely related to *B. henselae* than *B. quintana*, and like *B. henselae* it is associated with cats (Droz et al. 1999). However, previous studies indicated that its genome size was more similar to that of *B. quintana* (Yamamoto et al. 2002). We performed CGH to the *B. henselae* microarray in order to determine the genome content of this recently discovered species. Since the average sequence similarity to the reference species *B. henselae* is much higher for *B. koehlerae* than for *B. quintana*, the -2.0 cutoff used for *B. quintana* corresponds to -2.7 for *B. koehlerae*.

Using the -2.7 cutoff, 164 kb of the *B. henselae* genome was estimated to be absent from *B. koehlerae* (or 226 kb with the -2.0 cutoff), compared with 391 kb missing sequences in *B. quintana*. Supposing that no additional DNA is present in *B. koehlerae*, the CGH results give an estimated genome size of 1.70 to 1.77 Mb, which is intermediate between *B. henselae* (1.93 Mb) and *B. quintana* (1.58 Mb).

The genome size of *B. koehlerae* was also studied with RFLP-PFGE with two different restriction enzymes (NotI and AscI), which led to a genome size estimate of 1.7 – 1.8 Mb, similarly to the CGH results and indicating that no or only little additional sequences are present in *B. koehlerae* compared to *B. henselae*.

As for *B. quintana*, the absent sequences are located in the prophage (almost all of which is absent), the genomic islands (which are partly absent) and in the ChrII-like region (see figure 3 of paper I). In the ChrII-like region, sequence loss was particularly seen in the intergenic, noncoding sequences. However, the apparent absence of the noncoding sequence may be caused by their different degradation patterns in the three species. The ChrII-like segment is believed to stem from integration of a second chromosome, and the low coding content and large amount of noncoding sequences is probably due to the gradual degradation of sequences that were once genes.
All but 16 kb (or 4 kb with the -2.0 cutoff) of the sequences missing in *B. koehlerae* were also missing from *B. quintana*. Since *B. quintana* diverged from *B. henselae* and *B. koehlerae* before *B. koehlerae* split from *B. henselae*, this indicates that similar deletions have occurred independently in both species. Another possibility is that the sequences have not been lost from *B. quintana* and *B. koehlerae*, but rather gained by *B. henselae*. However, both *B. quintana* and *B. koehlerae* contain small remains of the prophage and genomic islands, and most of those sequences have also been found in the distantly related species *B. grahamii*, arguing against that hypothesis. Independent excisions of phage and genomic islands are not unlikely given the high number of integrase genes and gene remnants in *B. henselae*.

We could not identify any genes that explain the different host and preferences of the *Bartonella* species. Almost all genes that are shared only by the feline species *B. henselae* and *B. koehlerae* are of unknown function. However, host (or vector) specificity may also be caused by sequence differences in outer membrane proteins and secretion systems, rather than absence or presence of genes. Such genes may be under strong selection pressure and thus evolve rapidly, making it difficult to judge by CGH whether they are absent or highly diverged.

*B. koehlerae* does however seem to be rare both in cats and cat fleas (although its prevalence may be underestimated in culture-based surveys), suggesting that it may be less fit than *B. henselae*, which has the larger genome.

**Genomic diversity within *Bartonella henselae* (Paper II)**

In order to study the extent and sources of genomic diversity within a species, we studied a sample of 38 *B. henselae* isolates from different geographic regions and host species with several techniques.

**Sequence diversity**

Sequence diversity and population structure was studied using MLST, with the primers constructed by Iredell et al. (Iredell et al. 2003). We identified 6 of the 7 sequence types reported in the initial study in our sample. The sequence type we did not find in our sample was present in only one isolate in the initial study. The most common sequence type was ST1, which is that of the sequenced Houston-1 isolate. One isolate (Marseille) had a novel mutation in the *rpoB* allele, and was thus classified as a novel sequence type (ST8). However, the same isolate had also been typed in the initial study, where that mutation was absent, suggesting that it had probably arisen during laboratory cultivation. From the distribution of sequence types, we concluded that our sample is representative of the global *B. henselae* population.
Genome content diversity

CGH of the isolates showed moderate differences in genome content, with variation mostly confined to the prophage and the genomic islands, parts of which were missing in some isolates (see figure 1 in paper II). Three isolates appeared to have lost almost all of the prophage; however, a few genes were still present (although possibly at different positions in the genome). The isolates (Cheetah, GreekCat-23 and ZimCat) were of different sequence types, differing in up to four out of nine alleles, suggesting independent excision events. Several isolates also gave higher signals than the reference Houston-1 strain of the prophage region, suggesting that the prophage may be active and replicating in those isolates.

To compare the genome content with the sequence types, a phylogenetic tree was made based on presence or absence of contiguous genome regions (to avoid bias caused by gain or loss of multiple genes in single events, which may occur independently in several isolates) and compared to a phylogeny of the concatenated MLST sequences. The phylogenies showed an overall high concordance, suggesting that genome content correlates with sequence divergence in B. henselae (see figure 2 of paper II), and that recombination between isolates is rare, at least for the genes included in the MLST study. However, there were a few inconsistencies (for example IndoCat-2 and Indocat-11, which are of ST 1 but a genome content similar to isolates of ST 5), which may be explained either by independent mutation or indel events in different isolates, or by recombination between isolates with different sequence types and gene content.

Genome structure diversity

RFLP-PFGE was performed with the restriction enzymes NotI, AscI, SgfI and a combination of NotI and AscI. Initially the aim was to estimate the genome size of the isolates, in order to see whether they contained additional DNA not present in the sequenced strain and in the genome, and because a previous study had suggested that it varied widely (Maruyama et al. 2001). However, since the isolates yielded very varying restriction patterns, despite being identical in gene content and MLST sequences, we became interested in the diversity of genome structures and the rearrangement events responsible for generating them, and studied this by hybridizing gene-specific probes (PCR-products) to blots of the PFGE gels. In total, 24 probes from different genomic locations were hybridized. The results were however very difficult to interpret, partly due to inconsistencies between different restriction digests, the presence of multiple fragments of similar size seeming to be a single band, incomplete cleavage resulting in both cleaved and uncleaved fragments, and occasional probe hybridization to multiple bands, and loss or gain of restriction sites (in addition to rearrangements). In several cases, we believe that the studied culture comprises a mixed population with different
genome structures, or that the genome structure has changed between different experiments, as has previously been observed (Arvand et al. 2006).

The rearrangement events and resulting genome structures best matching the observed PFGE bands and probe hybridizations were inferred with the program XPulSee, by comparing the observed bands and probe bindings to those expected after a series of hypothetical rearrangement events.

In total, we inferred 22 different genome structures among the 35 isolates that were analyzed (see figure 4 of paper II). There was no clear correlation between sequence type or genome content and genome structure, suggesting that genome rearrangements occur frequently in *B. henselae*. Although we were unable to identify exactly which and where rearrangements had occurred, most of the data suggested inversions across the terminus of replication or between the genomic islands (which are located around the terminus of replication).

Rearrangements may cause changes in gene expression and/or sequence diversity as two homologous genes exchange parts. Thus, the frequent genome rearrangements in *B. henselae* may play a role in generating diversity in order to escape detection by the immune system of the host.

**Amplification – escape replication**

A region of the genome gave elevated signals in several isolates, suggesting higher copy numbers than in the reference (see figure 1 of paper II). The region varied in size between the isolates, ranging from ten to over 300 kb, thus spanning a large part of the genome. However, it was always centered at the same genomic location at around 1650 kb, containing a few genes with homologues to phage genes (a DNA helicase, an exonuclease and a lysozyme) and several genes of unknown function. Since the genes are located between two tRNA-genes they are probably the remains of a degraded prophage, as tRNA-genes are preferred integration sites for lysogenic phages. The relative signal (M-value) showed a peaked distribution, with a maximum at the center. This pattern suggested that the amplification was a result of bidirectional replication originating within the degraded prophage. However, the genes in the degraded prophage are probably not able to initiate replication on their own, let alone form phage particles.

To study whether the different sizes of the amplification in different isolates might be due to different culture times of the isolates, as prophages are frequently induced in stationary phase, several isolates were re-cultured and DNA was harvested after 5 and 10 days, respectively. CGH experiments showed clearly that the amplification increased in both extent and amplitude with growth time (see figure 3 in paper II), suggesting that the phenomenon is indeed linked to phage induction.
The amplification may also have contributed to the complex PFGE patterns, since the M-values in the peak suggested that this region is amplified up to 10-fold compared to the rest of the genome.

Spacer typing of *Bartonella quintana* (Paper III)

Spacer sequences, i.e. noncoding sequences between genes, can be expected not to be under selection and thus exhibit more variation than coding sequences. The new method multi-spacer typing was applied to *B. quintana*, since earlier studies suggested that the genomic diversity of this species was smaller than that of *B. henselae* (Roux and Raoult 1995). To identify suitable spacers, the sequenced *B. quintana* genome was scanned for spacers of lengths between 500 and 1500 bp, which could be amplified and sequenced in a single run with primers located in flanking genes or pseudogenes. 178 candidate spacers were found, and primers designed for 34 of these.

To screen for variable spacers, the 10 spacers with highest divergence between *B. henselae* and *B. quintana* were sequenced in three *B. quintana* isolates that differed significantly in an earlier PFGE analysis. This screen identified two variable spacers, which were found in two variants each. One of these spacers was sequenced in 64 and the other in 18 additional *B. quintana* isolates. This analysis revealed a third variant of each spacer. The different variants always co-occurred in the isolates where both spacers were sequenced, although they are located far from each other in the genome.

The remaining spacers were sequenced in three isolates, and four additional variable spacers were identified. These four and one of the two first identified were sequenced in a broader sample of 15 isolates, selected for their different PFGE profiles, disease manifestation and geographic origin. Several new variants were discovered for some of the spacers. Three spacers were considered redundant as they yielded the same results in all 15 isolates. No correlation could be found between MST genotype and PFGE pattern, geographic origin or disease symptom.

Two of the spacers were finally used on a sample of 9 lice and a patient with endocarditis, all of which had tested positive for *B. quintana* but not yielded bacterial cultures, where a fifth genotype was identified in a louse from Burundi.

The observed sequence variability could be ascribed to 12 SNPs and seven insertion/deletion events. All insertions/deletions were associated with repeated sequences.

The observed variability was nevertheless much less than expected, given that spacer sequences are generally not under selection. In addition, many of the observed mutations and indels were seen within short genes located in the spacer, or in flanking pseudogenes. The low variation might be explained by recent emergence of *B. quintana* as a species, initiated by host switch and
followed by genome reduction. The new species would then have little time to accumulate mutations. However, the studied sample is small and the geographic origin of the isolates is limited, so a larger set has to be examined.

The high variability of PFGE patterns in *B. quintana* is surprising, given that the sequenced genome lacks most of the repeated sequences present in *B. henselae* and believed to be responsible for the frequent rearrangements in that species. However, the *B. quintana* genome contains remains of the *B. henselae* prophage and genomic islands and these sites are associated with rearrangement between the two sequenced genomes. It is possible that other *B. quintana* isolates contain these sequences, although a CGH survey of ten *B. quintana* isolates to the *B. henselae* microarray showed no differences in genome content in that sample (Lindroos and Raoult, unpublished).

**Phage DNA content and growth-time related gene expression (Paper IV)**

Phage-like particles have been observed in cultures of *B. bacilliformis* (Barbian and Minnick 2000; Umemori et al. 1992), *B. henselae* (Anderson et al. 1994; Chenoweth et al. 2004) and *B. vinsonii* subsp. *berkhoffii* (Maggi and Breitschwerdt 2005). While the morphology of the phage particles differ between the species, they all contain 14 kb fragments of linear DNA. 14 kb is little to contain all genes required for infection and reproduction of a phage, and several studies have suggested that the DNA fragments are also not identical, as would be expected for a true phage. In *B. henselae*, digestion of the phage DNA by restriction enzymes followed by gel electrophoresis produced a smear, suggesting that the phage contains random samples of genomic DNA. This suggestion was strengthened by cloning and end-sequencing of individual phage DNA molecules, where each clone gave different sequences and different patterns after restriction enzyme digestion. In *B. bacilliformis* however, RFLP did not produce a smear but rather distinguishable bands, the total sum of which was however bigger than 14 kb, suggesting that the phage DNA is heterogenic but not random.

To resolve the issue of the content of the phage particles, phage DNA was isolated and hybridized against the microarray. Since it was more convenient to work with *B. grahamii*, most experiments were performed with that species. The phage DNA of *B. grahamii* was found to be of two sizes; one of 14 kb and the other bigger than 30 kb (size estimates from gel electrophoreses is more difficult for large fragments). DNA-DNA microarray hybridizations showed that the 14 kb phage DNA does indeed contain DNA from the bacterial chromosome, although with certain parts over-represented. The amplified DNA discovered in Paper II was preferentially packaged in phage particles, in congruence with its appearance after prolonged growth time and thus
probable phage induction, and its origin in a defective phage remnant. In addition, one of the *B. grahamii* prophages also appears functional and corresponds to the bigger fragment. Also in *B. henselae* the prophage was somewhat amplified in the phage DNA compared to genomic DNA although no band could be seen on the gel, which is however not unexpected since it must be a very small fraction of the total phage DNA.

Since prophages are often induced in stationary phase, which has also been observed for *B. henselae* (Chenoweth et al. 2004), we analyzed gene expression after different culture times by microarray hybridizations. The analysis revealed that phage-related genes and genes in the genomic island were induced after long culture times. However, no obvious genes responsible for the DNA amplification or for the 14 kb phage particles could be identified, as genes for the latter would be expected to be present in the genomes of *B. bacilliformis* and *B. quintana* as well. On the other hand, the presence of the same prophage in *B. henselae* and *B. grahamii* and its absence in *B. quintana*, *B. koehlerae* and some strains of *B. henselae* indicates that it is an element that may easily be lost and gained, and thus it is not impossible that strains that were used in the phage studies contained phage genes absent from the sequenced genomes. In addition all the genomes contain many genes of unknown function.

**B. henselae** gene expression during infection (Paper V)

In collaboration with a research group in Switzerland, I analyzed *B. henselae* gene expression in a time series of infection in human endothelial cells. Bacterial RNA was isolated 15 minutes, 6, 15, 24 and 48 hours after infection, and each sample and hybridized against RNA from bacteria grown on agar plates. Approximately 100 genes were differentially expressed (i.e. at least two-fold up- or down-regulated and a t-test p-value below 0.05) in any of the 5 time points. Cluster analysis revealed four groups of genes with different expression patterns; one group was consistently down-regulated, another was first up-regulated and then down-regulated, a third was up-regulated late in the time course and showed increased expression with time, and a fourth group was up-regulated almost immediately upon transfer to the intracellular niche but decreased in expression with time.

The cluster of genes that were up-regulated late in the time course contained several genes known to be involved in virulence and the expression profile matched that expected from genes involved in an adaptive response. A transcriptional regulator with a similar expression pattern was identified.

A mutant lacking the transcriptional regulator was created and the expression pattern of this mutant was compared to that of wild-type bacteria after both intracellular growth and growth in medium. The approximately 50
genes differentially expressed between mutant and wild-type cells overlapped well with those up-regulated late in the time course.

Although the transcriptional regulator is conserved in alpha-proteobacteria, the gene set it regulates contains more than ten *Bartonella*-specific and many horizontally transferred genes.

The microarray results were validated by quantitative real-time PCR which supported the micrarray results but indicated much (5- to 10-fold) higher levels of up-regulation, suggesting that microarrays might under-report expression changes. Compression of fold changes in microarray measurements compared to real-time PCR has previously been reported (Wang et al. 2006).
5. Summary and discussion

In this work I have studied genome diversity at different levels – genome content, genome structure and gene sequences – in different species and isolates of Bartonella, intracellular bacteria that inhabit the erythrocytes and endothelial cells of many mammals. The differences in host specificity, along with their unique infection strategies, make Bartonella an interesting genus for evolutionary studies.

The main focus has been on the feline-associated species B. henselae, where the variability in gene sequences (as determined by multi-locus sequence typing, MLST) was low compared to other species, and the variability in genome content mostly restricted to the prophage and genomic islands, most of which had also been lost from the closely related species B. koehlerae, which is also associated with felines, and which are absent from the human-specific pathogen B. quintana. It thus seems that these mobile elements have been lost (and possibly gained) in multiple independent events, and that neither of them are required for infection of cats or humans. However, none of the B. henselae isolates was even remotely similar in genome content as B. koehlerae or B. quintana, supporting their status as separate species. There was no correlation between sequence type or genome content and the animal host (cat or human) of isolation or its geographic origin, suggesting a global distribution and transmission of different genotypes and indicating that all genotypes are able to infect humans. However, there was a clear correlation between genome content and sequence type, indicating that recombination between isolates of different sequence types is rare. A few isolates did however indicate that such recombinations had occurred.

In contrast to the modest variation in gene sequences and genome content, genome structure was very variable and did not correlate with either gene sequence or genome content, but rearrangements appeared to occur frequently in all sequence types, probably by inversions between repeated sequences in the genomic islands, which are located around the terminus of replication. Since rearrangements may alter the sequence, context and/or expression pattern of genes, frequent rearrangement may contribute to antigenic variation allowing the bacteria to escape the immune system of the host.

Also B. quintana, which is restricted to humans, showed low sequence variability, although intergenic spacers rather than housekeeping genes were
studied. This low diversity is probably due to the recent emergence and host switch of this species from the feline-associated ancestor of *B. henselae*, *B. koehlerae* and *B. quintana*. However, PFGE profiling of *B. quintana* also showed evidence of frequent rearrangements, even during cultivation in the laboratory, which is surprising given the low amount of repeat in its genome but underlines the probable importance of this phenomenon.

Another surprising finding was that a large region of the *B. henselae* chromosome was amplified up to 10-fold compared to the rest of the chromosome in longtime cultures. The amplification is probably caused by phage escape replication originating within what is probably the remains of a degraded prophage (containing only a DNA helicase, a lysozyme and an exonuclease), and occurs also in the distantly related rodent-borne species *B. grahamii*. Analysis of the DNA content of phage particles from *B. henselae* and *B. grahamii* showed that the amplified DNA is preferentially packaged into phage particles, consistent with its probable phage-associated induction. The rest of the chromosome was also cleaved and packaged into page particles although at a much lower frequency. This could possibly explain previous inconsistent results regarding the phage DNA content, and might also be a way of achieving higher transmission and recombination rates for genes in the ChrII-like region than the rest of the genome. Gene expression analysis showed that phage-related genes in the prophages and genomic islands of *B. grahamii* were induced in longtime cultures, although no obvious candidates for the production of the 14 kb phage particles could be identified since they would be expected to be present in the genomes of all *Bartonella* species which produce similar phage particles. The genes in the degraded prophage did not increase in expression with growth time, suggesting that they are not actively involved in the amplification but rather serve as a replication origin that is recognized by other phage genes.

The retention of these genes in all sequenced *Bartonella* genomes, even the reduced genome of the very distantly related species *B. bacilliformis* (which also produces phage particles with similar properties), is intriguing unless they have some value for the bacterium (or have some other means of enforcing their continual presence). In general, the production of phage particles containing bacterial DNA is intriguing in itself, as it can hardly be of value to neither the bacterium (which dies in the process) nor the phage (which cannot propagate itself since it doesn’t contain its own DNA). The phenomenon may be of value to the bacterial community in increasing the recombination rate, but for such a system to be maintained through kin selection would require that only bacteria that contain the phage production genes (whichever they are) are able to take up the transduced DNA.

Finally I analyzed *B. henselae* gene expression during infection of human endothelial cells. A cluster of genes that increased in expression after infection contained many genes known to be involved in infection and pathogenesis. A transcriptional regulator with a similar expression profile was identi-
fied and demonstrated through mutation experiments to be responsible for their induction. Despite the conservation of the regulator in alphaproteobacteria, the regulated gene set contains many Bartonella-specific and horizontally transferred genes, pointing to the adaptive nature of regulatory systems.

In summary, genome rearrangements appear to be common in both B. henselae and B. quintana, while the sequence diversity is rather low. The prophage and genomic islands have been lost, and possibly gained, several times, suggesting that they are functional mobile elements and re-stating the importance of such elements in bacterial evolution.
Bakterier har olika mekanismer för genetisk utveckling (evolution), varav de flesta även förekommer i andra organismer. Förutom mutuationer, där en enskild bas i DNA-kedjan byts ut p.g.a. DNA-skada eller misstag i DNA-kopieringen inför celldelning, kan stora delar av kromosomen klippas ut, dupliceras eller flyttas till en annan plats. Omflyttningar av sekvenser inom kromosomen kan dels skapa nya gener eller gen-kombinationer, och dels andra betingelserna för när olika gener används. Detta kan vara till nytta bl.a. för bakterier som lever i en värdorganism och behöver ändra på sig för att inte bli igenkända och eliminerade av värdjurets immunförsvar. Repeterade sekvenser är ofta inblandade i både omarrangemang av kromosomen, dupliceringar och förluster av DNA.

Bakterier kan även ta upp DNA från andra bakterier, antingen direkt via en tunnel mellan två närliggande bakterier eller indirekt genom att antingen ta upp fritt DNA från omgivningen eller via ett bakterievirus (fag) som ibland av misstag innehåller bakterie-DNA istället för sitt eget DNA. DNA som tillkommit utifrån skiljer sig ofta från det som nedärvts, och bildar ”genomiska öar” i bakteriens kromosom. Sådana öar innehåller ofta gener som är av värde för bakterien under vissa omständigheter (t.ex. infektion) och som därför bevarats. Fager kan även integreras i bakteriens kromosom och kopieras med denna när bakterien förökar sig, i stället för att använda bakteriens maskineri för att kopiera sig själv och därefter spränga bakterien. Integrerade fager kallas profager, och när bakterien upplever någon form av stress, såsom minskade näringsresurser eller ogynnsamma förhållanden, tar profagen till flykten genom att hoppa ut från kromosomen och börja kopiera sig själv och slutligen döda bakterien. Många bakteriers kromosomer innehåller profager, eller rester av nedbrutna sådana.

Mitt doktorandarbete har i stort syftat till att studera den naturliga variationen i arvsmassan (genomet) i och mellan olika arter av bakterien *Bartonella*, för att ta reda på vilka mekanismer för genetisk förändring som förekommer och om de genetiska förändringarna har något samband med bakteriens värdjur eller ursprungsland.

*Bartonella* är bakterier som infekterar de röda blodkropparna och cellerna som omger blodkärlen i många däggdjur, från möss till människor. De sprids mellan värdjurven av blod sugande insekter, som t.ex. löss och loppor. Olika *Bartonella*-arter har specialiserat sig till olika värdjur, och i dessa gör de
normalt ingen skada. Ibland infekteras dock ett annat värddjur och då kan en mängd olika sjukdomssymtom uppstå.

Huvudsakligen har jag fokuserat på *Bartonella henselae*, som normalt finns i kattdjur. I vissa delar av världen är över 90% av alla katter infekterade av *Bartonella*, men i Sverige bara 1-2%. Ibland infekterar *B. henselae* även människor (genom ett rivas från en katt, eller eventuellt genom ett insektsbett), och kan då ge upphov olika symptom beroende på personens immunförsvar. Personer med normalt immunförsvar får ofta inga symptom alls, men kan ibland drabbas av kattklädersjuka (som karakteriseras av feber och svullna lymfkörtlar), medan personer med nedsatt immunförsvar kan få allvarligare symptom. Bl.a. kan *Bartonella*-bakterier förmå de celler de lever i att föröka sig och bilda tumörer, för att utöka bakteriens livsutrymme.

Med olika tekniker har jag undersökt skillnader i geninnehåll, ordning i kromosomen och genevektorer i 38 isolat av *B. henselae*, tagna från katter och människor i olika delar av världen och från en afrikansk gepard. Det visade sig vara relativt liten variation i geninnehåll, och de gener som saknades i några isolat fanns framför allt i en profag, och i genomiska öar. Det var stor samstämmighet mellan hur isolaten liknade varandra baserat på geninnehåll och genevektorer, vilket tyder på att det inte är så stor rekombination mellan olika isolat (d.v.s. att bakterier byter ut delar av arvsmassan mot motsvarande del från en annan individ). Det fanns dock några undantag som pekade på att viss rekombination förekommer, och även att profagen har förlorats i flera oberoende isolat. Det fanns inte något samband mellan geninnehåll eller -vektorer och värddjur eller geografiskt ursprung, vilket tyder på att alla olika sekvenstyper kan infektera både människor och katter.

Skillnaderna i kromosomens arkitektur var däremot stora, och få isolat verkade ha samma genordning. Dessa skillnader hade inget samband med skillnaderna i geninnehåll eller sekvenstyper, vilket tyder på att kromosomen omarangeras ofta. Eftersom *Bartonella* lever i däggdjur kan det vara ett sätt att undkomma immunsystemet. En annan upptäckt var att en stora del av bakteriens kromosom fanns i många kopior i flera isolat. Ytterligare studier visade på att denna amplifiering startade i vad som förmodligen är resterna av en nedbruten profag, som börjat kopiera sig själv men sedan fortsatt att kopiera även de kringliggande bakteriella generna, och att amplifieringen ökade efter längre odlingstid, antagligen eftersom den sätts igång av en fag som reagerar på den minskade näringsstillgången och försöker fly.

En nära släkting till *B. henselae* är det nyupptäckta arten *B. koehlerae*, som också finns i katter men verkar vara mer sällsynt. Eftersom tidigare studier tydde på att den hade en mycket mindre kromosom än *B. henselae* (mer lik den hos *B. quintana*, som har specialiserat sig på människor och troligen förlorat många gener i den processen), studerade vi dess geninnehåll i förhållande till *B. henselae*. Även *B. koehlerae* hade förlorat profagen och delar av de genomiska öarna, som även saknas i *B. quintana*. De måste dock ha förlorats oberoende av varandra eftersom *B. koehlerae* är närmare släkt
med *B. henselae* än *B. quintana*. *B. koehlerae*-kromosomen var dock större än de tidigare uppskattningarna tydde på.

Eftersom *B. quintana* tros ha skilt sig från *B. henselae* och *B. koehlerae* relativt nyligen och därför inte hunnit samla på sig så många olika mutatio-

ener, studerades variationen i denna art genom att sekvsera DNA mellan
gener, d.v.s. DNA som inte kodar för något. Det kan förväntas vara större
variation i dessa regioner än i gener eftersom det inte finns någon selektion
på dessa sekvser. Variationen var dock liten även i dessa regioner i *B.
quintana*. Däremot verkade kromosomen omarvarerat ofta även i denna art,
vilket var förväntade eftersom den har väldigt lite repeterade sekvser.

Flera tidigare studier har visat att olika *Bartonella*-arter bildar fag-
partiklar som innehåller 14 kb DNA (d.v.s. 14.000 nukleotid-baspar), men
DNA-innehållet i dessa har inte kunnat säkerställas. Vi isolerade DNA från
fag-partiklar från *B. henselae* och från *B. grahamii*, en art som normalt lever
i möss, och jämförde det med bakterie-DNA från samma odling. Det visade
sig att det framför allt var den del av kromosomen som amplifierades efter
långa odlingstider som klövs i småbitar och packades i fag-partiklar. Resten
av kromosomen klövs och packades också men i mindre utsträckning. Även
DNA från profagen var överreprisererat i fag-partiklarna, vilket tyder på att
profagen är funktionell och kan kopiera och sprida sig själv. För att ta reda
på vilka gener som ligger bakom amplifieringen och produktionen av fag-
partiklar undersökta vi genernas aktivitet efter olika odlingstider. Vi fann att
gener i profagen och de genomiska öarna ökade i aktivitet efter långa od-
lingstider, men kunde trots det inte identifiera vilka som orsakade bildandet
av fag-partiklarna som innehöll 14 kb DNA, eftersom sådana partiklar även
bildas i *B. quintana* och *B. bacilliformis*, som saknar intakta profager.

Slutligen bidrog jag också till att analysera data från experiment där *B.
henselae* infekterat mänskliga celler. Analyserna visade att en grupp med
gener som sedan tidigare var kända för att vara inblandade i infektion ökade
i uttryckt efter infektionen, och att en transkriptionsfaktor följe samma
mönster och visade sig styra uttrycket av dessa gener. Trots att transkrip-
tionsfaktorn finns i alla närbesläktade bakteriearter är många av generna den
reglerar i *B. henselae* specifika för *Bartonella*, vilket visar på ännu en typ av
genetisk variation och anpassning.

Sammanfattningsvis verkar omarvaremang av kromosomen vara den
primära orsaken till genetisk variation hos *Bartonella*, medan rekombination
mellan olika celler är mer sällsynt för de gener vi studerat. Vissa sekvser
har förlorats (och eventuellt återfåtts) flera gånger, i arter och isolat som
lever både i katter och människor. Att en viss del av kromosomen amplifier-
ras och packas i fag-partiklar kan bidra till ökad rekombination och variation
för dessa sekvser.
Acknowledgements

A lot of people have helped me, directly or indirectly, in my PhD work and in life, and I would like to express my deep-felt gratitude to all of you.

First of all, **THANK YOU Siv** for your support, advice and encouragement during the years, in both scientific and personal matters. Your enthusiasm, optimism and (sometimes a bit crazy) ideas is amazing!!! I wish every PhD student had a supervisor like you...

Thanks a lot also to **Kristina**, who has helped me a lot with the laboratory work – and looked after my cats when I’ve been away.
Thanks to **Cissie** for starting up the microarray project when the technology was still in its infancy, and for getting me started in the lab.
Thanks to **Alex** for your good company and hard work (especially with the d**d** phage) during my first year. I’ll never forget your UV-tan!
Thanks **Dirk** for starting me up with microarray analysis and R, and for discussions about normalization.
Thanks **Håkan** for helping me with all the strange computer problems I’ve encountered – you may get out of work now that I’m leaving... :-)

Thanks also for your cheerful contributions to coffee breaks and beer clubs!
Thanks **Olga** for all the PFGE gels and blots – and for checking and re-checking when I couldn’t believe them... Thanks also for helping with the phage project, for your cheerfulness – and for the monster cake!
Thanks **Eva** for the phage work, and for skype messages when I’ve worked from home – good luck with the grahamii strains (and the ever-almost-finished genome...)! Thanks **Ann-Sofie** for helping me in the lab, looking after my cats and hugging me when I needed it.
Thanks **Ola** for help with figures and Mac, and for being a nice room mate during the last months.
Thanks **Karin O.** for help with paperwork and reservations, and for lending me the cat-gloves.
Thanks to all project student who have participated in one way or the other in this project; **Aurelie, Alix, Marcel, Na, Olle, Kirsten** and **Christan**.

Thanks **Magnus** for lending Cy-dyes when I had forgotten to order, for help and advice on array matters, and for handling the functional genomics course
– I would have been lost without you. I admire your engagement in important issues in the “real world”!

Thanks **Mickie** for your expertise in everything regarding perl, bioinformatics and computers in general – you are a genius! :-) And a very nice one, taking time for to help when you already have an overfull agenda!

Thanks **Robban** for discussions about pedagogics and teaching, and for your immense knowledge of microbiology and its history.

Thanks **Johan** (mini-Håkan) for introducing the Friday Tie and for being helpful when I’ve had problems with bioperl and blast.

Thanks **Hans-Henrik** for computer expertise and good mood. Good luck Fuxen, soon its your turn!

Thanks **Jonas** for your entertaining corridor dances (or whatever you call it).

Thanks **Björn I** for teaching me Perl and HTML during my degree project – and for persuading me to stay on another five years!

Thanks **Björn II** for taking an interest in my (and others) work, for help with my new Mac, and for occasional discussions in the corridor about science, ethics, love and life in general. Good luck with your very promising work!

Thanks **Björn III** for discussions about the projects, for arranging activities and for making us laugh during coffee breaks. And for showing the proper use of the lab ethanol :-)

And thanks **Björn IV** for your smiles!

Thanks **Rolf** for array discussions, **Otto** for discussions on normalization and teaching; and both of you for wise comments on journal club and research presentations.

Thanks all my previous room mates for your nice company; **Idress**, **Ramy** (thanks for the bread – and the secret recipe!), **Ylva**, **Alexandra** (good luck with your new job, and pet Tigrou from me!).

Thanks also to everyone, past or present, at MolEv for nice company during coffee breaks and contributing to the nice atmosphere at the department: **Haleh**, **Ivan**, **Cessie**, **Karín H.**, **Boris**, **Wagied** (thanks for your long-distance pep-mails – I hope I’ll have more time for writing back once this is over...), **Gabor**, **Lisa** (I miss your laughs!), **Carolin** (thanks for answering all my questions about the genomes, and your encouragement), **Dave**, **Sean**, **Olof**, **Karin E.** (thanks for the chocolate cake – and for the tough Sculp-classes...), **Stefan**, **Erica A.**, **Erika G.**, **Gustav**, **Maria**, **Erik**, **Joaakim**, **Fredrik**, **Thijs**, **Wei**, **Katarzyna**, **Yun**... and everyone else I have momentarily forgotten while I’m writing this!

Thanks to the nice and helpful personnel at UAS, in particular **Elin** and **Kristina**, and **Ingrid M.**

Thanks also to my friends, for company, distraction and support through the years: **Katja**, **Linda R.** and **Marie** (“Fyrklövern” – we’ve held together for long now and I hope it will last many years more!), **Linda J.** (the partycat – Ich vermiss Dich sehr!), **Lisa** (the one with the hair – we ought to meet more often!), **Ingrid** for being there in sun and rain and always having your door open, and your lovely family (thanks **Jonathan** for the rides in the sky
and for your jokes (although I seldom get them), Emelie for being so kind and Johannes for being so sweet and for the honor of being your Godmother; I wish you all the best in life), Kaisa and Ulrika.

Thanks to my foreign friends I met during my year in Norwich: Miriam (my little rabbit), Susanne, Elisabeth (my little pony) and Johanna.

Thanks to you who helped me through the teenage years; Jennifer (forever in my memory, now an angel – as in life), Carin, Elisabet, and most of all Helena.

Thanks Philochoros for the fun dance (although I do it rather than well – a special thanks to all the patient partners!) and other nice activities.

Thanks to my parents Eva and Peter (you have certainly contributed to my interest in science, although in a different subject) and my brother Hannes for bright childhood memories and for all your love, help and support – I love you!

And thanks to all my relatives (too many to list here!) for nice parties and get-togethers, and showing you care.

Thanks also to my cats for filling my life with so much love and joy (and occasional troubles). Special thanks to Pippi for loudly reminding me of taking breaks from writing to cuddle with you (what could be more relaxing?) – and to Ronja for helping me in writing (a pity you’re so bad at spelling) and for tearing up and eating my references if I don’t read them...

And finally, again: THANK YOU SIV!!!
References


Ixodes pacificus ticks in California. *Journal of Clinical Microbiology* 39:1221-6


Moreno, E. (1998) Genome evolution within the alpha Proteobacteria: why do some bacteria not possess plasmids and others exhibit more than one different chromosome? *FEMS Microbiology Reviews* 22:255-75


58


Ward, N. and Fraser, C. M. (2005) How genomics has affected the concept of microbiology. *Current Opinion in Microbiology* 8:564-71


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 317

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)