JOURNAL OF BACTERIOLOGY, Sept. 2005, p. 6155–6165 Vol. 187, No. 17
0021-9193/05/$08.00+0 doi:10.1128/JB.187.17.6155–6165.2005

Characterization of the Genome Composition of *Bartonella koehlerae* by Microarray Comparative Genomic Hybridization Profiling†

Hillevi L. Lindroos,[1] Alex Mira,[1]‡ Dirk Repsilber,[1]§ Olga Vinnere,[1] Kristina Näslund,[1] Michaela Dehio,[2] Christoph Dehio,[2] and Siv G. E. Andersson[1]*

Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, 752 36 Uppsala, Sweden,[1] *and Division of Molecular Microbiology, Biozentrum of the University of Basel, Basel, Switzerland*[2]

Received 17 February 2005/Accepted 27 May 2005

Bartonella henselae **is present in a wide range of wild and domestic feline hosts and causes cat-scratch disease and bacillary angiomatosis in humans. We have estimated here the gene content of *Bartonella koehlerae*, a novel species isolated from cats that was recently identified as an agent of human endocarditis. The investigation was accomplished by comparative genomic hybridization (CGH) to a microarray constructed from the sequenced 1.93-Mb genome of *B. henselae*. Control hybridizations of labeled DNA from the human pathogen *Bartonella quintana* with a reduced genome of 1.58 Mb were performed to evaluate the accuracy of the array for genes with known levels of sequence divergence. Genome size estimates of *B. koehlerae* by pulsed-field gel electrophoresis matched that calculated by the CGH, indicating a genome of 1.7 to 1.8 Mb with few unique genes. As in *B. quintana*, sequences in the prophage and the genomic islands were reported absent in *B. koehlerae*. In addition, sequence variability was recorded in the chromosome II-like region, where *B. koehlerae* showed an intermediate retention pattern of both coding and noncoding sequences. Although most of the genes missing in *B. koehlerae* are also absent from *B. quintana*, its phylogenetic placement near *B. henselae* suggests independent deletion events, indicating that host specificity is not solely attributed to genes in the genomic islands. Rather, the results underscore the instability of the genomic islands even within bacterial populations adapted to the same host-vector system, as in the case of *B. henselae* and *B. koehlerae*.**

Comparative whole-genome microarray hybridization has revealed differences in the contents of genes and genomic islands (GEIs) in closely related strains or species with different host ranges and virulence characteristics (11, 15, 17, 22, 23, 25, 29, 35, 44, 52, 55). The GEIs may contain bacteriophage and plasmid genes and are typically flanked on one side by a tRNA gene and on the other by an integrase gene (21). The islands are integrated into the chromosome by site-specific recombination, stabilized by loss of genes for replication, and modified by insertion of other DNA elements (14). After excision from the chromosome with the aid of the integrase protein, GEIs may be exchanged among bacterial cells via horizontal gene transfer (14). Proteins encoded by the GEIs are thought to increase fitness characteristics in one or more growth environments and in the case of intracellular bacteria they may play a role in host adaptation.

The genus *Bartonella* belongs to the α-proteobacterial subdivision and contains approximately 20 arthropod-borne facultative intracellular species identified in a broad range of mammals. Small rodents, squirrels, dogs, cats, deer, and mice are common hosts for *Bartonella* species. Wild and domestic cats serve as the natural reservoir for *Bartonella henselae*, which establish long-term infections in erythrocytes typically without inducing disease symptoms. If transmitted to humans, which are only an incidental host, *B. henselae* may cause cat-scratch disease, bacteremia with fever, bacillary angiomatosis, and peliosis, endocarditis, and neuroretinitis (12). *Bartonella quintana* is a louse-borne human pathogen and the causative agent of trench fever, as well as bacillary angiomatosis and endocarditis (12).

Complete genome sequence data are available for the 1.93-Mb genome of *B. henselae* and the 1.58-Mb genome of *B. quintana* (1). Present in the *B. henselae* genome is a prophage and three main GEIs (HGIa to -c) that are flanked by tRNA and integrase genes. The largest genomic island (HGIa) is 72 kb and contains genes for phage proteins interspersed with multiple genes for filamentous hemagglutinin. Surprisingly, the *B. henselae* GEIs are absent from the louse-borne human pathogen *B. quintana*, suggesting that they are not pathogenicity islands (PAIs) of the classical type that encode virulence factors causing human disease (1).

Another interesting region is the so-called chromosome II-like segment, which is defined as the region flanked on one side by an rRNA operon and on the other side by a $tRNA^{Leu}$ gene (from 1,409 to 1,673 kb in the *B. henselae* genome) (1). This segment contains a second rRNA operon, as well as *Bartonella*-specific genes and homologs to genes located on the second replicon of *Brucella*. It is hypothesized that it was acquired by recombination of a second replicon into the main chromosome of an ancestral *Bartonella* genome at the site of the rRNA operon (1).

A few genes in this segment show similarity to genes normally located on plasmids, a finding indicative of frequent

* Corresponding author. Mailing address: Department of Molecular Evolution, Evolutionary Biology Center, Norbyvägen 18C, 752 36 Uppsala, Sweden. Phone: 46 (0)18-4714379. Fax: 46 (0)18-4716404. E-mail: Siv.Andersson@ebc.uu.se.

† Supplemental material for this article may be found at http://jb.asm.org/.

‡ Present address: División de Microbiología, Universidad Miguel Hernández, 03550 Alicante, Spain.

§ Present address: Institute for Medical Biometry and Statistics, University Lübeck, D-23538 Lübeck, Germany.

exchanges with plasmids. One of these, the *vapA*5 gene, encodes a protein with similarities to the *higA* gene of plasmid Rts1, which together with the killer protein HigB forms a plasmid maintenance system (8, 50). The *vapA*5 gene is located within a cluster of more than 10 short open reading frames in *B. henselae*, none of which is present in *B. quintana* (1). It is conceivable that the *vapA*5 gene and the surrounding short open reading frames are the remainders of an integrated plasmid that is no longer essential and therefore in the process of being lost.

The chromosome II-like segments are 195 and 274 kb and have coding contents of only 40 and 50% in *B. quintana* and *B. henselae*, respectively. As in the case of GEIs, the rate of sequence loss in the chromosome II-like region is more pronounced in the louse-borne human pathogen *B. quintana* (1). Dramatic gene loss has previously been associated with the emergence of the louse-borne human pathogen *Rickettsia prowazekii*, the agent of epidemic typhus (4, 33). Comparative sequence analysis has shown that sequence elimination in *Rickettsia* is mediated by a mutational bias for short deletions, leading to a gradual deterioration of inactivated gene sequences (3, 5, 6). A hypothesis is that adaptations to host-restricted vectors, such as lice, are associated with accelerated rates of genome degradation (1).

Genome size variations for species with different host preference patterns have also been reported for *Bordetella*, where gene loss, rather than gene gain, has been associated with the emergence of human pathogens (11, 34). Other paradigmatic examples are those of *Shigella flexneri* and *Salmonella enterica* serovar Typhi that have accumulated many pseudogenes and presumably have specialized in human infections (13, 51).

Because it is of general interest to examine the genome content of closely related bacterial species with different host association patterns, we compared the contents of the sequenced genomes of *B. henselae* and *B. quintana* with that of a third species, *B. koehlerae*, which has been isolated from domestic cats (16, 41) and was recently associated with a case of human endocarditis (7). Thus, the definitive host for both *B. henselae* and *B. koehlerae* is the cat, with incidental infections in humans. Phylogenetic analyses suggests that *B. koehlerae* is an early diverging sister clade of *B. henselae* to the exclusion of *B. quintana* (16, 19, 24, 26, 54). We refer here to the sister taxa *B. henselae* and *B. koehlerae* as feline pathogens, with the outgroup species *B. quintana* being referred to as a human pathogen.

Based on the genome sequence of *B. henselae*, we have constructed a whole-genome microarray, which was here used to investigate the gene content of *B. koehlerae*. Sequences absent in *B. koehlerae* were mostly confined to regions in the chromosome II-like region and the GEIs of *B. henselae*. The results are discussed in the context of the possible role of GEIs in *Bartonella* with respect to virulence, host specificity, and persistence of the infection.

MATERIALS AND METHODS

Culture of the strains and DNA isolation. The strains used in the present study—*B. quintana* (strain Toulouse), *B. koehlerae*, and *B. henselae* (strain Houston-1)—were kindly provided by Scott A. Handley (National Center for Infectious Disease, Atlanta, GA) and Martin Holmberg (Uppsala University Hospital, Uppsala, Sweden). Bacteria were grown for 7 to 14 days on blood (*B. henselae* and *B. quintana*) or chocolate (*B. koehlerae*) agar plates containing 5% horse blood at 35°C in a 5% CO_2-enriched atmosphere. Bacteria from each plate were collected with the help of sterile cotton applicators and suspended in 500 μl of phosphate-buffered saline. After centrifugation, DNA was isolated with Aqua-Pure DNA Isolation Kit (Bio-Rad, Hercules, CA) according to the manufacturer's instructions. The precipitated DNA was resuspended in 100 μl of DNA hydration solution (Bio-Rad) to a final concentration of 300 to 1,000 μg/ml per plate.

PFGE analysis. Bacteria grown on horse blood agar plates for 2 weeks at 35°C were washed in TNE buffer (10 mM Tris [pH 8.0], 150 mM NaCl, 1 mM EDTA) and centrifuged; washes were repeated twice. Plugs were made by using 2% SeaPlaque GTG agarose (Cambrex BioScience) in 0.5× Tris-borate-EDTA (TBE) buffer as described previously (10). Ready plugs were then stored in 0.2 M EDTA (pH 8.0) at 4°C until used. Prior to digestion with restriction endonucleases, the plugs were cut into 1- to 2-mm-thick slices, and each slice was preequilibrated in 1 ml of TE buffer for 30 min on ice and then in 100 μl of the appropriate restriction buffer for 15 min. Restriction reactions were then performed with 20 U of NotI or AscI (New England Biolabs) in a total volume of 100 μl at 37°C overnight. The slices were then preequilibrated in 1 ml of gel tank running buffer (0.5× TBE) for 30 min on ice. The DNA fragments were separated by electrophoresis in 1.2% agarose (SeaKem Gold; Cambrex Bio Science) in 0.5× TBE buffer in GenNavigator System apparatus (Amersham Biosciences, Uppsala, Sweden) at 14°C and 5.1 V/cm for a total of 55 h split into six phases: 5 s for 6 h, 60 s for 8 h, 90 s for 10 h, 100 s for 12 h, 120 s for 120 h, and 150 s for 5 h. The gels were then stained with ethidium bromide (0.5 μg/ml) for 30 min and destained in two washes of deionized water, with 20 min for each wash. The sizes of the fragments were estimated by using pulsed-field gel electrophoresis (PFGE) λ-ladder and Yeast Chromosome PFG marker (New England Biolabs).

***Bartonella* DNA array design and construction.** A total of 1,650 sequences from the *B. henselae* Houston-1 strain were amplified by PCR. These included 174 sequences from 139 noncoding regions and 1,476 sequences from 1,365 genes, which represent 92% of the *B. henselae* predicted coding regions. Primers were designed with PrimeArray (37) and Primer3 (43). External tags were appended to all primers, and the original PCR products were subjected to a second amplification with primers complementary to the tags in order to achieve even product concentrations. The sizes of the PCR products ranged from 75 to 1,745 bp, with a mean of 502 bp. PCR products covering areas of homology between genes were included in the analysis, but these were handled separately from single-copy sequences (see Results).

PCR products were purified with the Multiscreen PCR 96-well filtration system (Millipore), dissolved in 30% dimethyl sulfoxide, and spotted on UltraGAPS-coated slides (Corning, Inc.) in six replicates. Slides were UV cross-linked at 300 × 100 μJ/cm^2 in a Spectrolinker XL-1500 UV cross-linker (Spectronics Corp.) and prehybridized for 45 min with 3× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate (SDS)–0.1 mg of bovine serum albumin (BSA)/ml at 50°C and then washed with water and isopropanol and centrifuged dry.

Fluorescence labeling of genomic DNA. We mixed 2 μg of genomic DNA from *B. henselae*, *B. quintana*, or *B. koehlerae* with 8 μl of a 2.5× random primer-reaction mix (BioPrime DNA labeling kit; Invitrogen) in a total volume of 7.7 μl of water, boiled the mixture for 5 min, and placed it on ice. The primed DNA was mixed with 0.8 μl of a 25× nucleotide mix (12.5 mM dATP, dCTP, and dGTP; 5 mM dTTP), 2 μl of 1 mM Cy3- or Cy5-dUTP (Amersham Biosciences), and 0.5 μl of Klenow enzyme (BioPrime DNA labeling kit; Invitrogen). The sample was incubated for 1 h at 37°C, and the reaction was stopped with 2 μl of Stop Buffer (BioPrime DNA labeling kit; Invitrogen). Labeled samples to be cohybridized were combined and cleaned with the MinElute reaction cleanup kit (QIAGEN) and then eluted in 10 μl of buffer EB.

Microarray hybridization. The labeled DNA was mixed with 90 μl of hybridization solution (3× SSC, 0.1% SDS, 0.1 mg of sonicated salmon sperm DNA/ml) and heated at 95°C 1 min before application to the slide. Slides were hybridized for 15 h at 50°C and washed in the dark by first dipping them in 2× SSC–0.1% SDS at 50°C until the coverslip moved away and then keeping them in 2× SSC–0.1% SDS at 50°C for 5 min, followed by treatment for 10 min with 0.1× SSC–0.1% SDS at room temperature, four 1-min washes in 0.1× SSC at room temperature, and finally a rinse in 0.01× SSC for 5 to 10 s. Samples were then spun dry by using a slide centrifuge.

Scanning and quantification. Slides were scanned with a ScanArray 4000 scanner (Packard BioChip Technologies/Perkin-Elmer, Inc.) using the ScanArray Express (version 2.1) software (PackardElmer, Inc.) at a resolution of 10 μM. Photo multiplier tube voltage (PMT) was manually adjusted to balance intensities in the two channels while avoiding a high number of saturated spots. Quantification of scanned images was also performed with ScanArray Express

with the following settings: spot diameter, 130 μm; maximum diameter, 120%; and minimum diameter, 80%. Spots were automatically flagged as bad if the spot mean was ≤1.70 times the background mean or if the spot mean was ≤400 plus the background mean (limits chosen after manual inspection). Spots were also visually inspected and adjusted or flagged if necessary. For all hybridizations (except the comparison of the sequenced versus array reference strains of Houston-1) Ch1 denotes the reference *B. henselae* Houston-1 strain and Ch2 is the test strain (*B. quintana* or *B. koehlerae*).

Filtering and normalization. Data were analyzed by using the statistics program R (http://www.R-project.org) with custom-made scripts. Spots flagged as "bad" or as "not found" during quantification were removed before analysis, as were also spots with very low Ch1 intensities because low-intensity spots are less reliable due to the large relative contribution of background noise (28, 36). Spots were removed if Ch1 median spot intensity was less than three times higher than Ch1 median background intensity; if less than 90% of the spot pixels had an intensity that was higher than the background intensity plus one standard deviation or if less than 70% of the spot pixels had an intensity higher than the background intensity plus two standard deviations. Spots that were saturated in either channel (more than 15% of the pixels saturated) were also removed since they may give erroneous intensity measurements (36).

For hybridizations with *B. koehlerae* and *B. quintana*, only 3 and 4%, respectively, of the 9,882 spots with successful PCRs were filtered out compared to 97% of the 60 negative control spots containing only H_2O or the PCR mix and 32% of the 1,530 spots represented by negative PCRs. The filtered spots, as well as the spots representing negative controls and PCRs, were removed prior to the analysis.

After correction of median intensities, M values were calculated as $\log_2$(Ch2/Ch1). Normalization was performed by shifting the distribution of M values such that the peak (identified by Gaussian kernel density estimation), corresponding to the majority of genes, was centered at M = 0. The density peak was used instead of the mean or median since the distribution of M values was skewed due to the high proportion of absent or divergent sequences. Plots of M versus $\log_2$Ch1 did not reveal any significant intensity-dependent dye bias, and lowess normalization was not performed [plots of M value versus average intensity A, where A = 0.5($\log_2$Ch1 + $\log_2$Ch2), were biased by the high proportion of absent or diverged genes having low values for both M and A].

Identification of putatively cross-hybridizing sequences. Sequences amplified by PCR were scanned for potential cross-hybridization through a BLASTN (2) search of the sequence fragment against the sequenced Houston-1 genome. For each hit, a global alignment of the PCR-amplified sequence to the corresponding region of the genome was performed with the Needleman-Wunsch algorithm (32) with the EMBOSS program needle (39). The copy number of an amplified sequence in the Houston-1 genome was defined as the number of BLAST hits with lengths of more than 100 bp and at least 80% sequence identity, or a global alignment identity of >75%, based on previous reports on hybridization specificity (9, 11, 15, 18, 35, 53) and the control hybridizations with *B. quintana*. The results for sequences with a copy number higher than one were considered uncertain since a low M value could be due to a lower copy number rather than total absence of the sequence in the tested strain.

In addition, the experiments showed that the Houston-1 strain used as a reference for the microarray hybridizations (here called Aref) was different in two regions from the Houston-1 genome sequence, which was used for primer design and PCR amplification. A 10-kb region (bases 1159960 to 1169928 in the genome sequence, GenBank BX897699) was found to be duplicated in the Houston-1 (Aref) strain compared to the sequenced strain. In addition, the Houston-1 (Aref) strain was found to have a deletion of 10 kb (bases 204395 to 214491 in the genome sequence, corresponding to surface proteins). Both the deletion and the duplication were flanked by repeats and were confirmed by PCR and sequencing, as well as by a cohybridization of the two Houston-1 strains to the microarray (which revealed no further differences). PCR products in these regions were assigned copy numbers 0 and 2, respectively.

Absence/presence analysis. Normalized M values from four hybridization experiments, with each array containing six replicates were combined, by taking the mean value. Sequences were considered absent if the mean M value was less than −2 (49). At this cutoff level, the specificity and detection sensitivity were estimated to be 99.0 and 87.1%, respectively, for single-copy sequences in *B. quintana* with a sequence identity of ≥75% to the orthologous *B. henselae* sequences (see Results).

The sequence divergence for orthologs in *B. koehlerae* and *B. henselae* is slightly lower than for *B. quintana* and *B. henselae*; the mean substitution frequency at nonsynonymous sites is 0.018 for *B. koehlerae* and *B. henselae* compared to 0.023 for *B. quintana* and *B. henselae*, as inferred from a concatenated sequence consisting of the *ftsZ*, *gltA*, *groEL*, *ribC*, and *rpoB* genes. To adjust for this lower divergence level, it may be more appropriate to use a slightly lower cutoff level for *B. koehlerae* since the M values are normalized at 0 for the "typical" sequences. In *B. quintana*, an M value of 0 corresponds to ca. 88% sequence identity, with orthologous sequences of >96% identity having a median M value of 1.12. The corresponding orthologs in *B. koehlerae* have a median M value of only 0.41. From this comparison, we have estimated that a cutoff level of −2 in *B. quintana* should correspond to an M value of −2.7 in *B. koehlerae*.

Because the cutoff levels are set somewhat arbitrarily, we included estimates of gene absence based on M values of both −2 and −2.7 for *B. koehlerae* sequences, so as to reflect the upper and lower boundaries of the inferred gene absences. For genes represented by several PCR products, a majority decision was made based on the absence or presence designation of the individual single-copy sequences. PCR products derived from multicopy sequences were used only when no PCR products from single-copy sequences were available for a gene.

PCR verification. Attempts were made to verify the presence of genes in *B. koehlerae* that are absent in *B. quintana* with PCR using selected combinations of nondegenerate primers designed from the *B. henselae* genome sequence and used to generate the spotted PCR products (see Table S1 in the supplemental material).

Deposition of microarray data in the public databases. The microarray data have been deposited in the microarray database at EBI under the accession numbers E-MEXP-207 for the array design and E-MEXP-331 for the experimental data.

RESULTS

Estimated accuracy of the *Bartonella* genome hybridization experiments. To screen for regions that are putatively absent in strains and species that are closely related to *B. henselae*, we developed a microarray chip containing PCR products from 1367 genes and 112 noncoding regions of *B. henselae*. The genes included in the array represent 92% of the 1,491 genes identified in the *B. henselae* genome (1). The performance of the array was estimated through a series of control hybridizations with *B. quintana* DNA to the *B. henselae* microarray. The hybridization results were verified with the aid of the published genome sequence for the same *B. quintana* strain (1).

The identity for PCR-amplified genes that are positional orthologs in *B. henselae* and *B. quintana* ranged from 42.3 to 97.9%, with a mean of 86.8% and a median of 88.4%. For PCR-amplified sequences in *B. henselae* genes without a positional ortholog in *B. quintana*, the identity ranged from 34.6 to 77.7%, with a mean of 43.7% and a median of 42.5%. The control hybridizations showed that the M value (log ratio), which is a measure of the relative intensity of the signal, was approximately proportional to the percent sequence identity for sequences with >75% identity (Fig. 1). Here, sequence identity is defined as the number of identities divided by the length of a global alignment of the PCR-amplified sequence to the *B. quintana* genome in a region surrounding the first BLAST hit. Sequences with <75% identity were considered absent from *B. quintana* based on the distribution of sequence similarities for *B. henselae* genes with or without orthologs in *B. quintana*. Using a cutoff of 75%, which corresponds to an M value of −2 (Fig. 1), 1,215 sequences were considered present, whereas 431 sequences were considered absent in *B. quintana*.

A total of 1,428 successfully amplified sequences, representing 1,201 of the 1,491 genes and 59 intergenic regions, 19 pseudogenes, and 19 sequences bordering genes and intergenic regions had a copy number of 1. A total of 218 probe sequences representing 169 genes were present in the *B. henselae* Houston-1 genome more than once. Since cross-hybridization is known to confound the signals for repetitive sequences, a potential weakness is that single-copy genes in *B. quintana* and

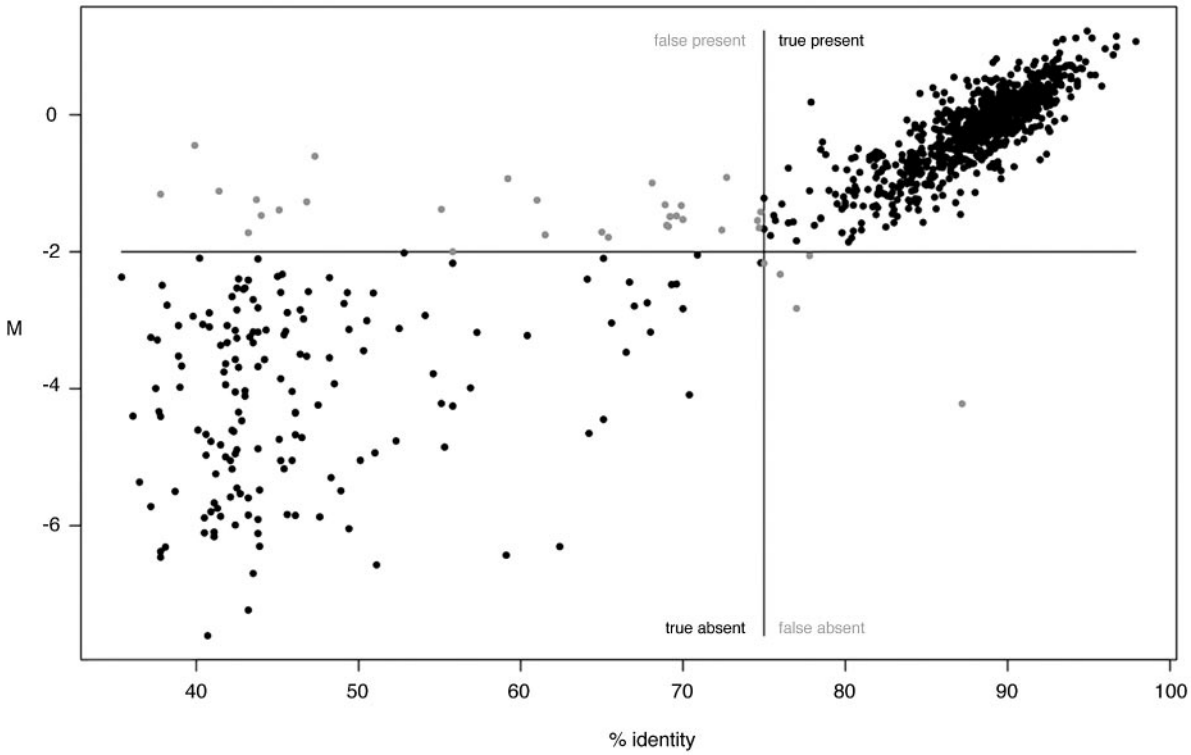

FIG. 1. Relationship between percent sequence identity in a global alignment of the *B. henselae* probe sequence to the *B. quintana* genome and the relative signal ($M = \log_2 BQ/\log_2 BH$) from control hybridizations of *B. quintana* versus *B. henselae*.

B. koehlerae that are represented by multicopy genes in *B. henselae* may be incorrectly classified as absent.

To estimate the severity of this limitation, we estimated the performance of the hybridization experiments separately for single- and multicopy sequences in *B. henselae*. Here, the specificity is defined as the number of PCR amplified sequences correctly detected on the array divided by the total number of such sequences considered present in the *B. quintana* genome. The sensitivity is defined as the number of PCR amplified sequences correctly reported as missing on the array divided by the total number of such sequences considered absent in the *B. quintana* genome.

For the single-copy sequences, we estimated the specificity to be 99.0% (1,136 correctly detected/1,147 classified as present by the 75% sequence criterion) and the detection sensitivity to be 87.1% (242 reported missing/278 classified as absent by the 75% sequence criterion). Only 52 of 65 sequences present in lower copy numbers in *B. quintana* were correctly recognized as present, corresponding to an average specificity for cross-hybridizing sequences of only 80%. The sensitivity for the cross-hybridizing sequences was 90.2%, with 138 probes correctly predicted to be absent out of the 153 multicopy probes missing from the *B. quintana* genome.

The single-copy sequences correctly recorded as present and absent on the array had average sequence identity values of 88.0 and 48.2%, respectively, whereas those incorrectly classified had average percent identity levels of 78.7% (falsely recorded as absent) and 60.3% (falsely recorded as present), i.e., the erroneous calls often represented borderline cases. PCR products giving very strong hybridization signals to the genomic DNA of *B. quintana* were found to correspond to sequences with atypically high levels of sequence identity to their homologs in *B. henselae*.

Genome size estimates of *B. koehlerae*. The genome size of *B. koehlerae* relative to *B. henselae* was estimated from hybridization experiments to the *B. henselae* custom microarray. The lengths of the segments reported as absent were estimated by defining the borders between absent and present sequences as the midpoints between the corresponding probes. Upper and lower boundaries in the classification of sequence absence were inferred by including estimates based on cutoff levels of both -2 and -2.7 in hybridizations with *B. koehlerae* DNA (see Materials and Methods).

Using these criteria, 164 to 226 kb of sequence data was estimated to be missing in *B. koehlerae* compared to 391 kb in *B. quintana*. Of the absent sequences in *B. quintana*, 177 to 234 kb generated a positive hybridization signal to the genomic DNA of *B. koehlerae*, whereas only 4 to 16 kb of the sequences reported to be missing in *B. koehlerae* were present in *B. quintana*. A more detailed comparison of the genome content of *B. koehlerae* to that of *B. henselae* revealed that most of the absent genes correspond to sequences in the GEIs or the chromosome II-like region in the *B. henselae* genome. Approximately 20 genes in other regions of the genome were also reported missing in *B. koehlerae*.

From these results, the size of the *B. koehlerae* genome was estimated to be 1.70 to 1.77 Mb compared to *B. henselae* and to 1.76 to 1.81 Mb compared to *B. quintana*, which should be considered a minimal estimate since genes uniquely present in *B. koehlerae* are not represented in the *B. henselae* microarray.

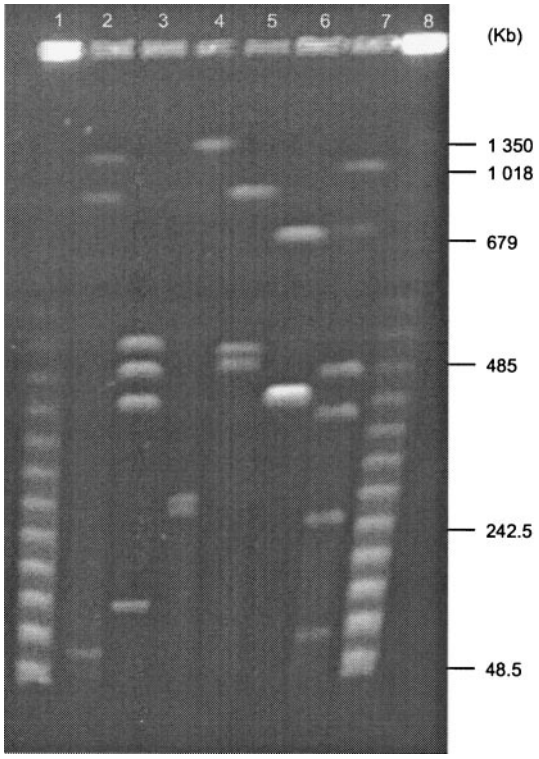

FIG. 2. PFGE-RFLP analysis of isolates of *Bartonella*. Lanes 1 and 8, DNA size standard (lambda ladder). Lanes 2 to 4 (digestion with AscI): lane 2, *B. koehlerae*; lane 3, *B. quintana* strain Toulouse; lane 4, *B. henselae* Houston 1-980517. Lanes 5 to 7 (digestion with NotI): lane 5, *B. koehlerae*; lane 6, *B. quintana* strain Toulouse; lane 7, *B. henselae* strain Houston 1-980517.

The genome size of *B. koehlerae* was also estimated to 1.7 to 1.8 Mb by PFGE analysis of DNA digested with the restriction enzymes AscI and NotI (Fig. 2). Thus, the estimates obtained from the microarray and PFGE analyses are in agreement and suggest that there are few, if any, unique sequences in *B. koehlerae*. We conclude that the *B. koehlerae* genome is intermediate in size (30) compared to the 1.58-Mb genome of *B. quintana* and the 1.93-Mb genome of *B. henselae* (1).

Sequence loss in the GEIs of *B. koehlerae*. Notable was the lack of hybridization of *B. koehlerae* to a majority of PCR products generated from the 57-kb prophage region (Fig. 3, left panel). Interspersed among putatively absent sequences in this region were 24 to 30 PCR products from 12 to 15 segments that yielded a positive hybridization signal, using M-value cutoffs of −2 and −2.7, respectively. However, several of these gave false-positive signals to the *B. quintana* DNA. Others were present in *B. henselae* in multiple copies, indicating that paralogs of these genes may be present in the phage region or elsewhere in the *B. koehlerae* genome. This suggests that the prophage is probably missing from the *B. koehlerae* genome but as in *B. quintana* remnants of phage genes may still be present in the genome.

In addition, *B. koehlerae* lacks most of the 72-kb genomic island (HGIa) in *B. henselae* (Fig. 3, left side). The missing genes include three copies of the *fhaC/hecB* and *fhaB* gene clusters that code for a two-partner secretion system where the transporter encoded by *fhaC/hecB* mediates the secretion of filamentous hemagglutinin encoded by *fhaB*. Each of the *fhaCB* gene clusters is flanked on one side by a gene that is repeated 20 times in the *B. henselae* genome; also, this gene yielded no hybridization signal to the genomic DNA of *B. koehlerae*. The middle segment of the genomic island HGIa, which contains genes of unknown function as well as genes coding for phage proteins, yielded a positive hybridization signal to *B. koehlerae*, but not to *B. quintana* consistent with the absence of this segment in the *B. quintana* genome (Fig. 3, left panel).

Of the 30 genes located within this segment, 14 are duplicated and form part of the 3′ end of the prophage in *B. henselae*. The hybridization experiments suggest that these are present as single-copy genes in *B. koehlerae*. Since all of the 30 genes in the 72-kb island of *B. henselae* gave a positive hybridization signal to *B. koehlerae*, it seems most likely that part of the island HGIa was retained in this species and that the entire prophage region, including the duplicated segment, was lost. Interestingly, the retained segment is flanked on one side by an integrase gene in *B. henselae*, a finding indicative of a phage integration event.

Approximately 20 short genes of unknown function at the 5′-terminal end of the 34-kb *B. henselae* genomic island (HGIb) were not identified, although another 30 genes at the 3′ end of the island, most of which are *Bartonella* specific, yielded a strong positive signal to the *B. koehlerae* genome (Fig. 3, left panel). In this case, a tRNA gene is located in the *B. henselae* genome at position corresponding to the downstream border of the segment that is missing in *B. koehlerae*. One of the smaller genomic islands (HGIc, 9 kb in size) contains a fourth *fhaC/hecB* gene, as well as disrupted fragments of the *fhaB* gene in *B. henselae*, and appears to be entirely missing from the *B. koehlerae* genome (Fig. 3, left panel).

However, since the *fhaBC* genes are multicopy sequences for which the specificity was estimated to be only 80%, it is not excluded that the *fhaBC* gene cluster is present as a single-copy gene and incorrectly predicted to be absent in the *B. koehlerae* genome. Partial retention of the prophage, as well as the genomic islands, lends additional support to the hypothesis that *B. koehlerae* is a genomic intermediate of *B. henselae* and *B. quintana*, corroborating that the absence of the prophage in *B. quintana* is due to gene deletion rather than to an infection in *B. henselae* after its divergence from the other two species.

Sequence divergence and loss of *Bartonella*-specific genes in the chromosome II-like segment. Overall, 75 and 98% of the 128 coding sequences in the chromosome II-like segment yielded a positive hybridization signal in *B. quintana* and *B. koehlerae*, respectively (Fig. 4a). A search for genes with annotated gene functions and host-specific hybridization patterns (i.e., present in *B. henselae* and *B. koehlerae* but absent from *B. quintana*) suggested five candidate genes at an M-value cutoff of −2; however, only one was considered significant after a more rigorous inspection.

Of these five genes, two were erroneously reported to be missing in the *B. quintana* genome because of low levels of sequence identity, 72 to 73%, which is below our cutoff level for gene presence. In another two cases the hybridization signals were weak and no PCR products were obtained, suggesting that the genes have been lost also from *B. koehlerae*. This

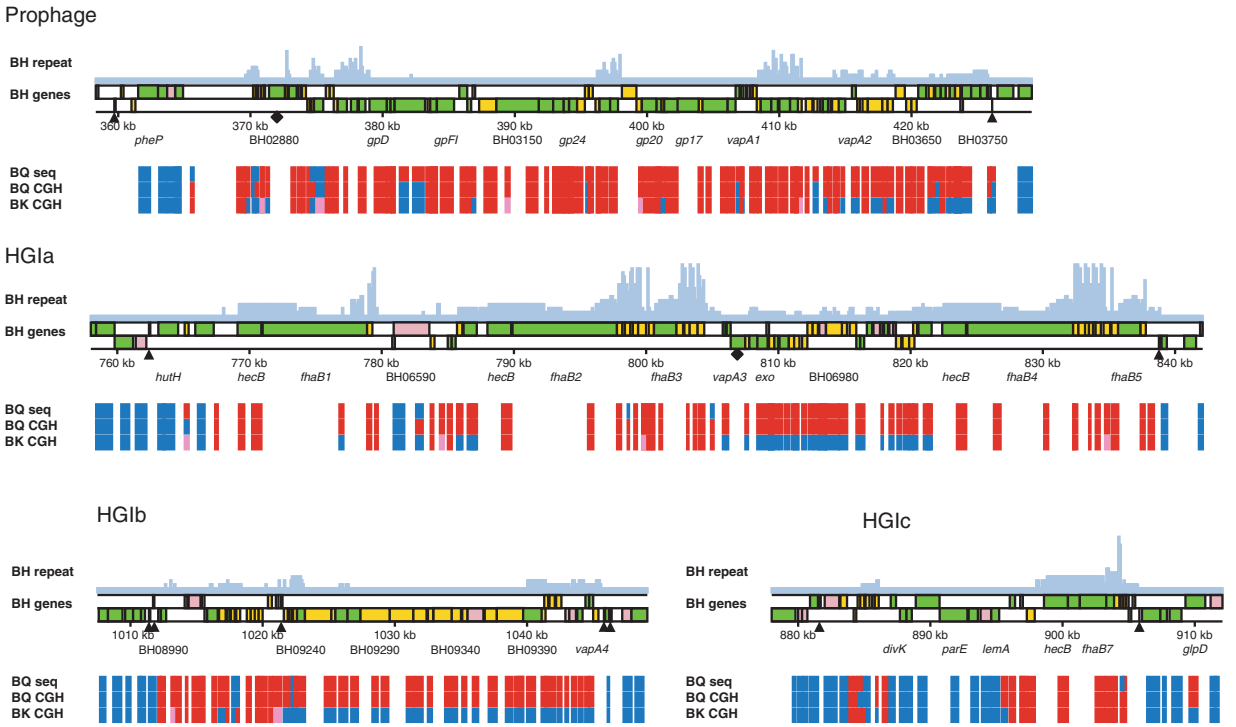

FIG. 3. Regions missing from *B. koehlerae* in the *B. henselae* prophage, the genomic islands HGIa, HGIb, and HGIc (left panel) and the chromosome II-like region (chrII) (right panel). Both panels show, from the top: repeat number of *B. henselae* sequences (the number at which each sequence occurs in the genome with more than 80% sequence identity over 100 bp); the *B. henselae* genes on "+" and "−" strands (green, annotated gene; pink, hypothetical conserved; yellow, *Bartonella* specific); expected array results for *B. quintana* (red, absent [<75% identity]; blue, present [>75% identity]); obtained array results for *B. quintana* (red, absent [M < -2]; blue, present [M > -2]); and obtained array results for *B. koehlerae* (red, absent [M < -2.7]; blue, present [M > -2.7]; pink, uncertain [M between -2.7 and -2]). In the right panel, the gene order in *B. quintana* is shown above the *B. henselae* genes. Orthologous sequences are highlighted by vertical lines between genes in *B. henselae* and *B. quintana*. Triangles show the position of tRNA genes and rhombi the position of integrase genes.

left one gene, *parA*1, which codes for a plasmid partitioning protein, as putatively present in *B. koehlerae* (Fig. 3, right panel), which was also confirmed by a PCR analysis. However, the *parA*1 gene in *B. henselae* is shorter than normal and is not flanked by a homolog to *parB*, as are most other *parA* genes, indicating that it may be a pseudogene. Thus, it is highly unlikely that the *parA1* gene per se is involved in host specificity.

However, this entire region may stem from an integrated plasmid, in which the *parA*1 gene mediated the partitioning of the ancestral plasmid or replicon. Interestingly, the genes flanking the *parA* gene display variable retention profiles (Fig. 3, right panel). These include seven genes in *B. henselae* and four genes in *B. quintana* that encode *Bartonella*-specific effector proteins BepA-BepG that are translocated into the host cell via the recently described type IV secretion system VirB/VirD4 (45, 47). Positive hybridization results were obtained for six of the *bep* genes in *B. koehlerae* compared to only one gene in *B. quintana*. The lack of hybridization signals in *B. quintana* is partially an effect of high sequence divergence levels, as inferred from a K_a value (substitution frequency at nonsynonymous sites) of 0.31 on the average compared to a K_a value of only 0.042 on the average for the genes *virB2-4*, *virB6*, and *virB8-11*. The substitution frequencies at synonymous sites (K_s) are similar for the *virB* and the *bep* genes (ranging from 0.40 to 0.82), with a K_a/K_s ratio of 0.107 for the *virB* genes and 0.44 for the *bep* genes, on average.

Upstream of the *parA* gene is another stretch of *Bartonella*-specific genes, most of which are truly absent in *B. quintana* but reported to be present in *B. koehlerae* by the microarray analysis (Fig. 3, right panel). This region contains several large noncoding spacer sequences in *B. quintana*, a finding indicative of recent gene degradation. Both of these two regions are solely present in the feline-associated species. Finally, we obtained conflicting hybridization signals for a few additional genes, such as the multiple *trwL* genes. This is probably due high copy numbers in *B. henselae* and high levels of sequence divergences for orthologs (48).

Sequence divergence and loss of noncoding DNA. Species-specific gene retention patterns may not solely be attributed to different functional constraints but may also reflect species-specific fixation rates for deletions. Of the 48 noncoding sequences in the chromosome II-like region, 52 to 71% yielded a positive hybridization signal in *B. koehlerae* compared to only 13% in *B. quintana* (Table 1 and Fig. 4). Most of the missing noncoding sequences in *B. koehlerae* were also considered to be absent in the *B. quintana* genome, with additional sequences reported missing in *B. quintana*. To distinguish the elimination of noncoding sequences from false negatives caused by highly diverged sequences, we examined the positions, lengths, and hybridization signals for each of the 71 spacers represented on the array.

Of the 54 spacers in all classified as absent in *B. quintana* by

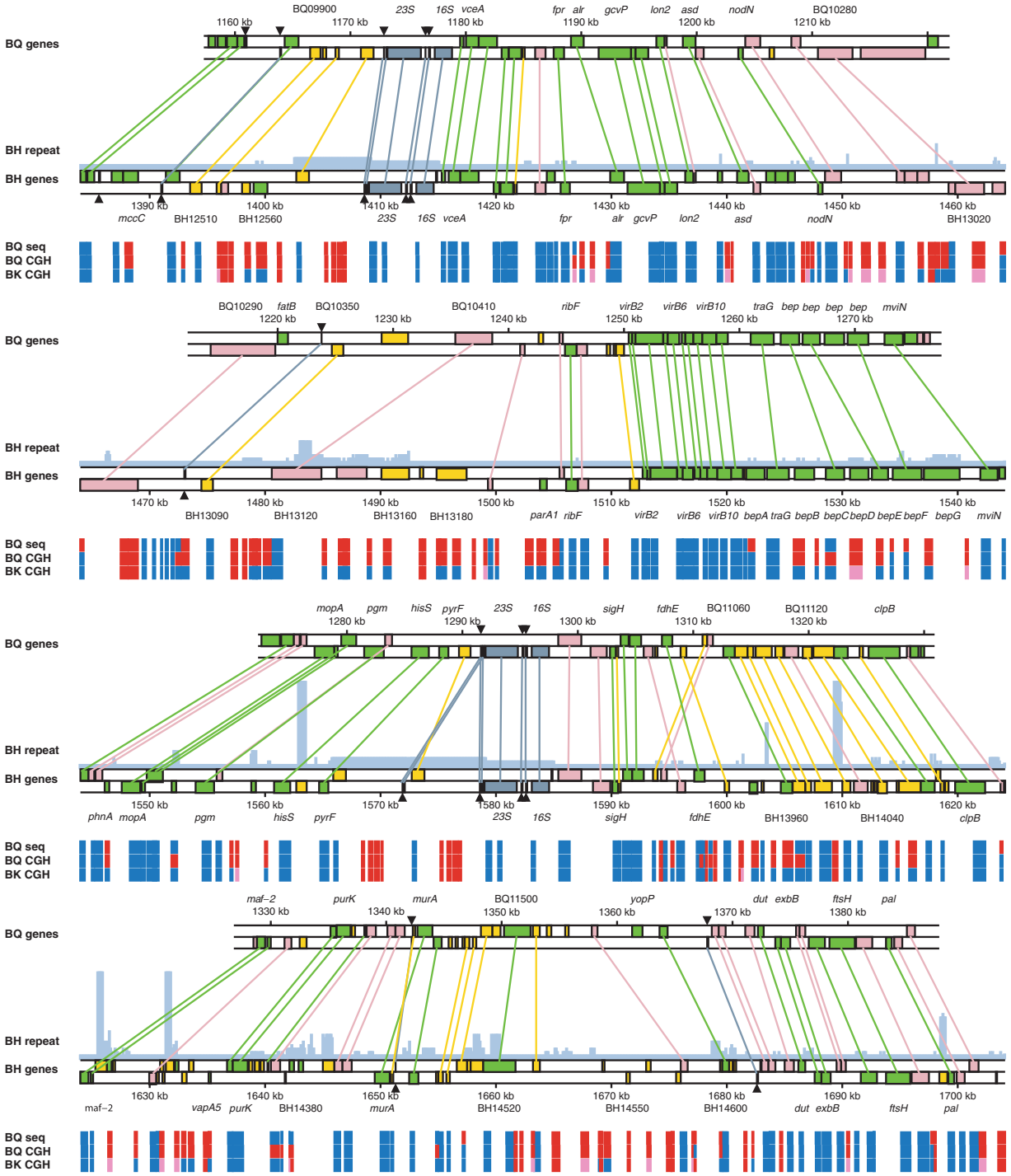

FIG. 3—*Continued*.

the microarray analysis, orthologs could be identified for the flanking genes in 14 cases. From the remaining 40 spacers, 25 are located in the chromosome II-like region and were lost due to deletions or rearrangements involving associated coding sequences. The mean lengths of the 14 "absent" spacers flanked by orthologous genes (only three of which are located in the chromosome II-like region) were 784 bp in *B. quintana* and 1,780 bp in *B. henselae*. The difference in size indicates that these spacers represent deteriorating sequences and that the rate of sequence elimination is much higher in *B. quintana* than in *B. henselae*.

Of the 17 spacers classified as present in *B. quintana* by the array, orthologs could be identified for the flanking genes in 13 cases. As expected, the difference in size between the two species was smaller for the present spacers with a mean length of 659 bp in *B. quintana* compared to a mean size of 877 bp in

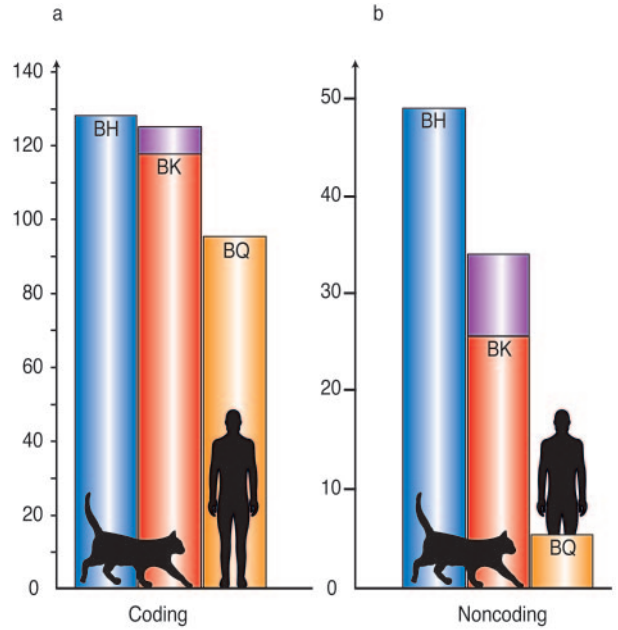

FIG. 4. Content of coding (a) and noncoding (b) DNA in the chromosome II-like region of *B. henselae* (BH), *B. koehlerae* (BK), and *B. quintana* (BQ). Upper and lower levels for *B. koehlerae* show the estimated number sequences classified as present based on M cutoff values of −2.7 and −2, respectively.

B. henselae. The divergences for *B. henselae* and *B. quintana* spacers reported to be present are on the average 60% identical over the entire spacer sequences or 74% identical over the array probes. In comparison, the sequence identity for spacers flanked by orthologs and reported to be absent by the array was only 37% on the average over the entire spacer sequences or 41% over the array probes. We speculate that the most highly diverged spacers represent the remains of longer sequences that are deteriorating differently in the three species. In effect, nonorthologous sequences will be located at homologous positions, leading to negative hybridization results and erroneously high estimates of sequence loss.

DISCUSSION

In this study, we have characterized the genome content of *B. koehlerae* by hybridization to a microarray containing PCR products amplified from the 1.93-Mb *B. henselae* genome. Since *B. koehlerae* has one of the smallest genomes of isolates from feline blood samples (30; the present study), we anticipated this analysis to reveal the full extent of sequence loss in bacterial populations adapted to hemotropism in feline hosts. Thus, the results obtained should help define the natural variation for zoonotic agents associated with coincidental infections in humans.

Our study identified the GEIs and the chromosome II-like region as the most variable segments among the three species here analyzed, with *B. koehlerae* showing an intermediate retention pattern of both coding and noncoding sequences. Uniquely present in both of the feline pathogens *B. koehlerae* and *B. henselae* are *Bartonella*-specific genes of unknown function in HGIb. Uniquely absent in both *B. koehlerae* and *B. quintana* is the prophage, a majority of genes in HGIa and HGIc, along with the 5′-terminal segment of HGIb, including genes for filamentous hemagglutinin, and many sequences in the chromosome II-like region.

When and how did these different genome variants arise? Since *B. koehlerae* is a sister clade with *B. henselae* to the exclusion of *B. quintana* (16, 19, 24, 26, 54), the absence of a shared set of genes in *B. koehlerae* and *B. quintana* is most readily explained by independent excision events mediated by integrases. Given that more than 40 integrase gene remnants are present in the *B. henselae* genome (1), many functional integrases were presumably encoded by the ancestors of both *B. quintana* and *B. koehlerae*, allowing independent excisions and transfers of GEIs to other recipients.

TABLE 1. Preferential loss of pseudogenes, intergenic DNA, and genes of unknown function in *B. koehlerae* and *B. quintana* in the chromosome II-like region (rRNA-tRNA)

Sequence type	No. of genes[a]					
	Single copy			Cross-hybridizing		
	B. henselae	*B. koehlerae*	*B. quintana*	*B. henselae*	*B. koehlerae*	*B. quintana*
Annotated	53	53–53	48	7	4–7	3
Conserved hypo[b]	25	23–24	20	5	5–5	3
Bartonella specific[c]	33	29–31	20	3	2–3	0
r/tRNA	0	0–0	0	2	2–2	4
Pseudogenes	5	3–3	3	6	1–3	0
Intergenic DNA	25	14–20	1	6	5	2
Mix[d]	3	0–1	0	3	2	0
Coding DNA	111	105–108	88	17	13–17	8
Noncoding DNA	33	17–24	3	15	8–10	2
Total	144	122–132	92	32	21–27	10

[a] The data are shown separately for sequences represented by single-copy and cross-hybridizing sequences in the 274-kb segment flanked by an rRNA and a tRNA-Leu gene in the *B. henselae* genome.
The lower and upper estimates in *B. koehlerae* are based on M-value cutoff levels of −2 and −2.7, respectively.
The number of sequences classified as present by microarray hybridization in *B. quintana* is based on M-value cutoff levels of −2.
[b] The conserved hypothetical genes have homologs in other species but their function is not known.
[c] The *Bartonella*-specific genes are only present in one or more of the *Bartonella* species.
[d] Mix, PCR products that cover both coding and noncoding sequences.

Alternatively, sequences reported to be "absent" from the *B. koehlerae* and *B. quintana* genomes may have been uniquely acquired in *B. henselae*. However, arguing against a recent phage integration event is the identification of phage remnants in *B. quintana* (1) and of prophage sequences in the genome of an outgroup species, *B. grahamii* (C. Frank and S. G. E. Andersson, unpublished data). Furthermore, a phylogenetic analysis of the phage lysozyme proteins indicate a complex evolutionary scenario with multiple, ancestral phage invasions followed by rearrangements (Frank and Andersson, unpublished). It is not excluded that some GEIs are the result of acquisition in *B. henselae*, followed by rearrangements, losses, and immobilization.

If we assume that most of the differences among the three species are caused by sequence loss, we may ask whether these occurred prior or subsequent to the host-species transmission event for *B. quintana*. Short transmission cycles among hosts due to the short generation time of lice could result in more generations per time unit for *B. quintana* and thereby lead to higher rates of sequence deterioration (1). However, if the cross-species transmission event of *B. quintana* coincided with events such as cat domestication (>5,000 years ago) or the divergence of the human body louse from the head louse (<100,000 years ago) (27), exceptionally high deletion rates have to be inferred to explain the reduced *B. quintana* genome solely as an effect of the short life cycle of lice.

Since we have shown in the present study that sequence loss also occurs in feline-associated *Bartonella* species, albeit to a lower extent than in *B. quintana*, it is perhaps more likely that an already-reduced clone entered the human-louse-human cycle or another enzoonotic cycle. Such reduced clones may arise frequently in the natural population of feline *Bartonella* species but, if less successful in establishing infections in the cat than their relatives with larger genomes, their frequency in the global pool of feline *Bartonella* species may be low, perhaps explaining the rare recovery of *B. koehlerae* from cats (16, 41) and their low frequency in cat fleas (42).

The different retention patterns of genes in the GEIs raise questions about their role in *Bartonella* spp. In many pathogens, GEIs have been shown to contribute virulence factors and thereby to play an important role in the development of disease (14, 46). However, since GEIs are missing from the human pathogen *B. quintana*, there is no obvious relationship between the presence of GEIs and virulence in humans in the case of *Bartonella* (1). Other human pathogens that also do not contain genomic islands include *Rickettsia*, *Chlamydia*, and *Mycobacterium* spp. (3–6, 9, 14). Furthermore, the presence or absence of GEIs in *Brucella* spp. does not mirror the virulence characteristics of the different strains; GEIs are present in one pathogenic species, *Brucella neotome*, but absent from another, *Brucella ovis* (38). All of these genera are obligate or facultative intracellular and infect a broad variety of hosts.

It is tempting to speculate that genes in the *Bartonella* GEIs play a role in host or vector specificities. However, we also observed no simple relationship between gene content and host adaptation patterns. Both *B. henselae* and *B. koehlerae* utilize the same host-vector system (42), and yet only very few genes with annotated functions are shared by the two species. Among those uniquely present in *B. henselae* and *B. koehlerae* are circa 20 genes located in HGIb and a stretch of genes of the *parA* gene in the chromosome II-like region, none of which encodes a known function.

Alternatively, host specificity may be conferred by sequence changes or gene expression differences of surface proteins. Putative candidates are proteins that mediate host cell binding, such as, for example, the highly repetitive nonfimbrial adhesin, designated *badA* in *B. henselae* (40) and *Vomp* in *B. quintana* (56). For this protein, expression level variation is achieved by deletions across the repeated sequences (56). Adhesins and other surface proteins encoded by multicopy genes with intragenic repeats may easily be modified to fit any given host cell structure.

In this context, it is interesting to recall that although most of the observed variation in the hybridization patterns reflects true differences in gene contents, the signals were occasionally lost because of high rates of sequence evolution. Such erroneous results were observed mostly for spacer sequences but also in some rare cases for very rapidly evolving genes, such as, for example, the *trwL* genes which encode the structural component of the pilus of the type IV secretion system (48). Inconsistent hybridization results were also observed for the *bep* genes, which encode translocated effector proteins known to mediate cellular changes after bacterial invasion, such as F-actin rearrangements, apoptotic cell death, proinflammatory response, and cytotoxic effects (47). To the extent that these negative results reflect adaptive sequence evolution in response to host specific receptors and structures, erroneous hybridization signals may prove to hide some of the most interesting genes from the perspective of host-pathogen interactions.

A third hypothesis is that the content and organization of the GEIs is associated with the persistence of the infection rather than with vector utilization and host specificity patterns per se. If sequences in the GEIs are important for the establishment of long-term bacteremia, a prediction is that GEIs should be present in species that cause chronic infections irrespective of whether the host is a cat or a human. On the other hand, if the emergence of human pathogens correlates with genome size reductions (31), we expect the GEIs to contain fewer genes or be absent from *Bartonella* species for which the human is the definitive host. To distinguish between these scenarios, the genomes of other feline and human *Bartonella* pathogens need to be determined.

One species of particular interest is *Bartonella clarridgeiae*, which is adapted to feline hosts but is phylogenetically unrelated to *B. henselae* and *B. koehlerae*. A recent study showed that 3.7 and 11.1% of cat fleas were infected with *B. koehlerae* and *B. henselae* compared to 67.9% with *B. clarridgeiae* (42). Another investigation of blood samples from domestic French cats suggested that 16.5% were *Bartonella* bacteremic, 65% of which were infected with *B. henselae* and 20% of which were infected with *B. clarridgeiae* (20). It should be of interest to find out whether GEIs are present in *B. clarridgeiae* and, if so, whether they confer a selective advantage to *Bartonella* strains and species that are successful colonizers of cats.

Another species that should be the focus of further examination is *Bartonella bacilliformis*, the agent of Carrions disease, which causes both acute and chronic infections. The acute phase of the disease, Oroya fever, has one the highest mortality rates of infectious diseases if left untreated. *B. bacilliformis* and

B. quintana are the only *Bartonella* species for which humans are the definitive host. The influence of genome sizes and structures on the transmission dynamics, host range, and duration of the infection can be substantiated by investigating the genomes of these additional *Bartonella* species and their intraspecies heterogeneity.

ACKNOWLEDGMENTS

We thank Carolin Frank and members of the European Network on Postgenomic Analysis of Human Pathogenic Bacteria for many interesting discussions.

This study was supported by the Swedish Research Council, the Swedish Foundation for Strategic Research, the Knut and Alice Wallenberg Foundation, and the European Union.

REFERENCES

1. **Alsmark, U. C. M., A. C. Frank, E. O. Karlberg, B. Legault, B. Canbäck, D. Ardell, A.-S. Eriksson, A. K. Näslund, S. Handley, M. Huvet, B. La Scola, M. Holmberg, and S. G. E. Andersson.** 2004. The louse-borne human pathogen *Bartonella quintana* is a genomic derivative of the zoonotic agent *Bartonella henselae.* Proc. Natl. Acad. Sci. USA **101:**9716–9721.
2. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
3. **Amiri, H., W. Davids, and S. G. E. Andersson.** 2003. Birth and death of orphan genes in *Rickettsia.* Mol. Biol. Evol. **20:**1575–1587.
4. **Andersson, S. G. E., A. Zomorodipour, J. O. Andersson, T. Sicheritz-Ponten, U. C. M. Alsmark, R. M. Podowski, A. K. Näslund, A.-C. Eriksson, H. H. Winkler, and C. G. Kurland.** 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature **396:**133–140.
5. **Andersson, J. O., and S. G. E. Andersson.** 1999. Genome degradation is an ongoing process in *Rickettsia.* Mol. Biol. Evol. **16:**1178–1191.
6. **Andersson, J. O., and S. G. E. Andersson.** 2001. Pseudogenes, junk DNA, and the dynamics of *Rickettsia* genomes. Mol. Biol. Evol. **18:**829–839.
7. **Avidor, B., M. Graidy, G. Efrat, C. Leibowitz, G. Shapira, A. Schattner, O. Zimhony, and M. Giladi.** 2004. *Bartonella koehlerae*, a new cat-associated agent of culture-negative human endocarditis. J. Clin. Microbiol. **42:**3462–3468.
8. **Bloomfield, G. A., G. Whittle, M. B. McDonagh, M. F. Katz, and B. F. Cheetham.** 1997. Analysis of sequences flanking the *vap* regions of *Dichelobacter nodosus:* evidence for multiple integration events, a killer system, and a new genetic element. Microbiology **143:**553–562.
9. **Brunelle, B. W., T. L. Nicholson, and R. S. Stephens.** 2004. Microarray-based genomic surveying of gene polymorphisms in *Chlamydia trachomatis.* Genome Biol. **5:**R42.
10. **Cornillot, E., C. Croux, and P. Soucaille.** 1997. Physical and genetic map of *Clostridium acetobutylicum* ATCC 824 chromosome. J. Bacteriol. **179:**7426–7434.
11. **Cummings, C. A., M. M. Brinig, S. van de Pas, and D. A. Relman.** 2004. *Bordetella* species are distinguished by patterns of substantial gene loss and host adaptation. J. Bacteriol. **186:**1484–1492.
12. **Dehio, C.** 2004. Molecular and cellular basis of *Bartonella* pathogenesis. Annu. Rev. Microbiol. **58:**365–390.
13. **Deng, W., S. R. Liou, G. Plunkett III, G. F. Mayhew, D. J. Rose, V. Burland, V. Kodoyianni, and F. R. Blattner.** 2003. Comparative genomics of *Salmonella* enterica serovar Typhi strains Ty2 and CT18. J. Bacteriol. **185:**2330–2337.
14. **Dobrindt, U., B. Hochhut, U. Hentschel, and J. Hacker.** 2004. Genomic islands in pathogenic and environmental microorganisms. Nat. Rev. Microbiol. **2:**414–424.
15. **Dong, Y., J. D. Glasner, F. R. Blattner, and E. W. Triplett.** 2001. Genomic interspecies microarray hybridization: rapid discovery of three thousand genes in the maize endophyte, *Klebsiella pneumoniae* 342, by microarray hybridization with *Escherichia coli* K-12 open reading frames. Appl. Environ. Microbiol. **67:**1911–1921.
16. **Droz, S., B. Chi, E. Horn, A. G. Steigerwalt, A. M. Whitney, and D. J. Brenner.** 1999. *Bartonella koehlerae* sp. nov., isolated from cats. J. Clin. Microb. **37:**1117–1122.
17. **Dzieman, M., E. Balon, D. Boyd, C. M. Fraser, J. F. Heidelberg, and J. J. Mekalanos.** 2002. Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. Proc. Natl. Acad. Sci. USA **99:**1556–1561.
18. **Evertsz, E. M., J. Au-Young, M. V. Ruvolo, A. C. Lim, and M. A. Reynolds.** 2001. Hybridization cross-reactivity within homologous gene families on glass cDNA microarrays. BioTechniques **31:**1182–1186.
19. **Gundi, V. A., B. Davoust, A. Khamis, M. Boni, D. Raoult, and B. La Scola.** 2004. Isolation of *Bartonella rattimassiliensis* sp. nov. and *Bartonella phoceensis* sp. nov. from European *Rattus norvegicus.* J. Clin. Microbiol. **42:**3816–3818.
20. **Gurfield, A. N., H. J. Boulouis, B. B. Chomel, R. W. Kasten, R. Heller, C. Bouillin, C. Gandoin, D. Thibault, C. C. Chang, F. Barrat, and Y. Piemont.** 2001. Epidemiology of *Bartonella* infection in domestic cats in France. Vet. Microbiol. **21:**185–198.
21. **Hacker, J., G. Blum-Oehler, I. Muhldorfer, and H. Tschape.** 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol. Microbiol. **23:**1089–1097.
22. **Hakenbeck, R., N. Balmelle, B. Weber, C. Gardés, W. Keck, and A. de Saizieu.** 2001. Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of *Streptococcus pneumoniae.* Infect. Immun. **69:**2477–2486.
23. **Hinchliffe, S. J., K. E. Isherwood, R. A. Stabler, M. B. Prentice, A. Rakin, P. C. Oyston, J. Hinds, R. W. Titball, and B. W. Wren.** 2003. Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis.* Genome Res. **13:**2018–2029.
24. **Houpikian, P., and D. Raoult.** 2001. Molecular phylogeny of the genus *Bartonella*: what is the current knowledge? FEMS Microbiol. Lett. **200:**1–7.
25. **Israel, D. A., N. Salama, U. M. Krishna, U. Rieger, J. C. Atherton, S. Falkow, M. Richard, and J. Peek.** 2001. *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. Proc. Natl. Acad. Sci. USA **98:**14625–14630.
26. **Jacomo, V., P. J. Kelly, and D. Raoult.** 2002. Natural history of *Bartonella* infections (an exception to Koch's postulate). Clin. Diagn. Lab. Immunol. **9:**8–18.
27. **Kittler, R., M. Kayser, and M. Stoneking.** 2003. Molecular evolution of *Pediculus humanus* and the origin of clothing. Curr. Biol. **13:**1414–1417.
28. **Leung, Y. F., and D. Cavalieri.** 2003. Fundamentals of cDNA microarray data analysis. Trends Genet. **19:**649–659.
29. **Marokhazi, J., N. Waterfield, G. LeGoff, E. Feil, R. Stabler, J. Hinds, A. Fodor, and R. H. ffrench-Constant.** 2003. Using a DNA microarray to investigate the distribution of insect virulence factors in strains of photorhabdus bacteria. J. Bacteriol. **185:**4648–4656.
30. **Maruyama, S., R. W. Kasten, H. J. Boulouis, N. A. Gurfield, Y. Katsube, and B. B. Chomel.** 2001. Genomic diversity of *Bartonella henselae* isolates from domestic cats from Japan, the USA and France by pulsed-field gel electrophoresis. Vet. Microbiol. **19:**337–349.
31. **Maurelli, A. T., R. E. Fernandez, C. A. Bloch, C. K. Rode, and A. Fasano.** 1998. "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli.* Proc. Natl. Acad. Sci. USA **95:**3943–3948.
32. **Needleman, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48:**443–453.
33. **Ogata, H., S. Audic, P. Renesto-Audiffren, P. E. Fournier, V. Barbe, D. Samson, V. Roux, P. Cossart, J. Weissenbach, J. M. Claverie, and D. Raoult.** 2001. Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii.* Science **293:**2093–2098.
34. **Parkhill, J., M. Sebaihia, A. Preston, L. D. Murphy, N. Thomson, D. E. Harris, M. T. Holden, C. M. Churcher, S. D. Bentley, K. L. Mungall, A. M. Cerdeno-Tarraga, L. Temple, K. James, B. Harris, M. A. Quail, M. Achtman, R. Atkin, S. Baker, D. Basham, N. Bason, I. Cherevach, T. Chillingworth, M. Collins, A. Cronin, P. Davis, J. Doggett, T. Feltwell, A. Goble, N. Hamlin, H. Hauser, S. Holroyd, K. Jagels, S. Leather, S. Moule, H. Norberczak, S. O'Neil, D. Ormond, C. Price, E. Rabbinowitsch, S. Rutter, M. Sanders, D. Saunders, K. Seeger, S. Sharp, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, L. Unwin, S. Whitehead, B. G. Barrell, and D. J. Maskell.** 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica.* Nat. Genet. **35:**32–40.
35. **Porwollik, S., R. M.-Y. Wong, and M. McClelland.** 2002. Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. Proc. Natl. Acad. Sci. USA **99:**8956–8961.
36. **Quackenbush, J.** 2002. Microarray data normalization and transformation. Nat. Genet. **32:**496–501.
37. **Raddatz, G., M. Dehio, T. F. Meyer, and C. Dehio.** 2001. PrimeArray: genome-scale primer design for DNA-microarray construction. Bioinformatics **17:**98–99.
38. **Rajashekara, G., J. D. Glasner, D. A. Glover, and G. A. Splitter.** 2004. Comparative whole-genome hybridization reveals genomic islands in *Brucella* species. J. Bacteriol. **186:**5040–5051.
39. **Rice, P., I. Longden, and A. Bleasby.** 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. **16:**276–277.
40. **Riess, T., S. G. E. Andersson, A. Lupas, M. Schaller, A. Schafer, P. Kyme, J. Martin, J. H. Walzlein, U. Elehalt, H. Lindroos, M. Schirle, A. Nordheim, I. B. Autenrieth, and V. A. Kempf.** 2004. *Bartonella* adhesin A mediates a proangiogenic host cell response. J. Exp. Med. **200:**1267–1278.
41. **Rolain, J.-M., P. E. Fournier, D. Raoult, and J. J. Bonerandi.** 2003. First isolation and detection by immunofluorescence assay of *Bartonella koehlerae* in erythrocytes from a French cat. J. Clin. Microbiol. **41:**4001–4002.
42. **Rolain, J.-M., M. Franc, B. Davoust, B., and D. Raoult.** 2003. Molecular

detection of *Bartonella quintana*, *B. koehlerae*, *B. henselae*, *B. clarridgeiae*, *Rickettsia felis*, and *Wolbachia pipientis* in cat fleas, France. Emerg. Infect. Dis. **9:**338–342.

43. **Rozen, S., and H. J. Skaletsky.** 1998. Primer3. [Online.] http://www-genome.wi.mit.edu/genome_software/other/primer3.html.
44. **Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and S. Falkow.** 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. Proc. Natl. Acad. Sci. USA **97:**14688–14693.
45. **Schmid, M. C., R. Schulein, M. Dehio, G. Denecker, I. Carena, and C. Dehio.** 2004. The *virB* type IV secretion system of *Bartonella henselae* mediates invasion, proinflammatory activation and antiapoptotic protection of endothelial cells. Mol. Microbiol. **52:**81–92.
46. **Schmidt, H., and M. Hensel.** 2004. Pathogenicity islands in bacterial pathogenesis. Clin. Microbiol. Rev. **17:**14–56.
47. **Schulein, R., P. Guye, T. Rhomberg, M. C. Schmid, G. Schröder, A. C. Vergunst, I. Carena, and C. Dehio.** 2005. A bipartite signal mediates the transfer of type IV secretion substrates of *Bartonella henselae* into human cells. Proc. Natl. Acad. Sci. USA **102:**856–861.
48. **Seubert, A., R. Hiestand, F. de la Cruz, and C. Dehio.** 2003. A bacterial conjugation machinery recruited for pathogenesis. Mol. Microbiol. **49:**1253–1266.
49. **Storey, J. D., and R. Tibshirani.** 2003. Statistical significance for genome-wide studies. Proc. Natl. Acad. Sci. USA **100:**9440–9445.
50. **Tian, Q. B., M. Ohnishi, A. Tabuchi, and Y. Terawaki.** 1996. A new plasmid-encoded proteic killer gene system: cloning, sequencing, and analyzing *hig* locus of plasmid Rts1. Biochem. Biophys. Res. Commun. **220:**280–284.
51. **Wei, J., M. B. Goldberg, V. Burland, M. M. Venkatesan, W. Deng, G. Fournier, G. F. Mayhew, G. Plunkett III, D. J. Rose, A. Darling, B. Mau, N. T. Perna, S. M. Payne, L. J. Runyen-Janecky, S. Zhou, D. C. Schwartz, and F. R. Blattner.** 2003. Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. Infect. Immun. **71:**2775–2786. (Erratum, Infect. Immun. **71:**4223.)
52. **Wolfgang, M. C., B. R. Kulasekara, XLiang, D. Boyd, K. Wu, Q. Yang, C. G. Miyada, and S. Lory.** 2003. Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa.* Proc. Natl. Acad. Sci. USA **100:**8484–8489.
53. **Wu, L., D. K. Thompson, G. Li, R. A. Hurt, J. M. Tiedje, and J. Zhou.** 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl. Environ. Microbiol. **67:**5780–5790.
54. **Zeaiter, Z., Z. Liang, and D. Raoult.** 2002. Genetic classification and differentiation of *Bartonella* species based on comparison of partial *ftsZ* gene sequences. J. Clin. Microbiol. **40:**3641–3647.
55. **Zhang, C., M. Zhang, J. Ju, J. Nietfeldt, J. Wise, P. M. Terry, M. Olson, S. D. Kachman, M. Wiedmanna, M. Samadpour, and A. K. Benson.** 2003. Genome diversification in phylogenetic lineages I and II of *Listeria monocytogenes*: identification of segments unique to lineage II populations. J. Bacteriol. **185:**5573–5584.
56. **Zhang, P., B. B. Chomel, M. K. Schau, J. S. Goo, S. Droz, K. L. Kelminson, S. S. George, N. W. Lerche, and J. E. Koehler.** 2004. A family of variably expressed outer-membrane proteins (Vomp) mediates adhesin and autoaggregation in *Bartonella quintana.* Proc. Natl. Acad. Sci. USA **101:**13630–13635.