Journal of Bacteriology and Mycology

Open Access

Austin Publishing Group

Research Article

The Pan-Genome of the Zoonotic Neglected Pathogen *Bartonella henselae* Reveals Two Groups with Different Patterns of Adaptation to Hosts

Alves LG[1], Oliveira LC[2], Benevides LJ[3], Zen FL[2], Tiwari S[1], Jaiswal A[1], Ghosh P[4], Barh D[5], Azevedo V[1] and Soares SC[2]*

[1]Department of Microbiology, Immunology and Parasitology, Institute of Biological Sciences, Federal University of Minas Gerais, Brazil

[2]Department of General Biology, Institute of Biological Sciences and Natural Sciences, Federal University of Triângulo Mineiro, Brazil

[3]Bioinformatics Laboratory, National Laboratory for Scientific Computing, Brazil

[4]Department of Computer Science, Virginia Commonwealth University, USA

[5]Center for Genomics and Applied Gene Technology, Institute of Integrative Omics and Applied Biotechnology (IIOAB), India

***Corresponding author:** Soares SC, Department of General Biology, Institute of Biological Sciences and Natural Sciences, Federal University of Triângulo Mineiro, 30 Frei Paulino St, Uberaba, MG, 38025180, Brazil

Received: May 15, 2020; **Accepted:** June 04, 2020; **Published:** June 11, 2020

Abstract

The genus *Bartonella* is comprised of Gram-negative re-emerging bacteria like *Bartonella henselae*, which mainly infects humans and survives inside erythrocytes. This species is transmitted by scratches and bites from domestic cats and usually causes a symptomatic infection in humans, known as Cat-Scratch Disease (CSD). The disease causes multiple clinical signs in humans, such as dermatic, cardiovascular, lymphatic, hepatic and nervous system diseases in immunosuppressed individuals. Although the bacteria are highly relevant for its zoonotic importance worldwide, few studies aimed at characterizing these species genomes and there is still no pan-genome study available. Here, we performed phylogenomic, pan-genome and genome plasticity analyses to determine the epidemiological aspects, the size of the pan-genome and its variability in the identified pathogenicity and resistance islands. Altogether, our results showed that the genomes are highly similar, with an almost closed pan-genome. Also, we found two subsets of genomes, composed of 7 and 17 genomes of bacteria. Our results point to the need of sequencing more genomes worldwide to better characterize these variations in the pan-genome and understand the patterns of adaptation of this species. The highly conserved genomes from this species are very important for the development of new vaccines and analyses of drug targets against this pathogen. Furthermore, these data may then be used in future works, which will be highly relevant for containing the disease worldwide.

Keywords: *Bartonella henselae*; Cat-Scratch Disease; Pan-Genome; Genome Plasticity; Phylogenomics

Abbreviations

ACT (Artemis Comparison Tool); *B. apis* (*Bartonella apis*); *B. henselae* (*Bartonella henselae*); BRIG (BLAST Ring Image Generator); CDSs (Coding DNA Sequences); COG (Cluster of Orthologous Groups); CSD (Cat-Scratch Disease); DNA (deoxyribonucleic acid); GC-content (guanine-cytosine content); GIPSy (Genomic Island Prediction Software); LCBs (Locally Collinear Blocks); Mb (mega base pairs); kb (kilo base pairs); MCL (Markov Clustering); NCBI (National Center for Biotechnology Information); PAIs (Pathogenicity Islands); RIs (Resistance Islands); T4SS (type IV secretion system); USA (United States of America); VFDB (Virulence Factor Database); wgMLST (whole genome Multilocus Sequence Typing).

Introduction

The genus *Bartonella* is comprised of Gram-negative, fastidious, intracellular and reemerging bacteria [1]. Some species from this genus mainly infect cats, but is also an opportunistic pathogen for humans and are frequently acquired through hematophagous arthropod vectors. Those species are able to infect and survive inside erythrocytes, through a long intra-erythrocytic and intra-endothelial infection, which results in recidivist bacteremia both in humans and other mammals [2]. The genus presents at least 13 human pathogenic species [3], among which the 3 most relevant ones being: *Bartonella bacilliformis*, which causes the Oroya fever [4]; *Bartonella quintana*, causing trench fever [5]; and, *Bartonella henselae*, the causative agent of the Cat-Scratch Disease (CSD).

Domestic cats are the main reservoirs of *B. henselae* and both immunocompetent and immunosuppressed cats are frequently asymptomatic to the infection, though they can present subclinical infections and suffer from recidivist bacteremia through their lives [6]. *B. henselae* transmission between cats occurs through arthropod vectors, mainly fleas [7] and dogs are also potential reservoirs of *B. henselae* [8]. The transmission route between humans and cats occurs mainly through the dermic inoculation of the bacteria through scratches or bites of contaminated cats [9]. However, other studies show *B. henselae* infections of immunocompetent humans through the possible inoculation by ticks and spider bites [10,11]. In the bloodstream, the bacteria invade erythrocytes, where they persist intracellularly, causing erythrocytic and vascular endothelial alterations [12].

B. henselae infection in humans causes multiple clinical signs, such as dermatic, cardiovascular, lymphatic, hepatic and nervous system diseases [13]. CSD is characterized by circumscribed regional lymphadenopathy in the inoculation site and previous studies revealed an immune-dependent pattern of clinical manifestations of

Citation: Alves LG, Oliveira LC, Benevides LJ, Zen FL, Tiwari S, Jaiswal A, et al. The Pan-Genome of the Zoonotic Neglected Pathogen *Bartonella henselae* Reveals Two Groups with Different Patterns of Adaptation to Hosts. J Bacteriol Mycol. 2020; 7(3): 1133.

CSD, mainly in immunosuppressed humans [14].

Although the bacteria present a high veterinary, medical, and zoonotic importance, there are currently few studies about the genomic profile of the species and the few existing ones aim at characterizing specific genes and analyzing the recombination and mutation rates [15-18]. Previous studies revealed the possibility of a horizontal gene transfer of adaptability and virulence genes in the species from the genus *Bartonella* [19]. Altogether, the fact that *B. henselae* is distributed worldwide and considered as a neglected zoonotic pathogen, opens doors for new genome plasticity studies of the species. In this work, we performed phylogenomics, pan-genomics, and genome plasticity analyses to find the possible epidemiological relationships between the strains, the conserved and variable subsets of genes and the pathogenicity and antibiotic resistance islands that may be involved in the pathogenesis process.

Material and Methods

Genomes

We retrieved 24 genomes of different strains of *B. henselae* publicly available at the GenBank on NCBI (https://www.ncbi.nlm.nih.gov). The genomes were retrieved through FTP in .gbff, .fna and .faa file formats. The genome of *Bartonella apis* was used as a non-pathogenic reference where applicable.

Phylogenomic analyses

The software Gegenees was used to compare the percentage of similarity between the 24 genomes [20]. The .fna files containing the complete and draft genomes of *B. henselae* were imported to Gegenees and analyzed using the default parameters. Briefly, Gegenees fragments all genomes in pre-defined sizes; performs similarity analyses using BLASTn among the genomes to identify the commonly shared regions; and, finally, the software creates a heatmap with the percentage of similarity among the genomes. Here, we used a fragmentation size of 200 bp and a similarity threshold of 40%.

The similarity matrix generated by Gegenees was then exported in .nexus format and further used in the software SplitsTree [21] to create a neighbour-joining phylogenetic tree based on the similarity among the strains using the option Equal Angle.

For a greater resolution on the phylogenomic tree analysis, a whole genome Multilocus sequence typing (wgMLST) analysis [22] was performed *via* the online software PGAdb-builder [23]. The 24 genomes of *B. henselae*, in the .fasta format, were compared with the PGAdb profile through BLASTn, using the module Build_PGAdb. Next, the PGAdb profile of the genomes, in the .scheme format, was used to build a wgMLST tree using the module Build_wgMLSTtree, with a 90% coverage and 90% identity filter. After the BLASTn analyses, the output file .newick was exported and used as input file in the software MEGA7 [24] for the reconstruction of the phylogenomic tree.

Gene Synteny

The gene synteny analysis was performed using the software Mauve [25]. Briefly, Mauve fragments the genomes in pre-defined sizes, creates Locally Collinear Blocks (LCBs) using the sequence alignments and exports the results as a figure where the rearrangement events are represented. Here, we imported the .fna files from the 24 strains and used the Houston-1 strain as reference along with the "progressiveMauve" algorithm.

Pan-genomic analyses

The prediction of orthologous genes was performed with the software OrthoFinder [26] and further classified with the use of an in-house script in: core genome, containing only genes that are commonly shared by all strains; shared genome, containing genes that are shared by two or more strains, but not all; and singletons, with strain-specific genes. Briefly, OrthoFinder used the .faa files with the amino acid sequences of all Coding DNA Sequences (CDSs) in each genome to perform an *all-vs-all* BLASTp analysis. The sequences were then grouped using the Markov Clustering (MCL) algorithm to determine the orthologous genes [27].

Pan-genome and subsets development

The pan-genome, core genome and singletons development were calculated based on the mean values of the permutations of all genomes using the method described in Soares et al. [28]. The final curves were then fitted using an *in-house* script to estimate the fixed parameters for Heap's Law (pan-genome analyses) and least-squares fit of the exponential regression decay (core-genome and singletons). The extrapolations of the pan-genomes from the different datasets were calculated by curve fitting based on Heap's Law with the formula $n=\kappa *N^{-\alpha}$, where n is the expected number of genes for a given number of genomes, N is the number of genomes, and the other terms are constants defined to fit the curve. The extrapolations of the core genomes and singletons for all datasets were calculated by curve fitting based on least-squares fit of the exponential regression decay with the formula $n=\kappa*\exp[-x/\tau]+tg(\theta)$, where n is the expected subset of genes for a given number of genomes, x is the number of genomes, exp is Euler's number, and the other terms are constants defined to fit the curve.

We used 7 subsets of genes for all the analyses: the complete genome dataset; 2 subsets with genomes isolated from different hosts (cats and humans); two subsets of genomes from different locations (France and USA); and, two subsets containing the genomes of the two groups identified on Gegenees and PGAdb (genomes 2-8 and 9-25 on Gegenees). The group 2-8 includes the strains: A242, A244, A121, A112, A233, U4 and BM1374165, while the group 9-25 consists of the strains: A235, JK41, JK42, Zeus, JK53, A76, A74, BM1374163, MVT02, A20, A71, F1, FDAARGOS175, Houston-I, Houston-1, JK51, and JK50.

Classification of CDSs of the pan-genome subsets into the cluster of orthologous groups

The subsets of the pan-genome (core, shared and singletons) were classified according to the functional categories of the cluster of orthologous groups (COG) into 1 – Information storage and processing; 2 – signaling and cellular processes; 3 – metabolism; and, 4 – poorly characterized. To perform this classification, the CDSs of the subsets were BLAST aligned against the myva database of COG, using an e-value of 1e-6, and the result was crosschecked with the WHOG information from COG [29].

Horizontally acquired regions

The software GIPSy (Genomic Island Prediction Software) [30,31] was used to perform the prediction of Pathogenicity Islands

Table 1: Genomic features of the strains used in this study.

Species	Strain	Assembly	Genome size (pb)	%GC	No. of contigs	No. of scaffolds	No. of genes	No. of proteins	Isolation country	Host	Collection date
Bartonella henselae	Houston-1	GCA_000046705	1.931.050	38.2	1	1	1.743	1.52	USA	Human	
	BM1374165	GCA_000612765	1.975.500	38.1	1	1	1.767	1.538	France	Cat	
	BM1374163	GCA_000612965	1.905.380	38.2	1	1	1.705	1.505	France	Cat	
	MVT02	GCA_001291465	1.905.380	38.2	1	1	1.706	1.505	France	Human	
	Houston-I	GCA_001525625	2.014.710	38.3	1	1	1.757	1.52	USA	Human	Jan-95
	JK 51	GCA_000516675	1.926.860	38.1	26	7	1.731	1.511	USA	Cat	Oct-94
	JK 50	GCA_000516695	1.921.450	38.2	28	5	1.732	1.505	USA	Human	Sep-94
	JK 42	GCA_000516715	1.921.970	38.3	47	13	1.691	1.469	USA	Cat	Jul-93
	JK 41	GCA_000516735	1.875.760	38	33	9	1.675	1.466	USA	Human	Jul-93
	Zeus	GCA_000708485	1.940.920	38.2	60	19	1.727	1.465	USA	Cat	1990/1995
	JK 53	GCA_000708545	1.940.190	38.3	53	12	1.711	1.462	USA	Human	1995/2005
	FDAARGOS_175	GCA_001525625	2.014.750	38.3	2	2	1.803	1.547	USA	Human	1990
	A20	GCA_001932075	1.840.740	38	66	66	1.596	1.454	France	Cat	Apr-96
	A74	GCA_001932085	1.855.280	38	71	71	1.61	1.471	France	Cat	Mar-96
	A76	GCA_001932095	1.862.680	37.9	58	58	1.613	1.466	France	Cat	Mar-96
	A71	GCA_001932135	1.839.690	38	74	74	1.607	1.463	France	Cat	Apr-96
	A112	GCA_001932145	1.838.410	37.9	64	64	1.63	1.467	France	Cat	May-96
	A233	GCA_001932165	1.835.600	37.9	46	46	1.618	1.461	Denmark	Cat	Oct-98
	A121	GCA_001932175	1.837.810	37.9	53	53	1.615	1.458	France	Cat	May-96
	A242	GCA_001932215	1.860.020	38	46	46	1.638	1.486	Denmark	Cat	Mar-98
	A235	GCA_001932225	1.821.810	38	42	42	1.584	1.448	Denmark	Cat	Mar-98
	A244	GCA_001932235	1.857.850	38	43	43	1.628	1.47	Denmark	Cat	Mar-98
	F1	GCA_001932245	1.864.200	38	67	67	1.621	1.479	USA	Cat	1996
	U4	GCA_001932295	1.812.180	38	56	56	1.576	1.42	USA	Cat	1996
Bartonella apis	BBC0122	GCA_002007565	2.907.210	45.7	1	1	2.546	2.445	Switzerland	*Apis mellifera*	Dec-14

(PAIs) and Resistance Islands (RIs) in the genomes of *B. henselae*. Briefly, GIPSy is a multi-pronged tool that identifies the most common features of the pathogenicity islands, such as genomic signature deviation, tRNA flanking genes, transposases and a high concentration of virulence factors. We used .gbk files and default parameters in our analyses. The genome of *Bartonella apis* [32] was used as the non-pathogenic reference.

The phage regions, i.e. regions harboring phage sequences, were predicted and annotated by the software PHASTER [33] using the .fasta file of the Houston-1 strain. PHASTER performs gene prediction in bacterial genome followed by BLAST analyses against a customized phage database on NCBI and prophage database developed by Srividhya et al. [34,35], which contains phage genes. PAI and phage regions were plotted in comparative circular genome representations on the software BRIG (BLAST Ring Image Generator) [36].

We used the software BRIG to create a circular genome mapping using different genomes as references. The genomic sequences in. fasta format were used on BRIG, which performed the comparative analysis using BLAST to compare the strains of *B. henselae* with the reference Houston-1 strain. Furthermore, the species *B. apis* was also added to the analysis. The results were plotted on a circular comparative map. The image generated is composed of rings, where each genome is represented by a ring of a given color. The intensity of such color is related to the degree of similarity with the reference genome. Deleted regions of the genomes, compared with the reference, are represented by empty regions (blanks). The coordinates of the islands predicted by GIPSy and the phages predicted by PHASTER were added to BRIG's circular map for visualization of the genomic plasticity events. Two additional plasticity analyses were performed using, as references, strains isolated from cats and humans (BM1374165 and Houston-I, respectively). Both were selected as references because they are strains with the largest complete genomes among the 24 available genomes for each host group.

The coordinates of the pathogenicity islands predicted by GIPSy were analyzed at the ACT (Artemis Comparison Tool) software [37] and exported to a list of the genes present in each region. The tool PATRIC [38] was used to find information on the virulence genes found in the analysis. The tool uses BLASTp to compare the genes found on the islands against a database that integrates the Virulence Factor Database (VFDB) [39], MvirDB [40] and manually annotated virulence genes, mainly of the genus *Mycobacterium, Shigella, Salmonella, Escherichia, Listeria* and *Bartonella* [38].

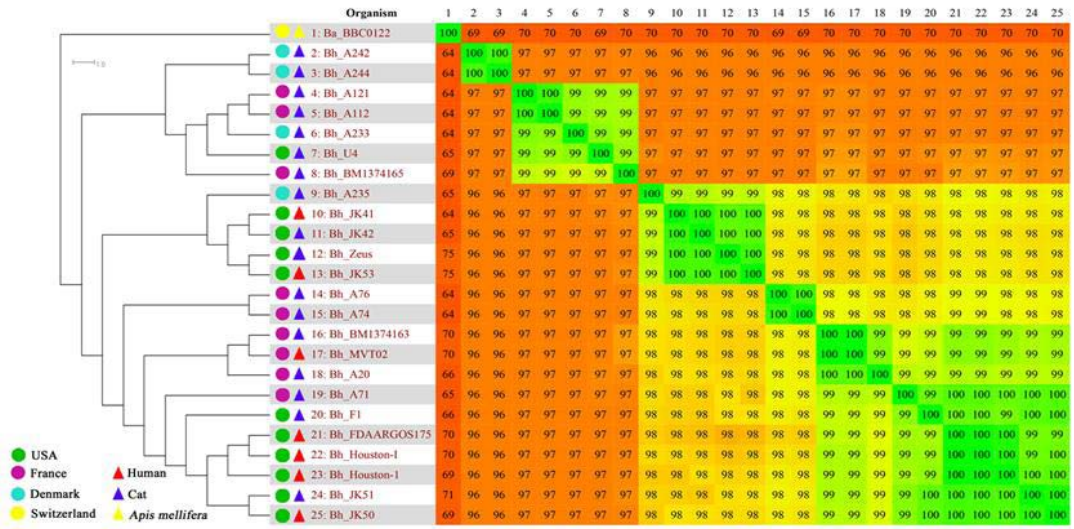

Figure 1: Phylogenomic tree correlating the evolutionary distance among the strains of ***Bartonella henselae***. The genomes of the bacteria are enumerated and named in the 25 lines, whereas the genomes are found in the same order in the columns at the upper region of the chart. The heatmap values indicate the percentage of similarity among the genomes. The colors indicate the degree of similarity, varying from red (low similarity) to green (high similarity). The phylogenomic reconstruction was performed using the SplitsTree software based upon the distance matrix generated by the Gegenees software. The circle next to the name of the strain indicates the country of isolation of the same, according to the respective color of the legend. The triangle represents the host from which the bacterium was isolated for sequencing, also following the color of the legend. Bh: *Bartonella henselae*. Ba: *Bartonella apis*. N/A: Not applicable. USA: United States of America.

Results

General Features

The genomic features of all the genomes used for the analyses are presented in Table 1. Regarding the species *B. henselae*, the genomes have a relatively homogenous size, varying between 1,81 and 2,02 Mb, with a gene content in the range from 1576 to 1803 genes and a GC-content that varies from 37,9% to 38,3%. The amount of contigs on the draft genomes varies from 2 to 74 (Table 1).

Phylogenomic analyses

The similarity analysis was made using Gegenees software to determine the degree of genomic variability among the strains of *B. henselae*. The generated heatmap (Figure 1) shows a high similarity among the strains of *B. henselae*, varying from 96 to 100%. The strains were sub-grouped on clusters, generating 2 clusters on the heatmap. The first cluster comprises of the strains 2 to 8, with the similarity varying from 99% to 100% among them, where all the strains were isolated from cats. The second cluster is formed by the other strains (9-25), with similarities varying from 98% to 100%, where the strains were isolated both from cats or humans. A phylogenomic tree based on Gegenees analysis was also plotted on Figure 1. Noteworthy, the strains A244 and A242, isolated from cats, are more distantly related to the others, whereas they cluster together with other 5 strains (A121, A112, A233, U4 and BM1374165), also isolated from cats.

A phylogenomic tree based on wgMLST analysis is shown in Figure 2, explaining the evolutionary correlation among the strains, where the short evolutionary distance among the various strains is noticeable. As observed at the SplitsTree reconstruction, the correlation of ancestry among the strains corresponds to the same pattern previously found on Gegenees, with a high bootstrap value, while the strains A112, A121, A233, BM1374165, U4, A242, and

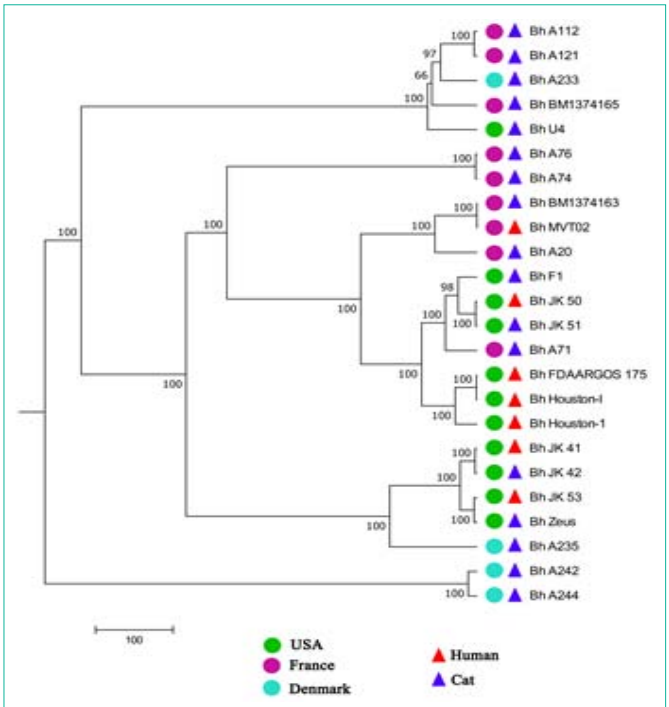

Figure 2: Dendrogram created with the wgMLST profile of the 24 genomes of ***Bartonella henselae***. The PGAdb profile of the genomes of *Bartonella henselae* was used to construct a wgMLST tree using the Build_wgMLSTtree module. The bootstrap values are shown at the nodes. The phylogenomic tree was built *via* the online software PGAdb-builder and the file ".newick" was exported and used in MEGA7 for the tree visualization. The circle next to the name of the strain indicates the country of isolation of the same, according to the respective color of the legend. The triangle represents the host from which the bacterium was isolated for sequencing, also following the color of the legend. Bh: *Bartonella henselae.* N/A: Not applicable. USA: United States of America.

Figure 3: Gene synteny analysis of the complete and draft genomes of ***Bartonella henselae***. The limits among the contigs are represented by red lines. The different Locally Collinear Blocks (LCBs) that are conserved among the strains are represented by different colors. The deleted regions of the genome are represented by blank spaces between the blocks, and the low similarity regions among the genomes are represented by blank spaces inside the blocks. The enumerated ruler above the LCBs represents the genomic position.

A244 are more distantly related from the others, including the strains isolated from humans. According to these results, other analyses were carried out using subsets based on the strains present in clusters 2-8 and 9-25.

Gene Synteny Analysis

At Figure 3, the prediction of LCBs on all the strains has shown large and numerous regions of homology, mainly among the complete genomes. All the draft genomes present small and large inverted regions represented by a change in the orientation of the LCBs from one DNA strand to another. The strain BM1374165 has a large new region compared to Houston-1.

The use of fragmented genomes hinders the visualization of the synteny among the strains. A more accurate analysis was made using only the completed genomes, as represented in Figure 4. There is a predominance of 3 LCBs: the first one (represented by the red color) with approximately 1.46 Mb, a second one (blue) with around 0.43 Mb and a small one (green) between the other two, with just 4.85 kb. Only the strain BM1374165 presents small deletions on a pathogenicity island (highlighted by the black box). Besides that, it is interesting to notice that MVT02 and BM1374163, although presenting 100% similarity on Gegenees, are still very similar on the analysis run by Mauve, as opposed to Houston 1 and Houston I, that present regions of deletion on different regions of the genome.

Pan-genomic analysis

The pan-genomic analysis through Ortho Finder software identified a pan-genome composed of 1655 non-redundant genes.

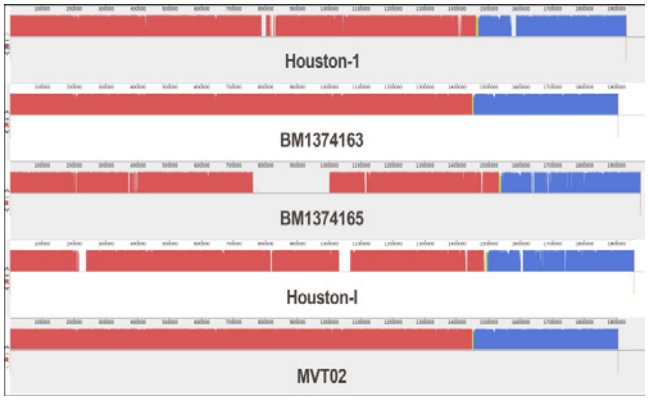

Figure 4: Gene synteny analysis of the completed genomes of ***Bartonella henselae***. The genomes of the strains of *B. henselae* were aligned using the Mauve software. The limits among the contigs are represented by red lines. The different Locally Collinear Blocks (LCBs), that are conserved among the strains are represented by different colors. The deleted regions of the genome are represented by blank spaces between the blocks. The low similarity regions between the genomes are represented by blank spaces inside the block. The enumerated ruler above the LCBs represents the genomic position.

The core genome, composed of genes commonly shared by all species, contains 1160 genes, or 70.1% of the total number of genes in the pan-genome. The shared genome, with genes shared by two or more strains, but not all, has 388 genes (23.44% of the total pan-genome). Lastly, 107 genes (6.46%) are strain specific, being denominated singletons.

Pan-genome, core genome and singletons development

Considering $\alpha=1-\gamma$, where the γ value for the extrapolation of the pan-genome of the complete dataset was 0.06, the α value of the complete dataset is 0.94, whereas the subsets USA, France, cats, humans, 2-8 and 9-25 have the following α values, respectively: 0.94, 0.93, 0.93, 0.92, 0.90 and 0.95. According to Heap's Law, an α lower than 1 is representative of an open pan-genome, where each added genome will contribute some genes to the pan-genome, whereas an α higher than 1 is representative of a closed pan-genome, where newly added genomes will not contribute significantly to the pan-genome. Considering this rule, all pan-genomes are still open, although varying at different rates, where the subset of 2-8 presents the fastest changing pan-genomes, opposing to the subset 9-25, which present the slowest growing pan-genome.

The core genomes of the complete dataset and the subsets USA, France, cats, humans, 2-8 and 9-25 have the following $tg(\theta)$ values, respectively: ~1136, ~1023, ~1141, ~1165, ~1230, ~1137 and ~1211. Concerning the singletons of the complete dataset and the subsets USA, France, Cats, humans, 2-8 and 9-25, they have the following $tg(\theta)$ values, respectively: ~1, ~6, ~3, ~3, ~12, ~16 and ~4.

According to the least-squares fit of the exponential regression decay, the $tg(\theta)$ represents the point where the curve will stabilize, which may be correlated to the number of genes in the core genome after stabilization and the number of singletons that will be added to the pan-genome for each newly sequenced genome. Considering this rule, the core genome of the complete dataset will have 1136 genes after stabilization, whereas the subset from the USA will be the

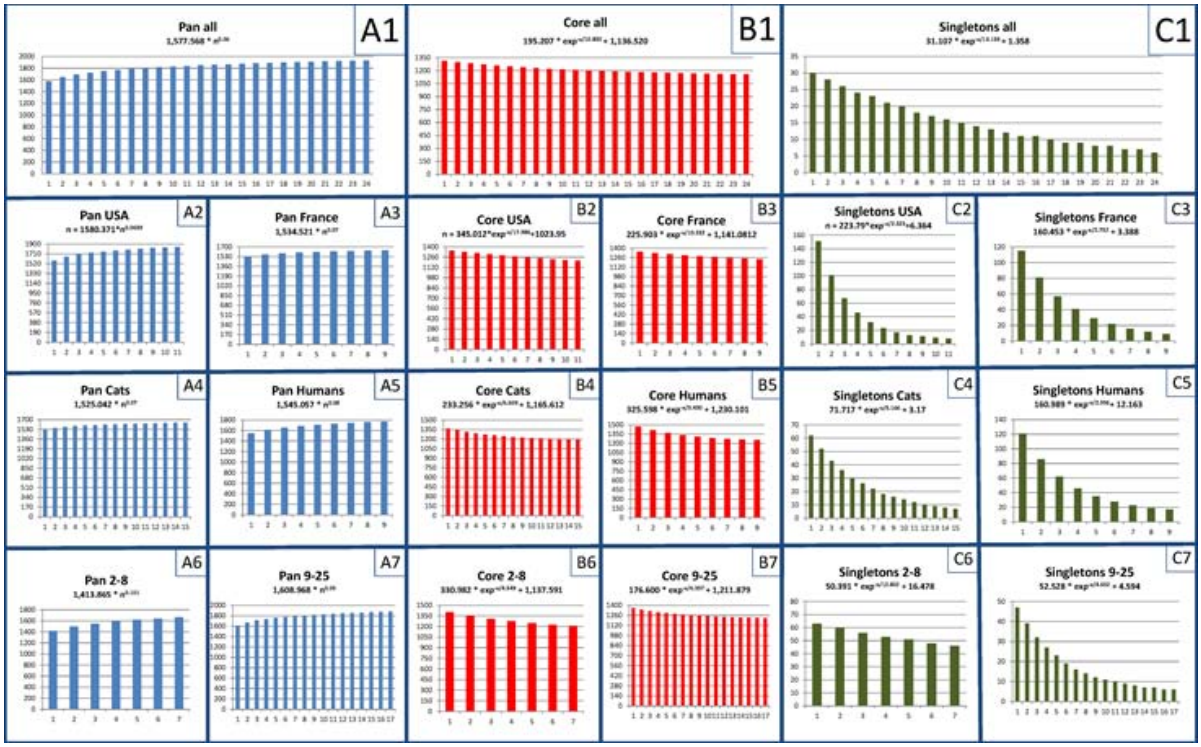

Figure 5: Pan-genome, core genome and singletons development of ***Bartonella henselae*.** A1/B1/C1, respectively, the pan-genome, core genome and singletons development using all 24 strains of *B. henselae*; A2/B2/C2, pan-genome, core genome and singletons development of the strains isolated from USA; A3/B3/C3, the pan-genome, core genome and singletons development of the strains isolated from France; A4/B4/C4, the pan-genome, core genome and singletons development of the strains isolated of cats; A5/B5/C5, the pan-genome, core genome and singletons development using the strains isolated of humans; A6/B6/C6, the pan-genome, core genome and singletons development using of the clusters 2-8 generated from the software Gegenees; A7/B7/C7, the pan-genome, core genome and singletons development using the clusters 9-25 generated from Gegenees.

Table 2: Locations and sizes of pathogenicity islands predicted in the reference strain *Bartonella henselae* Houston-1.

ISLAND	LOCATION		SIZE (pb)
PAI 1	126.888	136.627	9.739
PAI 2	356.6	428.14	71.54
PAI 3	1.277.004	1.300.302	23.298
PAI 4	1.437.361	1.505.591	68.23
PAI 5	1.524.451	1.542.024	17.573
PAI 6	1.799.481	1.819.889	20.408

smallest one, with 1023 genes and the core genome of humans will be the largest one, with 1230 genes. On the singletons subset, on the other hand, only one gene will be added to the complete dataset for each newly sequenced genome, whereas the subsets from France and Cats will have 3 newly added genes and the subsets humans and 2-8 will have 12 and 16 newly added genes, respectively (Figure 5).

Classification of the CDSs in the subsets of the pan-genome on the cluster of orthologous groups

The classification of subsets of the pan-genome on the functional categories of the COG is represented in Figure 6. Most of the genes at the subsets shared and singletons are composed of "poorly characterized" genes. On the other hand, the core genome is mostly composed of genes that integrate the cellular "metabolism", in addition to genes responsible for "cellular processes and signaling" and "information storage and processing".

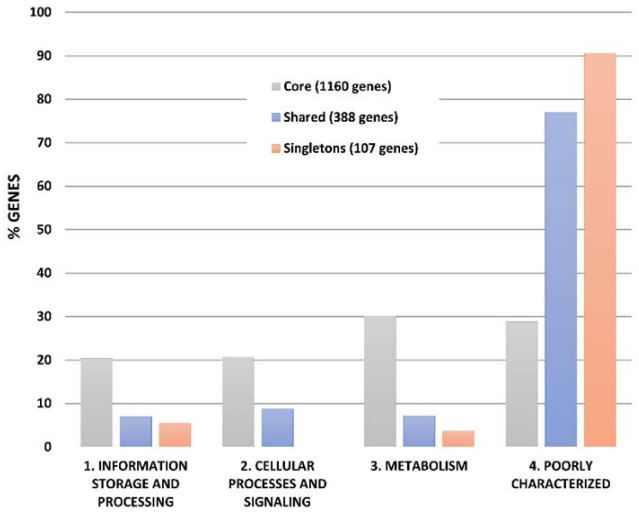

Figure 6: Subset classification of the pan-genome according to the functional categories of the Cluster of Orthologous Groups. The figure represents the classification of the gene sets of the pan-genome according to the functional categories of the Cluster of Orthologous Groups (COG) on a percentage scale.

Genomic Plasticity

The analysis of the circular comparison of the genomes of the

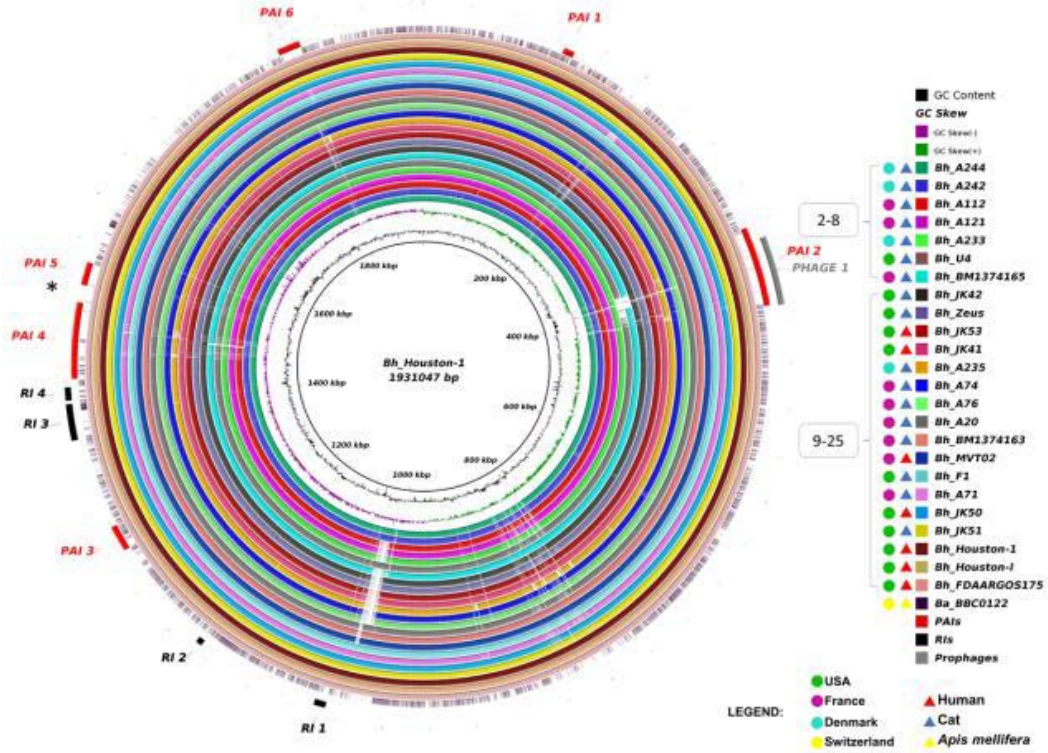

Figure 7: Circular representation of the genomic comparison among the strains of the *Bartonella henselae* species. The circular genomic comparisons were created with the BRIG software. The image represents the genomic plasticity among the genomes of *B. henselae* using the strain Houston-1 as a reference. Each ring in the image corresponds to one genome of one strain of *B. henselae*, with the respective color corresponding to the legend at the right. The deleted regions are represented by blank spaces inside the circles, while the shared genes regions are filled with color. The PAIs, RIs, and phage regions were predicted by the GIPSy and PHASTER software, respectively. The circle next to the name of the strain indicates the country of isolation of the same, according to the respective color of the legend. The triangle represents the host from which the bacterium was isolated for sequencing, also following the color of the legend. The numbers next to the keys indicate the numbering of the cluster generated by Gegenees and PGAdb analysis. The asterisk (*) represents the location of the cluster of virulence genes *virB.* Bh: *Bartonella henselae.* Ba: *Bartonella apis*. PAI: pathogenicity island. RI: resistance island N/A: Not applicable. USA: United States of America.

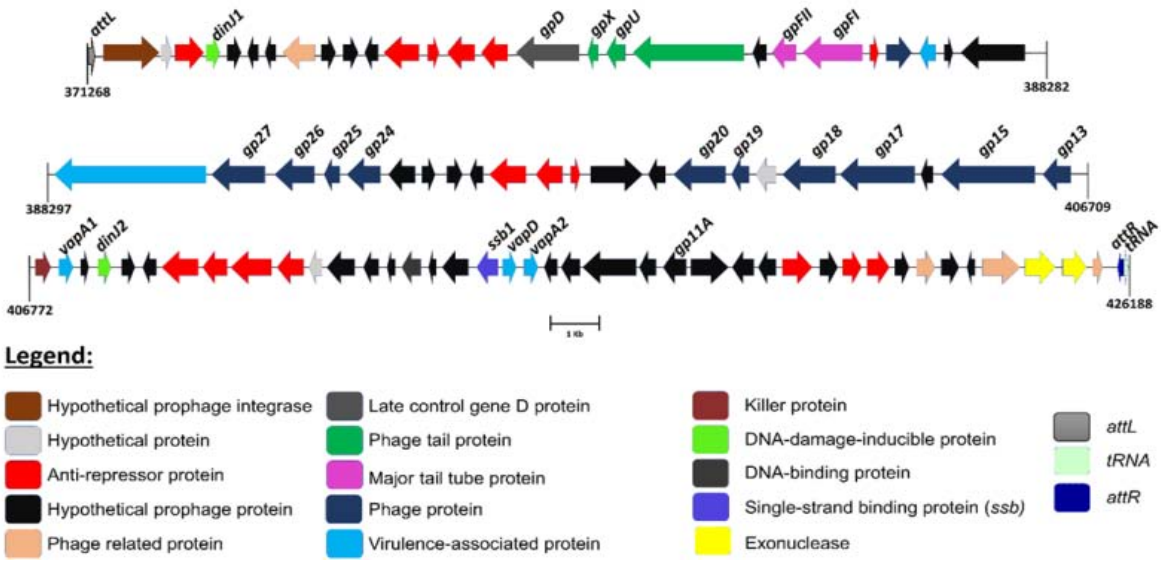

Figure 8: Gene content of the phage region found in *Bartonella henselae* Houston-1. The phage region determined by the PHASTER software, with approximately 54.9 Kb, possesses 89 genes that are represented by colored arrows. The segment was cut in 3 parts to facilitate the representation. Arrows pointing to the right and left sides represent the forward and reverse strands of DNA. Gene names are represented right above their correspondent arrows. To facilitate the representation, the intergenic regions were not dimensioned. The final region of the segment, composed of *attR* and *tRNA* are overlapping on the genome.

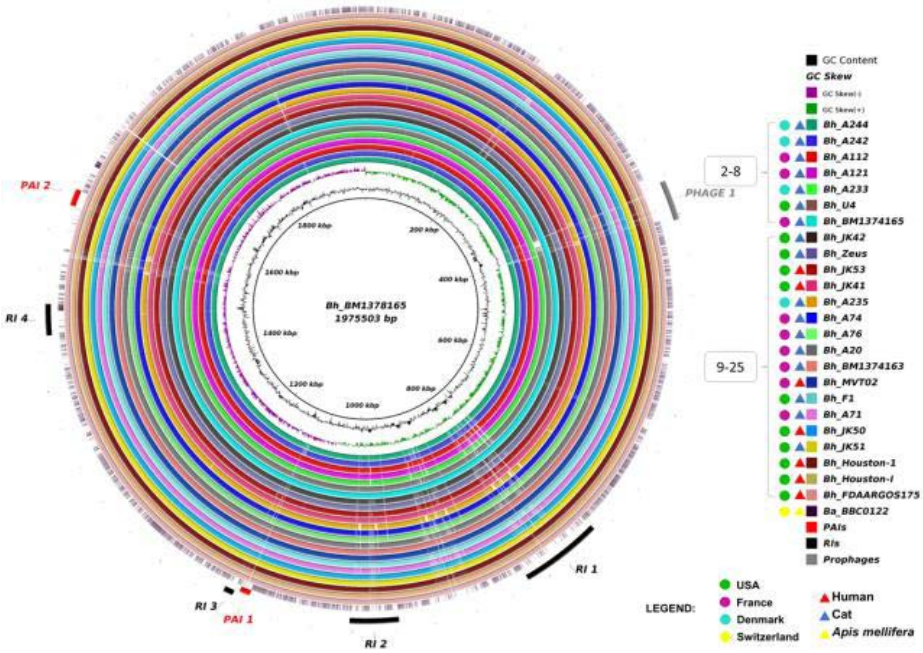

Figure 9: Circular representation of the genomic comparison between all strains of *Bartonella henselae* using the strain BM1374165 as reference. The image represents the genomic plasticity among the genomes of *B. henselae* using the strain BM1374165 as a reference. Each ring in the image corresponds to one genome of one strain of *B. henselae*, with the respective color corresponding to the legend at the right. The deleted regions are represented by blank spaces inside the circles, while the shared genes regions are filled with color. The circle next to the name of the strain indicates the country of isolation of the same, according to the respective color of the legend. The triangle represents the host from which the bacterium was isolated for sequencing, also following the color of the legend. The numbers next to the keys indicate the numbering of the clusters generated by Gegenees and PGAdb analysis. Bh: *Bartonella henselae.* Ba: *Bartonella apis*. PAI: pathogenicity island. RI: resistance island N/A: Not applicable. USA: United States of America.

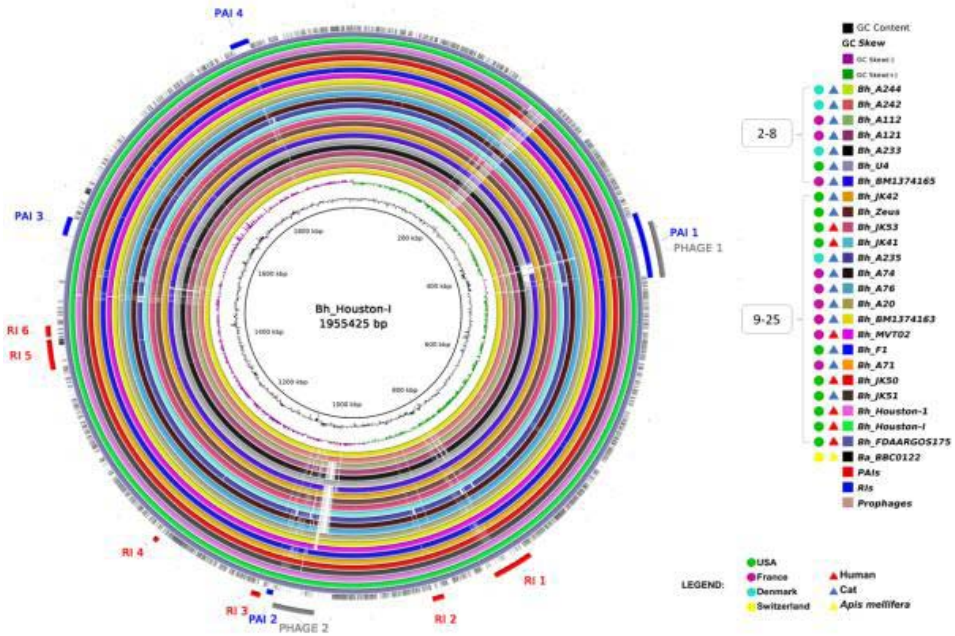

Figure 10: Circular representation of the genomic comparison between all strains of the *Bartonella henselae* using the strain Houston-I as reference. The image represents the genomic plasticity among the genomes of *B. henselae* using the strain Houston-I as a reference. Each ring in the image corresponds to one genome of one strain of *B. henselae*, with the respective color corresponding to the legend at the right. The deleted regions are represented by blank spaces inside the circles, while the shared genes regions are filled with color. The circle next to the name of the strain indicates the country of isolation of the same, according to the respective color of the legend. The triangle represents the host from which the bacterium was isolated for sequencing, also following the color of the legend. The numbers next to the keys indicate the numbering of the cluster generated by Gegenees and PGAdb analysis. Bh: *Bartonella henselae.* Ba: *Bartonella apis*. PAI: pathogenicity island. RI: resistance island N/A: Not applicable. USA: United States of America.

Table 3: List of all genes located in the pathogenicity islands of the reference strain *Bartonella henselae* Houston-1.

ISLAND	GENE	LOCATION		PRODUCT
PAI 1	*rpsA*	127083	128783	30S ribosomal protein S1
	cmk	128916	129563	Cytidylate kinase
	aroA	129560	130888	3-phosphoshikimate 1-carboxyvinyltransferase
	BH00960	131075	131455	Hypothetical protein
	ilvD	132510	134348	Dihydroxy-acid dehydratase
	BH00990	134491	134639	Prophage integrase
PAI 2	*lldD*	357436	358587	L-lactate dehydrogenase
	BH02730	360289	360414	Hypothetical protein
	BH02740	361075	361395	Hypothetical protein
	pheP	361576	363003	Phenylalanine-specific permease
	BH02760	363045	363725	ABC transporter ATP-binding protein
	BH02770	363722	364360	Hypothetical protein
	BH02780	364374	364928	Biotin synthase
	BH02790	366252	366632	exonuclease
	BH02800	367274	367522	[no /product]
	BH02810	368313	368600	[no /product]
	BH02820	368624	368975	Phage related
	BH02830	369436	370097	Phage related
	BH02840	370293	370475	Hypothetical protein
	BH02850	370522	370674	Hypothetical protein
	BH02860	370808	371101	Hypothetical protein
	BH02870	371023	371178	Hypothetical protein
	BH02880	371485	372642	Hypothetical phage integrase
	BH02890	372656	372886	Hypothetical protein
	BH02900	372914	373504	Anti-repressor protein
	dinJ1	373624	373899	DNA-damage-inducible protein
	BH02920	373912	374202	Hypothetical prophage protein
	BH02930	374325	374555	Hypothetical prophage protein
	BH02940	374666	374893	Hypothetical prophage protein
	BH02950	374890	375552	Phage related lysozyme
	BH02960	375712	376014	Hypothetical prophage protein
	BH02970	376001	376321	Hypothetical prophage protein
	BH02980	376365	376643	Hypothetical prophage protein
	BH02990	376640	377371	Anti-repressor protein
	BH03000	377391	377618	Anti-repressor
	BH03010	377623	378195	Anti-repressor protein
	BH03020	378223	378771	Anti-repressor protein
	gpD	379083	380390	Late control gene D protein
	gpX	380387	380611	Phage tail protein
	gpU	380608	380988	Phage tail protein
	BH03060	380994	383294	Phage tail protein
	BH03070	383404	383688	Hypothetical prophage protein
	gpFII	383690	384196	Major tail tube protein FII
	gpFI	384196	385428	Major tail sheath protein FI
	BH03100	385668	385844	Anti-repressor
	BH03110	385898	386419	Phage protein
	BH03120	386465	386794	Virulence-associated protein
	BH03130	386882	387061	Hypothetical prophage protein
	BH03140	387357	388682	Hypothetical prophage protein
	BH03150	388697	391840	Virulence-associated protein
	gp27	391842	392948	Phage protein gp27
	gp26	392948	393775	Phage protein gp26
	gp25	393772	394101	Phage protein gp25
	gp24	394110	394799	Phage-related baseplate assembly protein
	BH03200	394780	395322	Hypothetical prophage protein
	BH03210	395374	395646	Hypothetical prophage protein
	BH03220	395636	395953	Hypothetical prophage protein
	BH03230	396194	396472	Hypothetical prophage protein
	BH03240	396469	397227	Anti-repressor protein
	BH03250	397246	397800	Anti-repressor protein
	BH03260	397821	398012	Anti-repressor
	BH03270	398185	399258	Hypothetical prophage protein
	BH03280	399294	399647	Hypothetical prophage protein
	gp20	399649	400725	Phage protein gp20
	gp19	400738	401106	Phage protein gp19
	BH03310	401103	401516	Hypothetical protein
	gp18	401366	402475	Phage protein gp18
	gp17	402399	403955	Phage protein gp17
	BH03340	403955	404203	Hypothetical prophage protein
	gp15	404207	406135	Phage terminase large subunit (gp15)
	gp13	406128	406709	Phage protein gp13
	BH03370	406772	407086	Killer protein
	vapA1	407103	407399	Virulence-associated protein a
	BH03390	407511	407684	Hypothetical prophage protein
	dinJ2	407809	408072	DNA-damage-inducible protein
	BH03410	408059	408340	Hypothetical prophage protein
	BH03420	408383	408661	Hypothetical prophage protein
	BH03430	408658	409416	Anti-repressor protein
	BH03440	409476	409991	Anti-repressor protein
	BH03450	410008	410847	Anti-repressor protein
	BH03460	410879	411421	Anti-repressor protein
	BH03470	411438	411704	Hypothetical protein
	BH03480	412018	412596	Hypothetical prophage protein
	BH03490	412600	412920	Hypothetical prophage protein
	BH03500	413179	413355	Hypothetical prophage protein
	BH03510	413393	413788	DNA-binding protein
	BH03520	413785	413943	Hypothetical prophage protein
	BH03530	414079	414618	Hypothetical prophage protein
	ssb1	414599	415045	Single-strand binding protein SSB
	BH03550	415237	415517	*vapD*
	vapA2	415540	415830	Virulence-associated protein A
	BH03570	415927	416199	Hypothetical prophage protein
	BH03580	416364	416762	Hypothetical prophage protein
	BH03590	416776	417888	Hypothetical prophage protein
	BH03600	417878	418225	Hypothetical prophage protein
	gp11A	418218	418694	Hypothetical prophage protein
	BH03620	418822	419598	Hypothetical prophage protein
	BH03630	419607	420059	Hypothetical prophage protein
	BH03640	420072	420437	Hypothetical prophage protein
	BH03650	420515	421132	Phage repressor protein
	BH03660	421374	421748	Hypothetical prophage protein
	BH03670	421813	422205	Anti-repressor protein
	BH03680	422215	422682	Anti-repressor protein
	BH03690	422672	422974	Hypothetical prophage protein
	BH03700	422987	423370	Phage related protein
	BH03710	423374	423739	Hypothetical prophage protein
	BH03720	423736	423903	Hypothetical prophage protein
	BH03730	423812	424591	Phage related protein
	BH03740	424593	425216	Exonuclease
	BH03750	425245	425736	Exonuclease

	BH03760	425809	426012	Phage related protein
	serC	426588	427751	Phosphoserine aminotransferase
PAI3	**BH11550**	1277140	1278357	Hypothetical protein
	BH11560	1278427	1280100	Hypothetical protein
	BH11570	1280717	1280968	Hypothetical protein
	ubiC	1282863	1283378	Chorismate-pyruvate lyase
	carA	1284573	1285772	Carbamoyl-phosphate synthase small chain
	BH11600	1285837	1286849	putative transport protein
	BH11610	1287680	1288519	Hypothetical protein
	carB	1289195	1292680	Carbamoyl-phosphate synthase large chain
	BH11630	1293067	1293276	Cold shock protein
	aatA	1293574	1294776	Aspartate aminotransferase a
	BH11650	1295556	1296461	LysR family transcriptional regulator
	trxB	1296518	1297471	
	trkH	1297784	1299178	trk system potassium uptake protein trkh
	BH11680	1299870	1299953	[no /product]
PAI 4	BH12870	1439270	1439626	DNA-binding protein
	BH12880	1440117	1440584	Prophage integrase
	asd	1440905	1441921	Aspartate-semialdehyde dehydrogenase
	BH12900	1442300	1442926	Hypothetical protein
	yaeC	1443453	1444271	Lipoprotein
	BH12920	1444299	1445333	ABC transporter, ATP-binding protein
	BH12930	1445326	1445985	ABC transporter, permease protein
	BH12940	1446396	1447533	Prophage integrase
	nodN	1447916	1448395	Nodulation protein n
	BH12960	1448563	1449900	Hypothetical protein
	BH12970	1454740	1455480	Hypothetical protein
	BH12980	1455496	1456524	Hypothetical protein
	BH12990	1456535	1457506	Hypothetical protein
	BH13000	1458309	1458431	Prophage integrase
	BH13010	1459285	1459866	Probable surface protein
	BH13020	1459863	1462346	Hypothetical protein
	BH13030	1463064	1469090	Probable surface protein
	BH13040	1470331	1471284	fatD
	BH13050	1471232	1472220	fatC
	BH13060	1472178	1472962	ceuE
	BH13080	1473184	1473395	[no /product]
	BH13090	1474501	1475505	Hypothetical protein
	BH13100	1476693	1477628	[no /product]
	BH13110	1477431	1479672	[no /product]
	BH13120	1480666	1484991	Hypothetical protein
	BH13130	1485515	1485661	Prophage integrase
	BH13140	1486276	1488897	Probable surface protein
	BH13150	1489421	1489567	Prophage integrase
	BH13160	1490182	1492524	Hypothetical protein
	BH13170	1493412	1493714	Hypothetical protein
	BH13180	1494963	1497581	Hypothetical protein
	BH13190	1499323	1499766	Hypothetical protein
	parA1	1503848	1504477	parA protein
PAI 5	**BH13360**	1521519	1521725	Hypothetical protein
	BH13370	1521736	1523370	Putative cell filamentation protein
	traG	1523533	1525452	Conjugal transfer protein trag
	BH13390	1525994	1527622	Hypothetical protein
	BH13400	1528692	1530290	Putative cell filamentation protein
	BH13410	1530764	1532368	Hypothetical protein
	BH13420	1532694	1534088	Hypothetical protein
	BH13430	1534424	1536928	Hypothetical protein
	BH13440	1537113	1540244	Hypothetical protein
	BH13450	1541044	1541217	Prophage integrase
PAI 6	BH15520	1799985	1801265	Oxidoreductase
	korB	1802021	1802290	korB protein
	BH15540	1802337	1803065	Hypothetical protein
	trwN	1803062	1803730	trwN protein
	korA	1803874	1804167	korA protein
	trwL1	1804130	1804447	trwL1 protein
	trwL2	1804494	1804805	trwL2 protein
	trwL3	1804829	1805152	trwL3 protein
	trwL4	1805190	1805507	trwL4 protein
	trwL5	1805661	1805969	trwL5 protein
	trwL6	1806088	1806429	trwL6 protein
	trwL7	1806567	1806881	trwL7 protein
	trwL8	1807042	1807353	trwL8 protein
	trwM	1807366	1807674	trwM protein
	trwK	1807677	1810148	trwK protein
	trwJ1	1810145	1810933	trwJ1 protein
	trwI1	1811167	1812051	trwI1 protein
	trwH1	1812216	1812362	Hypothetical protein
	trwJ2	1812359	1812985	trwJ2 protein
	trwI2	1813241	1814125	trwI2 protein
	trwH2	1814293	1814439	Hypothetical protein
	trwG	1814436	1815134	trwG protein
	trwF	1815148	1815954	trwF protein
	trwE	1815954	1817141	trwE protein
	trwD	1817116	1818177	trwD protein

strains of *B. henselae* (Figure 7) shows various regions shared by all the strains, as well as few deleted regions on them. We identified 6 pathogenicity islands (PAIs) through predictions made by the GIPSy software. Their lengths vary from ~9.7 kb (PAI 1) to ~68.2 kb (PAI 4). The location and size of the PAIs are shown in Table 2 and the list of all the genes located at PAIs is presented at Table 3. We also identified one phage region considered intact by the PHASTER software, whose location corresponds, approximately, to the same PAI 2 region (Figure 8). Among the 191 genes predicted into all the islands, 73 (38%) are "hypothetical proteins", and, among these, 36 are "hypothetical prophage protein". The strain *B. henselae* U4 presents a deletion at the region predicted as PAI 2 on the reference genome (*B. henselae* Houston-1), that corresponds, also, to the region predicted as Phage 1.

According to the analysis of the genome BM1374165 (Figure 9) and Houston-I (Figure 10), isolated from cat and human, respectively, we predicted 4 RIs, 2 PAIs and one phage region on the former and 6 RIs, 4 PAIs and two phage regions on the latter. Interestingly, the region that is present on the strain BM1374165 and absent on Houston-1, observed on Mauve (Figure 3-4), mostly corresponded to RI 1 predicted by GIPSy on the former, as observed in Figure 9. This is a duplication of the region 840001-903302 of the strain Houston-1, which is equivalent to the regions 810628-873868 and 911748-976210 of the strain BM1374165.

Briefly, the phage region harbours the genes: phage integrase (int), which is required for the integration of the element, head genes, terminase subunits, prohead protease, scaffold protein, major capsid subunit, and the tail genes, including the major and minor

tail subunit proteins, overlapping open reading frames, and the tape measure protein [41].

Discussion

Our analyses suggest that the *B. henselae* species present a high similarity among its strains, indicating a certain stability of the genome and low degree of diversity. This fact corroborates previous studies that demonstrated the low rate of recombination of the intraspecies gene content [17,18].

Bouchouicha et al. [42] demonstrate the use of MLVA for classification of *B. henselae* genomes into two genotypes. In the case of our study, where the strains are completely sequenced and deposited in public databases, it is more viable to use wgMLST approaches, which use the whole genome, instead of 16s. This methodology is more precise than MLVA and is able to identify clones and meroclones from the samples as shown by Maiden [22]. Although the strains of *B. henselae* have a high similarity at its genetic content, they can still have some small variations at the nucleotide level that cannot be traced by the Gegenees software. In order to identify the differences at such level, we performed a phylogenomic analysis based on the genome polymorphism of the species. As shown by the tree created by Splits Tree, the analysis of wgMLST *via* PGAdb-builder also presents a high proximity among the species, demonstrating the gene conservation throughout time. The strains A242 and A244 became further apart compared to the others. Both strains were isolated from cats in Denmark on the same period (March 1998) [43].

The proximity of the A74 and A76 strains could be associated to the fact that both were isolated and sequenced at the same period from feline hosts on France (March 1996). The host's data, as well as in the case described above, could not justify the phylogenomic analysis. But, an interesting fact is evidenced by analyzing the U4 and F1 strains. Both were isolated in 1996 from feline hosts in the USA. However, both in the SplitsTree and the PGAdb-builder analyses, the strains presented themselves as phylogenomically apart and present in clusters with strains from 2 different countries (France and the USA). From these results, it might be possible to hypothesize that the phylogenomic proximity among different strains is independent of the fact that they were isolated from the same country or year.

Although highly similar concerning their gene content, different strains of the same species may exhibit genetic rearrangements that allow them to develop different phenotypes. To verify the possible existence of rearrangement at the *B. henselae* genomes, a gene synteny analysis was performed by Mauve. This software identifies LCBs during the alignment process, defined as regions without local rearrangements of probable homologous sequences shared by two or more genomes. From the analysis considering only the completed genomes, we could observe that the species maintain its structure conserved among its strains since there are no inversions and translocations. Although the completed genomes present some deletion regions, none of them are inside pathogenicity islands.

Altogether, the data on the extrapolations show a very homogeneous variation on the pan-genome, core genome and singletons of the complete dataset, with an α value of 0.94, representing a slow growing open pan-genome and tg(θ) of 1136 and 1 for the core genome and singletons, respectively, which shows a tendency for a closed pan-genome. However, the singletons from the genome isolated from the USA and the core genome and singletons from the strains isolated from humans show fast changing values, showing a possible fast changing pan-genome, which is in agreement with the α value of 0.92 from the pan-genome of human. Also, the alpha value of the strains 2-8 was the smallest one, 0.90, whereas the core genome and singletons present only slight variations on the curve, which is in agreement with the grouping of those strains in the phylogenomic tree.

As expected, the analysis of the subsets of the pan-genome by the COG database demonstrated that the main components of the core genome are genes responsible for essential processes of the bacteria, as metabolism and cellular signaling. In contrast, genes that compose the shared and singletons are poorly characterized genes. This is in agreement with the high importance of the core genes subset, which is mostly composed of housekeeping genes of the species.

Genome plasticity analyses allow the search for regions on the genome called pathogenicity island (PAIs), which are very important for being acquired by horizontal gene transfer and for containing virulence genes, as initially observed by Hacker, in virulence studies with *Escherichia coli* [44] For PAIs prediction, we included the *B. apis* bacteria genome at GIPSy as a non-pathogenic reference. The choice was based on the fact that this species, until this moment, was not associated to a pathogenic response on a human host or on mammals in general. *B. apis* can be found in the posterior intestine of fully grown bees, where they live in a symbiotic relationship with its host [32]. Besides, the species is intimately related with the *B. henselae* species.

Six PAIs were identified from the analyses of all the *B. henselae* genome through GIPSy (Table 2). The gene list with the respective location can be found in Table 3. We analyzed gene by gene and we focused on the genes that code proteins involved in known virulence mechanisms and described in the literature.

PAI 1: The gene *aroA*, coding for the product 3-phosphoshikimate 1-carboxyvinyltransferase, is an integral part of the Shikimate pathway, responsible for the formation of aromatic amino acids in bacteria [45]. The mutation in this gene is used in *Salmonella* species for the generation of recombinant mutants used as adjuncts in attenuated vaccines [46]. Besides that, other studies using the gene *aroA* as a potential vaccine target were published in literature and used on a variety of microorganisms, such as *Pasteurella multocida* [47], *Bordetella pertussis* [48] and *Yersinia pestis* [49]. However, in the last study, although the mutation in the gene was able to elicit an immune response in guinea pigs, it presented virulence in mice. A recent study revealed that the elimination of the *aroA* gene of *Salmonella typhimurium* led to an increased virulence and consequent elevation of immunogenicity of the strain, such a fact being advantageous to the optimization of the vaccine agents and bacterial immunotherapy [50].

The gene *ilvD* coding for the product dihydroxy-acid dehydratase was described in *Mycobacterium tuberculosis* [51] and *Escherichia coli* [52] as being an enzyme essential to bacteria. Its main function is to take part in the biosynthesis of branched-chain amino acid. Previous studies with anti-tuberculosis agents revealed that the inactivation of the *ilvD* gene led to auxotrophy of three amino acids (isoleucine,

leucine, and valine) and reduced the surviving capabilities of the bacteria [53].

PAI 2: The second pathogenicity island corresponds to a phage region predicted by Phaster as phage integration regions. Several phage proteins can be found in this region, justifying the pathogenicity island predicted by GIPSy.

PAI 3: The operon *carAB* was the main component found in this island. In prokaryotes, the genes *carA* and *carB* are responsible for the coding of the minor and major subunits, respectively, of the product Carbamoyl-phosphate synthase large chain [54]. This enzyme is essential for the metabolism of arginine and pyrimidines. Previous studies revealed that mutations on the *carA* and *carB* genes produce a silenced mutant, which may be used as an oral vaccine against avian colibacillosis [55].

PAI 4: The gene *asd*, that code Aspartate-semialdehyde dehydrogenase, takes part in the formation of the cell wall and was described in studies of the immunologic response of *Salmonella enterica serovar Typhimurium* [56].

PAI 5: Six of the ten genes present in this island are genes that code hypothetical proteins, with no description of those in the literature. It was also found as a prophage integration region.

PAI 6: The genes *trwL1, trwL2, trwJ2, trwF, trwE*, and *trwD* were found in this island. The genes of the *trw* family are intimately related with the species of the *Bartonella* genus virulence, mainly *Bartonella tribocorum* [57], being an operon that codes the components of type IV secretion system (T4SS). The trw locus was acquired *via* horizontal gene transfer on *Bartonella* species and is currently present in approximately 13 species [58]. It is a conjugation system involved in the bacteria pathogenicity and related with the transport of proteins (mostly virulence factors) and DNA. The bacteria component as a whole eases the dispersion of bacterial genetic material, including the pathogenicity islands to other microorganisms, being able to transform non-virulent species into pathogenic ones, which makes it an important virulence factor present in *B. henselae* [59]. Besides, it was already described that the *trw* system mediates the specific erythrocyte infection of the host, attaching itself to the erythrocyte surface, which is the main developmental factor in the pathogeny of *B. henselae* [57].

It is interesting to notice that the main virulence genes of the T4SS, of the *virB* family, are not harbored by any of the predicted PAIs in our study. However, such genes are present among the PAI 4 and PAI 5, as represented by an asterisk in Figure 7, and such a fact is relevant due to the possibility that these genes are integrated into the islands and become transmissible by horizontal transference to non-pathogenic species.

The additional analysis of the genomes isolated from cat (BM1374165) and human (Houston-I), selected from clusters 2-8 and 9-25, respectively, showed that, in cat isolates, the amount of horizontally acquired regions (PAIs, Ris, Prophage) is lower in relation to the human isolate.

Conclusion

All the results presented here highlighted the high similarity among the genomes of *B. henselae*. We have found two clusters of genomes, one composed of strains isolated from cats (2-8) with more diverse pan-genome and another composed of strains isolated from cats or humans (9-25), with a more homogeneous pan-genome, where the extrapolations of the pan-genome, core genome, and singletons show that group 2-8 are varying faster than the strains from group 9-25. Also, although many PAIs and RIs were predicted, none of them present deletions in agreement with the subsets studied here and may not be correlated with the epidemiological hypothesis raised here. Altogether, there is still a need for sequencing more strains from other countries to better understand the pangenomic features of this organism. Finally, the highly conserved genomes from this species are very important for the development of new vaccines and analyses of drug targets against this pathogen that is of great importance for causing disease not only in domestic cats but also in humans, and these data may then be used in future works, which will be highly relevant for containing the disease worldwide.

References

1. Boulouis H-J, Chao-chin C, Henn JB, Kasten RW, Chomel BB. Factors associated with the rapid emergence of zoonotic Bartonella infections. Vet Res. 2005; 36: 383-410.
2. Eicher SC, Dehio C. Bartonella entry mechanisms into mammalian host cells. Cell Microbiol. 2012; 14: 1166-73.
3. Okaro U, Addisu A, Casanas B, Anderson B. Bartonella Species, an Emerging Cause of Blood-Culture-Negative Endocarditis. Clin Microbiol Rev. 2017; 30: 709-746.
4. Minnick MF, Anderson BE, Lima A, Battisti JM, Lawyer PG, Birtles RJ. Oroya Fever and Verruga Peruana: Bartonelloses Unique to South America. PLoS Negl Trop Dis. 2014; 8: e2919.
5. Foucault C, Brouqui P, Raoult D. Bartonella quintana characteristics and clinical management. Emerg Infect Dis. 2006; 12: 217-223.
6. Breitschwerdt EB. Feline bartonellosis and cat scratch disease. Vet Immunol Immunopathol. 2008; 123: 167-171.
7. Gutiérrez R, Nachum-Biala Y, Harrus S. Relationship between the Presence of Bartonella Species and Bacterial Loads in Cats and Cat Fleas (Ctenocephalides felis) under Natural Conditions. Appl Environ Microbiol. 2015; 81: 5613-5621.
8. Weeden AL, Cherry NA, Breitschwerdt EB, Cheves AG, Wamsley HL. Bartonella henselae in canine cavitary effusions: prevalence, identification, and clinical associations. Vet Clin Pathol. 2017; 46: 326-30.
9. Regier Y, O'Rourke F, Kempf VAJ. Bartonella spp. - a chance to establish One Health concepts in veterinary and human medicine. Parasit Vectors 2016; 9: 261.
10. Maggi RG, Ericson M, Mascarelli PE, Bradley JM, Breitschwerdt EB. Bartonella henselae bacteremia in a mother and son potentially associated with tick exposure. Parasit Vectors. 2013; 6: 101.
11. Mascarelli PE, Maggi RG, Hopkins S, Mozayeni BR, Trull CL, Bradley JM, et al. Bartonella henselae infection in a family experiencing neurological and neurocognitive abnormalities after woodlouse hunter spider bites. Parasit Vectors. 2013; 6: 98.
12. Dehio C. Bartonella interactions with endothelial cells and erythrocytes. Trends Microbiol. 2001; 9: 279-285.
13. Florin TA, Zaoutis TE, Zaoutis LB. Beyond Cat Scratch Disease: Widening Spectrum of Bartonella henselae Infection. Pediatrics. 2008; 121: e1413-e1425.
14. Chang C, Chomel BB, Kasten RW, Tappero JW, Sanchez MA, Koehler JE. Molecular Epidemiology of Bartonella henselae Infection in Human Immunodeficiency Virus–Infected Patients and Their Cat Contacts, Using Pulsed Field Gel Electrophoresis and Genotyping. J Infect Dis. 2002; 186: 1733-1739.

15. Fiskus W, Padmalayam I, Kelly T, Guibao C, Baumstark BR. Identification and Characterization of the DdlB, FtsQ and FtsA Genes Upstream of FtsZ in Bartonella bacilliformis and Bartonella henselae. DNA Cell Biol. 2003; 22: 743-752.

16. Riess T, Raddatz G, Linke D, Schäfer A, Kempf VAJ. Analysis of Bartonella adhesin A expression reveals differences between various B. henselae strains. Infect Immun. 2007; 75: 35-43.

17. Guy L, Nystedt B, Sun Y, Näslund K, Berglund EC, Andersson SGE. A genome-wide study of recombination rate variation in Bartonella henselae. BMC Evol Biol. 2012; 12: 65.

18. Lindroos H, Vinnere O, Mira A, Repsilber D, Naslund K, Andersson SGE. Genome Rearrangements, Deletions, and Amplifications in the Natural Population of Bartonella henselae. J Bacteriol. 2006; 188: 7426-7439.

19. Berglund EC, Frank AC, Calteau A, Vinnere Pettersson O, Granberg F, Eriksson A-S, et al. Run-Off Replication of Host-Adaptability Genes Is Associated with Gene Transfer Agents in the Genome of Mouse-Infecting Bartonella grahamii. PLoS Genet. 2009; 5: e1000546.

20. Ågren J, Sundström A, Håfström T, Segerman B. Gegenees: Fragmented Alignment of Multiple Genomes for Determining Phylogenomic Distances and Genetic Signatures Unique for Specified Target Groups. PLoS One. 2012; 7: e39107.

21. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 2006; 23: 254-267.

22. Maiden MCJ, van Rensburg MJJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol. 2013; 11: 728-736.

23. Liu Y-Y, Chiou CS, Chen CC. PGAdb-builder: A web service tool for creating pan-genome allele database for molecular fine typing. Sci Rep. 2016; 6: 36213.

24. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016; 33: 1870-1874.

25. Darling AE, Mau B, Perna NT. Progressive Mauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLoS One. 2010; 5: e11147.

26. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015; 16: 157.

27. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002; 30: 1575-1584.

28. Soares SC, Silva A, Trost E, Blom J, Ramos R, Carneiro A, et al. The Pan-Genome of the Animal Pathogen Corynebacterium pseudotuberculosis Reveals Differences in Genome Plasticity between the Biovar ovis and equi Strains. PLoS One. 2013; 8: e53818.

29. Galperin MY, Kristensen DM, Makarova KS, Wolf YI, Koonin EV. Microbial genome analysis: The COG approach. Brief Bioinform. 2018.

30. Soares SC, Geyik H, Ramos RTJ, de Sá PHCG, Barbosa EGV, Baumbach J, et al. GIPSy: Genomic island prediction software. J Biotechnol. 2016; 232: 2-11.

31. da Silva Filho AC, Raittz RT, Guizelini D, De Pierri CR, Augusto DW, dos Santos-Weiss ICR, et al. Comparative Analysis of Genomic Island Prediction Tools. Front Genet. 2018; 9.

32. Kešnerová L, Moritz R, Engel P. Bartonella apis sp. nov., a honey bee gut symbiont of the class Alphaproteobacteria. Int J Syst Evol Microbiol. 2016; 66: 414-421.

33. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016; 44: W16-W21.

34. Srividhya K V., Rao G V., L R, Mehta P, Prilusky J, Manicka S, et al. Database and Comparative Identification of Prophages. Intell. Control Autom., vol. 344, Springer Berlin Heidelberg. 2006; 863-868.

35. Srividhya K V., Alaguraj V, Poornima G, Kumar D, Singh GP, Raghavenderan L, et al. Identification of Prophages in Bacterial Genomes by Dinucleotide Relative Abundance Difference. PLoS One 2007; 2: e1193.

36. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 2011; 12: 402.

37. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics. 2012; 28: 464-469.

38. Mao C, Abraham D, Wattam AR, Wilson MJC, Shukla M, Yoo HS, et al. Curation, integration and visualization of bacterial virulence factors in PATRIC. Bioinformatics. 2015; 31: 252-258.

39. Chen L, Xiong Z, Sun L, Yang J, Jin Q. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res. 2012; 40: D641-D645.

40. Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, Slezak T. MvirDB--a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res. 2007; 35: D391-D394.

41. Hatfull GF. Bacteriophage genomics. Curr Opin Microbiol. 2008; 11: 447-453.

42. Bouchouicha R, Durand B, Monteil M, Chomel BB, Berrich M, Arvand M, et al. Molecular Epidemiology of Feline and Human Bartonella henselae Isolates. Emerg Infect Dis. 2009; 15: 813-816.

43. Woudstra C, Fach P, Chomel BB, Haddad N, Boulouis HJ. Draft Genome Sequences of 12 FelineBartonella henselaeIsolates. Genome Announc. 2017; 5: 12-13.

44. Hacker J, Bender L, Ott M, Wingender J, Lund B, Marre R, et al. Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vitro* and *in vivo* in various extraintestinal Escherichia coli isolates. Microb Pathog. 1990; 8: 213-225.

45. Bentley R, Haslam E. The Shikimate Pathway-A Metabolic Tree with Many Branche. Crit Rev Biochem Mol Biol. 1990; 25: 307-384.

46. Karasova D, Sebkova A, Vrbas V, Havlickova H, Sisak F, Rychlik I. Comparative analysis of *Salmonella enterica serovar Enteritidis* mutants with a vaccine potential. Vaccine. 2009; 27: 5265-5270.

47. Homchampa P, Strugnell RA, Adler B. Molecular analysis of the aroA gene of *Pasteurella multocida* and vaccine potential of a constructed aroA mutant. Mol Microbiol. 1992; 6: 3585-3593.

48. Roberts M, Maskell D, Novotny P, Dougan G. Construction and characterization *in vivo* of Bordetella pertussis aroA mutants. Infect Immun. 1990; 58: 732-739.

49. Oyston PCF, Russell P, Williamson ED, Titball RW. An aroA mutant of Yersinia pestis is attenuated in guinea-pigs, but virulent in mice. Microbiology. 1996; 142: 1847-1853.

50. Felgner S, Frahm M, Kocijancic D, Rohde M, Eckweiler D, Bielecka A, et al. aroA -Deficient *Salmonella enterica Serovar Typhimurium* Is More Than a Metabolically Attenuated Mutant. MBio. 2016; 7: e01220-e01216.

51. Singh V, Chandra D, Srivastava BS, Srivastava R. Downregulation of Rv0189c, encoding a dihydroxyacid dehydratase, affects growth of Mycobacterium tuberculosis in vitro and in mice. Microbiology. 2011; 157: 38-46.

52. Flint DH, Emptage MH, Finnegan MG, Fu W, Johnson MK. The role and properties of the iron-sulfur cluster in Escherichia coli dihydroxy-acid dehydratase. J Biol Chem. 1993; 268: 14732-14742.

53. Grandoni JA, Marta PT, Schloss J V. Inhibitors of branched-chain amino acid biosynthesis as potential antituberculosis agents. J Antimicrob Chemother. 1998; 42: 475-482.

54. GUO J, SONG X, ZOU L, ZOU H, CHEN G. The small and large subunits of carbamoyl-phosphate synthase exhibit diverse contributions to pathogenicity in Xanthomonas citri subsp. citri. J Integr Agric. 2015; 14: 1338-1347.

55. Kwaga JKP, Allan BJ, Van den Hurk J V., Seida H, Potter AA. A carAB mutant

of avian pathogenic Escherichia coli serogroup O2 is attenuated and effective as a live oral vaccine against colibacillosis in Turkeys. Infect Immun. 1994; 62: 3766-3772.

56. Piao HH, Tam VTM, Na HS, Kim HJ, Ryu PY, Kim SY, et al. Immunological responses induced by asd and wzy/asd mutant strains of Salmonella enterica serovar Typhimurium in BALB/c mice. J Microbiol. 2010; 48: 486-495.

57. Vayssier-Taussat M, Rhun D Le, Deng HK, Biville F, Cescau S, Danchin A, et al. The Trw type IV secretion system of Bartonella mediates host-specific adhesion to erythrocytes. PLoS Pathog. 2010; 6: e1000946.

58. Saenz HL, Engel P, Stoeckli MC, Lanz C, Raddatz G, Vayssier-Taussat M, et al. Genomic analysis of Bartonella identifies type IV secretion systems as host adaptability factors. Nat Genet. 2007; 39: 1469-1476.

59. Saisongkorh W, Robert C, La Scola B, Raoult D, Rolain J-M. Evidence of Transfer by Conjugation of Type IV Secretion System Genes between Bartonella Species and Rhizobium radiobacter in Amoeba. PLoS One. 2010; 5: e12666.