Received: 11 February 2022 | Revised: 12 May 2022 | Accepted: 20 May 2022

DOI: 10.1111/mec.16547

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Genomic structural plasticity of rodent-associated *Bartonella* in nature

Keyla Carstens Marques de Sousa[1] | Ricardo Gutiérrez[1] | Dayana Yahalomi[2] | Tali Shalit[2] | Barak Markus[2] | Yaarit Nachum-Biala[1] | Hadas Hawlena[3] | Evgeniya Marcos-Hadad[4] | Einat Hazkani-Covo[5] | Haroldo Henrique de Rezende Neves[6] | Shay Covo[4] | Shimon Harrus[1]

[1]Koret School of Veterinary Medicine, The Hebrew University of Jerusalem, Rehovot, Israel

[2]The Mantoux Bioinformatics Institute of the Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science, Rehovot, Israel

[3]Mitrani Department of Desert Ecology, Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel

[4]Department of Plant Pathology and Microbiology, Robert H. Smith Faculty of Agriculture, The Hebrew University of Jerusalem, Rehovot, Israel

[5]Department of Natural and Life Sciences, Open University of Israel, Raanana, Israel

[6]Gensys Associated Consultants, Porto Alegre, Rio Grande do Sul, Brazil

Correspondence
Shimon Harrus, Koret School of Veterinary Medicine, The Hebrew University of Jerusalem, Rehovot, Israel.
Email: shimon.harrus@mail.huji.ac.il

Present address
Ricardo Gutiérrez, National Reference Center for Bacteriology, Costa Rican Institute for Research and Teaching in Nutrition and Health (INCIENSA), Tres Ríos, Costa Rica

Funding information
Israel Science Foundation, Grant/Award Number: 688/17; NSF-BSF, Grant/Award Number: 2017708

Abstract

Rodent-associated *Bartonella* species have shown a remarkable genetic diversity and pathogenic potential. To further explore the extent of the natural intraspecific genomic variation and its potential role as an evolutionary driver, we focused on a single genetically diverse *Bartonella* species, *Bartonella krasnovii*, which circulates among gerbils and their associated fleas. Twenty genomes from 16 different *B. krasnovii* genotypes were fully characterized through a genome sequencing assay (using short and long read sequencing), pulse field gel electrophoresis (PFGE), and PCR validation. Genomic analyses were performed in comparison to the *B. krasnovii* strain OE 1–1. While, single nucleotide polymorphism represented only a 0.3% of the genome variation, structural diversity was identified in these genomes, with an average of 51 ± 24 structural variation (SV) events per genome. Interestingly, a large proportion of the SVs (>40%) was associated with prophages. Further analyses revealed that most of the SVs, and prophage insertions were found at the chromosome replication termination site (*ter*), suggesting this site as a plastic zone of the *B. krasnovii* chromosome. Accordingly, six genomes were found to be unbalanced, and essential genes near the *ter* showed a shift between the leading and lagging strands, revealing the SV effect on these genomes. In summary, our findings demonstrate the extensive genomic diversity harbored by wild *B. krasnovii* strains and suggests that its diversification is initially promoted by structural changes, probably driven by phages. These events may constantly feed the system with novel genotypes that ultimately lead to inter- and intraspecies competition and adaptation.

KEYWORDS
Bartonella, rodents, structural variation, phages

Handling Editor: Kin-Ming (Clement) Tsui

Keyla Carstens Marques de Sousa and Ricardo Gutiérrez contributed equally to this work.

1 | INTRODUCTION

The understanding of genetic mechanisms involved in bacterial diversification and adaptation is fundamental to identify imprints associated with the emergence of pathogenicity and antibiotic resistance. With the development of advanced sequencing platforms, trustworthy assemblies could be constructed from complex bacterial genomes (De Maio et al., 2019). Since genomic comparisons have been performed, the dynamic nature of bacterial genome organization was elucidated. In order to preserve important intrinsic mechanisms such as replication, transcription and translation, prokaryotic genomes endure high selection-pressure to keep their structural stability through generations (Touchon & Rocha, 2016). However, the constant DNA exchange due to horizontal gene transfer (HGT) and exposure to fluctuating ecological niches, cause the instability of genomes. Genome instability is marked by evolutionary differences between gene repertories and genomic organizations (Cui et al., 2012; Iguchi et al., 2006; Liu et al., 2006; Scott et al., 2007; Yan et al., 2018). The evolution of genomes is shaped by the equilibrium between the maintenance of the genome organization and tolerance to instability. Instability plays a key role in adaptation and survival of bacteria, particularly in double niche bacteria, such as *Bartonella* species that constantly circulate between arthropod vectors and mammal hosts (Bochkareva et al., 2018; Darmon & Leach, 2014; Dobrindt & Hacker, 2001).

The genomic instability of bacterial chromosomes caused by structural variations (SVs) of large DNA fragments, including deletions, insertions, inversions and translocations, might result in altered genetic information and gene clusters (i.e., operons), and may interfere directly in the transcription and translation patterns of the bacterial machinery (Cerdeño-Tárraga et al., 2005; Periwal & Scaria, 2015; Sheppard et al., 2018). Among the evolutionary elements involved in the genome instability, mobile elements can promote genomic changes directly through the exchange of genetic information between genomes (Cui et al., 2012; Nzabarushimana & Tang, 2018) and indirectly through the activation of DNA damage-repair mechanisms by its integration or excision within the bacterial chromosome (Argueso et al., 2008; Hoff et al., 2018). Bacteriophages are a type of mobile elements that have a key role in the ecology and evolution of bacterial populations through the incorporation of external DNA into the bacterial chromosome, as well as selective weapons between populations (Bobay et al., 2014). Integrated bacteriophages, or prophages, have a relevant function in the bacterial adaptation, potentially enhancing bacterial fitness and resistance to foreign bacteriophages, as well as the dissemination of genes that confer virulence and antibiotic resistance (Howard-Varona et al., 2018). Moreover, phages can induce SVs in the recipient bacteria through HGT (Ramisetty & Sudhakari, 2019) and vertically, without any influence of foreign genetic information, as shown in an evolutionary study with *Bartonella* spp. (Gutierrez, Markus, et al., 2018).

The *Bartonella* genus belongs to the order *Rhizobiales*, which comprises bacteria with diverse lifestyle (Neuvonen et al., 2016). It has been proposed that *Rhizobiales* evolved from free-living ancestors to host-associated bacteria (Wang et al., 2020). *Bartonella* species circulate between intracellular (within their mammalian hosts) and extracellular milieus (within arthropod guts), representing an intriguing genus-model to study adaptation and evolution within the order *Rhizobiales* (Gutiérrez et al., 2015; Kosoy et al., 2012). Moreover, the diversity among *Bartonella* is notable, with 37 validated *Bartonella* species and three subspecies, and over 25 organisms with a *Candidatus* status or unclarified phylogeny (https://www.bacterio.net/genus/bartonella). Among the extensive list of *Bartonella* reservoirs, wild rodents and their associated fleas harbor one of the largest *Bartonella* diversity described to date (Gutiérrez et al., 2014, 2015; Inoue et al., 2009; Morick et al., 2010; Paziewska et al., 2011). The diversity observed in rodent-associated *Bartonella* has been evident both at the nucleotide level (i.e. hundreds of genotypes identified based on genetic polymorphisms) and at the structural genomic level (Gutiérrez, Cohen, et al., 2018). Remarkably, recombination events between rodent-associated *Bartonella* spp. have been extensively reported, incriminating it as the main mechanism for the diversity observed in nature (Gutiérrez, Cohen, et al., 2018; Morick et al., 2010; Paziewska et al., 2011). Hence, the vast evidence accumulated in the last two decades on rodent-associated *Bartonella*, reflects their marked genomic plasticity, which deserves special attention due to their role as zoonotic and animal pathogens.

Our group has recently isolated and characterized four novel *Bartonella* species, represented by 38 different genotypes (Gutiérrez, Cohen, et al., 2018). These organisms were isolated from the blood and fleas of two sympatric rodent species, sampled in two sites in the Negev Desert, southern Israel, circumscribed within a 3-km distance area (Gutiérrez, Cohen, et al., 2018). Coinfections with these rodent-associated *Bartonella* spp. and genotypes were commonly observed among these hosts (in over 50% of the hosts), providing ideal conditions for HGT and recombination events (Gutiérrez et al., 2014; Gutiérrez, Cohen et al., 2018). From the *Bartonella* spp. identified, *Bartonella krasnovii* stands out as an example of the genomic plasticity within this genus, with 16 genotypes identified, all showing different genomic structures (i.e., different enzymatic restriction profiles) (Gutiérrez, Cohen, et al., 2018). Furthermore, under laboratory conditions, it was shown that *B. krasnovii* can undergo spontaneous SV events during a short period of evolution, while single-point mutations are kept restricted (Gutierrez, Markus, et al., 2018). Interestingly, these SVs were associated with prophages, raising the hypothesis that these genetic elements promote rapid occurrence of genomic variation in this bacterial genus (Gutierrez, Markus, et al., 2018). In accordance with the latter, SVs have been previously observed among *Bartonella* genomes (Alsmark et al., 2004; Gutierrez, Cohen, et al., 2018; Lindroos et al., 2006), however, the genomic extension of this phenomenon in wild strains has been only partially explored.

Considering the importance of rearrangements for bacterial adaptation and variant emergence, the present study explored the extent of genomic variation events in nature, by comparing the diversity observed at nucleotide and gene content levels with

structural genomic diversity. For this aim, a collection of *B. krasnovii* strains, isolated at the same period and location, was chosen, and the genetic content and genomic organization were evaluated both independently and by comparative analyses against OE 1–1 strain genome (isolated and sequenced 2 years earlier). In order to evaluate the association of the structural rearrangements and prophages, a genomic comparison with *Bartonella quintana* (an acknowledged *Bartonella* species lacking of prophages) was also performed. The results of this study shed light on the genetic elements involved in the variability and genomic plasticity in the *Bartonella* genus.

2 | MATERIALS AND METHODS

2.1 | Bacterial strains and culture conditions

Bartonella krasnovii strains were isolated from 10 wild *Gerbillus andersoni* and 10 *Gerbillus pyramidum* and 40 associated *Synosternus cleopatrae* fleas, captured during October 2016 in two different sand dune sites (~4 km apart) in the Negev Desert, Israel (34°23′E, 30°58′N and 34°23′E, 30°55′N) (Gutiérrez et al., 2018) (Figure 1a). These *B. krasnovii* strains were previously characterized into 16 genotypes (named A1-A5, B, C1-C3, D1-D5, G and H) based on the identification of nucleotide polymorphisms at three genetic and intergenetic loci (i.e., *gltA*, *rpoB* and ITS); (Gutiérrez, Cohen, et al., 2018; Figure 1b). All strains were cultured on chocolate agar plates (Novamed, Ltd., Jerusalem, ISR), incubated at 37°C with constant 5% CO_2 atmosphere. Despite the fact all strains belong to the same species, they presented differences in colony morphotype and growth rates (Table S1). To further study their genomic diversity, one to three different *B. krasnovii* strains per genotype (for a total of 46 strains) were selected. The genomes of a total of 46 *Bartonella* strains were analysed by the *SmaI* digestion profiles using pulse field gel electrophoresis (PFGE), following previously described methods (Gutiérrez, Markus, et al., 2018). The observed *SmaI* restriction profiles were carefully and manually checked, using *Salmonella* ser. Braenderup H9812 (i.e., plugs digested with *XbaI*) as molecular marker in each run. Seventeen strains with different PFGE profiles, and at least one different *Bartonella* strain per genotype (for a total of 20 strains) were chosen for sequencing analyses (Figures 1b–c).

Bartonella krasnovii strain OE 1–1 (Gutiérrez et al., 2020), isolated 2 years earlier, from a *S. cleopatrae* flea, which its genome was fully characterized and curated (CP031844.2, CP042965.1), was used as a reference for genome and genetic comparisons. This isolate was classified as genotype "A2" according to the polymorphism analysis of genetic and intergenic loci described above.

2.2 | Whole genome sequencing

The DNA from each pure strain (retrieved after 3–4 consecutive single-colony passages on chocolate agar plates) was extracted with DNeasy Blood and Tissue Kit (Qiagen), following the manufacturer's recommendations. The DNA concentration, as well as the presence of chemical contaminants, were measured using the spectrophotometer NanoDrop (Thermo Fisher Scientific), and the DNA integrity was checked on agarose gels. DNA extracted from the 20 selected strains were processed on two different sequencing platforms, namely short reads sequencing through Illumina MiSeq sequencing (150 bp paired), and long reads sequencing through MinION Mk1b (Oxford Nanopore Technologies, Oxford, GBR) using SpotON flow cell (see Methods S1).

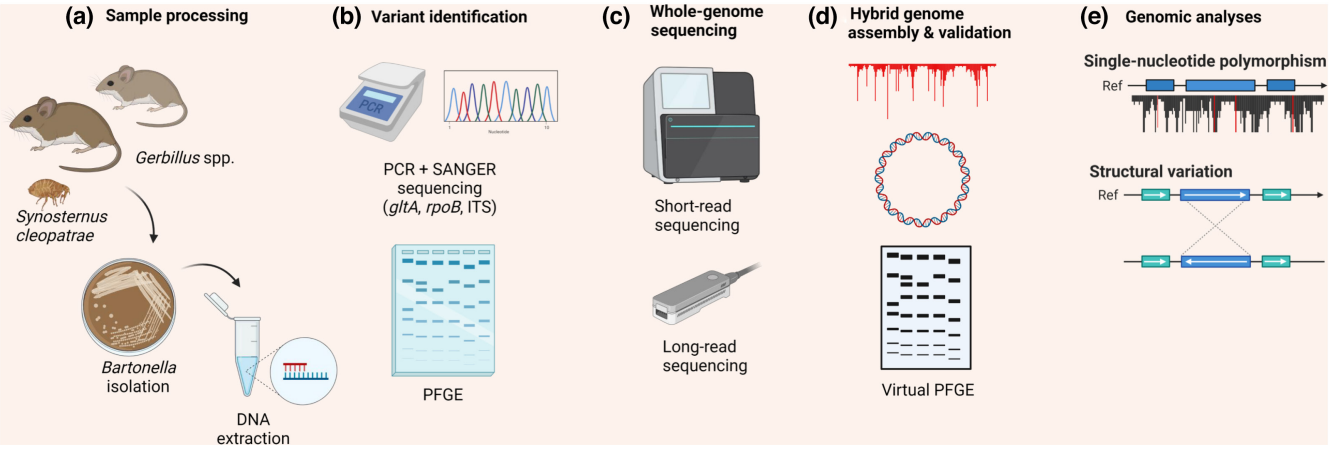

FIGURE 1 Study design. (a) Sample processing: Isolation of rodent-associated *Bartonella* from gerbils and their associated fleas and further DNA extraction. (b) Variant identification: Genotyping based on SNP of *Bartonella* conserved loci and *SmaI*-PFGE profiles. (c) Whole-genome sequencing: 20 different strains, comprising the 16 sequencing polymorphic genotypes and 17 different PFGE variants were selected to WGS through two different sequencing platforms, short reads (Illumina MiSeq sequencing) and long reads (MinION, Oxford Nanopore). (d) Genome assembly and validation: The genomes were assembled using pipelines which employ short and/or long reads, and the assembly outputs were validated by the comparison of in silico *SmaI*-digested profiles and the "real" PFGE profiles. PCR assays of flanking regions were carried out. (e) Genomic analyses: The genomes were submitted to SNP calling analysis and only the validated assemblies were submitted to SV analysis. This figure was created using Biorender (http://BioRender.com; last accessed 29 December 2021)

2.3 | Genome assembly validation

The genomes were assembled using both Unicycler version v0.4.9 (Wick et al., 2017), which combined short and long reads, and Canu version v2.2 (Koren et al., 2017), which uses long reads. Then, both resulting assemblies were compared to the PFGE profiles, as follows. First, the assembly outputs were characterized according to the quality and fragmentation status, using assembly-stats (version 3). Then, the fasta files from the largest contig obtained from the assemblies (if this was above 2.0 Mb length) were analysed in Geneious version 7.1.9 (Kearse et al., 2012). The linear contigs were circularized in Geneious and we set (arbitrarily) as the starting site the chromosomal replication initiator protein *dnaA* gene. Once circularized, the assemblies were in silico *SmaI*-digested in Geneious, producing a "virtual gel" of the restriction fragment profile. Genome assemblies were then validated (Figure 1d), testing whether their "virtual" restriction fragment profiles and estimated sizes match the "real" PFGE patterns. The resulting validated assemblies are shown in Table S2. Furthermore, a PCR assay per genome assembly was designed to amplify two flanking regions produced by *SmaI* restriction sites (Table S3). Therefore, only assemblies with matched PFGE and virtual gels and PCR-positive for the confirmation of flanking regions, were considered as structural-validated genomes.

All genome assemblies were included for nucleotide variation analyses, while only structural-validated genomes were analysed for the presence of SVs (Figure 1e).

2.4 | Average nucleotide identity

Average nucleotide identity (ANI) (Goris et al., 2007) was calculated between all the study *B. krasnovii* assemblies (i.e., validated and unvalidated) and the OE 1–1strain *B. krasnovii* genome (Gutiérrez et al., 2020), using FastANI (Jain et al., 2018).

2.5 | Single nucleotide polymorphisms (SNP) calling analysis

The SNP calling analysis was performed using the Illumina sequenced short-reads obtained from the 20 *B. krasnovii* strains mapped to the OE 1–1 strain genome (CP031844.2, CP042965.1), as previously described earlier (Gutiérrez, Markus, et al., 2018; see Methods S1).

2.6 | Identification of selection signatures

The McDonald and Kreitman Test (MKT) was performed to evaluate non-neutral evolution by comparing the ratio of nonsynonymous to synonymous polymorphisms within species (Pn/Ps) to the ratio of nonsynonymous to synonymous fixed differences between species (Dn/Ds), using the PopGenome package in R (McDonald & Kreitman, 1991; Pfeifer et al., 2014). The outgroup genome used for this test was the *Bartonella birtlesii* genome (GCA_000278095.1). In order to minimize any biases promoted by slightly deleterious mutations, low-frequency polymorphisms (MAF < 0.1) were removed from this analysis, following recommendation described elsewhere (Charlesworth & Eyre-Walker, 2008). The test was applied at the genomic level and for each coding-gene separately, in which case *p*-values were adjusted by false discovery rate (FDR).

2.7 | Genome annotation and characterization

All genomes were annotated with Prokka (Seemann, 2014). Additional annotations were performed on all structural-validated genomes using RASTtk (Brettin et al., 2015), implemented in the PATRIC portal (Wattam et al., 2017), and the Microbial Genomes Atlas Online (MiGA) (Rodriguez-R et al., 2018).

Prophages were annotated on all structural-validated genomes using Phaster annotation tool (Arndt et al., 2016; online server; http://phaster.ca, last accessed 3 November 2019). This tool identifies prophage regions from nucleotide sequences using gene prediction and local and GenBank annotation tools. In this way, Phaster is able to provide annotations regarding the phage position, length, boundaries, number of genes, attachment sites (*att*), tRNAs and phage-like genes. The annotations allow the prediction of completeness or potential viability of identified prophages and classifies them as intact, incomplete or questionable. Additionally, clusters of at least 4 phage-related genes were manually identified from the genome annotations (i.e., RASTtk), within a window of 5 up to 30 kb, and classified as "prophage-remnants". To identify possible relationships among the prophages, the principal component analysis (PCA) was done based on the distance matrix of the prophage nucleotide identities (see Methods S1).

Plasmids were screened in all genome assemblies as follows: (i) through the alignment of the short reads to the OE 1–1 strain plasmid (CP042965.1); (ii) searching conjugative transfer plasmid genes (*traJ*, *trbG*, *trbH*, *trbI*, *trbG*, *traF*, *traD*, *trbB*, *trbC*, *trbD*, *trbJ*, *trbL*) through tblastx (Boratyn et al., 2013) on the strain's assemblies; and (iii) through PlasmidVerify tool (Antipov et al., 2019).

2.8 | Pangenomic analyses

To estimate the pangenome of *B. krasnovii*, the genomes were analysed following Panaroo (Tonkin-Hill et al., 2020). To study further the sequence diversity at the gene level, predicted gene groups obtained from Panaroo were used to calculate ANI using pANIto (version 1), and the percentage of similarity of each core and soft genes was compared against the strain OE 1–1 genome.

2.9 | Replication origin and termination loci's prediction

The replication origin (*ori*) and termination (*ter*) loci were predicted for each structural-validated assembly using the GC skew analysis

(Lobry, 1996) as previously described by Neuvonen et al. (2016) for *Bartonella* species, through the g-language system (Arakawa et al., 2008), using the online server (http://soap.g-language.org/gembassy/emboss_explorer; last accessed 22 October 2019) and through Oriloc program (Frank & Lobry, 2000). The latter procedure was validated through the GC skew analysis performed for a curated genome of *Bartonella tribocorum* (NC_010161.1) and compared to an updated database of replication origins in prokaryotic genomes DoriC 10.0 (Luo & Gao, 2019) available online (http://tubic.org/doric/public/index.php/index; last accessed 25 October 2019). The starting position and sense for all structural-validated assemblies was defined using the annotation coordinates of the chromosomal replication initiator protein (*dnaA*) gene (position 1 of all genomes).

2.10 | Structural variation analyses

To identify SVs in the structural-validated genomes, pairwise whole genome alignments were obtained from each *B. krasnovii* assembly and the strain OE 1–1 genome using progressiveMauve (Darling et al., 2004, 2010), implemented in Geneious version 7.1.9 (Kearse et al., 2012). Accordingly, progressiveMauve alignments allowed the identification of inverted or translocated DNA regions (i.e., inversion and translocation events) within the genomes in comparison with the coordinates of its homologue regions within the strain OE 1–1 genome. On the other hand, deletions and insertions were identified using the short-read alignment data obtained through GATK pipeline (McKenna et al., 2010; see Methods S1). To minimize false SV events, a cutoff of ≥1000 bp length per event was employed (i.e., all events <1000 bp length were excluded). The flanking regions (around 1 kb) of the SVs were explored from the genome annotation to identify the loci involved in the events.

To estimate the impact of the SVs on the study genomes, we explored the distance between the *ori*- and *ter*-regions, which gave an estimate of the balance of the genomes. Accordingly, an off-balanced or unbalanced genome was defined when the replicores lengths (i.e., *ori*-to-*ter* and *ter*-to-*ori* distances) resulted in a difference ≥ 40 kb, which represents a ≥2% difference of a 2.0 Mb circular genome (Kothapalli et al., 2005). In addition, using the genomic coordinates of the essential genes (n = 105) predicted by MiGA, we explored whether SVs affected their distribution (as inversions may cause the shifting on genes from leading to lagging strand and vice versa) and distance between consecutive essential genes (as translocations/deletions/insertions may impact the distance between these genes). For this aim, using the predicted *ori*- and *ter*-regions, we defined the leading and lagging strands in the assemblies, being the leading strands of a circular genome the positive sense of the linear DNA segment from *ori* to *ter*, and the negative sense of the linear DNA segment from *ter* to *ori*. Then, using the genomic coordinates and strand directionality, we predicted the orientation on the leading or lagging strands of essential genes and the distance between the stop and starting position of consecutive essential genes.

Phylogenetic inference of the nucleotide-based tree and the SV-based tree was performed to estimate the implication of the SV events on the evolution of these *B. krasnovii* strains (see Methods S1). The plots corresponding to the genome comparisons and SVs phylogeny were generated through genoPlotR (Guy et al., 2011).

2.11 | Bartonella quintana comparison

To evaluate the potential implication of the prophages on the structural organization of *Bartonella* genomes, a comparison of intraspecies variation with the *B. quintana*, an acknowledged *Bartonella* species deprived of prophages (Alsmark et al., 2004) was chosen to serve as a control. Accordingly, five fully circulated *B. quintana* genomes available in GenBank (https://www.ncbi.nlm.nih.gov/; last accessed 4 October 2021) were downloaded and annotated (following the above-described methods). The included genomes represented a worldwide collection of *B. quintana*, as they were obtained from different hosts and geographical regions: *B. quintana* strain "Toulouse" (NC_005955), isolated from a human patient from France in 1993 (Alsmark et al. 2004), *B. quintana* strain "MF1-1" (NZ_AP019773.1), isolated from a monkey from Japan in 2012 (Sato et al., 2015), *B. quintana* strain "RM-11" (CP003784.1) a monkey, isolated from China in 2011 (Li et al., 2012), *B. quintana* strain KorN (CP041670), isolated from a human patient from South Korea in 2017; and *B. quintana* strain "NCTC12899" (NZ_LS483373.1), collection date 2016, for which isolation source is not available nor published.

3 | RESULTS

3.1 | Genome sequencing results

Illumina sequencing resulted in 2,889,782–5,647,059 paired-end 150-bp reads per sample (3,886,435 ± 828,284 reads; average ± SD). MinION Mk1b sequencing resulted in 90,011–501,921 long reads (with average of 200,463 ± 88,815) per sample, with an average read length of 6881.4 ± 1189.8 bp. The genome assembly lengths ranged from 2,023,468–2,260,108 bp, with an average of 2,143,012 ± 62,848 bp (±SD). GC% content varied from 37.9–38.2%, average 38.1 ± 0.11%.

3.2 | Validation of genomic assemblies

Fourteen out of 20 genomes (70%) were structurally-validated and confirmed as circular contigs. The structural-validated genomes ranged between 2,047,607–2,210,336 bp of length, GC content at the range of 37.9–38.2%, number of CDS from 2021 to 2347, repeat regions from 52 to 136, and all genomes carried 42 tRNAs and 6 rRNAs (Table S2).

3.3 | Nucleotide variation analyses

The average number of SNPs between the OE 1–1 strain and the *B. krasnovii* strains was 6381 ± 2125 SNPs per genome, ranging from 295–8667. On average, each genome carried 2756 ± 907 synonymous and 2015 ± 686 nonsynonymous mutations in their coding-sequences (CDS) (Table 1).

Subsequent SNP allele frequency analysis highlighted seven regions with greater alternative allele frequency across the 20 genomes (Figure 2a), as follows: (i) hemin binding protein genes (peak 1); (ii) locus of the cytochrome C-type biogenesis protein gene, DNA-methyltransferase and pyruvate phosphate dikinase (PPDK) genes (peaks 2 and 3); (iii) *Bartonella* effector protein genes (Beps) for intracellular delivery domain (BID) containing VirB/VirD4 type-IV-secretion system (T4SS) (peak 4); (iv) VirB/VirD4-homologous T4SS (*vbh*) regions (peaks 5 and 6); and (v) in Trw-T4SS (*trw*) genes region (peak 7) (Figure S1).

The McDonald and Kreitman test showed a low signal of positive selection with neutrality index (NI) of 0.91 (Fisher's test: $p = 9.2E\text{-}09$). The analysis of all CDS separately across the genomes by MKT, after FDR adjustment, resulted in only two genes with a significant selection signal. The first gene (identified under positive selection) was a tyrosine-type recombinase/integrase, located within an intact prophage (NI = 0.052, $\alpha = 0.9$, adjusted-$p = 0.03$). The second gene (identified under negative selection was *brrF* (*Bartonella* run-off replication of host-adaptability gene F; NI = 15.6, $\alpha = -14.6$, adjusted-$p = 0.03$; Figure 2b).

3.4 | Pangenomic analysis

An average of 1780 CDS, 42 tRNA, 6 rRNA and 1 tmRNA were predicted in the genomes. The pangenome analysis showed that from 2464 total genes, 1391 (60.0%) were classified as core genes (in ≥99% of the strains), 55 (2.4%) as soft-core genes (in ≥95 to <99% of the strains), 634 (27.4%) as shell genes (in ≥15 to <95% of the strains), and 403 (16.3%) as cloud genes (in <15% of the strains). The cloud gene pool was composed mainly by uncharacterized hypothetical proteins (37.5%) and phage associated genes (14.4%) followed by genes from other evolutionary origins, including genomic island proteins, transposon related proteins, T4SS-related, NADH-quinone related, adhesins, metabolic, transporters, filamentous haemagglutinin, and toxin-antitoxin systems. The great majority of the cloud genes (67.5%) were unique to a single genome, followed by 98 (24.3%) found in two genomes and only 33 (8.2%) found in three genomes (Figure S2). From a total of 1446 core and soft genes, only 33 genes (2%) showed ANI values below 97% in comparison to the strain OE 1–1 genome (Figure S3).

3.5 | Plasmid exploration

Plasmids were identified only in five strains (25%; BKF1, BKF2, BKR2, BKR8 and BKF9; PlasmidVerify log-likelihood ratio > 65; all circular predicted by Unicycler) that belonged to three genotypes (A2, A3, and A5). All genotypes carried a circular plasmid-contig of 29 kb length. The sequences showed 99.8%–100% ANI between strains and the *B. krasnovii* strain OE 1–1 plasmid (CP042965.1).

3.6 | *Bartonella* prophages

The annotations allowed the identification of 120 prophages among the 14 *B. krasnovii* structurally-validated genomes (Figure 3a). Most of the prophages were identified as "remnant prophage loci" (43.3%, 52/120), followed by "incomplete

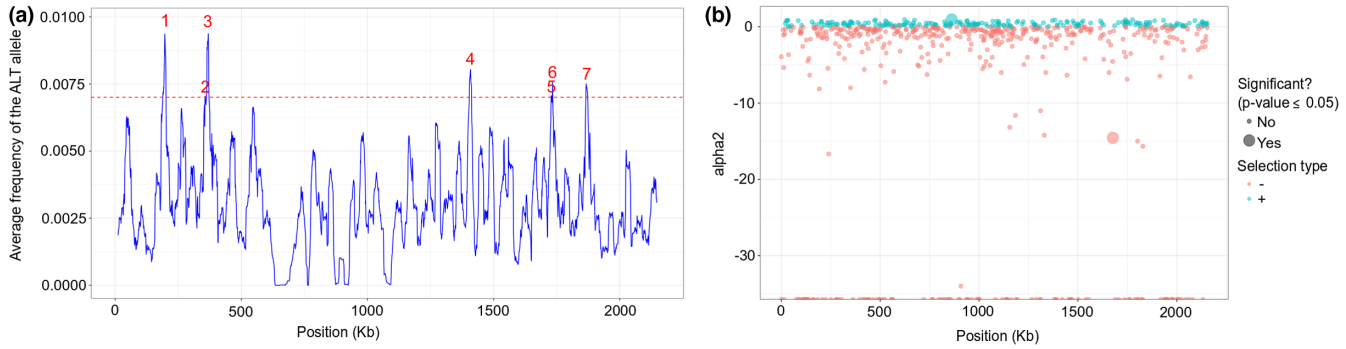

FIGURE 2 Nucleotide variation of wild *Bartonella krasnovii* strains is restricted to metabolic and virulence factor loci, exhibiting a minimal selection signature. (a) Frequency of SNPs on the studied genotypes per 20 KB window of the OE 1–1 *B. kransnovii* genome. Numbers refer to regions with greater alternative allele frequency ("peaks") that are above the threshold of distribution, accordingly the following numbers represent: 1. Hemin binding proteins; 2 & 3. Cytochrome C-type biogenesis proteins, DNA-methyl transferase and pyruvate phosphate dikinase; 4. BID domain containing VirB/VirD4 type-IV-secretion system (T4SS) effectors; 5 & 6. VirB/VirD4-homologous T4SS (Vbh) and 7. Trw-T4SS (trw) genes region (Trw). (b) Distribution of α values identified in coding genes analysed separately across the genomes through McDonald and Kreitman Test (MKT) using *Bartonella birtlesii* as an outgroup. The red dots represent genes under negative selection ($\alpha \leq 0$) and the blue dots represent genes under positive selection ($\alpha \geq 0$). The big dots represent genes with significant values ($p \leq .05$). All *p*-values were adjusted by false discovery rate (FDR)

prophages" (25%, 30/120), "intact prophages" (17.5%, 21/120) and "questionable prophages" (2.5%, 3/120). Notably, the *Bartonella* gene transfer agent (BaGTA) locus was identified in all genomes (Figure 3a). The genetic distances between the prophages of the 14 *B. krasnovii* validated genomes, showed three distinct clusters on the PCA plot: cluster (i) composed by 15 intact prophages and one questionable prophage; cluster (ii) composed by BaGTA loci only (14); and, cluster (iii) composed by all remnant and incomplete prophages together with some intact (6) and questionable (2) prophages (Figure 3b). The ANI pairwise comparison between the prophages, ranged from 74.4–100%. The ANI between the intact and questionable prophages from cluster 1 was 95.6% (84.2–99.9%), cluster 2 was 87.3% (74.4–100%) and cluster 3 (composed only by BaGTA) was 99.7% (99.5–100%).

3.7 | SV identification and effect estimation

From the 14 structural-validated genomes, an average of 51 SV events were identified per genome in comparison to the OE 1–1 strain (Table 2, Figure 4a). Accordingly, all but one genome showed different profiles than strain OE 1–1 in their PFGE restriction pattern. Deletions and insertions, namely loci (≥1000 bp) without mapped reads from the studied strains or OE 1–1 strain genome, respectively, were the most frequent events identified (Table 2). No SV events were identified in the plasmids (Figure S4).

The mauve-based multiple alignment (Figure 4b) of 23 locally collinear blocks (LCB) (Darling et al., 2004, 2010), did not enable a conclusive *Bartonella* phylogeny based on the SVs. The phylogentic tree based on the SVs was poorly resolved and could support only three clades with high bootstrap values (Figure S5). The nucleotide-based maximum likelihood (ML) phylogenetic tree based on 2042 orthologous groups (predicted by Microbializer; see Methods S1) showed high bootstrap values, except for the basal splits and the placements of BKF7, BKR3 and BKF3 strains (Figure 4c). The three clades supported by significant bootstrap values in the SV-based tree showed the same topology of the supported clades of the nucleotide-based tree. The SV tree suggests that strains BKR3 and BKF3, whose placement is poorly resolved in the nucleotide-based tree, are closely related to strains BKF5, BKR6, BKR7 and BKR5 (Figure 4c, Figure S5).

The exploration of the flanking areas of the identified rearrangements showed that prophages were the loci mostly associated with the events (Figure 5). Accordingly, 48% (15/31) and 42% (16/38) of the inversion and translocation events were associated with prophages, respectively. Similarly, 62% (31/50) of insertions and translocations co-occurring events were also associated with prophages. Uncharacterized hypothetical proteins were the second main element associated with 26% (8/31) of inversions, 34% (13/38) of translocations and 24% (12/50) of co-occurring events, followed by association with 14 different genes. Forty percent (113/284) and 56% (178/317) of the insertion and deletion events were also found to be associated with prophages, respectively (Table S4). Deletion and insertion events were also noticed among virulence factors associated genes, being 7.5% of the total events associated with *trw* genes, followed by 3.4% associated with *Bartonella* adhesins associated genes (Figures S6 and S7). Regarding the association of virulence factors associated genes with inversions and translocations, only one inversion event was flanked by *trwJ4* and *trwI4* genes in one sample (BKR1) solely.

The distribution of SV (especially inversions) and prophage insertions shows that most events occurred near the terminus (*ter*) region of replication of the chromosome (*ter*-region) (Figure 6a and b, Figure S8). Notably, while "intact" prophages were distributed in three different regions across the chromosomes, and the BaGTAs were located out of the *ter*-region, the majority of "remnant" and "incomplete" prophages were found in the *ter*-region (Figure 6b).

The SV events caused an effect in the balance level of some chromosomes (i.e., *ori*-to-*ter* distance). Accordingly, six genomes were classified as unbalanced, with up to 5.8% shifts between their theoretical middle position and the *ter*-region, representing a difference of up to 126 kb (Table 3). Remarkably, the highest shifts were recorded among the A2 genotypes. From 105 essential genes analysed, only genes found near the *ter*-region showed a re-distribution between leading and lagging strands (Table S5). These genes were the *gyrA*, arginine--tRNA ligase, aspartate--tRNA ligase, proline--tRNA ligase, tyrosine--tRNA ligase, 30S ribosomal protein S4 and threonine--tRNA ligase. Notably, the total number of essential genes in the leading and the lagging strands per genome was always found the same (78 and 27, respectable, Table 3). SVs also affected the distance between essential genes, increasing gene-to-gene distance in up to 120 Kb (Table S5).

3.8 | SV events in *Bartonella quintana*

The comparison of five *B. quintana* genomes isolated from different hosts and geographical locations showed only one SV event, that is, a large inversion in the genomes of the strains MF1-1, Korn and NCTC (Figure 7). The inverted segment was of approximately 700 Kb long and was flanked by an excinuclease subunit B and ribonuclease genes, in addition to repeated regions located upstream and downstream of the inverted event. Interestingly, Phaster annotation tool identified two small (6.7 and 7.1 kb) regions classified as "incomplete" prophages within the five *B. quintana* genomes studied (four genomes presented both regions, while one harbored only one). However, these regions contained a maximum of two genes with clear phage origin, namely genes showing similarities to the *gp245* and *gp59* proteins from the bacteriophage T4, and minor tail and baseplate wedge subunit related genes.

4 | DISCUSSION

This study offers a close look at the extent of the genomic diversity of co-circulating wild *B. krasnovii* genotypes in discrete ecological

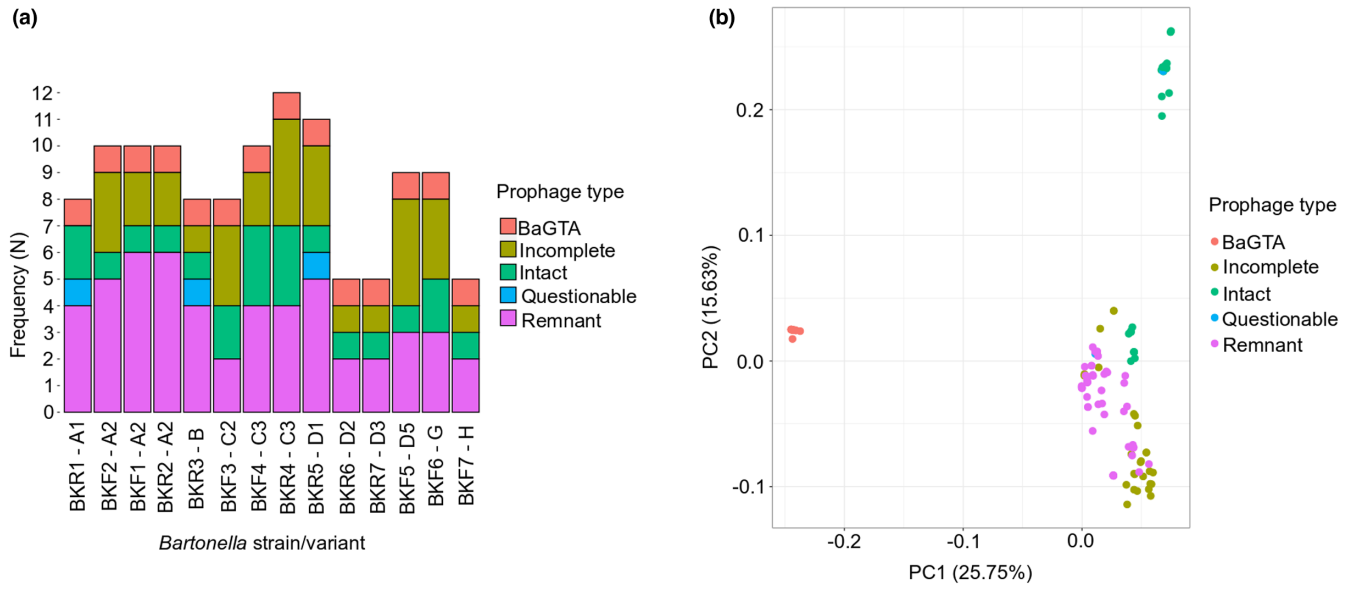

FIGURE 3 Distribution of prophages and similarity across the study genomes. (a) Frequency and type of prophages identified among 14 *Bartonella krasnovii* genotypes. (b) Principal component analysis (PCA) of the identified prophages based on nucleotide similarities. Dots are coloured according to the prophage type

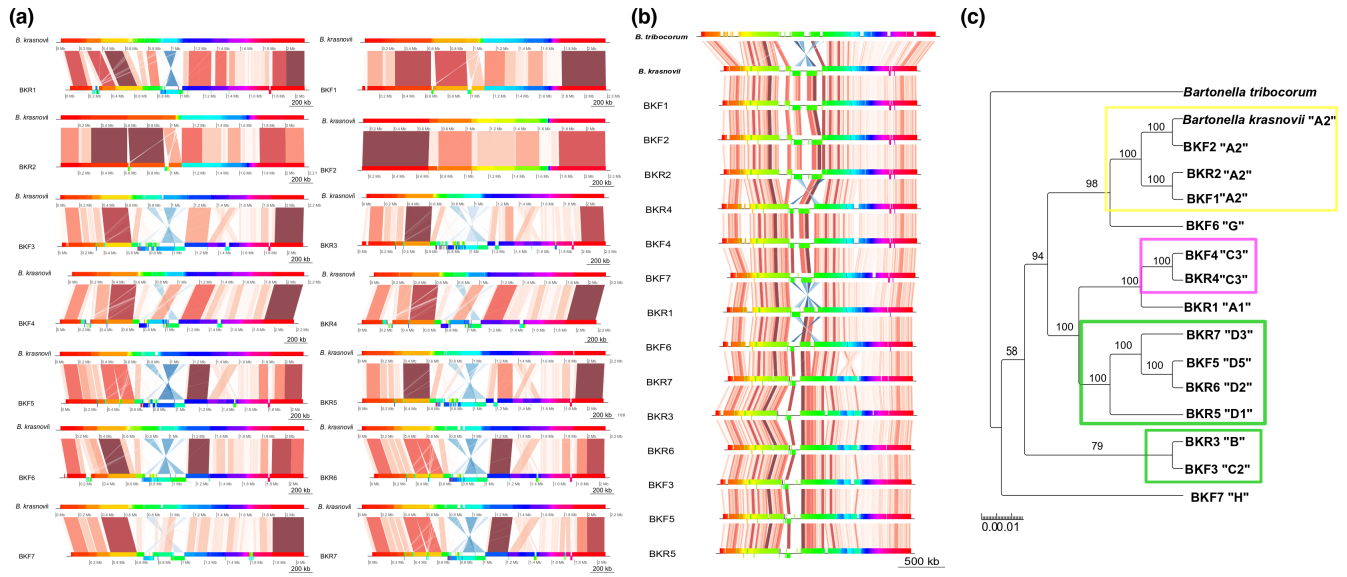

FIGURE 4 Pairwise alignments of each variant and OE 1–1 genome, multiple alignments of *B. krasnovii* genotypes and phylogenetic nucleotide-based tree. (a) Mauve alignment plots of each of the 14 *B. krasnovii* genotypes against the OE 1–1 *B. krasnovii*. (b) Mauve multiple alignments of *B. tribocorum*, *B. krasnovii* OE 1–1 genome and 14 *B. krasnovii* genotypes. The bars corresponding to the genomes were coloured according to the number of conserved segments or locally collinear blocks (LCBs) identified by the genome alignments, following a rainbow palette. The comparison features were coloured by shades of red and blue according to the length of the LCBs. The direct comparisons are coloured in red and the reversed ones in blue. (c) Phylogenetic ML tree of the nucleotides sequencing of the genomes assemblies, comprising the *B. krasnovii* OE 1–1 and 14 *B. krasnovii* genotypes, using bootstrap for supporting of clades and *B. tribocorum* as an outgroup. The three supported SV clades are coloured on the nucleotide-based tree

niches in nature. The results presented here confirm the outstanding genomic structural diversity of *Bartonella* genomes with a conserved nucleotide variation and support the association of the former with prophage elements (Gutiérrez, Markus, et al., 2018).

SVs were previously observed among the zoonotic cat associated *Bartonella*, *B. henselae* (Lindroos et al., 2006), and the rodent-associated *Bartonella*, *B. grahamii* (Berglund, Ehrenborg, et al., 2010a; Berglund, Ellegaard, et al., 2010b). The present study broadens and

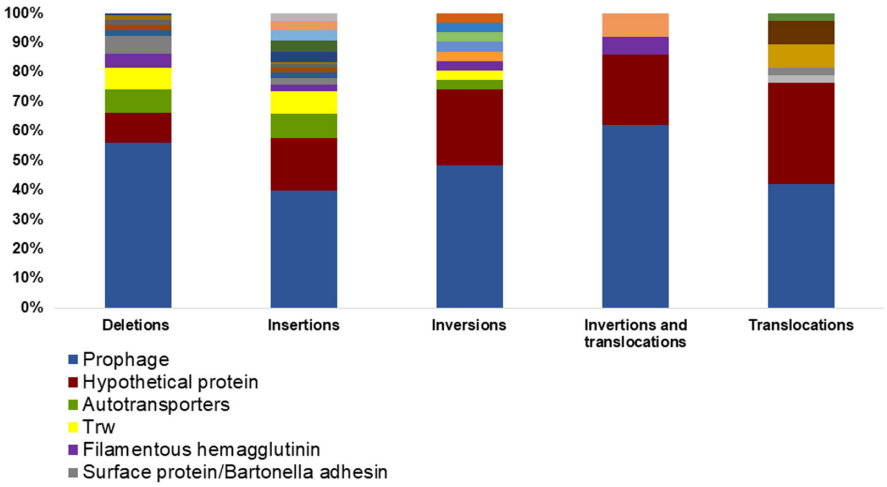

FIGURE 5 Genes associated with SV events. Percentage of genes associated with SV events accessed through the analysis of the SVs flanking regions

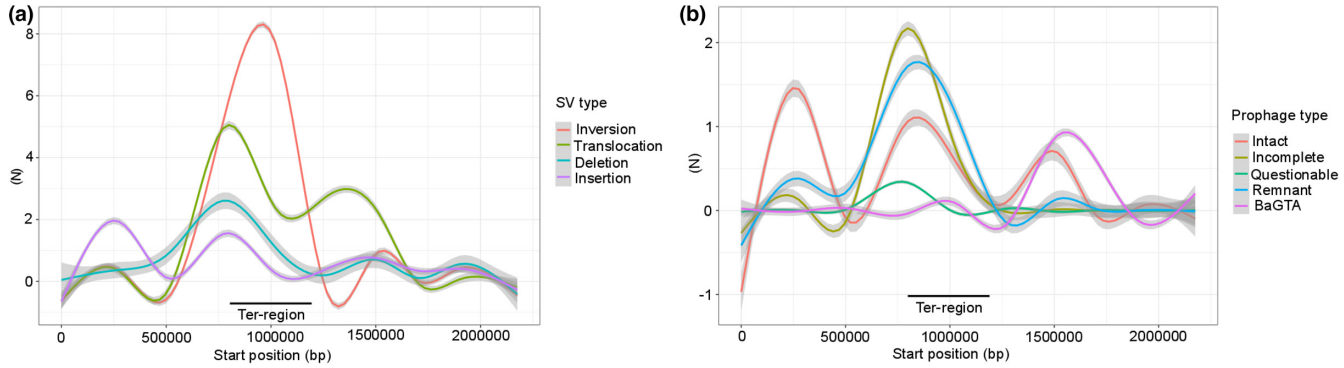

FIGURE 6 Distribution of all types of SV events and prophages across the 14 *B. krasnovii* genotypes. (a) Distribution of SV types across the genomes. (b) Distribution of prophage copies across the genomes

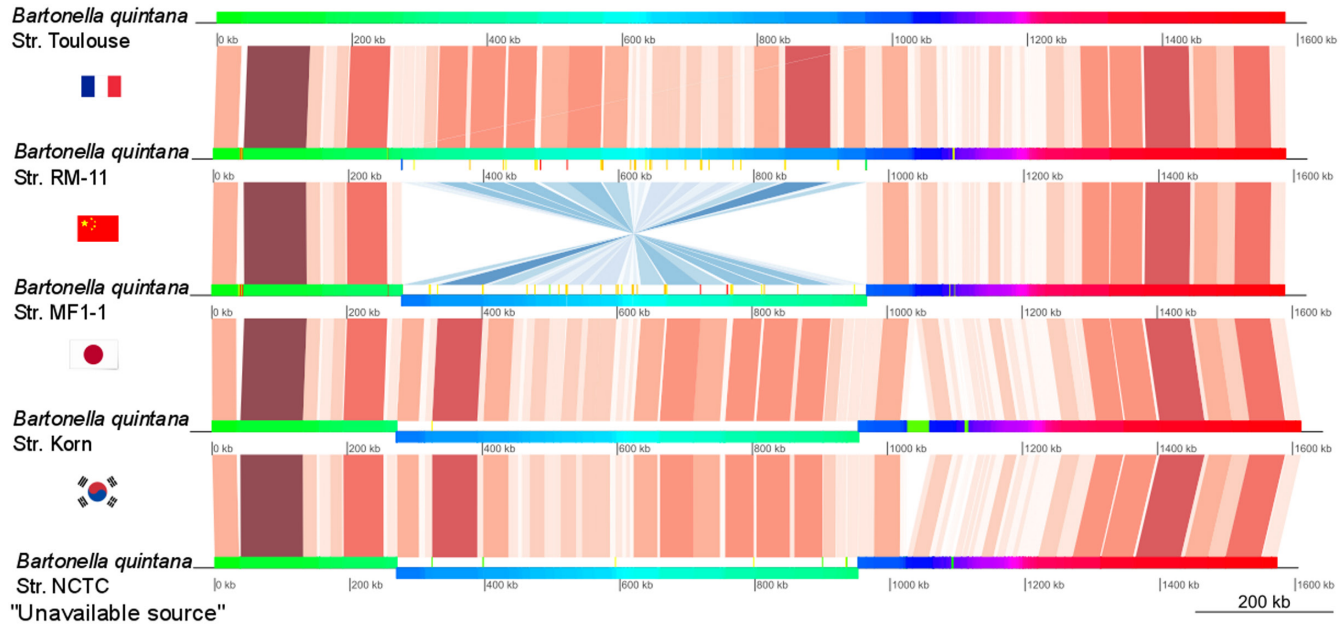

FIGURE 7 Whole genome comparison of five different strains of *Bartonella quintana*. Pairwise alignments of five strains of *B. quintana*. The bars corresponding to the genomes were coloured according to the number of conserved segments or locally collinear blocks (LCBs) identified by the genome alignments, following a rainbow palette. The comparison features were coloured by shades of red and blue according the length of the LCBs. The direct comparisons are coloured in red and the reversed ones in blue. The alignment is based on SVs identified by progressiveMauve

complements the knowledge on this subject by a detailed characterization of the genomic content of genotypes isolated at the same time and location, from two small communities of gerbils and their associated fleas. Particularly, this study reveals the implication of prophages and SVs at the *ter*-region of the *B. krasnovii* chromosome, as a plastic zone for genomic rearrangements and prophage integration, reinforcing the evident association of prophages and SV events in rodent-associated *Bartonella*. Furthermore, the evidence of six unbalanced genomes, and a strand-shift in some essential genes near the *ter*-region, suggests that some SV, probably driven by prophages, may lead to instability in *Bartonella* genomes. Given the differences in replication fidelity between the leading and the lagging strands, this observation may suggest a different rate of evolution in essential genes between different isolates (Fijalkowska et al., 1998).

The structural genomic diversity has been identified and well characterized in other bacterial species including *Escherichia coli*, *Salmonella typhi*, *Staphylococcus aureus*, *Campylobacter jejuni*, *Mycobacterium* spp. and other species (Cui et al., 2012; Garcia-Betancur et al., 2012; Iguchi et al., 2006; Liu et al., 2006; Scott et al., 2007). Particularly, high rates of genome rearrangements were identified in *Yersinia* spp. (Darling et al., 2008). *Yersinia pestis*, like *Bartonella*, is a flea-borne bacterium, that is constantly challenged to adapt to totally different niches. The genomic plasticity of this bacterium allowed the authors to perform a phylogenetic characterization, based on the SV data, of six genomes of *Yersinia pestis*, two *Yersinia pseudotuberculosis* and one genome of *Yersinia enterocolitica* (Darling et al., 2008). Contrary, the SVs-based phylogeny of *B. krasnovii* in our study was not fully conclusive, while the nucleotide-based tree was well supported, probably due to the fact that the phylogeny was done based on one single species, isolated from the same area at the same period, thus the events still are in progress to become fixed in the population. Studies involving isolates from small mammal communities, as the one presented here, represent an important addition of data to portray the ongoing genomic evolution in populations under constant natural selection. Under such conditions, confounding factors that could under- or overestimate the findings (e.g., different geographical locations or period of collection) are constrained, enabling a detailed genomic characterization, followed by a mechanistic investigation.

Rodent-associated *Bartonella* constitute a polymorphic group of species in nature (Gutiérrez et al., 2014, 2015; Gutiérrez, Cohen, et al., 2018). This has been demonstrated by the characterization of hundreds of strains, sequenced from wild rodent communities (Berglund, Ehrenborg, et al., 2010a; Berglund, Ellegaard, et al., 2010b; Gutiérrez et al., 2014, 2015; Gutiérrez, Cohen, et al., 2018; Morick et al., 2010; Paziewska et al., 2011). Interestingly, the present study shows that, although the genomes of *B. krasnovii* contain regions with greater alternative allele frequency, point mutations were responsible for only 0.3% of the genomic variation, with only two genes showing significant signal of selection. Contrarily, structural diversity was observed in all genotypes, with an average of 51 SVs per genome recorded. The SV effect was initially noticed in the patterns of the *SmaI*-restriction PFGE analysis, which correlated with several SV events identified through the comparison of each genome with the OE1-1 strain genome. Moreover, the predicted phylogeny based on SVs showed agreement with the nucleotide tree in three supported clades, and enabled the support of one clade split in the nucleotide-based tree. It is evident that SVs occur frequently in these genomes, affecting the overall *Bartonella* phylogeny. They may ultimately promote competitive interactions, resulting (in a later stage) in the selection of adaptive gene/nucleotide genotypes (Gutiérrez, Markus, et al., 2018) and in the speciation of *Bartonella*.

The genome plasticity observed in this study among *B. krasnovii* genotypes exceeded that of other *Bartonella* spp. previously investigated. A study that focused on *B. grahamii* showed high structural variation compared to nucleotide diversity (SNPs); however, the strains were isolated from different continents and from three different host families (Berglund, Ellegaard, et al., 2010b). Contrarily, another study based also on *B. grahamii* strains, but isolated from two different rodent families and collected from three near geographic locations (less than 30 kms apart) showed both low rates of SNPs and structural diversity (Berglund, Ellegaard, et al., 2010b). Structural variations were also observed among *Bartonella henselae*; however, in lower rates (106 events in 38 strains), including strains isolated from four different continents and collected from humans and cats (Lindroos et al., 2006). In the present study we identified a total of 702 SV events in 14 structurally-validated genomes in comparison to the OE1-1 genome. Remarkably, these strains were isolated from sympatric rodents that inhabit a restricted location. However, the higher diversity observed in our study might be a result of the smaller spatial scale. It is possible that in the comparison between *Bartonella* isolates across distant geographical locations and host groups, genetic drift played an additional role, reducing the observed genetic diversity.

In our study, most of the identified SV events were associated with prophages, confirming the implication of these elements in the onset of SVs, as previously shown in an *in vitro* evolutionary experiment with *B. krasnovii* (Gutiérrez, Markus, et al., 2018) and in other bacterial genera (Iguchi et al., 2006; Scott et al., 2007; Shen et al., 2020). Structural variations are a result of homologous and nonhomologous recombination events (Ottaviani et al., 2014; Treangen et al., 2009), that rather than point mutations, were previously suggested as the main source of genetic diversity among *Bartonella* spp. (Berglund, Ehrenborg, et al., 2010a; Berglund, Ellengaard, et al., 2010b; Gutiérrez, Cohen, et al., 2018; Guy et al., 2012; Paziewska et al., 2011). In *E. coli*, prophages play a key role in the host recombination events, shaping and affecting their evolvability (Bobay, Touchon, et al., 2013b). Prophages can employ the bacterial machinery to mediate their own recombination, replication and packing or impose their own recombination machinery by inhibiting the RecBCD enzyme subunit of the most important pathway of homologous recombination in the host (Bobay, Touchon, et al., 2013b). Interestingly, the prophages that confer the ability to promote recombination present mosaic genomes with a diverse gene content (Bobay et al., 2014). This pattern of

mosaicism is the major outcome of recombination events due to HGT (Bobay, Touchon, et al., 2013b; Dion et al., 2020). The prophages identified in our study evidenced two clear genetic clusters, suggesting accumulation of mutations and participation of HGT in the sequence composition. In fact, one of the clusters harbored most of the incomplete and remnant prophages, suggesting prophage degradation across the genomes. In addition to driving SVs, the role of *Bartonella* prophages in promoting recombination deserves further investigation.

We found a high concentration of prophages and SVs in the *ter*-region of the chromosome, suggesting that this region could be a plastic zone for integration and SV formation. Rearrangements at the *ter*-region were already reported in other *Bartonella* species (Alsmark et al., 2004; Lindroos et al., 2006; Segers et al., 2017) and in a broad range of different bacterial species (Bowden et al., 2016; Esnault et al., 2007; Hoff et al., 2018; Kresse et al., 2003; Repar & Warnecke, 2017). Replication in circular bacterial chromosomes occurs bidirectionally around the chromosome, starting from the *ori*-region and finishing in the *ter*-region, located about 180 degrees apart (Kothapalli et al., 2005). When the replication forks encounter transcriptional elements, conflicts may happen activating the DNA repair mechanisms and consequently recombination events, driving instability and SVs near this region (Merrikh et al., 2012). In our study, the majority of prophages and prophage remnants were located in the *ter*-region, and were associated with SVs. Prophages were suggested to provide conditions for bacteria to improve their fitness (i.e., protection against phages, provide virulence factors and/or antibiotic resistant genes) (Bobay et al., 2014). To keep this positive relationship, prophages may not integrate randomly, thus facilitating the coevolution of phages and bacteria to endure selection for integration sites (Bobay, Rocha, et al., 2013a). Temperate prophages show tropism to the *ter*-region of the chromosome where less abundance of coding genes is present, reducing the impact on the genome organization and bacterial fitness (Bobay, Rocha, et al., 2013a). However, in our study six strains presented unbalanced genomes, and impacted the distribution of some essential genes that had changed the strand orientation. These changes may affect the translation pattern in the bacteria, being evidence that prophages have the potential to cause the disruption of the physical balance in *Bartonella* genomes. One of the major outcomes of genome instability is the phase variation that modulates the extension of expression of operons and genes that can lead to antigenic variation and consecutively resulting in new phenotypes (Darmon & Leach, 2014). In *Salmonella typhi*, the disruption caused by insertion sequences was demonstrated. The instability observed drove genome diversification, with the outcome of new phenotypes, being these events implicated in the evolution of the species (Liu et al., 2006).

Bartonella quintana, the agent of trench fever, presents a low degree of genomic polymorphism, that is probably due to a recent genetic drift caused by host switch (Berglund, Ehrenborg, et al., 2010a; Foucault et al., 2006). It has been proposed that this species has lost its single prophage during the evolution (Alsmark et al., 2004). Based on the latter, *B. quintana* served as a prophage-free control for our study, to further investigate the association of SVs and prophages. In fact, the genome comparison of five different strains of *B. quintana* showed only one large inversion, demonstrating low degree of genomic plasticity compared with *B. henselae* (Lindroos et al., 2006) and rodent-associated *Bartonella* (Berglund, Ehrenborg, et al., 2010a; Berglund, Ellengaard, et al., 2010b; Gutiérrez, Markus, et al., 2018; Gutiérrez, Cohen, et al., 2018). However, prophage annotation revealed two small prophage-related regions in the *B. quintana* genomes. These genomic loci may represent prophage vestiges or remnants, and not truly temperate prophages. The inspection of *B. quintana* genomes complemented our *B. krasnovii* findings and conclusion that prophages are SV drivers and consecutively implicated in the diversity of *Bartonella* genomes.

In conclusion, the present study supports the positive association of SVs and prophages in the generation of diversity in rodent-associated *Bartonella* in nature, reflecting the dynamics of *Bartonella* lifestyle and the direct implication of these events in the adaptation and evolution of this genus.

AUTHOR CONTRIBUTIONS

The research was designed by Keyla Carstens Marques de Sousa, Ricardo Gutiérrez, Yaarit Nachum-Biala, Shay Covo and Shimon Harrus. The collection and sampling of animals were performed by Hadas Hawlena. The research was performed by Keyla Carstens Marques de Sousa, Ricardo Gutiérrez and Shimon Harrus. The bioinformatic support was done by Dayana Yahalomi, Tali Shalit, Barak Markus and Haroldo Henrique de Rezende Neves. The phylogenetic analysis was performed by Einat Hazkani-Covo. The analytic tools were assisted by Evgeniya Marcos-Hadad and Shay Covo. Data analysis and writing of the article were carried out by Keyla Carstens Marques de Sousa, Ricardo Gutiérrez and Shimon Harrus.

ACKNOWLEDGEMENTS

This research was supported by the Israel Science Foundation (grant no. 688/17 to S.H.) and the National Science Foundation - Binational Science Foundation (grant no. 2017708 to H.H. and S.H.).

CONFLICT OF INTEREST

The authors confirm no conflicting interests.

DATA AVAILABILITY STATEMENT

All genomes included in this study have been deposited in GenBank database under the BioProjet: PRJNA801099, genome accession numbers: CP093033 - CP093046.

ORCID

Ricardo Gutiérrez https://orcid.org/0000-0001-6988-0251
Hadas Hawlena https://orcid.org/0000-0002-0634-2920
Shimon Harrus https://orcid.org/0000-0003-0542-207X

REFERENCES

Alsmark, C. M., Frank, A. C., Karlberg, E. O., Legault, B. A., Ardell, D. H., Canbäck, B., Eriksson, A. S., Näslund, A. K., Handley, S. A., Huvet, M., La Scola, B., Holmberg, M., & Andersson, S. G. E. (2004). The louse-borne human pathogen *Bartonella quintana* is a genomic derivative of the zoonotic agent *Bartonella henselae*. *Proceedings of the National Academy of Sciences of The United States of America*, *101*(26), 9716–9721. https://doi.org/10.1073/pnas.0305659101

Antipov, D., Raiko, M., Lapidus, A., & Pevzner, P. A. (2019). Plasmid detection and assembly in genomic and metagenomic data sets. *Genome Research*, *29*(6), 961–968. https://doi.org/10.1101/gr.241299.118

Arakawa, K., Suzuki, H., & Tomita, M. (2008). Computational genome analysis using the G-language system. *Genes, Genomes and Genomics*, *2*, 1–13.

Argueso, J. L., Westmoreland, J., Mieczkowski, P. A., Gawel, M., Petes, T. D., & Resnick, M. A. (2008). Double-strand breaks associated with repetitive DNA can reshape the genome. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(33), 11845–11850. https://doi.org/10.1073/pnas.0804529105

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, *44*(W1), W16–W21. https://doi.org/10.1093/nar/gkw387

Berglund, E. C., Ehrenborg, C., Vinnere Pettersson, O., Granberg, F., Näslund, K., Holmberg, M., & Andersson, S. G. E. (2010a). Genome dynamics of *Bartonella grahamii* in micro-populations of woodland rodents. *BMC Genomics*, *11*(1), 152. https://doi.org/10.1186/1471-2164-11-152

Berglund, E. C., Ellegaard, K., Granberg, F., Xie, Z., Maruyama, S., Kosoy, M. Y., Birtles, R. J., & Andersson, S. G. E. (2010b). Rapid diversification by recombination in *Bartonella grahamii* from wild rodents in Asia contrasts with low levels of genomic divergence in Northern Europe and America. *Molecular Ecology*, *19*(11), 2241–2255. https://doi.org/10.1111/j.1365-294X.2010.04646.x

Bobay, L. M., Rocha, E. P. C., & Touchon, M. (2013a). The adaptation of temperate bacteriophages to their host genomes. *Molecular Biology and Evolution*, *30*(4), 737–751. https://doi.org/10.1093/molbev/mss279

Bobay, L. M., Touchon, M., & Rocha, E. P. C. (2013b). Manipulating or superseding host recombination functions: A dilemma that shapes phage evolvability. *PLoS Genetics*, *9*(9), e1003825. https://doi.org/10.1371/journal.pgen.1003825

Bobay, L. M., Touchon, M., & Rocha, E. P. C. (2014). Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(33), 12127–12132. https://doi.org/10.1073/pnas.1405336111

Bochkareva, O. O., Dranenko, N. O., Ocheredko, E. S., Kanevsky, G. M., Lozinsky, Y. N., Khalaycheva, V. A., Artamonova, I. I., & Gelfand, M. S. (2018). Genome rearrangements and phylogeny reconstruction in *Yersinia pestis*. *PeerJ*, *2018*(3), e4545. https://doi.org/10.7717/peerj.4545

Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., & Zaretskaya, I. (2013). BLAST: A more efficient report with usability improvements. *Nucleic Acids Research*, *41*(Web Server issue), 29–33. https://doi.org/10.1093/nar/gkt282

Bowden, K. E., Weigand, M. R., Peng, Y., Cassiday, P. K., Sammons, S., Knipe, K., Rowe, L. A., Loparev, V., Sheth, M., Weening, K., Tondella, M. L., & Williams, M. M. (2016). Genome structural diversity among 31 *Bordetella pertussis* isolates from two recent U.S. Whooping Cough Statewide Epidemics. *MSphere*, *1*(3), e00036-16. https://doi.org/10.1128/msphere.00036-16

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomason, J. A., Stevens, R., Vonstein, V., Wattam, A. R., & Xia, F. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*, *5*, 8365. https://doi.org/10.1038/srep08365

Cerdeño-Tárraga, A. M., Patrick, S., Crossman, L. C., Blakely, G., Abratt, V., Lennard, N., Poxton, I., Duerden, B., Harris, B., Quail, M. A., Barron, A., Clark, L., Corton, C., Doggett, J., Holden, M. T. G., Larke, N., Line, A., Lord, A., Norbertczak, H., … Parkhill, J. (2005). Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science*, *307*(5714), 1463–1465. https://doi.org/10.1126/science.1107008

Charlesworth, J., & Eyre-Walker, A. (2008). The McDonald-Kreitman test and slightly deleterious mutations. *Molecular Biology and Evolution*, *25*(6), 1007–1015. https://doi.org/10.1093/molbev/msn005

Cui, L., Neoh, H. M., Iwamoto, A., & Hiramatsu, K. (2012). Coordinated phenotype switching with large-scale chromosome flip-flop inversion observed in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(25), E1647–E1656. https://doi.org/10.1073/pnas.1204307109

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, *14*(7), 1394–1403. https://doi.org/10.1101/gr.2289704

Darling, A. E., Mau, B., & Perna, N. T. (2010). Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*, *5*(6), e11147. https://doi.org/10.1371/journal.pone.0011147

Darling, A. E., Miklós, I., & Ragan, M. A. (2008). Dynamics of genome rearrangement in bacterial populations. *PLoS Genetics*, *4*(7), e1000128. https://doi.org/10.1371/journal.pgen.1000128

Darmon, E., & Leach, D. R. F. (2014). Bacterial genome instability. *Microbiology and Molecular Biology Reviews*, *78*(1), 1–39. https://doi.org/10.1128/mmbr.00035-13

De Maio, N., Shaw, L. P., Hubbard, A., George, S., Sanderson, N. D., Swann, J., Wick, R., Oun, M. A., Stubberfield, E., Hoosdally, S. J., Crook, D. W., Peto, T. E. A., Sheppard, A. E., Bailey, M. J., Read, D. S., Anjum, M. F., Sarah Walker, A., & Stoesser, N. (2019). Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microbial Genomics*, *5*(9), e000294. https://doi.org/10.1099/mgen.0.000294

Dion, M. B., Oechslin, F., & Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nature Reviews Microbiology 18*(3), 125–138. Nature Research. https://doi.org/10.1038/s41579-019-0311-5

Dobrindt, U., & Hacker, J. (2001). Whole genome plasticity in pathogenic bacteria. *Current Opinion in Microbiology*, *4*(5), 550–557. https://doi.org/10.1016/S1369-5274(00)00250-2

Esnault, E., Valens, M., Espéli, O., & Boccard, F. (2007). Chromosome structuring limits genome plasticity in *Escherichia coli*. *PLoS Genetics*, *3*(12), e226. https://doi.org/10.1371/journal.pgen.0030226

Fijalkowska, I. J., Jonczyk, P., Tkaczyk, M. M., Bialoskorska, M., & Schaaper, R. M. (1998). Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(17), 10020–10025. https://doi.org/10.1073/pnas.95.17.10020

Foucault, C., Brouqui, P., & Raoult, D. (2006). *Bartonella quintana* characteristics and clinical management. *Emerging Infectious Diseases*, *12*(2), 217–223. https://doi.org/10.3201/eid1202.050874

Frank, A. C., & Lobry, J. R. (2000). Oriloc: Prediction of replication boundaries in unannotated bacterial chromosomes. *Bioinformatics*, *16*(6), 560–561. https://doi.org/10.1093/bioinformatics/16.6.560

Garcia-Betancur, J. C., Menendez, M. C., Del Portillo, P., & Garcia, M. J. (2012). Alignment of multiple complete genomes suggests that gene rearrangements may contribute towards the speciation of *Mycobacteria*. *Infection, Genetics and Evolution*, *12*(4), 819–826. https://doi.org/10.1016/j.meegid.2011.09.024

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, *57*(1), 81–91. https://doi.org/10.1099/ijs.0.64483-0

Gutiérrez, R., Cohen, C., Flatau, R., Marcos-Hadad, E., Garrido, M., Halle, S., Nachum-Biala, Y., Covo, S., Hawlena, H., & Harrus, S. (2018). Untangling the knots: Co-infection and diversity of *Bartonella* from wild gerbils and their associated fleas. *Molecular Ecology*, *27*(23), 4787–4807. https://doi.org/10.1111/mec.14906

Gutiérrez, R., Krasnov, B., Morick, D., Gottlieb, Y., Khokhlova, I. S., & Harrus, S. (2015). Bartonella infection in rodents and their flea ectoparasites: An overview. *Vector-Borne and Zoonotic Diseases*, *15*(1), 27–39. https://doi.org/10.1089/vbz.2014.1606

Gutierrez, R., Markus, B., De Sousa, K. C. M., Marcos-Hadad, E., Mugasimangalam, R. C., Nachum-Biala, Y., Hawlena, H., Covo, S., & Harrus, S. (2018). Prophage-driven genomic structural changes promote *Bartonella* vertical evolution. *Genome Biology and Evolution*, *10*(11), 3089–3103. https://doi.org/10.1093/gbe/evy236

Gutiérrez, R., Morick, D., Cohen, C., Hawlena, H., & Harrus, S. (2014). The effect of ecological and temporal factors on the composition of *Bartonella* infection in rodents and their fleas. *ISME Journal*, *8*(8), 1598–1608. https://doi.org/10.1038/ismej.2014.22

Gutiérrez, R., Shalit, T., Markus, B., Yuan, C., Nachum-Biala, Y., Elad, D., & Harrus, S. (2020). *Bartonella kosoyi* sp. nov. and *Bartonella krasnovii* sp. nov., two novel species closely related to the zoonotic Bartonella elizabethae, isolated from black rats and wild desert rodent-fleas. *International Journal of Systematic and Evolutionary Microbiology*, *70*(3), 1656–1665. https://doi.org/10.1099/ijsem.0.003952

Guy, L., Kultima, J. R., Andersson, S. G. E., & Quackenbush, J. (2011). GenoPlotR: comparative gene and genome visualization in R. *Bioinformatics*, *27*(13), 2334–2335. https://doi.org/10.1093/bioinformatics/btq413

Guy, L., Nystedt, B., Sun, Y., Näslund, K., Berglund, E. C., & Andersson, S. G. E. (2012). A genome-wide study of recombination rate variation in *Bartonella henselae*. *BMC Evolutionary Biology*, *12*(1), 65. https://doi.org/10.1186/1471-2148-12-65

Hoff, G., Bertrand, C., Piotrowski, E., Thibessard, A., & Leblond, P. (2018). Genome plasticity is governed by double strand break DNA repair in Streptomyces. *Scientific Reports*, *8*(1), 5272. https://doi.org/10.1038/s41598-018-23622-w

Howard-Varona, C., Hargreaves, K. R., Solonenko, N. E., Markillie, L. M., White, R. A., Brewer, H. M., Ansong, C., Orr, G., Adkins, J. N., & Sullivan, M. B. (2018). Multiple mechanisms drive phage infection efficiency in nearly identical hosts. *ISME Journal*, *12*(6), 1605–1618. https://doi.org/10.1038/s41396-018-0099-8

Iguchi, A., Iyoda, S., Terajima, J., Watanabe, H., & Osawa, R. (2006). Spontaneous recombination between homologous prophage regions causes large-scale inversions within the *Escherichia coli* O157:H7 chromosome. *Gene*, *372*(1–2), 199–207. https://doi.org/10.1016/j.gene.2006.01.005

Inoue, K., Maruyama, S., Kabeya, H., Hagiya, K., Izumi, Y., Une, Y., & Yoshikawa, Y. (2009). Exotic small mammals as potential reservoirs of zoonotic *Bartonella* spp. *Emerging Infectious Diseases*, *15*(4), 526–532. https://doi.org/10.3201/eid1504.081223

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 1–8. https://doi.org/10.1038/s41467-018-07641-9

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Kosoy, M., Hayman, D. T. S., & Chan, K. S. (2012). *Bartonella* bacteria in nature: Where does population variability end and a species start? *Infection, Genetics and Evolution*, *12*(5), 894–904. https://doi.org/10.1016/j.meegid.2012.03.005

Kothapalli, S., Nair, S., Alokam, S., Pang, T., Khakhria, R., Woodward, D., Johnson, W., Stocker, B. A. D., Sanderson, K. E., & Liu, S. L. (2005). Diversity of genome structure in *Salmonella enterica* serovar Typhi populations. *Journal of Bacteriology*, *187*(8), 2638–2650. https://doi.org/10.1128/JB.187.8.2638-2650.2005

Kresse, A. U., Dinesh, S. D., Larbig, K., & Römling, U. (2003). Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs. *Molecular Microbiology*, *47*(1), 145–158. https://doi.org/10.1046/j.1365-2958.2003.03261.x

Li, H., Tong, Y., Huang, Y., Bai, J., Yang, H., Liu, W., & Cao, W. (2012). Complete genome sequence of *Bartonella quintana*, a bacterium isolated from rhesus macaques. *Journal of Bacteriology*, *194*(22), 6347. https://doi.org/10.1128/JB.01602-12

Lindroos, H., Vinnere, O., Mira, A., Repsilber, D., Näslund, K., & Andersson, S. G. E. (2006). Genome rearrangements, deletions, and amplifications in the natural population of *Bartonella henselae*. *Journal of Bacteriology*, *188*(21), 7426–7439. https://doi.org/10.1128/JB.00472-06

Liu, G. R., Liu, W. Q., Johnston, R. N., Sanderson, K. E., Li, S. X., & Liu, S. L. (2006). Genome plasticity and ori-ter rebalancing in *Salmonella typhi*. *Molecular Biology and Evolution*, *23*(2), 365–371. https://doi.org/10.1093/molbev/msj042

Lobry, J. R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular Biology and Evolution*, *13*(5), 660–665. https://doi.org/10.1093/oxfordjournals.molbev.a025626

Luo, H., & Gao, F. (2019). DoriC 10.0: An updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acids Research*, *47*(D1), D74–D77. https://doi.org/10.1093/nar/gky1014

McDonald, J., & Kreitman, M. (1991). Adaptive protein evolution at Adh in *Drosophila*. *Nature*, *351*(June), 652–654. http://ib.berkeley.edu/labs/slatkin/popgenjclub/pdf/mcdonald-kreitman1991.pdf

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Merrikh, H., Zhang, Y., Grossman, A. D., & Wang, J. D. (2012). Replication-transcription conflicts in bacteria. *Nature Reviews Microbiology*, *10*(7), 449–458. https://doi.org/10.1038/nrmicro2800

Morick, D., Krasnov, B. R., Khokhlova, I. S., Shenbrot, G. I., Kosoy, M. Y., & Harrus, S. (2010). *Bartonella* genotypes in fleas (insecta: Siphonaptera) collected from rodents in the negev desert, Israel. *Applied and Environmental Microbiology*, *76*(20), 6864–6869. https://doi.org/10.1128/AEM.00879-10

Neuvonen, M. M., Tamarit, D., Näslund, K., Liebig, J., Feldhaar, H., Moran, N. A., Guy, L., & Andersson, S. G. E. (2016). The genome of Rhizobiales bacteria in predatory ants reveals urease gene functions but no genes for nitrogen fixation. *Scientific Reports*, *6*, 39197. https://doi.org/10.1038/srep39197

Nzabarushimana, E., & Tang, H. (2018). Insertion sequence elements-mediated structural variations in bacterial genomes. *Mobile DNA*, *9*(1), 29. https://doi.org/10.1186/s13100-018-0134-3

Ottaviani, D., LeCain, M., & Sheer, D. (2014). The role of microhomology in genomic structural variation. *Trends in Genetics*, *30*(3), 85–94. https://doi.org/10.1016/j.tig.2014.01.001

Paziewska, A., Harris, P. D., Zwolińska, L., Bajer, A., & Siński, E. (2011). Recombination within and between species of the alpha proteobacterium *Bartonella* Infecting Rodents. *Microbial Ecology*, *61*(1), 134–145. https://doi.org/10.1007/s00248-010-9735-1

Periwal, V., & Scaria, V. (2015). Insights into structural variations and genome rearrangements in prokaryotic genomes. *Bioinformatics*, *31*(1), 1–9. https://doi.org/10.1093/bioinformatics/btu600

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, *31*(7), 1929–1936. https://doi.org/10.1093/molbev/msu136

Ramisetty, B. C. M., & Sudhakari, P. A. (2019). Bacterial "grounded" prophages: Hotspots for genetic renovation and innovation. *Frontiers in Genetics*, *10*(FEB), 65. https://doi.org/10.3389/fgene.2019.00065

Repar, J., & Warnecke, T. (2017). Non-random inversion landscapes in prokaryotic genomes are shaped by heterogeneous selection pressures. *Molecular Biology and Evolution*, *34*(8), 1902–1911. https://doi.org/10.1093/molbev/msx127

Rodriguez-R, L. M., Gunturu, S., Harvey, W. T., Rosselló-Mora, R., Tiedje, J. M., Cole, J. R., & Konstantinidis, K. T. (2018). The Microbial Genomes Atlas (MiGA) webserver: Taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Research*, *46*(W1), W282–W288. https://doi.org/10.1093/nar/gky467

Sato, S., Kabeya, H., Yoshino, A., Sekine, W., Suzuki, K., Tamate, H. B., Yamazaki, S., Chomel, B. B., & Maruyama, S. (2015). Japanese Macaques (*Macaca fuscata*) as natural reservoir of *Bartonella quintana*. *Emerging Infectious Diseases*, *21*(12), 2168–2170. https://doi.org/10.3201/eid2112.150632

Scott, A. E., Timms, A. R., Connerton, P. L., Carrillo, C. L., Radzum, K. A., & Connerton, I. F. (2007). Genome dynamics of *Campylobacter jejuni* in response to bacteriophage predation. *PLoS Pathogens*, *3*(8), 1142–1151. https://doi.org/10.1371/journal.ppat.0030119

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Segers, F. H., Kešnerová, L., Kosoy, M., & Engel, P. (2017). Genomic changes associated with the evolutionary transition of an insect gut symbiont into a blood-borne pathogen. *ISME Journal*, *11*(5), 1232–1244. https://doi.org/10.1038/ismej.2016.201

Shen, J., Zhou, J., Xu, Y., & Xiu, Z. (2020). Prophages contribute to genome plasticity of *Klebsiella pneumoniae* and may involve the chromosomal integration of ARGs in CG258. *Genomics*, *112*(1), 998–1010. https://doi.org/10.1016/j.ygeno.2019.06.016

Sheppard, S. K., Guttman, D. S., & Fitzgerald, J. R. (2018). Population genomics of bacterial host adaptation. *Nature Reviews Genetics*, *19*(9), 549–565. https://doi.org/10.1038/s41576-018-0032-z

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., Gladstone, R. A., Lo, S., Beaudoin, C., Floto, R. A., Frost, S. D. W., Corander, J., Bentley, S. D., & Parkhill, J. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *BioRxiv*, 1–21. https://doi.org/10.1101/2020.01.28.922989

Touchon, M., & Rocha, E. P. C. (2016). Coevolution of the organization and structure of prokaryotic genomes. *Cold Spring Harbor Perspectives in Biology*, *8*(1), a018168. https://doi.org/10.1101/cshperspect.a018168

Treangen, T. J., Abraham, A. L., Touchon, M., & Rocha, E. P. C. (2009). Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiology Reviews*, *33*(3), 539–571. https://doi.org/10.1111/j.1574-6976.2009.00169.x

Wang, S., Meade, A., Lam, H.-M., & Luo, H. (2020). Evolutionary timeline and genomic plasticity underlying the lifestyle diversity in rhizobiales. *MSystems*, *5*(4), e00438-20. https://doi.org/10.1128/msystems.00438-20

Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E. M., Disz, T., Gabbard, J. L., Gerdes, S., Henry, C. S., Kenyon, R. W., Machi, D., Mao, C., Nordberg, E. K., Olsen, G. J., Murphy-Olson, D. E., Olson, R., … Stevens, R. L. (2017). Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Research*, *45*(D1), D535–D542. https://doi.org/10.1093/nar/gkw1017

Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology*, *13*(6), 1–22. https://doi.org/10.1371/journal.pcbi.1005595

Yan, W., Wei, S., Wang, Q., Xiao, X., Zeng, Q., Jiao, N., & Zhang, R. (2018). Genome rearrangement shapes *Prochlorococcus* ecological adaptation. *Applied and Environmental Microbiology*, *84*(17), e01178-18. https://doi.org/10.1128/AEM.01178-18

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: de Sousa, K. C. M., Gutiérrez, R., Yahalomi, D., Shalit, T., Markus, B., Nachum-Biala, Y., Hawlena, H., Marcos-Hadad, E., Hazkani-Covo, E., de Rezende Neves, H. H., Covo, S., & Harrus, S. (2022). Genomic structural plasticity of rodent-associated *Bartonella* in nature. *Molecular Ecology*, *31*, 3784–3797. https://doi.org/10.1111/mec.16547