Parasite contingency loci and the evolution of host specificity: Simple sequence repeats mediate *Bartonella* adaptation to a wild rodent host

Ruth Rodríguez-Pastor[1†], Nadav Knossow[1,2†], Naama Shahar[2], Adam Z. Hasik[1¶], Daniel E. Deatherage[3], Ricardo Gutiérrez[4,5], Shimon Harrus[6], Luis Zaman[7], Richard E. Lenski[8], Jeffrey E. Barrick[3]*, Hadas Hawlena[2]*

[1] Jacob Blaustein Center for Scientific Cooperation, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel

[2] Mitrani Department of Desert Ecology, Swiss Institute for Dryland Environmental and Energy Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Israel

[3] Department of Molecular Biosciences, Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, Texas, USA

[4] National Reference Center for Bacteriology, Costa Rican Institute for Research and Teaching in Nutrition and Health (Inciensa), Cartago, Costa Rica.

[5] Ross University School of Veterinary Medicine, Basseterre, St. Kitts and Nevis, West Indies

[6] Koret School of Veterinary Medicine, Faculty of Agricultural, Nutritional and Environmental Sciences, The Hebrew University of Jerusalem, Rehovot, Israel

[7] Department of Ecology and Evolutionary Biology, Center for the Study of Complex Systems, University of Michigan, Ann Arbor, Michigan, USA

[8] Department of Microbiology, Genetics, and Immunology, Michigan State University, East Lansing, Michigan, USA

[†] Equal contributions

[¶] Present address: Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

*Authors for correspondence:

Hadas Hawlena
hadashaw@bgu.ac.il

Jeffrey E. Barrick
jbarrick@cm.utexas.edu

Abstract

Parasites can adapt to better exploit their hosts on many timescales, ranging from within a single infection to across serial infections of multiple hosts. However, little is known about how the genomes of parasites in natural communities evolve when they face diverse hosts. We investigated how *Bartonella* bacteria that circulate in rodent communities in the dunes of the Negev Desert in Israel adapt to different hosts. We propagated 15 *Bartonella* populations through infections of either a single rodent species (*Gerbillus andersoni* or *Gerbillus pyramidum*) or alternating between the two. After 20 rodent passages, strains with *de novo* mutations outcompeted the ancestor in most lines. Mutations in two mononucleotide simple sequence repeats (SSRs) dominated the evolutionary dynamics and caused frameshifts in the same adhesin gene. These mutations appeared exclusively in populations that encountered *G. andersoni*, and they improved the ability to exploit this host. Similar SSRs in other genes are conserved and exhibit ON/OFF variation in *Bartonella* isolates from the Negev Desert dunes. Our results suggest that SSR-based contingency loci are important not only for rapidly and reversibly generating antigenic variation to escape immune responses but may also play a prominent role in the evolution of host specificity.

Parasites face constantly changing host environments. Individual hosts change within their lifetimes, different host individuals exhibit variation, and some parasites infect multiple host species. How parasites adapt to these dynamic environments is difficult to observe directly. Models and experiments related to the evolution of specialist versus generalist parasites find that the availability of different host species influences the trajectory of parasite evolution, both in terms of the level of host specificity and the frequency of host shifts[1–5]. These dynamics, especially in vertebrate hosts, are often determined by an interplay between host immune factors and parasite antigenic elements[6]. For example, parasites may respond to changes in host availability by altering the surface features that are targeted by the host immune system (i.e., through antigenic variation[7]). However, only limited information is available about how parasites adapt to vertebrate host diversity in natural communities.

To begin addressing this gap, we studied *Bartonella* bacteria that are flea-borne parasites of rodents living in sand dunes in the northwestern Negev Desert in Israel. This system offers the opportunity to investigate how parasites adapt in the context of a complex host community. In this region, a diverse collection of *Bartonella* strains circulates in communities with varying compositions of rodent host species including mainly *Gerbillus andersoni* and *Gerbillus pyramidum*[8,9]. *Bartonella* are prevalent in these rodent populations despite high strain turnover rates[10–12]. Infections with any individual strain are cleared by the host immune system, and that strain is subsequently unable to reinfect the same host[10]. This strong selection pressure from immune responses, coupled with the availability of multiple hosts, may favor *Bartonella* that have evolved the capacity to readily adapt to different host environments.

In line with this hypothesis, several genetic mechanisms that can facilitate rapid adaptation by generating antigenic variation have been noted in *Bartonella* species. First, *Bartonella* have high rates of intragenomic recombination events that copy, delete, and hybridize virulence genes with other nearby copies[13,14]. Second, *Bartonella* share a domesticated prophage that acts as a gene transfer agent, which can lead to recombination by exchanging DNA between co-infecting strains[15]. Finally, it has been reported that mononucleotide simple sequence repeats (SSRs) are overrepresented in *Bartonella* genomes compared to other prokaryotes[16]. In other parasites, genes containing SSRs have been shown to function as so-called "contingency" loci[17]. SSRs mutate at unusually high rates as a result of strand slippage during DNA replication, and these mutations often toggle expression of a gene between ON and OFF states[18]. When and how each of these mechanisms contributes to *Bartonella* adaptation to different hosts is unclear.

To examine how *Bartonella* adapt to different host environments, we passaged a strain isolated from the Negev Desert dunes (originally sampled from a *G.andersoni* individual) through rodents of the two main host species. We passaged *Bartonella* through each host species individually and, in a separate treatment, alternating between each host species. We tracked *Bartonella* evolution by whole-genome sequencing and asked two main questions: (i) Do the mutation rates and targets differ with host species availability? (ii) Do these mutations contribute to host adaptation? We found that mutations in SSRs in an adhesin gene dominated adaptation, arising in multiple lines independently and improving the ability of these *Bartonella* to exploit the more challenging (i.e., less favorable) host species. By sequencing the genomes of 38 *Bartonella* strains isolated in this region from wild gerbils, we found that similar SSRs are widespread and exhibit variation in their ON/OFF states. These results suggest that SSR contingency loci play an important role in allowing *Bartonella* to rapidly adapt to frequent host changes and thereby contribute to the remarkable local and global diversity of this genus of parasites.

Results

***Bartonella* evolution experiment in rodent hosts.** We evolved populations of *Bartonella krasnovii* OE-11 A2 for 363 days in three different rodent host scenarios: infecting *G. andersoni* only, infecting *G. pyramidum* only, or alternating between the two hosts (**Fig. 1**). We started all lines from the same single colony isolate. For each of the three treatments, we started five independent lines and then passaged each line through 20 individual rodents by isolating and injecting bacteria into a naïve host. At each passage, bacteria were isolated from blood after 15 days of infection, and we used this population to inoculate the next rodent. Similar infection dynamics were observed across all passages and host treatments, and bacterial loads in infected rodents at 15 days post-inoculation were comparable across all passages (**Supplementary Fig. 1a**). Moreover, what little variation there was in the number of bacteria inoculated at different passages did not lead to major differences in final bacterial loads in the recipient rodents (**Supplementary Fig. 1b**). The absence of any change in bacterial loads in the alternating host environment across passages, compared to the single host environments, suggests that trade-offs across host species did not play an important role in *Bartonella* evolution during the experiment.

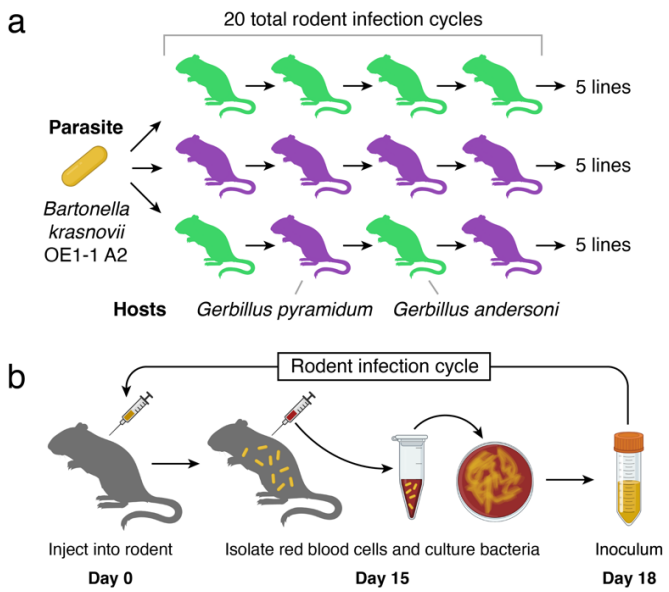

Fig. 1. *Bartonella* evolution experiment in rodent hosts. a, The design included 20 serial passages of *Bartonella krasnovii* OE1-1 A2 through individuals of one of the two host species (*G. andersoni* in green or *G. pyramidum* in purple) or alternating between them, with five independent rodent lines per treatment. **b**, In each infection cycle, captive *Bartonella*-negative rodents were inoculated intradermally with bacteria. Bacteria were cultured from red blood cells sampled 15 days later to create the inoculum for the next passage. A portion of each inoculum was archived for further study. Yellow rods and shading indicate bacteria in animal infections or samples. Figures created with artwork from Biorender.com.

Rates of genome evolution in different host treatments. We used whole-genome sequencing of 162 population samples and 147 endpoint clonal isolates to track new mutations that appeared during the evolution experiment (**Fig. 2**). The overall genetic dynamics were consistent with selective sweeps by lineages that acquired new beneficial mutations, including several instances of lineages with different mutations competing and displacing one another (**Fig. 2a**). Across all lines, we identified 27 unique mutations (**Supplementary Data 1**), which included 17 base substitutions, 7 small insertions or deletions of 18 or fewer bases, 1 three-base substitution, 1 gene conversion that changed three bases in a 44-bp stretch to a homologous sequence found elsewhere in the chromosome, and

1 large deletion of 33,117 bases. Of these mutations, 13 swept to high frequency and appeared to reach fixation in at least one experimental line, 2 were polymorphic in a final population (occurring in 2 or 8 of the 10 clones sequenced), 11 were found in only one sequenced endpoint clone and not observed in any population samples, and 1 was found only in population samples but not in a sequenced endpoint clone.

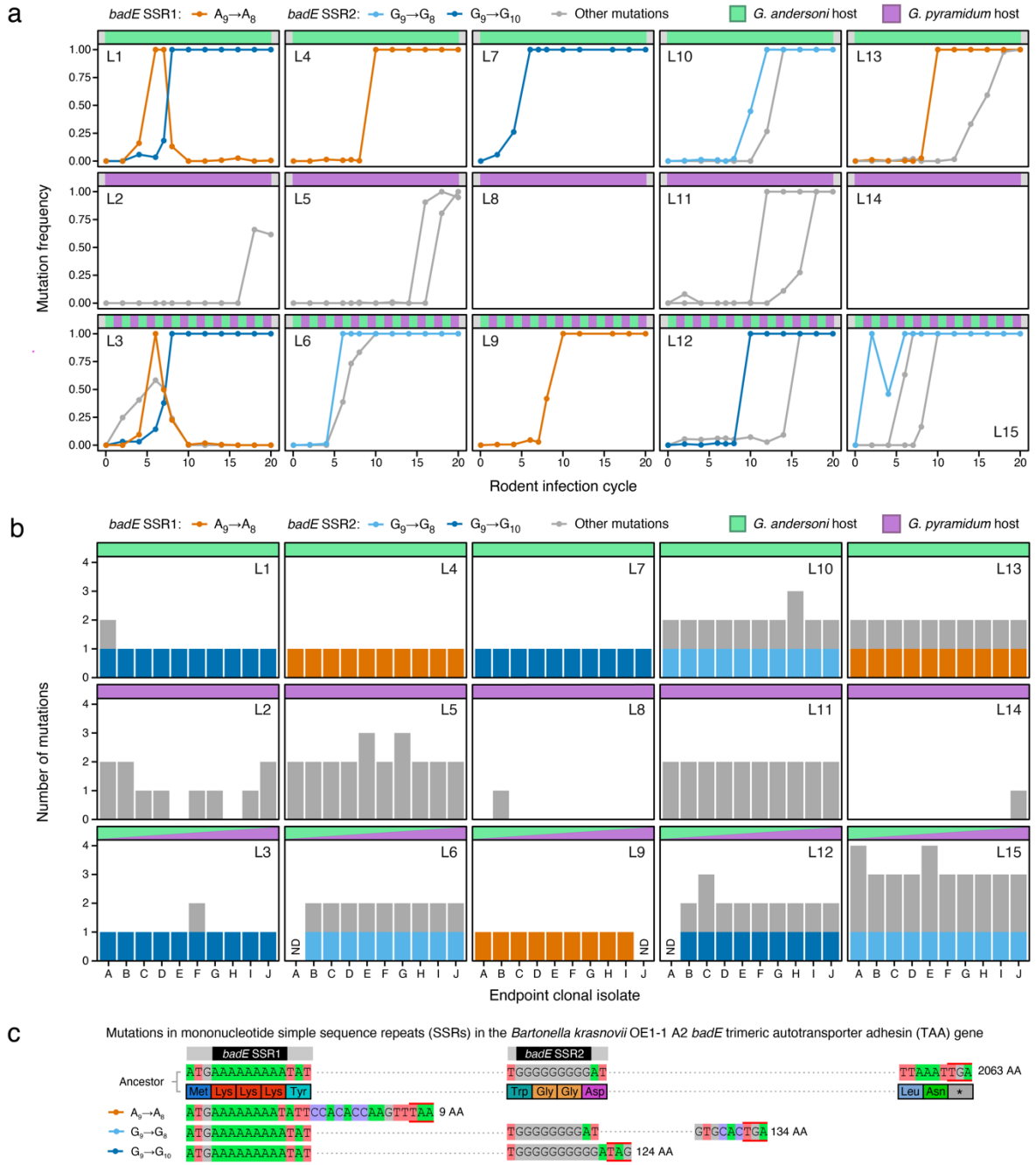

Fig. 2. *Bartonella* evolution is dominated by mutations in mononucleotide simple sequence repeats (SSRs) in the *badE* trimeric autotransporter adhesin (TAA) gene. a, Dynamics of competition between bacteria with new mutations in each of 15 experimental populations (L1–L15) over 20 rodent infection cycles. The relative abundance of each mutation was determined from metagenomic sequencing data. **b**, Number of mutations in each of 10 clones isolated at the end of the evolution experiment from each of 15 populations. Each panel shows results for the five populations that were passaged only in *G. andersoni* (L1, L4, L7, L10, L13), the five passaged only in *G. pyramidum* (L2, L5, L8, L11, L14), and the five passaged alternately in the two rodent species (L3, L6, L9, L12, L15). Sequencing data was insufficient to

call mutations in the three clones marked ND (not determined). In **a** and **b,** mutations that altered the lengths of two mononucleotide SSRs in *badE* are shown in color. All other mutations are displayed in grey. For full details see **Supplementary Data 1**. **c**, Effects of *badE* SSR mutations. Key portions of the ancestral gene's nucleotide and amino acid sequences are shown in the first two rows. SSR mutations and how they lead to frameshifts and premature stop codons that inactivate the gene are shown in the three bottom rows. Stop codons are indicated by red bars above and below nucleotide sequences. Dotted lines represent portions of the reading frames that are not shown.

We found up to four new mutations in each of the final clonal isolates (**Fig. 2b**). There were fewer mutations, on average, in an evolved *Bartonella* clone isolated from a population that was passaged only through *G. andersoni* (1.44 per clone) or only *G. pyramidum* (1.10 per clone) compared to a clone from a population passaged alternately through both rodent species (1.89 per clone). However, these differences among treatments in the rate of genome evolution were not statistically significant (repeated measures ANOVA: $F_{2,12} = 1.1$, $p = 0.37$).

Recurrent and parallel mutations in mononucleotide repeats in an adhesin gene. Indels in mononucleotide SSRs were the only successful mutations that recurred (independently arose in multiple lines) and exhibited genetic parallelism (with multiple distinct mutations affecting the same gene). Specifically, we detected changes in the lengths of two separate SSRs located within the first portion of the reading frame of the same trimeric autotransporter adhesin (TAA) gene (**Fig. 2c**), which we named *badE* for *Bartonella* adhesin that evolved. Within the *badE* open reading frame, we observed the deletion of a single A from a nine-base adenosine repeat in five populations, the deletion of a single G from a nine-base guanosine repeat in three populations, and the insertion of a single G in the same repeat in four populations. All three mutations result in frameshifts in *badE* that presumably inactivate its function. TAAs are considered virulence factors in *Bartonella*, and those that have been characterized in other *Bartonella* species mediate binding or adhesion to host cells[19–21].

These *badE* mutations occurred in 10 of the 15 lines, all of which experienced *G. andersoni* hosts. The SSR mutations increased in frequency until they were present in all or nearly all cells in these populations (**Fig. 2a**), and they were found in all of the corresponding endpoint clones (**Fig. 2b**). No mutations in these SSRs or elsewhere in *badE* were observed in any samples from the five lines that were passaged only in *G. pyramidum*. In some lines (L1 and L3), multiple mutations in the *badE* SSRs competed for dominance (**Fig. 2a**). The genome of the ancestor of the evolution experiment has 85 other A/T homopolymer repeats and 13 other G/C homopolymer repeats of 9 or more bases, but no mutations in these other SSRs were observed during the evolution experiment. Taken together, the recurrence, parallelism, temporal dynamics, and functional consequences of mutations in the two *badE* SSRs strongly suggest that they improved the fitness of *Bartonella* in the *G. andersoni* host.

Adhesin mutations affect infection dynamics. To test the effects of a *badE* SSR mutation under our experimental conditions, we compared the dynamics of infections of each rodent host species with the *B. krasnovii* OE1-1 A2 ancestor and an evolved clone in which the only observed mutation was a G insertion in the second *badE* SSR (**Fig. 3**). The duration of infections was significantly longer for both strains in *G. pyramidum* than in *G. andersoni* (ANOVA results for infection duration: $F_{1,21} = 14$, $p < 0.005$), indicating that the latter host presents a more challenging environment for the parasite. We found significant differences between the infection dynamics of the mutant and ancestor strains, with a faster increase in bacterial loads earlier in infections of *G. andersoni* but not *G. pyramidum* hosts (ANOVA, $F_{1,20} = 4.2$, $p = 0.05$ for host species × strain interaction and $p = 0.02$ for a planned comparison between the ancestor and evolved strains in *G. andersoni* at five days post-inoculation). This result confirms the expectation that this *badE* mutation is beneficial to

bacterial fitness during infections of *G. andersoni*. Additionally, this mutation does not seem to negatively impact bacterial fitness during *G. pyramidum* infections.

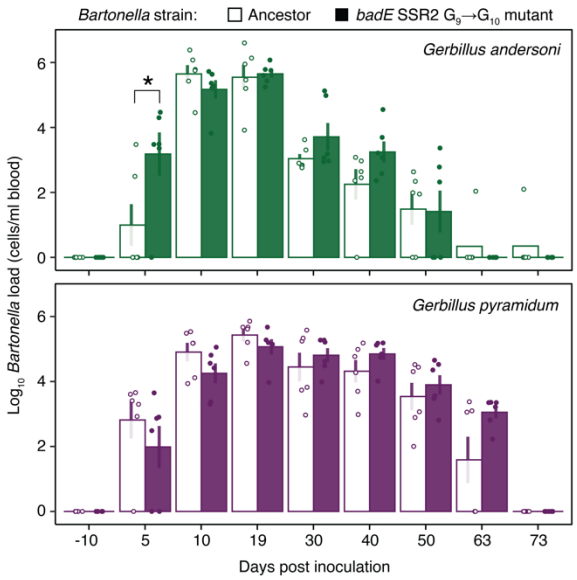

Fig. 3. Evolved SSR mutation in *badE* alters infection dynamics. The *B. krasnovii* OE1-1 A2 ancestor strain or an evolved strain, differing only by a mutation in *badE* SSR2, were inoculated into three pairs of female and three pairs of male *G. andersoni* and *G. pyramidum* twins (one sibling was inoculated with the ancestor and the other with the evolved strain). Bars show mean ± s.e. of the log-transformed bacterial titer in blood samples collected on the specified days after inoculation. Points show values for individual hosts. *Significant difference between the mutant and ancestor groups. See the main text for statistics.

One of the two *badE* SSRs is conserved in related *Bartonella* species. Our results suggest that the SSRs in the mutated *badE* TAA gene could make it function as a contingency locus, in which high rates of SSR-mediated mutations rapidly turn gene expression OFF and ON in a population of cells through reversible frameshifts. These changes might alter the parasite's localization within hosts and/or provide a mechanism for evading host immune responses. To examine whether these mononucleotide SSRs are found in homologous TAA genes in other *Bartonella*, we sequenced and assembled the genomes of 38 strains collected in the Negev Desert dunes from wild gerbils. We first compared TAA gene sequences in these genomes to those in genomes of representatives of eight *Bartonella* species not from this environment (**Fig. 4**). Protein clustering analyses indicate that the TAA gene in *B. krasnovii* is more closely related to a different TAA gene in *B. henselae* than it is to the highly characterized *badA* TAA gene[19,21]. Similarly, the TAA gene in *B. krasnovii* does not exhibit close homology to the *Vomp* TAA genes that have been characterized in *B. quintana*[20].

The Negev Desert *Bartonella* are classified into four species groups that are closely related to one another and to *Bartonella* from other locations that also infect rodents. Of the 38 Negev Desert strains we sequenced, 18 are *B. krasnovii*, 14 are *B. gerbillinarum*, 3 are *B. khokhlovae*, and 3 are *B. negeviensis*[8]. The A_9 SSR that mutated in the evolution experiment encodes three consecutive lysines immediately after the start codon in the reading frame of the *B. krasnovii* ancestor's *badE* gene (SSR1 in **Fig. 2a**). This homopolymer repeat is conserved in the TAA genes most similar to *badE* in all 18 *B. krasnovii* genomes. It is also present in the corresponding TAA genes in three of the four *Bartonella* species that are most closely related to *B. krasnovii*. By contrast, the G_9 SSR, which occurs later in the gene and overlaps four codons (TAA SSR2 in **Fig. 2a**), is present in only two *B. krasnovii* strains, the OE1-1 A2 ancestor and OE1-1 G.

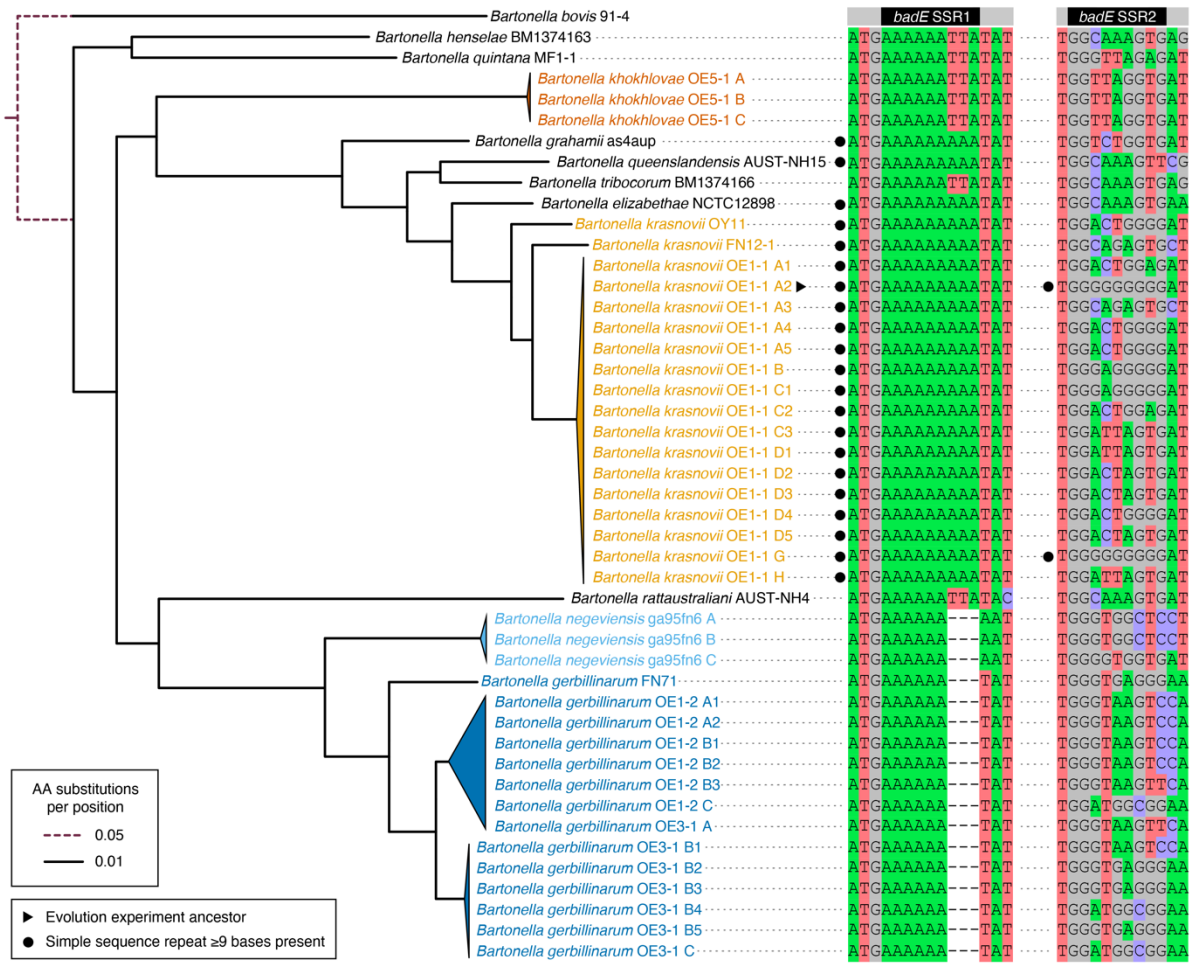

Fig 4. Conservation of mononucleotide SSRs in TAA genes most similar to the mutated *badE* gene in Negev Desert *Bartonella* and their relatives. The maximum likelihood phylogenetic tree of *Bartonella* strains on the left was constructed from a concatenated amino acid alignment of single-copy orthologs, with *B. bovis* 91-4 as the outgroup. Portions of the tree with very similar strains are collapsed into triangles. Every node in the tree not collapsed into a triangle was supported in all 100 bootstrap trees. The TAA gene in each *Bartonella* genome with the highest similarity to the first 200 amino acids of the *B. krasnovii* OE1-1 A2 *badE* gene that mutated during the evolution experiment was identified, and this region was aligned. The columns from this alignment that include or are adjacent to the two simple sequence repeats that mutated (SSR1 and SSR2) are shown on the right.

Other candidate SSR contingency loci in Negev Desert *Bartonella*. To examine whether the SSR mechanism that promotes evolvability might be more widespread, we searched the 38 Negev Desert *Bartonella* genomes for protein-coding genes containing mononucleotide repeats of nine bases or more. Each of these genomes contained roughly 50 such SSRs overlapping protein-coding genes (**Fig. 5a**), and very few of these repeats were longer than nine bases. Adenosine repeats predominated over repeats of other bases on the coding strand (**Fig. 5b**). They comprised 62–79% of all SSRs, depending on the species. SSRs were also noticeably concentrated at the 5′ ends of reading frames. We observed 20–35% of SSRs within the first 5% of the lengths of the genes they overlapped (**Fig. 5c**). This concentration of SSRs early in protein-coding genes was significantly greater than expected by chance in all four species ($p < 0.05$, one-tailed binomial tests).

Of the 181 different gene families that contained a mononucleotide SSR in at least two strains of the same Negev Desert *Bartonella* species, 41 were also found in at least two strains of one or more additional species (**Fig. 5d**). Some of these conserved SSRs are in genes with functions related to the bacterial cell surface and virulence, which suggests they could operate as contingency loci for the evolution of antigenic variation or changes in host specificity (**Table 1, Supplementary Data 2**). For example, one adenosine repeat shared by all four Negev Desert *Bartonella* species is in a homolog of the lectin-like protein BA14k, which was first described as an immunoreactive protein in *Brucella* infections of rodents[22]. We did not see variation in the lengths of most conserved SSRs in the Negev Desert *Bartonella* that would lead to ON/OFF variation in gene expression due to frameshifts.

Other genes, however, contained SSRs that did exhibit length variation within strains from a single Negev Desert *Bartonella* species (**Fig. 5e, Table 1, Supplementary Data 2**). These included a component of the VirB/VirD4 Type-IV secretion system (T4SS)[23], a T4SS effector, and three families of conserved *Bartonella* proteins of unknown function. We named one of these latter proteins variable *Bartonella* membrane protein A (*vbmA*). It contains a domain of unknown function (DUF6622), is predicted to have transmembrane helices, and has sequence similarity to NAD(P)/FAD-dependent oxidoreductases. The other *Bartonella* specific proteins with SSRs are from the same gene family. We named them variable *Bartonella* lipoproteins A and B (*vblA* and *vblB*) because they are predicted to be secreted by the Sec-system via signal peptides with cysteine at the cleavage site. Overall, these patterns of conservation and variation in SSRs (which are similar to the two SSRs that mutated during the evolution experiment) suggest that some operate as contingency loci that can rapidly and reversibly mutate to toggle gene expression between ON/OFF states.

Table 1. Highlighted genes with mononucleotide SSRs in Negev Desert *Bartonella* genomes

Orthogroup[§]	Gene[#]	SSR[†]	States[‡]	Species[¶]	Predicted function
OG0001164	–	A_{9-10}	ON	Bh, Bk, Bg, Bn	Lectin-like protein BA14k
OG0000503	*lpxD*	T_9	ON	Bk, Bg, Bn	UDP-3-*O*-acyl-glucosamine *N*-acyltransferase
OG0000315	*clsC*	T_9	ON	Bk, Bn	Cardiolipin synthase C
OG0000892	*lolA*	A_9, G_9	ON	Bk, Bn	Outer membrane lipoprotein carrier protein
OG0000423	*pssA*	A_9	ON	Bg, Bn	Phosphatidylserine synthase
OG0000008	–	C_{11}	ON	Bk, Bh	Autotransporter outer membrane β-barrel protein
OG0000618	*yhhT*	A_9	ON	Bg, Bn	Autoinducer-2 exporter (AI-2E) family protein
OG0001448	*pbuE*	G_{7-12}	ON/OFF	Bh, Bk	Purine efflux pump
OG0001520	*shlB*	T_9	ON	Bk, Bg	Hemolysin secretion/activation protein
OG0000099	*cdiA*	A_{7-9}	ON/OFF	Bk, Bg	T4SS effector, HNH/ENDO VII family nuclease
OG0000011	*vblA**	G_{10-11}	ON/OFF	Bk, Bg	Lipoprotein
OG0000011	*vblB**	G_{8-10}	ON/OFF	–	Lipoprotein
OG0001151	*acs*	A_{8-9}	ON/OFF	Bk	Acetyl-coenzyme A synthetase
OG0001185	*glnB*	A_{8-9}	ON/OFF	Bk	Nitrogen regulatory protein P-II
OG0000437	*gmk*	A_{8-9}	ON/OFF	Bh	Guanylate kinase
OG0001369	*virB9*	A_{8-9}	ON/OFF	Bk	T4SS system channel component
OG0000957	*vbmA**	C_{6-14}	ON/OFF	Bk	Membrane-localized NAD(P)/FAD-dependent oxidoreductase

[§]OrthoFinder orthogroup into which *Bartonella* genes containing the specified SSR were categorized.
[†]SSR base and lengths observed in this gene.
[‡]Predicted state of gene function based on whether any examples of the SSR generate frameshifts.
[¶]Species in which the genomes of at least two strains contain the same SSR with a length of ≥9 bases.
Abbreviations: Bh, *B. khokhlovae* Bk, *B. krasnovii*; Bg, *B. gerbillinarum*; Bn, *B. negeviensis*.
*Conserved *Bartonella* gene of unknown function named in this study.

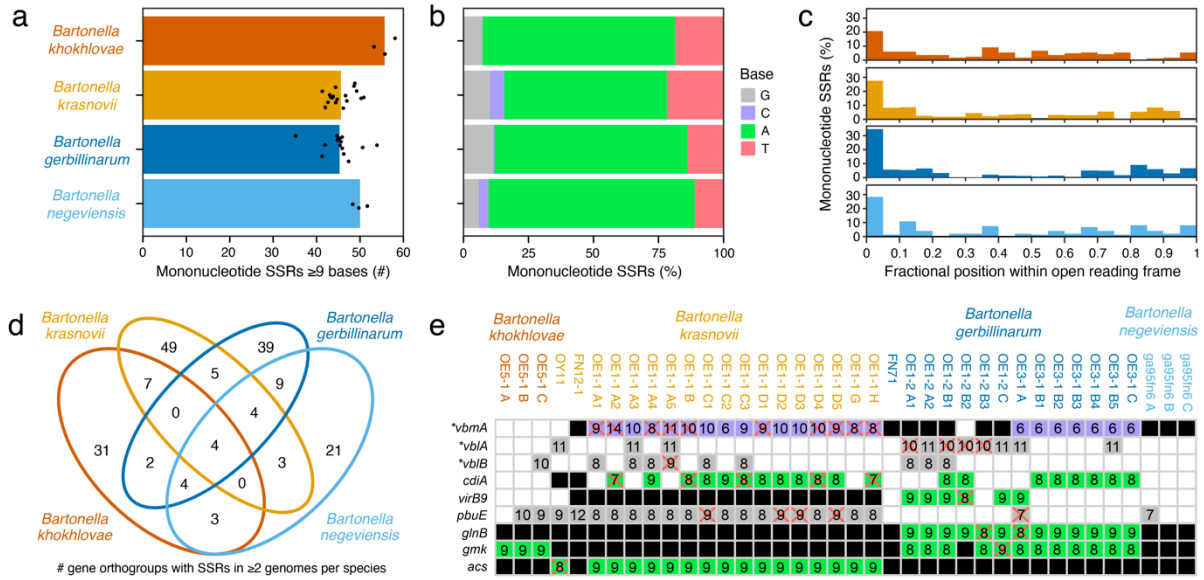

Fig 5. Mononucleotide simple sequence repeats (SSRs) in the genomes of Negev Desert *Bartonella*. a, Number of mononucleotide SSRs with lengths ≥9 bases in the genomes of 38 strains of *Bartonella* belonging to four species that were isolated from rodents in the Negev Desert dunes. Only SSRs overlapping protein-coding genes are included. Points are values for individual strains. Bars are averages within each species. **b**, Fraction of mononucleotide SSRs in each species that are repeats of each base. Bases are specified on the coding strand of the gene containing the SSR. **c**, Distribution of mononucleotide SSRs at different normalized positions within protein-coding open reading frames in each species. **d**, Venn diagram showing the number of SSRs that occurred in at least two genomes from each of the four *Bartonella* species and how they are shared between these species. **e**, Protein-coding genes exhibiting variation in the lengths of a mononucleotide SSR that result in frameshifts within the Negev Desert *Bartonella* strains. Numbers in boxes are the lengths of SSRs, with their backgrounds shaded by the repeated base as in **b**. A red X indicates a frameshift caused by the SSR in that gene. White boxes indicate that a gene was not identified in a genome. Black boxes indicate that the gene is present but does not have a mononucleotide SSR of ≥6 bases at that location. Asterisks indicate conserved genes of unknown function that were named in this study.

Discussion

To assess how differences in host species availability would influence parasite evolution, we passaged *Bartonella* populations through either a single rodent host species or alternating between two host species. We found that new genotypes with *de novo* mutations arose and outcompeted the ancestor strain in most bacterial populations by the end of the experiment. Mutations in two SSRs (simple sequence repeats) in the same *badE* TAA (trimeric autotransporter adhesin) gene dominated in lines that encountered one of the two rodent hosts. Evolution in changing host environments may favor different adaptive outcomes depending on mutation rates, bottleneck population sizes during infection, the strength of selection, the time spent in each host, the predictability of host shifts, and other factors. Our results provide insights into the evolutionary dynamics of *Bartonella*, specifically in the Negev Desert dunes, where these bacteria circulate within complex rodent communities.

Theoretical and empirical studies have shown that host diversity can either accelerate or impede the rate of parasite adaptation and can favor the evolution of parasites that are either specialists or generalists, depending on the rates of parasite mutation and host switching and whether adaptation to one host results in trade-offs that reduce fitness in another host. When the dynamics with which new parasite mutants arise and compete are slow compared to the rate of host switching, or when there are strong trade-offs, the evolution of generalists is favored, because any new parasite genotype must succeed in multiple host environments[24–26].

By contrast, when mutants that are more fit on one host can arise and dominate parasite populations before they switch hosts, and these mutants are less fit on other hosts, one expects parasites to evolve rapidly as they specialize on each successive host[27].

We found more mutations, on average, in the *Bartonella* lines that were passaged alternately through two rodent species than in lines that experienced only one host species. However, these rate differences were not statistically significant because the rates were low and fairly uniform. More tellingly, we observed that multiple host passages were required for new mutations to increase in frequency and approach fixation within the evolving bacterial populations. This result indicates that the bacteria's evolutionary dynamics could not keep pace with the rate of host switching in the alternating host treatment. We did not detect strong trade-offs across the host species in the effects of successful mutations. Such trade-offs would have been expected to impede the spread of new beneficial mutations in the alternating host environment compared to the single host environments, which we did not observe.

These evolutionary dynamics suggest that our alternating host treatment selected for *Bartonella* that were better generalists, which is expected to proceed with adaptation first and foremost to the host that is initially less favorable[24–26]. It is likely that *G. andersoni* is a more challenging host for the *B. krasnovii* strain that we used as the ancestor. We found that *G. andersoni* infections have a shorter duration than *G. pyramidum* infections, and the latter species is the preferred host for fleas that transmit these *Bartonella* strains in nature[28]. This model also agrees with how SSR mutations in *badE* evolved in all ten lines that included *G. andersoni* hosts, either alone or as part of the alternating treatment, but in none of the five lines that faced only *G. pyramidum* hosts. The *badE* mutant strain also reached a higher titer in red blood cells than the ancestor early in *G. andersoni* infections. This difference in dynamics is relevant to our experiment because we isolated bacteria to continue propagating a line from each host 15 days after it was inoculated. Adaptation to *G. andersoni* apparently occurred without a trade-off, as we did not see changes in bacterial loads in infected *G. pyramidum* hosts during the evolution experiment, nor did we see a difference between the TAA mutant and ancestor strain during infections of *G. pyramidum*.

The experimentally evolved SSR mutations in *badE* cause frameshifts, leading to truncated products that are almost certainly nonfunctional. The *badE* TAA gene product in *B. krasnovii* is homologous to key virulence factors in other *Bartonella* species. These homologs include *badA* from *B. henselae*, which is necessary for infecting endothelial cells and inducing a proangiogenic response[19], and *Vomp*s from *B. quintana*, which mediate adhesion to collagen and autoaggregation[20]. However, the TAA gene family that acquired the SSR mutations in our evolution experiment and in the natural *Bartonella* communities is distinct from these characterized families. Though the function of this *badE* TAA gene family is unknown, it is likely to mediate binding to the mammalian extracellular matrix or otherwise affect the interactions between bacterial and host cells[29]. TAA genes can be immunodominant antigens, and loss of Vomp expression has previously been observed during recurrent *B. quintana* bacteremia[20]. Loss of function mutations in the mutated *badE* TAA gene may be adaptive in some rodent hosts because they lead to different bacterial localization during infections or avoidance of the host's immune responses.

Our results suggest that SSR-based contingency loci play a more prominent role in *Bartonella* adaptation and, more generally, in the evolution of host specificity than previously appreciated. We observed that mononucleotide SSRs as long as those that mutated in the evolution experiment (nine or more bases) are common in the genomes of *Bartonella* isolates from the Negev Desert dunes, as had been previously noted for other *Bartonella* species[16]. These SSRs are concentrated at the 5′ ends of reading frames, as they are in other bacterial groups in which SSRs function as contingency loci[30,31]. Furthermore, we detected length variation in some SSRs in *Bartonella* species sampled from the Negev Desert, supporting the

hypothesis that they mediate rapid ON/OFF switching of gene expression in nature. Some of the conserved and varying SSRs are associated with known virulence factors (e.g., type IV secretion systems and their effectors); others are associated with conserved *Bartonella* genes of unknown function that are predicted to be membrane-localized or secreted, which may interact with hosts and/or elicit immune responses. The fixation of SSR mutations in our experimental *Bartonella* populations contrasts with the findings from an experiment with *Campylobacter jejuni* in a mouse model. In the latter system no single SSR mutation reached 100% frequency in the evolved populations[32]. This difference suggests there were more extreme population bottlenecks, less within-host replication, or both during the *Bartonella* infections in our experiments.

Our experiment reproduces some of the host complexity experienced by *Bartonella* in the Negev Desert, but it does not reflect other aspects of the natural environment, including transmission by fleas, co-infection of rodents with multiple *Bartonella* strains and species, and re-infection of rodents with immune systems that have cleared prior infections[8,12,33,34]. Fine-scale longitudinal sampling and genome sequencing of *Bartonella* in wild rodents could provide insights into the importance of SSR mutations under these conditions relative to other mechanisms that can promote evolvability, such as gene transfer agent-mediated recombination between co-infecting strains and species[15] and within-genome recombination in virulence gene arrays[13,14]. *Bartonella* bacteria possess many evolvability mechanisms that may contribute to the remarkable prevalence and diversity of these parasites, underscoring the danger of zoonotic disease through spillover into humans.

Our findings broaden the typical view of the importance of contingency loci in parasite adaptation. Even when the rates of fixation of mutations in them cannot keep pace with host immune responses or movement to new hosts, these loci may provide an advantage for parasites. They can do so by generating focused genetic variation that allows for rapid adaptation to a challenging host species in a way that also leaves open the possibility of restoring the inactivated gene if it is needed for success in future host environments.

Methods

Evolution experiment. We serially passaged *B. krasnovii* strain OE-11 A2 (originally isolated from a *G. andersoni* host in the Negev Desert) either through 20 individuals of one of the two host species (*G. andersoni* and *G. pyramidum*) or by alternating between them (10 individuals of each host species). Each of the three treatments had five replicate lines. We also included a negative control line that was transmitted from one uninfected host to the next throughout the experiment and negative control hosts that were inoculated in each passage with phosphate-buffered saline (PBS). Neither control showed any evidence of *Bartonella* infection at any point. A full description of the protocol that explains the rationale for specific decisions (e.g., the ancestor strain, inoculation procedure, bacterial quantification methods, inoculation source, and day of sampling) has been published elsewhere[35]. Briefly, in each passage, *Bartonella*-negative rodents were intradermally inoculated with 100 µl of a PBS suspension of bacteria isolated from the previously infected rodent ($5 \times 10^6 \pm 5 \times 10^5$ total cells/inoculum; **Supplementary Fig. 1b**). Rodents were bled by cardiac puncture 15 days after inoculation to collect *Bartonella* at the approximate peak of bacteremia. Red blood cells from each sample were spread on two chocolate agar (CA) plates. These plates were incubated for 3 days at 37°C and 5% CO_2. Then, bacterial cells were harvested by scraping the plates and resuspending the cells in PBS. This solution was homogenized by vortexing it with glass beads and passing it through a 5-µm filter. These samples were used to inoculate rodents in the next passage.

To assess bacterial loads in the rodent blood 15 days after each inoculation, we extracted DNA from the blood samples using a QIAamp BiOstic Bacteremia DNA Kit (Qiagen), following the manufacturer's instructions. To assess bacterial titers in the PBS suspensions used for inoculation, we extracted DNA using a thermal lysis procedure, as described previously[8]. In each set of extractions, we included a negative control, in which all the reagents were added to PBS but without adding the blood or bacterial suspension. We quantified bacterial loads by real-time quantitative PCR (qPCR), following Eidelman *et al.*[36]. The standard curve was created using DNA extracted from the ancestor strain and was calibrated to the count of colony-forming units of this sample.

Genome sequencing. Bacterial populations isolated from rodents after passages 2, 4, 6, 7, 8, 10, 12, 14, 16, 18, and 20 were preserved from the PBS solution that remained after inoculating the next rodents by centrifugation, resuspending them in 1 ml of 20% (w/v) glycerol in lysogeny broth (LB), and then storing the samples at −80°C. DNA was extracted from these samples using the DNeasy Blood & Tissue Kit (Qiagen), following the protocol for gram-negative bacteria with slight modifications: centrifugation of the initial solution was at 5400 × g for 10 min, and DNA elution was repeated twice in 50 µl of AE buffer that was preheated to 56°C before adding it to the purification column and incubating it for 5 min at room temperature. We also spread 100 µl of 10^{-5} and 10^{-6} dilutions of the bacterial solution that we cultivated from rodents after the last passage on CA plates. After 5 days of incubation at 37°C and 5% CO_2, we isolated 10 endpoint colonies for each experimental line. Each colony was propagated three times by streaking on a new CA plate and regrowing to ensure a clonal isolate. Following the last regrowth, 15 colonies were scraped from the plate for each clonal isolate and preserved in 500 µl of 20% (w/v) glycerol in LB. DNA extractions from these clone samples were performed as described above for the population samples.

In total, DNA was extracted from 318 samples: 165 populations (15 lines × 11 time points) and 153 clones (15 lines × 10 clonal isolates per population plus three replicates of the ancestor). Up to 100 ng of purified gDNA from each sample was put into the 2S Turbo DNA Library Kit (Swift Biosciences). All reactions were carried out at 20% of the manufacturer's recommended volumes, with dual 6-bp indexes incorporated during the final 10 cycle PCR step. The resulting libraries were pooled and sequenced on an Illumina HiSeq X Ten instrument by Psomagen (Rockville, MD) to generate 151-base paired-end reads. One population sample that was contaminated with DNA from a different *Bartonella* species was discarded before analysis.

Mutation identification. We removed adaptors from the Illumina reads using *trimmomatic* (v0.39) in paired-end mode with the following settings: 4 allowed mismatches to the seed, a palindrome clip threshold of 30, a simple clip threshold of 10, and discarding trimmed reads ≤ 30 bases. To improve mutation calling, we assembled and annotated an updated version of the previously reported genome sequence of *Bartonella krasnovii* strain OE-11 A2[37]. We used Trycycler (v0.5.3)[38] on nanopore long reads from a prior study (SRA:SRR6873507)[39] to generate a consensus assembly from Flye[40], Miniasm+Minipolish[41,42], and Raven[43] assemblies. Then, we polished the assembly by comparing it to our Illumina short reads for the ancestor strain using *breseq* (v0.36.1)[44] in consensus mode. Discrepancies between reads and the assembly were iteratively corrected by using the gdtools APPLY command and re-running *breseq* until no further improvements were found. Lastly, we used Prokka (v1.14.6)[45] to annotate the final assembly. The updated *B. krasnovii* OE-11 A2 reference sequence includes a chromosome with 2,163,391 base pairs and a plasmid with 29,057 base pairs.

To call mutations in evolved clones and populations, we compared trimmed Illumina reads to the reference genome using *breseq* (v0.36.1)[44]. This pipeline identifies point

mutations, small insertions and deletions, and structural mutations that can be predicted from split-read alignments, including insertion sequence (IS) element movements and large deletions. We first ran *breseq* in polymorphism mode with default parameters separately on each sample. Then, we combined the candidate mutations predicted by all *breseq* runs using gdtools MERGE and ran *breseq* again on each sample with the merged file provided as user evidence, so that it reported read counts supporting all candidate mutations in all samples. After this point, we did not further analyze samples that did not have ≥18-fold average read-depth coverage of the chromosome; one ancestor sample, two population samples, and three clone samples did not pass this cutoff.

Mis-mapping of reads to the reference genome, low coverage, and sequencing errors can lead to spurious mutation predictions. We employed several strategies to detect and disregard mutation predictions that are likely false-positives due to these or other biases. They rely on the following assumptions. First, mutations are expected to appear in clonal isolates with frequencies of either 0% or 100% and not at intermediate frequencies. Second, mutations should exhibit a frequency trajectory in population samples that begins at 0% and changes gradually over time in a correlated fashion, as subpopulations with the mutation outcompete others to reach high frequency or are outcompeted and diminish in frequency during the evolution experiment. Third, it is unlikely that the exact same mutation would be observed in all or most of the independent lines of the evolution experiment.

To implement these criteria, we began by keeping all mutations that were predicted at ≥90% frequency in any clone. Then, other predictions were examined for hallmarks of systematic errors of the types described above by analyzing all samples except for the ancestor controls. First, we disregarded mutation predictions in clonal or population samples that did not have at least 10 total reads supporting either the reference or mutated sequence because estimates of their frequencies are subject to large errors. If >5% of all samples did not meet this threshold, that mutation candidate was not considered further. Then, we removed mutations that did not reach a frequency of 20% in at least one population sample. Next, we filtered out mutations that were observed exclusively in the population samples and were predicted to have a frequency of ≥5% in at least one population from 8 or more of the 15 different experimental lines. Additionally, we required the summed frequency of a mutation across all clone and population samples to be ≥100% for it to be considered further.

Finally, we fit a null Poisson regression model with a uniform rate of generating reads supporting the variant across all population samples. The offset in this model was the total number of reads that were informative about whether that mutation was present or not. For mutations supported by new sequence junctions the offset was corrected for the number of read start position registers that would lead to reads that could be definitively assigned to the junction. We compared this null model to a Poisson regression model that allowed for per-sample variation in the rate of generating reads with a mutation. If the likelihood ratio test comparing these two models was rejected after performing the Benjamini-Hochberg correction for multiple testing across all candidate mutations at a 10^{-6} false-discovery rate (indicating insufficient evidence for a mutation), then it was removed from consideration.

Mutation candidates that remained after these steps were manually examined and merged when they were linked by proximity and matching frequency time courses. For example, three nearby base substitutions that all seem to have resulted from one gene conversion event in line L11 were merged. The frequency of the merged mutation is reported as the average of the frequencies predicted for each individual mutation in these cases. Two mutations that passed all criteria peaked at frequencies of ~33% and ~25% in the population samples from lines L11 and L13, respectively, and these mutations were also found at roughly the same frequencies in the endpoint clonal isolates from these lines. These mutations may be in near-identical regions of the genome that were not individually resolved during assembly of the

reference sequence. We corrected the frequencies of these mutations by dividing by the estimated peak value and capping the maximum frequency at 100%. For clarity, the frequencies of mutations were also assigned to be 0% in all samples outside of lines in which they were clearly present because they reached a frequency of >20% and did not exhibit sporadic jumps in frequency across population and clone samples. One mutation in L8 that appeared at 100% frequency in all of this line's population samples and clones was not counted or analyzed. Of the 27 mutations in the final list, 24 were at a >90% frequency in a clonal sample, 12 were predictions in population samples that passed the filters that eliminated systematic errors, and 9 were in both categories (**Supplementary Data 1**).

Infection experiment with *badE* mutant and ancestor strains. We inoculated three male twins and three female twins from each rodent species with either the ancestor or a mutant strain, and we compared their infection dynamics until the rodents cleared the infections. The mutant strain was a colony isolated after passage 20 from the L1 line, which was passaged exclusively in *G. andersoni* rodents. It was genetically identical to the ancestor except for a $G_9 \rightarrow G_{10}$ mutation in *badE* SSR2. We also had one control rodent of each species that was inoculated with PBS. To assess the rodents' bacterial loads, we bled them (50 µl of blood per sample) on days 5 and 10 post-inoculation and then every 10 days until day 73 post-inoculation. Blood was taken from the retro-orbital sinus under general anesthesia, and a drop of local anesthesia (Localin, Fischer Pharmaceutical Labs, Tel Aviv, ISR) was placed in the eye. We used the same DNA extraction and qPCR protocols described in the section on the evolution experiment.

Negev Desert *Bartonella* genome assembly and annotation. We assessed conservation and diversity in SSRs that may function as contingency loci across *B. krasnovii* OE-11 A2 and 37 additional *Bartonella* isolates classified into different genotypes. These isolates were cultured from fleas and blood sampled from 33 wild *G. andersoni* and 33 wild *G. pyramidum* rodents that were captured during October 2016 in two different sites in the northwestern Negev Desert sand dunes in Israel (34°23′E, 30°58′N and 34°23′E, 30°55′N)[8]. We performed DNA extractions and Illumina sequencing of these clonal isolates as described above for samples from the evolution experiment. Then, we assembled draft genome sequences using Unicycler (v0.4.9b)[46] on read subsets downsamled to 90× nominal coverage. Contigs that matched the genome of *Acinetobacter baylyi* ADP1 (GenBank: NC_005966.1) contaminated the initial 75C-4a isolate assembly due to barcode misassignment. They were identified and removed based on BLAST matches in Bandage (v0.8.1)[47]. Genes were annotated in the assemblies using Prokka (v1.14.6)[45] with the option to include Rfam predictions of noncoding RNAs. For all comparative analyses, we analyzed the *B. krasnovii* OE1-1 A2 sequence assembled in this same way, rather than the closed assembly we used to analyze the evolution experiments.

Orthogroup identification and phylogenetic tree construction. To identify orthologous gene groups, we used OrthoFinder (v.2.5.5)[48] on proteins predicted in the genome assemblies of the 38 Negev Desert *Bartonella* isolates and representatives of 8 *Bartonella* species that infect other hosts. The phylogenetic tree was created by aligning the amino acid sequences of the 808 predicted single-gene orthologs with MAFFT (v7.508)[49], trimming these alignments with BMGE (v1.12)[50] using the BLOSUM62 substitution matrix, and inferring a maximum likelihood tree with IQ-Tree (v2.2.3)[51] from the concatenated alignment with 265,987 columns using a JTT+FO+R10 model and 100 bootstraps. We used *B. bovis* 91-4 as the outgroup to root the resulting tree.

Trimeric autotransporter adhesin (TAA) sequence analysis. We aligned the amino acid sequences of proteins assigned to the orthogroup containing both the TAA gene *PRJBM_RS00735* from *B. henselae* BM1374163 and the *B. krasnovii* OE-11A2 *badE* TAA gene using MUSCLE (v3.8.1)[52], with further manual refinement of the resulting alignment in AliView[53]. To recover incomplete reading frames interrupted by contig boundaries, we also added the best TBLASTN (v2.12.0+)[54] matches in each genome to the first 200 amino acids of the *B. krasnovii* OE-11A2 TAA sequence. To further delineate TAA protein families, we used InterProScan (v5.65-97.0)[55] to identify all proteins with matches to the head, stalk, and membrane anchor domains of YadA-like trimeric autotransporter adhesins (IPR008640, IPR008635, and IPR005594, respectively) in the 46 genomes that were analyzed. We aligned and clustered the sequences of these proteins and examined their placement relative to *B. henselae* BadA and *B. quintana* Vomp TAAs[19,20].

Mononucleotide simple sequence repeat (SSR) survey. Mononucleotide repeats of ≥9 bases in the 46 *Bartonella* genomes were identified and associated with orthogroups using custom Python scripts that used Biopython (v1.81)[56] and pandas[57] functions and R scripts that used tidyverse packages[58]. Orthogroups were assigned initial gene names and descriptions based on annotations in the input genomes, with further refinement through NCBI BLAST[54] and InterProScan[55] searches. We calculated overall statistics for SSRs that overlap protein-coding genes in the Negev Desert *Bartonella* species. Then, we further analyzed orthogroups with the SSRs found in at least two strains of any of the four Negev Desert *Bartonella* species (**Supplementary Data 2**). We manually examined SSRs in sequence alignments for evidence of frameshifts caused by addition or deletion of the repeated bases using AliView[53]. For the *Bartonella* gene families of unknown function that contained SSRs with length variation, we predicted signal peptides using SignalP-6.0 (v0.0.56)[59] and transmembrane helices using DeepTMHMM (v1.0.24)[60].

Data availability

Raw sequencing data are available from the US National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProjects PRJNA914220 and PRJNA1061316. Biological materials are available upon reasonable request to H.H and S.H.

Code availability

Analysis scripts and source data are available at https://github.com/barricklab/ND-Bartonella-EE and https://github.com/barricklab/ND-Bartonella-SSR and have been archived in Zenodo at https://doi.org/10.5281/zenodo.10467299 and https://doi.org/10.5281/zenodo.10467301.

References

1. Kassen, R. The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* **15**, 173–190 (2002).
2. Cornwall, D. H. *et al.* Experimental manipulation of population-level MHC diversity controls pathogen virulence evolution in *Mus musculus*. *J. Evol. Biol.* **31**, 314–322 (2018).
3. Elena, S. F. Local adaptation of plant viruses: lessons from experimental evolution. *Mol. Ecol.* **26**, 1711–1719 (2017).
4. Visher, E. & Boots, M. The problem of mediocre generalists: population genetics and eco-evolutionary perspectives on host breadth evolution in pathogens. *Proc. Biol. Sci.* **287**, 20201230 (2020).

5. Betts, A., Rafaluk, C. & King, K. C. Host and parasite evolution in a tangled bank. *Trends Parasitol.* **32**, 863–873 (2016).
6. Cobey, S. Pathogen evolution and the immunological niche. *Ann. N. Y. Acad. Sci.* **1320**, 1–15 (2014).
7. Deitsch, K. W., Lukehart, S. A. & Stringer, J. R. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat. Rev. Microbiol.* **7**, 493–503 (2009).
8. Gutiérrez, R. *et al.* Untangling the knots: Co-infection and diversity of *Bartonella* from wild gerbils and their associated fleas. *Mol. Ecol.* **27**, 4787–4807 (2018).
9. Kedem, H. *et al.* Multiple effects of host-species diversity on coexisting host-specific and host-opportunistic microbes. *Ecology* **95**, 1173–1183 (2014).
10. Rodríguez-Pastor, R. *et al. Bartonella* infections are prevalent in rodents despite efficient immune responses. *Parasit. Vectors* **16**, 315 (2023).
11. Cohen, C., Einav, M. & Hawlena, H. Path analyses of cross-sectional and longitudinal data suggest that variability in natural communities of blood-associated parasites is derived from host characteristics and not interspecific interactions. *Parasit. Vectors* **8**, 429 (2015).
12. Gutiérrez, R., Morick, D., Cohen, C., Hawlena, H. & Harrus, S. The effect of ecological and temporal factors on the composition of *Bartonella* infection in rodents and their fleas. *ISME J.* **8**, 1598–1608 (2014).
13. Engel, P. *et al.* Parallel evolution of a type IV secretion system in radiating lineages of the host-restricted bacterial pathogen *Bartonella*. *PLoS Genet.* **7**, e1001296 (2011).
14. Thibau, A. *et al.* Long-read sequencing reveals genetic adaptation of *Bartonella* adhesin A among different *Bartonella henselae* isolates. *Front. Microbiol.* **13**, 838267 (2022).
15. Québatte, M. & Dehio, C. *Bartonella* gene transfer agent: Evolution, function, and proposed role in host adaptation. *Cell. Microbiol.* **21**, e13068 (2019).
16. Coenye, T. & Vandamme, P. Characterization of mononucleotide repeats in sequenced prokaryotic genomes. *DNA Res.* **12**, 221–233 (2005).
17. Moxon, E. R., Rainey, P. B., Nowak, M. A. & Lenski, R. E. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4**, 24–33 (1994).
18. Moxon, R., Bayliss, C. & Hood, D. Bacterial contingency loci: The role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* **40**, 307–333 (2006).
19. Riess, T. *et al. Bartonella* adhesin A mediates a proangiogenic host cell response. *J. Exp. Med.* **200**, 1267–1278 (2004).
20. Zhang, P. *et al.* A family of variably expressed outer-membrane proteins (Vomp) mediates adhesion and autoaggregation in *Bartonella quintana*. *Proc. Natl. Acad. Sci.* **101**, 13630–13635 (2004).
21. Kaiser, P. O. *et al.* The head of *Bartonella* adhesin A is crucial for host cell interaction of *Bartonella henselae*. *Cell. Microbiol.* **10**, 2223–2234 (2008).
22. Chirhart-Gilleland, R. L., Kovach, M. E., Elzer, P. H., Jennings, S. R. & Roop, R. M. Identification and characterization of a 14-kilodalton *Brucella abortus* protein reactive with antibodies from naturally and experimentally infected hosts and T lymphocytes from experimentally infected BALB/c mice. *Infect. Immun.* **66**, 4000–4003 (1998).
23. Dehio, C. Infection-associated type IV secretion systems of *Bartonella* and their diverse roles in host cell interaction. *Cell. Microbiol.* **10**, 1591–1598 (2008).
24. Cooper, T. F. & Lenski, R. E. Experimental evolution with *E. coli* in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. *BMC Evol. Biol.* **10**, 11–11 (2010).

25. Sandberg, T. E., Lloyd, C. J., Palsson, B. O. & Feist, A. M. Laboratory evolution to alternating substrate environments yields distinct phenotypic and genetic adaptive strategies. *Appl. Environ. Microbiol.* **83**, e00410-17 (2017).
26. Deatherage, D. E., Kepner, J. L., Bennett, A. F., Lenski, R. E. & Barrick, J. E. Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E1904–E1912 (2017).
27. Palmer, M. E., Lipsitch, M., Moxon, E. R. & Bayliss, C. D. Broad conditions favor the evolution of phase-variable loci. *mBio* **4**, e00430-12 (2013).
28. Messika, I. *et al.* From endosymbionts to host communities: factors determining the reproductive success of arthropod vectors. *Oecologia* **184**, 859–871 (2017).
29. Kiessling, A. R., Malik, A. & Goldman, A. Recent advances in the understanding of trimeric autotransporter adhesins. *Med. Microbiol. Immunol. (Berl.)* **209**, 233–242 (2020).
30. Orsi, R. H., Bowen, B. M. & Wiedmann, M. Homopolymeric tracts represent a general regulatory mechanism in prokaryotes. *BMC Genomics* **11**, 102–102 (2010).
31. Lin, W.-H. & Kussell, E. Evolutionary pressures on simple sequence repeats in prokaryotic coding regions. *Nucleic Acids Res.* **40**, 2399–2413 (2012).
32. Jerome, J. P. *et al.* Standing genetic variation in contingency loci drives the rapid adaptation of *Campylobacter jejuni* to a novel host. *PLoS ONE* **6**, e16399 (2011).
33. Morick, D., Krasnov, B. R., Khokhlova, I. S., Gottlieb, Y. & Harrus, S. Investigation of *Bartonella* acquisition and transmission in *Xenopsylla ramesis* fleas (Siphonaptera: Pulicidae). *Mol. Ecol.* **20**, 2864–2870 (2011).
34. Cohen, C., Toh, E., Munro, D., Dong, Q. & Hawlena, H. Similarities and seasonal variations in bacterial communities from the blood of rodents and from their flea vectors. *ISME J.* **9**, 1662–1676 (2015).
35. Rodríguez-Pastor, R. *et al.* A road map for in vivo evolution experiments with blood-borne parasitic microbes. *Mol. Ecol. Resour.* **22**, 2843–2859 (2022).
36. Eidelman, A. *et al.* The dynamics between limited-term and lifelong coinfecting bacterial parasites in wild rodent hosts. *J. Exp. Biol.* **222**, jeb.203562 (2019).
37. Gutiérrez, R. *et al. 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. *Int. J. Syst. Evol. Microbiol.* **70**, 1656–1665 (2020).
38. Wick, R. R. *et al.* Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biol.* **22**, 266 (2021).
39. Gutiérrez, R. *et al.* Prophage-driven genomic structural changes promote *Bartonella* vertical evolution. *Genome Biol. Evol.* (2018) doi:10.1093/gbe/evy236.
40. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
41. Li, H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016).
42. Wick, R. R. & Holt, K. E. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Research* **8**, 2138 (2019).
43. Vaser, R. & Šikić, M. Time- and memory-efficient genome assembly with Raven. *Nat. Comput. Sci.* **1**, 332–336 (2021).
44. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using *breseq*. *Methods Mol. Biol.* **1151**, 165–188 (2014).
45. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

46. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput. Biol.* **13**, e1005595 (2017).
47. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350–3352 (2015).
48. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
49. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
50. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
51. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
52. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
53. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinforma. Oxf. Engl.* **30**, 3276–3278 (2014).
54. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421–421 (2009).
55. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
56. Cock, P. J. A. *et al.* Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
57. McKinney, W. Data structures for statistical computing in Python. in *Proc. of the 9th Python in Science Cond.* 56–61 (2010). doi:10.25080/Majora-92bf1922-00a.
58. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
59. Teufel, F. *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**, 1023–1025 (2022).
60. Hallgren, J. *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural network. Preprint at https://doi.org/10.1101/2022.04.08.487609 (2022).

Acknowledgements

This work was funded by grants from the U.S.-Israel Binational Science Foundation (2017708 to H.H. and S.H.), the U.S. National Science Foundation (DEB-1813069 to L.Z., R.E.L. and J.E.B.), and Israel Foundation Science grant (1391/15 to H.H.). A.Z.H. was supported by the Zuckerman STEM Leadership Program. R.R.P. was supported by fellowships from the Kreitman School of Advanced Graduate Studies and the Swiss Institute for Dryland Environmental and Energy Research. R.E.L. was supported by a Hatch grant from the U.S. Department of Agriculture (MICL02253) to Michigan State University. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high-performance computing resources.

Author Contributions

Conceptualization (S.H., L.Z., R.E.L., J.E.B., and H.H), investigation (R.R-P., N.K., N.S., A.Z.H., D.E.D., R.G., J.E.B., and H.H.), formal analysis (J.E.B. and H.H.), visualization (A.Z.H., J.E.B., and H.H.), software (J.E.B.), writing – original draft (J.E.B. and H.H.), writing – review & editing (R.R-P., N.K., R.G., L.Z., A.Z.H., S.H., R.E.L., J.E.B., and H.H.) and funding acquisition (S.H., L.Z., R.E.L., J.E.B., and H.H).

Competing Interests

The authors declare no competing interests.