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[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

Keywords: Experimental evolution, disease ecology, eco-evolutionary dynamics, fluctuating environments, host-pathogen interactions, trimeric autotransporter adhesin

Abstract

Parasites can adapt to better exploit their hosts on many scales, ranging from within an infection of a single individual to series of infections spanning multiple host species. 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 species of rodent hosts. We propagated 15 *Bartonella* populations through infections of either a single host species (*Gerbillus andersoni* or *Gerbillus pyramidum*) or alternating between the two. After 20 rodent passages, strains with *de novo* mutations replaced the ancestor in most populations. Mutations in two mononucleotide simple sequence repeats (SSRs) that caused frameshifts in the same adhesin gene dominated the evolutionary dynamics. They appeared exclusively in populations that encountered *G. andersoni* and altered the dynamics of infections of 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 could be important not only for rapidly and reversibly generating antigenic variation to escape immune responses but that they may also mediate the evolution of host specificity.

Introduction

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

64 dynamics, especially in vertebrate hosts, are often determined by an interplay between host
65 immune factors and parasite antigenic elements [6]. For example, parasites may respond to
66 changes in host availability by altering the surface features that are targeted by the host
67 immune system, i.e., through antigenic variation [7]. However, only limited information is
68 available about how parasites adapt to vertebrate host diversity in natural communities.
69 To begin addressing this gap, we studied *Bartonella* bacteria that are flea-borne parasites
70 of rodents living in sand dunes in the northwestern Negev Desert in Israel. This system offers
71 the opportunity to investigate how parasites adapt in the context of a complex host
72 community. In this region, a diverse collection of *Bartonella* strains circulates in
73 communities with varying compositions of rodent host species including mainly *Gerbillus*
74 *andersoni* and *Gerbillus pyramidum* [8,9]. *Bartonella* are prevalent in these rodent
75 populations despite high strain turnover rates [10–12]. Infections with an individual strain are
76 cleared by the host immune system, and that strain is subsequently unable to reinfect the
77 same host[10]. This strong selection pressure from immune responses, coupled with the
78 availability of multiple hosts, may favor *Bartonella* that have evolved the capacity to readily
79 adapt to different host environments.
80 In line with this hypothesis, several genetic mechanisms that can facilitate rapid
81 adaptation by generating antigenic variation have been noted in *Bartonella* species. First,
82 *Bartonella* have high rates of intragenomic recombination events that copy, delete, and
83 hybridize virulence genes with other nearby copies [13,14]. Second, *Bartonella* share a
84 domesticated prophage that acts as a gene transfer agent, which can lead to recombination by
85 exchanging DNA between co-infecting strains [15]. Finally, it has been reported that
86 mononucleotide simple sequence repeats (SSRs) of five or more bases are overrepresented in
87 *Bartonella* genomes compared to other prokaryotes [16]. In other parasites, genes containing
88 SSRs have been shown to function as so-called “contingency” loci [17]. SSRs mutate at

89 unusually high rates as a result of strand slippage during DNA replication, and these
90 mutations often toggle expression of a gene between ON and OFF states [18]. When and how
91 each of these mechanisms contributes to *Bartonella* adaptation is unclear.

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

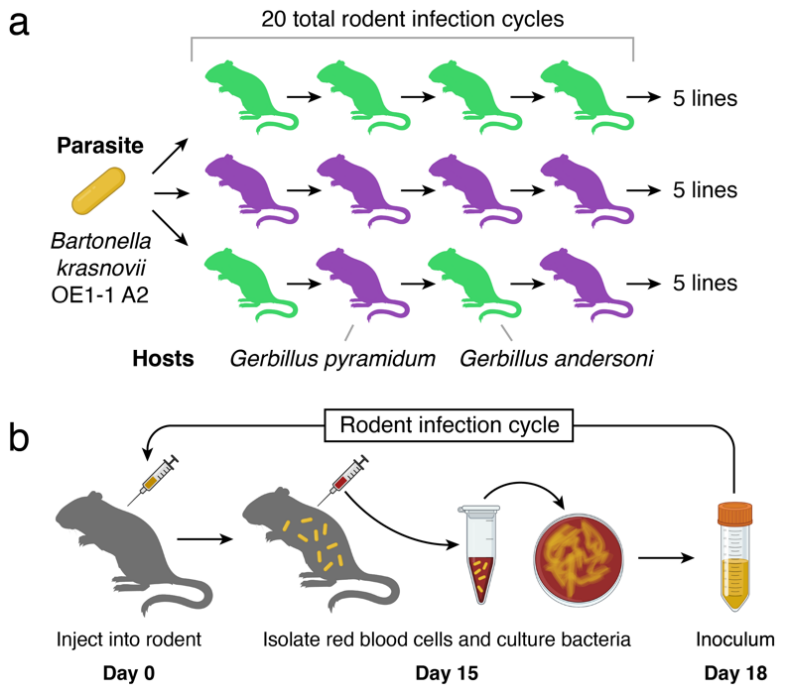
106 Results

107 *Bartonella* evolution experiment in rodent hosts

108 We evolved populations of *Bartonella krasnovii* OE-11 A2 for 363 days in three different
109 rodent host scenarios: infecting *G. andersoni* only, infecting *G. pyramidum* only, or
110 alternating between the two hosts (**Fig. 1**). We started all lines from the same single colony
111 isolate. For each of the three treatments, we started five independent lines and then passaged
112 each line through 20 individual rodents by isolating and injecting bacteria into a naïve host.

113 At each passage, bacteria were isolated from blood after 15 days of infection, and we used
114 this population to inoculate the next rodent. Similar infection dynamics were observed across
115 all passages and host treatments, and bacterial loads in infected rodents at 15 days post-
116 inoculation were comparable across all passages (**Fig. S1a**). Moreover, what little variation
117 there was in the number of bacteria inoculated at different passages did not lead to major
118 differences in final bacterial loads in the recipient rodents (**Fig. S1b**). The absence of any
119 change in bacterial loads in the alternating host environment across passages, compared to the
120 single host environments, suggests that trade-offs across host species did not play an
121 important role in *Bartonella* evolution during the experiment.

122



123

124 **Fig. 1. *Bartonella* evolution experiment in rodent hosts.** (**a**) The design included 20 serial passages of
125 *Bartonella krasnovii* OE1-1 A2 through individuals of one of the two host species (*G. andersoni* in green
126 or *G. pyramidum* in purple) or alternating between them, with five independent rodent lines per treatment.
127 (**b**) In each infection cycle, captive *Bartonella*-negative rodents were inoculated intradermally with
128 bacteria. Bacteria were cultured from red blood cells sampled 15 days later to create the inoculum for the
129 next passage. A portion of each inoculum was archived for further study. Yellow rods and shading indicate
130 bacteria in animal infections or samples. Figures incorporate 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 (**File S1**), 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.

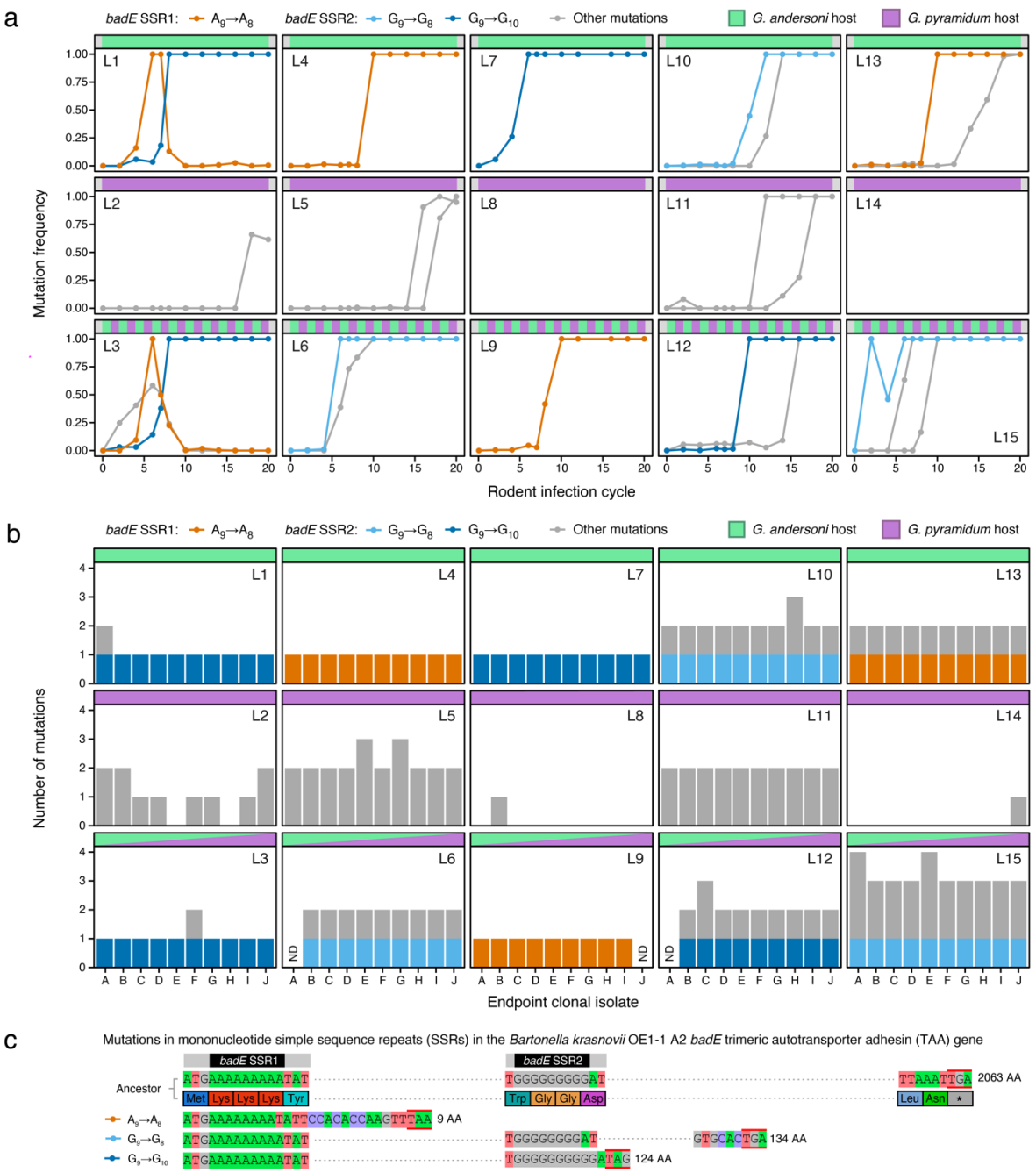


Fig. 2. *Bartonella* evolution is dominated by mutations in mononucleotide simple sequence repeats (SSRs) in the *badE* trimeric autotransporter adhesin (TAA) gene. (**a**) Dynamics of competition between bacteria with new mutations in each of 15 experimental populations (L1–L15) over 20 rodent infection 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), or the five passaged alternately in the two rodent species (L3, L6, L9, L12, L15). Sequencing data was

155 insufficient to call mutations in the three clones marked ND (not determined). In **a** and **b,** mutations that
156 altered the lengths of two mononucleotide SSRs in *badE* are shown in color. All other mutations are
157 displayed in grey. For full details see **File S1**. **c**, Effects of *badE* SSR mutations. Key portions of the
158 ancestral gene's nucleotide and amino acid sequences are shown in the first two rows. SSR mutations and
159 how they lead to frameshifts and premature stop codons that inactivate the gene are shown in the three
160 bottom rows. Stop codons are indicated by red bars above and below nucleotide sequences. Dotted lines
161 represent portions of the reading frames that are not shown.
162
163 We found up to four new mutations in each of the final clonal isolates (**Fig. 2b**). There
164 were fewer mutations, on average, in an evolved *Bartonella* clone isolated from a population
165 that was passaged only through *G. andersoni* (1.44 per clone) or only *G. pyramidum* (1.10
166 per clone) compared to a clone from a population passaged alternately through both rodent
167 species (1.89 per clone). However, these differences among treatments in the rate of genome
168 evolution were not statistically significant (repeated measures ANOVA: $F_{2,12} = 1.1$, $p = 0.37$).

169 Recurrent and parallel SSR mutations occur in an adhesin gene

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

181 These *badE* mutations occurred in 10 of the 15 lines, all of which experienced *G.*
182 *andersoni* hosts. The SSR mutations increased in frequency until they were present in all or
183 nearly all cells in these populations (**Fig. 2a**), and they were found in all of the corresponding
184 endpoint clones (**Fig. 2b**). No mutations in these SSRs or elsewhere in *badE* were observed
185 in any samples from the five lines that were passaged only in *G. pyramidum*. In some lines
186 (L1 and L3), multiple mutations in the *badE* SSRs competed for dominance (**Fig. 2a**). The
187 genome of the ancestor of the evolution experiment has 85 other A/T homopolymer repeats
188 and 13 other G/C homopolymer repeats of 9 or more bases, but no mutations in these other
189 SSRs were observed during the evolution experiment.

190 We also re-analyzed sequencing data from a prior experiment that passaged populations
191 of a closely related strain of *B. krasnovii* through 50 single colony passages *in vitro* on agar
192 plates [22]. There were no changes in the lengths of the same two SSRs in the *badE* gene
193 under the conditions of relaxed selection in this experiment. Thus, hypermutability of these
194 SSRs is unlikely to be sufficient for explaining why they consistently mutated in our *in vivo*
195 evolution experiment. Taken together, the recurrence, temporal dynamics, and host-
196 specificity of mutations in the two *badE* SSRs suggest that they improved the fitness of
197 *Bartonella* in the *G. andersoni* host under the infection conditions of our experiment.

198 Adhesin mutations affect infection dynamics

199 To test the effects of a *badE* SSR mutation under our experimental conditions, we compared
200 the dynamics of infections of each rodent host species with the *B. krasnovii* OE1-1 A2
201 ancestor and an evolved clone in which the only observed mutation was a G insertion in the
202 second *badE* SSR (**Fig. 3**). The duration of infections was significantly longer for both strains
203 in *G. pyramidum* than in *G. andersoni* (ANOVA results for infection duration: $F_{1,21} = 14$, $p <$
204 0.005), indicating that the latter host presents a more challenging environment for the
205 parasite. We found significant differences between the infection dynamics of the mutant and

206 ancestor strains, with a faster increase in bacterial loads earlier in infections of *G. andersoni*

207 but not *G. pyramidum* hosts (ANOVA, $F_{1,20} = 4.2$, $p = 0.05$ for host species × strain

208 interaction and $p = 0.02$ for a planned comparison between the ancestor and evolved strains in

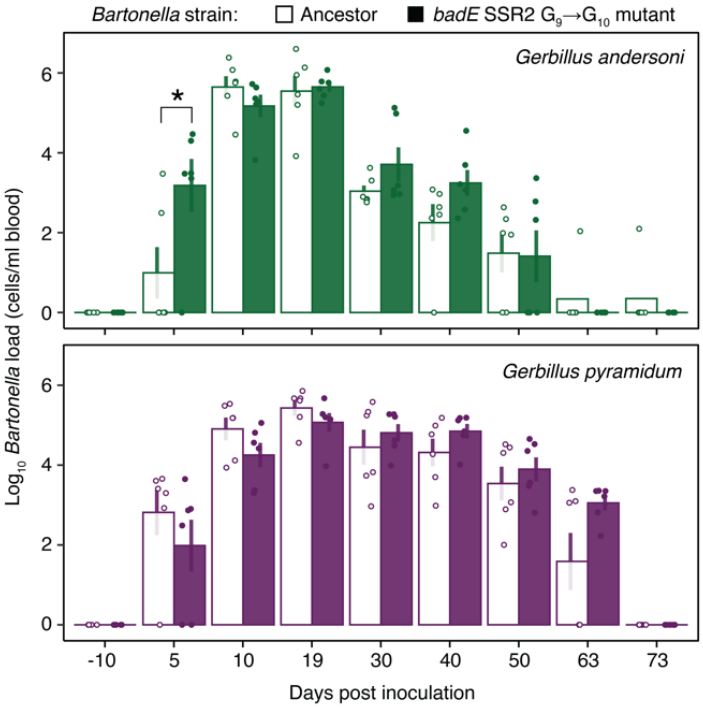
209 *G. andersoni* at five days post-inoculation). This result confirms the expectation that this

210 *badE* mutation is beneficial to bacterial fitness during infections of *G. andersoni*.

211 Additionally, this mutation does not seem to negatively impact bacterial fitness during *G.*

212 *pyramidum* infections.

213



214

215 **Fig. 3. Evolved SSR mutation in *badE* alters infection dynamics.** The *B. krasnovii* OE1-1 A2 ancestor

216 strain or an evolved strain, differing only by a mutation in *badE* SSR2, were inoculated into three pairs of

217 female and three pairs of male *G. andersoni* and *G. pyramidum* twins (one sibling was inoculated with the

218 ancestor and the other with the evolved strain). Bars show mean ± s.e. of the log-transformed bacterial titer

219 in blood samples collected on the specified days after inoculation. Points show values for individual hosts.

220 *Significant difference between the mutant and ancestor groups. See the main text for statistics.

221

222 ## Evidence that *badE* is a SSR-mediated contingency locus

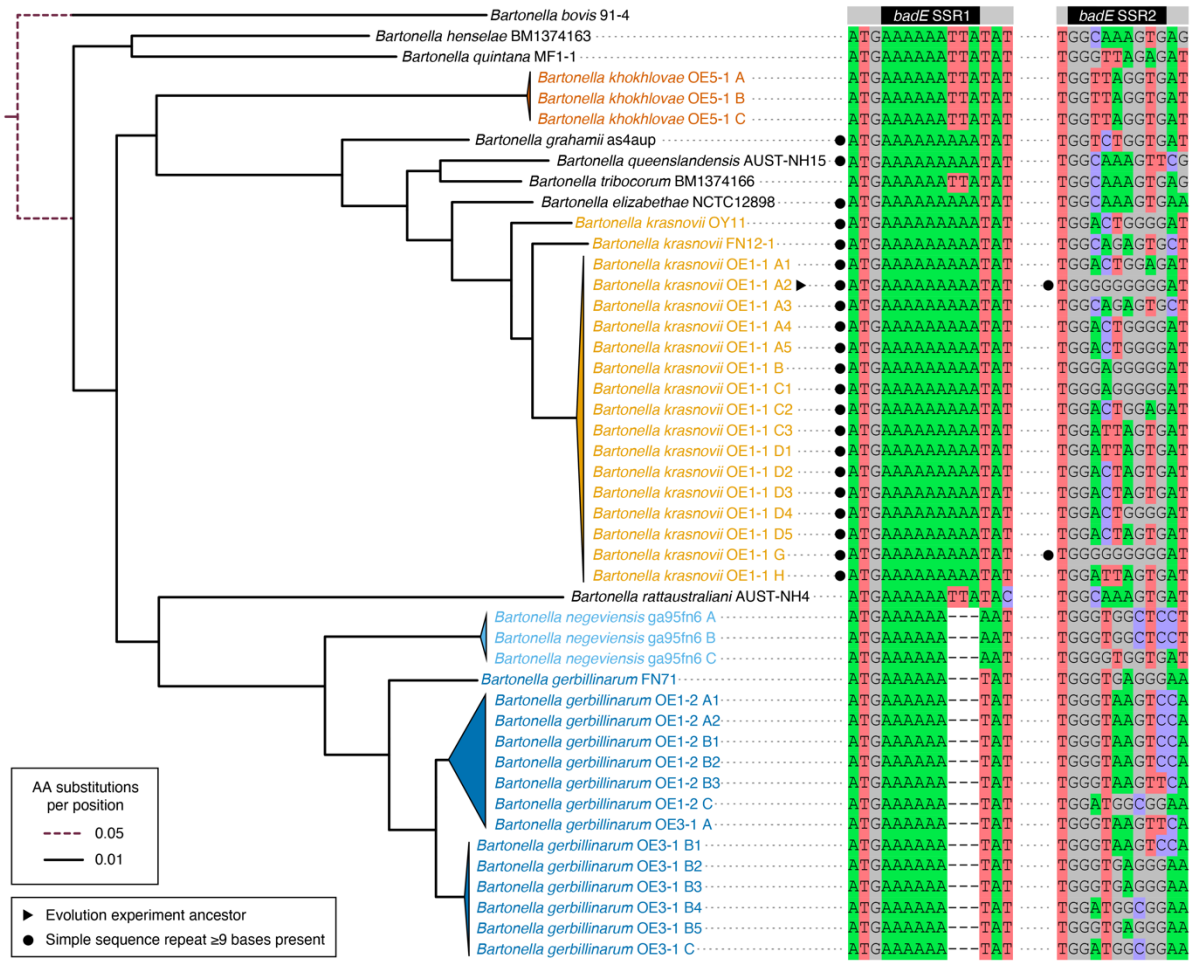
223 Our results suggest that the SSRs in the *badE* gene that mutated in the evolution experiment
224 might have been selected and preserved by evolution so that this TAA gene can function as a
225 contingency locus, in which high rates of SSR-mediated mutations can rapidly toggle gene
226 expression OFF and ON within a population of cells through reversible frameshifts.
227 Alternatively, it is possible that the SSRs in the *badE* gene mediate evolvability in our
228 experiment simply because they are mutational hotspots that happen to be there by chance.
229 The hypothesis that the *badE* gene evolved to act as a continency locus would be further
230 supported if: (1) the prevalence of hypermutable SSRs in its sequence was unusual and (2)
231 the presence of these SSRs was conserved in related strains and species.

232 To examine the first question, we performed randomization tests in which we preserved
233 the amino acid sequences of genes in the *B. krasnovii* OE1-1 A2 ancestor of the evolution
234 experiment but shuffled the codons used to encode them within each reading frame (see
235 **Methods**). Codon-shuffled versions of *badE* have 0.57 mononucleotide SSRs of nine or more
236 bases, on average, compared to the two in the actual gene. This overrepresentation is only
237 marginally significant ($p = 0.080$, one-tailed randomization test). However, this model does
238 not take into account that there are typically fewer long mononucleotide SSRs than expected
239 by chance in functionally important portions of microbial genomes [16,23–25], which has
240 been attributed to selection against the genetic load of having hypermutable sites in essential
241 genes [17]. In line with this expectation, the protein-coding genes of *B. krasnovii* OE1-1 A2
242 have just 50 mononucleotide SSRs of nine bases or more compared to the average of 206 that
243 are present in codon-shuffled gene sets, which is a significant depletion ($p < 0.001$, one-tailed
244 randomization test). If we re-evaluate the SSRs in *badE* relative to this depleted baseline,
245 then two SSRs is significantly more than one would expect ($p = 0.006$, one-tailed
246 randomization test), and observing even one would be somewhat unusual (13.5% of codon-

247 shuffled genes). In conclusion, there is some evidence that the *B. krasnovii* OE1-1 A2 *badE*
248 gene has more SSRs than expected by chance.
249 To further examine the evolutionary conservation of *badE* and its SSRs, we sequenced
250 and assembled the genomes of 38 *Bartonella* isolates collected in the Negev Desert dunes
251 from wild gerbils. We identified and aligned the most closely related TAA gene sequences in
252 each of these genomes and from eight representative *Bartonella* species that are not from this
253 environment (**Fig. 4, File S2**). Protein clustering based on sequence homology revealed that
254 the TAA gene in *B. krasnovii* is more closely related to a different TAA gene in *B. henselae*
255 than it is to the highly characterized *badA* TAA gene [19,21]. Similarly, the TAA gene in *B.*
256 *krasnovii* does not exhibit close homology to the *Vomp* TAA genes of *B. quintana* [20].
257 The Negev Desert *Bartonella* are classified into four species groups that are closely
258 related to one another and to *Bartonella* from other locations that also infect rodents. Of the
259 38 Negev Desert strains we sequenced, 18 are *B. krasnovii*, 14 are *B. gerbillinarum*, 3 are *B.*
260 *khokhlovae*, and 3 are *B. negeviensis* [8]. The A_9 SSR that mutated in the evolution
261 experiment encodes three consecutive lysines immediately after the start codon in the reading
262 frame of the *B. krasnovii* ancestor's *badE* gene (SSR1 in **Fig. 2a and 4**). This SSR is
263 conserved in the TAA genes most similar to *badE* in all 18 *B. krasnovii* genomes. It is also
264 present in the corresponding TAA genes in three of the four *Bartonella* species that are most
265 closely related to *B. krasnovii*. By contrast, the G_9 SSR is present in only two strains: the
266 OE1-1 A2 ancestor and OE1-1 G (SSR2 in **Fig. 2a**). To put evolutionary conservation of the
267 A_9 SSR into perspective, we examined an alignment of the 18 *B. krasnovii badE* genes and
268 the most closely related *badE* gene from *B. elizabethae.* In this alignment, there are a total of
269 7762 nine-base windows, of which 3742 do not contain gaps. Only 358 of these (9.57%) have
270 the same sequence in all 17 genes, as is the case for the A_9 SSR. In summary, we find that the

271 A_9 SSR in *badE* is conserved within *B. krasnovii* and related species and that this is unusual

272 but not exceedingly so compared to conservation of other sequences within this gene.



273

274 **Fig 4. Conservation of mononucleotide SSRs in TAA genes most similar to the mutated *badE* gene in**

275 **Negev Desert *Bartonella* and their relatives.** The maximum likelihood phylogenetic tree of *Bartonella*

276 strains on the left was constructed from a concatenated amino acid alignment of single-copy orthologs,

277 with *B. bovis* 91-4 as the outgroup. Portions of the tree with very similar strains are collapsed into

278 triangles. Every node in the tree not collapsed into a triangle was supported in all 100 bootstrap trees. The

279 TAA gene in each *Bartonella* genome with the highest similarity to the first 200 amino acids of the *B.*

280 *krasnovii* OE1-1 A2 *badE* gene that mutated during the evolution experiment was identified, and this

281 region was aligned. The columns from this alignment that include or are adjacent to the two simple

282 sequence repeats that mutated (SSR1 and SSR2) are shown on the right.

283

284 ## Other candidate SSR contingency loci in Negev Desert *Bartonella*

285 To examine whether SSRs that might promote the evolvability of contingency loci were more
286 widespread, we searched the 38 Negev Desert *Bartonella* genomes for protein-coding genes
287 containing mononucleotide repeats of nine bases or more (**File S3**). As was the case for *B.*
288 *krasnovii* OE1-1 A2, each of these genomes contained roughly 50 such SSRs (**Fig. 5a**).
289 Adenosine repeats predominated over repeats of other bases on the coding strand (**Fig. 5b**).
290 They comprised 62–79% of all SSRs, depending on the species. SSRs were also noticeably
291 concentrated at the 5′ ends of reading frames. We observed that 20–35% of SSRs within the
292 first 5% of the DNA sequence of the gene they overlapped (**Fig. 5c**). This concentration of
293 SSRs very early in genes was significantly greater than expected by chance in all four species
294 ($p < 0.05$, one-tailed binomial tests). This trend also recapitulates what we observed in the *B.*
295 *krasnovii* OE1-1 A2 *badE* gene where the A_9 SSR was immediately after the start codon and
296 the G_9 SRR was within the first 6% of its length (**File S2**).

297 Of the 181 different gene families that contained a mononucleotide SSR in at least two
298 strains of the same Negev Desert *Bartonella* species, 41 were also found in at least two
299 strains of one or more additional species (**Fig. 5d**). Some of these conserved SSRs are in
300 genes with functions related to the bacterial cell surface and virulence, which suggests they
301 could operate as contingency loci for the evolution of antigenic variation or changes in host
302 specificity (**Table 1, File S3**). For example, one adenosine repeat shared by all four Negev
303 Desert *Bartonella* species is in a homolog of the lectin-like protein BA14k, which was first
304 described as an immunoreactive protein in *Brucella* infections of rodents [26]. We did not see
305 variation in the lengths of most conserved SSRs in the Negev Desert *Bartonella* that would
306 lead to ON/OFF variation in gene expression due to frameshifts.

307 Other genes, however, contained SSRs that did exhibit length variation within strains
308 from a single Negev Desert *Bartonella* species (**Fig. 5e, Table 1, File S3**). These included a

309 component of the VirB/VirD4 Type-IV secretion system (T4SS) [27], a T4SS effector, and
310 three families of conserved *Bartonella* proteins of unknown function. We named one of these
311 latter proteins variable *Bartonella* membrane protein A (*vbmA*). It contains a domain of
312 unknown function (DUF6622), is predicted to have transmembrane helices, and has sequence
313 similarity to NAD(P)/FAD-dependent oxidoreductases. The other *Bartonella* specific
314 proteins with SSRs are from the same gene family. We named them variable *Bartonella*
315 lipoproteins A and B (*vblA* and *vblB*) because they are predicted to be secreted by the Sec-
316 system via signal peptides with cysteine at the cleavage site. Overall, these patterns of
317 conservation and variation in SSRs (which are similar to the two SSRs that mutated during
318 the evolution experiment) suggest that some could operate as contingency loci to rapidly and
319 reversibly mutate to toggle gene expression between ON/OFF states.
320

321 **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

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

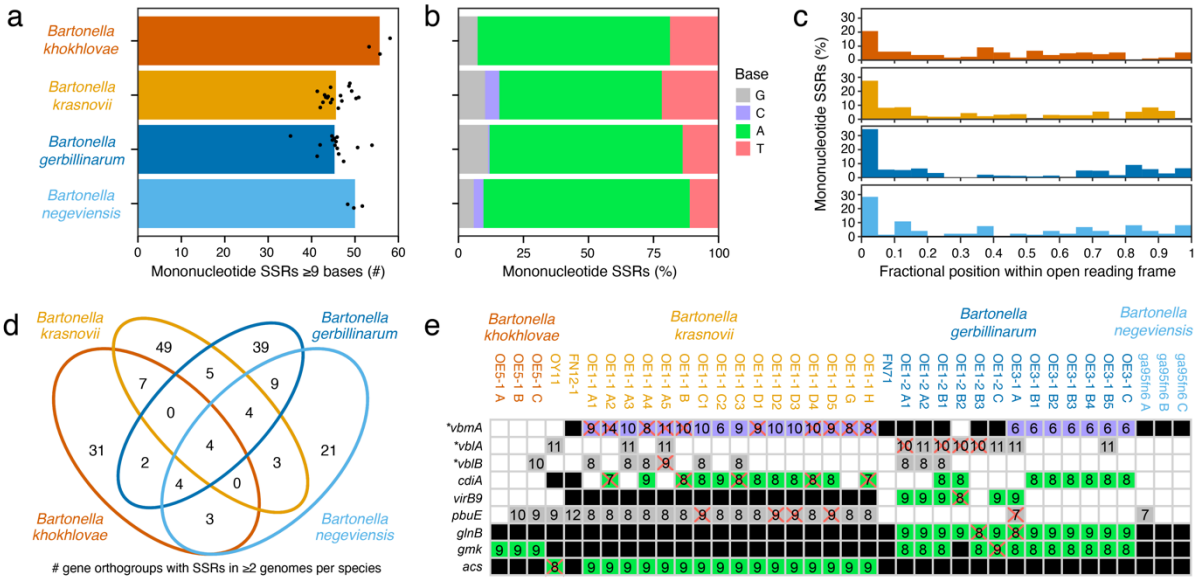


Fig 5. Mononucleotide simple sequence repeats (SSRs) in the genomes of Negev Desert *Bartonella*. (**a**) Number of mononucleotide SSRs with lengths ≥9 bases in the genomes of 38 strains of *Bartonella* belonging to four species that were isolated from rodents in the Negev Desert dunes (**File S3**). 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.

345 # Discussion

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

357 Theoretical and empirical studies have shown that host diversity can either accelerate or
358 impede the rate of parasite adaptation and can favor the evolution of parasites that are either
359 specialists or generalists, depending on the rates of parasite mutation and host switching and
360 whether adaptation to one host results in trade-offs that reduce fitness in another host. When
361 the dynamics with which new parasite mutants arise and compete are slow compared to the
362 rate of host switching, or when there are strong trade-offs, the evolution of generalists is
363 favored, because any new parasite genotype must succeed in multiple host environments [28–
364 30]. By contrast, when mutants that are more fit on one host can arise and dominate parasite
365 populations before they switch hosts, and these mutants are less fit on other hosts, one
366 expects parasites to evolve rapidly as they specialize on each successive host [31].

367 We found more mutations, on average, in the *Bartonella* lines that were passaged
368 alternately through two rodent species than in lines that experienced only one host species.
369 However, these rate differences were not statistically significant because the rates were low

370 and fairly uniform. More tellingly, we observed that multiple host passages were required for
371 new mutations to increase in frequency and approach fixation within the evolving bacterial
372 populations. This result indicates that the bacteria's evolutionary dynamics could not keep
373 pace with the rate of host switching in the alternating host treatment. We did not detect strong
374 trade-offs across the host species in the effects of successful mutations. Such trade-offs would
375 have been expected to impede the spread of new beneficial mutations in the alternating host
376 environment compared to the single host environments, which we did not observe.

377 These evolutionary dynamics suggest that our alternating host treatment selected for
378 *Bartonella* that were better generalists, which is expected to proceed with adaptation first and
379 foremost to the host that is initially less favorable [28–30]. It is likely that *G. andersoni* is a
380 more challenging host for the *B. krasnovii* strain that we used as the ancestor. We found that
381 *G. andersoni* infections have a shorter duration than *G. pyramidum* infections, and the latter
382 species is the preferred host for fleas that transmit these *Bartonella* strains in nature [32]. This
383 model also agrees with how SSR mutations in *badE* evolved in all ten lines that included *G.*
384 *andersoni* hosts, either alone or as part of the alternating treatment, but in none of the five
385 lines that faced only *G. pyramidum* hosts. The *badE* mutant strain also reached a higher titer
386 in red blood cells than the ancestor early in *G. andersoni* infections. This difference in
387 dynamics is relevant to our experiment because we isolated bacteria to continue propagating
388 a line from each host 15 days after it was inoculated. Adaptation to *G. andersoni* apparently
389 occurred without a trade-off, as we did not see changes in bacterial loads in infected *G.*
390 *pyramidum* hosts during the evolution experiment, nor did we see a difference between the
391 TAA mutant and ancestor strain during infections of *G. pyramidum*.

392 The experimentally evolved SSR mutations in *badE* cause frameshifts, leading to
393 truncated products that are almost certainly nonfunctional. The *badE* TAA gene product in *B.*
394 *krasnovii* is homologous to key virulence factors in other *Bartonella* species. These homologs

395 include *badA* from *B. henselae*, which is necessary for infecting endothelial cells and
396 inducing a proangiogenic response [19], and *Vomp*s from *B. quintana*, which mediate
397 adhesion to collagen and autoaggregation [20]. However, the TAA gene family that acquired
398 the SSR mutations in our evolution experiment and in the natural *Bartonella* communities is
399 distinct from these characterized families. Though the function of this *badE* TAA gene family
400 is unknown, it is likely to mediate binding to the mammalian extracellular matrix or
401 otherwise affect the interactions between bacterial and host cells [33]. TAA genes can be
402 immunodominant antigens, and loss of Vomp expression has previously been observed
403 during recurrent *B. quintana* bacteremia [20]. Loss of function mutations in the mutated *badE*
404 TAA gene may be adaptive in some rodent hosts because they lead to different bacterial
405 localization during infections or avoidance of the host's immune responses.
406 Our results suggest that SSRs acting as contingency loci could have a more prominent
407 role in *Bartonella* adaptation and, more generally, in the evolution of host specificity than
408 previously appreciated. We observed conservation of mononucleotide SSRs of nine or more
409 bases in additional genes in the genomes of *Bartonella* from the Negev Desert dunes. These
410 SSRs are concentrated at the 5′ ends of reading frames, as they are in other bacterial groups in
411 which SSRs within genes function as contingency loci [25,34]. Furthermore, we detected
412 length variation in some SSRs in *Bartonella* isolates from the Negev Desert, supporting the
413 hypothesis that they evolved to mediate rapid ON/OFF switching of gene expression in
414 nature. Some of the conserved and varying SSRs are associated with known virulence factors
415 (e.g., type IV secretion systems and their effectors). Others are associated with conserved
416 *Bartonella* genes of unknown function that are predicted to be membrane-localized or
417 secreted, which could mediate interactions with hosts and/or elicit immune responses.
418 Additional experiments will be needed to understand if the TAA SSRs and SSRs in other
419 genes act as contingency loci and their significance for parasite evolution.

420 Our experiment reproduces some of the host complexity experienced by *Bartonella* in
421 the Negev Desert, but it does not reflect other aspects of the natural environment, including
422 transmission by fleas, co-infection of rodents with multiple *Bartonella* strains and species,
423 and re-infection of rodents with immune systems that have cleared prior infections
424 [8,12,35,36]. Fine-scale longitudinal sampling and genome sequencing of *Bartonella* in wild
425 rodents could provide insights into the importance of SSR mutations under these conditions
426 relative to other mechanisms that can promote evolvability, such as gene transfer agent-
427 mediated recombination between co-infecting strains and species [15] and within-genome
428 recombination in virulence gene arrays [13,14]. *Bartonella* bacteria possess many
429 evolvability mechanisms that may contribute to the remarkable prevalence and diversity of
430 these parasites, underscoring the danger of zoonotic disease through spillover into humans.
431 The fixation of SSR mutations in our experimental *Bartonella* populations contrasts with
432 the findings from an experiment that passaged *Campylobacter jejuni* in a mouse model [37].
433 In that study, no single SSR mutation reached 100% frequency in the evolved bacterial
434 populations. This difference suggests there were more extreme population bottlenecks, less
435 within-host replication, fewer beneficial pathways for SSR-based adaptation to follow, or
436 some combination thereof during the *Bartonella* infections in our experiments. Overall, our
437 findings broaden the typical view of the importance of SSRs in parasite adaptation. Even
438 when the rates of fixation of mutations at these loci cannot keep pace with host immune
439 responses or movement to new hosts, they may provide an advantage for parasites.
440 Importantly, they can do so by generating focused genetic variation that allows for rapid
441 adaptation to a challenging host species in a way that also leaves open the possibility of
442 restoring the inactivated gene if it is needed for success in future host environments.

443 # Methods

444 ## Evolution experiment

445 We serially passaged *B. krasnovii* strain OE-11 A2 (originally isolated from a *G. andersoni*
446 host in the Negev Desert) either through 20 individuals of one of the two host species (*G.*
447 *andersoni* and *G. pyramidum*) or by alternating between them (10 individuals of each host
448 species). Each of the three treatments had five replicate lines. All lines were initiated from
449 separate bacterial colony picks to ensure they did not share any initial genetic diversity. We
450 also included a negative control line that was transmitted from one uninfected host to the next
451 throughout the experiment and negative control hosts that were inoculated in each passage
452 with phosphate-buffered saline (PBS). Neither control showed any evidence of *Bartonella*
453 infection at any point. A full description of the protocol that explains the rationale for specific
454 decisions (e.g., the ancestor strain, inoculation procedure, bacterial quantification methods,
455 inoculation source, and day of sampling) has been published elsewhere [38]. Briefly, in each
456 passage, *Bartonella*-negative rodents were intradermally inoculated with 100 μl of a PBS
457 suspension of bacteria isolated from the previously infected rodent ($5 \times 10^6 \pm 5 \times 10^5$ total
458 cells/inoculum). Rodents were bled by cardiac puncture 15 days after inoculation to collect
459 *Bartonella* at the approximate peak of bacteremia. Red blood cells from each sample were
460 spread on two chocolate agar (CA) plates. These plates were incubated for 3 days at 37°C and
461 5% CO_2. Then, bacterial cells were harvested by scraping the plates and resuspending the
462 cells in PBS. This solution was homogenized by vortexing it with glass beads and passing it
463 through a 5-μm filter. These samples were used to inoculate rodents in the next passage.

464 To assess bacterial loads in the rodent blood 15 days after each inoculation, we extracted
465 DNA from the blood samples using a QIAamp BiOstic Bacteremia DNA Kit (Qiagen),
466 following the manufacturer's instructions. To assess bacterial titers in the PBS suspensions

467 used for inoculation, we extracted DNA using a thermal lysis procedure, as described
468 previously [8]. In each set of extractions, we included a negative control, in which all the
469 reagents were added to PBS but without adding the blood or bacterial suspension. We
470 quantified bacterial loads by real-time quantitative PCR (qPCR), following Eidelman *et al.*
471 [39]. The standard curve was created using DNA extracted from the ancestor strain and was
472 calibrated to the count of colony-forming units of this sample.

473 Genome sequencing

474 Bacterial populations isolated from rodents after passages 2, 4, 6, 7, 8, 10, 12, 14, 16, 18, and
475 20 were preserved from the PBS solution that remained after inoculating the next rodents by
476 centrifugation, resuspending them in 1 ml of 20% (w/v) glycerol in lysogeny broth (LB), and
477 then storing the samples at −80°C. DNA was extracted from these samples using the DNeasy
478 Blood & Tissue Kit (Qiagen), following the protocol for gram-negative bacteria with slight
479 modifications: centrifugation of the initial solution was at 5400 × g for 10 min, and DNA
480 elution was repeated twice in 50 µl of AE buffer that was preheated to 56°C before adding it
481 to the purification column and incubating it for 5 min at room temperature. We also spread
482 100 µl of 10^{-5} and 10^{-6} dilutions of the bacterial solution that we cultivated from rodents after
483 the last passage on CA plates. After 5 days of incubation at 37°C and 5% CO_2, we isolated 10
484 endpoint colonies for each experimental line. Each colony was propagated three times by
485 streaking on a new CA plate and regrowing to ensure a clonal isolate. Following the last
486 regrowth, 15 colonies were scraped from the plate for each clonal isolate and preserved in
487 500 µl of 20% (w/v) glycerol in LB. DNA extractions from these clone samples were
488 performed as described above for the population samples.

489 In total, DNA was extracted from 318 samples: 165 populations (15 lines × 11 time
490 points) and 153 clones (15 lines × 10 clonal isolates per population plus three replicates of the

491 ancestor). Up to 100 ng of purified gDNA from each sample was put into the 2S Turbo DNA
492 Library Kit (Swift Biosciences). All reactions were carried out at 20% of the manufacturer's
493 recommended volumes, with dual 6-bp indexes incorporated during the final 10 cycle PCR
494 step. The resulting libraries were pooled and sequenced on an Illumina HiSeq X Ten
495 instrument by Psomagen (Rockville, MD) to generate 151-base paired-end read data sets
496 (**File S4**). One population sample that was contaminated with DNA from a different
497 *Bartonella* species was discarded before analysis.

498 Mutation identification

499 We removed adaptors from the Illumina reads using *trimmomatic* (v0.39) in paired-end mode
500 with the following settings: 4 allowed mismatches to the seed, a palindrome clip threshold of
501 30, a simple clip threshold of 10, and discarding trimmed reads ≤ 30 bases. To improve
502 mutation calling, we assembled and annotated an updated version of the previously reported
503 genome sequence of *Bartonella krasnovii* strain OE-11 A2 [40]. We used Trycycler (v0.5.3)
504 [41] on nanopore long reads from a prior study (SRA:SRR6873507) [22] to generate a
505 consensus assembly from Flye [42], Miniasm+Minipolish [43,44], and Raven [45]
506 assemblies. Then, we polished the assembly by comparing it to our Illumina short reads for
507 the ancestor strain using *breseq* (v0.36.1) [46] in consensus mode. Discrepancies between
508 reads and the assembly were iteratively corrected by using the gdtools APPLY command and
509 re-running *breseq* until no further improvements were found. Lastly, we used Prokka
510 (v1.14.6) [47] to annotate the final assembly. The updated *B. krasnovii* OE-11 A2 reference
511 sequence includes a chromosome with 2,163,391 base pairs and a plasmid with 29,057 base
512 pairs (**File S5**).

513 To call mutations in evolved clones and populations, we compared trimmed Illumina
514 reads to the reference genome using *breseq* (v0.36.1) [46]. This pipeline identifies point
515 mutations, small insertions and deletions, and structural mutations that can be predicted from

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

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

534 To implement these criteria, we began by keeping all mutations that were predicted at
535 ≥90% frequency in any clone. Then, other predictions were examined for hallmarks of
536 systematic errors of the types described above by analyzing all samples except for the
537 ancestor controls. First, we disregarded mutation predictions in clonal or population samples
538 that did not have at least 10 total reads supporting either the reference or mutated sequence
539 because estimates of their frequencies are subject to large errors. If >5% of all samples did
540 not meet this threshold, that mutation candidate was not considered further. Then, we

541 removed mutations that did not reach a frequency of 20% in at least one population sample.
542 Next, we filtered out mutations that were observed exclusively in the population samples and
543 were predicted to have a frequency of ≥5% in at least one population from 8 or more of the
544 15 different experimental lines. Additionally, we required the summed frequency of a
545 mutation across all clone and population samples to be ≥100% for it to be considered further.
546 Finally, we fit a null Poisson regression model with a uniform rate of generating reads
547 supporting the variant across all population samples. The offset in this model was the total
548 number of reads that were informative about whether that mutation was present or not. For
549 mutations supported by new sequence junctions the offset was corrected for the number of
550 read start position registers that would lead to reads that could be definitively assigned to the
551 junction. We compared this null model to a Poisson regression model that allowed for per-
552 sample variation in the rate of generating reads with a mutation. If the likelihood ratio test
553 comparing these two models was rejected after performing the Benjamini-Hochberg
554 correction for multiple testing across all candidate mutations at a 10^{-6} false-discovery rate
555 (indicating insufficient evidence for a mutation), then it was removed from consideration.
556 Mutation candidates that remained after these steps were manually examined and merged
557 when they were linked by proximity and matching frequency time courses. For example,
558 three nearby base substitutions that all seem to have resulted from one gene conversion event
559 in line L11 were merged. The frequency of the merged mutation is reported as the average of
560 the frequencies predicted for each individual mutation in these cases. Two mutations that
561 passed all criteria peaked at frequencies of ~33% and ~25% in the population samples from
562 lines L11 and L13, respectively, and these mutations were also found at roughly the same
563 frequencies in the endpoint clonal isolates from these lines. These mutations may be in near-
564 identical regions of the genome that were not individually resolved during assembly of the
565 reference sequence. We corrected the frequencies of these mutations by dividing by the

566 estimated peak value and capping the maximum frequency at 100%. For clarity, the
567 frequencies of mutations were also assigned to be 0% in all samples outside of lines in which
568 they were clearly present because they reached a frequency of >20% and did not exhibit
569 sporadic jumps in frequency across population and clone samples. One mutation in L8
570 appeared at 100% frequency in all of this line's population samples and clones, indicating
571 that it was already present in the progenitor cell that grew into the colony that was picked to
572 initiate this population. Because this mutation did not arise during the evolution experiment,
573 it was not counted or analyzed. Of the 27 mutations in the final list, 24 were at a >90%
574 frequency in a clonal sample, 12 were predictions in population samples that passed the
575 filters that eliminated systematic errors, and 9 were in both categories (**File S1**).

576 Infection experiment with *badE* mutant and ancestor strains

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

588 ## Negev Desert *Bartonella* genome assembly and annotation

589 We assessed conservation and diversity in SSRs that may function as contingency loci across
590 *B. krasnovii* OE-11 A2 and 37 additional *Bartonella* isolates classified into different
591 genotypes. These isolates were cultured from fleas and blood sampled from 33 wild *G.*
592 *andersoni* and 33 wild *G. pyramidum* rodents that were captured during October 2016 in two
593 different sites in the northwestern Negev Desert sand dunes in Israel (34°23′E, 30°58′N and
594 34°23′E, 30°55′N) [8]. We performed DNA extractions and Illumina sequencing of these
595 clonal isolates as described above for samples from the evolution experiment. Then, we
596 assembled draft genome sequences using Unicycler (v0.4.9b) [48] on read subsets
597 downsamled to 90× nominal coverage. Contigs that matched the genome of *Acinetobacter*
598 *baylyi* ADP1 (GenBank: NC_005966.1) contaminated the initial 75C-4a isolate assembly due
599 to barcode misassignment. They were identified and removed based on BLAST matches in
600 Bandage (v0.8.1) [49]. Prokka (v1.14.6) [47] was used with the option to include Rfam
601 predictions of noncoding RNAs to create the annotated assemblies that were analyzed. For all
602 comparative analyses, we used the *B. krasnovii* OE1-1 A2 sequence assembled in this same
603 way, rather than the closed assembly we used to analyze the evolution experiments.

604 ## Evolution of *badE* SSRs in prior colony passage experiment

605 We re-analyzed data from a study that passaged a *B. krasnovii* A2-type strain that was
606 isolated from a flea parasitizing a *G. andersoni* rodent in the Negev Desert through single
607 colony transfers on agar [22]. Illumina sequencing reads were downloaded from the NCBI
608 Sequence Read Archive (SRP136159) and trimmed as described above. We assembled the
609 genome of the ancestor of this experiment using Unicycler, as described for the Negev Desert
610 *Bartonella*. Then, we identified the *badE* TAA homolog using a BLAST search and
611 confirmed that it had both ancestral SSRs that mutated during our evolution experiment with

612 *B. krasnovii* OE1-1 A2. Finally, we used *breseq* to call mutations in all nine of the clonal
613 isolates that were sequenced at the end of the experiment after 50 colony passages. No
614 mutations in either SSR were found in any of these samples.

615 Codon randomization tests

616 Custom Python scripts relying on Biopython (v1.81) [57] were used to load protein-coding
617 gene sequences from our closed assembly of the *B. krasnovii* OE1-1 A2 genome (**File S4**)
618 and perform randomization tests that kept the amino acid sequences of each protein in the
619 genome fixed while randomly shuffling codons within each gene. This approach is similar to
620 how prior studies have accounted for codon usage bias affecting SSR prevalence in microbial
621 genomes [24,25]. We report 95% confidence intervals and *p*-values estimated from the
622 characteristics of 10,000 shuffled versions of the very large *badE* TAA gene and 1,000
623 shuffled sets of all proteins in the genome. For comparing overrepresentation of SSRs in
624 *badE* relative to the overall depletion of SSRs genome-wide, we performed the same
625 procedure but added a step where each SSR in a shuffled gene had a 50/206 chance of
626 surviving selection to make it into the final simulated distribution.

627 Orthogroup identification and phylogenetic tree construction

628 To identify orthologous gene groups, we used OrthoFinder (v.2.5.5) [50] on proteins
629 predicted in the genome assemblies of the 38 Negev Desert *Bartonella* isolates and
630 representatives of 8 *Bartonella* species that infect other hosts. The phylogenetic tree was
631 created by aligning the amino acid sequences of the 808 predicted single-gene orthologs with
632 MAFFT (v7.508) [51], trimming these alignments with BMGE (v1.12) [52] using the
633 BLOSUM62 substitution matrix, and inferring a maximum likelihood tree with IQ-Tree
634 (v2.2.3) [53] from the concatenated alignment with 265,987 columns using a JTT+FO+R10
635 model and 100 bootstraps. We used *B. bovis* 91-4 as the outgroup to root the resulting tree.

636 ## TAA gene family sequence analysis

637 We aligned the amino acid sequences of proteins assigned to the orthogroup containing both
638 the TAA gene *PRJBM_RS00735* from *B. henselae* BM1374163 and the *B. krasnovii* OE-
639 11A2 *badE* TAA gene using MUSCLE (v3.8.1) [54], with further manual refinement of the
640 resulting alignment in AliView [55]. To recover incomplete reading frames interrupted by
641 contig boundaries, we also added the best TBLASTN (v2.12.0+) [56] matches in each
642 genome to the first 200 amino acids of the *B. krasnovii* OE-11A2 TAA sequence. To further
643 delineate TAA families, we used InterProScan (v5.65-97.0) [57] to identify all proteins with
644 matches to the head, stalk, and membrane anchor domains of YadA-like trimeric
645 autotransporter adhesins (IPR008640, IPR008635, and IPR005594, respectively) in the 46
646 genomes that were analyzed. We aligned and clustered the sequences of these proteins and
647 examined their placement relative to *B. henselae* BadA and *B. quintana* Vomp TAAs [19,20].
648 For the analysis of conserved amino acid triplets and nine-base subsequences, we used a
649 custom Python script to examine a MUSCLE amino acid alignment of just the 18 *B. krasnovii*
650 *badE* genes, which all have the A_9 SSR that mutated in the evolution experiment.

651 ## Mononucleotide SSR survey

652 Mononucleotide repeats of ≥9 bases in the 46 *Bartonella* genomes were identified and
653 associated with orthogroups using custom Python scripts that used Biopython and pandas
654 [58] functions and R scripts that used tidyverse packages [59]. Orthogroups were assigned
655 initial gene names and descriptions based on annotations in the input genomes, with further
656 refinement through NCBI BLAST [56] and InterProScan [57] searches. We calculated overall
657 statistics for SSRs that overlap protein-coding genes in the Negev Desert *Bartonella* species.
658 Then, we further analyzed orthogroups with the SSRs found in at least two strains of any of
659 the four Negev Desert *Bartonella* species (**Data 3**). We manually examined SSRs in sequence

660 alignments for evidence of frameshifts caused by addition or deletion of the repeated bases
661 using AliView [55]. For the *Bartonella* gene families of unknown function that contained
662 SSRs with length variation, we predicted signal peptides using SignalP-6.0 (v0.0.56) [60] and
663 transmembrane helices using DeepTMHMM (v1.0.24) [61].

664 Data availability

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

668 Code availability

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

672 Acknowledgments

673 We acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas
674 at Austin for providing high-performance computing resources.

675 Competing Interests

676 The authors declare no competing interests.

677 Author Contributions

678 Conceptualization (S.H., L.Z., R.E.L., J.E.B., and H.H), investigation (R.R-P., N.K., N.S.,
679 A.Z.H., D.E.D., R.G., J.E.B., and H.H.), formal analysis (J.E.B. and H.H.), visualization
680 (A.Z.H., J.E.B., and H.H.), software (J.E.B.), writing – original draft (J.E.B. and H.H.),

681 writing – review & editing (R.R-P., N.K., A.Z.H., D.E.D., R.G., S.H., L.Z., R.E.L., J.E.B.,
682 and H.H.) and funding acquisition (S.H., L.Z., R.E.L., J.E.B., and H.H).

683 # References

684 1. Kassen R. The experimental evolution of specialists, generalists, and the maintenance of
685 diversity. J Evol Biol. 2002;15: 173–190. doi:10.1046/j.1420-9101.2002.00377.x
686 2. Cornwall DH, Kubinak JL, Zachary E, Stark DL, Seipel D, Potts WK. Experimental
687 manipulation of population-level MHC diversity controls pathogen virulence evolution
688 in *Mus musculus*. J Evol Biol. 2018;31: 314–322. doi:10.1111/jeb.13225
689 3. Elena SF. Local adaptation of plant viruses: lessons from experimental evolution. Mol
690 Ecol. 2017;26: 1711–1719. doi:10.1111/mec.13836
691 4. Visher E, Boots M. The problem of mediocre generalists: population genetics and eco-
692 evolutionary perspectives on host breadth evolution in pathogens. Proc Biol Sci.
693 2020;287: 20201230. doi:10.1098/rspb.2020.1230
694 5. Betts A, Rafaluk C, King KC. Host and parasite evolution in a tangled bank. Trends
695 Parasitol. 2016;32: 863–873. doi:10.1016/j.pt.2016.08.003
696 6. Cobey S. Pathogen evolution and the immunological niche. Ann N Y Acad Sci.
697 2014;1320: 1–15. doi:10.1111/nyas.12493
698 7. Deitsch KW, Lukehart SA, Stringer JR. Common strategies for antigenic variation by
699 bacterial, fungal and protozoan pathogens. Nat Rev Microbiol. 2009;7: 493–503.
700 doi:10.1038/nrmicro2145
701 8. Gutiérrez R, Cohen C, Flatau R, Marcos-Hadad E, Garrido M, Halle S, et al. Untangling
702 the knots: Co-infection and diversity of *Bartonella* from wild gerbils and their
703 associated fleas. Mol Ecol. 2018;27: 4787–4807. doi:10.1111/mec.14906

9. Kedem H, Cohen C, Messika I, Einav M, Pilosof S, Hawlena H. Multiple effects of host-species diversity on coexisting host-specific and host-opportunistic microbes. Ecology. 2014;95: 1173–1183. doi:10.1890/13-0678.1
10. Rodríguez-Pastor R, Hasik AZ, Knossow N, Bar-Shira E, Shahar N, Gutiérrez R, et al. *Bartonella* infections are prevalent in rodents despite efficient immune responses. Parasit Vectors. 2023;16: 315. doi:10.1186/s13071-023-05918-7
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. 2015;8: 429. doi:10.1186/s13071-015-1029-5
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. 2014;8: 1598–1608. doi:10.1038/ismej.2014.22
13. Engel P, Salzburger W, Liesch M, Chang C-C, Maruyama S, Lanz C, et al. Parallel evolution of a type IV secretion system in radiating lineages of the host-restricted bacterial pathogen *Bartonella*. Guttman DS, editor. PLoS Genet. 2011;7: e1001296. doi:10.1371/journal.pgen.1001296
14. Thibau A, Hipp K, Vaca DJ, Chowdhury S, Malmström J, Saragliadis A, et al. Long-read sequencing reveals genetic adaptation of *Bartonella* adhesin A among different *Bartonella henselae* isolates. Front Microbiol. 2022;13: 838267. doi:10.3389/fmicb.2022.838267
15. Québatte M, Dehio C. *Bartonella* gene transfer agent: Evolution, function, and proposed role in host adaptation. Cell Microbiol. 2019;21: e13068. doi:10.1111/cmi.13068
16. Coenye T, Vandamme P. Characterization of mononucleotide repeats in sequenced prokaryotic genomes. DNA Res. 2005;12: 221–233. doi:10.1093/dnares/dsi009

729 17. Moxon ER, Rainey PB, Nowak MA, Lenski RE. Adaptive evolution of highly mutable
730 loci in pathogenic bacteria. Curr Biol. 1994;4: 24–33.
731 18. Moxon R, Bayliss C, Hood D. Bacterial contingency loci: The role of simple sequence
732 DNA repeats in bacterial adaptation. Annu Rev Genet. 2006;40: 307–333.
733 doi:10.1146/annurev.genet.40.110405.090442
734 19. Riess T, Andersson SGE, Lupas A, Schaller M, Schäfer A, Kyme P, et al. *Bartonella*
735 adhesin A mediates a proangiogenic host cell response. J Exp Med. 2004;200: 1267–
736 1278. doi:10.1084/jem.20040500
737 20. Zhang P, Chomel BB, Schau MK, Goo JS, Droz S, Kelminson KL, et al. A family of
738 variably expressed outer-membrane proteins (Vomp) mediates adhesion and
739 autoaggregation in *Bartonella quintana*. Proc Natl Acad Sci. 2004;101: 13630–13635.
740 doi:10.1073/pnas.0405284101
741 21. Kaiser PO, Riess T, Wagner CL, Linke D, Lupas AN, Schwarz H, et al. The head of
742 *Bartonella* adhesin A is crucial for host cell interaction of *Bartonella henselae*. Cell
743 Microbiol. 2008;10: 2223–2234. doi:10.1111/j.1462-5822.2008.01201.x
744 22. Gutiérrez R, Markus B, Carstens Marques de Sousa K, Marcos-Hadad E,
745 Mugasimangalam RC, Nachum-Biala Y, et al. Prophage-driven genomic structural
746 changes promote *Bartonella* vertical evolution. Genome Biol Evol. 2018;10: 089–3103.
747 doi:10.1093/gbe/evy236
748 23. Field D, Wills C. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*,
749 and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*,
750 result from strong mutation pressures and a variety of selective forces. Proc Natl Acad
751 Sci U S A. 1998;95: 1647–1652. doi:10.1073/pnas.95.4.1647
752 24. Mrázek J, Guo X, Shah A. Simple sequence repeats in prokaryotic genomes. Proc Natl
753 Acad Sci U S A. 2007;104: 8472–8477.

754 25. Lin W-H, Kussell E. Evolutionary pressures on simple sequence repeats in prokaryotic
755 coding regions. Nucleic Acids Res. 2012;40: 2399–2413. doi:10.1093/nar/gkr1078
756 26. Chirhart-Gilleland RL, Kovach ME, Elzer PH, Jennings SR, Roop RM. Identification and
757 characterization of a 14-kilodalton *Brucella abortus* protein reactive with antibodies
758 from naturally and experimentally infected hosts and T lymphocytes from
759 experimentally infected BALB/c mice. Infect Immun. 1998;66: 4000–4003.
760 doi:10.1128/IAI.66.8.4000-4003.1998
761 27. Dehio C. Infection-associated type IV secretion systems of *Bartonella* and their diverse
762 roles in host cell interaction. Cell Microbiol. 2008;10: 1591–1598. doi:10.1111/j.1462-
763 5822.2008.01171.x
764 28. Cooper TF, Lenski RE. Experimental evolution with *E. coli* in diverse resource
765 environments. I. Fluctuating environments promote divergence of replicate populations.
766 BMC Evol Biol. 2010;10: 11–11.
767 29. Sandberg TE, Lloyd CJ, Palsson BO, Feist AM. Laboratory evolution to alternating
768 substrate environments yields distinct phenotypic and genetic adaptive strategies. Appl
769 Environ Microbiol. 2017;83: e00410-17. doi:10.1128/AEM.00410-17
770 30. Deatherage DE, Kepner JL, Bennett AF, Lenski RE, Barrick JE. Specificity of genome
771 evolution in experimental populations of *Escherichia coli* evolved at different
772 temperatures. Proc Natl Acad Sci U S A. 2017;114: E1904–E1912.
773 doi:10.1073/pnas.1616132114
774 31. Palmer ME, Lipsitch M, Moxon ER, Bayliss CD. Broad conditions favor the evolution of
775 phase-variable loci. Kolter R, editor. mBio. 2013;4: e00430-12.
776 doi:10.1128/mBio.00430-12

32. Messika I, Garrido M, Kedem H, China V, Gavish Y, Dong Q, et al. From endosymbionts to host communities: factors determining the reproductive success of arthropod vectors. Oecologia. 2017;184: 859–871. doi:10.1007/s00442-017-3906-4
33. Kiessling AR, Malik A, Goldman A. Recent advances in the understanding of trimeric autotransporter adhesins. Med Microbiol Immunol (Berl). 2020;209: 233–242. doi:10.1007/s00430-019-00652-3
34. Orsi RH, Bowen BM, Wiedmann M. Homopolymeric tracts represent a general regulatory mechanism in prokaryotes. BMC Genomics. 2010;11: 102–102.
35. Morick D, Krasnov BR, Khokhlova IS, Gottlieb Y, Harrus S. Investigation of *Bartonella* acquisition and transmission in *Xenopsylla ramesis* fleas (Siphonaptera: Pulicidae). Mol Ecol. 2011;20: 2864–2870. doi:10.1111/j.1365-294X.2011.05033.x
36. 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. 2015;9: 1662–1676. doi:10.1038/ismej.2014.255
37. Jerome JP, Bell JA, Plovanich-Jones AE, Barrick JE, Brown CT, Mansfield LS. Standing genetic variation in contingency loci drives the rapid adaptation of *Campylobacter jejuni* to a novel host. PLoS ONE. 2011;6: e16399.
38. Rodríguez-Pastor R, Shafran Y, Knossow N, Gutiérrez R, Harrus S, Zaman L, et al. A road map for in vivo evolution experiments with blood-borne parasitic microbes. Mol Ecol Resour. 2022;22: 2843–2859. doi:10.1111/1755-0998.13649
39. Eidelman A, Cohen C, Navarro-Castilla Á, Filler S, Gutiérrez R, Bar-Shira E, et al. The dynamics between limited-term and lifelong coinfecting bacterial parasites in wild rodent hosts. J Exp Biol. 2019;222: jeb.203562. doi:10.1242/jeb.203562
40. Gutiérrez R, Shalit T, Markus B, Yuan C, Nachum-Biala Y, Elad D, et al. *Bartonella kosoyi* sp. nov. and *Bartonella krasnovii* sp. nov., two novel species closely related to

802 the zoonotic *Bartonella elizabethae*, isolated from black rats and wild desert rodent-
803 fleas. Int J Syst Evol Microbiol. 2020;70: 1656–1665. doi:10.1099/ijsem.0.003952
804 41. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, et al. Trycycler:
805 consensus long-read assemblies for bacterial genomes. Genome Biol. 2021;22: 266.
806 doi:10.1186/s13059-021-02483-z
807 42. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using
808 repeat graphs. Nat Biotechnol. 2019;37: 540–546. doi:10.1038/s41587-019-0072-8
809 43. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long
810 sequences. Bioinformatics. 2016;32: 2103–2110. doi:10.1093/bioinformatics/btw152
811 44. Wick RR, Holt KE. Benchmarking of long-read assemblers for prokaryote whole genome
812 sequencing. F1000Research. 2019;8: 2138.
813 45. Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. Nat
814 Comput Sci. 2021;1: 332–336. doi:10.1038/s43588-021-00073-4
815 46. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes
816 from next-generation sequencing data using *breseq*. Sun L, Shou W, editors. Methods
817 Mol Biol. 2014;1151: 165–188. doi:10.1007/978-1-4939-0554-6_12
818 47. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:
819 2068–2069. doi:10.1093/bioinformatics/btu153
820 48. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome
821 assemblies from short and long sequencing reads. Phillippy AM, editor. PLOS Comput
822 Biol. 2017;13: e1005595. doi:10.1371/journal.pcbi.1005595
823 49. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo
824 genome assemblies. Bioinformatics. 2015;31: 3350–3352.
825 doi:10.1093/bioinformatics/btv383

826 50. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative
827 genomics. Genome Biol. 2019;20: 238. doi:10.1186/s13059-019-1832-y
828 51. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
829 Improvements in performance and usability. Mol Biol Evol. 2013;30: 772–780.
830 doi:10.1093/molbev/mst010
831 52. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new
832 software for selection of phylogenetic informative regions from multiple sequence
833 alignments. BMC Evol Biol. 2010;10: 210. doi:10.1186/1471-2148-10-210
834 53. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et
835 al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the
836 genomic era. Mol Biol Evol. 2020;37: 1530–1534. doi:10.1093/molbev/msaa015
837 54. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and
838 space complexity. BMC Bioinformatics. 2004;5: 113. doi:10.1186/1471-2105-5-113
839 55. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large
840 datasets. Bioinforma Oxf Engl. 2014;30: 3276–3278. doi:10.1093/bioinformatics/btu531
841 56. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+:
842 architecture and applications. BMC Bioinformatics. 2009;10: 421–421.
843 doi:10.1186/1471-2105-10-421
844 57. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5:
845 genome-scale protein function classification. Bioinformatics. 2014;30: 1236–1240.
846 doi:10.1093/bioinformatics/btu031
847 58. McKinney W. Data structures for statistical computing in Python. Proc of the 9th Python
848 in Science Conf. Austin, Texas; 2010. pp. 56–61. doi:10.25080/Majora-92bf1922-00a
849 59. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to
850 the Tidyverse. J Open Source Softw. 2019;4: 1686. doi:10.21105/joss.01686

60. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40: 1023–1025. doi:10.1038/s41587-021-01156-3

61. Hallgren J, Tsirigos K, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H, et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural network. 2022. doi:10.1101/2022.04.08.487609

Supporting Information

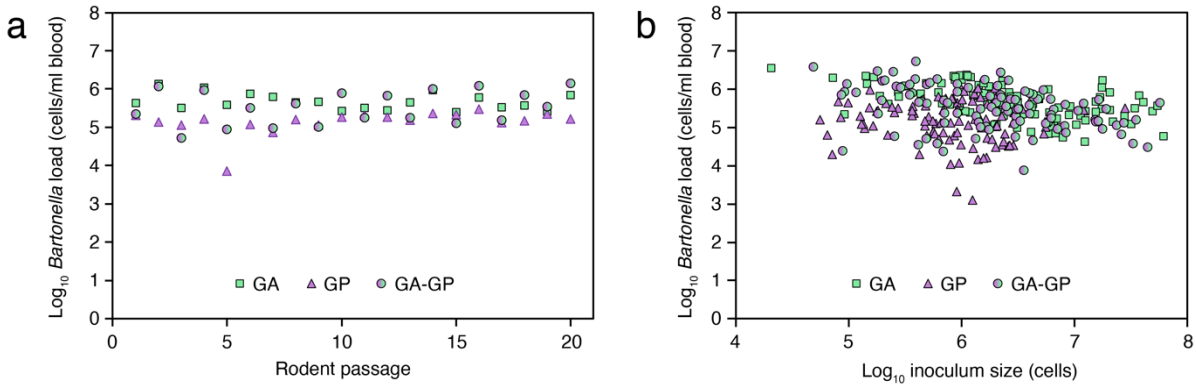


Fig. S1. *Bartonella* infection loads and inoculum sizes throughout the evolution experiment. We propagated *Bartonella* krasnovii OE1-1 A2 populations through rodents in three different host scenarios: infecting *Gerbillus andersoni* only (GA), infecting *G. pyramidum* only (GP), or alternating between the two hosts (GA-GP). (**a**) Mean of the log-transformed final *Bartonella* load in rodents from each host treatment at day 15 post infection, when blood was collected to culture bacteria for the next passage. (**b**) Log-transformed *Bartonella* loads in all rodents from every passage plotted against the log-transformed size of the *Bartonella* inoculum injected into that rodent.

S1 File. Evolution experiment mutations. Mutations identified in 162 population samples and 147 endpoint clonal isolates from whole-genome sequencing data.

870

871 **S2 File. *Bartonella badE* gene sequences.** Alignment of the *B. krasnovii* OE1-1 A2 *badE*

872 TAA gene and the most similar genes in 45 related *Bartonella* genomes.

873

874 **S3 File. Negev Desert *Bartonella* mononucleotide SSRs.** Characteristics of conserved SSRs

875 found in 38 wild *Bartonella* isolates classified into four species.

876

877 **S4 File. Genome sequencing data summary.** Characteristics of whole-genome sequencing

878 data sets for the ancestor, 162 population samples, and 147 endpoint clonal isolates.

879

880 **S5 File. *Bartonella* OE1-1 A2 reference genome.** Assembled and annotated reference

881 genome used to predict mutations in evolved populations and clones.