Strong within-host selection in a maternally inherited obligate symbiont: *Buchnera* and aphids

Julie Perreau[a,1], Bo Zhang[a,b,1], Gerald P. Maeda[a], Mark Kirkpatrick[a], and Nancy A. Moran[a,2]

[a]Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712; and [b]Institute of Plant Protection, Chinese Academy of Agricultural Sciences, 100193 Beijing, China

Edited by Michael Lynch, Arizona State University, Tempe, AZ, and approved July 15, 2021 (received for review February 20, 2021)

Numerous animal lineages have maternally inherited symbionts that are required for host reproduction and growth. Endosymbionts also pose a risk to their hosts because of the mutational decay of their genomes through genetic drift or to selfish mutations that favor symbiont fitness over host fitness. One model for heritable endosymbiosis is the association of aphids with their obligate bacterial symbiont, *Buchnera*. We experimentally established heteroplasmic pea aphid matrilines containing pairs of closely related *Buchnera* haplotypes and used deep sequencing of diagnostic markers to measure haplotype frequencies in successive host generations. These frequencies were used to estimate the effective population size of *Buchnera* within hosts (i.e., the transmission bottleneck size) and the extent of within-host selection. The within-host effective population size was in the range of 10 to 20, indicating a strong potential for genetic drift and fixation of deleterious mutations. Remarkably, closely related haplotypes were subject to strong within-host selection, with selection coefficients as high as 0.5 per aphid generation. In one case, the direction of selection depended on the thermal environment and went in the same direction as between-host selection. In another, a new mutant haplotype had a strong within-host advantage under both environments but had no discernible effect on host-level fitness under laboratory conditions. Thus, within-host selection can be strong, resulting in a rapid fixation of mutations with little impact on host-level fitness. Together, these results show that within-host selection can drive evolution of an obligate symbiont, accelerating sequence evolution.

endosymbiosis | within-host selection | levels of selection | mutualism | maternal transmission

Eukaryotic hosts often depend on symbiotic microorganisms for survival, growth, and reproduction. The most ancient of such symbionts, the mitochondria, are central to metabolism of almost all eukaryotic cells. A myriad of other bacterial symbionts have evolved intimate partnerships with protozoans, fungi, plants, and animals. Obligate symbionts, including mitochondria and many others, are often maternally transmitted to offspring, a feature that ensures the short-term benefit of reliable inoculation of progeny but incurs the long-term drawback of enforcing strict clonality and small genetic population size on symbionts, leading to increased fixation of deleterious mutations (1, 2). In the long term, exclusively uniparental transmission results in degraded symbiont genomes featuring fewer intact genes and reduced thermal stability of remaining gene products (1, 3). Furthermore, selection on symbiont populations within individual hosts potentially leads to the spread of selfish symbiont mutations that benefit symbionts but reduce host fitness (4). In human mitochondria, for example, mutations favored at the organelle level can increase in frequency in the mother's germline, causing deleterious effects in progeny (5).

The spread of symbiont alleles harmful to hosts, whether broadly deleterious or selfish, will be countered by selection on individual hosts. The effectiveness of host-level selection depends on several parameters. First, the host effective population size (N_{HOST}) affects efficacy of host-level selection (and for maternally transmitted symbionts only the female population size is relevant). Multicellular hosts (i.e., animals and plants) typically have small population sizes compared to bacteria, so asexual genomic components, such as maternally transmitted symbionts or Y chromosomes, will usually accumulate deleterious mutations (6, 7).

A second factor that influences the effectiveness of host-level selection on symbiont alleles is the size of the genetic bottleneck (N_{SYM}) imposed by the transmission cycle (8–10). N_{SYM} reflects both the size of the inoculum transferred from mother to daughter and also the packaging of symbionts into host cells during development. A large N_{SYM} enables more effective selection on symbionts within hosts, with contradictory effects depending on the type of symbiont mutations: it will enhance purifying selection against mutations deleterious to both symbiont and host but will promote the spread of selfish mutations that harm hosts. Examples of such mutations include any that increase symbiont replication rate and deplete host resources.

The symbiosis of the pea aphid, *Acyrthosiphon pisum*, with its mutualistic bacterial symbiont *Buchnera aphidicola* is a model for maternally transmitted obligate symbiosis (11). *Buchnera* has a small genome encoding only 583 genes but retains pathways for synthesis of essential amino acids needed by hosts (12). *Buchnera* is packaged within specialized aphid cells, bacteriocytes, and is transmitted to progeny within the mother's body. Most aphid reproduction is asexual and viviparous, with female embryos developing prenatally and born in order as they descend the maternal ovariole. *Buchnera* colonizes through exocytosis from a maternal bacteriocyte and

Significance

Many animals depend on maternally transmitted symbiotic bacteria that provide nutrients or other benefits. The evolution of these symbionts is complicated: natural selection can act on hosts, favoring symbionts that increase host reproduction, or on symbionts, favoring symbionts that spread within hosts. Furthermore, transmission bottlenecks can facilitate the spread of mutations deleterious to both. By measuring changes in frequencies of symbiont genotypes within individual insect matrilines, we estimated within-host selection and transmission bottleneck size. Results revealed surprisingly strong selection, with some symbiont genotypes more successful in colonizing progeny, as well as a severe transmission bottleneck, consistent with observations of deleterious mutation accumulation in symbiont genomes. Findings elucidate the forces driving evolution of heritable symbionts and generating their distinctive genomic features.

Author contributions: J.P., B.Z., G.P.M., M.K., and N.A.M. designed research; J.P., B.Z., G.P.M., M.K., and N.A.M. performed research; J.P., B.Z., G.P.M., M.K., and N.A.M. analyzed data; and J.P., B.Z., G.P.M., M.K., and N.A.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]J.P. and B.Z. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: nancy.moran@austin.utexas.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2102467118/-/DCSupplemental.

Published August 24, 2021.

endocytosis by the syncytial cell of a nearby early-stage embryo (13). *Buchnera* replicates within the embryo where it is packaged into about 80 bacteriocytes (14). Each embryo is colonized by symbionts from one, or possibly two, maternal bacteriocyte(s), so the inoculum is not randomly drawn from the entire maternal population (13). Thus, N_{SYM} is a function of the number of *Buchnera* cells transferred and also of the packaging that occurs during development.

In this study, we address three questions regarding *Buchnera* evolution within host matrilines: 1) Does within-host selection occur? 2) What is the effective bottleneck size (N_{SYM}), imposed by a transmission cycle within a matriline? 3) Does within-host selection oppose or support selection at the host level; that is, do symbionts evolve to be selfish, neutral, or beneficial? To address these questions, we generated *A. pisum* clones heteroplasmic for closely related *Buchnera* haplotypes using a modification of a previously established microinjection method (15). We then measured haplotype frequencies over five generations in mothers and daughters sampled at the same developmental stage using deep sequencing of a diagnostic nucleotide difference to obtain precise estimates of frequencies. We measured intergenerational changes in frequencies in order to estimate N_{SYM} and s_{SYM}, the magnitude of within-host selection on symbionts. We demonstrate that within-host selection can be surprisingly strong. Drift due to transmission bottlenecks is also substantial, as N_{SYM} is far smaller than the number of *Buchnera* cells transferred to daughters. Finally, in order to address whether a mutation strongly favored within hosts affects fitness of host individuals, we assessed selection on aphids fixed for *Buchnera* haplotypes that differed strongly in within-host fitness.

Results

Visualizing Haplotype Frequencies within Individual Aphids over Generations. Using microinjection, we established aphid matrilines with intermediate frequencies of two *Buchnera* haplotypes, then sampled successive female generations at the same developmental stage (adults at initiation of reproduction) in the absence of host-level selection (Fig. 1). The *Buchnera* haplotypes were near identical (*SI Appendix*, Fig. S1 and Table S1). We performed two experiments corresponding to two haplotype pairs, all within a single background aphid genotype. One experiment included *Buchnera* genotypes (5AY and LSR1) that differ at 19 single base indels or changes (12 intergenic or silent) including a single base difference known to sharply affect thermal tolerance. A second experiment included *Buchnera* genotypes (5A and LSR1) that differ at 17 locations (10 intergenic or silent) not predicted to affect fitness. Both experiments were carried out under two thermal environments: constant cool conditions (20 °C) or daily 4-h exposure to heat (33 °C). We obtained accurate estimates of haplotype frequencies in each aphid using high-throughput sequencing of amplicons across a diagnostic polymorphic site. In all, we sampled 732 aphids, which gave 165, 165, 182, and 175 mother–daughter pairs for the four combinations of *Buchnera* haplotypes and environments.

Haplotype frequencies for each aphid are presented in two formats: changes in frequencies within individual matrilines over five aphid generations (Fig. 2) and haplotype frequencies in mothers versus their daughters for each mother–daughter pair (Fig. 3). Visual inspection suggests strong within-host selection for both haplotype combinations and for both temperature conditions. As expected, *Buchnera* 5AY increases under heat exposure, where it approaches fixation in all lines after five generations and has a disadvantage at constant 20 °C, although all matrilines are still polymorphic at the end of the experiment (Fig. 2*A*). For the pair for which no difference in fitness was predicted, we found instead that *Buchnera* 5A increases relative to *Buchnera* LSR1 under both environments, with daughters having higher *Buchnera* 5A frequencies than their mothers (Fig. 3*B*). *Buchnera* 5A is fixed or nearly fixed after five generations in every matriline and at both temperatures (Fig. 2*B*).

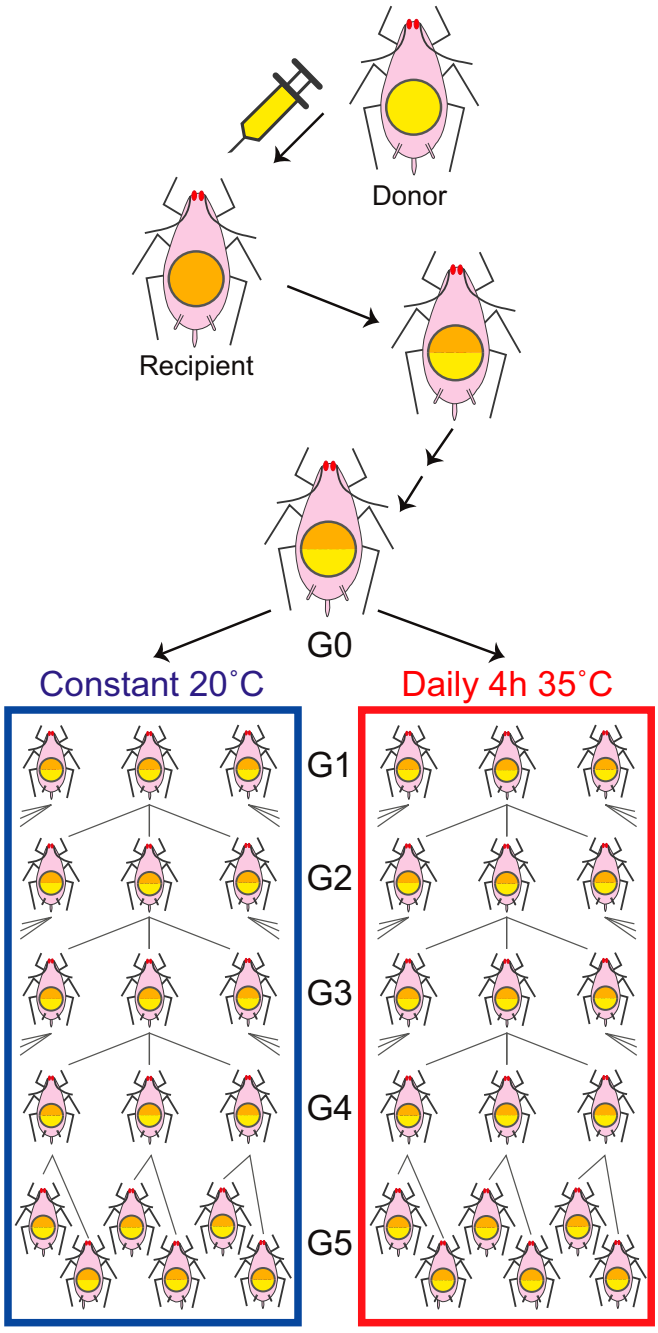

Fig. 1. Experimental design for measuring within-host selection and drift of *Buchnera*. A single female was injected with *Buchnera* from a donor line to establish a clonal aphid line heteroplasmic for two *Buchnera* haplotypes. After three to four generations, a founder female (G0) was used to initiate experiments in which successive female generations were allowed to reach adulthood and reproduce for 1 d and then harvested to estimate *Buchnera* haplotype frequencies. Three progeny were allowed to develop in each generation, and all aphids were sampled 1 d after beginning to reproduce. Experiments were performed under two thermal environments: constant 20 °C ("Cool") and constant 20 °C interrupted daily by 4 h of exposure to 33 °C ("Heat").

To understand the dynamics of haplotype frequencies within hosts, we estimated the effective bottleneck size (N_{SYM}) and the selection coefficient for the introduced haplotype (s_{SYM}) using a maximum likelihood approach. The estimated parameters summarize the net effects of drift and selection through a full generation cycle, with N_{SYM} and s_{SYM} assumed constant. For example, s_{SYM} is

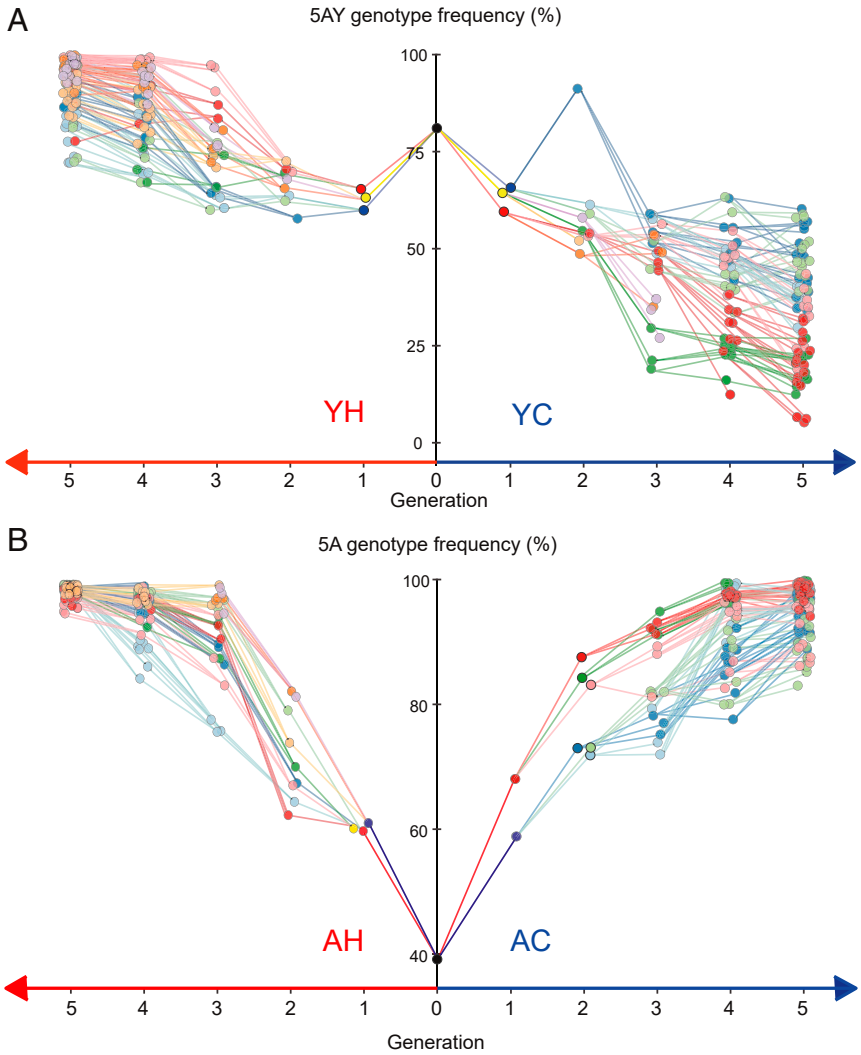

Fig. 2. *Buchnera* haplotype frequencies within individual mothers and daughters, over five generations. (*A*) Frequencies starting with a single female heteroplasmic for *Buchnera* haplotypes 5AY and LSR1. (*B*) Frequencies in aphids starting with a single female heteroplasmic for *Buchnera* haplotypes 5A and LSR1. (*A* and *B*, *Right*) Lines kept at constant 20 °C ("Cool" treatment), and (*Left*) lines subjected daily to 4 h at 33 °C ("Heat" treatment). Lines connect the mother–daughter pairs and are colored to enable visualization of individual matrilines within experiments. For Generations 0 to 3, each mother produced three sampled daughters, each of which became mothers for the next generation. Generation 4 mothers each produced two daughters, and Generation 5 was the last one sampled. In some cases, an aphid died, ending that subline prematurely. The overall death rate was 8.0%.

assumed to be frequency independent. We identified the values of N_{SYM} and s_{SYM} that maximize the likelihood of our observed frequencies and estimated the 95% confidence regions for both parameters (Fig. 4).

Within-Host Selection: Near-Identical *Buchnera* Haplotypes Differing in Thermal Tolerances. As expected, *Buchnera* 5AY has a clear advantage within matrilines when aphids are exposed to heat as juveniles: the maximum likelihood estimate for $s_{SYM} = 0.54$ (95% CI: ~0.41 to 0.67) (Fig. 4). This result is consistent with previous evidence that the responsive *ibpA* heat-shock promoter, present in *Buchnera* 5AY but absent in *Buchnera* LSR1, improves survivorship of *Buchnera* within aphid hosts following heat exposure (15, 16). In contrast, *Buchnera* LSR1 has the advantage under constant 20 °C, with the highest likelihood $s_{SYM} = -0.36$ (95% CI: ~ −0.30 to −0.41). Previously, the nonresponsive *Buchnera ibpA* promoter was shown to benefit aphid fitness under constant 20 °C (16). Our results suggest that the host-level advantage at each temperature reflects better survivorship of *Buchnera* corresponding to that condition. Thus, for this haplotype combination, within-host selection acts in the same direction as between-host selection but in opposite directions depending on temperature.

Unexpectedly Strong Within-Host Selection on Near-Identical *Buchnera* Haplotypes Expected to Be Neutral. Unexpectedly, *Buchnera* 5A had a strong selective advantage, under both the cool and heat environments. We estimated s_{SYM} for *Buchnera* 5A versus *Buchnera* LSR1 at 0.45 (95% CI: ~0.23 to 0.61) and 0.37 (95% CI: ~0.33 to 0.54) for the two environments (Fig. 4). To verify that we had a complete catalog of genetic differences, we resequenced representative lines at the end of the experiment and were surprised to discover four new mutations in the *Buchnera* 5A main chromosome and two mutations on the plasmid encoding *trpEG*. A haplotype with these new mutations had become fixed by the end of the experiment; we named the derived line $5A^E$. Chromosomal mutations included a 13-nucleotide duplication within an intergenic spacer plus three single-nucleotide substitutions (one intergenic, one synonymous, and one nonsynonymous); plasmid mutations included one synonymous and one single base deletion in an intergenic spacer (*SI Appendix*, Table S1). We designed PCR primers to investigate the

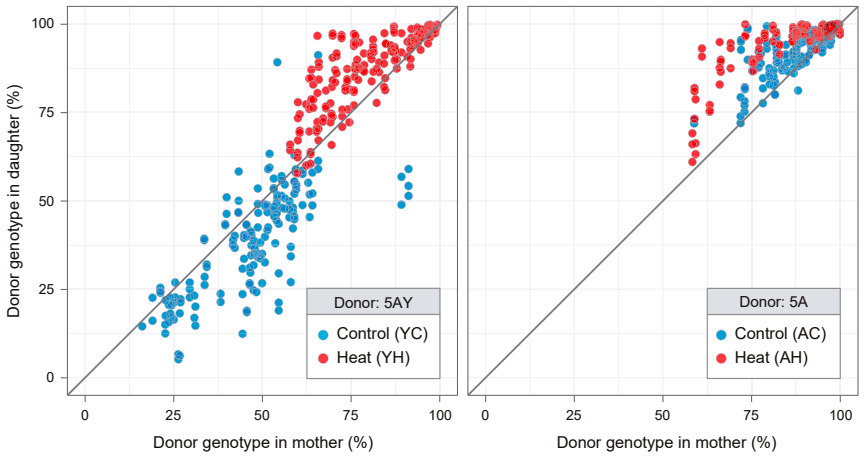

Fig. 3. *Buchnera* haplotype frequencies in mothers and their daughters. Frequencies in mothers are compared to frequencies in daughters for (*A*) 5AY-LSR1 heteroplasmic lines and (*B*) 5A-LSR1 heteroplasmic lines. The diagonal line corresponds to equal frequencies in mother and daughter, sampled at the same age. Samples for 5AY-Heat, 5AY-Cool, 5A-Heat, and 5A-Cool included 165, 165, 182, and 175 mother–daughter pairs, respectively.

time of origin of the new mutations and discovered that the heteroplasmic female initiating the experiment (Generation 0) contained a mixture of *Buchnera* 5A and *Buchnera* $5A^E$, along with *Buchnera* LSR1. *Buchnera* $5A^E$ increased in frequency over time, such that both the original *Buchnera* 5A and *Buchnera* LSR1 were eliminated or nearly eliminated by Generation 5. Although our diagnostic polymorphism for distinguishing *Buchnera* 5A from *Buchnera* LSR1 did not distinguish 5A from $5A^E$, the fitness advantage must have been one of the new mutations in *Buchnera* $5A^E$, which eliminated or nearly eliminated the original *Buchnera* 5A haplotype.

Our model for estimating N_{SYM} and s_{SYM} assumes that these parameters are constant during our experiment. However, for the *Buchnera* 5A/*Buchnera* LSR1 experiment, this assumption is likely violated since *Buchnera* 5A was itself a mixture of haplotypes (5A and $5A^E$) of changing frequencies. Clearly, selection strongly favors *Buchnera* $5A^E$, but the maximum likelihood estimates of s_{SYM} and N_{SYM} may be inaccurate for the *Buchnera* 5A-*Buchnera* LSR1 combination.

Measuring the Transmission Genetic Bottleneck (N_{SYM}). We expected N_{SYM} to be similar across experimental trials, reflecting mechanisms for symbiont packaging and transmission in this aphid species. Indeed, N_{SYM} estimates fell into a fairly narrow range for most trials, with maximum likelihood estimates of 15 and 20 for *Buchnera* 5AY under the two conditions and 37 and 13 for *Buchnera* 5A under the two conditions (Fig. 4). The values for the *Buchnera* 5A combination are less reliable since the magnitude of selection on the *Buchnera* 5A+$5A^E$ combined haplotype changed over time as the $5A^E$ haplotype increased in frequency. In contrast, no new mutations occurred in the *Buchnera* 5AY trials (in either haplotype), so the assumption of a constant selective coefficient is reasonable. Thus, the value of N_{SYM} appears to be in the range of 10 to 20. Effectively, this means that the level of drift is equivalent to drawing 10 to 20 *Buchnera* genomes at random from the maternal population for inoculation of each progeny.

Between-Host Selection on *Buchnera* Haplotypes Differing in Within-Host Fitness. Because *Buchnera* $5A^E$ had a strong within-host fitness advantage under both thermal environments, we wondered whether this haplotype was beneficial or neutral at the host level or whether it was selfish (i.e., harmful to hosts). We performed experiments to measure whether aphid fitness differed between aphids with the same host genotype (*A. pisum* LSR1) but fixed for the two *Buchnera* haplotypes ($5A^E$ versus LSR1). First, we compared three fitness-related parameters—time from birth to first reproduction, adult mass, and fecundity—for the two matrilines. Aphids fixed for *Buchnera* $5A^E$ developed faster, with average first reproduction of 9.5 versus 10.2 d for aphids fixed for *Buchnera* LSR1 ($P < 0.001$, analysis of covariance (ANCOVA); *SI Appendix*, Fig. S2). They were also heavier at an average of 4.34 mg as compared to 3.98 mg ($P < 0.001$, ANCOVA; *SI Appendix*, Fig. S2). However, greater body mass did not correspond to greater fecundity, measured as the number of progeny in first 7 d of reproduction; in fact, the trend was in the opposite direction [i.e., higher fecundity for the line bearing *Buchnera* LSR1 ($P = 0.077$ by ANCOVA) (*SI Appendix*, Fig. S2)]. These measures of aphid fitness parameters showed considerable variation.

To obtain a direct measure of selection on aphid lines fixed for the two *Buchnera* haplotypes, we allowed the lines to compete

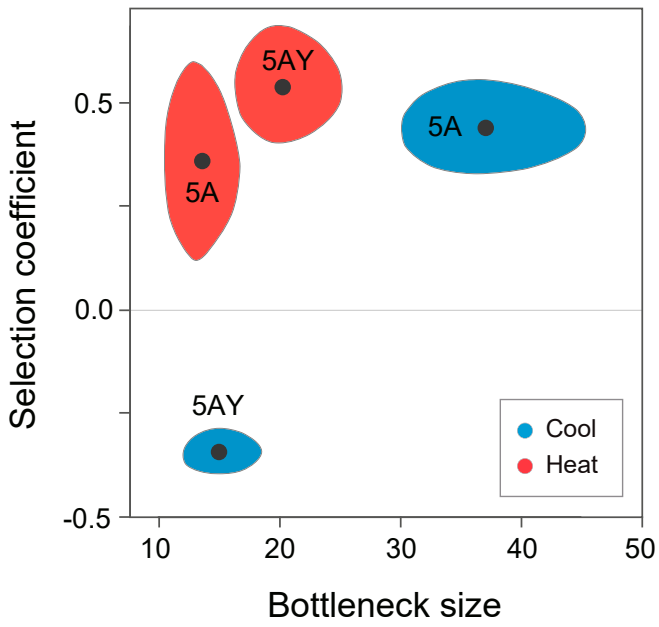

Fig. 4. Bottleneck size and selection coefficient estimates. The black dots are maximum likelihood estimates of N_{SYM} and s_{SYM} for the two combinations of *Buchnera* haplotypes and the two temperature treatments. The shaded areas correspond to the 95% confidence regions for the estimates.

directly in larger cages, transferring a large number of randomly sampled, mixed-age aphids to new plants weekly to avoid population bottlenecks. This experiment to estimate s_{HOST} paralleled those for within-host selection to measure s_{SYM}, but the selection interval was 1 wk rather than one aphid generation (Fig. 5*A*). We used weekly pooled aphid samples to estimate frequencies of the two *Buchnera* haplotypes in the sample and verified this approach with a standard curve based on known ratios of aphids from the two lines (*SI Appendix*, Fig. S3). We thus obtained frequencies in weekly population samples for the 18 replicate competition experiments (*SI Appendix*, Fig. S4) and used these to calculate pairwise frequencies for successive weeks for each replicate, over 12 wk. We used *SI Appendix*, Eq. 1 to calculate the selection coefficient at the host level, s_{HOST}, as applicable to the weekly sampling time. We obtained a maximum likelihood $s_{HOST} = 0.0016$ (95% CI: ~−0.061 to 0.063). Thus, the *Buchnera* $5A^E$ haplotype appears neutral or near neutral at the host level, despite its strong within-host advantage over *Buchnera* LSR1.

Recombination. We resequenced fixed haplotypes at the end of the experiments and found no recombinants. Thus, even when *Buchnera* haplotypes reside in the same host, they do not recombine, at least not often enough to appear in our experiments. The polymorphisms separating the haplotypes are scattered around the *Buchnera* chromosome, so any longer-range recombinants would likely have been detected.

Discussion

Within-Host Selection and Consequences for Host Fitness. The most surprising result from these experiments was the occurrence of very strong within-host selection in every trial: some *Buchnera* haplotypes are much better able than others to compete within heteroplasmic mothers for transmission to the next generation. In the case of the 5AY–LSR1 heteroplasmy (Figs. 2 and 3), within-host selection is in the same direction as between-host selection, and the direction of selection depends on the thermal environment. Previous studies, based on *Buchnera* haplotypes in their native hosts, showed that the *ibpA* promoter mutation had a strong effect on the heat-shock response of IbpA, a small heat-shock protein, and that the single base deletion increased or decreased aphid survival, developmental time, and fecundity depending on heat exposure (16, 17). We found that haplotypes that differed in this mutation were selected within hosts and in the same direction as between hosts in each environment. In this case, within-host selection benefits hosts, enabling each mother to produce daughters with fitness higher than her own, provided they live under the same thermal conditions. Since hosts require *Buchnera* for nutrient provisioning and normal development, it is logical that some mutations that affect *Buchnera* fitness within hosts would affect aphid fitness in the same direction. This within-host selection is similar to germline selection within individuals, in the sense that the large number of genes needed for basic cellular functions undergo purifying selection during replication within hosts (18). Such germline selection also appears to operate on human mitochondrial haplotypes in heteroplasmic human mothers (5).

In the case of the *Buchnera* $5A^E$–5A–LSR1 heteroplasmy of the second experiment, we expected no selection, given the near identity of the haplotypes (*SI Appendix*, Table S1). Surprisingly, we found strong within-host selection favoring *Buchnera* $5A^E$ over both *Buchnera* 5A and *Buchnera* LSR1, under both temperature conditions. Using a long-term selection experiment on aphids fixed for alternative *Buchnera* haplotypes, we found that the $5A^E$ haplotype was neutral or close to neutral at the host level.

The absence of a host-level cost raises the question of why the advantageous mutation (or combination of mutations) of *Buchnera* $5A^E$ has not arisen and become fixed through within-host selection in natural populations. A possible explanation is that neutrality at the host level reflects the favorable and invariant conditions of our laboratory selection experiment: these include a constant supply of nutrient-rich seedling host plants, absence of natural enemies, and absence of some life stages (winged females or sexual generations). The *Buchnera* $5A^E$ haplotype might be deleterious to hosts during nutrient limitation or during the sexual stages.

Our experiments did not identify the nature of the advantage of *Buchnera* $5A^E$ over *Buchnera* LSR1 within hosts. Potentially, *Buchnera* $5A^E$ replicates faster during some or all stages of host development; alternatively, *Buchnera* $5A^E$ may have some advantage in colonizing embryos. None of the five observed mutations of *Buchnera* $5A^E$ (*SI Appendix*, Table S1) has an obvious functional implication. One speculation is that one of the two mutations on the Trp plasmid reduces tryptophan provisioning, a host-beneficial contribution of *Buchnera* to hosts; the energy savings might thereby enable these *Buchnera* cells to replicate faster.

Buchnera $5A^E$ arose from *Buchnera* 5A at the beginning of our second selection experiment and featured six mutations in the founding females. This sudden appearance of multiple mutations is curious, as previous deep sequencing of *Buchnera* genomes of lines descending from a single female and maintained separately in the laboratory revealed only two single base substitutions during 14 y of evolution in the laboratory (19). Our resequencing of *Buchnera* haplotypes is otherwise consistent with this low rate

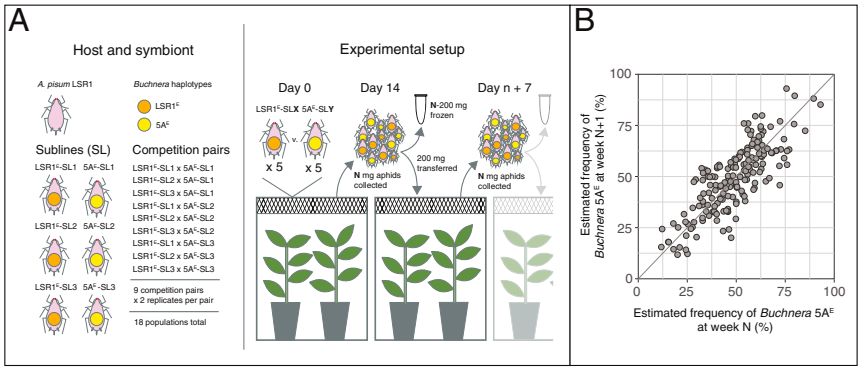

Fig. 5. Measuring host-level selection on *Buchnera* haplotypes $5A^E$ and LSR1 in a common aphid genotype background. (*A*) Experimental design included first growing each of the two lines in sublines to control for maternal effects due to variable rearing conditions. Aphids fixed for alternative *Buchnera* haplotypes were then competed in cages with plants, sampled and transferred weekly for 12 wk, in 18 replicate populations. (*B*) *Buchnera* haplotype frequencies in pooled population samples in successive weeks. The diagonal line corresponds to equal frequencies in successive samples. $n = 204$ population samples.

[e.g., three base substitutions occurred during 18 y in the lineage leading to *Buchnera* 5AY (*SI Appendix*, Fig. S1)]. We verified that the *Buchnera* $5A^E$ haplotype arose from the *Buchnera* 5A haplotype, as these share several unique polymorphic sites in common (*SI Appendix*, Table S1). Some mutagenic condition seems to have impacted *Buchnera* $5A^E$, potentially during the microinjection protocol, which may impose cellular stress when *Buchnera* cells are released from bacteriocytes of their original host and subjected to artificial conditions.

Within-Host Genetic Drift. Another surprising result from these experiments is the small estimate for N_{SYM}, at 10 to 20. Previously published direct counts of *Buchnera* cells entering developing *A. pisum* embryos show that ~200 *Buchnera* cells are transferred, although this number is difficult to ascertain as the cells are actively dividing as they enter the embryo (20). Flow cytometry–based counts indicate that *Buchnera* multiplies rapidly after being packaged into bacteriocytes, with numbers increasing by 460-fold from the late embryo stage to the adult stage (21). A smaller N_{SYM} is likely enforced by developmental packaging; the inoculum for each embryo is drawn from one or sometimes two bacteriocytes (13). *Buchnera* replicates actively within bacteriocytes; therefore, in heteroplasmic mothers, *Buchnera* haplotypes are expected to be nonrandomly sorted into bacteriocytes, lowering N_{SYM}. However, results also imply that single bacteriocytes contain both haplotypes, since we would otherwise expect complete loss of one or the other haplotype during a single transmission cycle.

The low estimate of N_{SYM} implies that many mutations that are deleterious within hosts can be fixed, since such deleterious mutations will be effectively neutral if the selection coefficient is less than $1/N_{SYM}$ (22). A large proportion of new mutations are likely deleterious for both symbiont and host; these would include any mutations that reduce the function or stability of essential genes. Essential genes make up most of the *Buchnera* genome (11, 12) and thus present a large mutational target. *Buchnera* and many other obligate symbiont lineages exhibit a distinctive genomic syndrome of accelerated sequence evolution and reduction in genome size and number of genes. This syndrome is attributed to increased rates of fixation of slightly deleterious mutations, leading to increased ratios of nonsynonymous to synonymous changes in protein-coding genes (higher dN/dS) and to inactivation and loss of many genes (23). Small values of N_{SYM} are likely a major basis of this genomic syndrome. Variation in N_{SYM} among systems of maternally transmitted symbioses, due to differences in transmission mechanisms or symbiont packaging within hosts, may contribute to variation in the extent of this genomic degradation. Within-host selection leading to rapid fixation of mutations, as we observed for *Buchnera* $5A^E$, could also contribute to the observed high rates of sequence evolution in symbiont genomes.

A small N_{SYM} also leads to low within-host polymorphism, consistent with observations that obligate symbionts are almost always genetically homogeneous. For example, with N_{SYM} of 10, the time to fixation for a completely neutral mutation is about 20 aphid generations, about the length of one season (24). Mutations under selection are fixed more rapidly. These estimates apply to dynamics within a single matriline and not to the aphid population: drift may fix different *Buchnera* mutations in different matrilines.

The estimate of N_{SYM} for *Buchnera* is close to the estimated bottleneck size of 10 to 20 for a transmission cycle of human mitochondria, in which drift can lead to high frequencies of deleterious haplotypes in offspring (5, 25). In humans, drift in heteroplasmic mothers can elevate frequencies of deleterious mitochondrial haplotypes, such that offspring exhibit fitness deficits when mothers were not affected.

Lack of Recombination within Hosts. Recombination will only make a difference if at least two loci are polymorphic within a host, but this situation is likely rare in *Buchnera* given the general lack of within-host polymorphism. A study using deep sequencing of *Buchnera* genomes of several *A. pisum* lines revealed no sites with intermediate frequency polymorphisms (19). Thus, *Buchnera* is effectively clonal, because haplotypes diverging at more than one locus are almost always segregated into separate matrilines. In our experiments, we had the opportunity to test for the occurrence of recombination within hosts because we generated heteroplasmic matrilines that contained *Buchnera* haplotypes differing at several sites around the genome. We resequenced lines that became fixed for a haplotype and observed no recombinants, implying that recombination does not occur within hosts or occurs rarely enough that we did not observe it. *Buchnera* has a very stable genome architecture, with almost no genomic rearrangements or gene acquisition across more than 100 million y of evolution (26, 27). Both the absence of recombination and the stability of the genome may reflect the loss of recombination pathways. In particular, *Buchnera* is one of the few bacteria known to lack the *recA* pathway, which is required in *Escherichia coli* for homologous recombination.

Implications of Within-Host Selection and Drift for the Evolution of Symbiosis. The clonality and small population size of obligate endosymbiont genomes results in rapid evolution of both DNA and protein sequences and loss of many ancestral genes (6, 7). These features are characteristic of all obligately maternally transmitted symbionts for which genomes have been sequenced and has resulted in the tiniest known bacterial genomes (23). The extent of within-host selection has not been quantified for any obligate endosymbionts, except for mitochondria (5). We observed strong evidence of within-host selection in every trial, often resulting in fixation within a few generations. Positively selected mutations are of course part of *Buchnera* evolution, and our results suggest they may be frequent. However, these likely do not cause selective sweeps that affect other loci: multiple *Buchnera* loci are rarely polymorphic within a host.

For a maternally transmitted symbiont, the optimal N_{SYM} from the host perspective is a compromise between increasing N_{SYM} to control spread of uniformly deleterious mutations and decreasing N_{SYM} to control spread of selfish mutations. Inoculation appears to be a largely host-controlled process (e.g., ref. 13), but selection on hosts to optimize N_{SYM} may be ineffective if deleterious consequences of symbiont evolution are mostly long term. We found that N_{SYM} is very small but that within-host selection was nonetheless effective due to large s_{SYM} values. Potentially, within-host selection has both positive and negative effects on aphid fitness in this symbiosis. Mutations deleterious to both symbiont and host are almost certainly frequent, as most mutations in coding genes will destabilize or inactivate the encoded protein; thus, within-host selection benefits hosts. However, selfish symbiont mutations also might be common: symbionts generally have low replication rates adjusted to host development, and mutations that release this control will spread within hosts, despite host fitness costs. Frank (28) suggested that in some maternally inherited symbioses in insects, hosts sequester a population of symbionts to serve as inoculum, thereby limiting within-host selection, but separation of germline and somatic symbionts is not apparent from detailed studies of transmission in the aphid/*Buchnera* symbiosis (13).

A somewhat parallel case of strong within-host selection has been documented for facultative symbionts within *Drosophila melanogaster*. *Wolbachia* haplotypes with increased copy numbers of a 21-kb region called "Octomom" proliferate more, with negative consequences for host survivorship (29). However, in the presence of viral pathogens, the higher titers can give an enhanced protective effect thereby increasing fly survivorship. Thus, the consequences of a selfish symbiont for host-level fitness can be dependent on the environmental conditions. The aphid/*Buchnera* symbiosis shows that within-host selection can occur even in all-female generations within an obligate symbiosis, in which host and symbiont fitness are expected to be tightly linked due to shared reproductive interests.

Our results show that, in the aphid/*Buchnera* symbiosis, a new mutation can undergo strong positive selection within hosts and quickly spread to fixation during a time span much shorter than an annual seasonal cycle. For the case we studied under laboratory conditions, no strong consequences for host fitness were observed. However, we did observe effects on host phenotypes (body mass and time to first reproduction), and this host-level selective neutrality may not apply to all mutations favored within hosts. Thus, attention to within-host dynamics is critical to understanding evolution of maternally transmitted symbioses.

Materials and Methods

A detailed account of methods is given in *SI Appendix*; only the general experimental framework is described here.

Buchnera haplotypes used in this study (LSR1, 5A, and 5AY) are closely related, with a maximum of 21 nucleotide differences genome wide (19) (*SI Appendix*, Table S1). A single aphid nuclear genotype, *A. pisum* LSR1, was used for all experiments. *Buchnera* LSR1 was the "resident" haplotype, and two foreign haplotypes, *Buchnera* 5A and *Buchnera* 5AY, were "donor" haplotypes injected to establish two heteroplasmic lines (LSR1-5AY and LSR1-5A) using a method similar to that previously reported (15). Prior to the experiment, *A. pisum* LSR1, 5A, and 5AY were screened for the presence of additional bacterial symbionts using universal bacterial primers (559F and 35R) that amplify essentially all bacteria excluding *Buchnera* (30). These aphid lines were found to be singly infected with *Buchnera.* We also resequenced representative lines at the end of the experiments to identify any new *Buchnera* mutations, and no other bacterial sequences were retrieved in screens of these datasets.

To generate aphid lines heteroplasmic for *Buchnera* haplotypes, reproductive *A. pisum* LSR1 were placed on seedlings and allowed to deposit nymphs for 24 h to generate age-controlled recipient aphids. At 4 d old, recipient aphids were exposed to 35 °C for 4 h to reduce resident *Buchnera* LSR1 titer and were returned to 20 °C for ∼24 h before injections were performed. On the day of the injections, adult donor aphids possessing *Buchnera* 5A or 5AY were ground individually in 30 μL Buffer A (25 mM KCl, 10 mM $MgCl_2$, 250 mM sucrose, 35 mM Tris HCl, adjusted to pH 7.5). Aphid homogenate was injected into the ventral abdominal segments nearest to the posterior legs, delivering ∼0.1 μL homogenate per injection. Injections of *Buchnera* 5A and 5AY were performed on separate days.

Injected aphids were placed on fava leaves in Petri dishes for 24 h and survivors transferred to seedlings. Offspring born 9 to 12 d after injection were transferred to individual dishes containing a single fava leaf in 1.5% agar, where they were allowed to mature to adulthood and reproduce for 1 d. Mothers were then screened for *Buchnera* 5A and 5AY using PCR and a diagnostic restriction digest, as previously described (15). Offspring from mothers that tested positive for both recipient *Buchnera* LSR1 and donor *Buchnera* 5A or 5AY were retained, including 3 LSR1-5A mothers (of 148 screened) and 5 LSR1-5AY mothers (of 152 screened). After three generations, we began experiments using individual aphids from matrilines in which sampled aphids had intermediate haplotype frequencies, based on band intensities following PCR and restriction digest. For both combinations (5A-LSR1 and 5AY-LSR1), a single female (G0) was used to initiate the experiment. G0 females were allowed to reproduce for several days and were then sampled; haplotype proportions were later determined to be 38.0% 5A in the LSR1-5A G0 female and 81.0% 5AY in the LSR1-5AY G0 female (as shown in Fig. 2). Because they were older when harvested, G0 females were not included in later analyses. All further generations were sampled after 1 d of reproduction.

We conducted two within-host selection experiments, each initiated with a single female. The LSR1-5AY haplotypes differed in a single base deletion in the promoter for a small heat-shock gene *ibpA*, and these haplotypes were expected to be under temperature-dependent selection, favoring *Buchnera* 5AY under heat and *Buchnera* LSR1 under cool conditions. The LSR1-5A haplotypes showed a few differences that were expected to be neutral, so we predicted no within-host selection. In each experiment, selection occurred under two environments: cool (constant 20 °C) and heat (20 °C interrupted by 4 h at 33 °C during each 24-h cycle). We estimated *Buchnera* haplotype frequencies in individual females over five generations by sampling females at the same developmental stage in each generation. In all, we estimated haplotype frequencies for 330 and 357 mother–daughter pairs for the LSR1-5AY and LSR1-5A experiments, respectively. Frequencies in individual females were estimated using PCR across a single-nucleotide polymorphism, then using Illumina sequencing of amplicons to compute the frequency of each type.

We used a likelihood approach to quantify selection and drift acting on *Buchnera*. Drift reflects the bottleneck that occurs during inoculation and during developmental packaging into bacteriocytes, and so we refer to this population size as the "bottleneck size," N_{SYM}. N_{SYM} can be regarded as an effective population size describing the drift acting on symbionts that occurs during a single aphid generation. Likewise, selection acts on *Buchnera* while it inhabits its host and during transmission to progeny, so s_{SYM} can be regarded as a selection coefficient describing the cumulative selective forces that occur throughout a single aphid generation. Details of the likelihood model are given in *SI Appendix*.

We resequenced *Buchnera* genomes at the end of the experiment using Illumina iSeq to verify all polymorphisms, including any that arose during the experiment. We discovered that the *Buchnera* 5A haplotype had produced a descendant with several mutations, and we named this haplotype *Buchnera* $5A^E$.

To determine whether the fitness advantage observed within hosts had an effect on between-host selection, we assessed fitness parameters of individual aphids with the same nuclear genotype (LSR1) but with different *Buchnera* haplotypes: *Buchnera* $5A^E$ and *Buchnera* LSR1. We also used these aphid lines in a long-term competition experiment in which we transferred aphids to new plants each week for 12 wk, using a large transfer population to minimize chance sampling effects. We sampled the population at each transfer date and estimated frequencies of the two aphid types by extracting DNA from these population samples, which pooled many individuals. Both transferred and harvested samples consisted of randomly selected, mixed-age aphids from the total population. For each sample, we performed PCR over a polymorphic 13 base pair insertion and sequenced amplicons in a single run of Illumina iSeq (as for the within-host experiment). We constructed a standard curve with known proportions of aphids of the two lines to verify this method of estimating population frequencies (*SI Appendix*, Fig. S3). We applied a similar maximum likelihood approach to that used for the within-host experiment, to obtain an estimate of the between-host selection coefficient, s_{HOST}.

Data Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. Funding came from NSF Award 1551092 to N.A.M., a Provost's Graduate Excellence Fellowship to J.P., NIH Grant R01-GM116853 to M.K., and the China Scholarship Council and the Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences to B.Z. We thank Kim Hammond for maintaining aphids, plants, and laboratory supplies and for work on figures, Eli Powell for technical assistance, and Yiyuan Li for sharing bioinformatic scripts.

1. J. P. McCutcheon, B. M. Boyd, C. Dale, The life of an insect endosymbiont from the cradle to the grave. *Curr. Biol.* **29**, R485–R495 (2019).
2. N. A. Moran, J. P. McCutcheon, A. Nakabachi, Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* **42**, 165–190 (2008).
3. J. J. Wernegreen, Ancient bacterial endosymbionts of insects: Genomes as sources of insight and springboards for inquiry. *Exp. Cell Res.* **358**, 427–432 (2017).
4. G. M. Bennett, N. A. Moran, Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10169–10176 (2015).
5. A. A. Zaidi *et al.*, Bottleneck and selection in the germline and maternal age influence transmission of mitochondrial DNA in human pedigrees. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25172–25178 (2019).
6. B. Charlesworth, D. Charlesworth, The degeneration of Y chromosomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**, 1563–1572 (2000).
7. N. A. Moran, Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2873–2878 (1996).
8. C. Rispe, N. A. Moran, Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. *Am. Nat.* **156**, 425–441 (2000).
9. M. E. Pettersson, O. G. Berg, Muller's ratchet in symbiont populations. *Genetica* **130**, 199–211 (2007).
10. B. O'Fallon, Population structure, levels of selection, and the evolution of intracellular symbionts. *Evolution* **62**, 361–373 (2008).
11. S. Shigenobu, A. C. C. Wilson, Genomic revelations of a mutualism: The pea aphid and its obligate bacterial symbiont. *Cell. Mol. Life Sci.* **68**, 1297–1309 (2011).
12. S. Shigenobu, H. Watanabe, M. Hattori, Y. Sakaki, H. Ishikawa, Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**, 81–86 (2000).
13. R. Koga, X.-Y. Meng, T. Tsuchida, T. Fukatsu, Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E1230–E1237 (2012).
14. C. Braendle *et al.*, Developmental origin and evolution of bacteriocytes in the aphid-*Buchnera* symbiosis. *PLoS Biol.* **1**, E21 (2003).

15. N. A. Moran, Y. Yun, Experimental replacement of an obligate insect symbiont. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2093–2096 (2015).
16. H. E. Dunbar, A. C. C. Wilson, N. R. Ferguson, N. A. Moran, Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biol.* **5**, e96 (2007).
17. B. Zhang, S. P. Leonard, Y. Li, N. A. Moran, Obligate bacterial endosymbionts limit thermal tolerance of insect host species. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24712–24718 (2019).
18. S. P. Otto, I. M. Hastings, Mutation and selection within the individual. *Genetica* **102-103**, 507–524 (1998).
19. N. A. Moran, H. J. McLaughlin, R. Sorek, The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science* **323**, 379–382 (2009).
20. A. Mira, N. A. Moran, Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb. Ecol.* **44**, 137–143 (2002).
21. P. Simonet *et al*., Direct flow cytometry measurements reveal a fine-tuning of symbiotic cell dynamics according to the host developmental needs in aphid symbiosis. *Sci. Rep.* **6**, 19967 (2016).
22. H. Akashi, N. Osada, T. Ohta, Weak selection and protein evolution. *Genetics* **192**, 15–31 (2012).
23. J. P. McCutcheon, N. A. Moran, Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2011).
24. M. Kimura, T. Ohta, The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763–771 (1969).
25. B. Rebolledo-Jaramillo *et al*., Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15474–15479 (2014).
26. I. Tamas *et al*., 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296**, 2376–2379 (2002).
27. R. A. Chong, H. Park, N. A. Moran, Genome evolution of the obligate endosymbiont *Buchnera aphidicola*. *Mol. Biol. Evol.* **36**, 1481–1489 (2019).
28. S. A. Frank, Perspective: Repression of competition and the evolution of cooperation. *Evolution* **57**, 693–705 (2003).
29. E. Chrostek, L. Teixeira, Mutualism breakdown by amplification of *Wolbachia* genes. *PLoS Biol.* **13**, e1002065 (2015).
30. J. A. Russell, A. Latorre, B. Sabater-Muñoz, A. Moya, N. A. Moran, Side-stepping secondary symbionts: Widespread horizontal transfer across and beyond the Aphidoidea. *Mol. Ecol.* **12**, 1061–1075 (2003).