Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments

Liana T. Burghardt[a], Brendan Epstein[a], Joseph Guhlin[a], Matt S. Nelson[b], Margaret R. Taylor[b], Nevin D. Young[a,c], Michael J. Sadowsky[a,b,d], and Peter Tiffin[a,1]

[a]Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN 55108; [b]BioTechnology Institute, University of Minnesota, St. Paul, MN 55108; [c]Department of Plant Pathology, University of Minnesota, St. Paul, MN 55108; and [d]Department of Soil, Water, and Climate, University of Minnesota, St. Paul, MN 55108

Edited by Detlef Weigel, Max Planck Institute for Developmental Biology, Tübingen, Germany, and approved January 17, 2018 (received for review August 11, 2017)

Assays to accurately estimate relative fitness of bacteria growing in multistrain communities can advance our understanding of how selection shapes diversity within a lineage. Here, we present a variant of the "evolve and resequence" approach both to estimate relative fitness and to identify genetic variants responsible for fitness variation of symbiotic bacteria in free-living and host environments. We demonstrate the utility of this approach by characterizing selection by two plant hosts and in two free-living environments (sterilized soil and liquid media) acting on synthetic communities of the facultatively symbiotic bacterium *Ensifer meliloti*. We find (*i*) selection that hosts exert on rhizobial communities depends on competition among strains, (*ii*) selection is stronger inside hosts than in either free-living environment, and (*iii*) a positive host-dependent relationship between relative strain fitness in multistrain communities and host benefits provided by strains in single-strain experiments. The greatest changes in allele frequencies in response to plant hosts are in genes associated with motility, regulation of nitrogen fixation, and host/rhizobia signaling. The approach we present provides a powerful complement to experimental evolution and forward genetic screens for characterizing selection in bacterial populations, identifying gene function, and surveying the functional importance of naturally occurring genomic variation.

Medicago | *Ensifer meliloti* | evolve and resequence | facultative mutualism | synthetic community

High-throughput sequencing and new empirical approaches have led to new insights into the ways selection shapes microbial communities as well as the genomic basis of microbial adaptation (1). Characterizing how selection shapes microbial communities is primarily done by comparing the relative abundance of evolutionary lineages [e.g., in metagenomics (2, 3) or synthetic communities (4)]. By contrast, characterization of selection and the genomic basis of adaptation in microbial lineages is often done using experimental evolution approaches that track de novo mutations (5, 6). Experimental evolution has greatly advanced our understanding of adaptation in bacterial populations, including clonal interference, coexistence, and the role of selective sweeps (7). These experiments, however, do not provide direct insight into selection acting on naturally occurring, standing genetic variation. Constructing synthetic communities with a large number of closely related strains (i.e., strains of the same species) can provide this insight. Here, we introduce a variant of "evolve and resequence" experiments that directly estimates relative strain fitness using a large synthetic community of naturally variable bacterial strains. This "select and resequence" method relies on sequencing the genomes of each strain before forming the synthetic community, using pooled sequencing (8) to measure allele frequencies of this community both before and after exposure to selection, and then using bioinformatic analyses to estimate strain frequencies (9). This approach opens new avenues to characterize the strength of selection acting in multistrain competitive environments as well as to survey the fitness consequences of naturally occurring allelic variants.

We demonstrate this method by investigating selection acting on the rhizobial bacteria *Ensifer meliloti* when growing in symbiosis with plant hosts and in free-living environments. Like other rhizobia, *E. meliloti* forms a facultative mutualism with legume plant hosts, species in the genus *Medicago* in this case, although at any given time, the majority of the population is living in the soil. When growing symbiotically inside host organs called nodules, rhizobia convert atmospheric nitrogen into a plant-useable form, thereby supporting plant growth. In exchange, rhizobia receive nutritional resources, mainly carbon, from the host (10). A complex signal exchange occurs between hosts and rhizobia to determine whether a functional symbiosis can be established (11). While each host can form scores of nodules, each nodule generally houses the progeny of a single rhizobium. As these indeterminate nodules develop, rhizobia form two distinct populations within the host: Some progeny become incorporated into host cells and differentiate into nonreproductive bacteroids that fix nitrogen, while others remain undifferentiated, continue to reproduce in the nodule tip, and are released into the soil when the nodule senesces (10). These reproductively competent cells represent the fitness outcome of the mutualism from the rhizobial perspective. The symbiosis has been well studied in single-strain experiments, but little is known about the consequences of competition among strains for strain fitness or the mutualism. To fill this gap, we

Significance

We describe an empirical approach to measure the outcomes of selection and competition in bacterial populations. This approach differs from others in that it examines selection acting on naturally occurring variation rather than new mutations. We demonstrate this method by examining selection on rhizobial bacteria living both in symbiosis with leguminous plants and independently in the soil. We identify fitness correlations across environments that could affect the maintenance of the mutualism and natural genomic variants underlying bacterial fitness. Identifying selection inside and outside of hosts may lead to future manipulation of the mutualism to increase agricultural yields.

Author contributions: L.T.B. and P.T. designed research; L.T.B., M.S.N., M.R.T., and P.T. performed research; L.T.B., B.E., and J.G. analyzed data; N.D.Y. and M.J.S. contributed to project development; and L.T.B., B.E., N.D.Y., M.J.S., and P.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: All datasets generated during and/or analyzed during the current study have been made publicly available. Illumina short read sequences for individual strains and pools have been deposited at the National Center for Biotechnology Information (pools: accession nos. SRR6029825–SRR6029912, individual strains: accession nos. SRR6055493–SRR6055493), and variant files, phenotypic data, metadata about the 101 strains and all major scripts used to analyze the data and generate figures have been deposited in a Dryad repository (doi:10.5061/dryad.fp1bg).

[1]To whom correspondence should be addressed. Email: ptiffin@umn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1714246115/-/DCSupplemental.

Published online February 16, 2018.

employed the select and resequence strategy in three distinct environments (Fig. 1). Specifically, we investigate (*i*) host-dependent selection, (*ii*) alignment of rhizobial fitness and host benefit derived from the mutualism, and (*iii*) selection acting on rhizobial populations when they are living outside of hosts. Our experiments provide insight into selection acting on bacterial populations and identify naturally occurring alleles associated with fitness variation in specific environments.

We first ask to what extent host-imposed selection on rhizobial populations depends on competition among rhizobial strains. Experiments that have inoculated plants with single or, at most, a few strains have revealed extensive host genotype-by-rhizobia genotype interactions in the fitness outcomes of the symbiosis, suggesting that selection acting on rhizobia depends on the host genotype (12–14). Although single-strain experiments are empirically tractable, fitness estimates from single-strain experiments may be biased because they do not incorporate the effects of interstrain competition to form nodules (15) or of hosts preferentially rewarding strains following nodule formation (16). If either of these effects is strong, then rhizobial fitness estimated in single-strain environments, and genome-by-genome interactions inferred from them, will not be representative of fitness in more ecologically realistic communities, which can harbor high levels of genetic variation (17). Second, we ask if strains that provide the most host benefit have the greatest relative fitness in hosts when competing with other strains. Mechanisms that provide greater rewards to more beneficial partners or fewer rewards to less beneficial partners are predicted to promote the long-term evolutionary persistence of mutualisms (18, 19). There is some support for such fitness alignment in the legume/rhizobia system (20); however, the data in support of fitness alignments come largely from single-strain inoculation experiments (15). Estimating rhizobial fitness in mixed-strain experiments, as we do here, avoids the biases of single-strain experiments. Lastly, we measure the strength of selection on rhizobial populations when those populations are growing outside of plant hosts. Soil contains large populations of rhizobia (10), and, at any given time, much of the rhizobial population is being exposed to selection outside of the host. The large size of populations and the strength of selection on bacteria in soil (21) suggest that soil selection could be a major driver of rhizobial evolution. Moreover, tradeoffs between fitness in hosts and fitness in free-living environments could constrain adaptation to hosts.

Results

To investigate selection imposed by plant hosts, we inoculated each of two *Medicago* genotypes, A17 (22) and R108 (23), growing in sterilized peat media with a synthetic community of 101 *E. meliloti* strains. From each replicate, we harvested an average of 250 nodules from 5-wk-old plants and then used a series of centrifugation steps to separate undifferentiated bacteria from plant material and bacteroids, extracted and sequenced DNA, and used bioinformatic analyses to estimate strain frequencies (9). The separation of bacteria from plant cells is central to the success of the method because it allows for efficient estimates of strain frequencies from allele frequencies using pooled sequencing. Changes in strain frequencies inside nodules provide a cumulative measure of selection accounting for survival before forming nodules (because hosts were grown in the same growth media, this should affect fitness in both hosts equally), competition between strains for nodulation, and differential rewards/sanctions after nodules are formed. To complement these host environments, we measured selection in two disparate free-living contexts. To estimate selection in a natural soil, we inoculated sterilized field soil with the same synthetic community; after 8 wk, we extracted and sequenced DNA from surviving rhizobia and estimated strain frequencies. By using sterilized soil, we were able to estimate the effect of the specific abiotic soil environment on rhizobial fitness but may miss potentially important selection imposed by interspecies competition and predators. We similarly measured selection in liquid tryptone-yeast (TY) medium by growing the rhizobial community for 3 d and extracting and sequencing DNA. This resource-rich, homogeneous environment is commonly used to grow *Ensifer* bacteria in the laboratory. Since strains did not all begin at the exact same frequency in the initial community, we converted strain frequencies to relative fitness estimates by dividing the strain frequency after selection in each environment by initial strain frequency and then taking the $\log_2$ (Fig. 2).

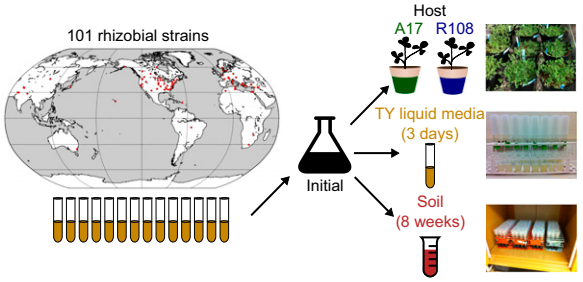

Fig. 1. Experimental design schematic. A total of 101 *E. meliloti* strains originally collected from geographically widespread locations (red dots) were acquired from the USDA and individually sequenced. To create the synthetic community, each strain was grown individually in liquid media and then combined in approximately equal proportions. We exposed this community to selection in four environments (two host genotypes, TY liquid medium, and sterilized field soil) and used PoolSeq to estimate allele frequencies and strain frequencies before and after selection.

Selection Strength and Fitness Correlations Across Environments. The four selective environments had strong and highly reproducible effects on relative strain fitness (Figs. S1 and S2). Redundancy analyses indicate that the environment explained a large and significant proportion of variation in relative fitness ($r^2_{adj} = 0.75$, $p_{df\ =\ 3,23} = 0.001$; Fig. 3*B*). Selection imposed by plant hosts was much stronger than selection in either soil or TY medium; the variance in relative fitness was approximately fourfold greater in plant hosts than soil and >1.5-fold greater in plant hosts than the liquid media. In the strongest selective environment, the A17 host, the distribution of relative fitness values was strongly skewed and only 15% of strains increased in frequency relative to the starting population (Fig. 3*A*). Strain fitness in the hosts was significantly positively correlated, although fitness in one host was not strongly predictive of fitness in the other host ($r^2 = 0.27$; Fig. 3*C*). Single-strain inoculation experiments revealed that all but one strain were capable of forming nodules with both hosts (Fig. 4 *A* and *B*). Thus, differences in rhizobial fitness inside nodules of the two host genotypes were not simply caused by rhizobia/host incompatibility. The correlations between rhizobial fitness in each host and in each of the free-living environments were weak (all $r^2 < 0.05$); however, there was a significant negative correlation between rhizobial fitness in soil and in nutrient-rich liquid media ($r^2 = 0.43$).

Comparison of Single-Strain and Multistrain Experiments. In the absence of rhizobial competition to form nodules or differential rewarding of strains by plant hosts, we would expect relative fitness in the multistrain communities to be strongly positively correlated with fitness measured in single-strain inoculations (Fig. 4 *A* and *B*). By contrast, there was only a weak (R108: $r^2 = 0.13$, $p_{df\ =\ 63} = 0.002$) or no (A17: $r^2 = 0.004$, $p_{df\ =\ 96} = 0.47$) relationship between nodule number in single-strain inoculations and relative fitness in the multistrain community (Fig. 4 *A* and *B*). The correlations between relative fitness estimates from the multistrain community and the number of pink (putatively nitrogen-fixing) nodules in single-strain inoculations were slightly stronger ($r^2 = 0.34$, $p_{df\ =\ 96} < 0.001$ for A17; $r^2 = 0.18$, $p_{df\ =\ 63} << 0.001$ for R108; Fig. 4 *C* and *D*) than the correlations between relative fitness and total nodule number. Similarly, rhizobial fitness in multistrain communities in both plant hosts was significantly, although not necessarily strongly, correlated with the benefits plants obtained, as estimated by plant dry weight, when grown with each strain individually ($r^2 = 0.54$,

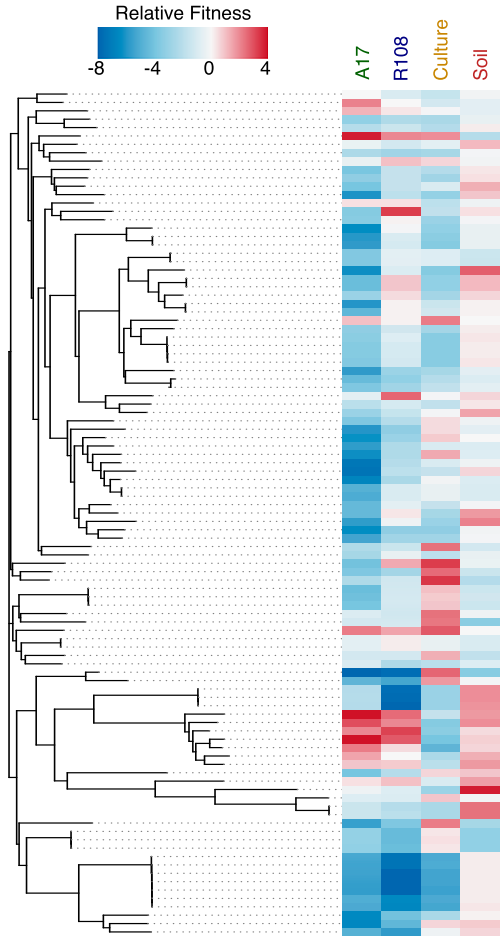

Fig. 2. Neighbor-joining tree of strain relationships paired with a heat map of relative strain fitness (strain $\log_2$-fold change) in each environment.

$p_{df=96} < 0.001$ for A17; $r^2 = 0.14$, $p_{df=63} = 0.002$ for R108; Fig. 4 *E* and *F*). The stronger correlations suggest that more helpful strains (in terms of nitrogen fixation or contributing to host biomass) are more competitive at forming nodules or are preferentially rewarded by hosts, particularly in A17. Interestingly, although putatively unhelpful strains had lower fitness in multistrain communities than beneficial strains, unhelpful strains did not necessarily have the lowest fitness (Fig. 4 *C*–*F*).

Genes with the Strongest Changes in Allele Frequency. Because we sequenced the genomes of each of the strains before forming the initial rhizobia community, and estimated allele frequencies in pre- and postselection samples, we were able to identify the variants that changed strongly and consistently in response to selection. We identified SNPs using a generalized linear model with a binomial link (Fig. S3). To account for nonindependence of allelic variants (i.e., SNPs), we grouped variants that were in strong linkage disequilibrium (LD; $r^2 > 0.75$) into variant groups. The alleles showing the greatest change in frequency in response to selection are likely those that are most responsible for variation in relative fitness.

The variant groups showing the greatest changes in frequency in response to hosts included genes with well-established functions related to cell motility, nitrogen fixation, and nodule formation that occur on all three primary, single-copy replicons that comprise the *Ensifer* genome (Table S1). Most striking, among the SNPs with the greatest change in allele frequencies in both plant hosts were three nonsynonymous variants in the highly conserved backbone region of a gene encoding flagellin B, a major component of the extracellular flagellar filament (24). Mutations in this region can reduce motility, which can, in turn, affect nodule formation and interstrain competitiveness (25). Flagellin proteins also trigger host immune responses (26). In contrast, other variants were detected in only one host: in R108, variants in the N_2 fixation-related transcriptional regulators FixJ and FixL (27) nearly became fixed, whereas in A17, variants in genes implicated in exopolysaccharide composition, which affects the outcome of host interactions (28), experienced strong shifts. Interestingly, in liquid media, the largest frequency changes involved nonsynonymous SNPs in genes directly involved in nitrogen fixation and nodulation (29), *fixQ2*, *fixA*, and *fixG*, as well as in nodulation protein NoeA. These variants may have pleiotropic effects in nutrient-rich, free-living environments or be linked to variants that are under selection.

As expected, the magnitude of changes in the frequencies of alleles most strongly affected by selection largely mirrored variation in the strength of selection and strain-level fitness correlations across environments (Fig. S4). Alleles that experienced strong shifts in response to selection by one host tended to experience similar shifts in the other host (positive pleiotropy), and there was a negative correlation between changes in allele frequencies in soil compared with liquid media (negative pleiotropy). Alleles strongly affected by plant hosts showed a diversity of effects in free-living environments: Some were conditionally neutral, showing no change in frequency, while others were positively or negatively pleiotropic. For instance, a nonsynonymous variant in *fixQ2* strongly increased in frequency in liquid media ($\Delta P = 0.49$) and hosts ($\Delta P = 0.31$ and $\Delta P = 0.55$), but strongly decreased in frequency in the soil ($\Delta P = -0.26$).

An alternative approach to identify genes that contribute to fitness variation is to use a standard association analyses that implements a linear mixed model (LMM) with fitness as a response variable, genomic variants as potential explanatory variables, and a relatedness matrix included as a covariate to statistically control for unequal relatedness among strains (30). Including a relatedness matrix can reduce the probability of identifying alleles that change in frequency because they are in LD with causative alleles due to population history, although it comes at the potential expense of

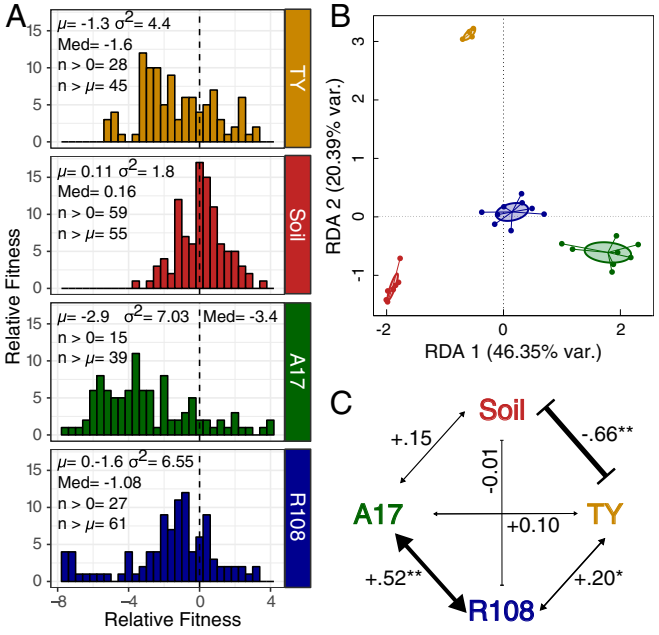

Fig. 3. (*A*) Relative strain fitness in each of the four selective environments. (*B*) First two axes of variation from RDA showing high repeatability of community composition within treatments (dots of the same color) and strong differences among treatments (ellipses represent the 95% confidence interval). (*C*) Pearson correlation coefficients (df = 99) of among-treatment differences in strain fitness (*$p < 0.05$ and **$p < 0.001$, respectively).

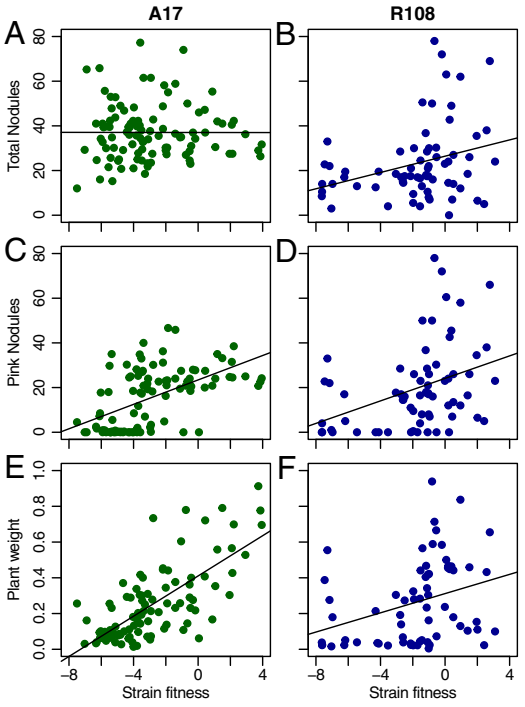

Fig. 4. Correlations between strain fitness in multistrain communities and rhizobial and host fitness proxies in single-strain inoculation experiments: total nodules (*A* and *B*), pink (likely nitrogen-fixing) nodules (*C* and *D*), and plant dry weight (*E* and *F*). Data with host A17 (*Left*) and host R108 (*Right*) are shown. Raw data are shown, but correlations were calculated from residuals after controlling for experimental structure.

an increased number of false-negative results if fitness and relatedness are strongly correlated (31, 32). The LMM analysis revealed that the top variant associations in the A17 host, the genetic background in which most functional genetic work has been conducted, were in transcriptional regulators of both host recognition and nitrogen fixation: NifA (27, 33) and a NodW homolog (34) (Table S2). In R108, the association analyses identified at least one strong candidate associated with motility, a component of the flagellar motor (MotB) (35). In free-living environments, we identified a candidate variant in KpsT, a gene encoding part of a protein involved in export of polysialic acid, which affects the exopolysaccharides surrounding cells (36) and desiccation tolerance (37), that was negatively associated with fitness in the soil and positively associated with fitness in the liquid media. We note that our association analyses are complicated by extensive LD, presence-absence variants (PAVs), and population structure (Fig. S7), which are not fully accounted for by inclusion of the among-strain relatedness matrix. For these reasons, the association analyses results should be viewed more as exploratory rather than as strong evidence for the functional importance of the identified candidate genes.

Discussion

In nature, rhizobia resemble other facultative symbionts: They live in multistrain, competitive communities and experience selection in both host and free-living environments. By contrast, most empirical work investigating rhizobial genetics and evolution is conducted in single-strain environments and is focused on interactions with plant hosts. Here, we present a select and resequence approach that can be used both to efficiently assay bacterial relative fitness in multistrain communities and to characterize naturally occurring allelic variants. We demonstrated this approach by investigating the strength of selection on a synthetic rhizobial community in three distinct environments: plant hosts, sterile soil, and a nutrient-rich liquid media. While the experiments we describe better resemble natural conditions experienced by rhizobia and their hosts, the ecological realism of these experiments is still limited. With these limitations in mind, our results identify avenues for future study.

Our results reveal only weak correlations between rhizobial fitness estimated in multistrain environments and either the number of nodules or the number of pink nodules, both of which are used as proxies of rhizobial fitness in single-strain experiments. The weak correlations confirm that competition for host access and/or host rewards is an important determinant of *Ensifer* fitness when growing symbiotically inside *Medicago* hosts (15). Consequently, experiments that fail to consider the multistrain environments may yield biased inferences of fitness and fitness alignments between plants and rhizobia, and thus mutualism evolution (15). Our results show that the extent of such biases likely depends not only on the measure of rhizobial fitness used in single-strain experiments (i.e., total nodules, pink nodules) but also on the genetic identity of the plant host. The A17 host indiscriminately formed nodules in single-strain inoculations, but in mixed communities, it was far better at rewarding, in terms of fitness, strains that provided more benefit (16).

We found that selection in plant hosts was stronger than selection in the soil. While this result may be expected, given the potential rewards that rhizobia gain from plant hosts (10), our data allowed us to determine that the variance in relative fitness was approximately fourfold greater with plant hosts than in soil [the variance in fitness is proportional to the strength of selection (38)]. Despite this difference in selection strength, the soil environment may be important in shaping the adaptive trajectory of rhizobial populations, because it is likely that much more of the rhizobial population is found in the soil than in association with plant hosts (10). Indeed, symbiosis genes from agriculturally introduced rhizobia often transfer into related strains adapted to local soil environments (39), and a soil generalist haplotype rapidly expanded across the California range of its legume host (40). Resolving the relative importance of selection in hosts and soil will require not only robust estimates of the strength of selection but also knowledge of the frequency with which plant and soil selection is imposed and the proportion of the population experiencing each of those environments. Moreover, our measurement of selection in the soil comes with the caveat that we examined selection in only a single, sterilized soil type; selection imposed by interspecific competition or soil predators is not incorporated into this estimate.

Strong selection is expected to cause shifts in the frequency of variants responsible for variation in relative fitness as well as the frequency of noncausative alleles that hitch-hike along with causative alleles (41). The LD we observe among strongly shifted variants illustrates empirically how strong selection can drive allelic changes in noncausative variants (42), even at genes far from selected sites (43). While the extensive LD among the alleles most strongly affected by selection can complicate the identification of causative variants, many of the variants showing large changes in allele frequencies were in genes that have been identified through forward or reverse genetics to affect symbiotic interactions or survival in free-living environments (44). While we did not functionally validate candidates, these provide an excellent starting point for future studies of the mechanisms that are responsible for naturally occurring variation in rhizobia fitness and variation in the outcomes of legume-by-rhizobia interactions.

The select and resequence approach we present provides an efficient and empirically tractable way to estimate relative fitness of bacterial strains growing in complex, multistrain communities. Moreover, by providing an efficient screen for the functional importance of naturally occurring alleles, it provides a potentially powerful complement to forward genetic screens, including transposon sequencing (Tn-seq) experiments (45), that infer gene function on the basis of knockout mutations, as well as experimental evolution approaches that rely on de novo mutations (46). It should be possible to expand into more ecologically relevant environments to further our understanding of selection acting on rhizobial populations, including, for instance, a diversity of soil types, host genotypes, and biotic communities. Pairing the

select-and-resequence approach with new methods to infer strain frequencies from large metagenomic samples (47) could provide additional ecological insights. Perhaps most importantly, this assay should be applicable to other bacterial populations in a range of free-living and symbiotic environments.

Methods

We created a synthetic community by pooling 101 strains from the US Department of Agriculture (USDA) *Ensifer* strain collection (strain relationships are illustrated in Fig. S5, and information on strain sampling is provided in *SI Methods*). Isolates of each strain were grown separately in 3 mL of liquid TY medium (6 g of tryptone, 3 g of yeast extract, and 0.38 g of $CaCl_2$ in 1 L of dH_20) at 28 °C with shaking. After 3 d, we combined an equal volume of each culture to create an initial rhizobial community with approximately equal representation from each of the 101 strains (Fig. S2).

Relative Fitness *in Planta*. Seeds of *Medicago truncatula* Jemalong A17 (Medicago Hapmap ID no. HM101) and R108 (Medicago Hapmap ID no. HM340) were sterilized in 10% bleach for 90 s, washed, and scarified with a razor blade. Seeds were incubated in the dark for 2 d at 4 °C before being moved to room temperature for a day to germinate. We planted eight to 10 replicate pots (1 L) of each genotype with 10–12 pregerminated seeds. Each pot was filled with autoclaved Sunshine Mix LP5 (SunGro Horticulture). At the time of planting, each pot was inoculated with 100 mL of the initial community (diluted in 0.85% NaCl solution to $\sim 10^6$ bacteria per milliliter). Pots were fertilized with N-free fertilizer (48), and thereafter watered with sterile water as needed. After 5 wk, we harvested nodules from eight pots of A17 and nine pots of R108 by gently separating the roots from the soil and removing all visible nodules with tweezers. Nodules were pooled across all plants (n = 8–13) in each pot. A17 plants had more and smaller nodules (average: 300 nodules, 198 mg per pot, range: 253–370 nodules, 115–274 mg) than R108 plants (average: 189 nodules, 453 mg, range: 89–343 nodules, 230–642 mg). We surface-sterilized nodules by soaking them in 10% bleach solution for 10–20 s before rinsing them thoroughly in sterile dH_20. We crushed nodules with a pestle, added 1 mL of 0.85% NaCl solution to a 1.5-mL microcentrifuge tube, and then centrifuged the nodules for 10 min at 400 × g. The centrifuging creates a pellet enriched for plant tissue and bacteroids and a supernatant enriched for undifferentiated microbial cells. We pipetted the supernatant into a sterile tube and repeated the crushing, diluting, and centrifuging on the pellet. The supernatants from the two rounds were combined and frozen for future DNA extraction.

Relative Fitness in Culture. To measure the relative fitness (i.e., relative growth rates) of the *E. meliloti* strains in rich media, we inoculated four replicate tubes containing 1 mL of liquid TY medium with 20 μL of the initial community mixture ($\sim 10^6$ cells). These cultures were grown for 72 h at 28 °C with shaking. After 72 h, the liquid cultures were frozen. Dilution plating estimates indicated a final population size of $\sim 10^9$ cells per tube.

Relative Fitness in Soil. To measure relative fitness of rhizobia in sterile field soil, we constructed soil mesocosms of 10 g of local field soil, Hubbard loamy sand (Udorthentic Haploboroll), and 2 mL of water in 25 × 200-mm Pyrex culture tubes (Corning, Inc.). The mesocosms were sterilized by autoclaving (two 1-h cycles); after sterilization, we added 1 mL of sterile water and 100 μL of a 1:10 dilution of the initial community mixture to create a starting population size of $\sim 10^6$ cells in each mesocosm. The tubes were vortexed thoroughly to mix, randomized, and placed in a dark cabinet at room temperature. After 8 wk, we sampled rhizobia from seven tubes. To separate rhizobia from the soil, we added 20 mL of 0.85% NaCl solution to each mesocosm and vortexed for 10 s, let the solution sit for 5 min, and vortexed again. After the soil settled, we pipetted off 1 mL of the liquid. We used dilution plating to estimate a final population size $\sim 10^5$ cells per mesocosm. To have sufficient DNA concentrations for sequencing, 1 mL of liquid TY medium was inoculated with 100 μL of soil extract, grown for 24 h (28 °C, 200 rpm), and frozen. To control for differences in strain growth rates in culture, we grew four replicates of the initial community in TY medium for 24 h. Relative fitness for the field soil treatment was estimated by comparing strain frequency after soil selection to this control.

Variant Calling from Individual Strains. We sequenced each of the 101 strains individually to a median coverage of 19.4 using Illumina short-read technology, mapped the reads from each strain to *E. meliloti* USDA1106 [National Center for Biotechnology Information (NCBI) accession nos. CP021797–CP021799], and identified SNPs (details on variant calling are provided in *SI Methods*). For estimating strain frequency from SNPs with Haplotype Analysis of Reads in Pools (HARP) (9), we used all 345,000 variant sites that passed read quality and heterozygosity filters. For allele frequency and genome-wide association (GWA) analyses, we further filtered variant sites based on missing data and minor allele frequency (MAF), resulting in 130,000 variant sites. For the LMM-based association analysis, we used these variant sites as well as 12,613 PAVs with MAF > 0.05 that we identified using de novo assemblies of the individual strains. Fig. S6*A* shows the MAF spectra for each variant type in the sample of 101 strains.

Pooled Sequencing and Data Processing. DNA was extracted from all samples (initial community, nodule, TY, and soil) using an UltraClean Microbial DNA Isolation Kit (no. 12224; Mo Bio Laboratories) following the manufacturer's instructions and used to make NexteraXT sequencing libraries. We obtained 250-bp paired-end reads from an Illumina HiSeq 2500 instrument for 32 libraries, and 125-bp paired-end reads for eight libraries, yielding 1.9–6.9 million read pairs per library [mean = 3.2 million read pairs, corresponding to a read depth of 140 to 260 (mean = 197)]. Application of Super-Deduper (dstreett.github.io/Super-Deduper/) revealed few putative PCR duplicates [0.9–3.3% of reads per library (mean = 1.6%)]; therefore, we did not remove these from the data. The reads were trimmed with TrimGalore! (v0.4.1; www.bioinformatics.babraham.ac.uk/projects/trim_galore/) using default settings, except that the quality parameter was set to 30, the minimum length was set to 150 bp (100 bp for the shorter reads), and the minimum overlap with an adapter sequence required to trigger adapter trimming was set to 3. Trimming reduced the number of read pairs per library to 1.5–5.9 million (mean = 2.6 million). These read pairs were aligned to the USDA1106 genome using bwa mem (v0.7.15) (49) with default settings, and the minimum mapping quality was set to 30, thereby eliminating reads that align to more than one location. A total of 78–90% of the reads aligned, yielding final mean read depth per pool ranging from 70 to 160 reads. Supporting the efficacy of the enrichment approach at excluding other microbes and host cells, the vast majority of reads isolated from nodules map to the *E. meliloti* reference genome (mean = 84%, range: 72–90%), only slightly less than the 85–90% of reads that map in pure culture (that <100% of reads map is likely due to presence/absence polymorphisms absent from the reference genome). Further, only 10–20% of reads from nodules mapped to the host (vs. ∼8% from pure rhizobial cultures).

Strain Frequency and Relative Fitness. We used the maximum-likelihood approach implemented in HARP (9) to estimate the strain frequencies in each population. In brief, HARP first calculates the probability that each read came from each of the strains and then finds the combination of strain frequencies that maximizes the likelihood across all reads. If all strains were in equal frequency in the initial community, we would expect each to be at a frequency just <1%; all of the individually grown strains were found in the initial community, with the least frequent strain at an estimated frequency of 0.24% and 90% of strains estimated to occur at frequencies between 0.55% and 1.5% (Fig. S2). An independent maximum likelihood method yielded highly similar results (Table S3). We calculated mean relative fitness for each strain as $\log_2$(mean selected strain frequency/mean initial strain frequency). The log transformation puts increases and decreases in proportional representation on the same scale (e.g., a fourfold decrease and increase have a value of −2 and 2, respectively). Strains with frequency estimates of 0 in the selected communities were assigned a fitness value of −8, a value less than the lowest measurable reductions in other strains.

Treatment Effects on Relative Fitness. To visualize repeatability of treatment effects among replicates, we performed an unconstrained principal component analysis (PCA) using the "prcomp" function in R. To estimate and visualize the contribution of specific covariates to changes in relative fitness across environments (R108, A17, soil, and TY), we used a constrained, multivariate approach commonly used in community ecology: redundancy analysis (RDA) (50). RDA fits a multivariate linear regression to centered and scaled data, and then uses PCA to decompose the major axis of variation in the fitted parameters. We used the "rda" function in the vegan R package (51) and permutations to determine the probability that differences in relative fitness between environments occurred by chance ("anova" function, 999 permutations).

Allele Frequency Change. We counted the number of reads supporting the reference allele or an alternate allele at each SNP using PoPoolation2 (52) (full details are provided in *SI Methods*). We then used a logistic regression implemented with the R function "glm" (53) to test for a difference in allele frequency between initial and selected communities. Each pool was treated as a replicate, with the count of reads supporting the reference or alternate alleles as the response (ignoring reads with an ambiguous base call) and environment as the predictor. We obtained a P value by using a likelihood ratio test comparing a model with an intercept and selection effect (initial vs. final) to a

model with just an intercept. This method will give higher priority (i.e., smaller *P* values) to sites with more reads because there is more power at those sites to detect allele frequency change. As expected, the top 2,000 SNPs tended to have slightly higher coverage than the genome-wide average, but included many sites with lower than median coverage. Further, many of the most highly covered sites were not among the top 2,000 variants (Fig. S6*B*).

Because variants in strong LD are not statistically distinguishable, we took the 2,000 variants with the lowest *P* values and grouped together those that were in high LD. We started with the variant with the lowest (strongest) *P* value, and calculated the correlation between that variant and each of the other top variants. Any variant with an $r^2 > 0.75$ was grouped with the focal variant. We iterated this procedure with the next lowest *P* value variant until all variants were grouped. Because we are concerned with nonindependence in the selected community, we calculated r^2 using only high-fitness strains ($\omega > 0$) in each environment. Variants for which more than two strains had an ambiguous base call were not used as seeds but were allowed to group with other variants.

GWA. We used LMMs, implemented in GEMMA (v0.94.1) (30) to measure the strength of the association between genotypes and strain fitness at each SNP/multinucleotide polymorphism and PAV. A likelihood ratio test was used to calculate *P* values, and we used the standardized kinship matrix created with GEMMA as a covariate to control for unequal relatedness among strains (QQ plots are shown in Fig. S7). As described previously, we grouped variants based on LD. This time, we calculated LD based on all strains, and variants were not used as seeds if more than 15 strains had ambiguous calls.

Single-Strain Plant Growth and Nodule Data. For 67 (R108) or 98 (A17) strains, we measured plant biomass and counted both total and pink nodules after 5 wk of growth with each strain individually in a separate experiment (*SI Methods*). The phenotypes were adjusted for experimental structure (e.g., batch effects) before calculating Pearson correlations with multistrain fitness estimates from the select and resequence experiments.

Data Availability. Illumina short-read sequences for individual strains and pools are available from the NCBI (pools: accession nos. SRR6029825–SRR6029912, individual strains: accession nos. SRR6055493–SRR6055493), and variant files, phenotypic data, metadata about the 101 strains, and all major scripts used to analyze the data are available via Dryad (doi:10.5061/dryad.fp1bg).

ACKNOWLEDGMENTS. We thank the editor and four anonymous reviewers for comments that greatly improved the manuscript. Funding was provided by NSF Grant IOS-1237993 and USDA-HATCH Award MIN-71-030.

1. Koskella B, Vos M (2015) Adaptation in natural microbial populations. *Annu Rev Ecol Evol Syst* 46:503–522.
2. Thompson LR, et al.; Earth Microbiome Project Consortium (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551:457–463.
3. Wagner MR, et al. (2016) Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat Commun* 7:12151.
4. Grosskopf T, Soyer OS (2014) Synthetic microbial communities. *Curr Opin Microbiol* 18:72–77.
5. Schlötterer C, Kofler R, Versace E, Tobler R, Franssen SU (2015) Combining experimental evolution with next-generation sequencing: A powerful tool to study adaptation from standing genetic variation. *Heredity (Edinb)* 114:431–440.
6. Long A, Liti G, Luptak A, Tenaillon O (2015) Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nat Rev Genet* 16:567–582.
7. Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM (2017) The dynamics of molecular evolution over 60,000 generations. *Nature* 551:45–50.
8. Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals - Mining genome-wide polymorphism data without big funding. *Nat Rev Genet* 15:749–763.
9. Kessner D, Turner TL, Novembre J (2013) Maximum likelihood estimation of frequencies of known haplotypes from pooled sequence data. *Mol Biol Evol* 30:1145–1158.
10. Denison RF, Kiers ET (2011) Life histories of symbiotic rhizobia and mycorrhizal fungi. *Curr Biol* 21:R775–R785.
11. Oldroyd GED, Murray JD, Poole PS, Downie JA (2011) The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet* 45:119–144.
12. Heath KD, Tiffin P (2007) Context dependence in the coevolution of plant and rhizobial mutualists. *Proc Biol Sci* 274:1905–1912.
13. Burghardt LT, et al. (2017) Transcriptomic basis of genome by genome variation in a legume-rhizobia mutualism. *Mol Ecol* 26:6122–6135.
14. Heath KD, Burke PV, Stinchcombe JR (2012) Coevolutionary genetic variation in the legume-rhizobium transcriptome. *Mol Ecol* 21:4735–4747.
15. Kiers ET, Ratcliff WC, Denison RF (2013) Single-strain inoculation may create spurious correlations between legume fitness and rhizobial fitness. *New Phytol* 198:4–6.
16. Oono R, Anderson CG, Denison RF (2011) Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proc Biol Sci* 278:2698–2703.
17. Bailly X, Olivieri I, De Mita S, Cleyet-Marel JC, Béna G (2006) Recombination and selection shape the molecular diversity pattern of nitrogen-fixing *Sinorhizobium* sp. associated to *Medicago*. *Mol Ecol* 15:2719–2734.
18. Akay E (2015) Evolutionary models of mutualism. *Mutualism*, ed Bronstein JL (Oxford Univ Press, Oxford), pp 57–76.
19. Sachs JL, Mueller UG, Wilcox TP, Bull JJ (2004) The evolution of cooperation. *Q Rev Biol* 79:135–160.
20. Friesen ML (2012) Widespread fitness alignment in the legume-rhizobium symbiosis. *New Phytol* 194:1096–1111.
21. van Veen JA, van Overbeek LS, van Elsas JD (1997) Fate and activity of microorganisms introduced into soil. *Microbiol Mol Biol Rev* 61:121–135.
22. Young ND, et al. (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480:520–524.
23. Pislariu CI, et al. (2012) A *Medicago truncatula* tobacco retrotransposon insertion mutant collection with defects in nodule development and symbiotic nitrogen fixation. *Plant Physiol* 159:1686–1699.
24. Tambalo DD, et al. (2010) Characterization and functional analysis of seven flagellin genes in *Rhizobium leguminosarum* bv. *viciae*. Characterization of R. leguminosarum flagellins. *BMC Microbiol* 10:219.
25. Caetano-Anollés G, et al. (1988) Role of motility and chemotaxis in efficiency of nodulation by *Rhizobium meliloti*. *Plant Physiol* 86:1228–1235.
26. Petutsching EK, Katharina M, Parniske M (2014) Knowing your friends and foes–Plant receptor-like kinases as initiators of symbiosis or defence. *New Phytol* 204:791–802.
27. Bobik C, Meilhoc E, Batut J (2006) FixJ: A major regulator of the oxygen limitation response and late symbiotic functions of *Sinorhizobium meliloti*. *J Bacteriol* 188:4890–4902.
28. Skorupska A, Janczarek M, Marczak M, Mazur A, Król J (2006) Rhizobial exopolysaccharides: Genetic control and symbiotic functions. *Microb Cell Fact* 5:7.
29. Becker A, et al. (2004) Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. *Mol Plant Microbe Interact* 17:292–303.
30. Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44:821–824.
31. Atwell S, et al. (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:627–631.
32. Yeaman S, et al. (2016) Convergent local adaptation to climate in distantly related conifers. *Science* 353:1431–1433.
33. Gong ZY, He ZS, Zhu JB, Yu GQ, Zou HS (2006) *Sinorhizobium meliloti nifA* mutant induces different gene expression profile from wild type in Alfalfa nodules. *Cell Res* 16:818–829.
34. Loh J, Garcia M, Stacey G (1997) NodV and NodW, a second flavonoid recognition system regulating nod gene expression in *Bradyrhizobium japonicum*. *J Bacteriol* 179:3013–3020.
35. Platzer J, Sterr W, Hausmann M, Schmitt R (1997) Three genes of a motility operon and their role in flagellar rotary speed variation in *Rhizobium meliloti*. *J Bacteriol* 179:6391–6399.
36. Cuthbertson L, Kos V, Whitfield C (2010) ABC transporters involved in export of cell surface glycoconjugates. *Microbiol Mol Biol Rev* 74:341–362.
37. Vanderlinde EM, et al. (2010) Identification of a novel ABC transporter required for desiccation tolerance, and biofilm formation in *Rhizobium leguminosarum* bv. *viciae* 3841. *FEMS Microbiol Ecol* 71:327–340.
38. Fisher RA (1930) *The Genetical Theory of Natural Selection* (Clarendon, Oxford).
39. Barcellos FG, Menna P, da Silva Batista JS, Hungria M (2007) Evidence of horizontal transfer of symbiotic genes from a *Bradyrhizobium japonicum* inoculant strain to indigenous diazotrophs *Sinorhizobium (Ensifer) fredii* and *Bradyrhizobium elkanii* in a Brazilian Savannah soil. *Appl Environ Microbiol* 73:2635–2643.
40. Hollowell AC, et al. (2016) Metapopulation dominance and genomic-island acquisition of *Bradyrhizobium* with superior catabolic capabilities. *Proc Biol Sci* 283:20160496.
41. Hill WG, Robertson A (1966) The effect of linkage on limits to artificial selection. *Genet Res* 8:269–294.
42. Nuzhdin SV, Turner TL (2013) Promises and limitations of hitchhiking mapping. *Curr Opin Genet Dev* 23:694–699.
43. Skelly DA, Magwene PM, Stone EA (2016) Sporadic, global linkage disequilibrium between unlinked segregating sites. *Genetics* 202:427–437.
44. Levy A, et al. (2017) Genomic features of bacterial adaptation to plants. *Nat Genet* 50:138–150.
45. van Opijnen T, Camilli A (2013) Transposon insertion sequencing: A new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* 11:435–442.
46. Hoang KL, Morran LT, Gerardo NM (2016) Experimental evolution as an underutilized tool for studying beneficial animal-microbe interactions. *Front Microbiol* 7:1444.
47. Albanese D, Donati C (2017) Strain profiling and epidemiology of bacterial species from metagenomic sequencing. *Nat Commun* 8:2260.
48. Bucciarelli B, Hanan J, Palmquist D, Vance CP (2006) A standardized method for analysis of *Medicago truncatula* phenotypic development. *Plant Physiol* 142:207–219.
49. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2.
50. Ramette A (2007) Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol* 62:142–160.
51. Oskanen J, et al. (2017) vegan: Community ecology package, Version 2.4-3. Available at https://CRAN.R-project.org/package=vegan. Accessed February 1, 2017.
52. Kofler R, Pandey RV, Schlötterer C (2011) PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27:3435–3436.
53. Development Core Team R (2016) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).