Testing for divergence in a plant symbiont across two natural environmental gradients

Chang-Yu Chang[1]†, Terrence Topping-Brown[1], Jazmine L. Rud[1], McCall B. Calvert[1], Gerardo Bencosme[1], Linda J. Robinson[1], Corlett W. Wood[1]†

[1]*Department of Biology, University of Pennsylvania, Philadelphia, PA, USA*

†Correspondence: C.-Y. C., C. W. W.

Abstract

Microbes dominate the biological diversity found in natural landscapes. Yet, studies of microbial population differentiation across environmental gradients have been surprisingly scarce. In this study, we tested for phenotypic and genetic divergence in rhizobia bacteria (genus *Sinorhizobium*) associated with the legume *Medicago lupulina* across two environmental gradients: an elevation gradient in the Appalachian Mountains and an urbanization gradient in Philadelphia. Common garden experiments in *Medicago sativa* revealed that symbiotic traits differed significantly between high- and low-elevation rhizobia populations, but not between urban and suburban populations. Nitrogen supplementation erased the trait differences between low- and high-elevation populations, consistent with the hypothesis that these differences arise from the nitrogen that rhizobia provide. In contrast to symbiotic traits, free-living growth traits showed minimal differentiation across gradients. To test the hypothesis that greater gene flow constrains phenotypic divergence in the urbanization gradient, we compared the strength of isolation by distance in the core and accessory genomes across the two gradients. Contrary to our expectation, isolation by distance was significantly stronger in the urbanization gradient in both the core and accessory genomes. These results suggest that our focal elevation gradient imposed stronger divergent selection on rhizobia than our focal urbanization gradient. Together, this study highlights the need for a holistic understanding of microbial evolution with regard to the traits across life stages, molecular evolution across genomic levels, and evolutionary processes across environmental gradients.

Keywords: host-microbe symbiosis, urbanization, elevation gradient, microbial populations

INTRODUCTION

Divergence is the process through which populations become increasingly phenotypically and genetically distinct over time, driven by various evolutionary processes (Clausen et al. 1948; Endler 1977). Microbes are a long-standing model system for studying divergence in the laboratory (Kawecki et al. 2012; Lenski 2017). For example, experimental evolution has clearly shown that identical microbial populations subjected to different selective pressures can rapidly diverge, leading to populations with distinct functions and genetic characteristics (Cvijović et al. 2018). However, empirical tests for population divergence in wild microbes are rare, primarily because studying microbes in their natural habitats presents significant challenges (Chase et al. 2021). As a result, it is unclear whether the rapid microbial adaptation documented in the laboratory is representative of microbial evolution in the wild (Giraud et al. 2017; Kraemer and Boynton 2017).

Environmental gradients are excellent natural laboratories to investigate evolution in the wild (Verheyen et al. 2019). Elevation gradients, for example, have been a model for exploring evolutionary processes in plants and animals for more than 80 years (Clausen et al. 1948; Keller et al. 2013; Halbritter et al. 2018). Many environmental factors, most notably temperature, land area, and UV radiation (Körner 2007), but also biotic factors like community composition, vary predictably across elevation gradients (Lomolino 2001). More recently, urbanization gradients have emerged as highly replicated natural evolution experiments, which provide a perfect playground for studying the evolutionary response of organisms to human-modified environments (Johnson and Munshi-South 2017; Santangelo et al. 2018, 2022). Urbanization, the process by which humans establish dense settlements, results in significant environmental changes from the surrounding rural and suburban areas, including elevated temperatures, increased pollution and nutrient levels, and a higher percentage of impervious surfaces, each of which can drive the evolution of life in the cities (Alberti et al. 2020; Murray-Stoker and Johnson 2021).

To date, evolutionary studies across environmental gradients have predominantly focused on animals and plants. While microbial community ecologists have documented significant changes in microbial *community composition* across environmental gradients (Fuhrman et al. 2008; Fierer et al. 2011; Hendershot et al. 2017), microbial *evolution* (i.e., phenotypic and genetic divergence among populations of the same microbial species) remains largely overlooked. Given the rapid microbial adaptation observed in lab experiments, we might expect microbial divergence across natural environmental gradients to be extensive and common (Koskella and Vos 2015).

However, translating laboratory results into field predictions is not straightforward because of three major differences between lab and wild microbes: the scale of gene flow, microbial generation time, and the complexity of the selective landscape. Gene flow determines the potential for divergence among populations in different environments. Yet the scale of gene flow in wild microbes remains an unresolved debate in microbial evolutionary ecology (Fenchel and Finlay 2004; Telford et al. 2006; Choudoir et al. 2018; Choudoir and DeAngelis 2022). Active dispersal in microbes is constrained to short distances (Custer et al. 2022), but microbes can disperse passively over long distances in rainwater (Evans et al. 2020), wind (Chaudhary et al. 2020), and soil (Walters et al. 2022). Studies of microbial population structure do show that microbial populations are genetically differentiated at large geographic scales (>1000 km) (Riley et al. 2023), suggesting that microbial gene flow can be sufficiently restricted to allow for divergence.

Second, microbial generation times may be longer in the field than in the laboratory (Caro et al. 2023; Walker et al. 2024). Typical laboratory culture conditions (e.g., rich media, minimal interspecific competition) favor rapid microbial growth relative to soil or seawater (Weissman et al. 2021). Furthermore, microbial dormancy—which lengthens generation time—is probably more common in the stressful environmental conditions encountered in the field (Lennon and Jones 2011). Finally, natural environmental gradients represent significantly more complex selective landscapes than the typical lab evolution experiment, which often manipulates only a single or few agents of selection (Kawecki et al. 2012). Conflicting selection imposed by multiple agents of selection could constrain population divergence in wild microbes (Wadgymar et al. 2022). Therefore, whether microbial populations exhibit phenotypic and genetic divergence across natural environmental gradients, like plants and animals, remains an open question.

Rhizobia are an exemplary model for studying microbial population divergence (Denison and Kiers 2011). These bacteria are known for their facultative symbiosis with leguminous plants, where they fix gaseous nitrogen in the root nodules in exchange for photosynthetic products from plants (Sachs et al. 2018). They also have a free-living life stage in the soil and rhizosphere, where the ecological environments are widely different from symbiosis and therefore can be subject to distinct agents of selection (Burghardt 2020; Burghardt and diCenzo 2023). The dual lifestyle of facultative symbionts enables experimental characterization of on- and off-host microbial traits in common garden laboratory conditions (Poole et al. 2018). In addition, many rhizobia taxa have a high degree of host specificity with particular legume species, i.e., such specificity variants are known as symbiovars (Martinez-Romero et al. 2024). This means that the host can be used to "bait" the microbes out of soil microbial communities across a wide geographic range (Harrison et al. 2017*b*; Riley et al. 2023), bypassing challenges usually associated with targeting specific microbial taxa for field sampling (Lopez et al. 2021; Tang et al. 2022). Rhizobia populations exhibit genetic structure across continental scales (Harrison et al. 2017*b*; Riley et al. 2023). Both sequence variation and gene content variation are important sources of genetic variation in symbiotic and free-living rhizobia traits (Sachs et al. 2010; Epstein et al. 2014; Epstein and Tiffin 2021).

In this study, we investigate phenotypic and genetic divergence of *Sinorhizobium* rhizobia populations along two distinct environmental gradients: an elevation gradient in the Appalachian Mountains and an urbanization gradient in Philadelphia, Pennsylvania, USA. Both gradients cover comparable geographic scales (~17 km). We isolated 38 *Sinorhizobium* strains from field-collected *Medicago lupulina* plants and measured divergence in symbiotic traits associated with two host species (the source host *M. lupulina* and an alternate host *M. sativa*) and free-living traits between rhizobia isolated from contrasting environments in a laboratory common garden. Furthermore, we performed de novo genome assemblies to evaluate population structure in the core and accessory genomes within each gradient, and to identify candidate genes that might underlie trait divergence. The common garden component of our study was carried out in partnership with the introductory organismal biology lab classes at the University of Pennsylvania, highlighting how deploying original research as course-based undergraduate research experiences (CUREs) can expand the scale of data collection and engage undergraduates in meaningful scientific investigation.

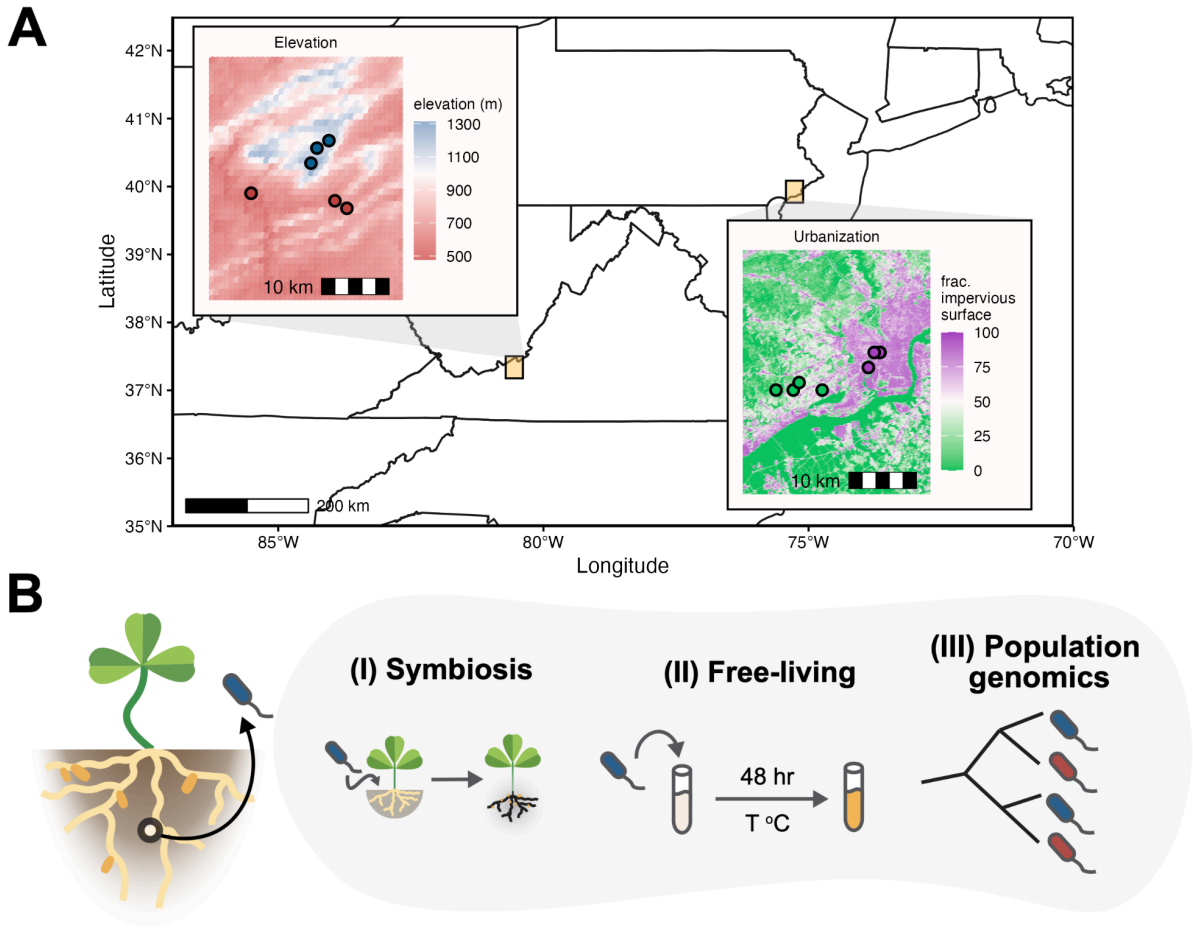

Figure 1. Examining phenotypic and genetic divergence of microbial populations across two environmental gradients. (A) We sampled *Sinorhizobium* isolates from the nodules of *M. lupulina* that inhabit two environmental gradients: an elevation gradient in the southern Appalachian Mountains (left inset) and an urbanization gradient in the greater Philadelphia area (right inset). The inset colors represent elevation (meters above sea level) in the elevation gradient and the fraction impervious surface (0-100) in the urbanization gradient. (B) We assessed the rhizobia (I) symbiosis traits with *M. lupulina* and *M. sativa*, (II) free-living growth, and (III) the population genomics of the core and accessory genomes.

METHODS

Study system

The *Sinorhizobium-Medicago* system is a model for investigating the evolution of symbiotic mutualism (Galibert et al. 2001; Kearsley et al. 2024). *M. lupulina,* an annual or short-lived perennial weed native to Eurasia, was introduced to the Americas in the eighteenth century (Turkington and Cavers 1979). This species is a pioneer plant found in natural environments such as open grasslands, as well as in disturbed areas like lawns and urban green spaces (Turkington and Cavers 1979). In North America, *M. lupulina* is commonly associated with two symbiotic *Sinorhizobium* species: *S. meliloti* and *S. medicae* (Harrison et al. 2017*b*). Both of these *Sinorhizobium* species form symbiotic relationships with various *Medicago* plants, including *M. sativa* (alfalfa) and the genetic model *M. truncatula* (Martinez-Romero et al. 2024).

Both the *S. meliloti* and *S. medicae* share similar genomic structures, comprising a chromosome (3.4-4.2 Mbp), a symbiosis megaplasmid pSymA (0.9-1.7 Mbp) that houses symbiosis genes essential for nodulation and nitrogen fixation (*nod*, *nif*, and *fix*), and a saprophytic chromid pSymB (1.6-2.0 Mbp) that contains chromosome-like and catabolic genes essential for saprophytic competence (diCenzo and Finan 2017; Geddes et al. 2020; Kearsley et al. 2024). Many strains also possess smaller accessory plasmids pAcce that encode a variety of functions (Geddes et al. 2020).

Field sampling

We sampled root nodules of the source plant *M. lupulina* from 13 sites across two environmental gradients between July-September of 2022 (Fig 1). The elevation gradient consisted of six sites (3 high elevation and 3 low elevation), from 490 to 1190 meters above sea level, in the Appalachian Mountains in Giles County, Virginia. The urbanization gradient consists of seven sites (three urban and four suburban) in the Philadelphia region. The paired sites (e.g., high-elevation versus low-elevation sites; suburban versus urban sites) differed in daily temperatures in 2022, the year rhizobia were collected: 2.9 ℃ in daily maximum (t_a) and 0.84℃ in daily minimum (t_i) for the elevation gradients and 0.40 ℃ in t_a and 0.28 ℃ in t_i for the urbanization gradients (Fig. S1-2; Supplementary Section 1).

We stored field-collected *M. lupulina* plants in zip-top bags at 4℃ for up to one week before we isolated the rhizobia strains from the nodules. We then isolated rhizobia with the following steps. First, we surface-sterilized nodules by submerging them in 100% ethanol for 20 seconds, followed by household bleach (5.25% sodium hypochlorite) for 20 seconds. We rinsed the nodules in distilled water and crushed them in a 20 µL droplet of phosphate buffer saline (PBS) on a Petri dish lid with flame-sterilized sharp tweezers. We then used a sterilized inoculation loop and streaked the inoculum on a solid tryptone-yeast (TY) agar plate, incubated the plates at 30℃ for three days until the visible colonies formed, and re-streaked morphologically distinguishable colonies on fresh TY agar plates. We repeated this re-streaking step three times until colonies in one agar plate had consistent morphologies. On the last day of isolation, we suspended the rhizobia colonies in 500 µL of PBS, mixed the suspension with an equal amount of 80% glycerol, and stored the mixture at -80℃ as stocks. The glycerol stocks were later used in the plant experiments, growth curve assays, and whole-genome long-read sequencing. In total, we obtained 38 distinct *Sinorhizobium* strains, including 6 and 9 strains from high- and low- elevation sites, respectively, and 12, and 11 strains from suburban and urban sites, respectively (Table S1).

Rhizobia symbiosis traits (I): alternate host *M. sativa*

University of Pennsylvania undergraduates assessed the rhizobia symbiosis traits in *M. sativa* (an alternate host) seedlings as part of a course-based undergraduate research experience (CURE) in Introductory Biology B (BIOL1102) course lab during the Spring 2023 (urbanization gradient) and Spring 2024 (elevation gradient) semesters. We used *M. sativa* for these experiments, rather than the source host *M. lupulina*, because we expected it to be equally unfamiliar to all field-collected rhizobia strains, and therefore is a reasonable "common garden" host.

We prepared 15 mL of sterile agar slants (30°) in 18 × 150 mm test tubes with nitrogen-free Fahraeus fertilizer (Barker et al. 2006) with 0.7% agar, capped with foam stoppers. We planted sterilized *M. sativa*

seeds purchased from Johnny's Selected Seeds (Product ID 2808), which consists of a heterogeneous mix of genotypes. The planted test tubes were incubated at 4℃ in the dark for two days to stratify the seeds. Test tubes were placed under constant bright fluorescent light at room temperature 22℃ for 3 days for seeds to germinate before inoculation. A nitrogen supplementation experiment was included in the Spring 2024 experiment, for the elevation gradient only. For plants in the nitrogen addition treatment, we added 52 μL of 1.25 mM nitrogen fertilizer (0.5 mM KNO_3 and 1 mM $Ca(NO_3)_2$) to test tubes in the nitrogen treatment immediately after seed stratification.

We selected 8 and 12 strains from the elevation and urbanization gradients, respectively, and diluted rhizobia cultures (grown in TY medium for two days) to OD_{600} = 0.1 with sterile deionized water on the inoculation day. Each group of BIOL1102 students inoculated each of three *M. sativa* seedlings that had germinated on the agar slants with 100 μL of rhizobia inoculum or sterile deionized water as a negative control. Student groups in different lab sections on different days were assigned the same rhizobia strains to control for block effects.

We grew the plants under constant fluorescent light at room temperature in the teaching lab room for 5 weeks. Each lab group (2-3 students) was assigned either 3 plants inoculated with the same strain or 3 uninoculated control plants. Strains and controls were replicated across groups, and the lab group was included in subsequent analyses as a random effect to control for lab group effects. At the end of the fifth week, BIOL1102 student groups measured eight traits: nodule number, nodule size (small or large), nodule color (pink or white), nodule shape (oval or round), stem height, leaf number, leaf color (1-8), and the length of the longest petiole. Each BIOL1102 student group recorded their measurements on a separate Google Form for each of their three plants. The final sample size was 414 plants (255 nitrogen-free (including 30 uninoculated controls) and 159 nitrogen-supplemented (23 controls)) for the elevation gradient and 157 plants for the urbanization gradient (including 52 controls). Detailed protocols for seed preparation, rhizobia inoculation, and plant host trait measurements are provided in Supplementary Section 2.

Rhizobia symbiosis traits (II): source host *M. lupulina*

We chose six (three also used in the alternate host experiment) and eight (two used in the alternate host) strains from elevation and urbanization gradients, respectively (Table S1 for a full list of strains). We inoculated them onto sterile *M. lupulina* seedlings collected from wild populations in Giles County, Virginia. The goal of this experiment was to measure symbiosis traits on the familiar source host background. The experiment employed a block design, with 15 plant maternal lines collected from low-, mid-, and high-elevation sites randomized across rhizobia treatments.

To initiate the experiment, we scarified the seeds with a razor blade, sanitized them using ethanol and household bleach, and stratified them in Petri plates wrapped in aluminum foil to block light at 4℃ for five days (see Supplementary Section 2 for details). After 12-16 hours at room temperature to encourage radicle elongation seedlings were planted in autoclaved cone-tainers (Stuewe & Sons, Inc., Tangent, OR, USA) filled with a 1:4 mix of turface and sand. The plants were maintained in a growth chamber with a 16:8 hour light:dark cycle, maintaining daytime temperatures at 22℃, nighttime temperatures at 18℃, and a relative humidity of approximately 50%. Five weeks after rhizobial inoculation, we top-fertilized

individual plants with 5 mL of one-fifth diluted nitrogen-free Fahraeus fertilizer (Barker et al. 2006) twice a week using an electronic peristaltic pump (Integra DOSE-IT with 4mm silicone tubing) and bottom-watered the plants once a week.

One week after planting, we inoculated each plant with 1 mL of rhizobium inoculum, which had been grown in TY medium at 30℃ for two days and then diluted to an OD_{600} = 0.1 using fresh TY medium. To prevent cross-contamination, we arranged the cone-tainers in groups of six, with each group sharing a single water reservoir to collect excess water. These blocks were included as a random effect in our analysis. Notably, the negative control, which received only fresh TY medium, showed some nodulation, suggesting that a low level of cross-contamination may have occurred.

At the end of the sixth week, we harvested 251 plants, consisting of 167 (including 8 controls) elevation plants and 84 urbanization (including 16 controls), and measured dry shoot biomass, dry root biomass, and counted the nodules. The above-ground portions were dried in a 60℃ oven, while the below-ground parts were frozen at -20℃ for up to six months prior to phenotyping. For phenotyping, we thawed each frozen root sample and counted the nodules under dissecting microscopes. The root samples were then dried at 60℃ and weighed using a scale with a resolution of 0.1 mg. Detailed protocols for seed germination, planting, growth conditions, block design, and rhizobium inoculation can be found in Supplementary Section 2.

Free-living traits: Growth assays

We measured the growth curves of 32 rhizobia strains at four temperatures (25℃, 30℃, 35℃, and 40℃) in the TY liquid medium. Three days before the experiments, we prepared the rhizobia inocula using a sterilized inoculation loop to streak the glycerol stock on solid TY agar plates, which were incubated at 30℃ for three days until visible colonies appeared. On the day of the experiment, we suspended each colony in a fresh TY medium and diluted the inoculum to a fixed optical density at OD_{600} = 0.1. We inoculated 2 μL of inocula to each well of a 400 μL 96-well clear culture plate (Thermo Scientific flat bottom microwell plates; catalog number 167008) containing 100 μL of TY medium. Each isolate was assayed in two replicates at each temperature. We measured the OD_{600} using a microplate reader (BioTek Epoch 2 Microplate Spectrophotometer) every five minutes for 48 hr with incubation set at a fixed temperature without shaking. We fitted each empirical growth curve to a generalized additive model and extracted the growth rate, lag time, and total yield for each temperature (Supplementary Section 3).

De novo genome assembly and genomic analysis

A detailed description of genome assembly, annotation, and replicon identification are included in Supplementary Section 4. In brief, to prepare genomic DNA for sequencing of *Sinorhizobium* genomes, we obtained cell pellets by streaking glycerol stocks on solid TY agar plates and incubated them at 30℃ for three days, followed by suspension in PBS and centrifugation. The washed pellets were resuspended in 500 μL of DNA/RNA shield for genomic DNA extraction and Oxford Nanopore long-read sequencing (Plasmidsaurus, Eugene, OR, USA). Raw reads were filtered using filtlong (ver. 0.2.1) (Wick and Menzel 2017) to remove the lowest 5% of quality reads, resulting in 38 samples with a median of 104 k reads, median read length of 2.7-5.3 kbp, and coverage estimates ranging from 48-171×. Downsampled reads were assembled into draft genomes using miniasm (ver. 0.3) (Li 2016). With this draft genome, we

re-downsampled the reads to a mean coverage of 100× or skipped this step if the realized coverage falls below 100×. We implemented a second downsampling using filtlong (ver. 0.2.1; options: –target_bases 100×G_d –mean_q_weigth 10) (Wick and Menzel 2017). From these downsampled reads, we then assembled the genomes using the assembler Flye (ver. 2.9.2; options: –meta –nano-corr) (Kolmogorov et al. 2019) and polished with medaka (ver. 1.8.0) (*medaka: Sequence correction provided by ONT Research* 2018), achieving N50 values between 1.54-4.31 Mbp and BUSCO (ver. 5.7.1; options: -m genome -l alphaproteobacteria_odb10) (Simão et al. 2015) completeness scores from 94.2-99.7%, indicating high assembly quality.

We determined the rhizobia taxonomy through two BLAST comparisons against a custom database composed of 18 strains belonging to nine *Sinorhizobium* species (*S. meliloti, S.medicae, S. adhaerens, S. mexicanum, S. canadensis, S. fredii, S. americanum, S. sojae,* and *S. alkalisoli*) (Table S2) and a generic 16S rRNA gene database, respectively. We also determined the replicon identity by first removing those contigs < 1 kbp and choosing the top blast hit against the custom *Sinorhizobium* database at the replicon level (Table S2). The contigs were categorized into five classes: chromosome, pSymA, pSymB, pAcce, or others.

Because *S. medicae* is the most common species in elevation gradient (N=10/15) and *S. meliloti* in urbanization gradient (N=17/23), we performed a pangenomic analysis for each gradient using the genomes of the most common species to avoid the low fraction of core gene orthologs caused by including heterospecifics. The genomes were first annotated using Prokka (ver. 1.14.5; options: -kingdom bacteria and --gcode 11) (Seemann 2014) and the pangenome were inferred using Panaroo (Tonkin-Hill et al. 2020). The results included a set of individual core-gene sequence alignments (based on MAFFT, default in Panaroo) and gene presence-absence matrix that were both used in the phylogenetics analysis.

Phylogenetic, population genetic, and GO enrichment analyses

We inferred the maximum likelihood phylogenies based on the single-copy core-gene alignments. We used a custom python script to concatenate the single-copy core-gene alignments into a single FASTA file. We then computed a maximum likelihood phylogeny using IQ-TREE2 (ver. 2.3.0; options: -nt AUTO -B 1000) with 1000 bootstrap replicates (Minh et al. 2020). We inferred the hierarchical clustering of accessory gene content using R function *hclust* with distance "ward.D2" based on the gene presence/absence matrix. Details are described in the Supplementary Section 4.

We computed the fixation index for each SNP using the R packages *poppr* (Kamvar et al. 2014) and *mmod* (Winter 2012). We used Hedrick's G'_{st} as our fixation index measure which ensures that the upper limit of G'_{st} =1 suggests the populations are fixed for a single allele and there are two alleles across two populations. (Hedrick 2005). From the Panaroo core-gene alignments as mentioned in the previous section, we detected 26042 SNPs in 3169/4701 single-copy core genes (those with SNPs / total) in elevation and 41899 SNPs in 3670/4483 genes in urbanization gradient (Fig. S8). Similarly, we computed an equivalence of fixation index for the accessory gene content, assuming the two allelic states of an accessory gene is its presence and absence. Lastly, we chose the top 5% of loci according to gene-level G'_{st} for Gene Ontology (GO) enrichment analysis, which resulted in 103 among 1872 core genes with GO terms in elevation and 107/2205 in urbanization, and 69/806 accessory genes in elevation and 64/1123 in

urbanization. The results from genes with identical G'_{st}. The GO terms were found using the UniProt "Retrieve/ID mapping" functionality (UniProt Consortium 2023) and then enrichment using the R package *topGO* (Adrian Alexa 2017). We repeated the same analysis for both elevation and urbanization gradients as well. Details are described in Supplementary Section 4.

Statistical analysis

We performed all analyses in R 4.3.2 (R Core Team 2023) with deviation coding ("contr.sum") for categorical variables. We computed effect sizes using the R packages *emmeans* (Lenth 2024) and *effectsize* (Ben-Shachar et al. 2020). Unless stated otherwise, we conducted all analyses and created figures using R packages *tidyverse* (Wickham et al. 2019), *janitor* (Firke 2023), *cowplot* (Wilke 2020), and *ggh4x* (van den Brand 2024). The R environment and package version were stored in the *renv* lockfile for reproducibility (Ushey and Wickham 2023).

We used permutation tests to test significance for all models because generalized linear models did not resolve violations of linear model assumptions (normality and homoscedasticity). Briefly, we performed 1000 permutations for each test to generate a null distribution of the relevant test statistics (e.g, χ^2 in a linear mixed model) by shuffling observations among populations (low- and high- elevation; or urban and suburban) within each gradient and assessed the P-value based on the percentile of the observation against the null distribution (i.e., a two-tailed test). We either performed linear mixed models using *lmer* from the R package *lme4* (Bates et al. 2015), or in the case of nodule count, we used generalized linear mixed models using *glmer* with a Poisson error distribution. Separate tests were conducted for the two gradients, with model specifications described below and detailed in the Supplementary Tables.

To examine whether symbiosis traits differ between populations within a gradient, we fit models with the individual symbiosis traits as response variables, population (low or high elevation; or urban or suburban) as a fixed effect, and rhizobia strain as a random effect. We included the relevant blocking factor (lab group in the *M. sativa* experiment, and experimental block in the *M. lupulina* experiment) as a random effect. We also included the plant maternal family to the *M. lupulina* in the elevation experiment as a random effect. To test whether low- and high-elevation strains responded similar to nitrogen supplementation, we included fixed effects of nitrogen and the population-by-nitrogen interaction and allowed for strain-specific responses to nitrogen (nitrogen-specific random intercepts; *(1|nitrogen:strain)* in R).

To evaluate whether free-living growth traits at 30℃ differ between populations within a gradient, we fit models with free-living growth traits (growth rate, lag time, and yield) as response variables, population as a fixed effect, and strain as a random effect. To evaluate whether growth traits exhibit adaptation to thermal environments, we included fixed effects of temperature and the population-by-temperature interaction, and allowed for strain-specific responses to temperature (temperature-specific random intercepts; *(1|temperature:strain)* in R)), and performed post hoc Tukey tests within temperatures. For permutational multivariate analyses involving multiple dependent variables, we used the *adonis2* function from the package *vegan* (Oksanen et al. 2022) with 1000 permutations. For the *M. sativa* experiment we included four traits: shoot height, nodule number, leaf color, leaf number as the response variables; for the *M. lupulina* experiment, three traits were included: shoot biomass, nodule number, root biomass.

RESULTS

Symbiosis traits differed between low- and high-elevation rhizobia, but not between urban and suburban rhizobia

We first compared the symbiosis traits between rhizobia populations using the dataset generated by the undergraduate laboratory courses in the alternate plant host *M. sativa*. In the elevation gradient, the plants grown with low-elevation rhizobia achieved greater shoot height than those with high-elevation strains (Wald $\chi^2_{df=1}$ = 32.33, P<0.001). This holds true in two leaf traits, but not in nodule number (Table S3) (Fig. 2A), and when these traits are included together as response variables in a PERMANOVA ($F_{df=1}$=37.58, P<0.001). It also holds in categorical traits such as nodule size and shape, but marginally significant in nodule color, resulting in larger, more oval, and pinker nodules in low-elevation rhizobia (Table S3). On the contrary, we did not detect a difference between urban and suburban rhizobia in their host plants' shoot height (Wald $\chi^2_{df=1}$=0.04, P=0.838), in any of other three symbiosis traits (nodule number, leaf number, leaf color) (Table S3) (Fig. 2B), or in a PERMANOVA with all symbiotic traits as response variables ($F_{df=1}$<0.01, P=0.96).

To account for the host's coevolutionary history with rhizobia, the second plant experiment used the source host *M. lupulina* and was performed by experienced graduate students and postdocs. We found that none of the three measured traits (shoot biomass, nodule number, and root biomass) differ between high- versus low-elevation populations, or between suburban versus urban populations (Table S4) (Fig. 2CD), and the results are consistent in the multivariate tests (elevation: $F_{df=1}$=0.72, P=0.49; urbanization: $F_{df=1}$=2.99, P=0.06). Accounting for the magnitude in variation, these results are consistent with the effect size for each trait (Fig. 2 panel right).

If the difference in symbiosis traits between low- and high-elevation strains is attributable to the nitrogen benefit they provide, then this difference should diminish when plants are supplemented with external nitrogen. To test this, we replicated the *M. sativa* experiment with rhizobia from the elevation gradient with nitrogen fertilizer (Fig. 3). As expected, the observed trait divergence in the absence of nitrogen disappeared when the plants were supplemented with nitrogen fertilizer for each trait (shoot height, leaf color, and leaf number) except for nodule number (Table S5), as well as in a PERMANOVA ($F_{df=1}$=118.2, P<0.001) (Fig. 3).

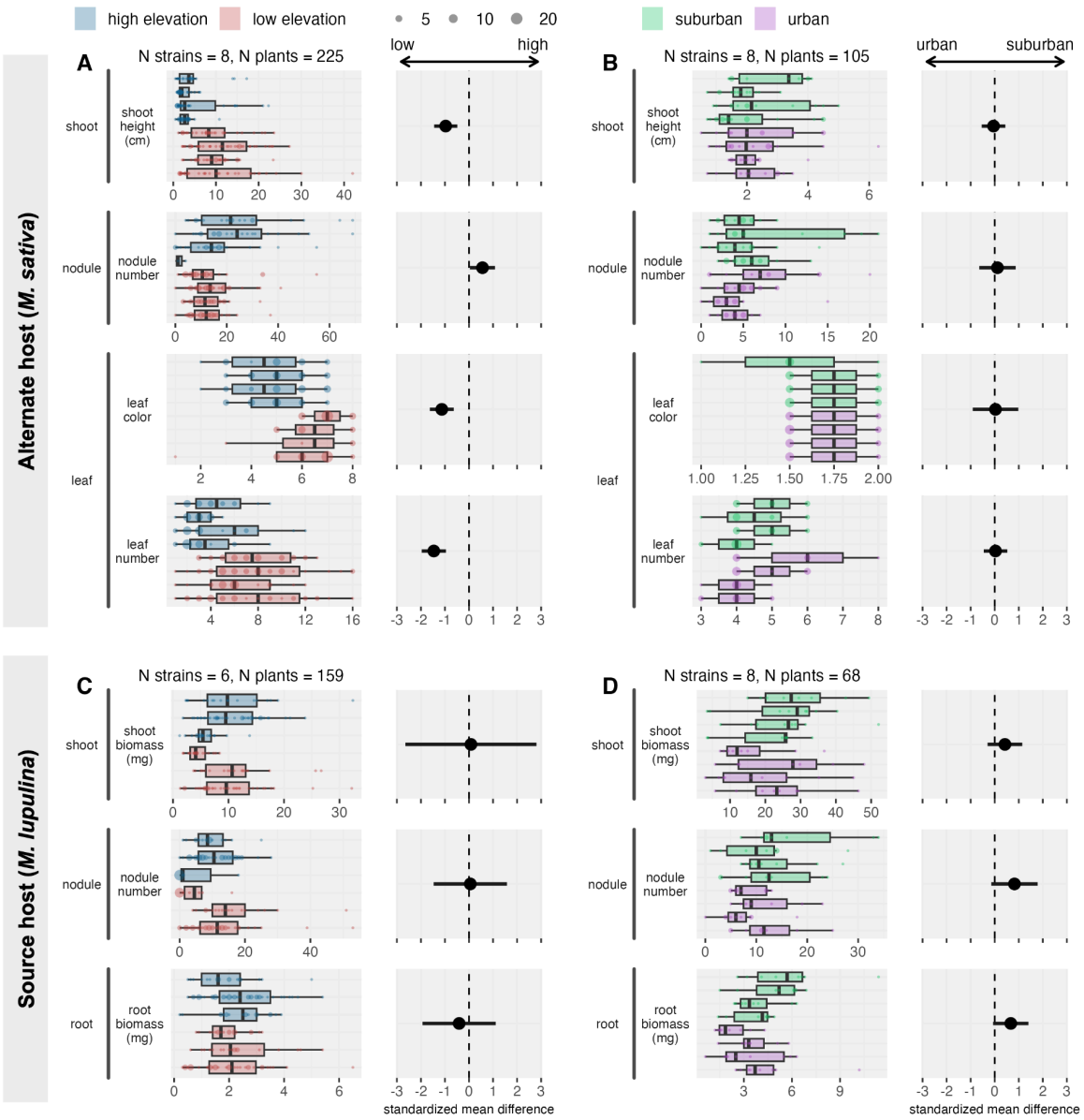

Figure 2. Symbiosis trait differences between low- and high elevation rhizobia (left) and urban and suburban rhizobia (right) in the two plant experiments. (A) and (B) represent the first plant experiment using the alternate host *M. sativa*. (C) and (D) represent the second experiment using the source host *M. lupulina*. In the box-and-whisker plots, each box represents one rhizobia strain, and the raw data are represented by points, the size of which is scaled by the number of overlaid points. The right part of each panel shows the standardized mean difference (Cohen's d) between populations at opposing ends of the relevant gradient, where the error bar represents the 95% confidence intervals. Note that three and two strains are used across both hosts in elevation and urbanization gradients, respectively. Also note that experiments shown in panels A and B were carried out in different semesters; panels C and D were carried out simultaneously under the same condition.

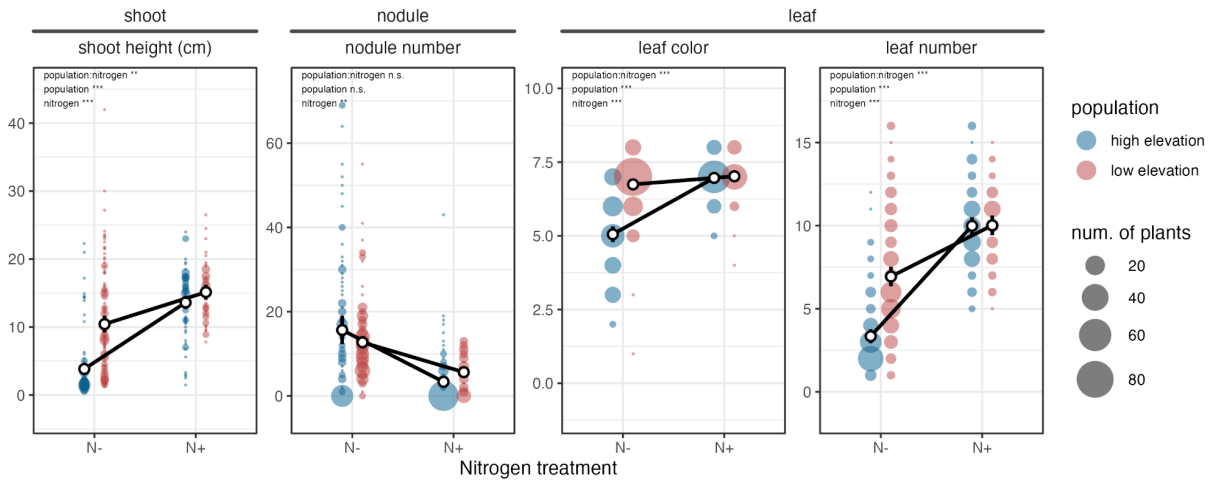

Figure 3. Nitrogen supplement reduces the difference in symbiosis traits between low- and high-elevation rhizobia. The white points represent the averages and the error bars are standard errors. The nitrogen-free data are identical to Fig 2A.

Free-living growth traits do not differ across either the elevation or urbanization gradients

We examined the free-living traits with growth curve assays in liquid culture. We did not detect a difference in growth rates at 30℃ ($\chi^2_{df=1}$ = 0.93, P=0.37), between either the high- and low-elevation rhizobia or suburban and urban rhizobia ($\chi^2_{df=1}$ = 0.31, P=0.57) (Table S6) (Fig. 4). When incorporating the three growth traits at 30℃: growth rate, lag time, and yield into a multivariate analysis, the pattern holds in both the elevation gradient ($F_{df=1}$=1.27, P=0.27) and urbanization gradient ($F_{df=1}$=1.77, P=0.18).

Next, we tested for thermal local adaptation. Temperature differs across both gradients, although more in the elevation gradient (mean difference between high and low-elevation sites: 3.1 ℃ in daily maximum; Methods) than in the urbanization gradient (mean difference between suburban and urban sites: 0.38 ℃ in daily maximum) (Figs. S1-2). Therefore, we hypothesized that low-elevation rhizobia would grow better than high-elevation rhizobia at high temperatures, and vice versa at low temperatures. Similarly we hypothesize that urban rhizobia would grow faster than suburban rhizobia at high temperatures, and vice versa at low temperatures. However, the temperature × population interaction were not significant in any of the three growth traits in either gradient (Table S7-8) (Fig. 4). High-elevation populations outperformed low-elevation populations at the coolest temperature (growth rate and yield at 25 ℃), but not the other way around. This result is partially consistent with rhizobia adaptation to the local thermal environments in the elevation gradient (Fig. 4A). There was no strong evidence for thermal local adaptation in the urbanization gradient. Although urban populations grew to a higher yield than suburban populations at the warm temperatures (30℃, 35℃, and 40℃), there was no difference between them at the coolest temperature (Fig. 4B).

It has been hypothesized that tradeoffs between symbiosis and free-living life stages maintain variation in rhizobia traits. To test this hypothesis, we tested for a genetic correlation between symbiotic and free-living growth traits for *M. sativa* (N=13 strains) and *M. lupulina* (N=11 strains) experiments. After applying a Bonferroni correction for multiple testing, there were no significant genetic correlations among individual free-living and symbiosis traits, nor among two principal components that summarized free-living and symbiosis traits, respectively (Fig. S3).

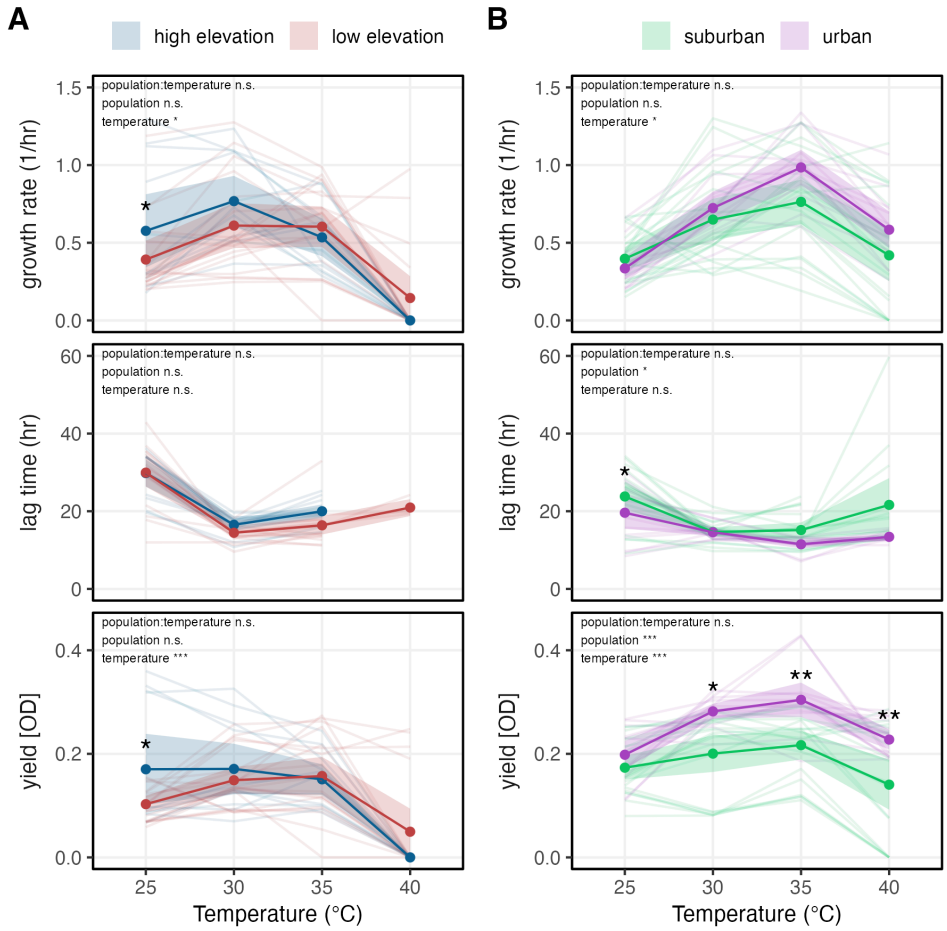

Figure 4. Divergence in free-living traits across the elevation and urbanization gradients. We measured three growth traits of free-living rhizobia from (A) elevation and (B) urbanization gradients at four temperatures. The semitransparent lines represent technical replicates (two per strain). The solid dots denote the mean values and shades represent the 95% confidence intervals. The statistical significance of each term of a linear mixed model is shown at the top left and Table S7. The asterisks above each temperature represent the statistical significance of pairwise comparison between populations using post hoc Tukey test: P<0.001***, P < 0.01**, P<0.05* (Table S8).

Relatively little genetic divergence in the core and accessory genomes across both gradients

We performed a series of population genomic analyses on the most common species within each gradient: 10 *S. medicae* strains in elevation gradient and 17 *S. meliloti* strains in urbanization gradient (Fig. S4). The phylogeny shows that there is little evidence for genome-wide divergence between contrasting populations in either environmental gradient: low-and high-elevation strains do not consistently segregate into different clades, nor do urban and suburban strains. This pattern holds true in an analysis based on sequence variation (SNPs) in single-copy core genes, as well as with variation in accessory gene content (Fig. 5). The lack of genetic population divergence is consistent at the replicon level (i.e., chromosome, pSymA, pSymB, and pAcce) (Fig. S5). The tree topology between the core gene tree and gene content tree is more congruent than the random expectation (Fig. S6).

We used two other metrics of pairwise whole-genome distance including average nucleotide identities (ANI) and the signatures of subsequences of length k (k-mers). Tree topology analysis shows that these metrics are congruent with one another, validating the finding where rhizobia from opposite ends of each gradient are not significantly divergent genome wide (Fig. S6-7).

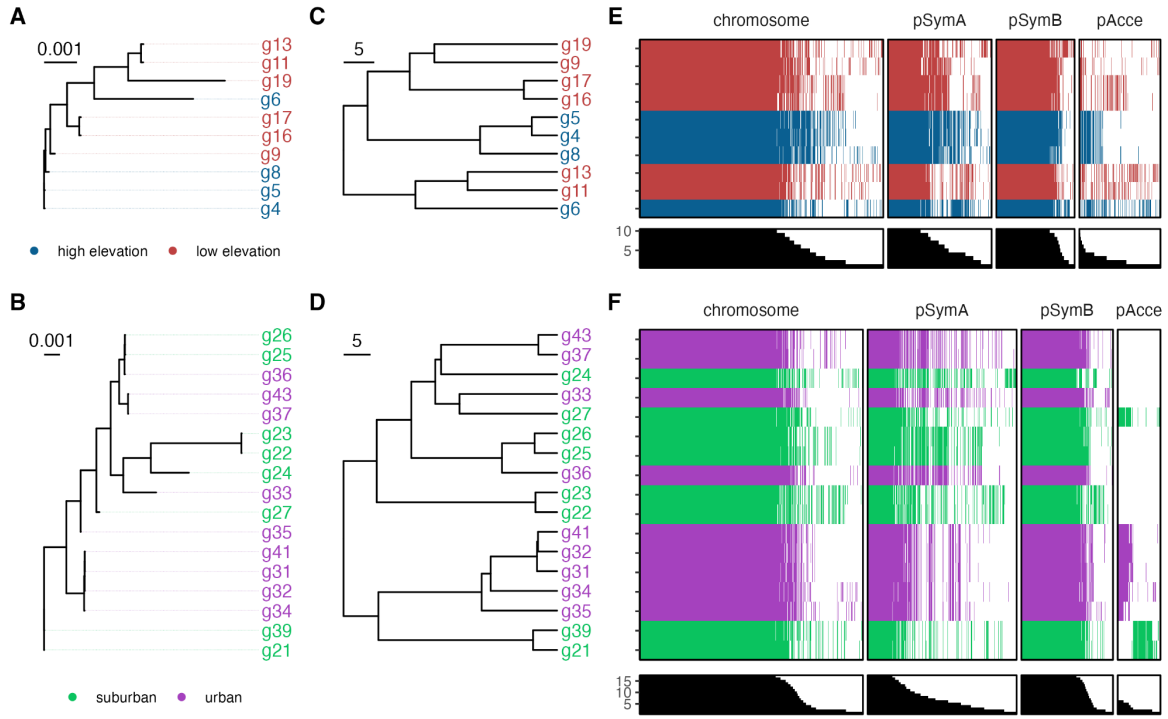

Figure 5. Core-gene phylogeny and gene content variation show no clear differentiation across the two environmental gradients. (A) and (B) show the phylogeny inferred from 4702 single-copy core genes for elevation and 4484 genes for urbanization gradient, respectively. The single-copy core-gene phylogeny represents the consensus tree of 1000 bootstrapped trees using the maximum likelihood approach. The branches with < 95% bootstrap support are collapsed. (C) and (D) denote the hierarchical clustering (ward.D2) of gene content variation inferred from the presence and absence of 4899 accessory genes for elevation and 5055 accessory genes for urbanization gradient. (E) and (F) show the presence and absence of orthologous gene clusters of the 10 rhizobial genomes for elevation gradient and 17 for urbanization gradient. The panel below each heatmap represents the number of genomes that share the same orthologous gene cluster.

Despite the lack of genome-wide differentiation, population divergence could still occur at loci under geographic site-specific selection within each gradient. To examine which loci are most differentiated between the populations at opposite ends of each gradient, we computed the fixation index (Hedrik's G'_{st}) for the SNPs in single-copy core genes (4701 in elevation and 4483 in urbanization). Average G'_{st} in the core genome was higher in the elevation gradient than the urbanization gradient (two-sample t-test: t = 4.83, df = 43362, P<0.001) (Fig. S8). G'_{st} outliers (top 5% of annotated genes in each gradient) included *ydhY* and *fdxB* in elevation gradient, both are ferredoxin-like proteins that may serve as a direct electron donor to the nitrogenase; in urbanization gradient, the top differentiated genes are *paoC*, an aldehyde oxidoreductase subunit and *nanR*, a transcriptional repressor. GO enrichment analysis indicates that G'_{st} outliers are enriched for amino acid and small molecule metabolism in the elevation gradient, and for transport and localization in urbanization gradient (Fig. S9). No canonical *Sinorhizobium* symbiosis and nitrogen fixation genes (e.g., *nif*, *nod*, and *fix*) were among the top 5% G'_{st} genes in either gradient.

We performed a similar analysis on gene content variation to test whether different accessory genes were enriched at opposite ends of the two gradients (4899 accessory genes in elevation and 5055 in urbanization). The elevation gradient exhibited a higher level of G'_{st} across the accessory genome than the urbanization gradient (two-sample t-test: t= 15.39, df = 7904, P<0.001) (Fig. S10). GO enrichment analysis shows that the top 5% outliers are enriched for functions related to organelle, nucleic acid binding, and glycan metabolism in elevation gradient, and monocarboxylic acid catabolism, antioxidant activity, and intracellular anatomical structure in urbanization gradient (Fig. S11).

Gene flow is more restricted in the urbanization gradient than the elevation gradient

Finally, we evaluated whether the difference in divergence along the two gradients can be explained by differences in gene flow. Because we detected greater phenotypic and genomic divergence in the elevation gradient than in the urbanization gradient, we hypothesized that greater gene flow may constrain divergence in the urbanization gradient. If gene flow constrains divergence in the urbanization gradient, isolation by distance across the urbanization gradient should be weaker than in the elevation gradient (Fig. 6). Contrary to this expectation, isolation by distance was significantly stronger in the urbanization gradient than in the elevation gradient for both the chromosome (gradient × distance interaction, SNP: $\chi^2_{df=1}$ =55.4, P <0.001; GCV: $\chi^2_{df=1}$ =29.3, P<0.001) and pSymB (SNP: $\chi^2_{df=1}$ =10.4, P=0.001; GCV: $\chi^2_{df=1}$ =20.4, P<0.001), but not in pSymA (SNP: $\chi^2_{df=1}$ =1.08, P=0.29; GCV: $\chi^2_{df=1}$ =1.13, P=0.28). In the urbanization gradient, we detected significant isolation-by-distance in core genes and gene content variation, with the exception of pSymA (Fig. 6BD). Within the urbanization gradient, the urban strains were also more genetically similar to each other than those within the suburban population (SNP: t=4.2, P < 0.001; GCV: t=4.4, P <0.001) (Fig. 6BD). By contrast, we did not detect significant isolation-by-distance for any genomic replicon in the elevation gradient, either as measured by the sequence similarity of core genes (Fig. 6A) or in gene content similarity (Fig. 6C).

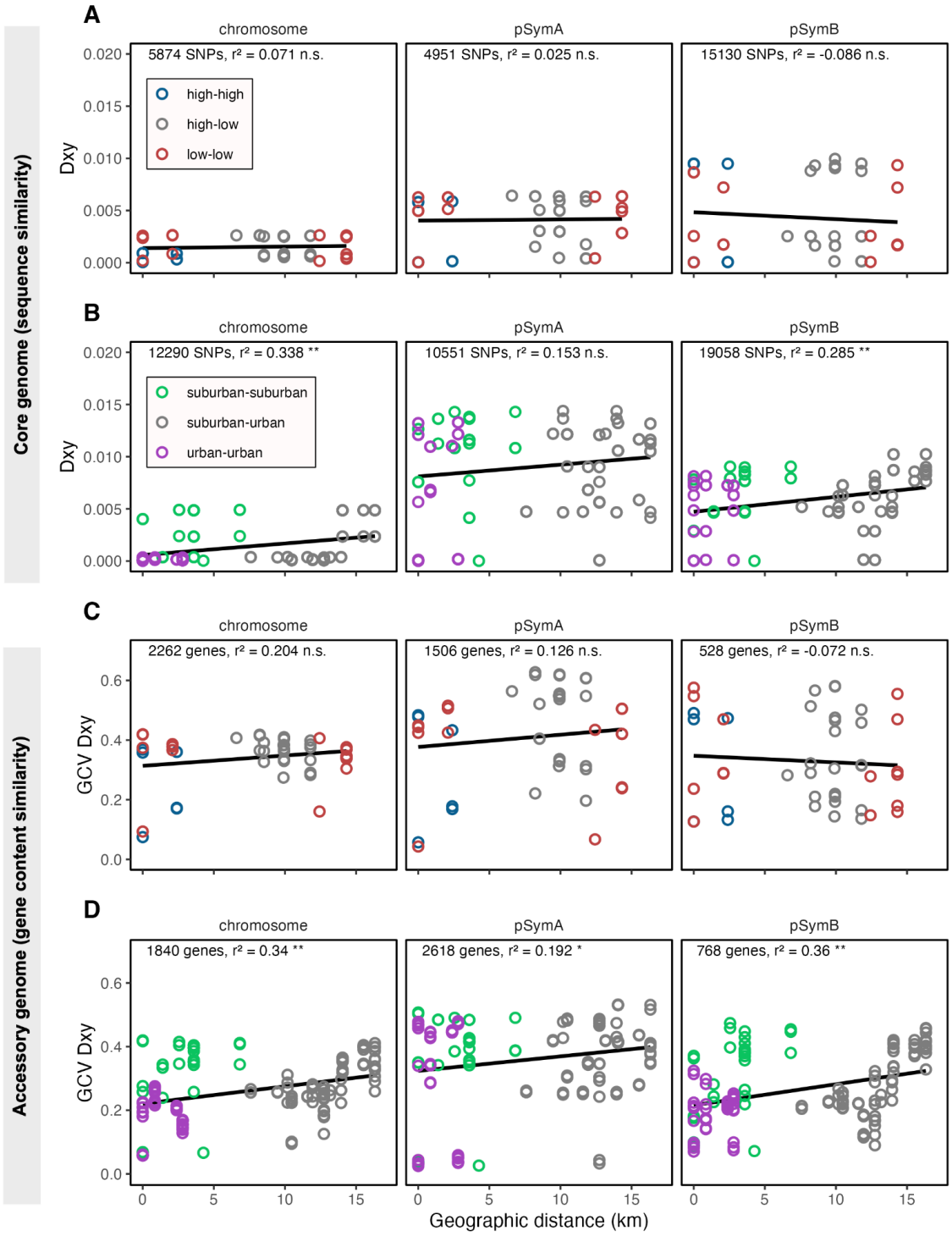

Figure 6. Genetic distance shows distinct patterns with geographic distance depending on the replicon types and gradients. (A) and (B) show the nucleotide distance of single-copy core genes between two rhizobia genomes in elevation and urbanization gradients, respectively. (C) and (D) show the Manhattan distance of gene content between two genomes in elevation and urbanization gradients, respectively. Mantel tests are reported in each panel shown with the number of loci (SNPs or accessory genes), and statistical significance (n.s. non-significant, * $p < 0.05$, ** $p < 0.01$). Note that the geographic distance spanned by both gradients is nearly identical (~15 km in both cases).

DISCUSSION

Environmental gradients serve as invaluable natural experiments for studying evolution in the wild, yet they have rarely been leveraged to explore microbial evolution. We investigated phenotypic and genetic differentiation of *Sinorhizobium* rhizobia across two environmental gradients: elevation, a longtime model in evolutionary ecology, and urbanization, an emerging model gradient. We found that symbiosis traits diverged in the elevation gradient, but not urbanization gradient. The difference in symbiosis traits between low- and high-elevation rhizobia diminished with supplemented nitrogen, consistent with that the difference results from the rhizobia-provided nitrogen. We also found the urbanization gradient exhibits stronger genetic differentiation than elevation gradient, despite their comparable geographic scales.

Symbiosis traits, but not free-living traits, diverged across elevations

We found that high-elevation rhizobia provided less benefit than low-elevation rhizobia in the alternate host plant *M. sativa*. Plants inoculated with low-elevation rhizobia grew taller than plants inoculated with high-elevation rhizobia (Fig. 2). Two observations suggest that this effect is attributable to the nitrogen benefits that low-elevation rhizobia provide. First, plants inoculated with low-elevation rhizobia produced significantly larger nodules that also tended to be pinker (Table S3), consistent with more active nitrogen fixation. Second, nitrogen supplementation erased the growth advantage of plants inoculated with low-elevation rhizobia (Fig. 3).

Given that the phenotypic differences between low- and high-elevation strains implicated nitrogen fixation, we hypothesized that the core *Sinorhizobium* nodulation and nitrogen fixation genes (*nif*, *nod*, and *fix*) would exhibit genetic divergence across elevations. While none of these genes were F_{st} outliers, the two top differentiated genes, *ydh* and *fdx*, are both implicated in nitrogen fixation. Both are putative ferredoxin-like proteins that regulate electron transfer, and the latter is necessary for nitrogen fixation in *S. meliloti* (Klipp et al. 1989) and other diazotrophs (Addison et al. 2024) suggest that regulation of nitrogen fixation might have differentiated across elevation. It is worth noting that genome-wide genetic divergence between low- and high-elevation populations was relatively weak, as measured by sequence similarity in core genes and by gene content similarity (Fig. 6). This indicates that the divergence in symbiosis traits occurred despite the homogenizing effect of gene flow.

It is important to note that the symbiotic trait differences between low- and high-elevation rhizobia were only apparent on the alternate host, *M. sativa* (Fig. 2A). In the source host, *M. lupulina*, there were no significant differences in any symbiotic traits (Fig. 2C). The absence of symbiotic trait differences in the source host could indicate that shared (co)evolutionary history between host and symbiont is an important determinant of the benefit provided by rhizobia. However, a previous study found no evidence for co-adaptation between *Sinorhizobium* rhizobia and *M. lupulina* plants in northeastern North America (Harrison et al. 2017*a*). Therefore, we think that a more likely explanation for the discrepancy between the experiments in source and alternate hosts in our study is statistical power. The *M. lupulina* experiment included fewer rhizobia strains and fewer replicates per strain, weakening our power to detect a difference between low- and high-elevation rhizobia.

Why might selection favor more cooperative rhizobia in low-elevation populations? The mutualism literature suggests at least three hypotheses. First, rhizobia evolve to provide more nitrogen under

nitrogen-limited conditions (Weese et al. 2015; Klinger et al. 2016). If soil nitrogen is scarce at low elevations, low-elevation rhizobia would experience stronger selection on their nitrogen-fixation ability. Second, rhizobia functional diversity influences the host plant's ability to selectively penalize ("sanction") uncooperative strains (Akçay and Simms 2011; Sachs et al. 2018; Westhoek et al. 2021). If rhizobia functional diversity is greater at low elevations, low-elevation plants could be more successful at sanctioning uncooperative strains. Biodiversity is generally higher at lower elevations due to greater area, productivity, and topographic complexity (Rahbek et al. 2019). However, it is unclear whether the altitudinal biodiversity gradient holds for functional diversity *within* species. Third, species interactions tend to be more mutualistic under stressful conditions, a phenomenon known as the stress gradient hypothesis (Bertness and Callaway 1994; O'Brien et al. 2018). If low-elevation environments are more stressful than high-elevation environments for plants and rhizobia, it could favor cooperative rhizobia. However, it isn't immediately apparent why this might be the case in our system.

In contrast to the symbiosis traits, our growth assays did not detect divergence between low- and high-elevation rhizobia in free-living growth traits. The lack of divergence could indicate that selection on free-living traits is similar in low- and high-elevation sites, or that our assay temperatures did not capture ecologically relevant thermal conditions. Alternatively, it is possible that the benign environment of our experimental assays masked existing differences in free-living growth. Our assays used rich medium, stable temperature settings, and single-strain cultures—conditions that do not accurately mirror the physically, chemically, and biologically heterogeneous soil environments (Kent and Triplett 2002; Baveye et al. 2018). High-elevation populations did perform significantly better at 25℃ than their low-elevation counterparts, suggestive of weak thermal local adaptation. Furthermore, the gene ontology analysis found that genes that are most divergent between low- and high-elevation populations are enriched for organic acid and small molecule metabolism, which are relevant to rhizobia competition in the rhizosphere (Wielbo et al. 2007). The fitness consequences of these differentiated genes should be further examined using multi-strain competition assays.

Neither symbiosis nor free-living traits diverged along the urbanization gradient

We did not detect divergence between urban and suburban populations in most of the symbiosis or free-living traits in our study. We initially expected urban strains to be less beneficial for their host plant than suburban strains because nitrogen is prevalent in urban soils (Zhang et al. 2012), and nitrogen disposition favors less mutualistic rhizobia (Weese et al. 2015; Regus et al. 2017). However, symbiosis traits were not different between urban and suburban populations.

Second, we expected urban rhizobia to outperform suburban rhizobia at high temperatures in the free-living growth assay. The urban heat island effect often results in thermal adaptation to ectothermic organisms living in the cities (Campbell-Staton et al. 2020). Our thermal growth assays did not support this hypothesis. The lack of thermal local adaptation may be due to the marginal temperature difference between our urban and suburban sites (1℃ in summer daily maximum) (Fig. S2), which may not be large enough to drive differences in selection on thermal performance. Instead, we found that urban rhizobia achieved higher yields than suburban rhizobia at three of the four growth temperatures. This difference in free-living growth was the only phenotypic difference that we detected between urban and suburban strains. It suggests the hypothesis that nutrient enrichment in urban soils —urban nitrogen deposition

(Zhang et al. 2010) and organic carbon in urban greenspaces (Guo et al. 2024)favors metabolically capable free-living microbes.

Few studies have explicitly investigated the evolutionary consequences of urbanization on mutualistic symbionts (Irwin et al. 2020)(Diamond and Martin 2021)(Irwin et al. 2020). As a result, it is not clear whether our results—minimal phenotypic divergence —are representative of microbial evolutionary responses to urbanization. Our urbanization gradient is an urban-suburban gradient, which might exhibit a weaker contrast than an urban-rural gradient. Indeed, a legume species along an urbanization gradient showed distinct nitrogen acquisition strategies between suburban and rural regions (Murray-Stoker and Johnson 2021), suggesting that ecological differences in these areas can impact the legume-rhizobia mutualism.

Finally, isolation by distance was significantly stronger in the urbanization gradient than the elevation gradient, despite their comparable geographic scales. These observations suggest that habitat fragmentation in urban environments impacts the genetic structure of microorganism populations, just as in macroorganisms (Miles et al. 2019). Furthermore, urban rhizobia were more genetically similar to one another than suburban rhizobia were to each other, as measured both by sequence variation in core genes and by accessory gene content (Fig. 6). This pattern is consistent with previous work in an invertebrate system, which also found that genetic similarity among urban populations is high (Miles et al. 2018). This pattern could arise from contemporary adaptation from standing genetic variation, as is expected to be common in evolution in cities (Johnson and Munshi-South 2017). However, this interpretation is not consistent with the weak phenotypic differences between urban and suburban strains in our experiment.

A roadmap for studying the evolutionary ecology of wild microbes

We summarized three emerging themes and questions for studying the evolutionary ecology of wild microbes and discussed the possible avenues to approach them, inspired by animal and plant studies.

What are the dominant agents of selection? Recent success integrating common garden, molecular, and genomic approaches have illuminated how wild microbes adapt to strong selective pressures across diverse landscapes, such as toxic heavy metals (Kehlet-Delgado et al. 2024). However, similar to macroorganisms, microbes are likely to encounter multiple selection agents acting on various traits throughout their life stages in the wild. This underscores the importance of understanding the diversity of selective pressures that operate on microorganisms in ecologically realistic environments (Wadgymar et al. 2022). To begin addressing this challenge, we can leverage established approaches to multivariate phenotypic characterization in microbial metabolism and host plant phenome (Yang et al. 2020; Calvert et al. 2024; Odriozola et al. 2024), and examine genetically based phenotypic variation along environmental gradients, which naturally harbors multiple agents of selection. Our common garden experiments examining multiple traits across rhizobia life stages advocated for the multivariate phenotypic approach (Santangelo et al. 2020), as only components in one life aspect (i.e., yield in free-living growth) but not all traits differentiated along urbanization gradient.

What contribution does accessory gene content variation have to phenotypic variation and population structure? The fluidity of prokaryotic genomes—i.e., the variation in gene content—is directly connected

to, and continuously influenced by, ecological and evolutionary processes (Azarian et al. 2020). In addition to nucleotide polymorphism, gene content variation is instrumental in inferring ecologically cohesive populations (Arevalo et al. 2019) and understanding contemporary adaptation (Choudoir et al. 2023), both of which are driven by extensive horizontal gene transfer. Recent advancements in low-cost long-read sequencing enable reference-free genome assemblies, which are crucial for exploring genetic diversity in accessory genomes in response to ecological selection.

Leveraging discovery-based research in introductory biology labs to explore rhizobia evolution

Our large common garden experiments to measure rhizobia symbiotic traits in *M. sativa* were made possible by deploying the experiment as discovery-based research in the introductory biology labs at the University of Pennsylvania. This collaboration capitalized on the fact that environmental gradients are an excellent model system for teaching evolutionary ecology. These gradients are natural experiments that lend themselves to straightforward hypothesis tests. Furthermore, culturable microbial symbionts, especially those associated with plants, are tractable to deploy in undergraduate lab classes. The rapid generation times of small plants and microbes allow these experiments to be conducted within a semester. And while most microbial traits are invisible to the naked eye (or require the capacity to precisely measure growth curves), microbial effects on plant hosts can be measured with common classroom tools. The pedagogical benefits of discovery-based undergraduate research have been well established, such as improving the students' competencies in scientific practices and appreciation of the processes of scientific research (Linn et al. 2015). Incorporating original research into lab classes further mitigates the gap between undergraduate courses and the research experiences often required by graduate school admissions (Bangera and Brownell 2014). Exposure to research early in college (i.e., in freshman year) is broadly beneficial to both STEM and non-STEM-focused students (Stanford et al. 2017)

A major scientific benefit is that deploying original research in large introductory biology courses increases the scale of experiments that are possible. However, generating publishable data from course-based research can be challenging (Buchanan and Fisher 2022). We designed the CURE in this study specifically to address these challenges. The partnership between the introductory biology lab course and the research laboratory allowed each student group to develop their own hypothesis and predictions based on a single predetermined research question derived from previous work in the research lab and informed by their review of the primary literature. Instead of allowing each of the approximately 140 student groups in the course to design their own experiment—which would have been logistically and financially impossible— we asked students to identify controls, variables, and replicates in the experiment that we had already designed, and to reflect on any experimental obstacles or problems encountered. The experiment was specifically designed to minimize logistical and financial obstacles, such as the need for greenhouse or growth chamber space and the watering of plants. We grew *M. sativa* plants in agar slants within capped test tubes, such that hundreds of test tubes containing seedlings can be grown under standard fluorescent lights on two shelves of a cart in the teaching laboratory classroom for five weeks without watering. Furthermore, students can easily visualize, count and measure root nodules through the glass tubes without removing plants from the agar. Finally, this experiment was designed specifically to dovetail with a lab module that we wrote to teach data analysis in R to introductory biology students. Since each student group measured many different traits and had access to the pooled data from all the groups in the course, each group had the flexibility and statistical power to test the hypothesis with

student-generated data. After instruction on data analysis in R, we allowed each student group to decide which data to analyze and how to analyze it for their oral presentation, where students communicated their results and conclusions. In summary, this approach allows for a scalable, student-driven research experience that benefits both the students and the research lab.

Conclusion

This study tested microbial population differentiation across two natural environmental gradients. Our findings on the trait and genetic divergence of a plant symbiont show that symbiosis trait divergence occurs along elevation gradient whereas non-adaptive processes like gene flow might be rather important in urbanization gradient. We highlight a holistic understanding of microbial evolution with regard to the ecological traits across life stages, the genetic signature across genomic levels and regions, and the evolutionary mechanisms across natural gradients. With increasing interest in urbanization gradients and advanced high-throughput methodologies, we anticipate more studies investigating microbial evolution across multiple environmental gradients in the future.

ACKNOWLEDGEMENT

We thank the members of the Wood lab for their feedback on this manuscript. We would like to thank the staff at Mountain Lake Biological Station for their support on the fieldwork. Growth chamber logistical support was provided by the University of Pennsylvania's Greenhouse support staff members Kathyrn Butler and Samara Gray. We would also like to thank the teaching lab staff, especially Svetlana Kozik, for help with development and preparation of the teaching lab materials for CURE, and the undergraduate students participating in a CURE in BIOL1102 course lab during the Spring 2023 and Spring 2024 semesters at the University of Pennsylvania for assessing the *M. sativa* traits. This project was supported by NSF-DEB 2118397 awarded to C.W.W, and startup funding from the University of Pennsylvania. C.-Y. C. was supported by The Data-Driven Discovery Postdoctoral fellowship by the University of Pennsylvania. Gerardo Bencosme was supported by the Mountain Lake REU program.

Statement of Authorship

C.-Y.C. and C.W.W conceived the idea and designed the study. C.-Y.C., T.T.-B., and L. J. R. performed the experiments. C.-Y.C., T.T.-B., L. J. R. and J.L.R. collected the data. C.-Y.C., M.B.C., and C.W.W discussed the results and drafted the paper. C.-Y.C. and C.W.W. wrote the final version of the paper.

Data and Code Availability

Data and code needed to replicate results contained in this article are available on the GitHub repository Chang-Yu-Chang/rhizobia-genomics. The raw sequence data are available upon acceptance.

Literature Cited

Addison, H., T. Glatter, G. K A Hochberg, and J. G. Rebelein. 2024. Two distinct ferredoxins are essential for nitrogen fixation by the iron nitrogenase in Rhodobacter capsulatus. mBio 15:e0331423.
Adrian Alexa, J. R. 2017. topGO. Bioconductor.

Akçay, E., and E. L. Simms. 2011. Negotiation, sanctions, and context dependency in the legume-Rhizobium mutualism. The American naturalist 178:1–14.
Alberti, M., E. P. Palkovacs, S. D. Roches, L. D. Meester, K. I. Brans, L. Govaert, N. B. Grimm, et al. 2020. The complexity of urban Eco-evolutionary dynamics. Bioscience 70:772–793.
Arevalo, P., D. VanInsberghe, J. Elsherbini, J. Gore, and M. F. Polz. 2019. A reverse ecology approach based on a biological definition of microbial populations. Cell 178:820–834.e14.
Azarian, T., I.-T. Huang, and W. P. Hanage. 2020. Structure and dynamics of bacterial populations: Pangenome ecology. Pages 115–128 *in*The Pangenome. Springer International Publishing, Cham.
Bangera, G., and S. E. Brownell. 2014. Course-based undergraduate research experiences can make scientific research more inclusive. CBE life sciences education 13:602–606.
Barker, D. G., T. Pfaff, D. Moreau, E. Groves, S. Ruffel, M. Lepetit, S. Whitehand, et al. 2006. Growing M. truncatula: choice of substrates and growth conditions. The Medicago truncatula handbook 1–26.
Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models Usinglme4. Journal of statistical software 67:1–48.
Baveye, P. C., W. Otten, A. Kravchenko, M. Balseiro-Romero, É. Beckers, M. Chalhoub, C. Darnault, et al. 2018. Emergent properties of microbial activity in heterogeneous soil microenvironments: Different research approaches are slowly converging, yet major challenges remain. Frontiers in microbiology 9:1929.
Ben-Shachar, M., D. Lüdecke, and D. Makowski. 2020. Effectsize: Estimation of effect size indices and standardized parameters. Journal of open source software 5:2815.
Bertness, M. D., and R. Callaway. 1994. Positive interactions in communities. Trends in ecology & evolution 9:191–193.
Buchanan, A. J., and G. R. Fisher. 2022. Current status and implementation of science practices in course-based undergraduate research experiences (CUREs): A systematic literature review. CBE life sciences education 21:ar83.
Burghardt, L. T. 2020. Evolving together, evolving apart: measuring the fitness of rhizobial bacteria in and out of symbiosis with leguminous plants. The new phytologist 228:28–34.
Burghardt, L. T., and G. C. diCenzo. 2023. The evolutionary ecology of rhizobia: multiple facets of competition before, during, and after symbiosis with legumes. Current opinion in microbiology 72:102281.
Calvert, M. B., M. Hoque, and C. W. Wood. 2024. Genotypic variation in resource exchange, use, and production traits in the legume-rhizobia mutualism. Ecology and evolution 14:e70245.
Campbell-Staton, S. C., K. M. Winchell, N. C. Rochette, J. Fredette, I. Maayan, R. M. Schweizer, and J. Catchen. 2020. Parallel selection on thermal physiology facilitates repeated adaptation of city lizards to urban heat islands. Nature ecology & evolution 4:652–658.
Caro, T. A., J. McFarlin, S. Jech, N. Fierer, and S. Kopf. 2023. Hydrogen stable isotope probing of lipids demonstrates slow rates of microbial growth in soil. Proceedings of the National Academy of Sciences of the United States of America 120:e2211625120.
Chase, A. B., C. Weihe, and J. B. H. Martiny. 2021. Adaptive differentiation and rapid evolution of a soil bacterium along a climate gradient. Proceedings of the National Academy of Sciences of the United States of America 118:e2101254118.
Chaudhary, V. B., S. Nolimal, M. A. Sosa-Hernández, C. Egan, and J. Kastens. 2020. Trait-based aerial dispersal of arbuscular mycorrhizal fungi. The new phytologist 228:238–252.
Choudoir, M. J., A. Barberán, H. L. Menninger, R. R. Dunn, and N. Fierer. 2018. Variation in range size and dispersal capabilities of microbial taxa. Ecology 99:322–334.
Choudoir, M. J., and K. M. DeAngelis. 2022. A framework for integrating microbial dispersal modes into soil ecosystem ecology. iScience 25:103887.
Choudoir, M. J., A. Narayanan, D. Rodriguez-Ramos, R. Simoes, A. Efroni, A. Sondrini, and K. M. DeAngelis. 2023. Pangenomes reveal genomic signatures of microbial adaptation to experimental soil warming. bioRxiv.
Clausen, J., D. Keck, and W. M. Hiesey. 1948. Experimental studies on the nature of species. III.

Environresponses of climatic races of Achillea.
Custer, G. F., L. Bresciani, and F. Dini-Andreote. 2022. Ecological and evolutionary implications of microbial dispersal. Frontiers in microbiology 13:855859.
Cvijović, I., A. N. Nguyen Ba, and M. M. Desai. 2018. Experimental studies of evolutionary dynamics in microbes. Trends in genetics: TIG 34:693–703.
Denison, R. F., and E. T. Kiers. 2011. Life histories of symbiotic rhizobia and mycorrhizal fungi. Current biology: CB 21:R775–85.
Diamond, S. E., and R. A. Martin. 2021. Evolution in cities. Annual review of ecology, evolution, and systematics 52:519–540.
diCenzo, G. C., and T. M. Finan. 2017. The divided bacterial genome: Structure, function, and evolution. Microbiology and molecular biology reviews: MMBR 81:10.1128/mmbr.00019–17.
Endler, J. A. 1977. Geographic variation, speciation and clines. (MPB-10), volume 10. Monographs in Population Biology. Princeton University Press, Princeton, NJ.
Epstein, B., M. J. Sadowsky, and P. Tiffin. 2014. Selection on horizontally transferred and duplicated genes in sinorhizobium (ensifer), the root-nodule symbionts of medicago. Genome biology and evolution 6:1199–1209.
Epstein, B., and P. Tiffin. 2021. Comparative genomics reveals high rates of horizontal transfer and strong purifying selection on rhizobial symbiosis genes. Proceedings. Biological sciences 288:20201804.
Evans, S. E., L. P. Bell-Dereske, K. M. Dougherty, and H. A. Kittredge. 2020. Dispersal alters soil microbial community response to drought. Environmental microbiology 22:905–916.
Fenchel, T., and B. J. Finlay. 2004. The Ubiquity of Small Species: Patterns of Local and Global Diversity. Bioscience 54:777–784.
Fierer, N., C. M. McCain, P. Meir, M. Zimmermann, J. M. Rapp, M. R. Silman, and R. Knight. 2011. Microbes do not follow the elevational diversity patterns of plants and animals. Ecology 92:797–804.
Firke, S. 2023. janitor: Simple Tools for Examining and Cleaning Dirty Data.
Fuhrman, J. A., J. A. Steele, I. Hewson, M. S. Schwalbach, M. V. Brown, J. L. Green, and J. H. Brown. 2008. A latitudinal diversity gradient in planktonic marine bacteria. Proceedings of the National Academy of Sciences of the United States of America 105:7774–7778.
Galibert, F., T. M. Finan, S. R. Long, A. Puhler, P. Abola, F. Ampe, F. Barloy-Hubler, et al. 2001. The composite genome of the legume symbiont Sinorhizobium meliloti. Science 293:668–672.
Geddes, B. A., J. Kearsley, R. Morton, G. C. diCenzo, and T. M. Finan. 2020. The genomes of rhizobia. Pages 213–249 *in* P. Frendo, F. Frugier, and C. Masson-Boivin, eds. Advances in Botanical Research, Advances in botanical research (Vol. 94). Elsevier.
Giraud, T., B. Koskella, and A.-L. Laine. 2017. Introduction: microbial local adaptation: insights from natural populations, genomics and experimental evolution. Molecular ecology 26:1703–1710.
Guo, H., E. Du, C. Terrer, and R. B. Jackson. 2024. Global distribution of surface soil organic carbon in urban greenspaces. Nature communications 15:806.
Halbritter, A. H., S. Fior, I. Keller, R. Billeter, P. J. Edwards, R. Holderegger, S. Karrenberg, et al. 2018. Trait differentiation and adaptation of plants along elevation gradients. Journal of evolutionary biology 31:784–800.
Harrison, T. L., C. W. Wood, I. L. Borges, and J. R. Stinchcombe. 2017*a*. No evidence for adaptation to local rhizobial mutualists in the legume Medicago lupulina. Ecology and evolution 7:4367–4376.
Harrison, T. L., C. W. Wood, K. D. Heath, and J. R. Stinchcombe. 2017*b*. Geographically structured genetic variation in the Medicago lupulina-Ensifer mutualism. Evolution; international journal of organic evolution 71:1787–1801.
Hedrick, P. W. 2005. A standardized genetic differentiation measure. Evolution; international journal of organic evolution 59:1633–1638.
Hendershot, J. N., Q. D. Read, J. A. Henning, N. J. Sanders, and A. T. Classen. 2017. Consistently inconsistent drivers of microbial diversity and abundance at macroecological scales. Ecology 98:1757–1763.
Irwin, R. E., E. Youngsteadt, P. S. Warren, and J. L. Bronstein. 2020. The evolutionary ecology of

mutualisms in urban landscapes. Pages 111–129 *in*Urban Evolutionary Biology. Oxford University PressOxford.
Johnson, M. T. J., and J. Munshi-South. 2017. Evolution of life in urban environments. Science (New York, N.Y.) 358.
Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281.
Kawecki, T. J., R. E. Lenski, D. Ebert, B. Hollis, I. Olivieri, and M. C. Whitlock. 2012. Experimental evolution. Trends in ecology & evolution 27:547–560.
Kearsley, J. V. S., L. M. Sather, and T. M. Finan. 2024. Sinorhizobium (Ensifer) meliloti. Trends in microbiology 32:516–518.
Kehlet-Delgado, H., A. P. Montoya, K. T. Jensen, C. E. Wendlandt, C. Dexheimer, M. Roberts, L. Torres Martínez, et al. 2024. The evolutionary genomics of adaptation to stress in wild rhizobium bacteria. Proceedings of the National Academy of Sciences of the United States of America 121:e2311127121.
Keller, I., J. M. Alexander, R. Holderegger, and P. J. Edwards. 2013. Widespread phenotypic and genetic divergence along altitudinal gradients in animals. Journal of evolutionary biology 26:2527–2543.
Kent, A. D., and E. W. Triplett. 2002. Microbial communities and their interactions in soil and rhizosphere ecosystems. Annual review of microbiology 56:211–236.
Klinger, C. R., J. A. Lau, and K. D. Heath. 2016. Ecological genomics of mutualism decline in nitrogen-fixing bacteria. Proceedings. Biological sciences 283:20152563.
Klipp, W., H. Reiländer, A. Schlüter, R. Krey, and A. Pühler. 1989. The Rhizobium meliloti fdxN gene encoding a ferredoxin-like protein is necessary for nitrogen fixation and is cotranscribed with nifA and nifB. Molecular & general genetics: MGG 216:293–302.
Kolmogorov, M., J. Yuan, Y. Lin, and P. A. Pevzner. 2019. Assembly of long, error-prone reads using repeat graphs. Nature biotechnology 37:540–546.
Körner, C. 2007. The use of “altitude” in ecological research. Trends in ecology & evolution 22:569–574.
Koskella, B., and M. Vos. 2015. Adaptation in natural microbial populations. Annual review of ecology, evolution, and systematics 46:503–522.
Kraemer, S. A., and P. J. Boynton. 2017. Evidence for microbial local adaptation in nature. Molecular ecology 26:1860–1876.
Lennon, J. T., and S. E. Jones. 2011. Microbial seed banks: the ecological and evolutionary implications of dormancy. Nature reviews. Microbiology 9:119–130.
Lenski, R. E. 2017. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. The ISME journal 11:2181–2194.
Lenth, R. V. 2024. Estimated Marginal Means, aka Least-Squares Means.
Li, H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics (Oxford, England) 32:2103–2110.
Linn, M. C., E. Palmer, A. Baranger, E. Gerard, and E. Stone. 2015. Education. Undergraduate research experiences: impacts and opportunities. Science (New York, N.Y.) 347:1261757.
Lomolino, M. V. 2001. Elevation gradients of species‐density: historical and prospective views: *Elevation gradients of species-density*. Global ecology and biogeography: a journal of macroecology 10:3–13.
Lopez, Z. C., M. L. Friesen, E. Von Wettberg, Leslie New, and S. Porter. 2021. Microbial mutualist distribution limits spread of the invasive legume Medicago polymorpha. Biological invasions 23:843–856.
Martinez-Romero, E., A. Peix, M. Hungria, S. A. Mousavi, J. Martinez-Romero, and P. Young. 2024. Guidelines for the description of rhizobial symbiovars. International journal of systematic and evolutionary microbiology 74.
medaka: Sequence correction provided by ONT Research. 2018. Github.
Miles, L. S., J. C. Johnson, R. J. Dyer, and B. C. Verrelli. 2018. Urbanization as a facilitator of gene flow in a human health pest. Molecular ecology 27:3219–3230.
Miles, L. S., L. R. Rivkin, M. T. J. Johnson, J. Munshi-South, and B. C. Verrelli. 2019. Gene flow and genetic drift in urban environments. Molecular ecology 28:4138–4151.

Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, and R. Lanfear. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular biology and evolution 37:1530–1534.
Murray-Stoker, D., and M. T. J. Johnson. 2021. Ecological consequences of urbanization on a legume–rhizobia mutualism. Oikos (Copenhagen, Denmark) 130:1750–1761.
O'Brien, A. M., R. J. H. Sawers, J. Ross-Ibarra, and S. Y. Strauss. 2018. Evolutionary responses to conditionality in species interactions across environmental gradients. The American naturalist 192:715–730.
Odriozola, I., J. A. Rasmussen, M. T. P. Gilbert, M. T. Limborg, and A. Alberdi. 2024. A practical introduction to holo-omics. Cell reports methods 4:100820.
Oksanen, J., G. L. Simpson, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, et al. 2022. vegan: Community Ecology Package.
Poole, P., V. Ramachandran, and J. Terpolilli. 2018. Rhizobia: from saprophytes to endosymbionts. Nature reviews. Microbiology 16:291–303.
Rahbek, C., M. K. Borregaard, R. K. Colwell, B. Dalsgaard, B. G. Holt, N. Morueta-Holme, D. Nogues-Bravo, et al. 2019. Humboldt's enigma: What causes global patterns of mountain biodiversity? Science (New York, N.Y.) 365:1108–1113.
R Core Team. 2023. The R Project for Statistical Computing. Vienna, Austria.
Regus, J. U., C. E. Wendlandt, R. M. Bantay, K. A. Gano-Cohen, N. J. Gleason, A. C. Hollowell, M. R. O'Neill, et al. 2017. Nitrogen deposition decreases the benefits of symbiosis in a native legume. Plant and soil 414:159–170.
Riley, A. B., M. A. Grillo, B. Epstein, P. Tiffin, and K. D. Heath. 2023. Discordant population structure among rhizobium divided genomes and their legume hosts. Molecular ecology 32:2646–2659.
Sachs, J. L., M. O. Ehinger, and E. L. Simms. 2010. Origins of cheating and loss of symbiosis in wild Bradyrhizobium. Journal of evolutionary biology 23:1075–1089.
Sachs, J. L., K. W. Quides, and C. E. Wendlandt. 2018. Legumes versus rhizobia: a model for ongoing conflict in symbiosis. The new phytologist 219:1199–1206.
Santangelo, J. S., R. W. Ness, B. Cohan, C. R. Fitzpatrick, S. G. Innes, S. Koch, L. S. Miles, et al. 2022. Global urban environmental change drives adaptation in white clover. Science (New York, N.Y.) 375:1275–1281.
Santangelo, J. S., L. R. Rivkin, C. Advenard, and K. A. Thompson. 2020. Multivariate phenotypic divergence along an urbanization gradient. Biology letters 16:20200511.
Santangelo, J. S., L. R. Rivkin, and M. T. J. Johnson. 2018. The evolution of city life. Proceedings. Biological sciences 285:20181529.
Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics (Oxford, England) 30:2068–2069.
Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics (Oxford, England) 31:3210–3212.
Stanford, J. S., S. E. Rocheleau, K. P. W. Smith, and J. Mohan. 2017. Early undergraduate research experiences lead to similar learning gains for STEM and Non-STEM undergraduates. Studies in higher education 42:115–129.
Tang, M., H. Wang, X. Qi, T. He, B. Zhang, E. Wang, M. Yu, et al. 2022. Diversification of Sinorhizobium populations associated with Medicago polymorpha and Medicago lupulina in purple soil of China. Frontiers in microbiology 13:1055694.
Telford, R. J., V. Vandvik, and H. J. B. Birks. 2006. Dispersal limitations matter for microbial morphospecies. Science (New York, N.Y.) 312:1015.
Tonkin-Hill, G., N. MacAlasdair, C. Ruis, A. Weimann, G. Horesh, J. A. Lees, R. A. Gladstone, et al. 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome biology 21:180.
Turkington, R., and P. B. Cavers. 1979. THE BIOLOGY OF CANADIAN WEEDS.: 33. Medicago lupulina L. Canadian journal of plant science 59:99–110.

UniProt Consortium. 2023. UniProt: The universal protein knowledgebase in 2023. Nucleic acids research 51:D523–D531.
Ushey, K., and H. Wickham. 2023. renv.
van den Brand, T. 2024. ggh4x: ggplot extension: options for tailored facets, multiple colourscales and miscellaneous. Github.
Verheyen, J., N. Tüzün, and R. Stoks. 2019. Using natural laboratories to study evolution to global warming: contrasting altitudinal, latitudinal, and urbanization gradients. Current opinion in insect science 35:10–19.
Wadgymar, S. M., M. L. DeMarche, E. B. Josephs, S. N. Sheth, and J. T. Anderson. 2022. Local adaptation: Causal agents of selection and adaptive trait divergence. Annual review of ecology, evolution, and systematics 53:87–111.
Walker, R. M., V. C. Sanabria, and H. Youk. 2024. Microbial life in slow and stopped lanes. Trends in microbiology 32:650–662.
Walters, K. E., J. K. Capocchi, M. B. N. Albright, Z. Hao, E. L. Brodie, and J. B. H. Martiny. 2022. Routes and rates of bacterial dispersal impact surface soil microbiome composition and functioning. The ISME journal 16:2295–2304.
Weese, D. J., K. D. Heath, B. T. M. Dentinger, and J. A. Lau. 2015. Long-term nitrogen addition causes the evolution of less-cooperative mutualists. Evolution; international journal of organic evolution 69:631–642.
Weissman, J. L., S. Hou, and J. A. Fuhrman. 2021. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. Proceedings of the National Academy of Sciences of the United States of America 118:e2016810118.
Westhoek, A., L. J. Clark, M. Culbert, N. Dalchau, M. Griffiths, B. Jorrin, R. Karunakaran, et al. 2021. Conditional sanctioning in a legume-Rhizobium mutualism. Proceedings of the National Academy of Sciences of the United States of America 118:e2025760118.
Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, et al. 2019. Welcome to the tidyverse. Journal of open source software 4:1686.
Wick, R. R., and P. Menzel. 2017. Filtlong. Available online: github. com/rrwick/Filtlong (accessed on 15 August 2021).
Wielbo, J., M. Marek-Kozaczuk, A. Kubik-Komar, and A. Skorupska. 2007. Increased metabolic potential of Rhizobium spp. is associated with bacterial competitiveness. Canadian journal of microbiology 53:957–967.
Wilke, C. O. 2020. cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2.”
Winter, D. 2012. mmod: an R library for the calculation of population differentiation statistics. Molecular ecology resources 12.
Yang, W., H. Feng, X. Zhang, J. Zhang, J. H. Doonan, W. D. Batchelor, L. Xiong, et al. 2020. Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. Molecular plant 13:187–214.
Zhang, L., D. J. Jacob, E. M. Knipping, N. Kumar, J. W. Munger, C. C. Carouge, A. van Donkelaar, et al. 2012. Nitrogen deposition to the United States: distribution, sources, and processes. Atmospheric chemistry and physics 12:4539–4554.