PNAS RESEARCH ARTICLE | EVOLUTION

The evolutionary genomics of adaptation to stress in wild rhizobium bacteria

Hanna Kehlet-Delgado[a], Angeliqua P. Montoya[b], Kyson T. Jensen[c], Camille E. Wendlandt[b], Christopher Dexheimer[b], Miles Roberts[b], Lorena Torres Martínez[d], Maren L. Friesen[a,e], Joel S. Griffitts[c], and Stephanie S. Porter[b,1]

Edited by Paul Schulze-Lefert, Max-Planck-Institut fur Pflanzenzuchtungsforschung, Cologne, Germany; received July 5, 2023; accepted February 8, 2024

Microbiota comprise the bulk of life's diversity, yet we know little about how populations of microbes accumulate adaptive diversity across natural landscapes. Adaptation to stressful soil conditions in plants provides seminal examples of adaptation in response to natural selection via allelic substitution. For microbes symbiotic with plants however, horizontal gene transfer allows for adaptation via gene gain and loss, which could generate fundamentally different evolutionary dynamics. We use comparative genomics and genetics to elucidate the evolutionary mechanisms of adaptation to physiologically stressful serpentine soils in rhizobial bacteria in western North American grasslands. In vitro experiments demonstrate that the presence of a locus of major effect, the *nre* operon, is necessary and sufficient to confer adaptation to nickel, a heavy metal enriched to toxic levels in serpentine soil, and a major axis of environmental soil chemistry variation. We find discordance between inferred evolutionary histories of the core genome and *nreAXY* genes, which often reside in putative genomic islands. This suggests that the evolutionary history of this adaptive variant is marked by frequent losses, and/or gains via horizontal acquisition across divergent rhizobium clades. However, different *nre* alleles confer distinct levels of nickel resistance, suggesting allelic substitution could also play a role in rhizobium adaptation to serpentine soil. These results illustrate that the interplay between evolution via gene gain and loss and evolution via allelic substitution may underlie adaptation in wild soil microbiota. Both processes are important to consider for understanding adaptive diversity in microbes and improving stress-adapted microbial inocula for human use.

adaptation | rhizobia | serpentine soil | population genomics | mobile genetic element

Adaptation is key to the persistence of microbial taxa across a variable landscape. The genetic variants that aid in environmental adaptation are often investigated in clonally evolving microbial populations in laboratory experiments (1, 2). However, conditions influencing evolutionary dynamics in these controlled experiments can differ dramatically from those in wild populations. For instance, natural microbial populations have access to novel genetic variants via horizontal gene transfer (HGT). In diverse communities, this can generate massive pangenomes with vast reservoirs of accessory functions that participate in complex mosaics of selection (3–5). We know surprisingly little about specific patterns of adaptation in wild microbes in response to environmental variation. In particular, how environmentally adaptive variants arise and spread in wild populations is poorly understood: Are adaptive variants lineage-specific mutations that arise independently in different species-like lineages or do divergent lineages share adaptive variants broadly via HGT?

The distribution of microbial population diversity across the landscape is impacted by evolutionary dynamics during adaptation. Novel adaptive genetic variants that confer stress tolerance will be rare but could confer dramatic fitness benefits in stressful environments. These locally adaptive variants can trigger genome-wide sweeps if low rates of HGT-driven recombination result in genome-wide linkage (6). Here, a clone bearing a locally adaptive variant may outcompete other local clones, leading to a locally homogeneous population, in a process similar to the founder effect. This can generate specialized subpopulations that occupy discrete ecological niches and display environmental differentiation at the whole-genome level (1, 4, 7–9). However, locally adaptive variants can result in gene-specific sweeps if high rates of HGT-driven recombination can free adaptive variants from linkage with other portions of the genome (6). This can allow an adaptive variant to go to high frequency locally but maintain local diversity for other portions of the genome. This can generate generalist quasi-sexual populations that together occupy a broad ecological niche with each lineage also accumulating niche-specific genes that confer a local fitness advantage (10–13). Understanding the contributions of these contrasting dynamics to adaptation in wild microbial populations is key to accurately modeling

Significance

There is great interest in investigating the genetic basis for adaptation in microbes, yet few studies reveal both the genes and evolutionary dynamics that allow microbes to adapt to natural environmental variation. We identify genes associated with the ability to tolerate stressful soil conditions in wild symbiotic bacteria and demonstrate that these genes drive replicated patterns of adaptation across the landscape. Phylogenetic evidence indicates that these adaptive genes are transferred among otherwise distinct lineages and drive parallel molecular solutions to stress among populations across large spatial scales. These findings reveal molecular processes of adaptation in wild microbes across the landscape and are widely applicable to efforts to understand the evolutionary origins of microbial diversity.

Author contributions: H.K.-D., A.P.M., K.T.J., C.E.W., M.L.F., J.S.G., and S.S.P. designed research; H.K.-D., A.P.M., K.T.J., C.E.W., C.D., M.R., L.T.M., J.S.G., and S.S.P. performed research; H.K.-D., A.P.M., K.T.J., C.E.W., M.L.F., J.S.G., and S.S.P. analyzed data; and H.K.-D., A.P.M., K.T.J., C.E.W., J.S.G., and S.S.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: stephanie.porter@wsu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2311127121/-/DCSupplemental.

Published March 20, 2024.

how variable selection can maintain microbial diversity across the landscape.

Due to HGT, different genomic compartments in microbes can experience distinct patterns of dispersal, selection, drift, and mutation, and thus can show distinct biogeographic patterns (14). The biogeography of environmentally adaptive loci in soil habitats is rarely elucidated, despite the importance of these loci to microbial persistence (10). If HGT of adaptive loci occurs primarily among close relatives with low dispersal, this could lead to unique solutions to stress among clades and among regions, respectively. Here, adaptive loci acquired by a lineage would tend to be limited to those from local close relatives. Alternatively, if HGT of adaptive loci spans highly divergent taxa with high dispersal, this could lead to globally shared solutions to stress among clades, and among regions, respectively. Such broad patterns of HGT could occur if adaptive loci are borne on excisable mobile genetic elements (MGEs) capable of inserting into globally conserved insertion sites. Shedding light on these potential biogeographic outcomes for environmentally adaptive loci will complement growing evidence of microbial biogeography.

Wild soil microbial communities are highly diverse, but the traits and evolutionary dynamics that contribute to adaptive diversity in the soil are challenging to study in wild systems (15). Serpentine soil outcrops provide a useful model system for studying environmental adaptation due to the steep shifts in selection pressure across their boundaries with adjacent soils and the fact that many outcrops occur across the landscape, providing spatial replication of shifts in selection (16–19). Serpentine soils are the product of naturally weathered ultramafic serpentine bedrock and are characterized by high and often toxic concentrations of heavy metals like nickel (Ni), and unusual ionic composition, such as low Ca^{2+} and high Mg^{2+} (17). Patches of serpentine soil are areas that impose extreme physiological stress, while the non-serpentine soils in which such outcrops are found are more physiologically benign. Adaptations to the extreme physiological challenges of serpentine soil can be found within the genomes of plant species that have adapted to such conditions. For example, transmembrane metal ion transporters such as Ni efflux transport proteins are highly differentiated between populations of *Arabidopsis lyrata* that span serpentine and non-serpentine soils (20). Some of these serpentine adaptive alleles appear to have introgressed to nearby congener *Arabidopsis arenosa* (16). Diverse prokaryotes inhabiting serpentine soils also contain heavy metal tolerance systems (21), but the evolutionary dynamics and biogeographic distributions of these adaptations are poorly resolved. Are adaptations to heavy metal shared among disparate lineages of soil microbes due to high rates of HGT across divergent taxa and gene-specific sweeps on serpentine soil outcrops? Or do distinct serpentine lineages of soil microbes inhabit adjacent non-serpentine soil and serpentine outcrops due to low HGT, HGT restricted to close relatives, and genome-wide sweeps?

The fate of soil microbes and the plant hosts they colonize on serpentine barrens are intertwined. Heavy metal-tolerant rhizobacteria can improve plant growth in serpentine (22, 23) and other stressful soils (24). Bacteria can evolve tolerance to excess heavy metals by removing metals from the cell, preventing ions from entering the cell, sequestering or detoxifying metals within the cell, or altering the redox state of the metal (25, 26). Such adaptations to serpentine soil could be costly and lead to trade-offs that render the fitness benefits of serpentine adaptations context dependent (27). This may generate clines in adaptive variants across the boundaries of serpentine soil outcrops (28). On the other hand, there is relatively little evidence for trade-offs associated with microbial adaptations in wild populations (29). If prokaryotes induce expression of costly adaptations only in response to strongly selective environmental cues, adaptations may incur little cost across environments (30). The absence of cross-environment costs could allow adaptations to go to fixation globally. However, the fitness effects and distributions of microbial adaptive variants across heterogenous environments are largely undescribed in soil.

To investigate the genetics, evolutionary dynamics, and biogeography of microbial adaptation to serpentine soil (Ca:Mg < 1), we examine the symbiotic rhizobium bacteria associated with native legume species both inside and outside of multiple serpentine barrens. Rhizobia are free-living bacteria in soil microbial communities capable of initiating symbiosis if they encounter the roots of compatible legume species. Rhizobia dispersal across soil types is likely common as they are mobile for short distances in their flagellated life history stage (31), and appear to disperse longer distances in wind and water flow (32, 33). Molecular plant-microbe signaling can induce the legume to form nodule organs on their roots within which rhizobia can fix atmospheric nitrogen in a protected intracellular environment, essentially fertilizing the plant (34). Western North American native grassland legumes, *Acmispon wrangelianus and Acmispon brachycarpus,* are symbiotic specialists that fix nitrogen with select strains of *Mesorhizobium* rhizobia in both physiologically harsh, heavy metal-enriched serpentine soil as well as various non-serpentine soils (27). This symbiotic specificity is useful for using plants to sample wild bacterial species with continental-scale distributions, providing a complement to the historic focus on microbial adaptation in vitro (28, 35).

We link patterns of traits, fitness, genomics, and genetics for these mesorhizobia to investigate the following: 1) Is there a genetic basis for adaptation to serpentine soils? Upon identifying the *nre* operon as necessary and sufficient to confer adaptation to Ni, a heavy metal enriched to toxic levels on serpentine soil, we ask, 2) What are the evolutionary dynamics of *nre* in the rhizobium pangenome? and 3) What is the biogeographic distribution of *nre* across the heterogeneous environments found in natural soils?

Results and Discussion

Genetic Basis for Microbial Adaptation to Serpentine Soils.

Rhizobial adaptation to spatially variable heavy-metal enrichment. We sampled sites to span a transect through most of the latitudinal extent of the geographic range of *A. wrangelianus* and *A. brachycarpus*, which occur from Northern Mexico to Southern Oregon, USA (*SI Appendix,* Table S1). Variation in soil chemistry among these sites is largely driven by two classic bioindicators of serpentine soil, high Ni and low calcium (Ca) (17, 19). We measured soil chemistry, including Ni, macro-, and micronutrients, at 55 grassland sites hosting populations of *A. wrangelianus* and/or *A. brachycarpus,* that were in 18 reserves located in California and Oregon, USA. Principal component analysis revealed two major axes of variation, PC1 and PC2, which explain 25% and 23% of soil chemistry variation, respectively (Fig. 1*A*). The top loadings for PC2 were calcium (–0.50) and Ni (0.39) and for PC1 are soil organic matter (0.43) and potassium (0.40). Soil Ni plays a key role in driving adaptation across soil variation in many taxa (17, 18), so to investigate adaptation to local soil conditions, we focused on this major axis of natural environmental variation.

Mesorhizobium from replicated sites within serpentine soil adapts to local Ni concentrations. We isolated rhizobia from 669 field-collected nodules from both *Acmispon* species from serpentine and non-serpentine soils across our 55 sites. We also included 46 *Mesorhizobium* strains from a similar study (27). We

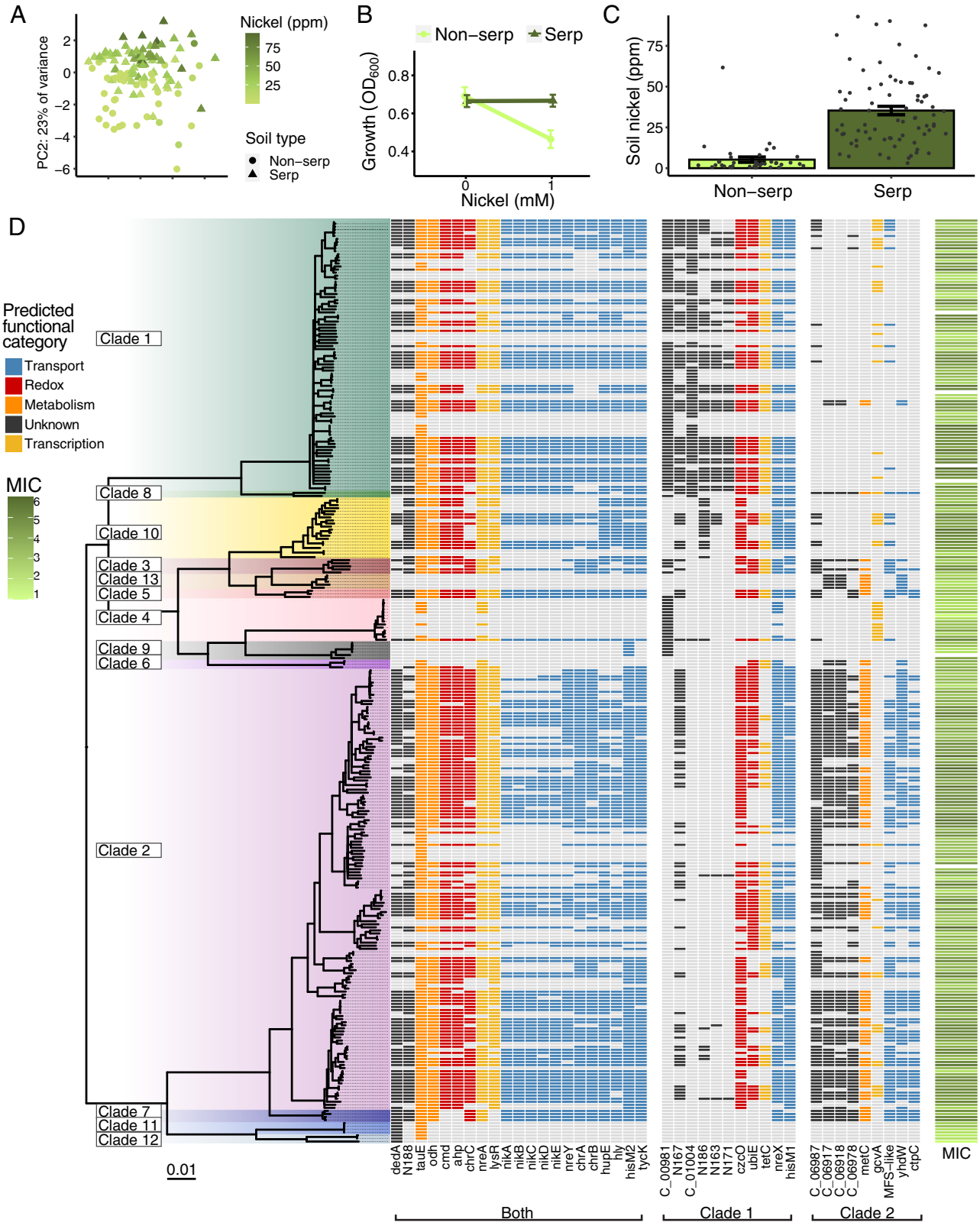

Fig. 1. *Mesorhizobium* from heavy-metal-enriched serpentine soil adapts to nickel (Ni) and shows a polyphyletic distribution of Ni tolerance candidate genes in a GWAS. (*A*) Ni enrichment reflects an important axis of soil niche space. Principal components analysis of 11 soil chemistry parameters across 55 sites inhabited by host legumes reveals that Ni (ppm) is a top loading for PC2. (*B*) The growth of *Mesorhizobium* strains from serpentine soils (n = 204) is insensitive to Ni; however, strains from non-serpentine soils (n = 92) grow slowly in the presence of Ni (Dataset S1). Shown is an interaction plot of the estimated marginal mean growth at 72 h ±SE based on a general linear model. SE; standard error. (*C*) Serpentine soils (serp) contain more Ni than non-serpentine soils (non-serp). Soils with a Ca:Mg < 1 are classified as serpentine. Bars indicate means ±1 SEM. Points indicate individual soil samples (n = 114). (*D*) A phylogeny based on amino acid sequences of 1,542 single copy core genes built with RAxML. The presence (colored tile) or absence (gray tile) of genes in each strain associated with Ni tolerance (MIC for Ni and/or growth in nickel media). Candidate genes are grouped by whether they were significant in both Clades, Clade 1 only, or Clade 2 only (FDR-adjusted *P*-value < 0.05). Predicted function of candidate genes is indicated by tile color. A strain's MIC is indicated by the shade of a green tile. Clades 1 to 13 are designated by background colors in the phylogeny (scale bar indicates substitutions per site). Branches with bootstrap support values of less than 70 are collapsed to polytomies. For strain identities and additional strain information, see *SI Appendix,* Fig. S3.

then measured each strain's tolerance to Ni via the minimum inhibitory concentration (MIC) of Ni grown on tryptone yeast agar supplemented with 0 to 5 mM Ni. For strains selected for genomic sequencing (see below), we also measured growth in Ni media as the OD_{600} after 72 h in tryptone yeast broth supplemented with 1 mM $NiCl_2$. Strains from soil with higher levels of Ni (and higher PC2 values) show higher Ni tolerance (Fig. 1 *A* and *B* and *SI Appendix*, Fig. S1 *A–C* and Table S2). This repeated pattern of adaptation was indistinguishable for strains isolated from different natural reserves and host species (*SI Appendix,* Table S2).

These environmentally adapted rhizobia comprise multiple clades of *Mesorhizobium.* We characterized the pangenome of the collected *Mesorhizobium* strains using PacBio (n = 16) and Illumina (n = 286) sequencing to provide complete (n = 3) or draft genomes (n = 299) for 302 strains spanning the geographical breadth of the collection (*SI Appendix*, Table S3 and Dataset S1). The pangenome was obtained through Markov clustering (MCL) (36) of protein sequences into orthologous clusters using Anvi'o (37). This *Mesorhizobium* pangenome comprises a core genome that includes 2,423 genes (orthologous clusters) conserved across the strains and a flexible genome that includes 55,613 genes (*SI Appendix,* Fig. S2 *A* and *B* and Table S4). A phylogeny built using a concatenated alignment of single copy core genes (Fig. 1*D* and *SI Appendix*, Fig. S3) and pairwise average nucleotide identity (ANI) distances (38, 39) of marker genes at a 95% threshold indicate that the majority of these genomes comprise two major phylogenetically distinct clades, Clade 1 (n = 89) and Clade 2 (n = 144), with the remaining strains comprising 11 smaller clades (Fig. 1*D* and *SI Appendix,* Fig. S4*A* and Table S3). This clade structuring was recaptured in an additional phylogeny based on marker genes universally conserved in bacteria (40–42) (*SI Appendix*, Fig. S5 *A* and *B*). Within our *Mesorhizobium* marker genes, 15.5% of the variation in amino acids is attributed to variation among reserves (PERMANOVA, $P < 0.01$), while host species and serpentine or non-serpentine soil type accounted for 4.9% and 2.3% ($P < 0.01$) (*SI Appendix,* Table S5), respectively. Therefore, we observe some regional differentiation among *Mesorhizobium,* but weak differentiation across soil types. However, dispersion, the variation within a group, differs among reserves ($df = 14$, $F = 7.652$, $P < 0.01$), and host species ($df = 1$, $F = 29.497$, $P < 0.01$), but not between soil types (40). Thus, differences in dispersion within these groups may contribute to measures of *Mesorhizobium* differentiation among reserves and hosts (41).

To identify genetic variants that underlie environmental adaptation, we sought locally abundant genetic variants associated with high fitness under conditions that reflect an important axis of local niche space (5). We used a genome-wide association study (GWAS) approach to associate genetic variants to Ni tolerance, an adaptation to an important axis of environmental variation across the landscape. We tested both patterns of gene presence/absence in the flexible genome and single nucleotide polymorphisms (SNPs) as sources of genetic variation. Since different *Mesorhizobium* clades could differ in which genetic variants are correlated with Ni tolerance, GWAS was performed separately for Clade 1 and Clade 2. To account for a clonal background among strains, which can reduce power due to large amounts of shared DNA (42), we implemented a linear mixed model (FaST-LMM) with phylogenetic population structure correction through pyseer (43–45) using the marker gene phylogeny. To add additional support, we repeated the pyseer GWAS analyses using first a FaST-LMM with population structure correction based on the single-copy core gene phylogeny and then again under a fixed-effects model with population structure correction using Mash-computed whole-genome sequence distances (46). These three methods for accounting for population structure yielded largely overlapping results (Dataset S2), and below, we discuss the candidate loci that were significant in all three analyses.

Our findings are consistent with a scenario in which the gain or loss of genes contributes to Ni adaptation in *Mesorhizobium*. After performing GWAS through pyseer and applying multiple test corrections, we found that the presence of 32 and 30 genes in Clades 1 and 2, respectively, were associated with Ni MIC, growth in Ni media, or both at the 5% significance level (Dataset S2). Twenty-one of these genes associate with Ni tolerance for both Clade 1 and Clade 2, consistent with a scenario in which genetic strategies for Ni tolerance are shared across divergent *Mesorhizobium* clades (Fig. 1*D*). Patterns of presence/absence of these Ni tolerance candidate loci in other clades showed these genes are not limited to Clades 1 and 2 (Fig. 1*D*). Our findings are also consistent with a model in which allelic variation in the core genome does not contribute substantially to Ni adaptation. After testing 197,600 and 38,665 SNPs for Clade 1 and Clade 2, respectively, and applying Benjamin–Hochberg FDR-correction for multiple testing, there were no SNPs associated with either Ni tolerance phenotype at the 5% significance level. Our findings suggest SNP alleles of large effect that segregate at intermediate frequencies in core genes are unlikely to underlie Ni adaptation. However, this association genetics approach could fail to detect small effect or rare alleles (47).

Many of the genes whose presence is associated with Ni tolerance phenotypes reside in close physical proximity in the genome. One strain in which the genetic proximity of these GWAS candidates is especially evident is C089B, in which 30 candidate genes are closely associated along the chromosome (*SI Appendix*, Fig. S6 *C* and *D*) in a region bearing similarities to a MGE (Fig. 2*A*). Further inspection in other genomes shows a similar pattern (Fig. 1*D*), indicating the possibility of a small genomic island (GI) bearing P4 integration motifs, with multiple groups of genes contributing to Ni tolerance. To determine whether genetic variants associated with a phenotype via GWAS that are co-localized contribute to the phenotype or are spuriously associated via physical linkage, we tested for a causal link to the phenotype for each locus (48) and used experimental genetic evidence to validate whether we observe an impact of genotype on the Ni adaptation phenotype.

Functional validation of the nre operon for Ni tolerance. To identify genes that contribute to Ni tolerance in the genomic region enriched for candidate loci, we conducted a functional analysis of individual genes and multi-gene transcription units (TUs) within this region, using C089B as a reference strain. Twenty regions (excluding some genes for DNA processing and transcriptional regulation) were defined within the C089B Ni tolerance gene cluster (Fig. 2*A*). These TUs were ligated into a broad host-range plasmid for expression in the fast-growing *Rhizobiaceae* strain *Agrobacterium fabrum* UBAPF2, which is sensitive to Ni. One transcription unit (#8) could not be cloned despite several attempts. Of the remaining 19 clones, only one (#12) conferred robust Ni tolerance to the *A. fabrum* test strain (Fig. 2*B* and *SI Appendix*, Fig. S7). This clone consisted of four putatively co-transcribed genes, designated *nreA*, *nreX*, *nreY*, and *dedA*. The protein encoded by *nreA* belongs to the CsoR family of transcriptional regulators and shares similarity to the *nreA* gene of *Cupriavidus metallidurans* 31A, which contributes to Ni tolerance in that organism (49). The proteins encoded by *nreX* and *nreY* are predicted to be proton:cation antiporters of the cation diffusion facilitator and major facilitator superfamily families, respectively. The *dedA* gene encodes a putative transmembrane transporter protein associated

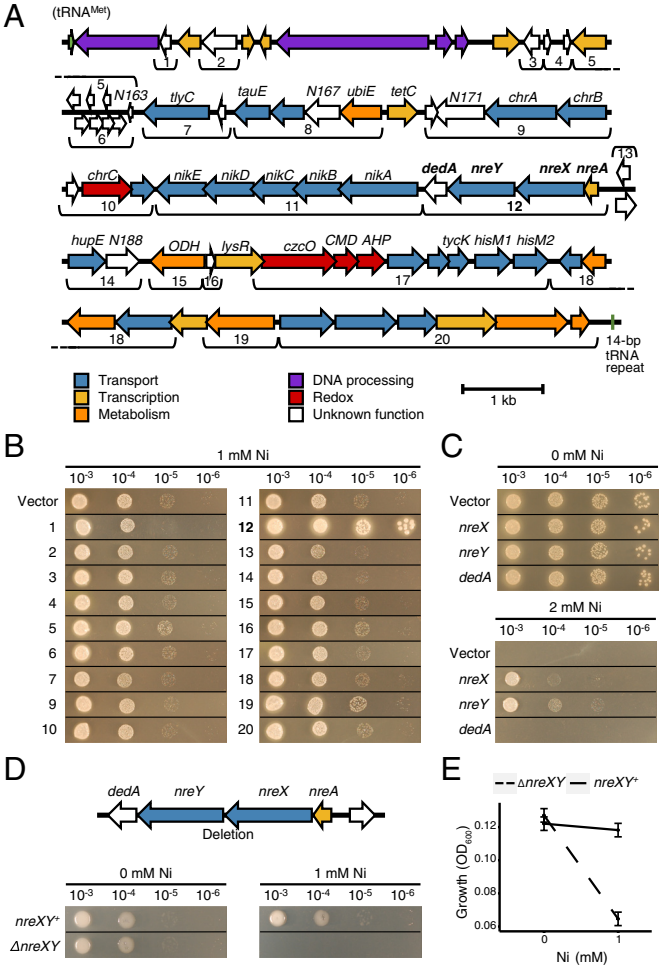

Fig. 2. The candidate gene island contains genes necessary and sufficient for *Mesorhizobium* nickel tolerance. (*A*) A *Mesorhizobium* GI from strain C089B enriched in Ni tolerance candidate genes from GWAS. Brackets group genes into 20 probable TUs for plasmid-based testing. (*B*) Spot dilution testing of *A. fabrum* clones expressing *Mesorhizobium* candidate Ni tolerance genes. A plasmid for expressing TU8 was not successfully constructed. (*C*) *Mesorhizobium* genes residing in TU12 (*nreX*, *nreY*, and *dedA*) were individually introduced into *A. fabrum* on an expression plasmid and the resulting strains were spotted onto medium with 0 and 2 mM $NiCl_2$. (*D*) Deletion of the two-gene *nreXY* region was accomplished in the *Mesorhizobium* C089B background and the resulting strain (compared to *nreXY*$^+$) was tested for Ni tolerance by spot dilution test. (*E*) Growth of *nreXY*$^+$ C089B and the isogenic Δ*nreXY* mutant in the presence and absence of Ni. Shown is an interaction plot of the estimated marginal mean growth at 48 h ±SE from a general linear model. SE; standard error.

with resistance to various metal and organic compounds, either indirectly through its effects on bacterial surface chemistry, or directly via efflux of the chemical stressor (50, 51).

Ni tolerance gene clusters in Clades 1 and 2 vary in *nreAX*(*Y*)-*dedA* gene organization. Of 246 occurrences of the *nre* gene cluster in 208 strains (some strains have more than one copy), 65% of loci had the *nreAXYdedA* organization, 19% had an *nreAXdedA* organization, 9% had *nreAX* occurring without *nreY* or *dedA*, 5% had an organization with *nreAXY* separated from *dedA* by 3 to 4 hypothetical genes, and 2% have undeterminable organization due to contig breaks.

The *nreX*, *nreY*, and *dedA* genes were tested individually for their ability to confer Ni tolerance in *A. fabrum* (Fig. 2*C*). *A. fabrum* expressing *nreX* or *nreY* supported strain growth at 2 mM $NiCl_2$, a Ni concentration that did not allow growth of the vector-only control strain, while *A. fabrum* expressing *dedA* did not exhibit Ni tolerance above background. The *nreX* and *nreY* genes are therefore each sufficient to confer Ni tolerance. Having demonstrated the sufficiency of *nreX* or *nreY* to confer Ni tolerance in the heterologous system, we investigated how crucial this gene pair is for Ni tolerance in the *Mesorhizobium* strain C089B, which is Ni tolerant. We created a precise deletion of the *nreX-nreY* gene pair in the C089B strain background using an allele exchange strategy. This deletion mutant lost considerable tolerance to Ni compared to the wild-type parent strain background (Fig. 2 *D* and *E*), highlighting the role of *nre* as necessary for high Ni tolerance in this serpentine soil-derived strain. Genes in the *nre* operon are thus primary determinants of C089B nickel tolerance, though additional untested loci could contribute to the phenotype as well.

Evolutionary History of the *nre* Gene Cluster.

Origins of nre. To investigate the origins of adaptive loci, we examined whether lineages tend to inherit adaptive loci vertically or acquire them from distantly related lineages. We find that many Ni tolerance candidate genes are shared across highly divergent *Mesorhizobium* clades (Fig. 1*D*), consistent with globally shared sweeps of adaptive loci, rather than lineage-specific innovation followed by clonal sweeps in response to environmental Ni. The phylogeny based on the *nreAX* locus is incongruent with the *Mesorhizobium* core and marker gene phylogenies, and strains that share high sequence similarity, genome-wide, differ in the presence/absence of the nreAXY locus (*SI Appendix,* Fig. S8 *A* and *B*). This incongruence, combined with the fitness advantage *nreAXY* confers in the presence of Ni, is consistent with a scenario in which diverse *Mesorhizobium* lineages acquire this niche-specific locus where it is locally advantageous via HGT-driven recombination that has freed nreAXY from tight linkage with the core genome (6, 10–13).

To explore the evolutionary history of the *nre* operon (and to test whether certain forms of the *nreAXY* genes associate with higher Ni tolerance, below), we clustered *nreA*, *nreX*, and *nreY* protein-coding sequences with a sequence similarity approach (referred to as "alleles" hereafter). Comparisons of the distributions of these alleles with respect to the phylogeny showed that although some alleles have a higher frequency in one clade or the other, there is imperfect sorting of alleles across clades (Fig. 3*A*). This is consistent with HGT of adaptive alleles primarily among close relatives, but also occasionally across deeply diverged *Mesorhizobium* clades.

The nre gene cluster often resides in a putative transmissible element. The mechanisms by which adaptive loci transfer between cells and integrate into the genome are predicted to shape biogeographic and phylogenetic distributions of the adaptations they encode (5). We find that candidate genes for *Mesorhizobium* Ni tolerance tend to be physically clustered in various putative GIs, such as putative MGEs, segments of DNA that carry genes for their own excision and transfer (53, 54). For example, the *nre* operon, and linked candidate genes, reside within an inferred mobile element region in all eight of the complete or nearly complete PacBio-sequenced *Mesorhizobium* genomes that contain the *nre* operon (*SI Appendix,* Table S7). This physical organization is recapitulated in eight draft genomes with contigs large enough to enable us to detect MGE-related genes in regions neighboring the *nre* operon and linked candidate genes (*SI Appendix,* Table S7). Introduction of DNA from a MGE can result in multiple gene modules that provide novel functions and potentially contribute to adaptation to a specific niche (5, 53, 55). In C089B, the putative mobile element containing the *nre* operon and candidate Ni tolerance genes resides 2.4 Mb from the symbiosis island (*SI Appendix,* Fig. S9) and is inserted at the 3′ end of a tRNA-Met gene with a 14-bp direct repeat 54 kbp downstream. This region contains an open reading frame (ORF) coding for a serine recombinase as well as ORFs with similarity to

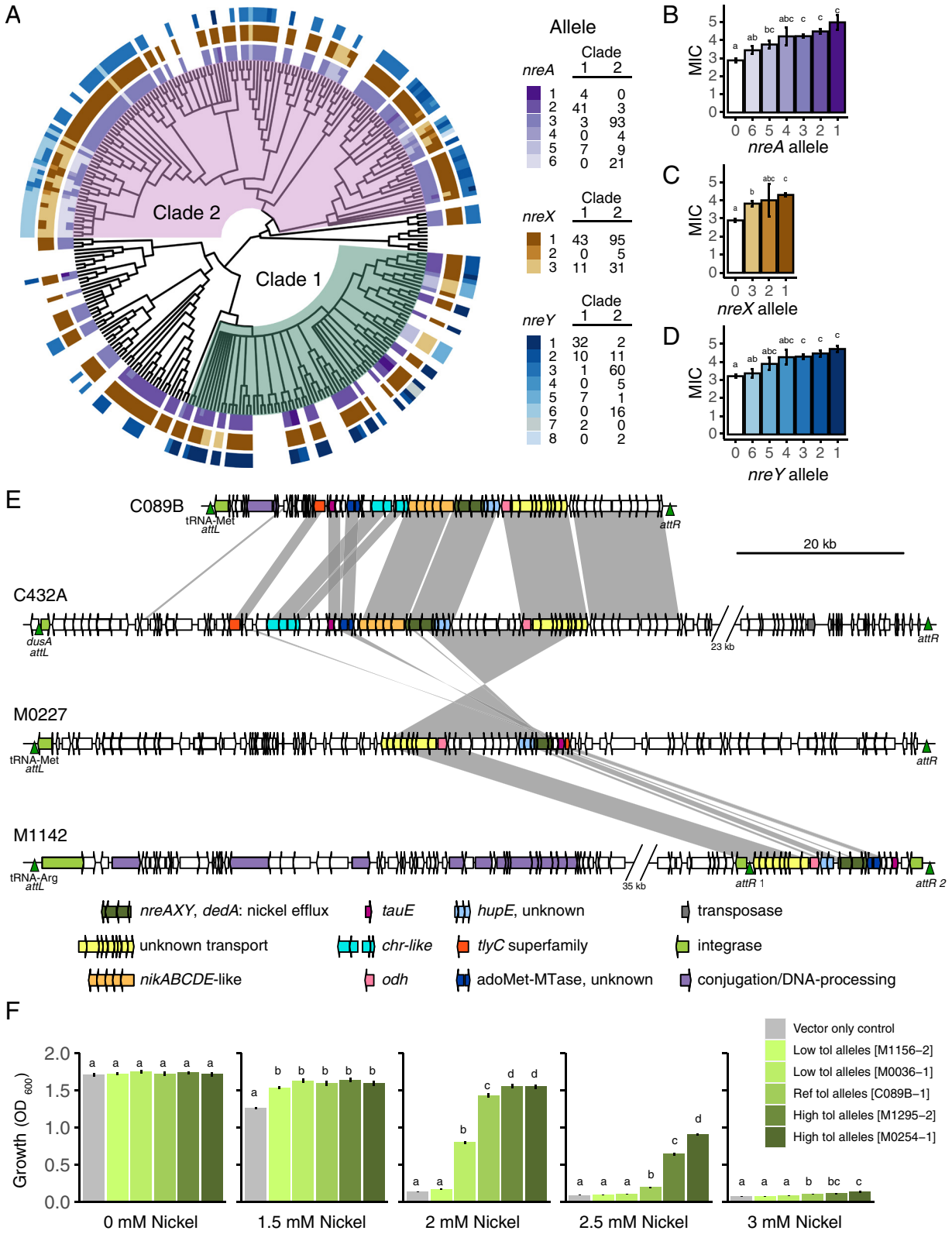

Fig. 3. *nre* Ni efflux genes reside within a variety of putative GIs and display allelic diversity. (*A*) Single copy core gene phylogeny of *Mesorhizobium* isolates with *nre* allele plotted adjacent showing a polyphyletic distribution of *nreA*, *nreX*, and *nreY* alleles. Clades 1 and 2 are highlighted. Allele counts per clade are displayed in the key (presence/absence). (*B–D*) Estimated marginal mean MIC ±SE of strains with *nre* (*B*: *nreA*, *C*: *nreX*, and *D*: *nreY*) allele number designations. Allele 0 refers to the absence of that gene. Alleles 7 and 8 of *nreY* were removed because of low abundance. (*E*) Diagram comparing GI-like regions containing Ni tolerance candidates from different genomes. Regions of similarity as computed by the Mauve aligner (52) connected by gray. Ni tolerance candidates and putative HGT-related genes are color coded and labeled with predicted function. Segments cut due to length in islands of C432A and M1142 are symbolized with diagonal lines. (*F*) Natural allelic variants of the *nreAXY* operon inserted into *A. fabrum* confer different growth in the presence of nickel. Bars represent mean growth ±SE after 24 h for low tolerance, reference strain, and high tolerance alleles. Inside brackets are the strain ID and clade the alleles were derived from. Letters above bars indicate significant differences at $P < 0.05$. SE; standard error.

the *traACD* mobilization locus of the integrative mobile element IME*Ml*^{R88B} of *Mesorhizobium* sp. R88B (56, 57). (Fig. 3*E*). We consider similar regions in other *Mesorhizobium* that share attributes such as insertion at a tRNA-met and *att* sites, and bear multiple Ni tolerance candidate genes, to be "C089B-type Ni islands." Out of the 29 strains with this organization that we examined, 27 belonged to Clade 1, one to Clade 2, and one to Clade 10. Among intact C089B-type Ni islands (i.e., those that lack contig breaks), island size ranges widely, from 53,319 bp in strain M0626 to 172,320 bp in strain M0025. While the Ni island of C089B does not contain genes with predicted conjugative functions, C089B-type Ni islands in several other genomes contained putative conjugative type IV secretion systems which could function to transfer the Ni island among lineages (58, 59). HGT of these putative GIs thus appears to rely on mobile element features and insertion sites that are broadly conserved among *Mesorhizobium* strains and present a mechanism by which an environmentally adaptive locus is distributed across phylogenetically distant and geographically widespread strains.

Across *Mesorhizobium* strains, *nre* loci occur in various putative mobile genomic contexts. Clusters of Ni tolerance genes localize in the same genomic region, but with variable gene order across divergent strains, even when a distinct GI-like region is not identifiable (Fig. 3*E*). The genomic context of *nreAXYdedA* and other Ni-associated gene clusters is not limited to C089B-type Ni islands, but also includes putative GIs with alternative insertion sites and features consistent with integrative and conjugative elements (60) (Fig. 3*E* and *SI Appendix,* Table S7). The variable co-occurrence and mosaic structure of Ni candidate gene modules in different MGE-like systems could result from recombination between MGEs, which can occur in *Mesorhizobium* (56). Additionally, insertion sequence (IS) transposases were sometimes found located near Ni candidate genes and IS-mediated transposition can result in gain and loss of genes (61). Thus, HGT of *nre,* which confers fitness benefits in Ni-enriched habitats, could drive gene transfer of variable numbers and identities of other loci in diverse mobile contexts and contribute to genomic diversity.

Of the 41 genes associated with Ni tolerance, only *nreX* and *nreY* contribute to functional Ni tolerance in vitro. The fitness impacts of the loci co-transferred as components of putative Ni GIs remain unknown. On one hand, these genes could provide little to no fitness benefit (5, 62). Their absence from some putative Ni GIs across different strain backgrounds could result from genome streamlining whereby selection has purged lineages with extra neutral DNA or a process by which shorter mobile elements have had a greater probability of integrating into the chromosome (5, 63, 64). Alternatively, co-transferred genes may confer tolerance to other stressors in serpentine soil. Beyond Ni enrichment, serpentine soil can impose stress due to excess Cu and Mg, dry conditions, and low levels of macronutrients (17, 19), The fact that gene clusters with potential roles in the transport of other metals occur within some Ni islands is consistent with the latter possibility (*SI Appendix*, Table S8). We note that while metal tolerance systems in bacteria sometimes confer tolerance to multiple metals (65, 66), Ni tolerance is uncorrelated with Co or Cr tolerance across 94 haphazardly selected strains we examined (*SI Appendix*, Fig. S10 *A* and *B*). Often the serpentine soils we examined contained low levels of Co or Cr, so selection for metal cross-tolerance may be weak in our system (*SI Appendix*, Table S9).

Allelic variation in nre genes may contribute to environmental adaptation. The *nre* alleles we identified associate with different levels of Ni tolerance. Based on patterns of amino acid divergence at each gene, we delineated 6 *nreA* alleles, 3 *nreX* alleles, and 8 *nreY* alleles. While some alleles were more common in one *Mesorhizobium* clade over another, other alleles occurred at similar frequencies in both Clade 1 and Clade 2 strains (Fig. 3*A*). General linear mixed models (GLMMs) indicate that the allelic identity of *nreA* (X^2 = 132.69, P < 0.001), *nreX* (X^2 = 114.56, P < 0.001), and nreY (X^2 = 100.78, P < 0.001) predicts the Ni minimal inhibitory concentration (MIC) of strains, which was confirmed with estimated marginal means post hoc analysis (Fig. 3 *B*–*D*).

Naturally variant *nreAXY* alleles that were expressed in *A. fabrum* conferred distinct degrees of Ni tolerance to elevated levels of nickel. We used *nreAXY* alleles that were predicted in the model above to confer high Ni tolerance or low Ni tolerance, from clade 1 and 2 strains (Fig. 3 *A*–*C*) as well as reference strain C089B, which we find to have intermediate Ni tolerance. While *nreAXY* alleles confer growth indistinguishable from that of the vector-only control, in the absence of Ni, all *nreAXY* alleles confer Ni resistance, compared to the vector-only control, in 1.5 mM $NiCl_2$ ($F_{5,66}$ = 27, P < 0.001; Fig. 3*F*). However, at higher Ni levels, the predicted high tolerance alleles confer greater tolerance to Ni than the predicted low tolerance alleles (2 mM nickel: $F_{5,66}$ = 1044.7, P < 0.001; 2.5 mM nickel: $F_{5,66}$ = 1342.8, P < 0.001; 3 mM nickel: $F_{5,66}$ = 14.28, P < 0.001; Fig. 3*F*). Genetic variation conferred by different alleles occurs across both clades 1 and 2.

These findings are consistent with a scenario in which *nre* alleles that confer a level of Ni tolerance just sufficient to withstand local levels of Ni enrichment could accumulate in a population. Thus, allelic variation at the adaptive *nre* locus contributes genetic variation that has the potential to fine-tune environmental adaptation driven by the gain or loss of the Ni GI.

nre across the Landscape.

Replicated clines in nre. Our findings document large-scale observations of repeated clinal variation in the frequency of an adaptive genetic variant in soil bacteria. Across the western United States, *Mesorhizobium* strains from sites with higher levels of Ni enrichment are more likely to contain *nreAX* or *nreY* ($X^2(1)$ = 48.2, P < 0.0001; $X^2(1)$ = 36.8, P < 0.0001; respectively; *SI Appendix,* Fig. S11 *A* and *B* and Tables S10 and S11). Even at fine spatial scales, including those separated by only hundreds of meters, frequencies of strains possessing *nre* were dramatically elevated on serpentine soils, often up to 100%, while strains from adjacent sites with non-serpentine soils rarely possessed *nre* (Fig. 4 and *SI Appendix*, Table S1). In fact, for every 1 ppm increase in soil Ni, the log-odds of an isolate bearing *nreAX* increased by 1.15, and the log-odds of an isolate bearing *nreY* increased by 1.11. Similarly, soil PC2, a major axis of soil chemistry variation heavily influenced by Ni enrichment, also had a positive relationship with the presence of *nreAX* [$X^2(1)$ = 21.3, P < 0.0001] and *nreY* [$X^2(1)$ = 21.4, P < 0.0001]. Our findings are consistent for symbiont populations collected from both *Acmispon* host species (*SI Appendix,* Tables S10 and S11). Thus, despite the potential for high rates of soil microbial migration via transport in wind, water flow, or on mobile animals (33), spatial variation in natural selection due to soil chemistry appeared to be sufficient to maintain the repeated evolution of environmental adaptation to Ni via clines in the frequency of the *nre* locus.

Conditional neutrality of nre. While adaptive loci that are transmitted horizontally can be key to persistence across environments, their fitness effects are rarely measured across environmental conditions (5). To understand the evolutionary dynamics that cause *nre* to segregate at high frequency (90%) in Ni enriched serpentine soil, but low frequency (23%) in low Ni non-serpentine soil, we tested whether *nre* imposes a detectable cost in the absence of Ni. Under antagonistic pleiotropy, a variant beneficial in one environment can be costly elsewhere

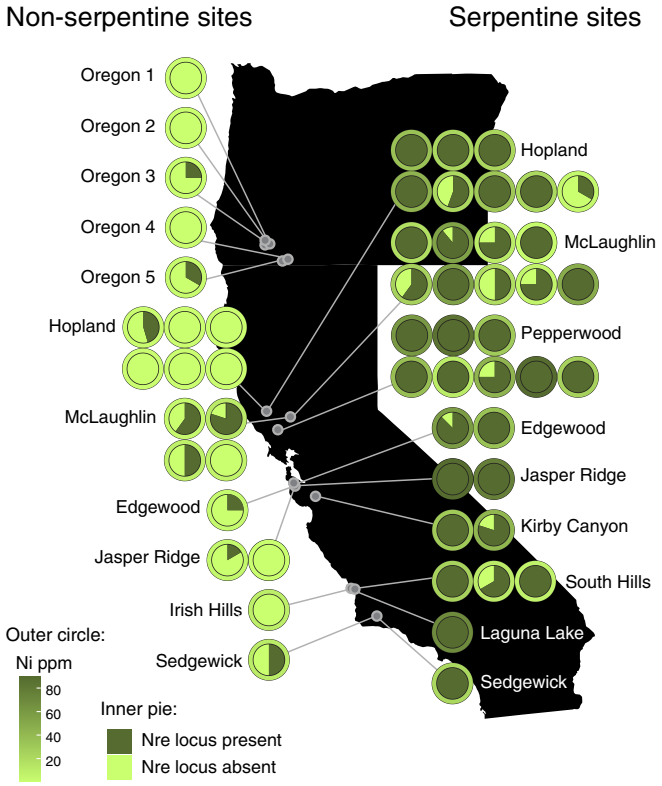

Fig. 4. Across natural *Mesorhizobium* populations in Oregon and California, USA, *nreAXY* presence is higher at sites with higher soil Ni concentrations. Pies indicate sites within a reserve. Reserve name is labeled by each cluster of sites. Colors within the pie indicate the proportion of sequenced strains with the *nre* locus (defined as genome containing *nreAX* and/or *nreAXY* locus) present (dark green) or absent (light green). A total of 56 sites with strains sequenced for this study are shown: 50 from the from this study (2019 sampling); and 6 from a similar study from 2009 (27). The outer circle around each pie indicates the level of Ni enrichment in soil at a site (darker green indicates higher Ni). Serpentine sites (defined as soil Ca:Mg < 1; *Right*) and non-serpentine soil (*Left*) are shown clustered by soil type. *SI Appendix*, Table S1 gives the number of strains sequenced per site and reserve.

and this genetic trade-off maintains the spatial distribution of adaptive variants. However, trade-offs are not ubiquitous. Under conditional neutrality, variants beneficial in one environment have no impact on fitness elsewhere, and here, barriers to gene flow can maintain the non-random spatial distribution of adaptive variants (67). We found that while the C089B Ni tolerant streptomycin-resistant strain, which bears the *nre* locus, grows more rapidly in the presence of Ni than does the isogenic *nreXY* deletion strain ($F_{3,60}$ = 51.3, P < 0.001), these strains show indistinguishable growth at 48 h in TYC media supplemented with streptomycin in the absence of Ni (Fig. 2*E*). Thus, while the *nre* locus increases fitness in the presence of Ni, it has no detectable cost in the absence of Ni under our experimental conditions, consistent with a conditionally neutral adaptive variant. In fact, across the set of 296 *Mesorhizobium* strains we tested, Ni tolerance and the presence/absence of *nre* do not predict a strain's growth rate in media that lacks Ni enrichment (*SI Appendix*, Fig. S12 *A–L* and Table S12), in contrast to findings from a smaller previous study (27). Thus, our present findings are consistent with growing evidence that horizontally transmitted adaptive loci can segregate across habitat types despite imposing no detectable costs (68). However, future experiments could test for other pleiotropic costs of metal efflux proteins like *nre* that might be missed in in vitro assays, such as a reduction of survival under natural biotic and abiotic conditions, due to factors such as increased susceptibility to biotic antagonists present in non-serpentine soil (29, 69).

The lack of a detected fitness trade-off at the single locus level (70) is consistent with findings from other systems in which low expression of gene modules in the absence of an environmental cue results in very weak costs for bearing inducible adaptations (29). Given the absence of detectable costs to bearing *nre,* it is possible that its low frequency in strains at sites with low levels of Ni could be explained by genomic streamlining, whereby genetic variants that are neutral in a given environment tend to be lost due to random genetic drift leading to the reduction of excess DNA (71, 72). The variable gene content in the putative GI in which *nre* resides may arise from HGT combined with "soft" selective sweeps, which preserve genomic diversity surrounding a sweeping locus that confers an environment-specific selective advantage (10, 73). Neutral variants, such as *nre* in a low-Ni context, can persist in populations long enough to recombine or undergo HGT into different genomic backgrounds. Despite the lack of benefit *nre* confers in the absence of Ni, *nre* operons present in strains from low-Ni non-serpentine soil still appear to confer functional Ni tolerance to the strains that bear them (*SI Appendix*, Fig. S13; P < 0.001; Wilcoxon two-sided test). If these lineages subsequently migrate onto high-Ni serpentine soil, the selective advantage *nre* would confer could cause them to rise to high frequency (10). This could cause different Ni island variants to rise to high frequency as strains bearing different *nre* alleles migrate onto Ni-enriched patches of serpentine soil, a process akin to adaptation via standing genetic variation in organisms like plants where HGT is uncommon.

Conclusions

Remarkably little is known about the impacts of horizontally transmitted loci on evolutionary dynamics and outcomes of adaptation in wild microbial communities across the landscape. This work elucidates the molecular genetics of environmental adaptation to stress in wild soil bacteria across variable selection in natural environments. We observe a repeated pattern along the western United States whereby wild *Mesorhizobium* bacteria adapt to the presence of nickel in their native soil conditions by acquiring *nre,* an operon that increases fitness in the presence of nickel. This key adaptive locus is harbored by diverse *Mesorhizobium* clades and resides in putative MGEs exchanged horizontally among lineages. Our results highlight the importance of HGT gene modules and GIs to microbial adaptation to spatially varying selection, analogous to inversions in organisms like plants, in that they are transferred by recombination but protected from being broken up by their genomic architecture.

Materials and Methods

Strain Isolation and Soil Analysis. *A. wrangelianus* and *A. brachycarpus* were gathered during March 2019 from natural reserves in California and May 2019 from areas in Oregon. Within reserves, both serpentine and non-serpentine sites were sampled for 1) plant root nodules, and 2) the soil directly below the plant, if available. Some strains collected in the same manner from additional sites in 2009 were also included in this study. Details on strain isolation and soil analysis procedures are in *SI Appendix*.

Phenotyping Ni Tolerance. The MIC of Ni was determined for 668 field-collected strains of *Mesorhizobium* by growing the strains on $NiCl_2$ enriched media plates that ranged from 0 mM to 5 mM increasing by 1 mM for each treatment with three replicates. Ni tolerance for 296 sequenced strains was also determined by the growth (OD_{600}) in liquid media containing 0 or 1 mM $NiCl_2$. Details of these procedures are in *SI Appendix*.

Genome Sequencing and Assembly. Genomic reads from Illumina sequencing were de novo assembled with SPAdes (v. 3.14.1) (74) within Shovill (75).

Long-read sequence data for 18 strains was generated with Pacific Biosciences' (PacBio) sequencing and assembled with the PacBio SMRT cell portal. For genome assemblies passing quality control, protein coding sequences were predicted and annotated with Prokka (v. 1.13.3) (76) using Prodigal (v. 2.6.3) (77). Details on genome sequencing and assembly are in *SI Appendix*.

Phylogenetics and Pangenomics. Amino acid sequences from 1,542 single-copy core genes were aligned with MUSCLE (78), concatenated, and trimmed with Trimal (79). We then constructed a ML tree using RAxML (80) with rapid bootstrapping. An additional phylogeny containing newly sequenced genomes and reference *Mesorhizobium* genomes was built with the PhyloPhlan pipeline (81, 82) (v. 3.0.58) with RAxML (80) from a set of 345 amino acid markers highly conserved in bacterial genomes (81–83). Trees were visualized with the R package "ggtree" (84). To delineate clades, pairwise calculations of ANI distances between the concatenated alignments of marker genes identified by PhyloPhlan were performed with the tool FastANI (85) and a cutoff of 95% for clustering of medoid genomes (86). For understanding the pangenome, we used MCL (36) to cluster amino acid sequences into orthologous clusters within the Anvi'o v7 (37) "anvi-pan-genome" pipeline. SNPs from the genomes of Clade 1 and 2 (*SI Appendix*, Table S13) were detected separately using Parsnp from the Harvest tools suite (87) to understand variation at the nucleotide level. Details on phylogenetics and pangenomics are in *SI Appendix*.

GWAS. To assess genomic patterns correlated with Ni tolerance, we used a linear mixed model (FaST-LMM) (44) as implemented in pyseer (v. 1.3.6) (43) using phylogenetic distances, first with the marker gene phylogeny, then with the core gene phylogeny, to control for population structure. To further bolster our confidence in our results, we used pyseer again, this time with Mash distances (46) and a fixed effects model. Clade 1 and Clade 2 were analyzed separately. Gene presence/absence and SNPs were both tested. Full GWAS methods are in *SI Appendix*.

Bacterial Strains and Culture Conditions for Ni Island Testing. Strains used in the process for testing of a putative Ni island were *Mesorhizobium* C089B (a streptomycin-resistant variant, KJ094, was used for convenience), *A. fabrum* D224 (a streptomycin-resistant variant of UBAPF2) (88), *Escherichia coli* DH5α, and helper strain B001 (89). Culture conditions are described in *SI Appendix*.

Plasmid Construction. Details on plasmid construction for testing of Ni TUs and *nre* genes are available in *SI Appendix*. A list of plasmids and primers used in this study are provided in *SI Appendix*, Tables S13 and S14, respectively.

Strain Construction. Modified *Mesorhizobium* and *A. fabrum* strains were created through triparental matings using the helper strain (B001) and donor strains (DH5α) containing the experimental plasmids. For conjugation, strain mixtures were co-cultured on TYC agar at 30 °C for 4 to 24 h and plated on TYC agar containing streptomycin (Sm) and neomycin (Nm). To construct the *nreXY* deletion by allele exchange, Nm-resistant (Nm^R) transconjugants were subsequently selected on TYC agar containing sucrose and X-Gluc to obtain potential deletion clones, which were subsequently verified by PCR. Strains used are provided in *SI Appendix*, Table S15. Ni tolerance test methods are described in *SI Appendix*.

Principal Component Analysis on Soil Data. To identify which soil chemistry traits (listed in *SI Appendix*) discriminate most among field sites, we performed principal component analysis on the correlation matrix of all soil chemistry traits for 114 soil samples using the "princomp" command in R version 4.1.2.

Statistical Analyses. We analyzed Ni tolerance of *Mesorhizobium* isolates using linear models and linear mixed models implemented in the R package lme4 v. 1.1–28 ("lmer" function) (90). We analyzed the presence of *nre* (*nreA*, *nreX*, and *nreY*) genes in *Mesorhizobium* isolates using the "glmer" function in lme4. Further details are provided in *SI Appendix*.

Assessing Allelic Variation in *nre* Genes. Amino acid sequences of *nreA*, *nreX*, and *nreY* were each aligned using MAFFT v. 7.475 (91) with the LINSI option for use with the command "IdClusters" from the DECIPHER package (92) in R for assignment of cluster IDs (referred to as "alleles" here). A GLMM (90) was used to test whether different alleles confer different levels of Ni tolerance (MIC). In addition, we used fixed effects linear models to test whether naturally variant *nreAXY* alleles expressed in vitro in *A. fabrum* conferred different levels of Ni tolerance. Full details are provided in *SI Appendix*.

Data, Materials, and Software Availability. Genome sequence data and Mesorhizobium phenotypes' data have been deposited in NCBI bioproject PRJNA852284 (93) and Dryad data repository (https://datadryad.org/stash/share/I2Wzvb8-q5d_WnFZrAQEh9QOcjXOKQf5lwXnvNvTwRI) (94), respectively.

ACKNOWLEDGMENTS. We thank Natural Reserves across CA and OR, USA (*SI Appendix*, Table S1), for providing access to field sites to support our research. This work was supported by NSF grants IOS-1755454 to M.L.F., J.S.G., and S.S.P., and DEB-1943239 to S.S.P., a Murdock Charitable Trust grant to S.S.P., and a CAS fellowship and Elling travel grants from Washington State University to A.P.M. We thank Z. C. Lopez for field assistance. We thank S. M. Rudman and two anonymous reviewers for comments that improved the manuscript.

Author affiliations: [a]Department of Plant Pathology, Washington State University, Pullman, WA 99164; [b]School of Biological Sciences, Washington State University, Vancouver, WA 98686; [c]Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT 84602; [d]Department of Biology, St. Mary's College of Maryland, St. Mary's City, MD 20686-3001; and [e]Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164

1. J. E. Barrick, R. E. Lenski, Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **14**, 827–839 (2013).
2. A. N. T. Nguyen *et al.*, Recombination resolves the cost of horizontal gene transfer in experimental populations of Helicobacter pylori. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2119010119 (2022).
3. J. J. Power *et al.*, Adaptive evolution of hybrid bacteria by horizontal gene transfer. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2007873118 (2021).
4. B. J. Shapiro, How clonal are bacteria over time? *Curr. Opin. Microbiol.* **31**, 116–123 (2016).
5. B. J. Arnold, I.-T. Huang, W. P. Hanage, Horizontal gene transfer and adaptive evolution in bacteria. *Nat. Rev. Microbiol.* **20**, 206–218 (2022).
6. B. J. Shapiro, M. F. Polz, Ordering microbial diversity into ecologically and genetically cohesive units. *Trends Microbiol.* **22**, 235–247 (2014).
7. K. J. Card, M. D. Thomas, J. L. Graves, J. E. Barrick, R. E. Lenski, Genomic evolution of antibiotic resistance is contingent on genetic background following a long-term experiment with Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2016886118 (2021).
8. D. E. Deatherage, J. L. Kepner, A. F. Bennett, R. E. Lenski, J. E. Barrick, Specificity of genome evolution in experimental populations of Escherichia coli evolved at different temperatures. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1904–E1912 (2017).
9. G. I. Lang *et al.*, Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571–574 (2013).
10. L. C. Woods *et al.*, Horizontal gene transfer potentiates adaptation by reducing selective constraints on the spread of genetic variation. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 26868–26875 (2020).
11. M. J. Rosen, M. Davison, D. Bhaya, D. S. Fisher, Fine-scale diversity and extensive recombination in a quasisexual bacterial population occupying a broad niche. *Science* **348**, 1019–1023 (2015).
12. A. Crits-Christoph, M. R. Olm, S. Diamond, K. Bouma-Gregson, J. F. Banfield, Soil bacterial populations are shaped by recombination and gene-specific selection across a grassland meadow. *ISME J.* **14**, 1834–1846 (2020).
13. J. Felsenstein, The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
14. A. Greenlon *et al.*, Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15200–15209 (2019).
15. S. A. Kraemer, P. J. Boynton, Evidence for microbial local adaptation in nature. *Mol. Ecol.* **26**, 1860–1876 (2017).
16. B. J. Arnold *et al.*, Borrowed alleles and convergence in serpentine adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8320–8325 (2016).
17. K. U. Brady, A. R. Kruckeberg, H. D. Bradshaw Jr., Evolutionary ecology of plant adaptation to serpentine soils. *Annu. Rev. Ecol. Evol. Syst.* **36**, 243–266 (2005).
18. E. Kazakou, P. G. Dimitrakopoulos, A. J. M. Baker, R. D. Reeves, A. Y. Troumbis, Hypotheses, mechanisms and trade-offs of tolerance and adaptation to serpentine soils: From species to ecosystem level. *Biol. Rev. Camb. Philos. Soc.* **83**, 495–508 (2008), 10.1111/j.1469-185X.2008.00051.x.
19. S. Harrison, N. Rajakaruna, *Serpentine: The Evolution and Ecology of a Model System* (University of California Press, 2011).
20. T. L. Turner, E. C. Bourne, E. J. Von Wettberg, T. T. Hu, S. V. Nuzhdin, Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nat. Genet.* **42**, 260–263 (2010).
21. A. Mengoni, C. Viti, R. J. Turner, L.-N. Huang, *Genomics of Bacterial Metal Resistance* (MDPI, 2021).
22. C. Fagorzi *et al.*, Harnessing rhizobia to improve heavy-metal phytoremediation by legumes. *Genes* **9**, 542 (2018).
23. M. Sujkowska-Rybkowska, D. Kasowska, K. Gediga, J. Banasiewicz, T. Stępkowski, Lotus corniculatus-rhizobia symbiosis under Ni, Co and Cr stress on ultramafic soil. *Plant Soil* **451**, 459–484 (2020).
24. C. Wu *et al.*, Bioremediation of mercury-polluted soil and water by the plant symbiotic fungus Metarhizium robertsii. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2214513119 (2022).
25. D. A. Rouch, B. T. O. Lee, A. P. Morby, Understanding cellular responses to toxic agents: A model for mechanism-choice in bacterial metal resistance. *J. Ind. Microbiol.* **14**, 132–141 (1995).
26. R. J. Turner, L.-N. Huang, C. Viti, A. Mengoni, Metal-resistance in bacteria: Why care? *Genes* **11**, 1470 (2020).

27. S. Porter, K. Rice, Trade-offs, spatial heterogeneity, and the maintenance of microbial diversity. *Evol. Int. J. Org. Evol.* **67**, 599–608 (2013).
28. S. S. Porter, P. L. Chang, C. A. Conow, J. P. Dunham, M. L. Friesen, Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic Mesorhizobium. *ISME J.* **11**, 248–262 (2017).
29. T. Ferenci, Trade-off mechanisms shaping the diversity of bacteria. *Trends Microbiol.* **24**, 209–223 (2016).
30. A. Schick, S. F. Bailey, R. Kassen, Evolution of fitness trade-offs in locally adapted populations of pseudomonas fluorescens. *Am. Nat.* **186** (suppl. 1), S48–S59 (2015).
31. W. Zhang *et al.*, Mycelial network-mediated rhizobial dispersal enhances legume nodulation. *ISME J.* **14**, 1015–1029 (2020).
32. S. T. N. Aroney, P. S. Poole, C. Sánchez-Cañizares, Rhizobial chemotaxis and motility systems at work in the soil. *Front. Plant Sci.* **12**, 725338 (2021).
33. Z. C. Lopez, M. L. Friesen, E. Von Wettberg, L. New, S. Porter, Microbial mutualist distribution limits spread of the invasive legume Medicago polymorpha. *Biol. Invasions* **23**, 843–856 (2021).
34. P. Poole, V. Ramachandran, J. Terpolilli, Rhizobia: From saprophytes to endosymbionts. *Nat. Rev. Microbiol.* **16**, 291–303 (2018).
35. L. Torres-Martínez *et al.*, Evolution of specialization in a plant-microbial mutualism is explained by the oscillation theory of speciation. *Evolution* **75**, 1070–1086 (2021).
36. S. Van Dongen, Graph clustering via a discrete uncoupling process. *SIAM J. Matrix Anal. Appl.* **30**, 121–141 (2008).
37. A. M. Eren *et al.*, Community-led, integrated, reproducible multi-omics with anvi'o. *Nat. Microbiol.* **6**, 3–6 (2021).
38. M. Kim, H.-S. Oh, S.-C. Park, J. Chun, Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **64**, 346–351 (2014).
39. J. Goris *et al.*, DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81–91 (2007).
40. J. Oksanen *et al.*, The Vegan Package, Community ecology package, R package, Version 2.6-0. https://cran.r-project.org/web/packages/vegan/index.html. Accessed 18 April 2022.
41. M. J. Anderson, Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62**, 245–253 (2006).
42. J. E. San *et al.*, Current affairs of microbial genome-wide association studies: Approaches, bottlenecks and analytical pitfalls. *Front. Microbiol.* **10**, 3119 (2020).
43. J. A. Lees, M. Galardini, S. D. Bentley, J. N. Weiser, J. Corander, pyseer: A comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* **34**, 4310–4312 (2018).
44. C. Lippert *et al.*, FaST linear mixed models for genome-wide association studies. *Nat. Methods* **8**, 833–835 (2011).
45. S. G. Earle *et al.*, Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat. Microbiol.* **1**, 1–8 (2016).
46. B. D. Ondov *et al.*, Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
47. E. A. Boyle, Y. I. Li, J. K. Pritchard, An expanded view of complex traits: From polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
48. M. M. Saber, B. J. Shapiro, Benchmarking bacterial genome-wide association study methods using simulated genomes and phenotypes. *Microb. Genom.* **6**, e000337 (2020).
49. G. Grass *et al.*, NreB from Achromobacter xylosoxidans 31A is a nickel-induced transporter conferring nickel resistance. *J. Bacteriol.* **183**, 2803–2807 (2001).
50. J. B. Caldeira, A. P. Chung, A. P. Piedade, P. V. Morais, R. Branco, A DedA family membrane protein in indium extrusion in Rhodanobacter sp. B2A1Ga4. *Front. Microbiol.* **12**, 772127 (2021).
51. P. R. Panta *et al.*, A DedA family membrane protein is required for Burkholderia thailandensis colistin resistance. *Front. Microbiol.* **10**, 2532 (2019).
52. A. E. Darling, B. Mau, N. T. Perna, progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**, e11147 (2010).
53. L. S. Frost, R. Leplae, A. O. Summers, A. Toussaint, Mobile genetic elements: The agents of open source evolution. *Nat. Rev. Microbiol.* **3**, 722–732 (2005).
54. D. J. Rankin, E. P. C. Rocha, S. P. Brown, What traits are carried on mobile genetic elements, and why? *Heredity* **106**, 1–10 (2011).
55. X. Jiang, A. B. Hall, R. J. Xavier, E. J. Alm, Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level niche-adaptive gene pools. *PLoS One* **14**, e0223680 (2019).
56. E. Colombi *et al.*, Comparative analysis of integrative and conjugative mobile genetic elements in the genus Mesorhizobium. *Microb. Genom.* **7**, 000657 (2021).
57. W. Reeve *et al.*, Genome sequence of the Lotus corniculatus microsymbiont Mesorhizobium loti strain R88B. *Stand. Genomic Sci.* **9**, 3 (2014).
58. E. Cabezón, J. Ripoll-Rozada, A. Peña, F. de la Cruz, I. Arechaga, Towards an integrated model of bacterial conjugation. *FEMS Microbiol. Rev.* **39**, 81–95 (2014).
59. P. J. Christie, K. Atmakuri, V. Krishnamoorthy, S. Jakubowski, E. Cascales, Biogenesis, architecture, and function of bacterial type Iv secretion systems. *Annu. Rev. Microbiol.* **59**, 451–485 (2005).
60. R. A. F. Wozniak, M. K. Waldor, Integrative and conjugative elements: Mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* **8**, 552–563 (2010).
61. P. Siguier, E. Gourbeyre, A. Varani, B. Ton-Hoang, M. Chandler, Everyman's guide to bacterial insertion sequences. *Microbiol. Spectr.* **3**, MDNA3-0030–2014 (2015).
62. S. Slomka *et al.*, Experimental evolution of Bacillus subtilis reveals the evolutionary dynamics of horizontal gene transfer and suggests adaptive and neutral effects. *Genetics* **216**, 543–558 (2020).
63. K. J. Apagyi, C. Fraser, N. J. Croucher, Transformation asymmetry and the evolution of the bacterial accessory genome. *Mol. Biol. Evol.* **35**, 575–581 (2018).
64. P. Puigbò, A. E. Lobkovsky, D. M. Kristensen, Y. I. Wolf, E. V. Koonin, Genomes in turmoil: Quantification of genome dynamics in prokaryote supergenomes. *BMC Biol.* **12**, 1–19 (2014).
65. M. Nanda, V. Kumar, D. K. Sharma, Multimetal tolerance mechanisms in bacteria: The resistance strategies acquired by bacteria that can be exploited to 'clean-up' heavy metal contaminants from water. *Aquat. Toxicol.* **212**, 1–10 (2019).
66. G. Maynaud *et al.*, CadA of Mesorhizobium metallidurans isolated from a zinc-rich mining soil is a PIB-2-type ATPase involved in cadmium and zinc resistance. *Res. Microbiol.* **165**, 175–189 (2014).
67. S. Yeaman, Evolution of polygenic traits under global vs local adaptation. *Genetics* **220**, iyab134 (2022).
68. S. M. Wadgymar, M. L. DeMarche, E. B. Josephs, S. N. Sheth, J. T. Anderson, Local adaptation: Causal agents of selection and adaptive trait divergence. *Annu. Rev. Ecol. Evol. Syst.* **53**, 87–111 (2022).
69. U. Halder *et al.*, Genomic, morphological, and biochemical analyses of a multi-metal resistant but multi-drug susceptible strain of Bordetella petrii from hospital soil. *Sci. Rep.* **12**, 8439 (2022).
70. P. Chen, J. Zhang, Antagonistic pleiotropy conceals molecular adaptations in changing environments. *Nat. Ecol. Evol.* **4**, 461–469 (2020).
71. S. J. Giovannoni *et al.*, Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**, 1242–1245 (2005).
72. A. K. Simonsen, Environmental stress leads to genome streamlining in a widely distributed species of soil bacteria. *ISME J.* **16**, 423–434 (2021), 10.1038/s41396-021-01082-x.
73. M. L. Bendall *et al.*, Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J.* **10**, 1589–1601 (2016).
74. A. Bankevich *et al.*, SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
75. T. Seemann, For draft genome analysis, Illumina reads were assembled de novo using Shovill version 1.1.0. Github. https://github.com/tseemann/shovill. Accessed 1 December 2020.
76. T. Seemann, Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
77. D. Hyatt *et al.*, Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
78. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
79. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
80. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
81. F. Asnicar *et al.*, Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat. Commun.* **11**, 2500 (2020).
82. N. Segata, D. Börnigen, X. C. Morgan, C. Huttenhower, PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* **4**, 2304 (2013).
83. Q. Zhu *et al.*, Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat. Commun.* **10**, 5477 (2019).
84. G. Yu, Using ggtree to visualize data on tree-like structures. *Curr. Protoc. Bioinformatics* **69**, e96 (2020).
85. C. Jain, L. M. Rodriguez-R, A. M. Phillippy, K. T. Konstantinidis, S. Aluru, High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
86. L. M. Carroll, M. Wiedmann, J. Kovac, Proposal of a taxonomic nomenclature for the Bacillus cereus group which reconciles genomic definitions of bacterial species with clinical and industrial phenotypes. *mBio* **11**, e00034-20 (2020), 10.1128/mbio.00034-20.
87. T. J. Treangen, B. D. Ondov, S. Koren, A. M. Phillippy, The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **15**, 524 (2014).
88. M. F. Hynes, R. Simon, A. Pühler, The development of plasmid-free strains of Agrobacterium tumefaciens by using incompatibility with a Rhizobium meliloti plasmid to eliminate pAtC58. *Plasmid* **13**, 99–105 (1985).
89. J. S. Griffitts *et al.*, A Sinorhizobium meliloti osmosensory two-component system required for cyclic glucan export and symbiosis. *Mol. Microbiol.* **69**, 479–490 (2008).
90. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
91. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
92. E. S. Wright, Using DECIPHER v2.0 to analyze big biological sequence data in R. *R. J.* **8**, 352 (2016).
93. H. Kehlet-Delgado *et al.*, Data from "Wild mesorhizobium populations of legumes in western North America." NCBI BioProject. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA852284/. Deposited 26 June 2023.
94. H. Kehlet-Delgado *et al.*, Data supporting: The evolutionary genomics of adaptation to stress in wild rhizobium bacteria [Dataset]. Dryad. https://doi.org/10.5061/dryad.vt4b8gtxz. Deposited 24 January 2024.