environmental microbiology

Environmental Microbiology (2018) **20**(11), 4112–4126 doi:10.1111/1462-2920.14405

Emergence of soil bacterial ecotypes along a climate gradient

Alexander B. Chase,[1*] **Zulema Gomez-Lunar**,[2] **Alberto E. Lopez**,[1] **Junhui Li**,[2] **Steven D. Allison**,[1,2] **Adam C. Martiny**[1,2] **and Jennifer B. H. Martiny**[1]

[1]*Department of Ecology and Evolutionary Biology, University of California, Irvine, California, USA.*

[2]*Department of Earth System Sciences, University of California, Irvine, California, USA.*

Summary

The high diversity of soil bacteria is attributed to the spatial complexity of soil systems, where habitat heterogeneity promotes niche partitioning among bacterial taxa. This premise remains challenging to test, however, as it requires quantifying the traits of closely related soil bacteria and relating these traits to bacterial abundances and geographic distributions. Here, we sought to investigate whether the widespread soil taxon *Curtobacterium* consists of multiple coexisting ecotypes with differential geographic distributions. We isolated *Curtobacterium* strains from six sites along a climate gradient and assayed four functional traits that may contribute to niche partitioning in leaf litter, the top layer of soil. Our results revealed that cultured isolates separated into fine-scale genetic clusters that reflected distinct suites of phenotypic traits, denoting the existence of multiple ecotypes. We then quantified the distribution of *Curtobacterium* by analysing metagenomic data collected across the gradient over 18 months. Six abundant ecotypes were observed with differential abundances along the gradient, suggesting fine-scale niche partitioning. However, we could not clearly explain observed geographic distributions of ecotypes by relating their traits to environmental variables. Thus, while we can resolve soil bacterial ecotypes, the traits delineating their distinct niches in the environment remain unclear.

Received 8 June, 2018; revised 3 August, 2018; accepted 6 September, 2018. *For correspondence. E-mail abchase@uci.edu

Introduction

A focus on traits can provide a mechanistic understanding of an organism's geographic distribution (McGill *et al*., 2006; Litchman and Klausmeier, 2008; Allison, 2012; Edwards *et al*., 2013) as traits underlie an organism's response to abiotic and biotic conditions (Lavorel and Garnier, 2002). In microbial communities, the link between traits and the distribution of microbial taxa remains poorly understood (Green *et al*., 2008; Litchman *et al*., 2015). While whole-genome and metagenomic data provide a sense for the potential types of traits of the microorganisms within an environmental sample (Raes *et al*., 2011), it is unclear how well this potential translates to actual phenotypic differences (Chase and Martiny, 2018). However, as in macroorganisms, the functional assessment of an ecosystem's abundant taxa is important to developing trait-based approaches to predict community and ecosystem dynamics (Lavorel and Garnier, 2002; Enquist *et al*., 2015).

Soil systems harbour incredible microbial diversity where high habitat heterogeneity promotes niche partitioning among bacterial taxa (Ranjard and Richaume, 2001; Nannipieri *et al*., 2003). Indeed, the biogeographic distributions of soil bacterial communities are correlated with environmental variables [e.g. pH (Fierer and Jackson, 2006; Yao *et al*., 2011) and nutrients (Leff *et al*., 2015)] suggesting that traits related to the response of these variables underlie bacterial distributions. However, most studies consider these patterns at a fairly broad genetic resolution, lumping taxa based on the sequence similarity of a highly conserved 16S rRNA region, and subsequently, mask a high degree of trait variation among distinct bacterial taxa (Chase *et al*., 2017; Larkin and Martiny, 2017). Such variation may be important for explaining the distribution of bacterial diversity in soil. In particular, some ecologically relevant traits, including a response to drought conditions may be shared at broad taxonomic levels (Amend *et al*., 2016), while others, such as temperature preference (Martiny *et al*., 2015), might be more variable. Thus, the ability to resolve the degree of trait variation requires linking genetic information with relevant phenotypic variation (McLaren and Callahan,

2018) to distinguish fine-scale niche partitioning contributing to the distribution and diversity of soil bacteria.

To illustrate the importance of trait variation among very closely related bacterial strains, we can consider the distribution of the abundant marine phototroph, *Prochlorococcus.* Strains of *Prochlorococcus* cluster into genetically distinct clades that share physiological traits, including light preference and nutrient utilization (Moore and Chisholm, 1999; Moore *et al*., 2002). Distinct fine-scale genetic clusters corresponding to ecologically relevant phenotypes have been defined as ecotypes (Rocap *et al*., 2003), where all strains within an ecotype occupy the same ecological niche (Cohan, 2001). Coexisting, but distinct *Prochlorococcus* ecotypes exhibit differential geographic distributions that are highly correlated with environmental variables, suggesting fine-scale niche partitioning (Moore *et al*., 1998; Johnson *et al*., 2006). Furthermore, the traits that underlie these correlations are relatively clear; for instance, the optimal temperature of an ecotype's growth in the laboratory corresponds to its distribution across oceanic temperature gradients (Johnson *et al*., 2006). Thus, coexisting individuals can be clustered based on genotypic and phenotypic information to provide a functional basis in driving ecotype differentiation and niche partitioning (Kent *et al*., 2016; Delmont and Eren, 2018).

Here, we sought to determine whether soil bacteria – like marine bacteria – form distinct ecotypes that are differentially distributed. To test this idea, we focused on the widespread soil taxon *Curtobacterium* (Chase *et al*., 2016). In a southern California grassland, *Curtobacterium* is highly abundant in leaf litter, the top layer of soil, suggesting a potential functional role in plant decomposition, and therefore, the carbon cycle (Matulich *et al*., 2015). In previous work, we identified the co-occurrence of multiple *Curtobacterium* clades in leaf litter, and we hypothesized that thermal adaptation might be contributing to ecological differences of co-occurring clades (Chase *et al*., 2017). Indeed, a recent study found that northern and southern *Streptomyces* lineages differed in their thermal tolerance across a latitudinal gradient (Choudoir and Buckley, 2018). However, it remains unclear what traits differentiate *Curtobacterium* clades and if this diversity is associated with niche partitioning in the environment. Therefore, in this study, we isolated *Curtobacterium* strains from leaf litter across six locations spanning a climate gradient and assayed four functional traits (growth, biofilm formation and depolymerization of xylan and cellulose). We further investigated the biogeographic distributions of *Curtobacterium* clades using metagenomic data collected from leaf litter across the gradient over 18 months. We hypothesized that (i) genetically distinct *Curtobacterium* lineages share functional traits, forming distinct ecotypes; (ii) these ecotypes have differential geographic distributions across the gradient; and (iii) the distribution of ecotypes is correlated with environmental variation (e.g. soil temperature).

Results

Identification of Curtobacterium ecotypes

We established six sites along an elevation gradient that covaried in precipitation and temperature (Supporting Information Table S1). We isolated *Curtobacterium* strains from leaf litter (decaying leaves that make up the topmost layer of soil) at all sites along this climate gradient except at the coldest and wettest Subalpine site. We sequenced 56 new *Curtobacterium* genomes with an average size of 3.6 Mbp and 70.7% GC content (Supporting Information Table S2). Incorporating all known *Curtobacterium* genomic diversity into a phylogenetic analysis, we established five major clades (I–V) delineating the genus (Fig. 1). These genomes were highly diverse, sharing as low as 79.2% average nucleotide identity (ANI) and 68.9% average amino acid identity (AAI). Most (93%) of the isolates from our sites fell within Clades I, IV, and V. The major clades further diverged into finer genomic clusters (assigned to subclade designation at ≥90% AAI; Fig. 1). Each site along the gradient exhibited vast genomic diversity, with most of the subclades including isolates from multiple sites. For instance, 25 of the isolates from four of the sites (excluding the higher elevation sites, Pine-Oak and Subalpine) fell into Subclade IB/C. In contrast, Subclade VA was only isolated from the Desert site.

Our analyses revealed extensive genomic diversity that would otherwise be masked using traditional genetic characterizations (i.e. 16S rRNA gene similarity) of bacterial operational taxonomic units (OTUs). Analysing the hypervariable V4/V5 region of the 16S rRNA region at 100% sequence similarity, we identified only four OTUs. All eight isolates from Subclade VA grouped together, while the majority of strains (45 of the 56 climate gradient strains), irrespective of phylogenetic relatedness, shared identical V4/V5 regions. An alignment of the full-length 16S rRNA gene region revealed some congruence between subclade diversity, but clustering into OTUs at 99% similarity [as recommended for full-length sequences (Schloss, 2010; Edgar, 2018)], still revealed only two OTUs (Supporting Information Fig. S1).

We next asked whether this genomic variation corresponded to phenotypic diversity by assaying a subset of isolates for four functional traits (growth, biofilm formation and depolymerization of xylan and cellulose). Given the stark gradient in temperature across our sites, we measured each trait at a range of temperatures to consider the response of traits to environmental variation (McGill

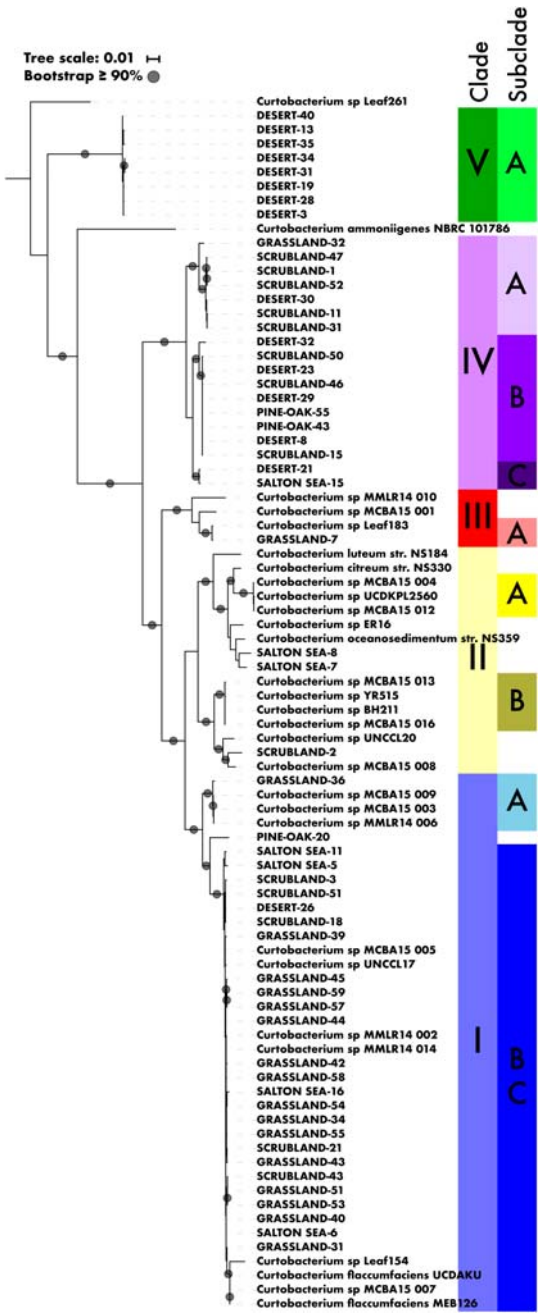

Fig. 1. Phylogeny of *Curtobacterium* clades and subclades constructed from a multilocus phylogenetic analysis of 21 single-copy marker genes. Subclade designations are assigned if all genomes within subclade share ≥90% average amino acid identity. Strains with assigned nomenclature beginning with 'Curtobacterium' are publicly available genomes, while the other labels designate the site of isolation along the climate gradient.

et al., 2006). Indeed, across all assays, temperature explained 15%–20% of the trait variation observed in the lab assays (linear regressions; all $p < 0.0001$; reporting adjusted R^2). Isolates depolymerized both cellulose and xylan at varying efficiencies but strains did not discriminate between carbon substrate utilization (linear regression; $p > 0.05$). Additionally, the ability to depolymerize specific carbon substrates was not a simple product of increased growth across temperatures (Supporting Information Fig. S2), underlining the differences in carbon utilization ability across strains. Total carbon depolymerization (combining cellulose and xylan assays) varied significantly by subclade designation (analysis of covariance (ANCOVA); $F_{5,195} = 175.7$, $p < 0.0001$) with a significant interaction between temperature ($F_{15,195} = 13.6$, $p < 0.0001$; Fig. 2). Maximum growth rate (μ_{max}) and biofilm formation followed similar statistical trends (Supporting Information Table S3) with subclade and temperature significantly explaining trait performance (Supporting Information Fig. S3A,B, respectively). However, temperature affected traits differently; for example, carbon depolymerization and maximum growth rate peaked at 28 °C while biofilm formation decreased linearly with increasing temperature (Fig. 2). Across all assays, trait performance was strongly influenced by temperature amplifying the degree of phenotypic variation among *Curtobacterium* isolates.

Together, the trait assays indicated that despite highly differential trait responses within *Curtobacterium,* isolates within the same subclade reflected similar phenotypic traits. By combining all observed trait variation, isolates clustered significantly by subclade, such that strains within the same genetic subclade shared more similar traits (analysis of similarities (ANOSIM); $R = 0.69$, $p < 0.001$; Fig. 3). Carbon depolymerization abilities and growth parameters [maximum absorbance (A_{max}), μ_{max} and lag phase] across temperatures best distinguished these strains while biofilm formation remained highly variable (Fig. 3). All isolates could depolymerize carbon, but the efficiency of carbon depolymerization, especially at higher temperatures, strongly differentiated subclades. For example, Subclade IVA was unable to degrade either cellulose or xylan at high temperatures, whereas Subclade IVB was generally the best degrader (Fig. 2). In contrast, the broader genetic clade designations (i.e. Clades I–V) were indistinguishable by the measured traits at varying temperatures (ANOSIM; $R = 0.14$, $p > 0.05$; Fig. 3). Thus, the degree of trait variation within subclades was highly correlated with fine-scale genetic clusters within *Curtobacterium*, denoting the existence of distinct *Curtobacterium* ecotypes.

To consider whether there was evidence for adaptation to the site from which the strains were isolated, we also tested whether the isolation site influenced trait performance. When we considered both subclade designation and the site of isolation along with the temperature of the assay, we accounted for 52% and 87% of the variation in maximum growth rate and carbon depolymerization, respectively (ANCOVA; all $p < 0.01$; reporting adjusted R^2; Supporting Information Table S3). However, subclade

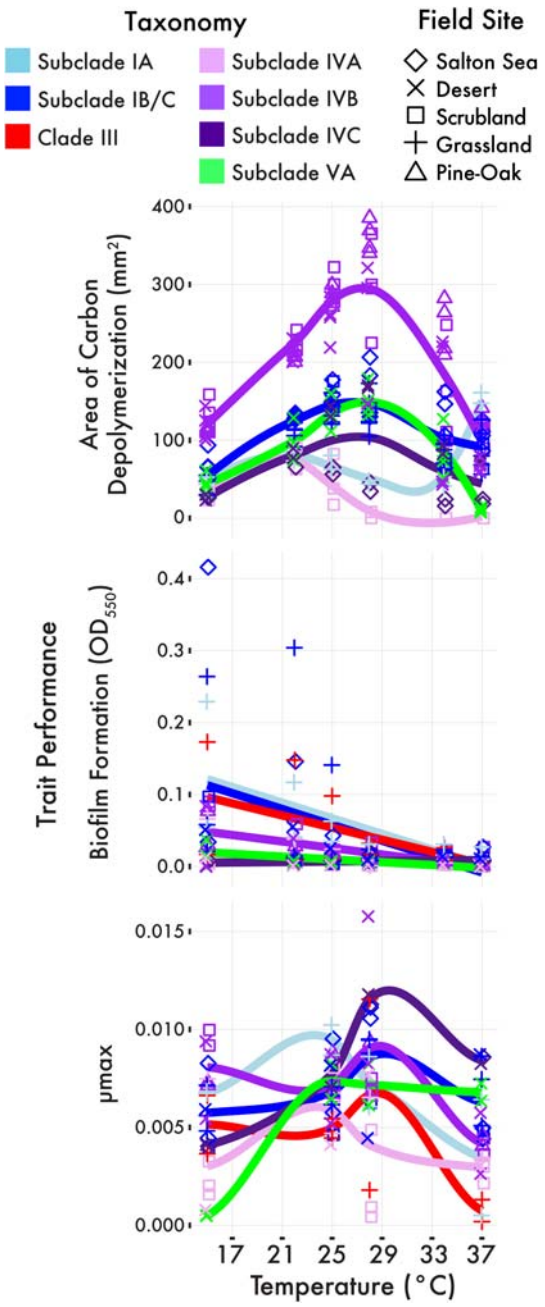

Fig. 2. Physiological response curves plotting functional traits (carbon depolymerization, biofilm formation and maximum growth rate, μ_{max}) versus temperature. The colours distinguish clade or subclade designation. The symbols represent the isolation site of each strain. Smoothed averages (lines) were calculated from locally weighted smoothing (LOESS) using either a polynomial (carbon and μ_{max}) or linear regression (biofilm).

designation explained more trait variation than site effects for both carbon depolymerization (ANCOVA; Ω^2 = 0.50 versus 0.02 respectively) and μ_{max} (ANCOVA; Ω^2 = 0.14 versus 0.06 respectively) across the temperature gradient (Fig. 2; Supporting Information Table S3). Biofilm formation, conversely, was only related to subclade and temperature (ANCOVA; adjusted R^2 = 0.29, $p < 0.001$) despite being highly variable across strains. For example, strains from both Subclades IA and IB/C were among the best biofilm producers at lower temperatures, while other strains from the same subclade and site of isolation produced minimal biofilms (Supporting Information Fig. S3B). Although site effects contributed to observed trait variation among isolates, site effects explained little additional variation beyond subclade designations to accurately distinguish ecotypes.

Biogeography of Curtobacterium and its ecotypes

To evaluate the biogeographic distribution of *Curtobacterium* ecotypes, we characterized the litter bacterial community across the climate gradient using 91 metagenomic libraries collected over 18 months. Total bacterial composition varied across the climate gradient, with Desert, Grassland and Scrubland communities more similar in composition to one another than to Pine and Subalpine communities (Supporting Information Fig. S4A). Notably, *Acidobacteria* were common in the colder, wetter sites (Pine-Oak and Subalpine), while *Actinobacteria* dominated in the hotter, drier sites (Supporting Information Fig. S4B). The Salton Sea bacterial composition was distinct from all other sites being dominated by *Proteobacteria* and the genus *Halomonas* (Supporting Information Fig. S4C). *Curtobacterium* (phylum: *Actinobacteria*) was the sixth most abundant genus across all sites and time points with an average relative abundance of 1.6% (Supporting Information Fig. S4C). Total *Curtobacterium* abundance was highest in the Grassland and decreased towards the extreme ends of the climate gradient (top line in Fig. 4A).

The geographic distribution of subclades within *Curtobacterium* – the genetic resolution at which we could distinguish ecotypes – also varied along the climate gradient (Fig. 4A). To identify subclade sequences in the metagenomic data, we extracted 830 orthologous protein groups belonging to *Curtobacterium* and *Frigoribacterium* (a sister genus) identified from the full genome sequences (see Experimental Procedures). The overall abundance of *Curtobacterium* was represented by multiple subclades, comprised primarily of six abundant ecotypes spanning the climate gradient (Fig. 4A). Nevertheless, subclade composition varied significantly by site (permutational multivariate analysis of variance (PERMANOVA; $p < 0.001$) such that some subclades were strongly correlated with site location (Fig. 4B). For example, Subclade IVB was the dominant *Curtobacterium* ecotype in the hot, dry Desert site and the cold, wet Pine-Oak and Subalpine sites; whereas, at the intermediate climate sites, Scrubland and Grassland, Subclade IB/C was the dominant ecotype (Fig. 4A). The less abundant ecotypes also exhibited differential distributions with Subclades IVB and IVC being

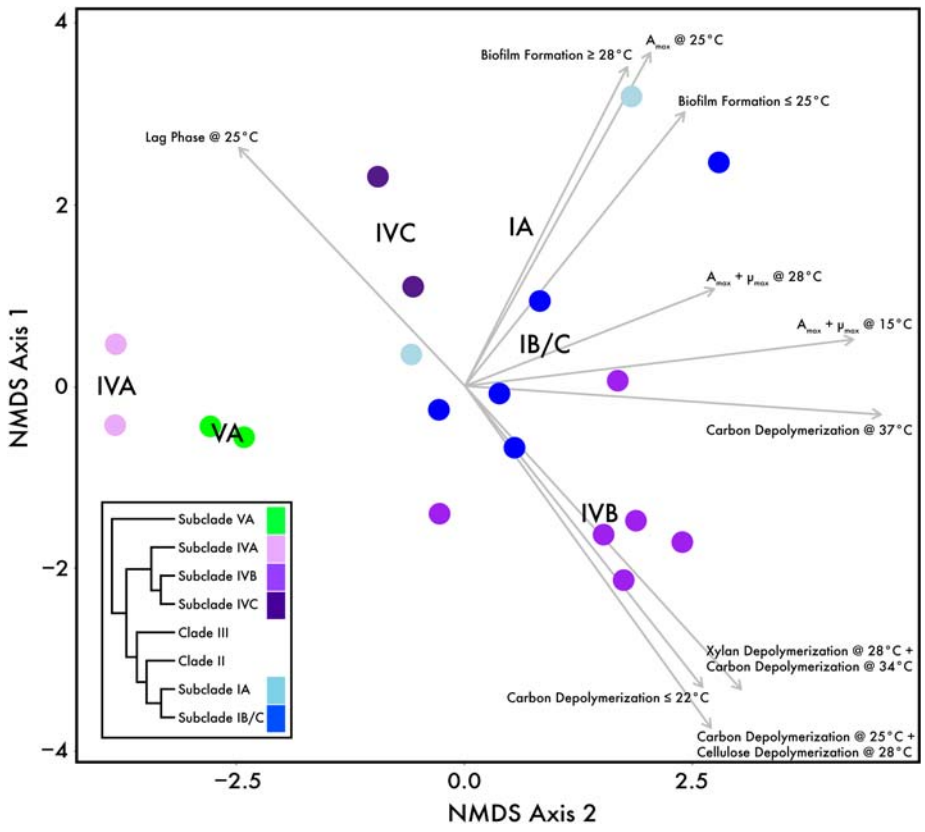

Fig. 3. Non-metric multidimensional scaling (NMDS) plot (Euclidean distance) depicting physiological variables correlated with variation in *Curtobacterium* isolates. Variables assigned as 'carbon' are collapsed to include both cellulose and xylan degradation. Each point represents an individual strain coloured by subclade. Inset is a cladogram of *Curtobacterium* clades and subclades.

more pronounced in the Desert and Subclade IA peaking in the Grassland (Fig. 4A,B). Along the climate gradient, we identified six abundant ecotypes co-occurring at each site with each ecotype exhibiting preferential distributions.

The relative abundance of the ecotypes remained relatively constant over the year and a half (Supporting Information Fig. S5A). Overall, the temporal effects were less pronounced than the site effects (PERMANOVA; $p > 0.05$), and accounted for only 0.8% of the observed variation in subclade composition across sites (as compared to 54.2% attributed to site effects). Therefore, *Curtobacterium* composition along the climate gradient was temporally stable over the course of the study.

Ecotype–environmental relationships

Since *Curtobacterium* ecotypes clearly differed in their geographic distribution, we next asked how their

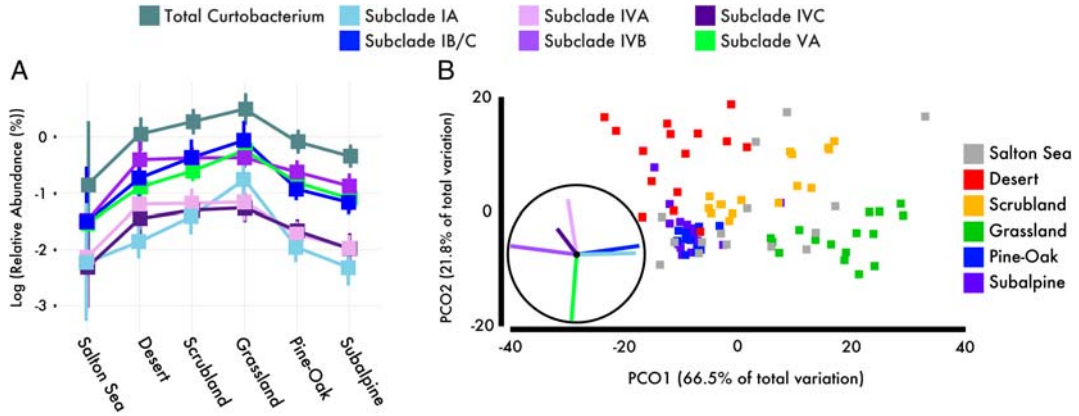

Fig. 4. The composition of *Curtobacterium* ecotypes along the climate gradient.
A. Log_{10} of the mean relative abundances of *Curtobacterium* and its subclades (± 1 SD) with respect to the entire bacterial community by site.
B. Principal coordinates (PCO) ordination plot depicting the ecotype composition in each metagenomic sample, coloured by site. Spearman correlation vectors illustrate the contribution of each subclade to compositional differences along the two PCO axes.

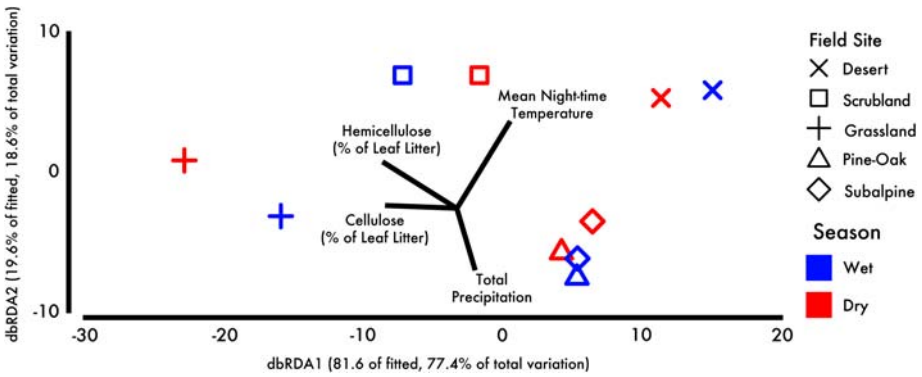

Fig. 5. Results of a distance-based redundancy analysis (dbRDA) showing ecotype composition of a sample by site (symbols) and season (colours). Ecotype composition was averaged across seasons with abiotic environmental variables averaged over the entire month of sampling. Vectors represent the direction and strength of correlations with environmental variables.

abundances were correlated with environmental variation. Ecotype composition varied over time, but this shift in composition was minimal relative to the site effects (Fig. 5). Indeed, the environmental factors measured at each site (leaf litter chemistry and abiotic parameters) largely explained the observed ecotype composition [distance-based linear model (distLM); adjusted R^2 = 0.91]. In particular, the proportion of hemicelluloses (e.g. xylan) in the leaf litter explained 41.5% of the variation in ecotype composition (distLM; $p < 0.05$; Supporting Information Table S4). The measured abiotic factors (precipitation and soil surface day- and night-time temperatures; Supporting Information Fig. S5B,C) explained an additional 38% of ecotype variation between sites and across seasons.

Despite identifying environmental factors related to ecotype composition, we were unable to link these patterns to the trait measurements. For instance, leaf litter from the Grassland and Scrubland sites contained the highest proportion of polymeric carbohydrates (cellulose and hemicelluloses; Supporting Information Fig. S5D). However, Subclade IVB, the most efficient degrader of cellulose and xylan (Fig. 2), was not the dominant ecotype at these sites; instead, Subclade IB/C was nearly twice as abundant as Subclade IVB (Fig. 4A). Furthermore, subclades whose trait performance peaked at warmer temperatures in the lab assays (e.g. Subclade IA in carbon depolymerization and Subclade VA in growth parameters; Fig. 2) were more abundant at the Grassland site rather than the warmer Desert site. Thus, while the phenotypic trait measurements strongly differentiated strains into ecotypes, these traits did not explain ecotype distribution across the sites.

Finally, given the variability in litter chemistry among sites (Supporting Information Fig. S4D), we considered whether the genetic potential to utilize a range of carbohydrates might be correlated with the distribution of *Curtobacterium* ecotypes. Specifically, we targeted the genomic diversity of glycoside hydrolase (GH) and carbohydrate binding module (CBM) proteins that potentially contribute to degradation of various carbon substrates in leaf litter. Overall, the composition of GH and CBM genes was correlated with phylogenetic distance between *Curtobacterium* strains (RELATE test; $\rho = 0.43$, $p < 0.0001$) such that more phylogenetically similar genomes encoded similar GH-CBM profiles. Furthermore, the genomic potential to degrade complex polymeric carbohydrates common in leaf litter (i.e. cellulose, chitin and xylan) differed significantly between subclades (Kruskal–Wallis test, $p < 0.0001$; Supporting Information Fig. S6A). However, the total abundance of polymeric GH/CBM did not clearly predict ecotype distribution along the gradient. We predicted that ecotypes with higher numbers of polymeric GH-CBMs would be more abundant on leaf litter; however, two of the rarer ecotypes, Subclades IA and IVA, contained the highest total number of polymeric GH-CBMs (Supporting Information Fig. S6A). Similarly, total GH and CBM composition also varied significantly by subclade (PERMANOVA; $p < 0.001$; Supporting Information Fig. S6B), but ecotypes with highly similar GH-CBM compositions, such as Subclades IVA and IVB, differed strikingly in their association with different sites (Fig. 4B). Therefore, while the abundance and composition of GH-CBMs in *Curtobacterium* genomes supported our ecotype designations (Supporting Information Fig. S6B), variation in these functional genes did not elucidate ecotype distributions along the climate gradient.

Discussion

In this study, we applied a trait-based framework (Diaz *et al*., 1998; Dıaz and Cabido, 2001; Cadotte *et al*., 2015) to identify ecological populations, or ecotypes, in a terrestrial bacterium and investigated the drivers of their biogeographic distribution. By sampling across a climate gradient varying in temperature and precipitation, we identified highly similar genomic clusters within *Curtobacterium* that corresponded to distinct phenotypes denoting the existence of bacterial ecotypes. These results contribute to the growing understanding that traditional taxonomic assignments (i.e. OTUs) mask bacterial 'microdiversity' that contributes to ecological differentiation (Jaspers and

Overmann, 2004; Larkin and Martiny, 2017). More broadly, our study highlights the application of a trait-based approach to microbial systems to assess the ecological and evolutionary mechanisms contributing to community assembly (McGill *et al*., 2006; Nemergut *et al*., 2013; Enquist *et al*., 2015).

A growing number of studies demonstrate that the fine-scale genomic structure of bacterial diversity reflects diverged populations (Polz *et al*., 2006; Connor *et al*., 2010) occupying separate ecological niches (Johnson *et al*., 2006; Hunt *et al*., 2008). Indeed, it appears that the total abundance of typically defined taxa (i.e. OTUs based on 16S rRNA sequence similarity) may often be comprised of distinct ecological populations that vary over a range of environments (Moore *et al*., 1998; Thompson *et al*., 2005). This emerging pattern has implications for how we interpret biogeographic patterns of bacterial diversity. In particular, phenotypic differences among ecotypes can permit the coexistence of fine-scale genetic diversity within an environment. In this study, we identified and observed six abundant *Curtobacterium* ecotypes at all sites along our gradient, suggesting fine-scale niche partitioning of environmental resources. In addition, ecotypic diversity may allow a taxon to persist in a broader range of environments than would be expected based on the phenotype of a single representative (Moore *et al*., 1998; Partensky *et al*., 1999). Therefore, the biogeographic distribution of a typical OTU or a representative strain may not be indicative of the range of genetic and phenotypic diversity encoded at finer taxonomic levels.

Whether an isolate is representative of its broader taxon will depend on the particular trait of interest (McLaren and Callahan, 2018), as different traits vary in the degree to which they are phylogenetically conserved (Martiny *et al*., 2013). Many traits are conserved across all *Curtobacterium* diversity including those contributing to its dominance as a leaf litter bacterium (Chase *et al*., 2016). For instance, all strains in this study shared the ability to degrade polymeric carbohydrates common in leaf litter, cellulose and xylan, and, relative to other genera in the Microbacteriaceae family, *Curtobacterium* has a high genomic potential for carbohydrate degradation (as assessed by the total number of GH-CBM genes; Chase *et al*., 2016). Additionally, the taxon generally appears to prefer relatively dry surface soil conditions (Lennon *et al*., 2012) as the relative abundance of *Curtobacterium* as a whole tends to increase in drier seasons (Chase *et al*., 2017). In contrast, traits that vary within the genus will contribute to ecological differences among ecotypes. Such fine-scale trait variation may often result in quantitative rather than qualitative differences. For example, *Curtobacterium* ecotypes varied in their growth rates at different temperatures. And while all *Curtobacterium* ecotypes could degrade cellulose and xylan, the degree to which they degraded these compounds in the lab varied significantly.

Although we were able to identify correlations between ecotype composition and environmental factors across the sites, it was not clear which traits underlie *Curtobacterium* ecotype distributions as has been resolved for marine bacteria (Johnson *et al*., 2006; Martiny *et al*., 2009). There are several possible reasons for this disconnect. One possibility is that we did not measure the correct traits. For example, we hypothesized that the ability to form biofilms might be important because biofilms can protect bacteria from desiccation and fluctuations in water potential (Hartel and Alexander, 1986; Roberson and Firestone, 1992) and are correlated with soil moisture adaptation (Lennon *et al*., 2012). Thus, we expected biofilm formation to be prevalent across all *Curtobacterium* strains, especially with higher production in strains abundant at drier sites. However, biofilm formation was highly variable among strains, so much so that subclade differences explained little-observed variation and there was no effect from the site of isolation. Of course, biofilm formation is just one trait that might contribute to moisture adaptation (Potts, 1994) and other traits related to moisture preference might be more predictive for assessing fine-scale niche partitioning. We also did not measure a variety of traits that known to be important to soil bacteria including nutrient uptake abilities and pH preferences (Fierer and Jackson, 2006; Leff *et al*., 2015). Environmental constraints clearly contribute to the distribution of soil bacterial taxa, however, the traits delineating these biogeographic patterns require further investigation.

A second reason that we may have missed ecotype–environment correlations is that we are not measuring the environment at the correct spatiotemporal scale. Soils are highly heterogeneous and differences in soil microhabitats are thought to contribute to the maintenance of soil diversity (Ranjard and Richaume, 2001; Nannipieri *et al*., 2003). Consequently, soil ecotypes are likely to respond to environmental variation at very small spatial scales. The existence of multiple *Curtobacterium* ecotypes co-occurring within a given site suggests that fine-scale environmental variation is contributing to niche partitioning in leaf litter. Indeed, even in the marine water column, which is thought to be more homogeneous than soil, strains of *Vibrio splendidus* partition resources to differentiate between particle-associated or free-living habitats (Hunt *et al*., 2008). On a similar spatial scale, variation in hemicellulose availability or temperature within a decomposing leaf may explain the coexistence of multiple *Curtobacterium* ecotypes. Thus, by sampling across a regional climate gradient, we may have masked much of the within-site environmental variation that contributes to soil ecotype distributions. A further possibility

is that *Curtobacterium* diversity is not at equilibrium in the sampled communities. Maladapted strains may be present and even abundant if environmental selection is weak and/or dispersal is high (Lenormand, 2002). Much more work is needed to understand the spatiotemporal scales of these mechanisms for soil bacterial diversity.

In sum, our study presents evidence that the genomic diversity within an abundant terrestrial bacterial taxon can be classified into ecotypes that vary in their biogeographic distribution across a climate gradient. Especially for terrestrial soil communities, we lack an understanding of the ecological and evolutionary processes governing the distribution and functioning of bacterial diversity. The results presented here are consistent with the growing understanding that fine-scale genomic diversity, and the traits encoded by this variation, is key to microbial biogeography. However, identifying and measuring relevant traits remains a distinct challenge for the application of trait-based frameworks to microbial communities.

Experimental procedures

Field sites

We characterized the microbial community on leaf litter by establishing four replicate plots (1 m^2) at six sites across a climate gradient in southern California from October 2015 to April 2017 (Glassman *et al*., in preparation). The five sites (from lowest to highest elevation) include the Sonoran desert (33.652 N, 116.372 W), pinyon-juniper scrubland (33.605 N, 116.455 W), coastal grassland (33.737 N, 117.695 W), pine-oak forest (33.808 N, 116.772 W) and subalpine forest (33.824 N, 116.755 W) as previously described in Baker and Allison, (2017). In addition, we sampled leaf litter near the highly-saline Salton Sea (33.518 N, 115.938 W) to extend the climate gradient further (Supporting Information Table S1). Sites are hereafter referred to as Desert, Scrubland, Grassland, Pine-Oak, Subalpine and Salton Sea, respectively. All sites experience Mediterranean climate patterns with a hot, dry summer and a cool, wet winter. The sites range in mean annual air temperature (MAT) from 10.3 to 24.6 °C and precipitation (MAP) from 80 to 400 mm. To characterize climate at the sites during the experiment, we collated precipitation data from nearby weather stations and collected surface soil temperature at 90 min intervals using two iButton temperature sensors (Maxim Integrated) from 4th April 2016 to 20th April 2017 at five of the sites (excluding Salton Sea; Glassman *et al*., in preparation). In addition to changes in temperature and precipitation, the sites differed greatly in the plant communities present and, therefore, the litter chemistry. Leaf litter chemistry was determined from samples in both the dry (June) and wet (December) seasons in 2015 using near-IR spectroscopy, as previously described (Baker and Allison, 2017).

Isolation and genomic characterization of Curtobacterium

To isolate *Curtobacterium* strains, we collected fresh leaf litter from the perimeter of the four plots at each site on 14th June 2016 to create a homogenized batch of litter from each site. We ground the litter in a sterile coffee grinder and vortexed 0.2 g of homogenized litter in 5 ml of 0.9% saline (NaCl) solution for 5 min. Samples were serially diluted and plated on grassland leaf litter leachate media (Chase *et al*., 2016). Colonies were visually screened for phenotypic characteristics ascribed to *Curtobacterium* (Evtushenko and Takeuchi, 2006), streaked on Luria Broth (LB) media agar plates, transferred three times and stored in glycerol solution at −80 °C. We identified each cultured isolate by PCR amplification and Sanger sequencing of a 1500 bp region of the 16S rRNA region. For each isolate, we used DNA extracted from a single colony that we added to a PCR cocktail containing 0.3 μl HotMaster Taq polymerase (5 units/μl), 15 μl 2× Premix F (Epicentre; Madison, WI), and 0.2 μl of 50 μM of each primer, pA (5′-AGAGTTTGATCCTGGCTCAG-3′) and pH' (5′-AAGGAGGTGATCCAGCCGCA-3′), under identical PCR conditions (Chase *et al*., 2016). The 16S rRNA sequence of each isolate was used to identify the taxonomy using the Ribosomal Database Project (RDP) database (Wang *et al*., 2007).

Identified *Curtobacterium* isolates were selected for whole-genome sequencing and grown on LB plates for 48–72 h. A single colony from each plate was transferred to 10 ml liquid LB media to grow for an additional 48 h. Genomic DNA extraction was performed using the Wizard Genomic DNA Purification Kit (Promega; Madison, WI) with the additional step of adding lysozyme for Gram-positive bacteria. Extracted DNA was quantified on the Qubit (BioTek; Winooski, VT), quality assessed on the Nanodrop (Thermo Fisher; Waltham, MA), and diluted to 0.5 ng/μl for library preparation. Next, we followed the protocol for the Nextera XT Library Preparation kit (Illumina Inc., San Diego, CA). Samples were pooled in equimolar portions and assessed using the High Sensitivity Bioanalyzer. The pooled library was sequenced using an Illumina HiSeq4000 instrument (Illumina Inc., San Diego, CA) with 150 bp paired-end reads. Demultiplexed sequence data were assembled using the SPAdes genome assembler (Bankevich *et al*., 2012) with a 'careful' iterative k-step ranging from k = 31 to 111. We assessed the quality of the assemblies by creating taxon-annotated-GC-coverage (TAGC) plots. Specifically, we calculated coverage for each contig by mapping back the raw sequence data to assembled contigs using Bowtie2

(Langmead and Salzberg, 2012) and taxonomic assignments were assigned using MegaBLAST against the NCBI nucleotide database (Federhen, 2012) with an *E* value of 1×10^{-5}. Based on the results from the TAGC-plots, we discarded all contigs with coverage < 30, length < 500 bp and GC% <55%. In total, we identified 56 high-quality *Curtobacterium* genomes to be included in this study, which are deposited at GenBank under BioProject PRJNA391502 with biosamples SAMN09009025–SAMN09009080.

We created a *Curtobacterium* phylogeny using a multilocus sequence alignment (MLSA) of 21 single-copy marker genes (Wu *et al*., 2013). For comparison of the climate gradient genomes (*N* = 56), we downloaded all publicly available *Curtobacterium* genomes (*N* = 30) and a *Frigoribacterium* genome (to serve as an outgroup), which included 14 previously identified *Curtobacterium* isolates from our previous work in leaf litter (Chase *et al*., 2016, 2017). Each of the 87 genomes was translated using Prodigal (Hyatt *et al*., 2010) and screened for the presence of the 21 marker genes using HMMER version 3.1b2 (Finn *et al*., 2011) with an *E* value of 1×10^{-10}. Each marker gene was independently aligned using ClustalO version 1.2.0 (Sievers *et al*., 2011) to create a concatenated protein alignment consisting of 3947 amino acids for phylogenetic analysis using RAxML version 8.0.0 (Stamatakis, 2014) under the PROTGAMMAWAG model for 100 replicates. We designated the major branching points in the resulting phylogeny into five distinct clades. To identify finer taxonomic groupings, we calculated pairwise average amino acid identity (AAI) and nucleotide identity (ANI) across all 87 genomes using the enveomics package (Rodriguez-R and Konstantinidis, 2016). Genomes that clustered at ≥90% AAI at the whole-genome level, the suggested boundaries for bacterial species groupings (Richter and Rosselló-Móra, 2009), were further designated into subclades. Subclade designations were also supported by the phylogeny.

To cluster genomes into operational taxonomic units (OTUs), we extracted 16S rRNA gene sequences from the full genomes using Barrnap (http://www.vicbioinformatics.com/software.barrnap.shtml) and conducted two analyses recommended for optimal assessment of taxonomic units (Edgar, 2018). First, we extracted the hypervariable V4/V5 region of the 16S rRNA gene and defined OTUs at 100% gene similarity with UCLUST (Edgar, 2010), also termed zero-radius OTUs (zOTUs) or exact sequence variants (ESVs). To include effects of alignment quality (Schloss, 2010), we aligned the full-length 16S rRNA gene region with SINA (Pruesse *et al*., 2012) then clustered at 99% gene similarity with mother (Schloss *et al*., 2009). We conducted a phylogenetic analysis of the full-length, aligned 16S rRNA gene region using RAxML version 8.0.0 (Stamatakis, 2014) under the GTRGAMMA model for 100 replicates.

We characterized the functional potential to degrade carbohydrates [glycoside hydrolase (GH) and carbohydrate binding module (CBM) proteins] within all *Curtobacterium* genomes. The predicted open reading frames generated from Prodigal were searched using HMMER against the Pfam-A version 30.0 database (Finn *et al*., 2016). GH and CBM genes and their targeted substrate were identified according to the Pfam identifiers as stated in Chase *et al*. (2016). Total GH and CBM gene composition profiles for each genome were normalized and used to construct a Euclidean distance matrix for producing an ordination plot.

Characterization of Curtobacterium *traits*

In the laboratory, we characterized the traits of a subset of the *Curtobacterium* isolates spanning across the climate gradient and phylogenetic clades. Specifically, we sought to measure four functional traits (growth, biofilm formation and depolymerization of cellulose and xylan) across a temperature range (15–42 °C) experienced along the climate gradient. We selected these traits because we speculated that they would influence competitive dynamics in the leaf litter community. The ability to degrade polymeric carbohydrates and, specifically, an increased degradation efficiency should provide a competitive advantage as the primary carbon supply in leaf litter is in the form of celluloses and hemicelluloses (e.g. xylan; Baker and Allison, 2017). Our sites experience long periods without precipitation and, therefore, the ability to form biofilms may prevent desiccation from water stress (Lennon *et al*., 2012). Increased growth, both in response (lag phase) and rate (μ_{max}), could allow for the competitive exclusion of other organisms. Traits were assayed along the temperature gradient to simulate abiotic conditions from the climate gradient.

For all assays, a subset of *Curtobacterium* strains and one *Escherichia coli* strain (as a control) were grown from −80 °C freezer stocks for 48 h in liquid LB media at 22 °C. Isolates were pelleted by spinning down at 4500 RPM for 10 min, washed three times with 0.9% saline solution to remove residual media, and resuspended in 10 ml of M63 minimal media (supplemented with 0.1% peptone and 1 μg/ml thiamine) with 0.5% (wt/vol) dextrose as the sole carbon source. After 24 h, isolates were washed again under identical conditions and diluted to an optical density of 0.1 OD_{600} to ensure equal cell density across all isolates.

For the growth rate and biofilm assays, we inoculated 10 μl of diluted isolates (*N* = 29 *Curtobacterium* isolates) into 96-well plates containing 190 μl of M63 media with 0.5% (wt/vol) dextrose. Each strain was grown in

triplicate on each plate for each assay. The inoculated plates for the growth rate assays were shaken at 200 RPM at four temperatures (15, 25, 28 and 37 °C) with OD_{600} being measured every 1–2 h for the first 48 h and every 4 h thereafter. Sampling was terminated if any of the six negative controls in any plate increased in OD_{600} measurements over the course of the experiment. To estimate growth parameters [max absorbance (A_{max}), max growth rate (μ_{max}) and lag phase], we fit OD_{600} measurements to either a logistic, gompertz or a locally weighted scatterplot (LOESS) regression model using the 'growthcurve' package in the R software environment (Pinheiro *et al*., 2011). For biofilm assays, inoculated plates were sealed and placed in incubators at six temperatures (15, 22, 25, 28, 34 and 37 °C) without shaking. After 4 days, we removed residual cells and media by submerging the microplates in deionized water. We then added 125 μl of 0.1% crystal violet solution to each well and incubated the plates for 15 min at room temperature. Plates were re-submerged in water and vigorously shaken to remove residual liquid (repeated 4×). We dried each plate for 2 h and added 125 μl of 30% acetic acid to solubilize the crystal violet. Plates were incubated for 15 min and absorbance was measured at OD_{550} for biofilm production (O'Toole, 2011).

We selected a subset of *Curtobacterium* isolates (N = 18) from various sites along the climate gradient to assess their ability to depolymerize cellulose and xylan as previously described (Chase *et al*., 2017). Briefly, we inoculated 10 μl of washed 0.1 OD_{600} cultures, in triplicate, onto solid M63 media with 0.5% (wt/vol) of either carboxymethylcellulose (CMC; catalogue no. 150560; MP Biomedicals, Santa Ana, CA) or xylan (catalogue no. X0502; Sigma, St. Louis, MO) and placed inoculated plates in incubators at seven temperatures (15, 22, 25, 28, 34, 37 and 42 °C). We classified the zones of depolymerization after 4 days using ImageJ (https://imagej.nih.gov/ij/) by subtracting the original colony area from the total area of carbohydrate degradation. An *E. coli* strain was included as a negative control and did not depolymerize either substrate at any temperature. No strains could depolymerize either substrate at 42 °C and, therefore, were removed from statistical analyses.

Metagenomic sequencing and analysis

Metagenomic samples. We sampled leaf litter from the four replicate plots at each site every 6 months until 20th April 2017 (6 sites × 4 time points × 4 replicate plots). We extracted DNA from 0.05 g of ground leaf litter using the FastDNA SPIN Kit for Soil (Mo Bio, Carlsbad, CA) and cleaned the DNA with the Genomic DNA Clean and Concentrator kit (Zymo Research, Irvine, CA). Cleaned samples were diluted to 0.5 ng/μl and 1 ng of DNA was used for input for the Nextera XT library Prep kit for sequencing on the Illumina HiSeq 4000 instrument with 150 bp paired-end reads. Due to the low quality sequence data, we excluded five libraries and, in total, analysed 91 metagenomic libraries. The raw data is deposited on the metagenomics analysis server (MG-RAST) (Meyer *et al*., 2008) under the project ID mgp17355.

Bacterial community analysis. To characterize the bacterial litter community, we built upon our previous pipeline (Chase *et al*., 2017) using phylogenetic inference to characterize conserved single-copy marker genes (Wu *et al*., 2013) within the metagenomic data. To compensate for the lack of genomic representation of soil microbes, we downloaded 7392 publicly available genomes that are designated as 'representative' genomes by the PATRIC database (Wattam *et al*., 2014) and included representative *Curtobacterium* genomes from the climate gradient (see above). We translated all genomes using Prodigal (Hyatt *et al*., 2010) and searched for the presence of 21 single-copy marker genes using HMMER version 3.1b2 (Finn *et al*., 2011) with an E value of 1×10^{-10}. Each protein was individually aligned with ClustalO version 1.2.0 (Sievers *et al*., 2011) and used to create a 12 271 amino acid concatenated alignment for phylogenetic analysis using FastTree2 (Price *et al*., 2010). The reference tree was manually curated for the misplacement of genomes based on assigned nomenclature, and a genome was removed if it did not group within its assigned family designation. This highly curated tree served as a reference tree to guide construction of each individual marker gene tree using RAxML version 8.0.0 (Stamatakis, 2014) under the PROTGAMMAWAG model for 100 replicates. The remaining 5433 genomes were used to construct BLASTp (Altschul *et al*., 1997) databases, HMMER profiles and pplacer (Matsen *et al*., 2010) reference packages (all databases available at https://github.com/alex-b-chase/elevation-community).

Metagenomic libraries were quality trimmed using BBMap (Bushnell, 2016) and filtered to remove eukaryotic DNA. Specifically, we mapped all reads to a reference genome using BWA (Li and Durbin, 2009) from both an abundant grass (*Lolium perenne*; Accession: MEHO01000000) and fungus (*Pyrenophora* teres; Accession: NZ_AEEY00000000) found at the grassland site. All filtered reads were then merged using BBMap (Bushnell, 2016) to form paired-end reads. If a read could not be merged with its counterpart, we included only the forward read in further analyses. Reads were then translated using Prodigal (Hyatt *et al*., 2010) with the metagenomic flag and searched against the reference marker gene databases, as previously described (Chase *et al*., 2017). Briefly, we imposed a primary filter against the reference BLASTp database with an E value of 1×10^{-5} and a secondary filter against the reference HMMER profiles with an E value ranging from 1×10^{-10} to 1×10^{-25} depending on the individual marker gene. Passed reads were aligned using ClustalO version 1.2.0 (Sievers *et al*., 2011) to the corresponding reference package and placed onto the reference phylogenies using pplacer v.1.1.alpha17 (Matsen *et al*., 2010). Relative abundances were calculated by generating single branch abundance matrices and normalizing to the total number of marker genes present in each library.

Curtobacterium ecotype abundances. The above analyses provided an estimate of the total abundance of *Curtobacterium*

and other taxa in the metagenomic libraries. However, to investigate the distribution of diversity within *Curtobacterium*, we first characterized *Curtobacterium* orthologous protein groups (orthologs) from the *Curtobacterium* genomes isolated from leaf litter. Publicly available *Frigoribacterium* genomes ($N = 5$) were also included to serve as outgroups. Orthologs were identified using Roary (Page *et al*., 2015) with coding regions predicted by Prokka (Seemann, 2014). Due to the diversity of these genomes, we decreased the percentage sequence identity to 50% to encompass all possible orthologs. The resulting 1075 orthologs were used to create a core-genome tree, using RAxML version 8.0.0 (Stamatakis, 2014) under the PROTGAMMAWAG model for 100 replicates, that was nearly identical to the reference tree derived from the genomic MLSA analysis. We built individual ortholog trees, using identical model parameters, with the core-genome tree as the guiding reference tree, to generate a *Curtobacterium* reference database (reference database can be found here: https://github.com/alex-b-chase/elevation-curto). We then removed orthologs that lacked a robust phylogenetic signal yielding a final set of 830 orthologs. We parsed the filtered metagenomic reads for the presence of each ortholog with a BLASTp *E* value of 1×10^{-20} and a secondary filter against the reference HMMER profiles with an *E* value of 1×10^{-40}. Each filtered metagenomic read was then placed onto the corresponding ortholog tree with pplacer version 1.1.alpha17 (Matsen *et al*., 2010) and classified to each clade and subclade. Clade and subclade relative abundances were normalized by the total abundance of *Curtobacterium* calculated from the community analyses above. For the remainder of the subclade compositional analyses, subclades were treated as the proportion to all *Curtobacterium*, not the entire community, to limit compositional biases.

Statistical analyses

Ecotype identification – linking traits to phylogeny. To tease apart the relative importance of isolation source (where the *Curtobacterium* strain was isolated from along the climate gradient) and phylogenetic relatedness (subclade designation) for each of our physiological assays, we implemented a statistical model with the site of isolation and subclade designation as dependent variables with temperature as a covariate. To start, we examined various regression models to test the best model fit using Bayesian information criterion (BIC) to confirm that temperature covaried with the other variables across all assays. For each assay, we then determined whether our regression models should be either linear or polynomial by comparing both BIC and residual values for each model. We constructed a linear regression model for biofilm formation and polynomial regression models for carbon degradation and growth rate. Finally, we used an analysis of covariance (ANCOVA) to test the effects of our main fixed factors, site and subclade, while controlling for the effects of the covariate, temperature. Within each ANCOVA design, we implemented a backward selection process (Mac Nally, 2002) to eliminate spurious relationships (Harrell, 2015) for each assay.

To further examine the physiological differences between *Curtobacterium* subclades, we constructed a non-metric multidimensional scales (NMDS) ordination plot of each strain using the physiological measurements. Specifically, we included biofilm formation (at 6 temperatures), cellulose degradation (6 temperatures), xylan degradation (6 temperatures), A_{max} (4 temperatures), μ_{max} (4 temperatures) and lag phase (4 temperatures). All variables were normalized by subtracting the mean from each measurement and dividing by the standard deviation. Before performing the NMDS analysis, we generated Spearman's correlation coefficients (ρ^2) for each physiological assay and clustered variables into groups when $\rho^2 > 0.6$. We kept one representative trait for each Spearman-defined cluster and generated a Euclidean similarity matrix across strains. Next, we fitted each physiological variable onto the ordination plot and calculated the significance of each variable over 9999 permutations. Finally, we removed nonsignificant variables to reduce spurious relationships (Harrell, 2015) and reran all analyses. We report only the ordination plot generated from the remaining significant variables for each strain. The significance of strain groupings was assessed using an analysis of similarities (ANOSIM) for subclade or clade designation and site of isolation for 9999 permutations. All analyses were performed in the R software environment.

Ecotype distributions along the climate gradient. To test the effects of site on the distribution of *Curtobacterium* subclade composition, we used a permutational multivariate analysis of variance (PERMANOVA; Clarke, 1993). The statistical model included the site along the climate gradient and season (wet or dry) as fixed effects. We generated a Bray–Curtis similarity matrix to run a type III partial sum of squares for 9999 permutations of residuals under a reduced PERMANOVA model. The Bray–Curtis matrix was also used to generate principal coordinates analysis (PCO) ordination plot. To assess the effects of the abiotic environment (surface soil day- and night-time temperature and total precipitation) and leaf litter chemistry (i.e. cellulose and hemicellulose) on subclade composition, we applied a distLM. Again, the Bray–Curtis matrix for subclade composition was analysed using a step-wise forward procedure with adjusted R^2 as the model selection criterion. All multivariate statistical analyses were conducted using PRIMER6 with the PERMANOVA+ function (Primer-E Ltd., Ivybridge, UK).

Acknowledgements

The authors thank Claudia Weihe for all her work organizing the climate gradient project and also thank Chamee Moua for her assistance in isolating strains, Brandon Gaut for his advisement on data analysis, Alyssa Kent and Andrew Oliver for computational assistance and Gavin Lear and the Martiny lab for their helpful comments on earlier revisions. The authors thank Michaeline Albright, Nameer Baker, Sydney Glassman, Mike Goulden, and Kathleen Treseder for their assistance in data collection. And finally, to the authors thank the High-Performance Computing Cluster at UCI, specifically Harry Mangalam and Joseph Farran, for their continued help and support for computational resources. This work was supported by an US Department of Education Graduate Assistance of National Need (GAANN) fellowship to ABC, a UC MEXUS-CONACYT post-doctoral research fellowship to ZGL, a National Institutes of Health Maximizing Access to

Research Careers (MARC) grant (GM-69337) to AEL, a National Science Foundation grant (DEB-1457160), and US Department of Energy Office of Science, Biological and Environmental Research awards (DE-PS02-09ER09-25 and DE-SC0016410).

References

Allison, S. D. (2012) A trait-based approach for modelling microbial litter decomposition. *Ecol Lett* **15**: 1058–1070.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST:a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Amend, A. S., Martiny, A. C., Allison, S. D., Berlemont, R., Goulden, M. L., Lu, Y., *et al*. (2016) Microbial response to simulated global change is phylogenetically conserved and linked with functional potential. *ISME J* **10**: 109–118.

Baker, N. R., and Allison, S. D. (2017) Extracellular enzyme kinetics and thermodynamics along a climate gradient in southern California. *Soil Biol Biochem* **114**: 82–92.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., *et al*. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.

Bushnell, B. (2016) BBMap short read aligner. Berkeley, CA: University of California. URL http//sourceforge.net/projects/bbmap.

Cadotte, M. W., Arnillas, C. A., Livingstone, S. W., and Yasui, S.-L. E. (2015) Predicting communities from functional traits. *Trends Ecol Evol* **30**: 510–511.

Chase, A. B., Arevalo, P., Polz, M. F., Berlemont, R., and Martiny, J. B. H. (2016) Evidence for ecological flexibility in the cosmopolitan genus *Curtobacterium*. *Front Microbiol* **7**: 1874.

Chase, A. B., Karaoz, U., Brodie, E. L., Gomez-Lunar, Z., Martiny, A. C., and Martiny, J. B. H. (2017) Microdiversity of an abundant terrestrial bacterium encompasses extensive variation in ecologically relevant traits. *MBio* **8**: e0180917.

Chase, A. B., and Martiny, J. B. H. (2018) The importance of resolving biogeographic patterns of microbial microdiversity. *Microbiol Aust* **39**: 5–8.

Choudoir, M. J., and Buckley, D. H. (2018) Phylogenetic conservatism of thermal traits explains dispersal limitation and genomic differentiation of *Streptomyces* sister-taxa. *ISME J* **12**: 2176–2186.

Clarke, K. R. (1993) Non-parametric multivariate analyses of changes in community structure. *Austral Ecol* **18**: 117–143.

Cohan, F. M. (2001) Bacterial species and speciation. *Syst Biol* **50**: 513–524.

Connor, N., Sikorski, J., Rooney, A. P., Kopac, S., Koeppel, A. F., Burger, A., *et al*. (2010) Ecology of speciation in the genus bacillus. *Appl Environ Microbiol* **76**: 1349–1358.

Delmont, T. O., and Eren, A. M. (2018) Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* **6**: e4320.

Diaz, S., Cabido, M., and Casanoves, F. (1998) Plant functional traits and environmental filters at a regional scale. *J Veg Sci* **9**: 113–122.

Dıaz, S., and Cabido, M. (2001) Vive la difference: plant functional diversity matters to ecosystem processes. *Trends Ecol Evol* **16**: 646–655.

Edgar, R. C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Edgar, R. C. (2018) Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* **1**: 5.

Edwards, K. F., Litchman, E., and Klausmeier, C. A. (2013) Functional traits explain phytoplankton community structure and seasonal dynamics in a marine ecosystem. *Ecol Lett* **16**: 56–63.

Enquist, B. J., Norberg, J., Bonser, S. P., Violle, C., Webb, C. T., Henderson, A., *et al*. (2015) Scaling from traits to ecosystems: developing a general trait driver theory via integrating trait-based and metabolic scaling theories. *Adv Ecol Res* **52**: 249–318.

Evtushenko, L., and Takeuchi, M. (2006) *The Prokaryotes*. New York: Springer.

Federhen, S. (2012) The NCBI taxonomy. *Nucleic Acids Res* **40**: D136–D143.

Fierer, N., and Jackson, R. B. (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* **103**: 626–631.

Finn, R. D., Clements, J., and Eddy, S. R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**: 29–37.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., *et al*. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**: D279–D285.

Glassman, S. I., Weihe, C., Li, J., Albright, M. B. N., Looby, C. I., Martiny, A. C., Treseder, K. K., Allison, S. D., and Martiny, J. B. H. (2018) Decomposition responses to climate depends on microbial community composition. *PNAS* (In revision).

Green, J. L., Bohannan, B. J. M., and Whitaker, R. J. (2008) Microbial biogeography: from taxonomy to traits. *Science* **320**: 1039–1043.

Harrell, F. (2015) *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. New York: Springer.

Hartel, P. G., and Alexander, M. (1986) Role of extracellular polysaccharide production and clays in the desiccation tolerance of cowpea *Bradyrhizobia*. *Soil Sci Soc Am J* **50**: 1193–1198.

Hunt, D. E., David, L. A., Gevers, D., Preheim, S. P., Alm, E. J., and Polz, M. F. (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 1081–1085.

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

Jaspers, E., and Overmann, J. (2004) Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. *Appl Environ Microbiol* **70**: 4831–4839.

Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S., and Chisholm, S. W. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**: 1737–1740.

Kent, A. G., Dupont, C. L., Yooseph, S., and Martiny, A. C. (2016) Global biogeography of *Prochlorococcus* genome diversity in the surface ocean. *ISME J* **10**: 1856–1865.

Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Larkin, A. A., and Martiny, A. C. (2017) Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environ Microbiol Rep* **9**: 55–70.

Lavorel, S., and Garnier, É. (2002) Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the holy grail. *Funct Ecol* **16**: 545–556.

Leff, J. W., Jones, S. E., Prober, S. M., Barberán, A., Borer, E. T., Firn, J. L., *et al*. (2015) Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proc Natl Acad Sci USA* **112**: 10967–10972.

Lennon, J. T., Aanderud, Z. T., Lehmkuhl, B. K., and Schoolmaster, D. R. (2012) Mapping the niche space of soil microorganisms using taxonomy and traits. *Ecology* **93**: 1867–1879.

Lenormand, T. (2002) Gene flow and the limits to natural selection. *Trends Ecol Evol* **17**: 183–189.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* **25**: 1754–1760.

Litchman, E., Edwards, K. F., and Klausmeier, C. A. (2015) Microbial resource utilization traits and trade-offs: implications for community structure, functioning, and biogeochemical impacts at present and in the future. *Front Microbiol* **6**: 254.

Litchman, E., and Klausmeier, C. A. (2008) Trait-based community ecology of phytoplankton. *Annu Rev Ecol Evol Syst* **39**: 615–639.

Martiny, A. C., Tai, A. P. K., Veneziano, D., Primeau, F., and Chisholm, S. W. (2009) Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ Microbiol* **11**: 823–832.

Martiny, A. C., Treseder, K., and Pusch, G. (2013) Phylogenetic conservatism of functional traits in microorganisms. *ISME J* **7**: 830–838.

Martiny, J. B. H., Jones, S. E., Lennon, J. T., and Martiny, A. C. (2015) Microbiomes in light of traits: a phylogenetic perspective. *Science* **350**: aac9323.

Matsen, F. A., Kodner, R. B., and Armbrust, E. V. (2010) Pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**: 538.

Matulich, K. L., Weihe, C., Allison, S. D., Amend, A. S., Berlemont, R., Goulden, M. L., *et al*. (2015) Temporal variation overshadows the response of leaf litter microbial communities to simulated global change. *ISME J* **9**: 2477–2489.

McGill, B. J., Enquist, B. J., Weiher, E., and Westoby, M. (2006) Rebuilding community ecology from functional traits. *Trends Ecol Evol* **21**: 178–185.

McLaren, M. R., and Callahan, B. J. (2018) In nature, there is only diversity. *MBio* **9**: e0214917.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., *et al*. (2008) The metagenomics RAST server: a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Moore, L. R., and Chisholm, S. W. (1999) Photophysiology of the marine cyanobacterium *Prochlorococcus*: ecotypic differences among cultured isolates. *Limnol Oceanogr* **44**: 628–638.

Moore, L. R., Post, A. F., Rocap, G., and Chisholm, S. W. (2002) Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol Oceanogr* **47**: 989–996.

Moore, L. R., Rocap, G., and Chisholm, S. W. (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**: 464–467.

Mac Nally, R. (2002) Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables. *Biodivers Conserv* **11**: 1397–1401.

Nannipieri, P., Ascher, J., Ceccherini, M., Landi, L., Pietramellara, G., and Renella, G. (2003) Microbial diversity and soil functions. *Eur J Soil Sci* **54**: 655–670.

Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., *et al*. (2013) Patterns and processes of microbial community assembly. *Microbiol Mol Biol Rev* **77**: 342–356.

O'Toole, G. A. (2011) Microtiter dish biofilm formation assay. *J Vis Exp* **47**: 2437.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., *et al*. (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**: 3691–3693.

Partensky, F., Hess, W. R., and Vaulot, D. (1999) Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63**: 106–127.

Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D. (2011) *R Development Core Team. 2010. nlme: linear and nonlinear mixed effects models. R package version 3.1-97*. Vienna, Austria: R Foundation for Satatistical Computing.

Polz, M. F., Hunt, D. E., Preheim, S. P., and Weinreich, D. M. (2006) Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Philos Trans R Soc B Biol Sci* **361**: 2009–2021.

Potts, M. (1994) Desiccation tolerance of prokaryotes. *Microbiol Rev* **58**: 755–805.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) FastTree 2: approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Pruesse, E., Peplies, J., and Glöckner, F. O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.

Raes, J., Letunic, I., Yamada, T., Jensen, L. J., and Bork, P. (2011) Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Mol Syst Biol* **7**: 473.

Ranjard, L., and Richaume, A. (2001) Quantitative and qualitative microscale distribution of bacteria in soil. *Res Microbiol* **152**: 707–716.

Richter, M., and Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* **106**: 19126–19131.

Roberson, E. B., and Firestone, M. K. (1992) Relationship between desiccation and exopolysaccharide production in a soil pseudomonas sp. *Appl Environ Microbiol* **58**: 1284–1291.

Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., *et al*. (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ* **4**: e1900v1.

Schloss, P. D. (2010) The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput Biol* **6**: e1000844.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., and Hollister, E. B. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., *et al*. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol Syst Biol* **7**: 539.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Thompson, J. R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D. E., Benoit, J., *et al*. (2005) Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**: 1311–1313.

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007) Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., *et al*. (2014) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* **42**: 581–591.

Wu, D., Jospin, G., and Eisen, J. A. (2013) Systematic identification of gene families for use as "markers" for phylogenetic and phylogeny-driven ecological studies of bacteria and archaea and their major subgroups. *PLoS One* **8**: e77033.

Yao, H., Gao, Y., Nicol, G. W., Campbell, C. D., Prosser, J. I., Zhang, L., *et al*. (2011) Links between ammonia oxidizer community structure, abundance, and nitrification potential in acidic soils. *Appl Environ Microbiol* **77**: 4618–4625.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Cladogram comparison of core genome (N = 1075 orthologous proteins) and 16S rRNA phylogenetic analyses. Terminal branches are coloured by subclade designation with lines connecting identical strains in each respective cladogram. Consensus 16S rRNA cladogram is further coloured by operational taxonomic unit (OTU) designations clustered at 99%.

Fig. S2. Plot of maximum growth rate (μ_{max}) versus carbon depolymerization at each assayed temperature. Points and lines are coloured by the temperature of each assay with linear regression lines depicting the relationship between the two physiological measurements. Symbols represent the site of isolation along the climate gradient.

Fig. S3. Trait measurements versus temperature. Lines and bar plots are coloured by the clade or subclade designation. **(A)** Growth curves across time (hrs) averaged by clade or subclade designation. Each time point plots a point for the mean ($\pm$ 1 SD) growth for all strains assayed within a clade or subclade. Growth curves are visualized using smoothed averages from locally weighted smoothing (LOESS) of the mean growth values for each clade or subclade. **(B)** Biofilm formation of each assayed strain, organized by site of isolation. Error bars indicate $\pm$1 SD from the triplicate assays for each strain. Note the data are the same as in Fig. 2 with additional strain details.

Fig. S4. Bacterial community composition along the climate gradient. **(A)** Multidimensional scaling (MDS) plot depicting the difference in microbial communities. Symbols represent sampling date and colours indicate the site. Relative abundances of the bacterial community at the **(B)** phylum level and **(C)** genus level (note abbreviated y-axis). Each bar represents one sample from each plot at each time point along the climate gradient. Taxonomic designations include the 7 and 14 most abundant phyla and genera, respectively.

Fig. S5. Detailed biotic and abiotic characteristics of five of the sites along the climate gradient over time. **(A)** Relative abundance of the six most abundant subclades with respect to the total abundance of *Curtobacterium* over time. Each point represents a sampled plot and lines were calculated from smoothed averages from locally weighted smoothing (LOESS) with confidence intervals (coloured areas). **(B)** Surface soil temperature coloured by day- or night-time measurement. **(C)** Total precipitation during the duration of the field experiment. **(D)** Mean leaf litter chemistry of non-ash dry weight from the dry season and wet season in 2015 (data from (Baker and Allison, 2017)).

Fig. S6. Genomic analyses for glycoside hydrolases (GH) and carbohydrate binding module (CBM) proteins. **A)** Average number of GH/CBM protein families targeting specific polymeric carbohydrates across strains by each designated subclade. **B)** Multidimensional scaling (MDS)

plot depicting all identified GH and CHM genes across *Curtobacterium* genomes at the subclade level. Insert denotes cladogram of *Curtobacterium.*

Table S1. General climate characteristics of field sites. Historical values represent five year average measurements, while site-level measurements occurred during the duration of the experiment. MAP = mean annual precipitation, MAT = mean annual temperature.

Table S2. General genomic characteristics of *Curtobacterium* isolates.

Table S3. ANCOVA results for factors affecting functional traits (biofilm formation, carbon degradation, and maximum growth rate). The measure of effect (Ω^2) for each factor (Subclade, Site, and Temperature (T)) is also designated.

Table S4. Results from a distance based linear model (distLM). Response variable is the relative composition of ecotypes. Sequential test denotes the best overall model, with each independent predictor variable depicted in marginal test below.