Latitude delineates patterns of biogeography in terrestrial *Streptomyces*

Mallory J Choudoir, James R Doroghazi, and Daniel H Buckley*

School of Integrative Plant Sciences, Cornell University, Ithaca, NY 14853 USA

* Corresponding Author: Daniel H Buckley
dbuckley@cornell.edu

Running Title: Biogeography of *Streptomyces*

Subject Category: Microbial population and community ecology

Conflict of Interest: None
Abstract

The evolutionary and ecological forces that govern microbial biogeography remain poorly characterized. We examined the biogeography of *Streptomyces* at regional spatial scales to identify factors that govern extant patterns of microbial diversity within the genus. *Streptomyces* are spore forming filamentous bacteria that are widespread in soil and are a predominant source of antibiotics. We applied population genetic approaches to analyze geographic and genetic population structure in six phylogroups of *Streptomyces* identified from a geographically explicit culture collection. *Streptomyces* strains were isolated from soils associated with perennial grass habitats sampled across a spatial scale of more than 5,000 km. We find that *Streptomyces* allelic diversity correlates with geographic distance and that gene flow between sites is constrained by latitude. In addition, we find that phylogroup nucleotide diversity is negatively correlated with latitude. These observations are consistent with the hypothesis that historical demographic processes have influenced the contemporary biogeography of *Streptomyces*.

Introduction

Widespread evidence documents microbial biogeography across varying spatial (Whitaker *et al.*, 2003; Vos and Velicer, 2008; Bissett *et al.*, 2010; Martiny *et al.*, 2011) and temporal scales (Gilbert *et al.*, 2012; Hatosy *et al.*, 2013), yet we are only beginning to understand the evolutionary forces that generate and maintain these patterns. Explorations of biogeography are valuable because biogeographical patterns illustrate fundamental principles of evolution and ecology. One of the earliest and most well documented patterns of biogeography is the latitudinal diversity gradient (Wallace, 1878), and this trend of decreasing species richness with increasing latitude is widely described in plant and animal systems (Hillebrand, 2004). We examine the
biogeography of *Streptomyces* to determine whether a latitudinal diversity gradient can be observed in terrestrial bacteria and to evaluate the underlying mechanisms that govern biogeography within this genus.

Patterns of microbial biogeography are determined by rates of dispersal and diversification (Martiny *et al*., 2006; Hanson *et al*., 2012). Rates of dispersal can be difficult to quantify for microbes because dispersal that does not contribute to the genetic diversity of future populations remains undetected. Gene flow results from the transfer of genetic material between populations, and gene flow between geographically distinct populations requires dispersal. Hence, patterns of microbial dispersal are typically inferred from extant patterns of genetic diversity. It has been hypothesized that microbes disperse ubiquitously due to their small cell size and massive population numbers (Finaly, 2002; Finlay and Fenchel, 2004). For example, Gibbons *et al*. (2013) determined that a single sampling site in the English Channel could encapsulate the total expected diversity of the Earth’s marine biosphere. Yet endemism and dispersal limitation have been observed for a range of microbes (Cho and Tiedge, 2000; Green and Bohannan, 2006; Telford *et al*., 2006; Boucher *et al*., 2011), and microbial dispersal limitation has been verified experimentally (Bell, 2010). Contrasting evidence for dispersal limitation in the literature can be explained by at least two factors: firstly, dispersal constraints are likely to vary between species and habitats; and secondly, the units used to define microbial diversity can vary dramatically in their phylogenetic resolution. Each of these factors has been discussed previously (Hanson *et al*., 2012; Choudoir *et al*., 2012), and we will only consider them briefly here.
Patterns of microbial dispersal and gene flow appear to differ considerably between habitats and species. At one end of the spectrum, globally widespread marine microbes such as Prochlorococcus and Pelagibacter show little variation in gene content between the Atlantic and Pacific Oceans suggesting that high dispersal can homogenize genetic diversity in pelagic marine systems, while adaptation to regional habitats occurs within limited regions of the genome (Coleman and Chisholm, 2010). At the other end of the spectrum are extremophiles such as Sulfolobus and thermophilic Synechococcus, which live in island-like volcanic habitats and exhibit strong patterns of allopatric divergence resulting from dispersal limitation (Papke et al., 2003; Whitaker et al., 2003). Both allopatric processes and local patterns of gene flow contribute to the biogeography and genetic diversity of Sulfolobus populations (Cadillo-Quiroz et al., 2012). Terrestrial microbes fall somewhere between these extremes. Soil dwelling microbes such as Burkholderia pseudomallei, Burkholderia mallei, and Bacillus anthracis exhibit biogeographical patterns governed by dispersal limitation at regional spatial scales (Kenefic et al., 2009; Pearson et al., 2009). For instance, the contemporary population structure of North American anthrax reflects historical migration across the Bering land bridge during the late Pleistocene followed by range expansion into the Southwestern United States and regional diversification due to drift (Kenefic et al., 2009).

The phylogenetic resolution at which microbial diversity is defined can have a profound impact on our ability to discern patterns of microbial biogeography (as reviewed Hanson et al., 2012). Observations of microbial biogeography are often based upon variation in SSU rRNA gene sequences. Surveys of SSU rRNA genes in terrestrial habitats indicate that environmental variables including temperature (Fierer et al., 2009; Miller et al., 2009), pH (Fierer and Jackson,
2006; Lauber et al., 2009; Rousk et al., 2010), and salinity (Lozupone and Knight, 2007) are more important than geographic distance in determining patterns of microbial diversity and biogeography. However, SSU rRNA gene sequences have an extremely low rate of nucleotide substitution, and the formation of discrete taxonomic units as defined by this locus (i.e. Operational Taxonomic Units defined at 97% nucleotide identity) requires millions of years (Ochman and Wilson, 1987; Ochman et al., 1999). Hence, microbes with similar or even identical SSU rRNA genes can have extensive genomic and ecological diversity (Welch et al., 2002; Jaspers and Overmann; 2004) which far exceeds the diversity found within plant and animal species. Neutral processes that drive patterns of biogeography, such as dispersal limitation and genetic drift, can only be detected using genetic techniques that have high resolution (Green and Bohannan, 2006; Hanson et al., 2012). Fortunately, the evolutionary and ecological forces that drive microbial biogeography are readily explored using geographically and ecologically explicit culture collections that can be characterized at high genetic resolution (Hunt et al., 2008; Shapiro et al., 2012).

Streptomyces are spore forming bacteria ubiquitous in soils, and many species are easily cultured. Furthermore, Streptomyces species have high rates of gene exchange both within and between species (Doroghazi and Buckley, 2010). Hence this genus represents a useful model for examining limits on dispersal and gene flow in terrestrial bacterial communities. Streptomyces are gram-positive Actinobacteria (Kâmpfer, 2006) known for their complex developmental cycle, which entails mycelial growth followed by formation of aerial hyphae and desiccation resistant spores which are readily dispersed (Keiser et al., 2000). Streptomyces are able to degrade recalcitrant carbon molecules like cellulose and play an important role within the
terrestrial carbon cycle (McCarthy and Williams, 1992; Takasuka et al., 2013). In addition, phytopathogenic *Streptomyces*, such as *S. scabies*, cause substantial economic losses in agriculture (Loria et al., 2006; Labeda, 2011). Finally, *Streptomyces* are prolific producers of secondary metabolites, and most antibiotics we use today were discovered in this genus (Watve et al., 2001). Despite their clinical, agricultural, and environmental importance, we still lack an evolutionary and ecological framework to understand *Streptomyces* biodiversity and biogeography.

*Streptomyces* diversity varies spatially, though the effects of geographic distance and ecological variation on *Streptomyces* diversity remains poorly resolved. *Streptomyces* genetic and phenotypic diversity can vary at small spatial scale (1 m²) in prairie soils (Davelos et al., 2004a and 2004b) and across dune and beach habitats (60 m) (Antony-Babu et al., 2008), suggesting that variation in local soil characteristics can select for species and populations with different ecological adaptations. There is also evidence that some *Streptomyces* species are endemic to North America or Central Asia, which suggests dispersal limitation at large scales (Wawrik et al., 2007). Analysis of genetic diversity within a single population of *Streptomyces pratensis*, isolated from soils that spanned 1,000 km, revealed a population in linkage equilibrium when assessed both by multi-locus sequence analysis (MLSA) and genome analysis (Doroghazi and Buckley, 2010 and 2014). This result indicates that either gene flow is unlimited in *S. pratensis* across large spatial scales, or that it has experienced an evolutionarily recent demographic expansion from a population that was once in linkage equilibrium. Furthermore, analysis of ancestral patterns of *Streptomyces* inter-species gene exchange suggests that the fixation of incongruent loci could be due to demographic range expansion into higher latitudes (Andam et
al., 2015). These latter data suggest a role for both adaptive and neutral processes in governing
the diversity and biogeography of *Streptomyces*, though the degree to which these forces
constrain extant patterns of genetic diversity within the genus remain poorly described.

We evaluated *Streptomyces* biogeography by constructing an isolate collection from grassland
sites spanning the United States of America. The most powerful approach for detecting neutral
evolutionary processes associated with dispersal limitation is to examine patterns of
biogeography across similar habitats (Hanson *et al.*, 2012). In a previous study which employed
a single locus (*rpoB*), we observed that *Streptomyces* species diversity exhibited a significant
distance decay relationship and was inversely correlated with latitude (Andam *et al.*, 2016).
Evidence for a latitudinal diversity gradient coupled with evidence of demographic range
expansion, while not conclusive, is consistent with predictions of phylogenetic niche
conservatism (Wiens and Donoghue, 2004; Stevens, 2006) which posits that these
biogeographical patterns result from the historical effects of changes in paleoclimate (Hewitt *et
al.*, 1996, 2000, and 2004). The hypothesis of a latitudinal diversity gradient in *Streptomyces*
predicts that, in addition to a gradient of species richness, there should also be latitudinal
differences in gene flow and genetic diversity within populations. Here we investigate genetic
diversity and geographic distribution within six species-like phylogenetic clusters, or
phylogroups, to determine whether patterns of gene flow and diversity corroborate the hypothesis
of a latitudinal diversity gradient for *Streptomyces*.

### Materials and Methods

*Strain isolation and DNA extraction*
We assembled a culture collection of more than 1,000 Streptomyces from 15 sites across the United States with soil sampled at 0-5 cm depth (Andam et al., 2016). Sites were selected to represent a narrow range of habitats dominated by perennial grasses and having neutral to slightly acidic pH (3.9-7.3). Soil was air dried, and Streptomyces strains were isolated on glycerol-arginine agar plates of pH 8.7 containing cycloheximide and Rose Bengal (El-Nakeeb and Lechevalier, 1963; Ottow, 1972) as previously described (Doroghazi and Buckley, 2010). DNA was extracted from purified cultures, which were grown by shaking at 30°C in liquid yeast extract-malt extract medium (YEME) containing 0.5% glycine (Kieser et al., 2000), by using a standard phenol/chloroform/isoamyl alcohol protocol.

The genetic diversity of isolates was initially assessed using partial rpoB sequences. Streptomyces species are typically delineated by 0.007 nucleotide dissimilarity across MLSA loci (Rong and Huang, 2012). Since we were using only a single locus we elected to define species-like phylogenetic clusters, or phylogroups, at 0.01 patristic distances with RAMI (Pommier et al., 2009). Using this approach, we identified 107 phylogenetic clusters based on rpoB sequences. We selected six of these phylogroups for further analysis. MLSA was performed on 17-47 isolates from each phylogroup. Phylogroup names are capitalized and based on a representative isolate; for example, phylogroup WA1063 is named for Streptomyces sp. wa1063. Isolates are identified with a lowercase letter code indicating site of origin followed by a strain number; for example, isolate wa1063 is strain 1063 isolated from Washington (WA) state.

Multilocus sequence analysis (MLSA)

We adapted the MLSA scheme developed for Streptomyces by Guo et al. (2008), which targets
the five housekeeping loci \textit{atpD}, \textit{gyrB}, \textit{recA}, \textit{rpoB}, and \textit{trpB} as described in Doroghazi and Buckley (2010). Partial SSU rRNA sequences were amplified using universal primers (8F: AGAGTTTGATCCTGGCTCAG; 1492R: CGGTTACCTTGTTACGACTT). Reactions for Sanger sequencing were performed using forward primers for all loci except \textit{rpoB}, for which the reverse primer was used. Trace files were uniformly screened using CAP3 (Huang and Madan, 1999) interfaced with a Perl script. Sequences were aligned using MUSCLE (Edgar, 2004), trimmed to 431 bp, 415 bp, 446 bp, 557 bp, and 489 bp, for genes \textit{atpD}, \textit{gyrB}, \textit{recA}, \textit{rpoB}, \textit{trpB}, respectively, and concatenated consistent with the genomic order in \textit{Streptomyces coelicolor} A3(2). SSU rRNA alignments were trimmed to 357 bp. Gene sequences are available on GenBank with accession numbers XXXX-XXXX.

Good’s coverage estimation and haplotype rarefaction was determined using mothur (Schloss et al., 2009). DNA polymorphism statistics including: number of segregating sites, nucleotide diversity, and Tajima’s D, were determined with DnaSP v5 (Librado and Rozas, 2009) and LDhat (McVean et al., 2002). Population scaled mutation rates (Watterson's theta; \( \theta_w = 2N_e \mu \)), recombination or gene conversion rates (\( \rho = 2N_e r / 2 \)), and relative rates of recombination (\( \rho / \theta_w \)) were estimated using LDhat (McVean et al., 2002) and are expressed per nucleotide site. The standard index of association (\( I_A \)) was calculated from allelic data with LIAN v3.5 (Haubold and Hudson, 2000) using the Monte-Carlo test and 100 iterations. The pairwise homoplasy index (PHI) statistic was determined using PhiPack (Bruen et al., 2006), and statistical significance was evaluated under a null hypothesis of no recombination. Sequence identity across phylogroups was calculated with mothur (Shloss et al., 2009)
Phylogeny reconstruction

Maximum likelihood (ML) trees were constructed from the nucleotide sequences of individual and concatenated MLSA loci using the generalized time reversible nucleotide substitution model (Tavaré, 1986) with gamma distributed rate heterogeneity among sites (GTRGAMMA) supported in RAxML v7.3.0 (Stamatakis, 2006). Bootstrap support was determined for the highest-scoring ML tree of 20 iterations, and the number of bootstrap replicates was determined using the extended majority rule (autoMRE) convergence criteria (Pattengale et al., 2010). Root placement is defined by Mycobacterium smegmatis. Significant phylogenetic incongruence between loci was determined using the Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999) implemented in the R package phangorn (Schliep, 2011).

Population structure

Concatenated MLSA sequences were analyzed using Structure v2.3.3 (Pritchard et al., 2000) to examine population affiliation, subdivision, and admixture within and between phylogroups. Structure was run using an admixture model with a burn-in length of $1.0 \times 10^6$ and $3.0 \times 10^6$ replicates. The most probable number of sub-populations (k) was chosen using the Evanno method (Evanno et al., 2005) from 10 independent runs with $k = 1$ through $k = 6$ within phylogroups, and through $k = 8$ between phylogroups implemented by Structure Harvester (Earl and vonHoldt, 2012). After choosing the most probable k-value, the program Clumpp was used to permute outputs of the independent runs (Jakobsson and Rosenberg, 2007).

Patterns of dispersal and gene flow

We used permutation tests to evaluate whether phylogroup distribution across sites could be
explained by panmixia. We compared the observed distributions to those expected under a
random distribution model using 1,000 permutations to assess significance. In addition,
correlations between geographic distance and allele composition between sites were assessed
using Mantel and partial Mantel tests (Mantel, 1967; Smouse et al., 1986). These tests were
performed with the R package ecodist (Goslee and Urban, 2007) using the Pearson correlation
method and 1,000 permutations. Bray-Curtis dissimilarity was calculated from unique allele
composition across sites.

Haplotype networks were created using a statistical parsimony procedure (Templeton et al.,
1987; Templeton et al., 1995) implemented in TCS v1.18 (Clement et al., 2000). Nested clade
information was used to infer processes that could explain the geographic and genetic
distribution of sequences using the program GeoDis v2.2 (Posada et al., 2000). Both TCS v1.18
and GeoDis v2.2 were performed in ANeCA (Panchal, 2007).

Results

Characterization of Streptomyces phylogroups

We identified a total of 308 Streptomyces isolates representing the six targeted phylogroups, and
these isolates spanned 13 sites (Figure 1, Table 1). The six phylogroups all share more than 97%
nucleotide identity in their SSU rRNA genes and greater than 88% average nucleotide identity
(ANI) across MLSA loci. Strains within each phylogroup share 99.4%-100% nucleotide identity
in their SSU rRNA genes (Figure 2) and 97.6%-99.5% ANI across concatenated MLSA loci
(Figure 1, Table 2). The genetic ancestry of phylogroup members as determined by Structure is
congruent with phylogroup boundaries defined by nucleotide cutoffs (Figure S1), and this
indicates that phylogroups approximate biological populations. There are a total of 122 unique MLSA haplotypes, with each phylogroup representing 15-26 haplotypes (Figure 1, Table 2). Good’s coverage for each of the six phylogroups ranges from 0.88-1.0 for individual loci and 0.94-1.0 for concatenated MLSA loci clustered at 99% nucleotide identity (Table S1). While unique haplotypes remain somewhat under sampled (Figure S2), allelic diversity at single loci is sampled well (Table S1). The per site nucleotide diversity (\(\pi\)) of each phylogroup ranges from 0.0026 to 0.011 (Table 2).

Four of our phylogroups do not match any *Streptomyces* species described in PubMLST (less than 97% ANI across MLSA loci, Figure S3A) (Jolley et al., 2004). However, phylogroups WA1063 and MS152 affiliate with strains that belong to the *S. griseus* species cluster (Rong and Huang, 2010). Isolates in phylogroup WA1063 share greater than 99% MLSA identity with *S. anulatus* and *S. praecox* and form a monophyletic clade with these type strains (Figure S3A). Isolates in phylogroup MS152 share greater than 98% MLSA identity with *S. mediolani*, *S. albovinaceus*, and *S. griseinus* and form a paraphyletic clade that includes these type strains (Figure S3A).

We find evidence of horizontal gene transfer consistent with previous observations of *Streptomyces* species (Doroghazi and Buckley, 2010; Andam et al., 2015). There is significant phylogenetic incongruence between MLSA loci (Figure S3, Table S2), suggesting that horizontal gene transfer has shaped the phylogeny of these groups. The six phylogroups present evidence of population structure and admixture, with each phylogroup composed of 3.2 ± 0.8 (mean ± s.d.) subpopulations (Figure 1). Evidence of admixture between populations suggests horizontal gene
transfer within phylogroups as has been previously described in *Streptomyces* (Doroghazi and Buckley, 2010). Likewise, there is evidence of significant levels of recombination within phylogroups MAN196, MS200, MS152, and F34 (PHI test, $p < 0.05$, Table 2). Phylogroups WA1063 and MS152 share two identical *atpD* alleles, with one of these alleles shared among strains isolated from MS, CA, WI (Brookfield), AK (Manley Hot Springs), and WA, and the other allele shared among strains isolated from MS and WI (Brookfield) (Figure 3, Figure S3B). It is not clear whether these alleles are shared as a result of contemporary horizontal gene transfer or vertical inheritance from the most recent common ancestor of the two clades, though the latter explanation is more parsimonious given the low level of polymorphism between the phylogroups.

*Dispersal limitation*

Strains of the six phylogroups were obtained from soil samples from 13 sites of diverse geographic origin (Table 1). Each phylogroup was detected in $4.2 \pm 0.4$ sites, and this distribution differs significantly from expectations for a random distribution of strains across sites (permutation test, $p < 0.0005$), thereby rejecting the hypothesis of panmixia. Each phylogroup subpopulation was observed in $2.2 \pm 0.9$ sites (mean $\pm$ s.d.), and this value is lower than expected if subpopulations are randomly distributed across the sites occupied by each phylogroup (permutation test, $p < 0.001$). These results indicate that phylogroup distribution is constrained geographically and that phylogroups have subpopulation structure that is also geographically explicit.

The geographic distribution of *Streptomyces* allelic diversity indicates dispersal limitation.
Identical alleles are shared among phylogroup members across each phylogroup’s geographic range, which can exceed 5,000 km (Figure 3, Figure S3). However, dissimilarity in allele composition increases with geographic distance, and this result is significant (Bray-Curtis dissimilarity, Mantel r = 0.29, p = 0.005) (Figure S4). Hence, alleles are more likely shared between geographically similar sites indicating dispersal limitation with potential for long range dispersal. This result is significant for all individual loci except recA (Bray-Curtis dissimilarity, aptD Mantel r = 0.31, p = 0.004; gyrB r = 0.22, p = 0.031; recA r = 0.16, p = 0.088; rpoB r = 0.27, p = 0.004; trpB r = 0.19, p = 0.047). Nearly all MLSA haplotypes (Figure 1, Figure 4) are unique to a single site, with the sole exception being a haplotype from phylogroup MS200 which is shared between strains ms53, ms30, and sun103 isolated from MS and WI (Sun Prairie). It is worth noting that all of the strains examined share > 97% SSU rRNA gene identity and these geographic patterns would not be detected through analysis of this locus. That is, everything would appear to be everywhere based on analysis of SSU rRNA genes.

Analysis of haplotype distribution is consistent with diversification due to spatial isolation resulting from dispersal limitation. Nested clade analysis (NCA) establishes significant phylogeographic inferences for phylogroups MAN196, MAN125, WA1063, and MS152 (Figure 4) but not for MS200 and F34. Nested clade phylogeographic inference postulates potential evolutionary and historical demographic processes that support extant patterns of diversity and biogeography. For instance, population subdivision of MAN125 across the Pacific Northwest (Figure 4A) and MAN196 between Maine and the Pacific Northwest (Figure 4B) is consistent with restricted gene flow due to historical long distance dispersal events. Likewise, population subdivision of MS152 and between the Southeast (MS and FL) and CA (Figure 4C) and WA1063
between WI (Brookfield) and OR (Figure 4D) is consistent with allopatric fragmentation.

**Latitudinal diversity gradient**

The distribution and diversity of the phylogroups reveals a latitudinal diversity gradient. Strains from MAN125, MAN196, WA1063 occur mostly north of 40°N latitude, while strains from MS200, MS152, F34 occur mostly south of this latitude (Table 1). This pattern of north/south partitioning is significant for each phylogroup when evaluated against the expectation of a random distribution across sites (permutation test, \( p < 0.01 \) for each phylogroup after Bonferroni correction). Furthermore, partial Mantel tests were performed to evaluate the latitudinal and longitudinal vector components of geographic distance in relation to the allele composition of sites. There remains a significant relationship between allele composition and geographic distance when we control for longitude, (Mantel \( r = 0.23, \ p = 0.022 \)), but this relationship is no longer significant when we control for latitude (Mantel \( r = 0.15, \ p = 0.12 \)). This result indicates that allele composition changes more across latitude than it does across longitude. The latitudinal partitioning of alleles can be readily observed in the pattern of allele sharing between sites (Figure 3). Finally, we also observed a significant relationship between per site nucleotide diversity of phylogroup MLSA loci and the average latitude of sites in which they are found (\( R = -0.91, \ p = 0.012 \); Figure 5). This result indicates that phylogroups recovered from lower latitudes have higher genetic diversity than those recovered from higher latitudes.

**Discussion**

We used population genetic approaches to analyze spatial patterns of genetic diversity for six *Streptomyces* phylogroups isolated from geographically disparate but ecologically similar sites.
The distribution of phylogroups is nonrandom and dispersal limited (Figure S4) with phylogroups inhabiting geographic ranges defined by latitude (Table 1, Figure 3). In addition, the genetic diversity of phylogroups is inversely proportional to the latitude from which they were isolated (Figure 5). These findings suggest that there are latitudinal barriers to dispersal, and that patterns of Streptomyces biogeography result from dispersal limitation and regional diversification due to genetic drift.

The phylogroups we describe are coherent phylogenetic groups that approximate biological populations (Figure S1). The members of each phylogroup share a distinguishable geographic range (Table 1, Figure 3), a recent common ancestor (Figure 1), greater than 97% ANI across MLSA loci (Table 2), and greater than 99.4% SSU rRNA gene nucleotide identity. Despite geographic and genetic subpopulation structure (Figure 1), phylogroup members frequently share identical alleles across demes. We infer the presence of recombination within all phylogroups using both nucleotide polymorphism and phylogenetic methods (Table 2, Table S2). For instance, MAN125 is nearly in linkage equilibrium (I_A = 0.09) though its members span a geographic range of over 2,000 km across the Pacific Northwest. The observation of linkage equilibrium indicates that either the population lacks contemporary barriers to gene flow or the population has experienced a recent demographic expansion in which there has been insufficient time to accumulate mutations and linkage disequilibrium. Significant recombination and linkage equilibrium have also been observed in a S. pratensis population which spanned 1,000 km across sites present in North Carolina and northern New York (Doroghazi and Buckley, 2010 and 2014).
Regional patterns of biogeography among phylogroups are consistent with limitations to dispersal and gene flow. Allopatric processes like genetic drift can drive diversification between populations that are geographically isolated. The geographic distribution of our phylogroups is nonrandom, and we find regional subpopulation structure within phylogroups (Figure 1, Figure 4). Although we find identical alleles in sites thousands of kilometers apart (Figure 3), MLSA haplotypes are not shared across sites (with the single exception of a haplotype shared between MS and WI) (Figure 4). We also observe a significant distance decay relationship for MLSA allele composition and geographic distance between sites (Figure S4). This data implies that while gene flow is moderate across the geographic range of a phylogroup, dispersal limitation and genetic drift have combined to produce regional populations whose distributions are constrained by latitude.

Streptomyces phylogroup diversity is consistent with a latitudinal diversity gradient. We find that latitude is a significant predictor of gene flow (Figure 3). Furthermore, intra-phylogroup nucleotide diversity has a significant negative relationship with average latitude (Figure 5), which is congruous with the latitudinal diversity gradient observed for diverse macroorganisms (Hillebrand, 2004). There is conflicting evidence for latitudinal diversity gradients among microorganisms. Evidence for microbial latitudinal diversity gradients comes from marine systems (Fuhrman et al., 2008; Sul et al., 2013; Swan et al., 2013), with contrary evidence obtained in terrestrial systems (Neufeld and Mohn, 2005; Chu et al., 2010). However, most analyses of terrestrial bacterial biogeography are derived from analyses of SSU rRNA genes, and we show that analyses of SSU rRNA genes lack the sensitivity needed to detect the biogeographic patterns that we observe for Streptomyces.
Several hypotheses have been advanced to explain the formation of latitudinal diversity gradients (Wiens and Donoghue, 2004; Mittelbach et al., 2007). Ecological hypotheses posit that factors such as carrying capacity, productivity, and niche availability vary across latitude and that these factors impose constraints on biodiversity (Currie et al., 2004; Mouchet et al., 2015). Evolutionary hypotheses invoke the positive relationship between temperature and the kinetics of metabolism to predict that evolutionary rates and cladogenesis correspond with temperature (Allen et al., 2002). Historical hypotheses propose that the latitudinal diversity gradient is the product of historical geological, ecological, or demographic events that have influenced dispersal and diversification (Wiens and Donoghue, 2004; Stevens, 2006). For example, the influence of Pleistocene glacial events on the biogeography of diverse species of terrestrial and aquatic plants and animals is well documented (Soltis et al., 1997; Bernatch and Wilson, 1998; Conroy and Cook, 2000; Milá et al., 2006; Maggs et al., 2008; Wilson and Veraguth, 2010). There is also evidence that Pleistocene glaciation events have impacted both microbial communities (Eisenlord et al., 2012) and populations (Kenefic et al., 2009; Mikheyev et al., 2008).

The biogeography of our *Streptomyces* phylogroups is consistent with historical demography and dispersal limitation resulting in a contemporary latitudinal diversity gradient. Eisenlord et al. (2012) also found that dispersal limitations shape biogeography of Actinobacteria across a glacial retreat transect in Michigan. However, Fuhrman et al. (2008) gives strong support for the hypothesis that the increased kinetics of metabolism at higher temperatures result in greater bacterial species richness in marine communities at tropical latitudes. Since latitude is correlated with temperature, we cannot dismiss the influence of kinetics on driving the diversity gradient in
terrestrial *Streptomyces*. Furthermore, it is possible that unappreciated ecological variables, such as the species composition of perennial grass communities, could shape the diversity gradient. However, phylogroup MAN125 from the Pacific Northwest (Table 2) and *S. pratensis* from the Eastern United States (Doroghazi and Buckley, 2010 and 2014) each have populations in linkage equilibrium though occupying a large geographic range over which gene flow is limited (Figure 3, Figure S4). The most parsimonious explanation for these conflicting results is that both taxa have independently experienced a recent demographic range expansion and that insufficient polymorphisms have accumulated since this expansion to allow for detection of disequilibrium due to isolation by distance. However, it is likely that latitudinal gradients are population specific and can arise from a combination of ecological, evolutionary, and historical processes.

The genetic consequences of glaciation events are described in depth by Hewitt (1996, 2000, and 2004), and the population structure we observe in our *Streptomyces* phylogroups is consistent with post-glacial range expansion. One of the expectations of post-glacial expansion is “southern richness versus northern purity” (Hewitt, 2004). This is evident in the negative correlation we observe between latitude and the nucleotide diversity of phylogroups (Figure 5). Williams *et al.* (1998) justifies 40˚N latitude as approximating late Pleistocene glacial and non-glacial regions with respect to species distributions in North America. Hence, the latitudinal delineation of allele distributions for *Streptomyces* phylogroups roughly corresponds to the extent of ice coverage during the late Pleistocene (Figure 3), which suggests historical population expansion from lower to higher latitudes.

Haplotype distributions of phylogroups MAN125, MAN196, WA1063, and MS152 are
consistent with allopatric diversification resulting from dispersal limitation (Figure 4). Haplotype nested clade analysis (NCA) predicts that historical dispersal events and range expansion across northern regions resulted in limits on dispersal during intermediate timescales allowing genetic drift to create the phylogroup population structures observed today. There is moderate criticism of NCA (Knowles and Maddison, 2002; Nielsen and Beumont, 2009) due to the subjective nature of inferring historical processes and the wide potential for stochastic processes creating similar patterns of biogeography. Yet these tools can provide useful hypotheses. Northern phylogroups MAN125 and MAN196 share a common ancestor with southern phylogroups MS200 and F34 (Figure 1). The contemporary population structure of MAN125 and MAN196 is consistent with a historical range expansion from a common ancestor shared by both clades (Figure 1, Figure 4). Further population structure within each phylogroup likely resulted from barriers to gene flow and historical dispersal events across the Pacific Northwest for MAN125 (Figure 4A) and between Maine and the Pacific Northwest for MAN196 (Figure 4B). Analysis of haplotype distribution within WA1063 and MS152 is also consistent with diversification of populations as a result of gene flow limitation between the Midwest and Pacific Northwest for WA1063 (Figure 4D) and between the Southeast and West for MS152 (Figure 4C).

Phylogroups WA1063 and MS152 share a recent common ancestor (Figure 1, Figure S1) and even share two identical alleles at the atpD locus (Figure 3). These phylogroups have distinct, non-overlapping geographic ranges with WA1063 found in higher latitudes and MS152 in lower latitudes (Figure 3). WA1063 and MS152 have 0.0205 net nucleotide substitutions per site across MLSA loci. We evaluate the possible time range for divergence between WA1063 and MS152 by extrapolating very roughly from the nucleotide substitution rate ($\mu = 4.5 \times 10^{-9}$) and generation...
time (100-300 generations per year) for *E. coli* (Ochman *et al*., 1999), since corresponding values are not available for *Streptomyces* or their relatives. Based upon these gross approximations, we would estimate that WA1063 and MS152 diverged 15,000-50,000 years ago, corresponding to events in the late Pleistocene (Clayton *et al*., 2006). Hence, it is likely that the identical *atpD* alleles found in both WA1063 and MS152 were inherited without modification from an ancestral population that both groups share (Figure S3B). If this is the case it suggests that the divergence of phylogroup WA1063 was driven by range expansion into higher latitudes and subsequent isolation by distance from the ancestral population it shares with phylogroup MS152.

Through population genetic analysis of six *Streptomyces* phylogroups, we find evidence for dispersal limitation associated with geographically explicit patterns of gene flow which manifest in a latitudinal gradient of nucleotide diversity. Furthermore, these data support the hypothesis that historical demographic processes influence the contemporary biogeography of *Streptomyces*. Due to their spore forming capabilities and potential for long range dispersal, *Streptomyces* are an ideal system for assessing limits on gene flow among terrestrial bacteria. We cannot yet determine whether the cause of dispersal limitation is due to limits on spore mobility or due to adaptive constraints as proposed by the hypothesis of phylogenetic niche conservatism. A better understanding of *Streptomyces* biogeography and the evolutionary processes driving population structure may ultimately assist in the discovery of novel genetic diversity and possibly novel antibiotics within this genus.
Acknowledgements

This material is based upon work supported by the National Science Foundation under Grants No. DEB-1050475 and DEB-1456821. We would like to thank Ashley Campbell for her help in generating MLSA sequences.

References


Figures and Tables
Figure 1. *Streptomyces* phylogroups have different geographic distributions defined by latitude with evidence for subpopulation structure and admixture. The colored bars in the outer ring indicate genetic contributions from different ancestral populations as inferred by Structure analysis. The shading of the inner ring indicates the latitude from which each strain was isolated according to the scale provided. The isolation site for each strain can be determined by isolate names as indicated in Table 1. The tree was constructed from concatenated MLSA loci nucleotide sequences using maximum likelihood with a GTRGAMMA evolution model. Scale bar represents nucleotide substitutions per site. The root was defined by *Mycobacterium smegmatis*. Nodes with bootstrap confidences > 80 are indicated with gray circles, and precise bootstrap values are found in Figure S3A.

Figure 2. Dissimilarity in SSU rRNA genes is less than 3% across all six phylogroups and ranges from 0-0.6% for representatives within each phylogroup, revealing that this locus lacks sufficient sensitivity to examine the biogeographic distribution of *Streptomyces* phylogroups. The tree was constructed using maximum likelihood with a GTRGAMMA evolution model using partial SSU rRNA gene sequences from three representatives for each phylogroup. Scale
bar represents nucleotide substitutions per site. Branches with bootstrap support values greater than 75% are labeled, and the outgroup is *Mycobacterium smegmatis*. Strains are colored to reveal phylogroup membership as indicated in the color scale provided.

**Figure 3.** Identical MLSA alleles are frequently shared between sites. Circles depict sample sites and are labeled according to site code (Table 1). The relative abundance of each phylogroup at each site is indicated by color according to the legend, with raw counts provided in Table 1. Solid colored lines represent identical alleles shared by phylogroup members across sites. Dashed multicolored lines depict identical *atpD* alleles shared by strains in phylogroups WA1063 and MS152 (Figure S3B).
**Figure 4.** Haplotype networks are depicted for phylogroups MAN125 (4A), MAN196 (4B), MS152 (4C), WA1063 (4D). Circles represent unique MLSA haplotypes whose radius is proportional to the number of isolates sharing that haplotype, and colors correspond to strain...
ancestry and subpopulation affiliation as determined by Structure analysis and defined in Figure 1. Haplotypes are labeled with a letter code referring to sample site location as indicated in Table 1. Black circles represent un-sampled, inferred haplotypes with each circle designating a single nucleotide polymorphism. The length of edges between nodes is uninformative. Dashed rectangles encompass clades that have significant phylogeographic inferences from nested clade analysis, as described in methods. Roman numerals correspond to the following inferences: I. Allopatric fragmentation; II. Long distance colonization and/or past fragmentation; III. Restricted gene flow with isolation by distance; IV. Restricted gene flow but with some long-distance gene flow over intermediate ranges not occupied by the species; or past gene flow followed by extinction of intermediate populations; V. Contiguous range expansion.

Figure 5. There is a significant inverse correlation between phylogroup average latitude and nucleotide diversity. Average latitude was determined using a weighted average based on the total number of isolates per site as indicated in Table 1. Nucleotide diversity was calculated using concatenated MLSA loci and is expressed per site.
Table 1. The six targeted phylogroups were observed in 13 sites. A total of 755 *Streptomyces* strains were isolated from these sites, and 308 of these strains were found to represent the six targeted phylogroups. The numbers of isolates per site that belonged to each of our target phylogroups is indicated. Isolate names begin with a letter code referring to the isolation site.

<table>
<thead>
<tr>
<th>Site</th>
<th>Code</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Isolates /Site</th>
<th>Isolates from <em>Streptomyces</em> sp. phylogroup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MAN 125</td>
</tr>
<tr>
<td>Manley Hot Springs, AK</td>
<td>man</td>
<td>63.87˚N</td>
<td>-149.02˚W</td>
<td>95</td>
<td>51</td>
</tr>
<tr>
<td>Denali Hwy, AK</td>
<td>den</td>
<td>63.22˚N</td>
<td>-147.68˚W</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Bothell, WA</td>
<td>wa</td>
<td>47.73˚N</td>
<td>-122.24˚W</td>
<td>105</td>
<td>21</td>
</tr>
<tr>
<td>Astoria, OR</td>
<td>or</td>
<td>46.18˚N</td>
<td>-123.85˚W</td>
<td>79</td>
<td>8</td>
</tr>
<tr>
<td>Kennebunk, ME</td>
<td>me</td>
<td>43.4˚N</td>
<td>-70.54˚W</td>
<td>85</td>
<td>13</td>
</tr>
<tr>
<td>Sun Prairie, WI</td>
<td>sun</td>
<td>43.17˚N</td>
<td>-89.24˚W</td>
<td>65</td>
<td>3</td>
</tr>
<tr>
<td>Brookfield, WI</td>
<td>b</td>
<td>43.06˚N</td>
<td>-88.13˚W</td>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td>Palo Alto, CA</td>
<td>st</td>
<td>37.43˚N</td>
<td>-122.17˚W</td>
<td>34</td>
<td>2</td>
</tr>
<tr>
<td>Greensboro, NC</td>
<td>gb</td>
<td>36.09˚N</td>
<td>-79.89˚W</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Troy (Uwharrie), NC</td>
<td>uw</td>
<td>35.71˚N</td>
<td>-79.88˚W</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>Starkville, MS</td>
<td>ms</td>
<td>33.46˚N</td>
<td>-88.8˚W</td>
<td>90</td>
<td>11</td>
</tr>
<tr>
<td>Austin, TX</td>
<td>t</td>
<td>30.2˚N</td>
<td>-97.77˚W</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Fort Pierce, FL</td>
<td>f</td>
<td>27.54˚N</td>
<td>-80.35˚W</td>
<td>70</td>
<td>7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td>755</td>
<td>115</td>
</tr>
</tbody>
</table>
Table 2. MLSA was performed on a total of 17-47 strains from each phylogroup. Summary statistics for concatenated MLSA nucleotide sequences were determined as described in methods. Phylogroup average latitude (Ave Lat) was determined using the total number of isolates per phylogroup provided in Table 1. The standard index of association is zero for a population in linkage equilibrium.

<table>
<thead>
<tr>
<th>Phylogroup</th>
<th>Ave Lat</th>
<th>n</th>
<th>Haplotypes</th>
<th>Min ANI</th>
<th>S</th>
<th>π</th>
<th>Tajima’s D</th>
<th>θ_W</th>
<th>ρ</th>
<th>ρ/θ_W</th>
<th>I_A</th>
<th>Φ_W</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAN125</td>
<td>60.15°N</td>
<td>45</td>
<td>25</td>
<td>99.5</td>
<td>23</td>
<td>0.0026</td>
<td>0.53</td>
<td>0.0023</td>
<td>0.0094</td>
<td>4.18</td>
<td>0.09</td>
<td>2.95E-01</td>
</tr>
<tr>
<td>MAN196</td>
<td>52.84°N</td>
<td>47</td>
<td>26</td>
<td>98.5</td>
<td>65</td>
<td>0.0075</td>
<td>0.67</td>
<td>0.0063</td>
<td>0.0043</td>
<td>0.068</td>
<td>0.49</td>
<td>4.24E-02*</td>
</tr>
<tr>
<td>WA1063</td>
<td>47.20°N</td>
<td>19</td>
<td>13</td>
<td>98.1</td>
<td>63</td>
<td>0.0052</td>
<td>-1.35</td>
<td>0.0077</td>
<td>0</td>
<td>0</td>
<td>0.43</td>
<td>4.33E-03</td>
</tr>
<tr>
<td>MS200</td>
<td>33.58°N</td>
<td>27</td>
<td>21</td>
<td>98.4</td>
<td>93</td>
<td>0.0088</td>
<td>-0.58</td>
<td>0.0103</td>
<td>0.0017</td>
<td>0.17</td>
<td>0.39</td>
<td>1.33E-03**</td>
</tr>
<tr>
<td>MS152</td>
<td>32.50°N</td>
<td>17</td>
<td>15</td>
<td>97.6</td>
<td>96</td>
<td>0.0107</td>
<td>-0.51</td>
<td>0.012</td>
<td>0.0013</td>
<td>0.11</td>
<td>0.35</td>
<td>1.078E-01**</td>
</tr>
<tr>
<td>F34</td>
<td>28.20°N</td>
<td>36</td>
<td>22</td>
<td>97.7</td>
<td>93</td>
<td>0.0109</td>
<td>0.52</td>
<td>0.0096</td>
<td>0.0011</td>
<td>0.11</td>
<td>0.46</td>
<td>1.79E-01**</td>
</tr>
</tbody>
</table>

Ave Lat: average latitude of phylogroup members
n: number of isolates subsampled for MLSA
Min ANI: minimum percent pairwise average nucleotide identity (ANI) within phylogroup members
S: segregating sites
π: per site nucleotide diversity
θ_W: per site Watterson's theta (2N_eµ)
ρ: per site rate of recombination/gene conversion (2N_eτ/2)
ρ/θ_W: relative rate of recombination to mutation
I_A: standard index of association
Φ_W: pairwise homoplasy index (Phi) statistic with p-value < 0.05*, 0.01** under the null hypothesis of no recombination