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

92 terrestrial carbon cycle (McCarthy and Williams, 1992; Takasuka *et al.*, 2013). In addition,
93 phytopathogenic *Streptomyces,* such as *S. scabies,* cause substantial economic losses in
94 agriculture (Loria *et al.*, 2006; Labeda, 2011). Finally, *Streptomyces* are prolific producers of
95 secondary metabolites, and most antibiotics we use today were discovered in this genus (Watve
96 *et al.*, 2001). Despite their clinical, agricultural, and environmental importance, we still lack an
97 evolutionary and ecological framework to understand *Streptomyces* biodiversity and
98 biogeography.
99
100 *Streptomyces* diversity varies spatially, though the effects of geographic distance and ecological
101 variation on *Streptomyces* diversity remains poorly resolved. *Streptomyces* genetic and
102 phenotypic diversity can vary at small spatial scale (1 m^2) in prairie soils (Davelos *et al.*, 2004a
103 and 2004b) and across dune and beach habitats (60 m) (Antony-Babu *et al.*, 2008), suggesting
104 that variation in local soil characteristics can select for species and populations with different
105 ecological adaptations. There is also evidence that some *Streptomyces* species are endemic to
106 North America or Central Asia, which suggests dispersal limitation at large scales (Wawrik *et al.*,
107 2007). Analysis of genetic diversity within a single population of *Streptomyces pratensis,*
108 isolated from soils that spanned 1,000 km, revealed a population in linkage equilibrium when
109 assessed both by multi-locus sequence analysis (MLSA) and genome analysis (Doroghazi and
110 Buckley, 2010 and 2014). This result indicates that either gene flow is unlimited in *S. pratensis*
111 across large spatial scales, or that it has experienced an evolutionarily recent demographic
112 expansion from a population that was once in linkage equilibrium. Furthermore, analysis of
113 ancestral patterns of *Streptomyces* inter-species gene exchange suggests that the fixation of
114 incongruent loci could be due to demographic range expansion into higher latitudes (Andam *et*

115 *al*., 2015). These latter data suggest a role for both adaptive and neutral processes in governing
116 the diversity and biogeography of *Streptomyces*, though the degree to which these forces
117 constrain extant patterns of genetic diversity within the genus remain poorly described.
118
119 We evaluated *Streptomyces* biogeography by constructing an isolate collection from grassland
120 sites spanning the United States of America. The most powerful approach for detecting neutral
121 evolutionary processes associated with dispersal limitation is to examine patterns of
122 biogeography across similar habitats (Hanson *et al.*, 2012). In a previous study which employed
123 a single locus (*rpoB*), we observed that *Streptomyces* species diversity exhibited a significant
124 distance decay relationship and was inversely correlated with latitude (Andam *et al.,* 2016).
125 Evidence for a latitudinal diversity gradient coupled with evidence of demographic range
126 expansion, while not conclusive, is consistent with predictions of phylogenetic niche
127 conservatism (Wiens and Donoghue, 2004; Stevens, 2006) which posits that these
128 biogeographical patterns result from the historical effects of changes in paleoclimate (Hewitt *et*
129 *al*., 1996, 2000, and 2004). The hypothesis of a latitudinal diversity gradient in *Streptomyces*
130 predicts that, in addition to a gradient of species richness, there should also be latitudinal
131 differences in gene flow and genetic diversity within populations. Here we investigate genetic
132 diversity and geographic distribution within six species-like phylogenetic clusters, or
133 phylogroups, to determine whether patters of gene flow and diversity corroborate the hypothesis
134 of a latitudinal diversity gradient for *Streptomyces*.
135
136 **Materials and Methods**
137 *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 phylogentic 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

161 the five housekeeping loci *atpD*, *gyrB*, *recA*, *rpoB*, and *trpB* as described in Doroghazi and
162 Buckley (2010). Partial SSU rRNA sequences were amplified using universal primers (8F:
163 AGAGTTTGATCCTGGCTCAG; 1492R: CGGTTACCTTGTTACGACTT). Reactions for
164 Sanger sequencing were performed using forward primers for all loci except *rpoB*, for which the
165 reverse primer was used. Trace files were uniformly screened using CAP3 (Huang and Madan,
166 1999) interfaced with a Perl script. Sequences were aligned using MUSCLE (Edgar, 2004),
167 trimmed to 431 bp, 415 bp, 446 bp, 557 bp, and 489 bp, for genes *atpD*, *gyrB*, *recA*, *rpoB*, *trpB*,
168 respectively, and concatenated consistent with the genomic order in *Streptomyces coelicolor*
169 A3(2). SSU rRNA alignments were trimmed to 357 bp. Gene sequences are available on
170 GenBank with accession numbers XXXX-XXXX.
171
172 Good's coverage estimation and haplotype rarefaction was determined using mothur (Schloss *et*
173 *al.*, 2009). DNA polymorphism statistics including: number of segregating sites, nucleotide
174 diversity, and Tajima's D, were determined with DnaSP v5 (Librado and Rozas, 2009) and LDhat
175 (McVean *et al.*, 2002). Population scaled mutation rates (Watterson's theta; $\theta_w = 2N_e\mu$),
176 recombination or gene conversion rates ($\rho = 2N_e r/2$), and relative rates of recombination (ρ/θ_w)
177 were estimated using LDhat (McVean *et al.*, 2002) and are expressed per nucleotide site. The
178 standard index of association (I_A) was calculated from allelic data with LIAN v3.5 (Haubold and
179 Hudson, 2000) using the Monte-Carlo test and 100 iterations. The pairwise homoplasy index
180 (PHI) statistic was determined using PhiPack (Bruen *et al.*, 2006), and statistical significance
181 was evaluated under a null hypothesis of no recombination. Sequence identity across
182 phylogroups was calculated with mothur (Shloss *et al.*, 2009)
183

184 *Phylogengy reconstruction*

185 Maximum likelihood (ML) trees were constructed from the nucleotide sequences of individual
186 and concatenated MLSA loci using the generalized time reversible nucleotide substitution model
187 (Tavaré, 1986) with gamma distributed rate heterogeneity among sites (GTRGAMMA)
188 supported in RAxML v7.3.0 (Stamatakis, 2006). Bootstrap support was determined for the
189 highest-scoring ML tree of 20 iterations, and the number of bootstrap replicates was determined
190 using the extended majority rule (autoMRE) convergence criteria (Pattengale *et al*., 2010). Root
191 placement is defined by *Mycobacterium smegmatis*. Significant phylogenetic incongruence
192 between loci was determined using the Shimodaira-Hasegawa test (Shimodaira and Hasegawa,
193 1999) implemented in the R package phangorn (Schliep, 2011).

194

195 *Population structure*

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

204

205 *Patterns of dispersal and gene flow*

206 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 (π) 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

253 transfer within phylogroups as has been previously described in *Streptomyces* (Doroghazi and
254 Buckley, 2010). Likewise, there is evidence of significant levels of recombination within
255 phylogroups MAN196, MS200, MS152, and F34 (PHI test, $p < 0.05$, Table 2). Phylogroups
256 WA1063 and MS152 share two identical *atpD* alleles, with one of these alleles shared among
257 strains isolated from MS, CA, WI (Brookfield), AK (Manley Hot Springs), and WA, and the
258 other allele shared among strains isolated from MS and WI (Brookfield) (Figure 3, Figure S3B).
259 It is not clear whether these alleles are shared as a result of contemporary horizontal gene
260 transfer or vertical inheritance from the most recent common ancestor of the two clades, though
261 the latter explanation is more parsimonious given the low level of polymorphism between the
262 phylogroups.
263
264 *Dispersal limitation*
265 Strains of the six phylogroups were obtained from soil samples from 13 sites of diverse
266 geographic origin (Table 1). Each phylogroup was detected in 4.2 ± 0.4 sites, and this
267 distribution differs significantly from expectations for a random distribution of strains across
268 sites (permutation test, $p < 0.0005$), thereby rejecting the hypothesis of panmixia. Each
269 phylogroup subpopulation was observed in 2.2 ± 0.9 sites (mean ± s.d.), and this value is lower
270 than expected if subpopulations are randomly distributed across the sites occupied by each
271 phylogroup (permutation test, $p < 0.001$). These results indicate that phylogroup distribution is
272 constrained geographically and that phylogroups have subpopulation structure that is also
273 geographically explicit.
274
275 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 expect *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

299 between WI (Brookfield) and OR (Figure 4D) is consistent with allopatric fragmentation.

300

301 *Latitudinal diversity gradient*

302 The distribution and diversity of the phylogroups reveals a latitudinal diversity gradient. Strains
303 from MAN125, MAN196, WA1063 occur mostly north of 40°N latitude, while strains from
304 MS200, MS152, F34 occur mostly south of this latitude (Table 1). This pattern of north/south
305 partitioning is significant for each phylogroup when evaluated against the expectation of a
306 random distribution across sites (permutation test, $p < 0.01$ for each phylogroup after Bonferroni
307 correction). Furthermore, partial Mantel tests were performed to evaluate the latitudinal and
308 longitudinal vector components of geographic distance in relation to the allele composition of
309 sites. There remains a significant relationship between allele composition and geographic
310 distance when we control for longitude, (Mantel r = 0.23, $p = 0.022$), but this relationship is no
311 longer significant when we control for latitude (Mantel r = 0.15, $p = 0.12$). This result indicates
312 that allele composition changes more across latitude than it does across longitude. The latitudinal
313 partitioning of alleles can be readily observed in the pattern of allele sharing between sites
314 (Figure 3). Finally, we also observed a significant relationship between per site nucleotide
315 diversity of phylogroup MLSA loci and the average latitude of sites in which they are found (R =
316 -0.91, $p = 0.012$; Figure 5). This result indicates that phylogroups recovered from lower latitudes
317 have higher genetic diversity than those recovered from higher latitudes.

318

319 **Discussion**

320 We used population genetic approaches to analyze spatial patterns of genetic diversity for six
321 *Streptomyces* phylogroups isolated from geographically disparate but ecologically similar sites

322 across the United States (Table 1). The distribution of phylogroups is nonrandom and dispersal
323 limited (Figure S4) with phylogroups inhabiting geographic ranges defined by latitude (Table 1,
324 Figure 3). In addition, the genetic diversity of phylogroups is inversely proportional to the
325 latitude from which they were isolated (Figure 5). These findings suggest that that there are
326 latitudinal barriers to dispersal, and that patterns of *Streptomyces* biogeography result from
327 dispersal limitation and regional diversification due to genetic drift.
328
329 The phylogroups we describe are coherent phylogenetic groups that approximate biological
330 populations (Figure S1). The members of each phylogroup share a distinguishable geographic
331 range (Table 1, Figure 3), a recent common ancestor (Figure 1), greater than 97% ANI across
332 MLSA loci (Table 2), and greater than 99.4% SSU rRNA gene nucleotide identity. Despite
333 geographic and genetic subpopulation structure (Figure 1), phylogroup members frequently share
334 identical alleles across demes. We infer the presence of recombination within all phylogroups
335 using both nucleotide polymorphism and phylogenetic methods (Table 2, Table S2). For instance,
336 MAN125 is nearly in linkage equilibrium (I_A = 0.09) though its members span a geographic
337 range of over 2,000 km across the Pacific Northwest. The observation of linkage equilibrium
338 indicates that either the population lacks contemporary barriers to gene flow or the population
339 has experienced a recent demographic expansion in which there has been insufficient time to
340 accumulate mutations and linkage disequilibrium. Significant recombination and linkage
341 equilibrium have also been observed in a *S. pratensis* population which spanned 1,000 km across
342 sites present in North Carolina and northern New York (Doroghazi and Buckley, 2010 and 2014).
343

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

436 time (100-300 generations per year) for *E. coli* (Ochman *et al*., 1999), since corresponding
437 values are not available for *Streptomyces* or their relatives. Based upon these gross
438 approximations, we would estimate that WA1063 and MS152 diverged 15,000-50,000 years ago,
439 corresponding to events in the late Pleistocene (Clayton *et al*., 2006). Hence, it is likely that the
440 identical *atpD* alleles found in both WA1063 and MS152 were inherited without modification
441 from an ancestral population that both groups share (Figure S3B). If this is the case it suggests
442 that the divergence of phylogroup WA1063 was driven by range expansion into higher latitudes
443 and subsequent isolation by distance from the ancestral population it shares with phylogroup
444 MS152.
445
446 Through population genetic analysis of six *Streptomyces* phylogroups, we find evidence for
447 dispersal limitation associated with geographically explicit patterns of gene flow which manifest
448 in a latitudinal gradient of nucleotide diversity. Furthermore, these data support the hypothesis
449 that historical demographic processes influence the contemporary biogeography of *Streptomyces*.
450 Due to their spore forming capabilities and potential for long range dispersal, *Streptomyces* are
451 an ideal system for assessing limits on gene flow among terrestrial bacteria. We cannot yet
452 determine whether the cause of dispersal limitation is due to limits on spore mobility or due to
453 adaptive constraints as proposed by the hypothesis of phylogenetic niche conservatism. A better
454 understanding of *Streptomyces* biogeography and the evolutionary processes driving population
455 structure may ultimately assist in the discovery of novel genetic diversity and possibly novel
456 antibiotics within this genus.
457
458

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

Allen AP, Brown JH, Gillooly JF. (2002). Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science* **297**: 1545–1548.

Andam CP, Choudoir MJ, Nguygen AVT, Park HS, Buckley DH. (2015). Contributions of ancestral inter-species recombination to the genetic diversity of extant *Streptomyces* lineages. *In Press.*

Andam CP, Doroghazi JR, Campbell AN, Kelly PJ, Choudoir MJ, Buckley DH. (2016). A latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*. *Personal communication.*

Antony-Babu S, Stach JEM, Goodfellow M. (2008). Genetic and phenotypic evidence for *Streptomyces griseus* ecovars isolated from a beach and dune sand system. *Antonie Van Leeuwenhoek* **94**: 63–74.

Bell T. (2010). Experimental tests of the bacterial distance-decay relationship. *ISME J* **4**: 1357–1365.

Bernatchez L, Wilson CC. (1998). Comparative phylogeography of Nearctic and Palearctic fishes. *Mol Ecol* **7**: 431–452.

480 Bissett A, Richardson AE, Baker G, Wakelin S, Thrall PH. (2010). Life history determines
481 biogeographical patterns of soil bacterial communities over multiple spatial scales. *Mol*
482 *Ecol* **19**: 4315–4327.
483 Boucher Y, Cordero OX, Takemura A, Hunt DE, Schliep K, Bapteste E, *et al*. (2011). Local
484 mobile gene pools rapidly cross species boundaries to create endemicity within global
485 *Vibrio cholerae* populations. *mBio* **2**: e00335–10.

486 Bruen TC, Philippe H, Bryant D. (2005). A simple and robust statistical test for detecting the
487 presence of recombination. *Genetics* **172**: 2665–2681.
488 Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, *et al.* (2012). Patterns
489 of gene flow define species of thermophilic Archaea. *PLoS Biol* **10**: e1001265.
490 Cho J, Tiedje JM. (2000). Biogeography and degree of endemicity of fluorescent *Pseudomonas*
491 strains in soil. *Appl Environ Microbiol* **66**: 5448-5456.
492 Choudoir MJ, Campbell AN, Buckley DH. (2012). Grappling with Proteus: population level
493 approaches to understanding microbial diversity. *Front Microbiol* **3**: 336.
494 Chu H, Fierer N, Lauber CL, Caporaso JG, Knight R, Grogan P. (2010). Soil bacterial diversity
495 in the Arctic is not fundamentally different from that found in other biomes. *Environ*
496 *Microbiol* **12**: 2998–3006.
497 Clayton L, Attig JW, Mickelson DM, Johnson MD, Syverson KM. (2006). Glaciation of
498 Wisconsin.
499 Clement M, Posada D, Crandall KA. (2000). TCS: A computer program to estimate gene
500 genealogies. *Mol Ecol* **9**: 1657–1659.
501 Coleman ML, Chisholm SW. (2010). Ecosystem-specific selection pressures revealed through
502 comparative population genomics. *Proc Natl Acad Sci U S A* **107**: 18634–18639.

503 Conroy CJ, Cook JA. (2000). Phylogeography of a post-glacial colonizer: *Microtus longicaudus*
504 (Rodentia: Muridae). *Mol Ecol* **9**: 165–175.
505 Currie DJ, Mittelbach GG, Cornell HV, Field R, Guégan JF, Hawkins BA*, et al.* (2004).
506 Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic
507 richness. *Ecol Lett* **7**: 1121–1134.
508 Davelos AL, Xiao K, Samac DA, Martin AP, Kinkel LL. (2004a). Spatial variation in
509 *Streptomyces* genetic composition and diversity in a prairie soil. *Microb Ecol* **48**: 601–
510 612.
511 Davelos AL, Kinkel LL, Samac DA. (2004b). Spatial variation in frequency and intensity of
512 antibiotic interactions among *Streptomycetes* from prairie soil. *Appl Environ Microbiol*
513 **70**: 1051–1058.
514 Doroghazi JR, Buckley DH. (2014). Intraspecies comparison of *Streptomyces pratensis* genomes
515 reveals high levels of recombination and gene conservation between strains of disparate
516 geographic origin. *BMC Genomics* **15**: 970.
517 Doroghazi JR, Buckley DH. (2010). Widespread homologous recombination within and between
518 *Streptomyces* species. *ISME J* **4**: 1136–1143.
519 Earl DA, vonHoldt BM. (2012). STRUCTURE HARVESTER: a website and program for
520 visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet*
521 *Resour* **4**: 359–361.
522 Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high
523 throughput. *Nucleic Acid Res* **32**: 1792–1797.
524 Eisenlord SD, Zak DR, Upchurch RA. (2012). Dispersal limitation and the assembly of soil
525 *Actinobacteria* communities in a long-term chronosequence. *Ecol Evol* **2**: 538–549.

526 El-Nakeeb MA, Lechevalier HA. (1963). Selective isolation of aerobic Actinomycetes. *Appl*
527 *Microbiol* 11: 75–77.
528 Evanno G, Regnaut S, Goudet J. (2005). Detecting the number of clusters of individuals using
529 the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.
530 Fierer N, Carney KM, Horner-Devine MC, Megonigal JP. (2009). The biogeography of
531 ammonia-oxidizing bacterial communities in soil. *Microb Ecol* **58**: 435–445.
532 Fierer N, Jackson RB. (2006). The diversity and biogeography of soil bacterial communities.
533 *Proc Natl Acad Sci U S A* **103**: 626–631.
534 Finlay BJ. (2002). Global dispersal of free-living microbial eukaryote species. *Science* **296**:
535 1061–1063.
536 Finlay BJ, Fenchel T. (2004). Cosmopolitan metapopulations of free-living microbial eukaryotes.
537 *Protist* **155**: 237–244.
538 Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown M V, Green JL, *et al.* (2008). A
539 latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci U S A*
540 **105**: 7774–7778.
541 Gibbons SM, Caporaso JG, Pirrung M, Field D, Knight R, Gilbert JA. (2013). Evidence for a
542 persistent microbial seed bank throughout the global ocean. *Proc Natl Acad Sci U S A*
543 **110**: 4651–4655.
544 Gilbert JA, Steele JA, Caporaso JG, Steinbrück L, Reeder J, Temperton B, *et al.* (2012).
545 Defining seasonal marine microbial community dynamics. *ISME J* **6**: 298–308.
546 Goslee SC, Urban DL. (2007). The ecodist package for dissimilarity-based analysis of ecological
547 data. *J Stat Softw* **22**: 1–19.

548 Green J, Bohannan BJM. (2006). Spatial scaling of microbial biodiversity. *Trends Ecol Evol* **21**:
549 501–507.
550 Guo Y, Zheng W, Rong X, Huang Y. (2008). A multilocus phylogeny of the *Streptomyces*
551 *griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete
552 systematics. *Int J Syst Evol Microbiol* **58**: 149–159.
553 Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. (2012). Beyond biogeographic
554 patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.
555 Hatosy SM, Martiny JBH, Sachdeva R, Steele J, Fuhrman JA, Martiny AC. (2013). Beta
556 diversity of marine bacteria depends on temporal scale. *Ecology* **94**: 1898–1904.
557 Haubold B, Hudson RR. (2000). LIAN 3.0: detecting linkage disequilibrium in multilocus data.
558 *Bioinformatics* **16**: 847–848.
559 Hewitt GM. (1996). Some genetic consequences of ice ages, and their role in divergence and
560 speciation. *Biol J Linnean Soc* **58**: 247–276.
561 Hewitt G. (2000). The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.
562 Hewitt GM. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philos*
563 *Trans R Soc Lond B Biol Sci* **359**: 183–195.
564 Hillebrand H. (2014). On the generality of the latitudinal diversity gradient. *Amer Nat* **163**: 192–
565 211.
566 Huang X, Madan A. (1999). CAP3: a DNA sequence assembly program. *Genome Res* **9**: 868–
567 677.
568 Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. (2008). Resource partitioning
569 and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 1081–
570 1085.

571 Jakobsson M, Rosenberg NA. (2007). CLUMPP: a cluster matching and permutation program
572 for dealing with label switching and multimodality in analysis of population structure.
573 *Bioinformatics* **23**: 1801–1806.
574 Jaspers E, Overmann J. (2004). Ecological significance of microdiversity: identical 16S rRNA
575 gene sequences can be found in bacteria with highly divergent genomes and
576 ecophysiologies. *Appl Environ Microbiol* **70**: 4831–4839.
577 Jolley KA, Chan MS, Maiden MCJ. (2004). mlstdbNet - distributed multi-locus sequence typing
578 (MLST) databases. *BMC Bioinformatics* **5**: 86.
579 Kâmpfer P. (2006). The family *Streptomycetaceae* Part I: Taxonomy. In: The Prokaryotes,
580 Volume 3. Dworkin, M, Falkow, S, Rosenberg, E, Schliefer, K, & Stackebrandt, E (eds).
581 Springer: New York, pp pp. 538–604.
582 Kenefic LJ, Pearson T, Okinaka RT, Schupp JM, Wagner DM, Hoffmaster AR, *et al.* (2009).
583 Pre-Columbian origins for North American anthrax. *PLoS One* **4**: e4813.
584 Kieser T, Bibb M, Buttner M, Charter K, Hopwood D. (2000). Practical *Streptomyces* Genetics.
585 John Innes Foundation: Norwich, UK.
586 Knowles L, Maddison W. (2002). Statistical phylogeography. *Mol Ecol* **11**: 2623–2635.
587 Labeda DP. (2011). Multilocus sequence analysis of phytopathogenic species of the genus
588 *Streptomyces*. *Int J Syst Evol Microbiol* **61**: 2525–2531.
589 Lauber CL, Hamady M, Knight R, Fierer N. (2009). Pyrosequencing-based assessment of soil
590 pH as a predictor of soil bacterial community structure at the continental scale. *Appl*
591 *Environ Microbiol* **75**: 5111–5120.
592 Librado P, Rozas J. (2009). DnaSP v5: a software for comprehensive analysis of DNA
593 polymorphism data. *Bioinformatics* **25**: 1451–1452.

Loria R, Kers J, Joshi M. (2006). Evolution of plant pathogenicity in *Streptomyces*. *Annu Rev Phytopathol* **44**: 469–487.

Lozupone CA, Knight R. (2007). Global patterns in bacterial diversity. *Proc Natl Acad Sci U S A* **104**: 11436–11440.

Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, *et al.* (2008). Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology* **89**: S108–S122.

Mantel N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**: 209–220.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, *et al.* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.

Martiny JBH, Eisen JA, Penn K, Allison SD, Horner-Devine MC. (2011). Drivers of bacterial β -diversity depend on spatial scale. *Proc Natl Acad Sci U S A* **108**: 7850–7854.

McCarthy AJ, Williams ST. (1992). Actinomycetes as agents of biodegradation in the environment - a review. *Gene* **115**: 189–192.

McVean G, Awadalla P, Fearnhead P. (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.

Mikheyev AS, Vo T, Mueller UG. (2008). Phylogeography of post-Pleistocene population expansion in a fungus-gardening ant and its microbial mutualists. *Mol Ecol* **17**: 4480–4488.

Mila B, Smith T, Wayne R. (2006). Postglacial population expansion drives the evolution of long-distance migration in a songbird. *Evolution* **60**: 2403–2409.

617 Miller SR, Strong AL, Jones KL, Ungerer MC. (2009). Bar-coded pyrosequencing reveals shared
618 bacterial community properties along the temperature gradients of two alkaline hot
619 springs in Yellowstone National Park. *Appl Environ Microbiol* **75**: 4565–4572.
620 Mittelbach GG, Schemske DW, Cornell H V., Allen AP, Brown JM, Bush MB, *et al.* (2007).
621 Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography.
622 *Ecol Lett* **10**: 315–331.
623 Mouchet M, Levers C, Zupan L, Kuemmerle T, Plutzar C, Erb K, *et al.* (2015). Testing the
624 effectiveness of environmental variables to explain European terrestrial vertebrate species
625 richness across biogeographical scales. *PLoS One* **10**: e0131924.
626 Neufeld JD, Mohn WW. (2005). Unexpectedly high bacterial diversity in arctic tundra relative to
627 boreal forest soils, revealed by serial analysis of ribosomal sequence tags. *Appl Environ*
628 *Microbiol* **71**: 5710–5718.
629 Nielsen R, Beaumont MA. (2009). Statistical inferences in phylogeography. *Mol Ecol* **18**: 1034–
630 1047.
631 Ochman H, Wilson AC. (1987). Evolution in cacteria: evidence for a universal substitution rate
632 in cellular genomes. *J Mol Evol* **26**: 74–86.
633 Ochman H, Elwyn S, Moran NA. (1999). Calibrating bacterial evolution. *Proc Natl Acad Sci U S*
634 *A* **96**: 12638–12643.
635 Ottow JCG. (2013). Rose bengal as a selective aid in the isolation of fungi and Actinomycetes
636 from natural sources. *Mycologia* **64**: 304–315.
637 Panchal M. (2006). The automation of Nested Clade Phylogeographic Analysis. *Bioinformatics*
638 **23**: 509–510.

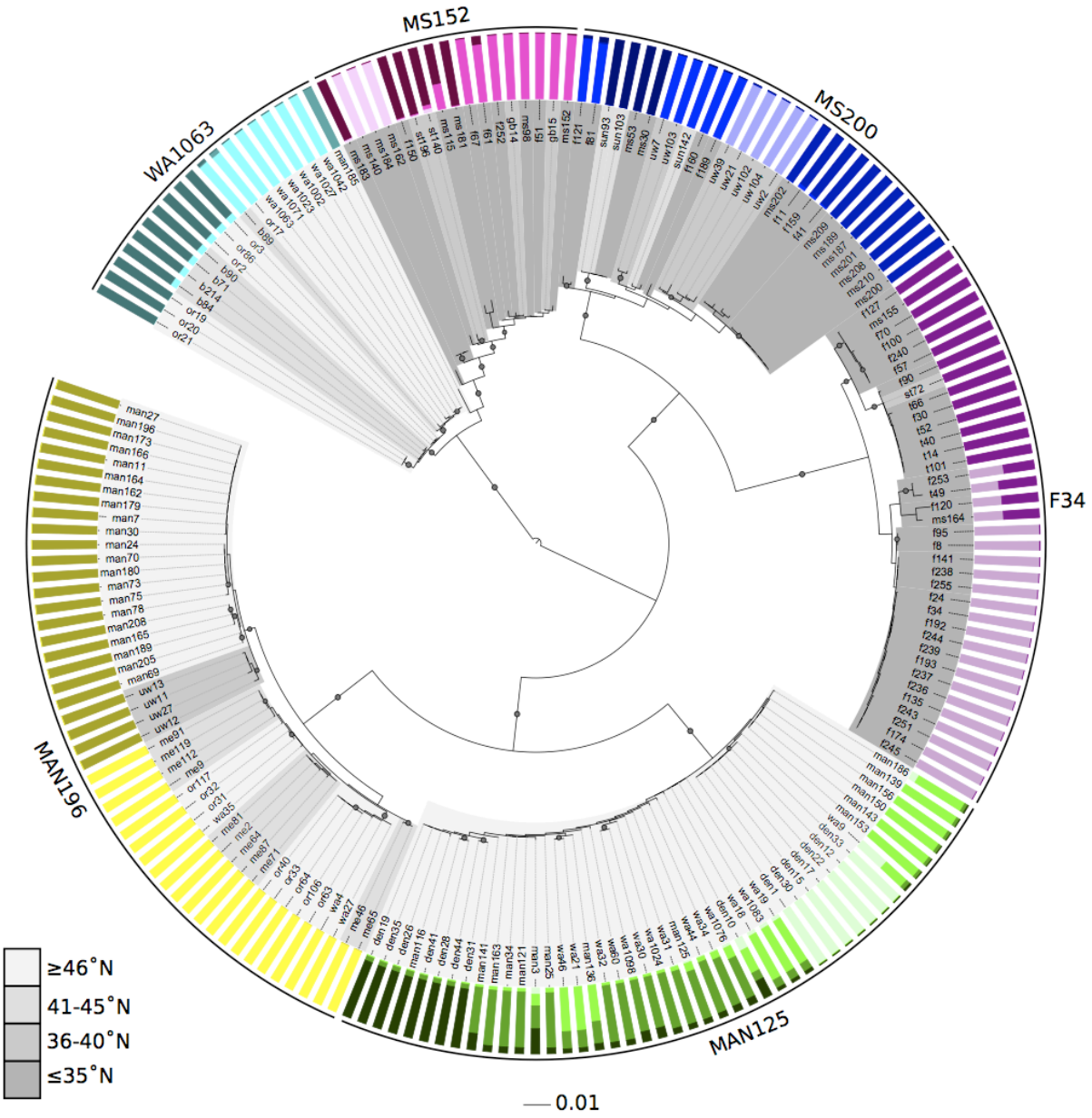
639 Papke RT, Ramsing NB, Bateson MM, Ward DM. (2003). Geographical isolation in hot spring
640 cyanobacteria. *Environ Microbiol* **5**: 650–659.
641 Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. (2010). How
642 many bootstrap replicates are necessary? *J Comput Biol* **17**: 337–354.
643 Pearson T, Giffard P, Beckstrom-Sternberg S, Auerbach R, Hornstra H, Tuanyok A*, et al*.
644 (2009). Phylogeographic reconstruction of a bacterial species with high levels of lateral
645 gene transfer. *BMC Biol* **7**: 78.
646 Pommier T, Canback B, Lundberg P, Hagstrom A, Tunlid A. (2009). RAMI: a tool for
647 identification and characterization of phylogenetic clusters in microbial communities.
648 *Bioinformatics* **25**: 736–742.
649 Posada D, Crandall KA, Templeton AR. (2000). GeoDis: A program for the cladistic nested
650 analysis of the geographical distribution of genetic haplotypes. *Mol Ecol* **9**: 487–488.
651 Pritchard JK, Stephens M, Donnelly P. (2000). Inference of population structure using multilocus
652 genotype data. *Genetics* **155**: 945–959.
653 Rong X, Huang Y. (2010). Taxonomic evaluation of the *Streptomyces griseus* clade using
654 multilocus sequence analysis and DNA-DNA hybridization, with proposal to combine 29
655 species and three subspecies as 11 genomic species. *Int J Syst Evol Microbiol* **60**: 696–
656 703.
657 Rong X, Huang Y. (2012). Taxonomic evaluation of the *Streptomyces hygroscopicus* clade using
658 multilocus sequence analysis and DNA-DNA hybridization, validating the MLSA
659 scheme for systematics of the whole genus. *Syst Appl Microbiol* **35**: 7–18.

660 Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, *et al*. (2010). Soil
661 bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* **4**: 1340–
662 1351.
663 Schliep KP. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics* **27**: 592–593.
664 Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, *et al*. (2009).
665 Introducing mothur: open-source, platform-independent, community-supported software
666 for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–
667 7541.
668 Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, *et al*. (2012).
669 Population genomics of early events in the ecological differentiation of bacteria. *Science*
670 **336**: 48–51.
671 Shimodaira H, Hasegawa M. (1999). Multiple comparisons of log-likelihoods with applications
672 to phylogenetic inference. *Mol Biol Evol* **16**: 1114–1116.
673 Smouse PE, Long JC, Sokal RR. (1986). Multiple regression and correlation extensions of the
674 Mantel test of matrix correspondence. *Syst Zool* **35**: 627–632.
675 Soltis D, Gitzendanner M, Strenge D, Soltis P. (1997). Chloroplast DNA intraspecific
676 phylogeography of plants from the Pacific Northwest of North America. *Plant Syst Evol*
677 **206**: 353–373.
678 Stamatakis A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
679 thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
680 Stevens RD. (2006). Historical processes enhance patterns of diversity along latitudinal
681 gradients. *Proc R Soc B Biol Sci* **273**: 2283–2289.

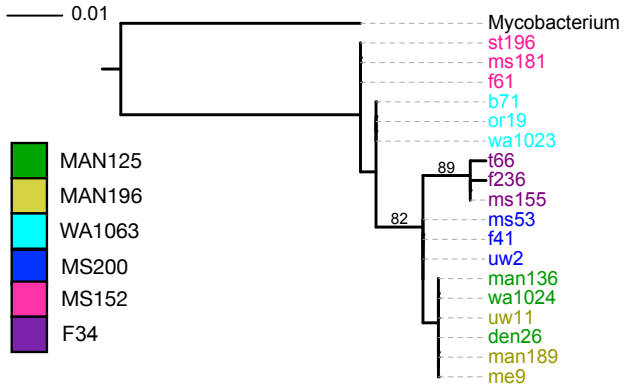
682 Sul WJ, Oliver TA, Ducklow HW, Amaral-zettler LA, Sogin ML. (2012). Marine bacteria
683 exhibit a bipolar distribution. *Proc Natl Acad Sci U S A* **110**: 2342–2347.
684 Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM, *et al*. (2013).
685 Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the
686 surface ocean. *Proc Natl Acad Sci U S A* **110**: 11463–11468.
687 Takasuka TE, Book AJ, Lewin GR, Currie CR, Fox BG. (2013). Aerobic deconstruction of
688 cellulosic biomass by an insect-associated *Streptomyces*. *Sci Rep* **3**: 1–10.
689 Tavaré S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences.
690 *Am Math Soc Lect Math Life Sci* **17**: 57–86.
691 Telford RJ, Vandvik V, Birks HJB. (2006). Dispersal limitations matter for microbial
692 morphospecies. *Science* **312**: 1015.
693 Templeton AR, Boerwinkle E, Sing CF. (1987). A cladistic analysis of phenotypic associations
694 with haplotypes inferred from restriction endonuclease mapping. i. basic theory and an
695 analysis of alcohol dehydrogenase activity in Drosophila. *Genetics* **117**: 343–351.
696 Templeton AR, Routman E, Phillips CA. (1995). Separating population structure from
697 population history: a cladistic analysis of the geographical distribution of mitochondrial
698 DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* **140**: 767–782.
699 Vos M, Velicer GJ. (2008). Isolation by distance in the spore-forming soil bacterium
700 *Myxococcus xanthus*. *Curr Biol* **18**: 386–391.
701 Wallace A. (1878). Tropical nature, and other essays. Macmillan and Co.: London.
702 Watve M, Tickoo R, Jog M, Bhole B. (2001). How many antibiotics are produced by the genus
703 *Streptomyces*? *Arch Microbiol* **176**: 386–390.

704 Wawrik B, Kutliev D, Abdivasievna UA, Kukor JJ, Zylstra GJ, Kerkhof L. (2007).
705 Biogeography of Actinomycete communities and type ii polyketide synthase genes in
706 soils collected in New Jersey and Central Asia. *Appl Environ Microbiol* **73**: 2982–2989.
707 Welch RA, Burland V, Plunkett G, Redford P, Roesch P, Rasko D, *et al*. (2002). Extensive
708 mosaic structure revealed by the complete genome sequence of uropathogenic
709 *Escherichia coli*. *Proc Natl Acad Sci U S A* **99**: 17020–17024.
710 Whitaker RJ, Grogan DW, Taylor JW. (2003). Geographic barriers isolate endemic populations
711 of hyperthermophilic archaea. *Science* **301**: 976–978.
712 Wiens JJ, Donoghue MJ. (2004). Historical biogeography, ecology and species richness. *Trends*
713 *Ecol Evol* **19**: 639–644.
714 Williams D, Dunkerley D, DeDecker P, Kershaw P, Chappell M. (1998). Quaternary
715 environments. Arnold: London.
716 Wilson AB, Veraguth IE. (2010). The impact of Pleistocene glaciation across the range of a
717 widespread European coastal species. *Mol Ecol* **19**: 4535–53.

Figures and Tables



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



728 **Figure 2.** Dissimilarity in SSU rRNA genes is less than 3% across all six phylogroups and
729 ranges from 0-0.6% for representatives within each phylogroup, revealing that this locus lacks
730 sufficient sensitivity to examine the biogeographic distribution of *Streptomyces* phylogroups.
731 The tree was constructed using maximum likelihood with a GTRGAMMA evolution model
732 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.

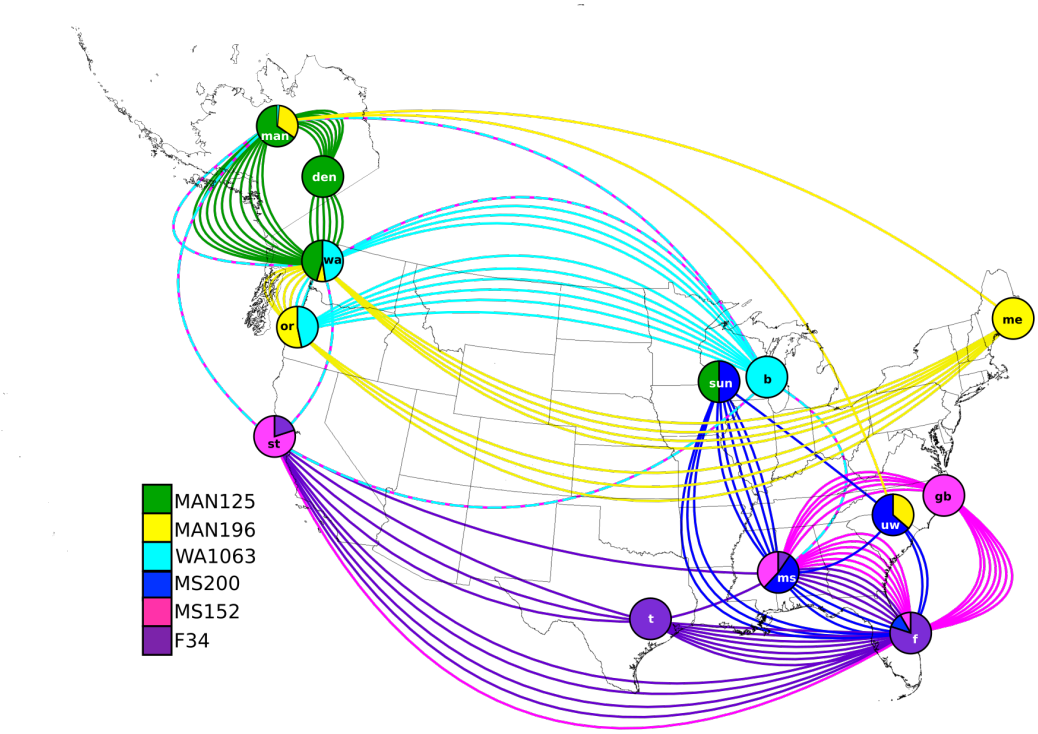
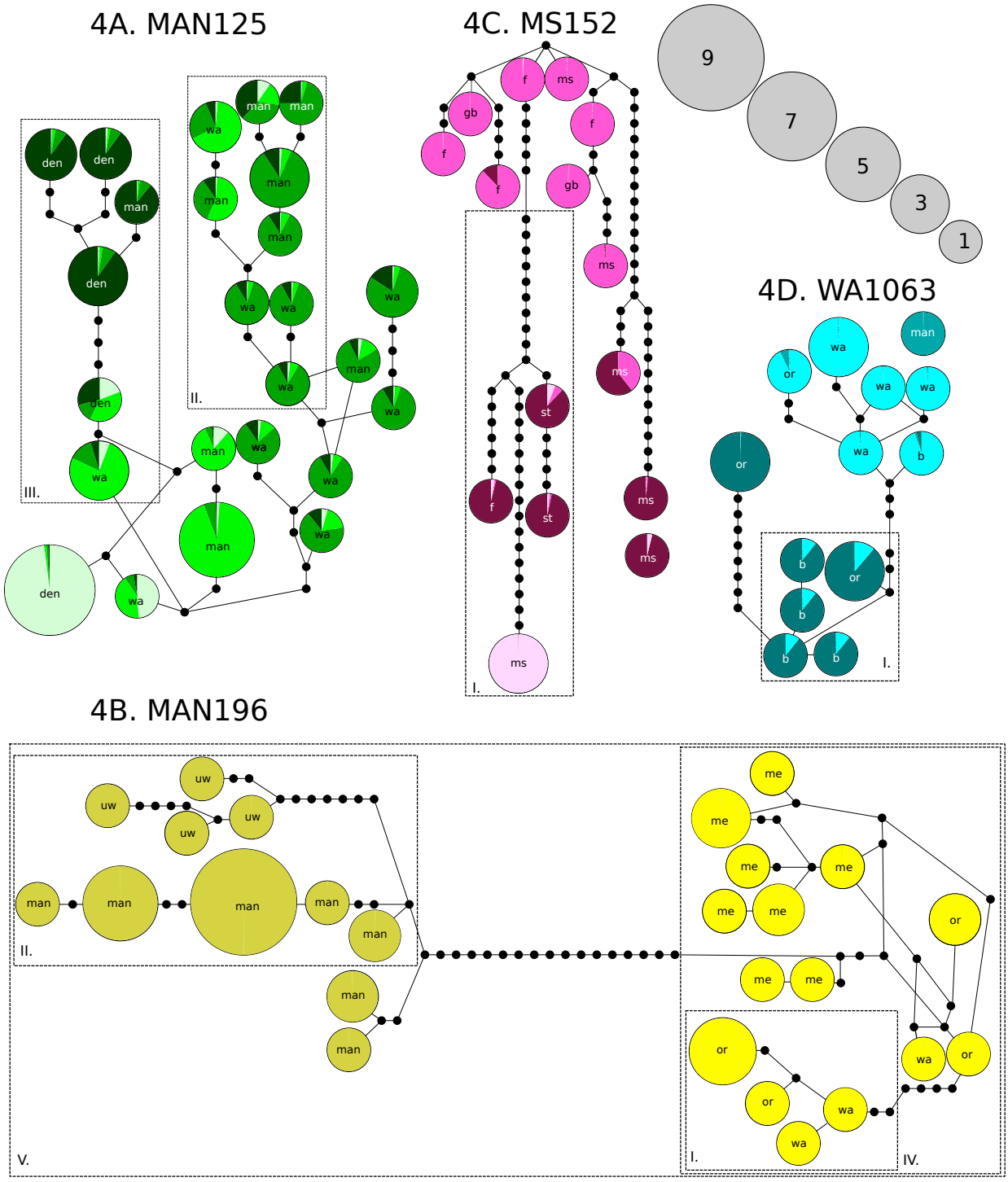
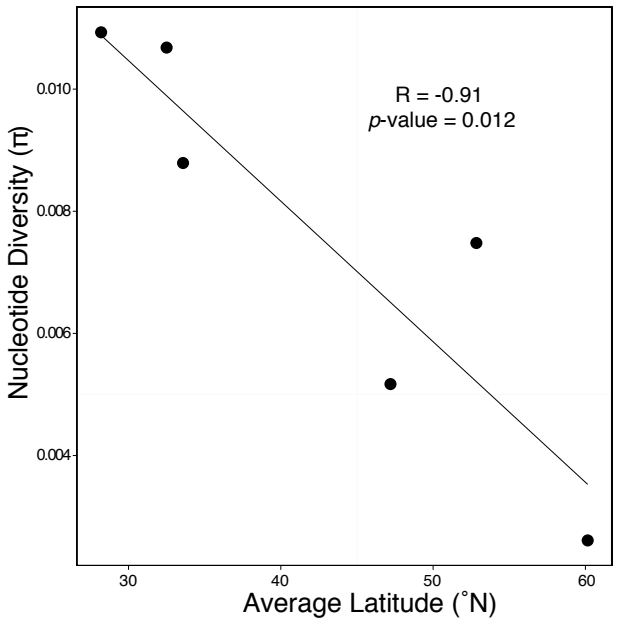


Figure 3. Identical MLSA alleles are frequently shared between sites. Circles depict sample sites and are labeled according to site code (Table 1). The relative abundance of each phylogroup at each site is indicated by color according to the legend, with raw counts provided in Table 1. Solid colored 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).



742 **Figure 4.** Haplotype networks are depicted for phylogroups MAN125 (4A), MAN196 (4B),
743 MS152 (4C), WA1063 (4D). Circles represent unique MLSA haplotypes whose radius is
744 proportional to the number of isolates sharing that haplotype, and colors correspond to strain

745 ancestry and subpopulation affiliation as determined by Structure analysis and defined in Figure
746 1. Haplotypes are labeled with a letter code referring to sample site location as indicated in Table
747 1. Black circles represent un-sampled, inferred haplotypes with each circle designating a single
748 nucleotide polymorphism. The length of edges between nodes is uninformative. Dashed
749 rectangles encompass clades that have significant phylogeographic inferences from nested clade
750 analysis, as described in methods. Roman numerals correspond to the following inferences: I.
751 Allopatric fragmentation; II. Long distance colonization and/or past fragmentation; III.
752 Restricted gene flow with isolation by distance; IV. Restricted gene flow but with some long-
753 distance gene flow over intermediate ranges not occupied by the species; or past gene flow
754 followed by extinction of intermediate populations; V. Contiguous range expansion.



755 **Figure 5.** There is a significant inverse correlation between phylogroup average latitude and
756 nucleotide diversity. Average latitude was determined using a weighted average based on the
757 total number of isolates per site as indicated in Table 1. Nucleotide diversity was calculated using
758 concatenated MLSA loci and is expressed per site.

759 **Table 1.** The six targeted phylogroups were observed in 13 sites. A total of 755 *Streptomyces*
760 strains were isolated from these sites, and 308 of these strains were found to represent the six
761 targeted phylogroups. The numbers of isolates per site that belonged to each of our target
762 phylogroups is indicated. Isolate names begin with a letter code referring to the isolation site.

Site	**Code**	**Latitude**	**Longitude**	**Isolates /Site**	**Isolates from *Streptomyces* sp. phylogroup**					
					MAN 125	MAN 196	WA 1063	MS 200	MS 152	F34
Manley Hot Springs, AK	man	63.87˚N	-149.02˚W	95	51	26	1			
Denali Hwy, AK	den	63.22˚N	-147.68˚W	40	40					
Bothell, WA	wa	47.73˚N	-122.24˚W	105	21	5	21			
Astoria, OR	or	46.18˚N	-123.85˚W	79		8	7			
Kennebunk, ME	me	43.4˚N	-70.54˚W	85		13				
Sun Prairie, WI	sun	43.17˚N	-89.24˚W	65				3		
Brookfield, WI	b	43.06˚N	-88.13˚W	22	3		5			
Palo Alto, CA	st	37.43˚N	-122.17˚W	34					2	1
Greensboro, NC	gb	36.09˚N	-79.89˚W	9		1			2	
Troy (Uwharrie), NC	uw	35.71˚N	-79.88˚W	19		4		7		
Starkville, MS	ms	33.46˚N	-88.8˚W	90				11	8	2
Austin, TX	t	30.2˚N	-97.77˚W	42						6
Fort Pierce, FL	f	27.54˚N	-80.35˚W	70				7	5	48
Total				755	115	57	34	28	17	57

763 **Table 2.** MLSA was performed on a total of 17-47 strains from each phylogroup. Summary
764 statistics for concatenated MLSA nucleotide sequences were determined as described in
765 methods. Phylogroup average latitude (Ave Lat) was determined using the total number of
766 isolates per phylogroup provided in Table 1. The standard index of association is zero for a
767 population in linkage equilibrium.

Phylo-group	Ave Lat	n	Haplo-types	Min ANI	S	π	Tajima's D	θ_W	ρ	ρ/θ_W	I_A	Φ_W
MAN125	60.15˚N	45	25	99.5	23	0.0026	0.53	0.0023	0.0094	4.18	0.09	2.95E-01
MAN196	52.84˚N	47	26	98.5	65	0.0075	0.67	0.0063	0.00043	0.068	0.49	4.24E-02*
WA1063	47.20˚N	19	13	98.1	63	0.0052	-1.35	0.0077	0	0	0.43	4.33E-03
MS200	33.58˚N	27	21	98.4	93	0.0088	-0.58	0.0103	0.0017	0.17	0.39	1.33E-03**
MS152	32.50˚N	17	15	97.6	96	0.0107	-0.51	0.012	0.0013	0.11	0.35	1.078E-01**
F34	28.20˚N	36	22	97.7	93	0.0109	0.52	0.0096	0.0011	0.11	0.46	1.79E-01**

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 ($2N_e\mu$)
ρ: per site rate of recombination/gene conversion ($2N_er/2$)
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