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
School of Integrative Plant Science
Bradfield Hall 705, Cornell University, Ithaca, NY 14853
607-255-1716

Running Title: Biogeography of *Streptomyces*

Conflict of Interest: None

Originality-Significance Statement

We provide the first population genetic evidence that patterns of *Streptomyces* biogeography, which manifest in geographically explicit patterns of gene flow and a latitudinal gradient of nucleotide diversity, result from dispersal limitation and regional diversification due to drift. This contribution elucidates evolutionary processes that underlie patterns of microbial biogeography.

Summary

We examined the biogeography of *Streptomyces* at regional spatial scales to identify factors that govern patterns of microbial diversity. *Streptomyces* are spore forming filamentous bacteria which are widespread in soil. *Streptomyces* strains were isolated from perennial grass habitats sampled across a spatial scale of more than 6,000 km. Previous analysis of this geographically explicit culture collection provided evidence for a latitudinal diversity gradient in *Streptomyces* species. Here we evaluate the hypothesis that this latitudinal diversity gradient is a result of evolutionary dynamics associated with historical demographic processes. Historical demographic phenomena have genetic consequences that can be evaluated through analysis of population genetics. We applied population genetic approaches to analyze population structure in six of the most numerically abundant and geographically widespread *Streptomyces* phylogroups from our culture collection. *Streptomyces* population structure varied at regional spatial scales and allelic diversity correlated with geographic distance. In addition, allelic diversity and gene flow are partitioned by latitude. Finally, we found that nucleotide diversity within phylogroups is negatively correlated with latitude. These results indicate that phylogroup diversification is constrained by dispersal limitation at regional spatial scales and they are consistent with the hypothesis that historical demographic processes have influenced the contemporary biogeography of *Streptomyces*.

Introduction

Patterns of microbial biogeography have been widely documented (Whitaker *et al*., 2003; Vos and Velicer, 2008; Bissett *et al*., 2010; Martiny *et al*., 2011; Gilbert *et al*., 2012; Hatosy *et al*., 2013), and 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. Biogeographical patterns are ultimately governed by rates of dispersal and diversification (Martiny *et al*., 2006; Hanson *et al*., 2012). Since microbial dispersal cannot be observed directly, rates of 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). 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). Contradictory findings in the literature can be explained by at least two factors: first, dispersal constraints are likely to vary with respect to different species and habitats; and second, methods used to define units of diversity vary dramatically in their taxonomic and 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 between habitats and species. At one end of the spectrum, globally widespread microbes such as *Prochlorococcus* and *Pelagibacter* show little variation in gene content between the Atlantic and Pacific Oceans suggesting that

47 dispersal can homogenize genetic diversity in pelagic systems (Coleman and Chisholm, 2010). At
48 the other end of the spectrum are extremophiles such as *Sulfolobus* and thermophilic
49 *Synechococcus*, which live in island-like volcanic habitats and exhibit strong patterns of allopatric
50 divergence resulting from dispersal limitation (Papke *et al*., 2003; Whitaker *et al*., 2003; Cadillo-
51 Quiroz *et al.*, 2012). Terrestrial microbes fall somewhere between these extremes. For example,
52 soil dwelling microbes such as *Burkholderia pseudomallei*, *Burkholderia mallei*, and *Bacillus*
53 *anthracis* exhibit biogeographical patterns governed by dispersal limitation at regional spatial
54 scales (Kenefic *et al*., 2009; Pearson *et al*., 2009).
55
56 The phylogenetic resolution at which microbial diversity is defined can have a profound impact
57 on our ability to discern patterns of microbial biogeography (as reviewed Hanson *et al.*, 2012).
58 Surveys of SSU rRNA genes in terrestrial habitats indicate that environmental variables including
59 temperature (Fierer *et al*., 2009; Miller *et al*., 2009), pH (Fierer and Jackson, 2006; Lauber *et al*.,
60 2009; Rousk *et al*., 2010), and salinity (Lozupone and Knight, 2007) are more important than
61 geographic distance or latitude in determining spatial patterns of microbial diversity. However,
62 SSU rRNA gene sequences have an extremely low rate of nucleotide substitution (Ochman *et al*.,
63 1999), and microbes with similar or even identical SSU rRNA genes can have extensive genomic
64 and ecological diversity (Welch *et al*., 2002; Jaspers and Overmann; 2004). Thus, this marker has
65 low sensitivity for detecting neutral processes that drive patterns of biogeography, such as dispersal
66 limitation and genetic drift (Green and Bohannan, 2006; Hanson *et al*., 2012). However, these
67 neutral processes are readily explored using geographically and ecologically explicit culture
68 collections characterized at high genetic resolution.
69

Streptomyces are ubiquitous across soil habitats, and many species are easily cultured, making this genus an excellent candidate for making a taxon-specific survey of biogeography. Furthermore, *Streptomyces* species have high rates of gene exchange both within and between species (Doroghazi and Buckley, 2010). Hence, this genus is an ideal model to explore dispersal limitations and gene flow in terrestrial systems. *Streptomyces* are gram-positive *Actinobacteria* (Kâmpfer, 2006) known for their complex developmental cycle, which entails filamentous growth and the formation of desiccation resistant spores which are readily dispersed (Keiser *et al.*, 2000). *Streptomyces* play a significant role in the terrestrial carbon cycle (McCarthy and Williams, 1992; Takasuka *et al.*, 2013), represent important agricultural pathogens (Loria *et al.*, 2006; Labeda, 2011), and are prolific producers of antibiotics (Watve *et al.*, 2001). Despite their importance, we lack an evolutionary framework to understand *Streptomyces* biogeography.

Streptomyces diversity varies spatially, though the influence of geographic distance and ecological variation remains poorly resolved. Ecological adaptation constrains the environmental distribution of *Streptomyces* and their genetic and phenotypic diversity can vary in relation to soil characteristics at small spatial scale (1 m – 60 m) in prairie soils (Davelos *et al.*, 2004a and 2004b), and in dune habitats (Antony-Babu *et al.*, 2008). There is also evidence for dispersal limitation at very large (continental) spatial scales with endemic species observed in North America and Central Asia (Wawrik *et al.*, 2007). Remarkably, genetic analysis of *Streptomyces pratensis* has revealed that strains of this species are in linkage equilibrium (i.e. random association of alleles at different loci) across a range that spans 1,000 km (Doroghazi and Buckley, 2010 and 2014). One interpretation of this finding is that *Streptomyces* are subject to widespread dispersal and unlimited gene flow at regional spatial scales. However, linkage equilibrium can also be observed for

93 dispersal limited species that have undergone a recent historical demographic range expansion
94 (Doroghazi and Buckley, 2010 and 2014). Demographic range expansion has previously been
95 implicated as a factor that can explain ancestral patterns of gene exchange in *Streptomyces* clades
96 (Andam *et al*., 2016a). Combined, these data suggest a role for both adaptive and neutral processes
97 in governing the diversity and biogeography of *Streptomyces*, but the evolutionary interpretation
98 of these data depends on the degree to which dispersal limitation drives patterns of diversity in
99 *Streptomyces*.
100
101 We evaluated *Streptomyces* biogeography to explore the ecological and evolutionary forces that
102 govern diversification within this genus. The most powerful approach for detecting evolutionary
103 patterns that result from dispersal limitation is to examine taxon-specific biogeography patterns
104 across ecologically similar sites. This approach controls for the effects of selection and provides a
105 powerful test of neutral evolutionary processes (Hanson *et al.*, 2012). We constructed a taxon-
106 specific isolate collection of *Streptomyces* found in ecologically similar grassland sites spanning
107 the United States of America. In a study of *Streptomyces* species-level diversity, which evaluated
108 polymorphism at a single locus (*rpoB*), we observed evidence for dispersal limitation manifesting
109 in a latitudinal diversity gradient (Andam *et al.,* 2016b). The latitudinal diversity gradient is one
110 of the earliest and most well documented patterns of biogeography (Wallace, 1878; Hillebrand,
111 2004), but it is essentially undocumented in terrestrial bacterial systems. We hypothesize that
112 historical demographic phenomena associated with Pleistocene glaciation generated the latitudinal
113 gradient of *Streptomyces* species diversity (Andam *et al.*, 2016b). This hypothesis predicts that
114 dispersal and gene flow will be limited and discontinuous across latitudes which correspond to the
115 maximal extent of historical glaciation. The hypothesis also predicts that intra-species diversity

116 will vary with latitude as a consequence of limited time for diversification in northern latitudes.
117 These hypotheses are best addressed using population genetic approaches. Here we test these
118 specific predictions by using multilocus sequence analysis (MLSA) to characterize population-
119 level patterns of gene flow and genetic diversity within six phylogroups of *Streptomyces* which are
120 distributed widely across the United States.
121
122 **Results**
123 *Characterization of Streptomyces phylogroups*
124 We identified a total of 308 isolates representing the six targeted phylogroups, and these isolates
125 spanned 13 sites (Table 1). Strains within phylogroups share 99.4-100% SSU rRNA gene sequence
126 identity, and all strains share greater than 97% nucleotide identity in their SSU rRNA gene
127 sequences (Figure S1). Strains within phylogroups also share 97.6%-99.5% average nucleotide
128 identity (ANI) across concatenated MLSA loci, and all strains share greater than 88% ANI (Figure
129 1, Table 2). The pattern of genetic ancestry as determined by population structure analysis (see
130 Structure analysis, Experimental Procedures) is congruent with phylogroup boundaries (Figure
131 S2), and this indicates that these phylogroups approximate biological populations. We observed
132 122 unique MLSA haplotypes, with each phylogroup represented by 13-26 haplotypes (Figure 1,
133 Table 2). Good's coverage for each of the six phylogroups ranges from 0.88-1.0 for individual loci
134 and 0.94-1.0 for concatenated MLSA loci clustered at 99% nucleotide identity (Table S1). Hence,
135 allelic diversity and nucleotide diversity are well sampled (Table S1), though unique haplotypes
136 remain under sampled (Figure S3). The per site nucleotide diversity (π) of each phylogroup ranges
137 from 0.0026 to 0.011 (Table 2).
138

Four of the phylogroups do not match any *Streptomyces* species described in PubMLST (less than 97% ANI across MLSA loci, Figure S4A) (Jolley *et al*., 2004). While the other two, phylogroups WA1063 and MS152 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 S4A). 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 S4A).

We find evidence of horizontal gene transfer consistent with previous observations of *Streptomyces* species (Doroghazi and Buckley, 2010; Andam *et al.*, 2016a). There is significant phylogenetic incongruence between MLSA loci (Figure S4, Table S2), suggesting that inter-species horizontal gene transfer has shaped the phylogeny of these groups. The six phylogroups exhibit evidence of population structure with each phylogroup composed of 3.2 ± 0.8 (mean ± s.d.) subpopulations and with evidence for admixture (Figure 1). Evidence of admixture suggests horizontal gene transfer within phylogroups and is consistent with previous evidence of gene exchange in *Streptomyces* (Doroghazi and Buckley, 2010). Evidence of recombination within phylogroups MAN196, MS200, MS152, and F34 (PHI test, $p < 0.05$, Table 2) further supports the conclusion of gene exchange within populations. Furthermore, the standard index of association (I_A) is 0.09 for MAN125 suggesting a freely recombining population in linkage equilibrium (Table 2). In addition, phylogroups WA1063 and MS152 share two identical *atpD* alleles (Figure S4B), but 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. The latter explanation is more parsimonious given the low level of polymorphism between phylogroups

WA1063 and MS152. We do not observe evidence of inter-group horizontal gene transfer between these six phylogroups.

Evidence for 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 ± 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 (*i.e.* the ability of organisms to move freely across habitats). Each phylogroup subpopulation was observed in 2.2 ± 0.9 sites (mean ± 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* alleles indicates dispersal limitation. Identical alleles are shared among phylogroup members across each phylogroup's geographic range, which can exceed 5,000 km (Figure 2, Table 1). 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 S5). 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). In addition, all MLSA haplotypes (Figure 1, Figure 3) are unique to a single site, with the sole

185 exception being a haplotype from phylogroup MS200 which is observed in both MS and WI (Sun
186 Prairie).
187
188 Analysis of haplotype distribution is consistent with diversification due to dispersal limitation. We
189 used nested clade analysis (NCA) combined with a rigorous statistical framework to evaluate
190 population structure and demography (see Experimental Procedures). NCA establishes significant
191 phylogeographic inferences for phylogroups MAN196, MAN125, WA1063, and MS152 (Figure
192 3) but not for MS200 and F34. Nested clade phylogeographic inference postulates potential
193 evolutionary and historical demographic processes that support extant patterns of diversity and
194 biogeography (see legend of Figure 3). For instance, population subdivision of MAN125 across
195 the Pacific Northwest (Figure 3A) and MAN196 between Maine and the Pacific Northwest (Figure
196 3B) is consistent with restricted gene flow due to historical long distance dispersal events.
197 Likewise, population subdivision of MS152 and between the Southeast (MS and FL) and CA
198 (Figure 3C) and WA1063 between WI (Brookfield) and OR (Figure 3D) is consistent with
199 allopatric fragmentation.
200
201 *Latitudinal diversity gradient*
202 The distribution and diversity of the phylogroups reveals a latitudinal diversity gradient. Strains
203 from MAN125, MAN196, WA1063 occur mostly north of 40°N latitude, while strains from
204 MS200, MS152, F34 occur mostly south of this latitude (Table 1, Figure 2). This pattern of
205 North/South partitioning is significant for each phylogroup when evaluated against the expectation
206 of a random distribution across sites (permutation test, $p < 0.01$ for each phylogroup after
207 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 2). 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 4). This result indicates 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 across the United States (Table 1). The distribution of phylogroups is nonrandom and likely dispersal limited (Figure S5) with phylogroups inhabiting geographic ranges defined by latitude (Table 1, Figure 2). In addition, the genetic diversity of phylogroups is negatively correlated to the latitude from which they were isolated (Figure 4). 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. Furthermore, these results are consistent with the hypothesis that historical demographic processes have influenced the contemporary biogeography of *Streptomyces*.

231 The phylogroups we describe are coherent phylogenetic groups that approximate biological
232 populations (Figure S2). The members of each phylogroup share a distinguishable geographic
233 range (Table 1, Figure 2), a recent common ancestor (Figure 1), and greater than 97% ANI across
234 MLSA loci (Table 2). Despite geographic and genetic subpopulation structure (Figure 1),
235 phylogroup members frequently share identical alleles across demes. It is worth noting that since
236 all of the strains examined share > 97% SSU rRNA gene identity, these geographic patterns would
237 not be detected using standard SSU rRNA analyses methods (Figure S1). We infer the presence of
238 recombination within all phylogroups using both nucleotide polymorphism and phylogenetic
239 methods (Table 2, Table S2).
240
241 Regional patterns of biogeography among phylogroups are consistent with limitations to dispersal
242 and gene flow. Allopatric processes like genetic drift can drive diversification between populations
243 that are geographically isolated. The geographic distribution of our phylogroups is nonrandom,
244 and we find regional subpopulation structure within phylogroups (Figure 1, Figure 3). Although
245 we find identical alleles in sites thousands of kilometers apart (Figure 2, Figure S4), MLSA
246 haplotypes are not shared across sites (with the single exception of a haplotype shared between
247 MS and WI) (Figure 3). We also observe a significant distance decay relationship for MLSA allele
248 composition and geographic distance between sites (Figure S5). Distance decay relationships
249 driven by neutral processes can be challenging to identify given that environmental variables are
250 often spatially structured (Nekola and White, 1999). However, Andam *et al.* (2016b) shows that
251 *Streptomyces* community phylogenetic diversity across sites correlates significantly with latitude
252 and temperature but not with soil pH, soil organic matter, or rainfall. Hence, it is unlikely that these
253 coarse environmental variables determine regional population structure. This data implies that

254 while gene flow is moderate across the geographic range of a phylogroup, dispersal limitation and
255 genetic drift create appreciable regional population structure.
256
257 *Streptomyces* phylogroup diversity is consistent with a latitudinal diversity gradient. While the
258 classical description of a latitudinal diversity gradient defines diversity at the level of species
259 richness, these diversity gradients are also apparent at the level of intra-species genetic diversity
260 (Hadly, 2013). We find that latitude is a significant predictor of gene flow (Figure 2). Furthermore,
261 intra-phylogroup nucleotide diversity has a significant negative relationship with average latitude
262 (Figure 4), which is congruous with the latitudinal diversity gradient observed for diverse
263 macroorganisms (Hadly, 2013; Hillebrand, 2004). There is conflicting evidence for latitudinal
264 diversity gradients among microorganisms. Evidence for microbial latitudinal diversity gradients
265 comes from marine systems (Fuhrman *et al*., 2008; Sul *et al*., 2013; Swan *et al*., 2013), with
266 contrary evidence obtained in terrestrial systems (Neufeld and Mohn, 2005; Chu *et al*., 2010).
267 However, most analyses of terrestrial bacterial biogeography are derived from analyses of SSU
268 rRNA genes, and we show that analyses of SSU rRNA genes lack the sensitivity needed to detect
269 the biogeographic patterns that we observe for *Streptomyces* (Figure S1).
270
271 Several hypotheses have been advanced to explain the formation of latitudinal diversity gradients
272 (Wiens and Donoghue, 2004; Mittelbach *et al*., 2007). Ecological hypotheses posit that factors
273 such as carrying capacity, productivity, and niche availability vary across latitude and that these
274 factors impose constraints on biodiversity (Currie *et al.*, 2004; Mouchet *et al.*, 2015). Evolutionary
275 hypotheses invoke the positive relationship between temperature and the kinetics of metabolism
276 to predict that evolutionary rates and cladogenesis correspond with temperature (Allen *et al*.,

277 2002). Historical hypotheses propose that the latitudinal diversity gradient is the product of
278 historical geological, ecological, or demographic events that have influenced dispersal and
279 diversification (Wiens and Donoghue, 2004; Stevens, 2006). For example, the influence of
280 Pleistocene glacial events on the biogeography of diverse species of terrestrial and aquatic plants
281 and animals is well documented, with postglacial range expansion giving rise to the latitudinal
282 partitioning of populations and species, and resulting in decreased molecular variation of northern
283 lineages relative to southern ancestral lineages (Soltis *et al*., 1997; Bernatch and Wilson, 1998;
284 Conroy and Cook, 2000; Milá *et al*., 2006; Maggs *et al*., 2008; Wilson and Veraguth, 2010). There
285 is also evidence that Pleistocene glaciation events have impacted both microbial communities
286 (Eisenlord *et al*., 2012) and populations (Kenefic *et al.*, 2009; Mikheyev *et al*., 2008).
287
288 The biogeography of our *Streptomyces* phylogroups is consistent with the hypothesis that historical
289 demography and dispersal limitation have produced the latitudinal diversity gradient that we
290 observe. For instance, MAN125 is nearly in linkage equilibrium (I_A = 0.09) though its members
291 span a geographic range of over 2,000 km across the Pacific Northwest. Similar patterns of
292 recombination have also been observed in a *S. pratensis* population which spanned 1,000 km
293 across sites present in North Carolina and northern New York (Doroghazi and Buckley, 2010 and
294 2014). The observation of linkage equilibrium indicates that either the population lacks
295 contemporary barriers to gene flow or the population has experienced a recent historical
296 demographic expansion. Coupled with evidence of limitations to gene flow at regional scales
297 (Figure 2, Figure S5), the most parsimonious explanation for these conflicting observations is a
298 recent historical demographic range expansion. Conversely, latitudinal gradients in marine
299 bacterial diversity have been attributed to correlations between temperature and the kinetics of

metabolism (Fuhrman *et al*., 2008). This kinetic effect has been hypothesized to increase evolutionary tempo and speciation rates in tropical latitudes and would be expected to generate latitudinal gradients of both species diversity and nucleotide diversity (Allen *et al*., 2002). Since latitude is correlated with temperature, we cannot completely dismiss the influence of kinetics as a cause of the intra-group latitudinal gradient of genetic diversity in terrestrial *Streptomyces*. However, the kinetic effects of temperature cannot account for partitioning of gene flow across latitude, while in contrast, this is a specific prediction of the historical demography hypothesis. It is possible, however, that unappreciated ecological variables, such as the species composition of perennial grass communities, could shape the diversity gradient. Ultimately, it is likely that latitudinal diversity gradients can arise from a combination of ecological, evolutionary, and historical processes that vary in their relative influence with respect to different species and different habitats.

Evidence for a latitudinal diversity gradient coupled with evidence of contemporary dispersal limitation following historical demographic expansion, while not conclusive, suggests that phylogenetic niche conservatism has contributed to the formation of the *Streptomyces* latitudinal diversity gradient. Phylogenetic niche conservatism can cause diversity gradients when contemporary species distributions are determined by historical climate regimes (Wiens and Donoghue, 2004; Stevens, 2006). Climactic regimes oscillate widely across geologic time scales and these historical changes in climate produce demographic phenomena that impact the evolutionary dynamics of species. In particular, the genetic consequences of glaciation events are described in depth by Hewitt (1996, 2000, and 2004). The population structure we observe in our *Streptomyces* phylogroups is consistent with the effects of post-glacial demographic range

323 expansion followed by dispersal limitation and regional diversification. Dispersal limitation may
324 occur following range expansion as a result of density dependent blocking or by adaptive barriers
325 that arise after the colonization of new habitat. One of the expectations of post-glacial expansion
326 is "southern richness versus northern purity" (Hewitt, 2004). This is evident in the negative
327 correlation we observe between latitude and the nucleotide diversity of phylogroups (Figure 4).
328 Similar relationships between intraspecific nucleotide diversity and average latitude as result of
329 post-glacial colonization are evident in other systems (Bernatchez and Wilson, 1998). Williams *et*
330 *al*. (1998) justifies 40˚N latitude as approximating late Pleistocene glacial and non-glacial regions
331 with respect to species distributions in North America. Hence, the latitudinal delineation of allele
332 distributions for *Streptomyces* phylogroups roughly corresponds to the extent of ice coverage
333 during the late Pleistocene (Figure 2), which suggests historical population expansion from lower
334 to higher latitudes.
335
336 Haplotype distributions of phylogroups MAN125, MAN196, WA1063, and MS152 are consistent
337 with allopatric diversification resulting from dispersal limitation (Figure 3). Haplotype nested
338 clade analysis (NCA) predicts that historical dispersal events and range expansion across northern
339 regions resulted in limits on dispersal during intermediate timescales, allowing genetic drift to
340 create the phylogroup population structures observed today. There is moderate criticism of NCA
341 (Knowles and Maddison, 2002; Nielsen and Beumont, 2009) due to the subjective nature of
342 inferring historical processes and the wide potential for stochastic processes creating similar
343 patterns of biogeography. Yet these tools can provide useful hypotheses. Northern phylogroups
344 MAN125 and MAN196 share a common ancestor with southern phylogroups MS200 and F34
345 (Figure 1). The contemporary population structure of MAN125 and MAN196 is consistent with a

346 historical range expansion from a common ancestor shared by both clades (Figure 1, Figure 3).
347 Further population structure within each phylogroup likely resulted from barriers to gene flow and
348 historical dispersal events across the Pacific Northwest for MAN125 (Figure 3A) and between
349 Maine and the Pacific Northwest for MAN196 (Figure 3B). Analysis of haplotype distribution
350 within WA1063 and MS152 is also consistent with diversification of populations as a result of
351 gene flow limitation between the Midwest and Pacific Northwest for WA1063 (Figure 3D) and
352 between the Southeast and West for MS152 (Figure 3C).
353
354 Phylogroups WA1063 and MS152 share a recent common ancestor (Figure 1, Figure S2) and also
355 share two identical alleles at the *atpD* locus (Figure S4B). These phylogroups have distinct, non-
356 overlapping geographic ranges with WA1063 found in higher latitudes and MS152 in lower
357 latitudes (Table 1, Figure 2). WA1063 and MS152 have 0.0205 net nucleotide substitutions per
358 site across concatenated MLSA loci. We evaluate the possible time range for divergence between
359 WA1063 and MS152 by extrapolating very roughly from the nucleotide substitution rate (μ =
360 4.5×10^{-9}) and generation time (100-300 generations per year) for *E. coli* (Ochman *et al*., 1999),
361 since corresponding values are not available for *Streptomyces* or their relatives. Based upon these
362 gross approximations, we would estimate that WA1063 and MS152 diverged 15,000-50,000 years
363 ago, corresponding to events in the late Pleistocene (Clayton *et al*., 2006). Hence, it is likely that
364 the identical *atpD* alleles found in both WA1063 and MS152 were inherited from a shared
365 ancestral population (Figure S3B) as opposed to inheritance from contemporary gene exchange.
366
367 Through population genetic analysis of six *Streptomyces* phylogroups, we find evidence for
368 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. Future research should seek to determine the degree to which dispersal limitation is due to limits on spore mobility, density dependent blocking, or a result of adaptive constraints relating to phylogenetic niche conservatism. A better understanding of *Streptomyces* biogeography and the evolutionary forces that govern *Streptomyces* diversification may ultimately assist in the discovery of novel genetic diversity and possibly novel antibiotics within this genus.

Experimental Procedures

Strain isolation and DNA extraction

We previously 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.,* 2016b). Sites were selected to represent a narrow range of ecological characteristics including meadow, pasture, or native grasslands dominated by perennial grasses and having moderately acidic soil (pH: 6.0 ± 1.0; ave. ± s.d.). Strains were isolated using uniform conditions and this will select for strains having similar physiological characteristics. The analysis of physiologically similar strains from ecologically similar sites improves our ability to detect biogeographical patterns that result from drift by minimizing the importance of selection (as reviewed by Hanson *et al.*, 2012). 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). Genomic DNA was extracted for each isolate 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 (Roberts and Crawford, 2000).

The gene encoding the RNA polymerase beta-subunit (*rpoB*) provides a robust, advantageous alternative to the SSU rRNA locus for phylogenetic analyses of the genera *Streptomyces* (Kim *et al.*, 2004). We previously assessed genetic diversity of our culture collection using partial *rpoB* sequences clustered at 0.01 patristic distances with RAMI (Pommier *et al.*, 2009), and using this approach we identified 107 species-like phylogenetic clusters, or phylogroups (Andam *et al*., 2016b). We selected six of these phylogroups for further analysis. These six phylogroups had the highest numerical abundance and widest geographical distribution in our isolate collection, representing 308 strains isolated across 13 sites (and representing 308 of the 755 strains isolated from these 13 sites). 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 *atpD*, *gyrB*, *recA*, *rpoB*, and *trpB* as described in Doroghazi and Buckley (2010) (Table S3). The V2 and V2 regions of SSU rRNA sequences were amplified using universal primers 8F and 1492R (Table S3). Reactions for Sanger sequencing were performed using forward

415 primers for all loci except *rpoB*, for which the reverse primer was used. Trace files were uniformly
416 screened using CAP3 (Huang and Madan, 1999), and sequences were trimmed as to discard
417 nucleobases with a Phred quality score below 23. Sequences were aligned using MUSCLE (Edgar,
418 2004), trimmed to 431 bp, 415 bp, 446 bp, 557 bp, and 489 bp, for genes *atpD*, *gyrB*, *recA*, *rpoB*,
419 *trpB*, respectively, and concatenated consistently with the genomic order in *Streptomyces*
420 *coelicolor* A3(2). SSU rRNA sequences were trimmed to 357 bp, creating an alignment spanning
421 the V1 and V2 regions. Gene sequences are available on GenBank with accession numbers
422 KX110408-KA111380.
423
424 Good's coverage estimation and haplotype rarefaction was determined using mothur (Schloss *et*
425 *al.*, 2009). DNA polymorphism statistics including, number of segregating sites, nucleotide
426 diversity, and Tajima's D, were determined with DnaSP v5 (Librado and Rozas, 2009) and LDhat
427 (McVean *et al.*, 2002). Population scaled mutation rates (Watterson's theta; $\theta_w = 2N_e\mu$),
428 recombination or gene conversion rates ($\rho = 2N_e r/2$), and relative rates of recombination (ρ/θ_w) were
429 estimated using LDhat (McVean *et al.*, 2002) and are expressed per nucleotide site. The standard
430 index of association (I_A) was calculated from allelic data with LIAN v3.5 (Haubold and Hudson,
431 2000) using the Monte-Carlo test and 100 iterations. The pairwise homoplasy index (PHI) statistic
432 was determined using PhiPack (Bruen *et al.*, 2006), and statistical significance was evaluated
433 under a null hypothesis of no recombination. Sequence identity across phylogroups was calculated
434 with mothur (Shloss *et al.*, 2009)
435
436 *Phylogenetic reconstruction*
437 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 $1.0E^6$ and $3.0E^6$ replicates. The most probable number of sub-populations (k) was evaluated with 10 independent runs and chosen using the Evanno method (Evanno *et al*., 2005), 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. The null model for the

permutation test assigned strains to sites as a random draw without replacement from the OTU table while holding the number of strains sampled at each site to be invariant. 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 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).

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.

483 Adams RI and Hadly EA. (2013) Genetic diversity within vertebrate species is greater at lower
484 latitudes. Evol Ecol **27**: 113-143.
485
486 Andam CP, Choudoir MJ, Nguygen AV, Park HS, Buckley DH. (2016a) Contributions of
487 ancestral inter-species recombination to the genetic diversity of extant *Streptomyces* lineages.
488 ISMEJ In Press. doi: 10.1038/ismej.2015.230.
489
490 Andam CP, Doroghazi JR, Campbell AN, Kelly PJ, Choudoir MJ, Buckley DH. (2016b). A
491 latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*. mBio **7**: e02200-
492 15.
493
494 Antony-Babu S, Stach JEM, Goodfellow M. (2008) Genetic and phenotypic evidence for
495 *Streptomyces griseus* ecovars isolated from a beach and dune sand system. Antonie Van
496 Leeuwenhoek **94**: 63–74.
497
498 Bell T. (2010) Experimental tests of the bacterial distance-decay relationship. ISME J **4**: 1357–
499 1365.
500
501 Bernatchez L, Wilson CC. (1998) Comparative phylogeography of Nearctic and Palearctic
502 fishes. Mol Ecol **7**: 431–452.
503

504 Bissett A, Richardson AE, Baker G, Wakelin S, Thrall PH. (2010) Life history determines
505 biogeographical patterns of soil bacterial communities over multiple spatial scales. Mol Ecol **19**:
506 4315–4327.
507
508 Boucher Y, Cordero OX, Takemura A, Hunt DE, Schliep K, Bapteste E, *et al*. (2011) Local
509 mobile gene pools rapidly cross species boundaries to create endemicity within global *Vibrio*
510 *cholerae* populations. mBio **2**: e00335–10.
511
512 Bruen TC, Philippe H, Bryant D. (2005) A simple and robust statistical test for detecting the
513 presence of recombination. Genetics **172**: 2665–2681.
514
515 Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, *et al.* (2012) Patterns of
516 gene flow define species of thermophilic Archaea. PLoS Biol **10**: e1001265.
517
518 Cho J, Tiedje JM. (2000) Biogeography and degree of endemicity of fluorescent *Pseudomonas*
519 strains in soil. Appl Environ Microbiol **66**: 5448-5456.
520
521 Choudoir MJ, Campbell AN, Buckley DH. (2012) Grappling with Proteus: population level
522 approaches to understanding microbial diversity. Front Microbiol **3**: 336.
523
524 Chu H, Fierer N, Lauber CL, Caporaso JG, Knight R, Grogan P. (2010) Soil bacterial diversity in
525 the Arctic is not fundamentally different from that found in other biomes. Environ Microbiol **12**:
526 2998–3006.

Clayton L, Attig JW, Mickelson DM, Johnson MD, Syverson KM. (2006) Glaciation of Wisconsin: Wisconsin Geological and Natural History Survey Educational Series 36. Wisconsin Geological and Natural History Survey. Madison, WI.

Clement M, Posada D, Crandall KA. (2000) TCS: A computer program to estimate gene genealogies. Mol Ecol **9**: 1657–1659.

Coleman ML, Chisholm SW. (2010) Ecosystem-specific selection pressures revealed through comparative population genomics. Proc Natl Acad Sci U S A **107**: 18634–18639.

Conroy CJ, Cook JA. (2000) Phylogeography of a post-glacial colonizer: *Microtus longicaudus* (Rodentia: Muridae). Mol Ecol **9**: 165–175.

Currie DJ, Mittelbach GG, Cornell HV, Field R, Guégan JF, Hawkins BA*, et al.* (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. Ecol Lett **7**: 1121–1134.

Davelos AL, Xiao K, Samac DA, Martin AP, Kinkel LL. (2004a) Spatial variation in *Streptomyces* genetic composition and diversity in a prairie soil. Microb Ecol **48**: 601–612.

548 Davelos AL, Kinkel LL, Samac DA. (2004b) Spatial variation in frequency and intensity of
549 antibiotic interactions among *Streptomycetes* from prairie soil. Appl Environ Microbiol **70**:
550 1051–1058.
551
552 Doroghazi JR, Buckley DH. (2014) Intraspecies comparison of *Streptomyces pratensis* genomes
553 reveals high levels of recombination and gene conservation between strains of disparate
554 geographic origin. BMC Genomics **15**: 970.
555
556 Doroghazi JR, Buckley DH. (2010) Widespread homologous recombination within and between
557 *Streptomyces* species. ISME J **4**: 1136–1143.
558
559 Earl DA, vonHoldt BM. (2012) STRUCTURE HARVESTER: a website and program for
560 visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour
561 **4**: 359–361.
562
563 Edgar RC. (2004) MUSCLE: multiple sequence alignment with high accuracy and high
564 throughput. Nucleic Acid Res **32**: 1792–1797.
565
566 Eisenlord SD, Zak DR, Upchurch RA. (2012) Dispersal limitation and the assembly of soil
567 *Actinobacteria* communities in a long-term chronosequence. Ecol Evol **2**: 538–549.
568
569 El-Nakeeb MA, Lechevalier HA. (1963) Selective isolation of aerobic Actinomycetes. Appl
570 Microbiol 11: 75–77.

Evanno G, Regnaut S, Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol **14**: 2611–2620.

Fierer N, Carney KM, Horner-Devine MC, Megonigal JP. (2009) The biogeography of ammonia-oxidizing bacterial communities in soil. Microb Ecol **58**: 435–445.

Fierer N, Jackson RB. (2006) The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A **103**: 626–631.

Finlay BJ. (2002) Global dispersal of free-living microbial eukaryote species. Science **296**: 1061–1063.

Finlay BJ, Fenchel T. (2004) Cosmopolitan metapopulations of free-living microbial eukaryotes. Protist **155**: 237–244.

Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown M V, Green JL, *et al.* (2008) A latitudinal diversity gradient in planktonic marine bacteria. Proc Natl Acad Sci U S A **105**: 7774–7778.

Gilbert JA, Steele JA, Caporaso JG, Steinbrück L, Reeder J, Temperton B, *et al.* (2012) Defining seasonal marine microbial community dynamics. ISME J **6**: 298–308.

Goslee SC, Urban DL. (2007) The ecodist package for dissimilarity-based analysis of ecological data. J Stat Softw **22**: 1–19.

Green J, Bohannan BJM. (2006) Spatial scaling of microbial biodiversity. Trends Ecol Evol **21**: 501–507.

Guo Y, Zheng W, Rong X, Huang Y. (2008) A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics. Int J Syst Evol Microbiol **58**: 149–159.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol **10**: 497–506.

Hatosy SM, Martiny JBH, Sachdeva R, Steele J, Fuhrman JA, Martiny AC. (2013) Beta diversity of marine bacteria depends on temporal scale. Ecology **94**: 1898–1904.

Haubold B, Hudson RR. (2000) LIAN 3.0: detecting linkage disequilibrium in multilocus data. Bioinformatics **16**: 847–848.

Hewitt GM. (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biol J Linnean Soc **58**: 247–276.

Hewitt G. (2000) The genetic legacy of the Quaternary ice ages. Nature **405**: 907–913.

Hewitt GM. (2004) Genetic consequences of climatic oscillations in the Quaternary. Philos Trans R Soc Lond B Biol Sci **359**: 183–195.

Hillebrand H. (2014) On the generality of the latitudinal diversity gradient. Amer Nat **163**: 192–211.

Huang X, Madan A. (1999) CAP3: a DNA sequence assembly program. Genome Res **9**: 868–677.

Jakobsson M, Rosenberg NA. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics **23**: 1801–1806.

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

Jolley KA, Chan MS, Maiden MCJ. (2004) mlstdbNet - distributed multi-locus sequence typing (MLST) databases. BMC Bioinformatics **5**: 86.

638 Kâmpfer P. (2006) The family *Streptomycetaceae* Part I: Taxonomy. In: The Prokaryotes,
639 Volume 3. Dworkin, M, Falkow, S, Rosenberg, E, Schliefer, K, & Stackebrandt, E. (eds). New
640 York: Springer, pp. 538–604.
641
642 Kenefic LJ, Pearson T, Okinaka RT, Schupp JM, Wagner DM, Hoffmaster AR, *et al.* (2009) Pre-
643 Columbian origins for North American anthrax. PLoS One **4**: e4813.
644
645 Kieser T, Bibb M, Buttner M, Charter K, Hopwood D. (2000) Practical *Streptomyces* Genetics.
646 Norwich, UK: John Innes Foundation.
647
648 Kim BJ, Kim CJ, Chun J, Koh YH, Lee SH, Hyun JW, *et al.*(2004) Phylogenetic analysis of the
649 genera *Streptomyces* and *Kitasatospora* based on partial RNA polymerase beta-subunit gene
650 (*rpoB*) sequences. Int J Syst Evol Microbiol. **54**: 593–598.
651
652 Knowles L, Maddison W. (2002) Statistical phylogeography. Mol Ecol **11**: 2623–2635.
653
654 Labeda DP. (2011) Multilocus sequence analysis of phytopathogenic species of the genus
655 *Streptomyces*. Int J Syst Evol Microbiol **61**: 2525–2531.
656
657 Lauber CL, Hamady M, Knight R, Fierer N. (2009) Pyrosequencing-based assessment of soil pH
658 as a predictor of soil bacterial community structure at the continental scale. Appl Environ
659 Microbiol **75**: 5111–5120.
660

661 Librado P, Rozas J. (2009) DnaSP v5: a software for comprehensive analysis of DNA
662 polymorphism data. Bioinformatics **25**: 1451–1452.
663
664 Loria R, Kers J, Joshi M. (2006) Evolution of plant pathogenicity in *Streptomyces*. Annu Rev
665 Phytopathol **44**: 469–487.
666
667 Lozupone CA, Knight R. (2007) Global patterns in bacterial diversity. Proc Natl Acad Sci U S A
668 **104**: 11436–11440.
669
670 Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, *et al.* (2008) Evaluating
671 signatures of glacial refugia for North Atlantic benthic marine taxa. Ecology **89**: S108–S122.
672
673 Mantel N. (1967) The detection of disease clustering and a generalized regression approach.
674 Cancer Research **27**: 209–220.
675
676 Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, *et al.* (2006)
677 Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol **4**: 102–112.
678
679 Martiny JBH, Eisen JA, Penn K, Allison SD, Horner-Devine MC. (2011) Drivers of bacterial β -
680 diversity depend on spatial scale. Proc Natl Acad Sci U S A **108**: 7850–7854.
681
682 McCarthy AJ, Williams ST. (1992) Actinomycetes as agents of biodegradation in the
683 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.

Miller SR, Strong AL, Jones KL, Ungerer MC. (2009) Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. Appl Environ Microbiol **75**: 4565–4572.

Mittelbach GG, Schemske DW, Cornell H V., Allen AP, Brown JM, Bush MB, *et al.* (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecol Lett **10**: 315–331.

Mouchet M, Levers C, Zupan L, Kuemmerle T, Plutzar C, Erb K, *et al.* (2015) Testing the effectiveness of environmental variables to explain European terrestrial vertebrate species richness across biogeographical scales. PLoS One **10**: e0131924.

706 Nekola JC and White PS. (1999) The distance decay similarity in biogeography and ecology. J
707 Biogeogr **26**: 867-878.
708
709 Neufeld JD, Mohn WW. (2005) Unexpectedly high bacterial diversity in arctic tundra relative to
710 boreal forest soils, revealed by serial analysis of ribosomal sequence tags. Appl Environ
711 Microbiol **71**: 5710–5718.
712
713 Nielsen R, Beaumont MA. (2009) Statistical inferences in phylogeography. Mol Ecol **18**: 1034–
714 1047.
715
716 Ochman H, Elwyn S, Moran NA. (1999) Calibrating bacterial evolution. Proc Natl Acad Sci U S
717 A **96**: 12638–12643.
718
719 Ottow JCG. (2013) Rose bengal as a selective aid in the isolation of fungi and Actinomycetes
720 from natural sources. Mycologia **64**: 304–315.
721
722 Panchal M. (2006) The automation of Nested Clade Phylogeographic Analysis. Bioinformatics
723 **23**: 509–510.
724
725 Papke RT, Ramsing NB, Bateson MM, Ward DM. (2003) Geographical isolation in hot spring
726 cyanobacteria. Environ Microbiol **5**: 650–659.
727

728 Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. (2010) How
729 many bootstrap replicates are necessary? J Comput Biol **17**: 337–354.
730
731 Pearson T, Giffard P, Beckstrom-Sternberg S, Auerbach R, Hornstra H, Tuanyok A*, et al*. (2009)
732 Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer.
733 BMC Biol **7**: 78.
734
735 Pommier T, Canback B, Lundberg P, Hagstrom A, Tunlid A. (2009) RAMI: a tool for
736 identification and characterization of phylogenetic clusters in microbial communities.
737 Bioinformatics **25**: 736–742.
738
739 Posada D, Crandall KA, Templeton AR. (2000) GeoDis: A program for the cladistic nested
740 analysis of the geographical distribution of genetic haplotypes. Mol Ecol **9**: 487–488.
741
742 Pritchard JK, Stephens M, Donnelly P. (2000) Inference of population structure using multilocus
743 genotype data. Genetics **155**: 945–959.
744
745 Roberts MA and Crawford DL. (2000) Use of randomly amplified polymorphic DNA as a means
746 of developing genus- and strain-specific *Streptomyces* DNA probes. Appl Environ Microbiol **66:**
747 2555-2564.
748

749 Rong X, Huang Y. (2010) Taxonomic evaluation of the *Streptomyces griseus* clade using
750 multilocus sequence analysis and DNA-DNA hybridization, with proposal to combine 29 species
751 and three subspecies as 11 genomic species. Int J Syst Evol Microbiol **60**: 696–703.
752
753 Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, *et al*. (2010) Soil
754 bacterial and fungal communities across a pH gradient in an arable soil. ISME J **4**: 1340–1351.
755
756 Schliep KP. (2011) phangorn: phylogenetic analysis in R. Bioinformatics **27**: 592–593.
757
758 Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, *et al*. (2009)
759 Introducing mothur: open-source, platform-independent, community-supported software for
760 describing and comparing microbial communities. Appl Environ Microbiol **75**: 7537–7541.
761
762 Shimodaira H, Hasegawa M. (1999) Multiple comparisons of log-likelihoods with applications
763 to phylogenetic inference. Mol Biol Evol **16**: 1114–1116.
764
765 Smouse PE, Long JC, Sokal RR. (1986) Multiple regression and correlation extensions of the
766 Mantel test of matrix correspondence. Syst Zool **35**: 627–632.
767
768 Soltis D, Gitzendanner M, Strenge D, Soltis P. (1997) Chloroplast DNA intraspecific
769 phylogeography of plants from the Pacific Northwest of North America. Plant Syst Evol **206**:
770 353–373.
771

772 Stamatakis A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
773 thousands of taxa and mixed models. Bioinformatics **22**: 2688–2690.
774
775 Stevens RD. (2006) Historical processes enhance patterns of diversity along latitudinal gradients.
776 *Proc R Soc B Biol Sci* **273**: 2283–2289.
777
778 Sul WJ, Oliver TA, Ducklow HW, Amaral-zettler LA, Sogin ML. (2012) Marine bacteria exhibit
779 a bipolar distribution. Proc Natl Acad Sci U S A **110**: 2342–2347.
780
781 Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM, *et al*. (2013)
782 Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface
783 ocean. Proc Natl Acad Sci U S A **110**: 11463–11468.
784
785 Takasuka TE, Book AJ, Lewin GR, Currie CR, Fox BG. (2013) Aerobic deconstruction of
786 cellulosic biomass by an insect-associated *Streptomyces*. Sci Rep **3**: 1–10.
787
788 Tavaré S. (1986) Some probabilistic and statistical problems in the analysis of DNA sequences.
789 Am Math Soc Lect Math Life Sci **17**: 57–86.
790
791 Telford RJ, Vandvik V, Birks HJB. (2006) Dispersal limitations matter for microbial
792 morphospecies. Science **312**: 1015.
793

794 Templeton AR, Boerwinkle E, Sing CF. (1987) A cladistic analysis of phenotypic associations
795 with haplotypes inferred from restriction endonuclease mapping. i. basic theory and an analysis
796 of alcohol dehydrogenase activity in Drosophila. Genetics **117**: 343–351.
797
798 Templeton AR, Routman E, Phillips CA. (1995) Separating population structure from population
799 history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in
800 the tiger salamander, *Ambystoma tigrinum*. Genetics **140**: 767–782.
801
802 Vos M, Velicer GJ. (2008) Isolation by distance in the spore-forming soil bacterium
803 *Myxococcus xanthus*. Curr Biol **18**: 386–391.
804
805 Wallace A. (1878) Tropical nature, and other essays. London: Macmillan and Co.
806
807 Watve M, Tickoo R, Jog M, Bhole B. (2001) How many antibiotics are produced by the genus
808 *Streptomyces*? Arch Microbiol **176**: 386–390.
809
810 Wawrik B, Kutliev D, Abdivasievna UA, Kukor JJ, Zylstra GJ, Kerkhof L. (2007) Biogeography
811 of Actinomycete communities and type ii polyketide synthase genes in soils collected in New
812 Jersey and Central Asia. Appl Environ Microbiol **73**: 2982–2989.
813
814 Welch RA, Burland V, Plunkett G, Redford P, Roesch P, Rasko D, *et al*. (2002) Extensive
815 mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*.
816 Proc Natl Acad Sci U S A **99**: 17020–17024.

817

818 Whitaker RJ, Grogan DW, Taylor JW. (2003) Geographic barriers isolate endemic populations
819 of hyperthermophilic archaea. Science **301**: 976–978.

820

821 Wiens JJ, Donoghue MJ. (2004) Historical biogeography, ecology and species richness. Trends
822 Ecol Evol **19**: 639–644.

823

824 Williams D, Dunkerley D, DeDecker P, Kershaw P, Chappell M. (1998) Quaternary
825 environments. London: Arnold.

826

827 Wilson AB, Veraguth IE. (2010) The impact of Pleistocene glaciation across the range of a
828 widespread European coastal species. Mol Ecol **19**: 4535–53.

Table and Figure Legends

Table 1. A total of 755 *Streptomyces* strains were isolated from 13 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 sample site.

Site	Code	Latitude	Longitude	Altitude	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	452 m	95	51	26	1			
Denali Hwy, AK	den	63.22˚N	-147.68˚W	894 m	40	40					
Bothell, WA	wa	47.73˚N	-122.24˚W	132 m	105	21	5	21			
Astoria, OR	or	46.18˚N	-123.85˚W	40 m	79		8	7			
Kennebunk, ME	me	43.4˚N	-70.54˚W	28 m	85		13				
Sun Prairie, WI	sun	43.17˚N	-89.24˚W	289 m	65	3			3		
Brookfield, WI	b	43.06˚N	-88.13˚W	277 m	22			5			
Palo Alto, CA	st	37.43˚N	-122.17˚W	24 m	34					2	1
Greensboro, NC	gb	36.09˚N	-79.89˚W	276 m	9		1			2	
Troy, NC	uw	35.71˚N	-79.88˚W	194 m	19		4		7		
Starkville, MS	ms	33.46˚N	-88.8˚W	119 m	90				11	8	2
Austin, TX	t	30.2˚N	-97.77˚W	199 m	42						6
Fort Pierce, FL	f	27.54˚N	-80.35˚W	0 m	70				7	5	48
Total					755	115	57	34	28	17	57

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.

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

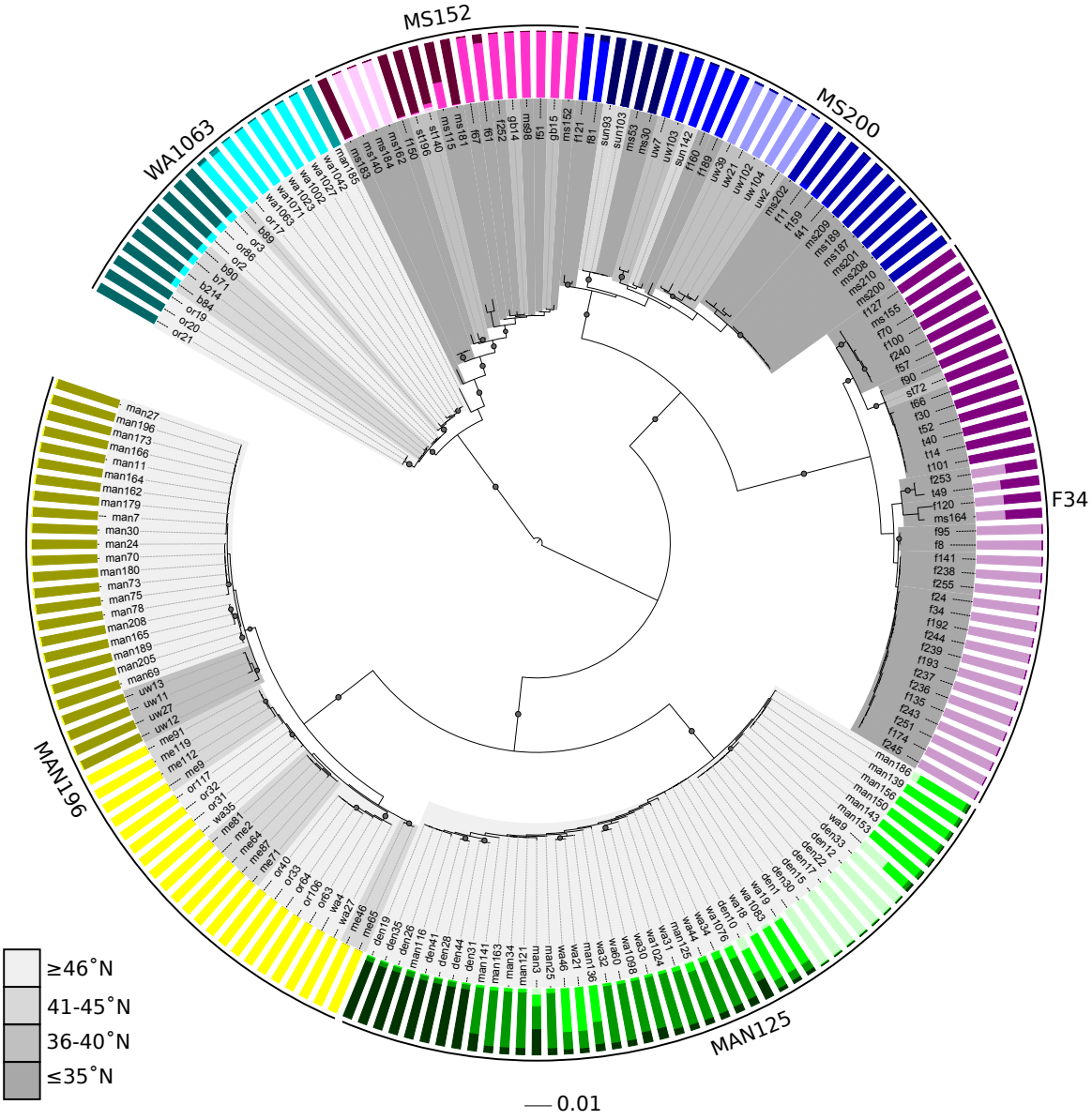

Figure 1. Tree was constructed from concatenated MLSA loci nucleotide sequences using maximum likelihood with a GTRGAMMA evolution model. Scale bar represents nucleotide substitutions per site. The root was defined by *Mycobacterium smegmatis*. Nodes with bootstrap confidences > 80 are indicated with gray circles, and precise bootstrap values are found in Figure S4A. 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.

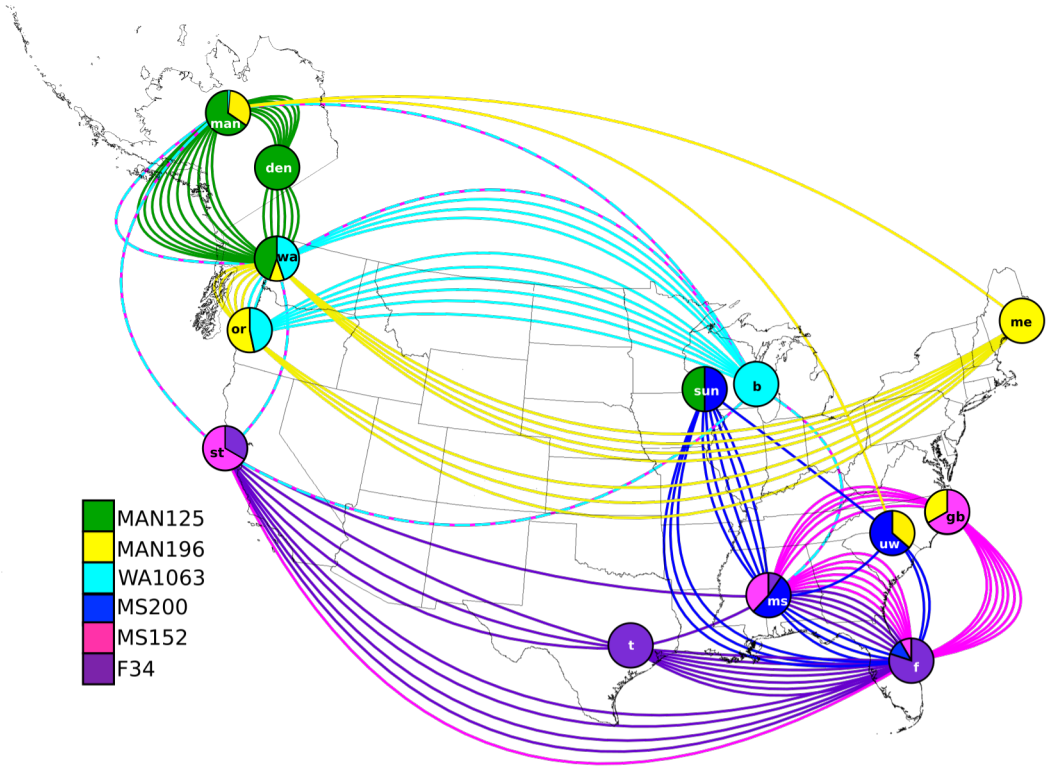

Figure 2. Circles depict sample sites and are labeled according to site code (Table 1). The relative abundance of each phylogroup at each site is indicated by color according to the legend, with raw counts provided in Table 1. Solid colored lines represent identical alleles shared by phylogroup members subsampled for MLSA across sites. Dashed multicolored lines depict identical *atpD* alleles shared by strains in phylogroups WA1063 and MS152 (Figure S4B).

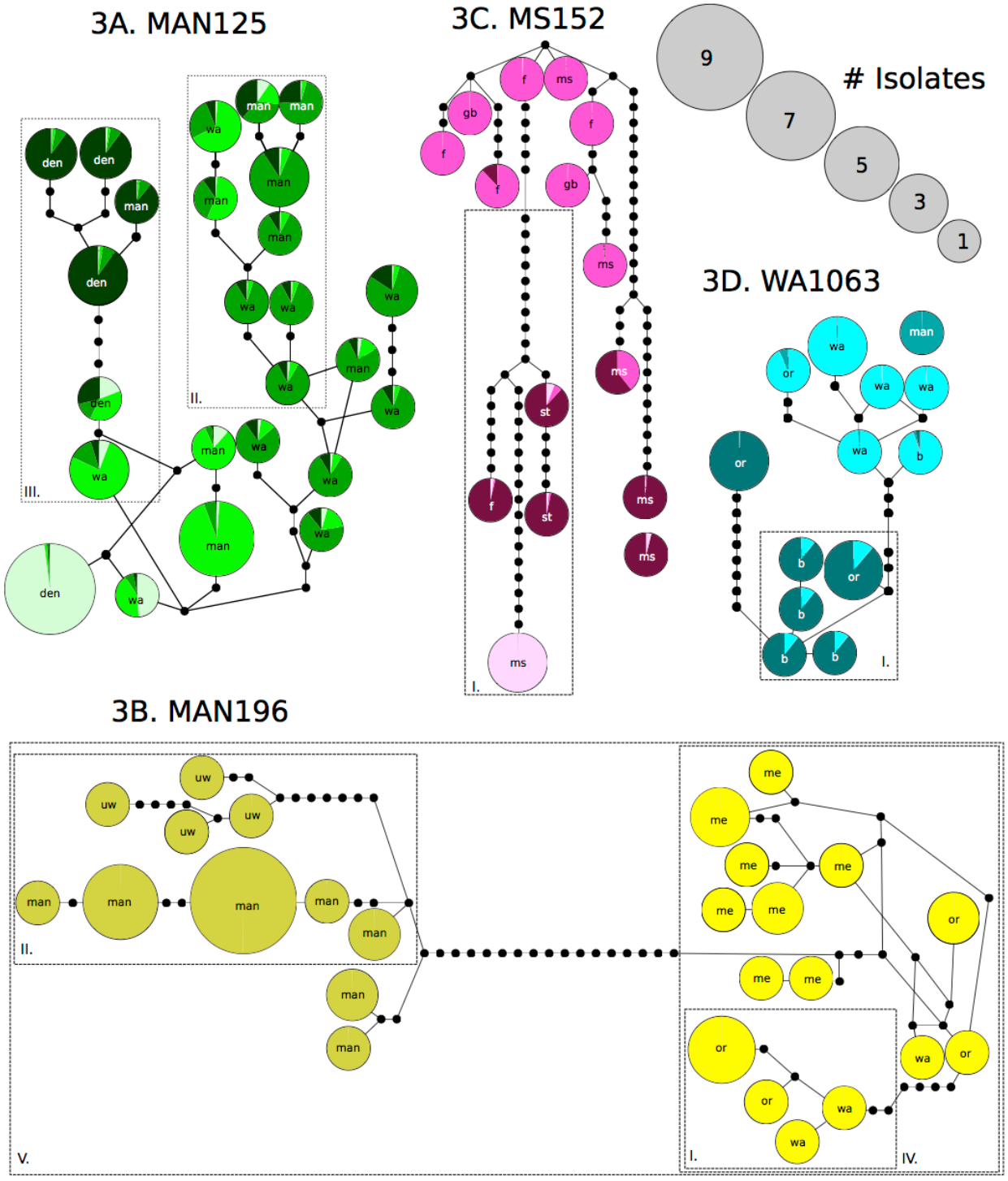

Figure 3. Haplotype networks for phylogroups MAN125 (3A), MAN196 (3B), MS152 (3C), WA1063 (3D). Circles represent MLSA haplotypes whose radius is proportional to the number of isolates having that haplotype, and colors correspond to strain ancestry and subpopulation affiliation as determined by Structure analysis and defined in Figure 1. Haplotypes are labeled with a letter code referring to sample site location as indicated in Table 1 (all haplotypes for MAN125,

MAN196, MS152, and WA1063 are found exclusively at a single site). Black circles represent unsampled, 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.

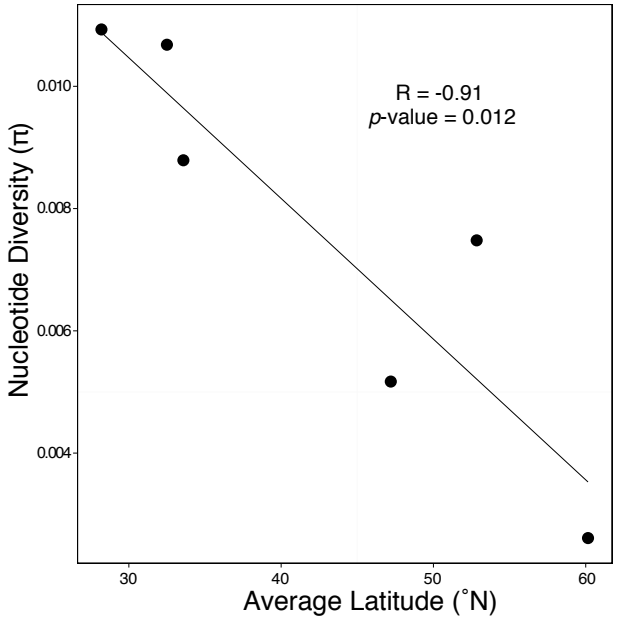

Figure 4. Average latitude was determined using a weighted average based on the total number of isolates per site as indicated in Table 1. Nucleotide diversity was calculated using concatenated MLSA loci and is expressed per site.