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.

Summary

The biogeography of *Streptomyces* was examined 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 6000 km. Previous analysis of this geographically explicit culture collection provided evidence for a latitudinal diversity gradient in *Streptomyces* species. Here the hypothesis that this latitudinal diversity gradient is a result of evolutionary dynamics associated with historical demographic processes was evaluated. Historical demographic phenomena have genetic consequences that can be evaluated through analysis of population genetics. Population genetic approaches were applied 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, it was found that nucleotide diversity within phylogroups was 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 (Finlay, 2002; Finlay and Fenchel, 2004). Yet endemism and dispersal limitation have been observed for a range of microbes (Cho and Tiedje, 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 dispersal can homogenize genetic diversity in pelagic systems (Coleman and Chisholm, 2010). At the other end of the spectrum are extremophiles such as *Sulfolobus* and thermophilic *Synechococcus*, which live in island-like hot spring habitats and exhibit strong patterns of allopatric divergence resulting from dispersal limitation (Papke *et al.*, 2003; Whitaker *et al.*, 2003; Cadillo-Quiroz *et al.*, 2012). Terrestrial microbes fall somewhere between these extremes. For example, 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).

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). 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 or latitude in determining spatial patterns of microbial diversity. However, SSU rRNA gene sequences have an extremely low rate of nucleotide substitution (Ochman *et al.*, 1999), and microbes with similar or even identical SSU rRNA genes can have extensive genomic and ecological diversity (Welch *et al.*, 2002; Jaspers and Overmann, 2004). Thus, this marker has low sensitivity for detecting neutral processes that drive patterns of biogeography, such as dispersal limitation and genetic drift (Green and Bohannan, 2006; Hanson *et al.*, 2012). However, these neutral processes are readily explored using geographically and ecologically explicit culture collections characterized at high genetic resolution.

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 (Kieser 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 (1–60 m) in prairie soils (Davelos *et al.*, 2004a,b) 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 1000 km (Doroghazi and Buckley, 2010; 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 dispersal limited species that have undergone a recent historical demographic range expansion (Doroghazi and Buckley, 2010; 2014). Demographic range expansion has previously been implicated as a factor that can explain ancestral patterns of gene exchange in *Streptomyces* clades (Andam *et al.*, 2016a). Combined, these data suggest a role for both adaptive and neutral processes in governing the diversity and biogeography of *Streptomyces*, but the evolutionary interpretation of these data depends on the degree to which dispersal limitation drives patterns of diversity in *Streptomyces*.

We evaluated *Streptomyces* biogeography to explore the ecological and evolutionary forces that govern diversification within this genus. The most powerful approach for detecting evolutionary patterns that result from dispersal limitation is to examine taxon-specific biogeography patterns across ecologically similar sites. This approach controls for the effects of selection and provides a powerful test of neutral evolutionary processes (Hanson *et al.*, 2012). We constructed a taxon-specific isolate collection of *Streptomyces* found in ecologically similar grassland sites spanning the United States. In a study of *Streptomyces* species-level diversity, which evaluated polymorphism at a single locus (*rpoB*), we observed evidence for dispersal limitation manifesting in a latitudinal diversity gradient (Andam *et al.*, 2016b). The latitudinal diversity gradient is one of the earliest and most well documented patterns of biogeography (Wallace, 1878; Hillebrand, 2014), but it is essentially undocumented in terrestrial bacterial systems. We hypothesize that historical demographic phenomena associated with Pleistocene glaciation generated the latitudinal gradient of *Streptomyces* species diversity (Andam *et al.*, 2016b). This hypothesis predicts that dispersal and gene flow will be limited and discontinuous across latitudes which correspond to the maximal extent of historical glaciation. The hypothesis also predicts that intra-species diversity will vary with latitude as a consequence of limited time for diversification in northern latitudes. These hypotheses are best addressed using population genetic approaches. Here we test these specific predictions by using multilocus sequence analysis (MLSA) to characterize population-level patterns of gene flow and genetic diversity within six phylogroups of *Streptomyces* which are distributed widely across the United States.

Table 1. Geographic distribution of *Streptomyces* phylogroups.

| Site | Code | Latitude | Longitude | Altitude | Isolates/site | Isolates from *Streptomyces* sp. phylogroup | | | | | |
						MAN125	MAN196	WA1063	MS200	MS152	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 (Uwharrie), 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

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.

Results

### Characterization of Streptomyces phylogroups

We identified a total of 308 isolates representing the six targeted phylogroups, and these isolates spanned 13 sites (Table 1). Strains within phylogroups share 99.4%–100% SSU rRNA gene sequence identity, and all strains share greater than 97% nucleotide identity in their SSU rRNA gene sequences (Supporting Information Fig. S1). Strains within phylogroups also share 97.6%–99.5% average nucleotide identity (ANI) across concatenated MLSA loci, and all strains share greater than 88% ANI (Fig. 1, Table 2). The pattern of genetic ancestry as determined by population structure analysis (see Structure analysis, Experimental procedures) is congruent with phylogroup boundaries (Supporting Information Fig. S2), and this indicates that these phylogroups approximate biological populations. We observed 122 unique MLSA haplotypes, with each phylogroup represented by 13–26 haplotypes (Fig. 1, Table 2). Good's coverage for each of the six phylogroups ranges from 0.88 to 1.0 for individual loci and 0.94 to 1.0 for concatenated MLSA loci clustered at 99% nucleotide identity (Supporting Information Table S1). Hence, allelic diversity and nucleotide diversity are well sampled (Supporting Information Table S1), though unique haplotypes remain under sampled (Supporting Information Fig. S3). The per site nucleotide diversity (π) of each phylogroup ranges from 0.0026 to 0.011 (Table 2).

Four of the phylogroups do not match any *Streptomyces* species described in PubMLST (< 97% ANI across MLSA loci; Supporting Information Fig. 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 (Supporting Information Fig. 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 (Supporting Information Fig. 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 (Supporting Information Fig. S4, Supporting Information 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 (Fig. 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 (Supporting Information Fig. S4B), but it is not clear whether these alleles are shared as a result of contemporary horizontal gene
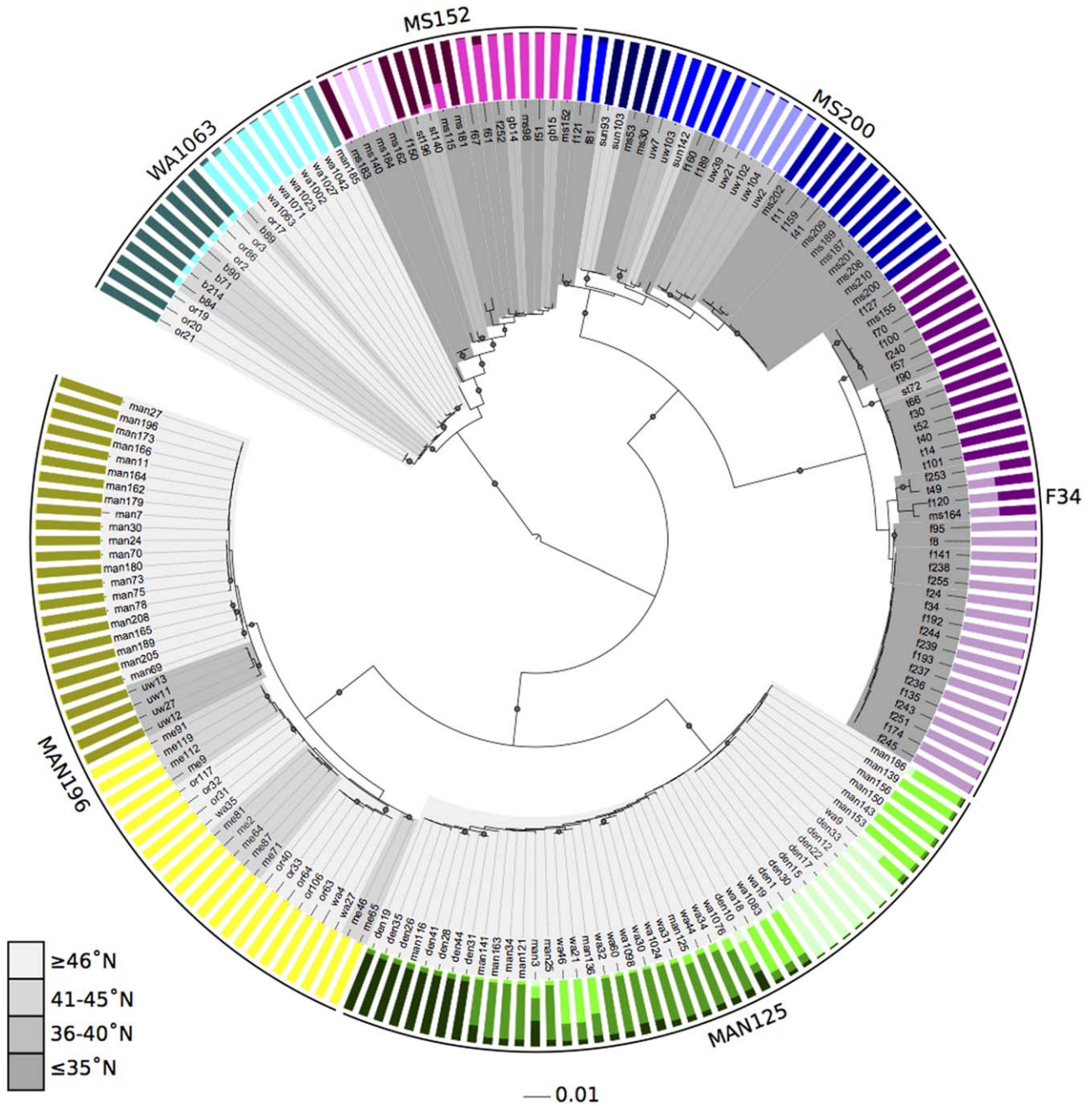
Fig. 1. Tree was constructed from concatenated MLSA loci nucleotide sequences using maximum likelihood with a GTRGAMMA evolution model. Scale bar represents nucleotide substitutions per site. The root was defined by *Mycobacterium smegmatis*. Nodes with bootstrap confidences > 80 are indicated with grey circles, and precise bootstrap values are found in Supporting Information Fig. S4A. The coloured 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.

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 approximately 4 sites (4.2 ± 0.4, mean + s.d.), and this distribution differs

Table 2. Phylogroup MLSA summary.

Phylogroup	Ave Lat	n	Haplotypes	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**

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.
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.

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 up to 3 sites (2.2 ± 0.9, mean $\pm$ s.d.), and this value is lower than expected if subpopulations are randomly distributed across the sites occupied by each phylogroup (permutation test, $P < 0.001$). These results indicate that phylogroup distribution is constrained geographically and that phylogroups have subpopulation structure that is also geographically explicit.

The geographic distribution of *Streptomyces* alleles indicates dispersal limitation. Identical alleles are shared among phylogroup members across each phylogroup's geographic range, which can exceed 5000 km (Fig. 2, Table 1). However, dissimilarity in allele composition increases with geographic distance, and this result is significant (Bray–Curtis dissimilarity, Mantel $r = 0.29$, $P = 0.005$) (Supporting Information Fig. S5). Hence, alleles are more likely shared between geographically similar sites suggesting 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 (Figs. 1 and 3) are unique to a single site, with the sole exception being a haplotype from phylogroup MS200 which is observed in both MS and WI (Sun Prairie).

We used nested clade analysis (NCA) combined with a rigorous statistical framework to evaluate population structure and demography (see Experimental procedures). Nested clade phylogeographic inference postulates potential evolutionary and historical demographic processes that support extant patterns of diversity and biogeography (see legend of Fig. 3). NCA establishes significant phylogeographic inferences for phylogroups MAN196, MAN125,

WA1063 and MS152 (Fig. 3) but not for MS200 and F34. Analysis of haplotype distribution is consistent with diversification due to dispersal limitation. For instance, population subdivision of MAN125 across the Pacific Northwest (Fig. 3A) and MAN196 between Maine and the Pacific Northwest (Fig. 3B) is consistent with restricted gene flow due to historical long distance dispersal events. Likewise, population subdivision of MS152 between the Southeastern sites in Mississippi and Florida (ms and f) and the site in California (st) (Fig. 3C), and subdivision of WA1063 between Brookfield, WI (b) and Oregon (or) (Fig. 3D) is consistent with allopatric fragmentation.

Latitudinal diversity gradient

The distribution and diversity of the phylogroups reveals a latitudinal diversity gradient. Strains from MAN125, MAN196, WA1063 occur mostly north of 40°N latitude, while strains from MS200, MS152, F34 occur mostly south of this latitude (Table 1, Fig. 2). This pattern of North/South partitioning is significant for each phylogroup when evaluated against the expectation of a random distribution across sites (permutation test, $P < 0.01$ for each phylogroup after Bonferroni correction). Furthermore, partial Mantel tests were performed to evaluate the latitudinal and longitudinal vector components of geographic distance in relation to the allele composition of sites. There remains a significant relationship between allele composition and geographic distance when we control for longitude, (Mantel $r = 0.23$, $P = 0.022$), but this relationship is no longer significant when we control for latitude (Mantel $r = 0.15$, $P = 0.12$). This result indicates that allele composition changes more across latitude than it does across longitude. The latitudinal partitioning of alleles can be readily observed in the pattern of allele sharing between sites (Fig. 2). Finally, we also observed a significant relationship between the nucleotide diversity of phylogroup MLSA loci and the average latitude of the locations in which they are
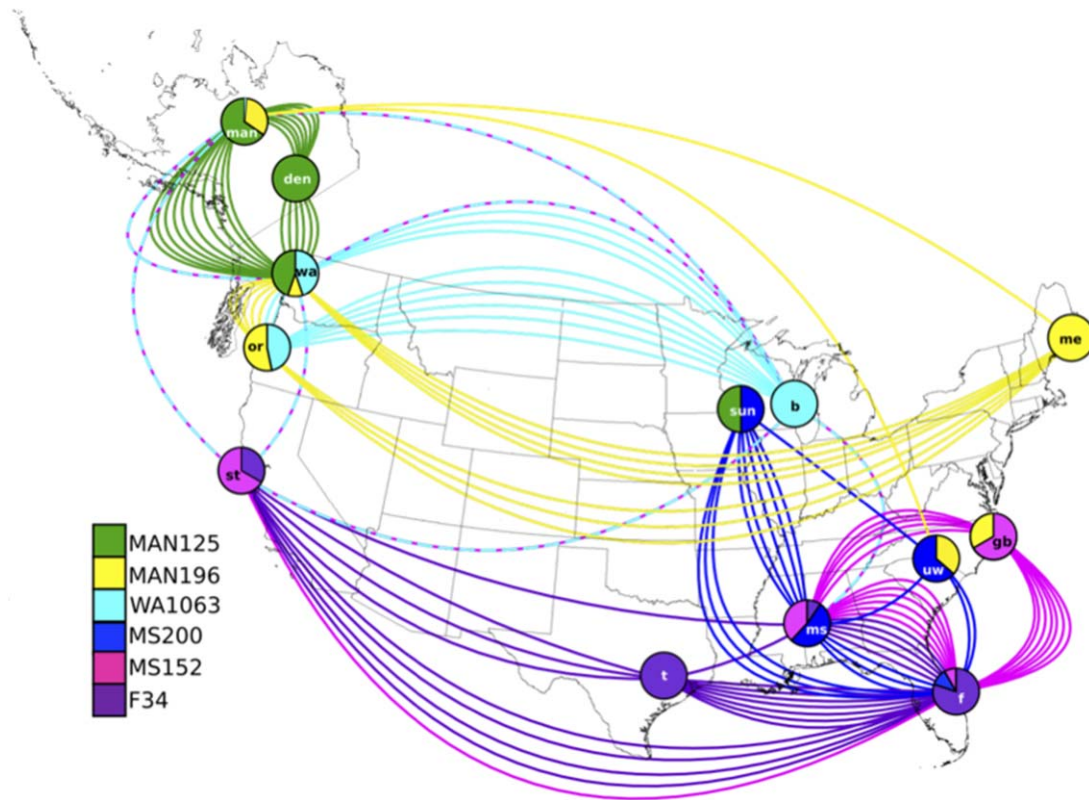
Fig. 2. Circles depict sample sites and are labelled according to site code (Table 1). The relative abundance of each phylogroup at each site is indicated by colour according to the legend, with raw counts provided in Table 1. Solid coloured lines represent identical alleles shared by phylogroup members subsampled for MLSA across sites. Dashed multicoloured lines depict identical *atpD* alleles shared by strains in phylogroups WA1063 and MS152 (Supporting Information Fig. S4B).

found (R = −0.91, *P* = 0.012; Fig. 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 (Supporting Information Fig. S5) with phylogroups inhabiting geographic ranges defined by latitude (Table 1, Fig. 2). In addition, the genetic diversity of phylogroups is negatively correlated to the latitude from which they were isolated, and on average, phylogroups isolated from lower latitudes were more genetically diverse than those isolated from higher latitudes (Fig. 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*.

The phylogroups we describe are coherent phylogenetic groups that approximate biological populations (Supporting Information Fig. S2). The members of each phylogroup share a distinguishable geographic range (Table 1, Fig. 2), a recent common ancestor (Fig. 1), and greater than 97% ANI across MLSA loci (Table 2). Despite geographic and genetic subpopulation structure (Fig. 1), phylogroup members frequently share identical alleles across demes. It is worth noting that since all of the strains examined share greater than 97% SSU rRNA gene identity, these geographic patterns would not be detected using standard SSU rRNA analyses methods (Supporting Information Fig. S1). We infer the presence of recombination within all phylogroups using both nucleotide polymorphism and phylogenetic methods (Table 2; Supporting Information Table S2).

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.
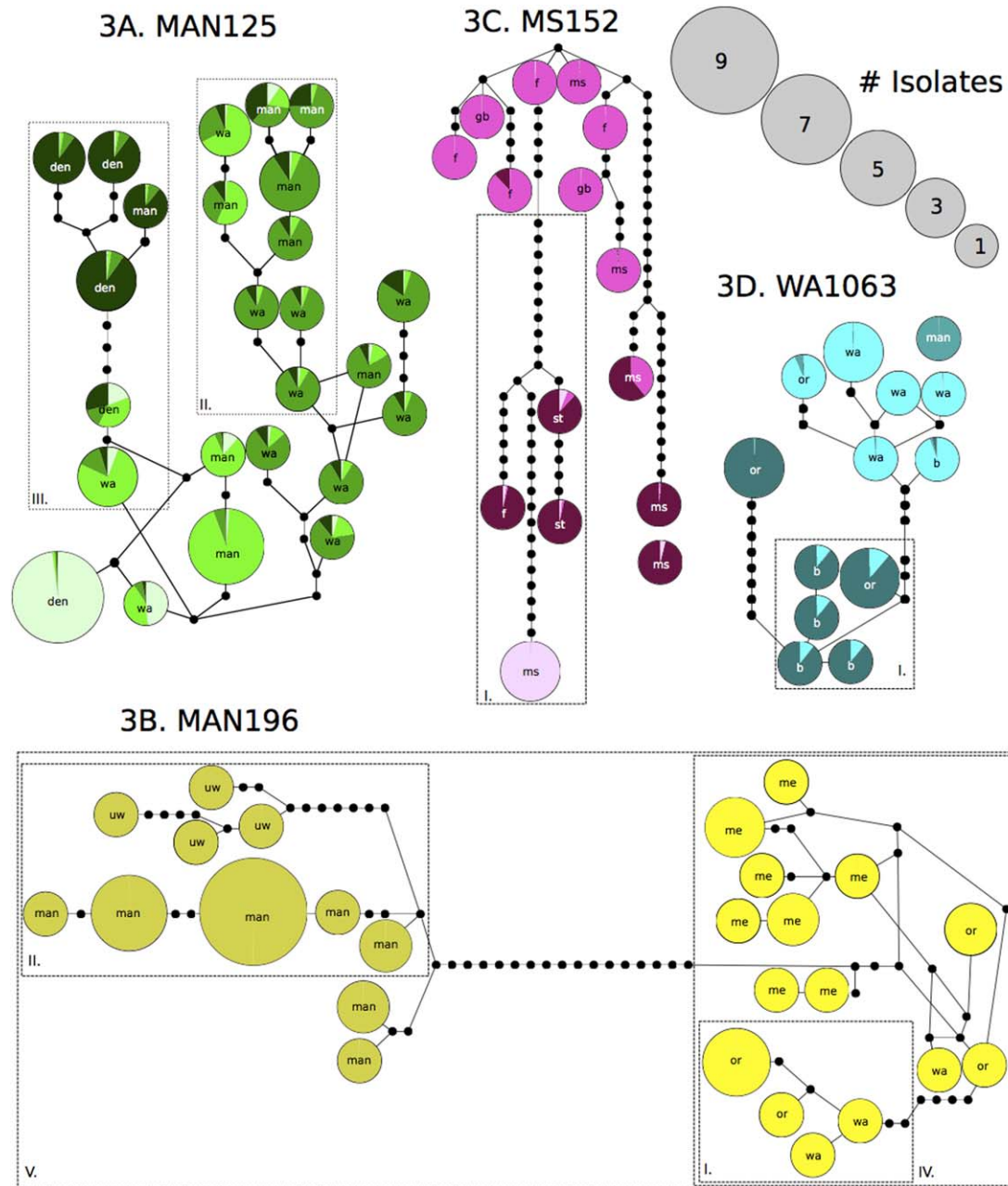
Fig. 3. Haplotype networks for phylogroups MAN125 (3A), MAN196 (3B), MS152 (3C), WA1063 (3D). Circles represent MLSA haplotypes whose radius is proportional to the number of isolates having that haplotype, and colours correspond to strain ancestry and subpopulation affiliation as determined by Structure analysis and defined in Fig. 1. Haplotypes are labelled 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 un-sampled, inferred haplotypes with each circle designating a single nucleotide polymorphism. The length of edges between nodes is uninformative. Dashed rectangles encompass clades that have significant phylogeographic inferences from nested clade analysis, as described in methods. Roman numerals correspond to the following inferences: I. Allopatric fragmentation; II. Long distance colonization and/or past fragmentation; III. Restricted gene flow with isolation by distance; IV. Restricted gene flow but with some long-distance gene flow over intermediate ranges not occupied by the species; or past gene flow followed by extinction of intermediate populations; V. Contiguous range expansion.

The geographic distribution of our phylogroups is nonrandom, and we find regional subpopulation structure within phylogroups (Figs. 1 and 3). Although we find identical alleles in sites thousands of kilometres apart (Fig. 2; Supporting Information Fig. S4), MLSA haplotypes are not shared across sites (with the single exception of a haplotype shared between MS and WI) (Fig. 3). We also observe a significant distance decay relationship for MLSA
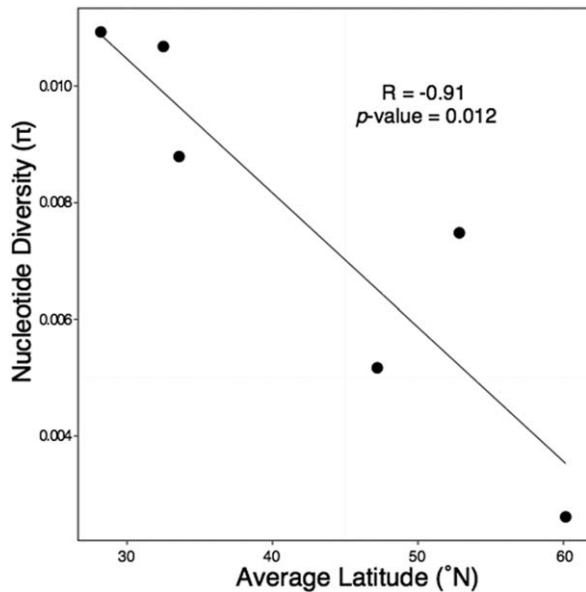
Fig. 4. Each point represents an individual phylogroup. Average latitude was determined using a weighted average based on the total number of isolates per site as indicated in Table 1. Nucleotide diversity was calculated using concatenated MLSA loci and is expressed per nucleotide site.

allele composition and geographic distance between sites (Supporting Information Fig. S5). Distance decay relationships driven by neutral processes can be challenging to identify given that environmental variables are often spatially structured (Nekola and White, 1999). However, Andam *et al.* (2016b) shows that *Streptomyces* community phylogenetic diversity across sites correlates significantly with latitude and temperature but not with soil pH, soil organic matter or rainfall. Hence, it is unlikely that these coarse environmental variables determine regional population structure. This data implies that while gene flow is moderate across the geographic range of a phylogroup, dispersal limitation and genetic drift create appreciable regional population structure.

Streptomyces phylogroup diversity is consistent with a latitudinal diversity gradient. While the classical description of a latitudinal diversity gradient defines diversity at the level of species richness, these diversity gradients are also apparent at the level of intra-species genetic diversity (Adams and Hadly, 2013). We find that latitude is a significant predictor of gene flow (Fig. 2). Furthermore, intra-phylogroup nucleotide diversity has a significant negative relationship with average latitude (Fig. 4), which is congruous with the latitudinal diversity gradient observed for diverse macroorganisms (Adams and Hadly, 2013; Hillebrand, 2014). 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.*, 2012;

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* (Supporting Information Fig. S1).

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, with post-glacial range expansion giving rise to the latitudinal partitioning of populations and species, and resulting in decreased molecular variation of northern lineages relative to southern ancestral lineages (Soltis *et al.*, 1997; Bernatchez and Wilson, 1998; Conroy and Cook, 2000; Mila 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 the hypothesis that historical demography and dispersal limitation have produced the latitudinal diversity gradient that we observe. For instance, MAN125 is nearly in linkage equilibrium ($I_A = 0.09$) though its members span a geographic range of over 2000 km across the Pacific Northwest. Similar patterns of recombination have also been observed in a *S. pratensis* population which spanned 1000 km across sites present in North Carolina and northern New York (Doroghazi and Buckley, 2010; 2014). The observation of linkage equilibrium indicates that either the population lacks contemporary barriers to gene flow or the population has experienced a recent historical demographic expansion. Coupled with evidence of limitations to gene flow at regional scales (Fig. 2; Supporting Information Fig. S5), the most parsimonious explanation for these conflicting observations is a recent historical demographic range expansion. Conversely, latitudinal gradients in marine 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; 2004).

The population structure we observe in our *Streptomyces* phylogroups is consistent with the effects of post-glacial demographic range expansion followed by dispersal limitation and regional diversification. Dispersal limitation may occur following range expansion as a result of density dependent blocking or by adaptive barriers that arise after the colonization of new habitat. 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 (Fig. 4). Similar relationships between intraspecific nucleotide diversity and average latitude as result of post-glacial colonization are evident in other systems (Bernatchez and Wilson, 1998). 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 (Fig. 2), 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 (Fig. 3). 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 Beaumont, 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 (Fig. 1). The contemporary population structure of MAN125 and MAN196 is consistent with a historical range expansion from a common ancestor shared by both clades (Figs. 1 and 3). Further population structure within these phylogroups likely resulted from historical dispersal events across the Pacific Northwest and Alaska for MAN125 (Fig. 3A) and between Maine and the Pacific Northwest and the East Coast and Alaska for MAN196 (Fig. 3B). Restricted gene flow likely influenced additional population subdivision within MAN125 between Washington and both Alaska sites (Fig. 3A). Analysis of haplotype distribution within WA1063 and MS152 is also consistent with allopatric fragmentation and diversification of populations as a result of geographic isolation between the Midwest and Pacific Northwest for WA1063 (Fig. 3D) and between the Southeast and West for MS152 (Fig. 3C).

Phylogroups WA1063 and MS152 share a recent common ancestor (Fig. 1 and Supporting Information Fig. S2) and also share two identical alleles at the *atpD* locus (Supporting Information Fig. S4B). These phylogroups have distinct, non-overlapping geographic ranges with WA1063 found in higher latitudes and MS152 in lower latitudes (Table 1, Fig. 2). WA1063 and MS152 have 0.0205 net nucleotide substitutions per site across concatenated MLSA loci. We evaluate the possible time range for divergence between WA1063 and MS152 by extrapolating very roughly from the nucleotide substitution rate ($\mu = 4.5 \times 10^{-9}$) and generation time (100–300 generations per year) for *Escherichia coli* (Ochman *et al.*, 1999), since corresponding values are not available for *Streptomyces* or their relatives. Based upon these gross approximations, we would estimate that WA1063 and MS152 diverged 15 000–50 000 years ago, corresponding to events in the late Pleistocene (Clayton *et al.*, 2006). Hence, it is likely that the identical *atpD* alleles found in both WA1063 and MS152 were inherited from a shared ancestral population

(Supporting Information Fig. S3B) as opposed to inheritance from contemporary gene exchange.

Through population genetic analysis of six *Streptomyces* phylogroups, we find evidence for dispersal limitation associated with geographically explicit patterns of gene flow which manifest in a latitudinal gradient of nucleotide diversity. Furthermore, these data support the hypothesis that historical demographic processes influence the contemporary biogeography of *Streptomyces*. Due to their spore forming capabilities and potential for long range dispersal, *Streptomyces* are an ideal system for assessing limits on gene flow among terrestrial bacteria. 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 1000 *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. $\pm$ 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, 2013) 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) (Supporting Information Table S3). The V2 and V2 regions of SSU rRNA sequences were amplified using universal primers 8F and 1492R (Supporting Information Table S3). Reactions for Sanger sequencing were performed using forward primers for all loci except *rpoB*, for which the reverse primer was used. Trace files were uniformly screened using CAP3 (Huang and Madan, 1999), and sequences were trimmed as to discard nucleobases with a Phred quality score below 23. Sequences were aligned using MUSCLE (Edgar, 2004), trimmed to 431 bp, 415 bp, 446 bp, 557 bp and 489 bp, for genes *atpD*, *gyrB*, *recA*, *rpoB* and *trpB*, respectively, and concatenated consistently with the genomic order in *Streptomyces coelicolor* A3(2). SSU rRNA sequences were trimmed to 357 bp, creating an alignment spanning the V1 and V2 regions. Gene sequences are available on GenBank with accession numbers KX110408-KA111380.

Good's coverage estimation and haplotype rarefaction was determined using mothur (Schloss *et al.*, 2009). DNA polymorphism statistics including, number of segregating sites, nucleotide diversity, and Tajima's D, were determined with DnaSP v5 (Librado and Rozas, 2009) and LDhat (McVean *et al.*, 2002). Population scaled mutation rates (Watterson's theta; $\theta_w = 2N_e\mu$), recombination or gene conversion rates ($\rho = 2N_e r/2$), and relative rates of recombination (ρ/θ_w) were estimated using LDhat (McVean *et al.*, 2002) and are expressed per nucleotide site. The standard index of association (I_A) was calculated from allelic data with LIAN v3.5 (Haubold and Hudson, 2000) using the Monte–Carlo test and 100 iterations. The pairwise homoplasy index (PHI) statistic was determined using PhiPack (Bruen et al., 2005), and statistical significance was evaluated under a null hypothesis of no recombination. Sequence identity across phylogroups was calculated with mothur (Schloss et al., 2009).

Phylogenetic reconstruction

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

Mycobacterium smegmatis. Significant phylogenetic incongruence between loci was determined using the Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999) implemented in the R package phangorn (Schliep, 2011).

Population structure

Concatenated MLSA sequences were analyzed using Structure v2.3.3 (Pritchard *et al.*, 2000) to examine population affiliation, subdivision, and admixture within and between phylogroups. Structure was run using an admixture model with a burn-in length of $1.0E^6$ 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 1000 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 1000 permutations. Bray–Curtis dissimilarity was calculated from allele composition across sites.

Haplotype networks were created using a statistical parsimony procedure (Templeton *et al.*, 1987; 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, 2006).

Acknowledgements

References

Allen, A.P., Brown, J.H., and Gillooly, J.F. (2002) Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science* 297: 1545–1548.

Adams, R.I., and Hadly, E.A. (2013) Genetic diversity within vertebrate species is greater at lower latitudes. *Evol Ecol* 27: 113–143.

Andam, C.P., Choudoir, M.J., Nguygen, A.V., Park, H.S., and Buckley, D.H. (2016a) Contributions of ancestral interspecies recombination to the genetic diversity of extant streptomyces lineages. *ISMEJ* 10: 1731–1741.

Andam, C.P., Doroghazi, J.R., Campbell, A.N., Kelly, P.J., Choudoir, M.J., and Buckley, D.H. (2016b) A latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*. *mBio* 7: e02200–e02215.

Antony-Babu, S., Stach, J.E.M., and 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., and Wilson, C.C. (1998) Comparative phylogeography of nearctic and palearctic fishes. *Mol Ecol* 7: 431–452.

Bissett, A., Richardson, A.E., Baker, G., Wakelin, S., and Thrall, P.H. (2010) Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. *Mol Ecol* 19: 4315–4327.

Boucher, Y., Cordero, O.X., Takemura, A., Hunt, D.E., Schliep, K., Bapteste, E., *et al.* (2011) Local mobile gene pools rapidly cross species boundaries to create endemicity within global *vibrio cholerae* populations. *mBio* 2: e00335–e00310.

Bruen, T.C., Philippe, H., and Bryant, D. (2005) A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172: 2665–2681.

Cadillo-Quiroz, H., Didelot, X., Held, N.L., Herrera, A., Darling, A., Reno, M.L., *et al.* (2012) Patterns of gene flow define species of thermophilic Archaea. *PLoS Biol* 10: e1001265.

Cho, J., and Tiedje, J.M. (2000) Biogeography and degree of endemicity of fluorescent *pseudomonas* strains in soil. *Appl Environ Microbiol* 66: 5448–5456.

Choudoir, M.J., Campbell, A.N., and Buckley, D.H. (2012) Grappling with proteus: population level approaches to understanding microbial diversity. *Front Microbiol* 3: 336.

Chu, H., Fierer, N., Lauber, C.L., Caporaso, J.G., Knight, R., and Grogan, P. (2010) Soil bacterial diversity in the arctic is not fundamentally different from that found in other biomes. *Environ Microbiol* 12: 2998–3006.

Clayton, L., Attig, J.W., Mickelson, D.M., Johnson, M.D., and Syverson, K.M. (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., and Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9: 1657–1659.

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

Conroy, C.J., and Cook, J.A. (2000) Phylogeography of a post-glacial colonizer: *microtus longicaudus* (rodentia: muridae). *Mol Ecol* 9: 165–175.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J.F., Hawkins, B.A., *et al.* (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol Lett* **7:** 1121–1134.

Davelos, A.L., Xiao, K., Samac, D.A., Martin, A.P., and Kinkel, L.L. (2004a) Spatial variation in *streptomyces* genetic composition and diversity in a prairie soil. *Microb Ecol* **48:** 601–612.

Davelos, A.L., Kinkel, L.L., and Samac, D.A. (2004b) Spatial variation in frequency and intensity of antibiotic interactions among *streptomycetes* from prairie soil. *Appl Environ Microbiol* **70:** 1051–1058.

Doroghazi, J.R., and Buckley, D.H. (2010) Widespread homologous recombination within and between *streptomyces* species. *ISME J* **4:** 1136–1143.

Doroghazi, J.R., and Buckley, D.H. (2014) Intraspecies comparison of *streptomyces pratensis* genomes reveals high levels of recombination and gene conservation between strains of disparate geographic origin. *BMC Genom* **15:** 970.

Earl, D.A., and vonHoldt, B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv Genet Resour* **4:** 359–361.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Res* **32:** 1792–1797.

Eisenlord, S.D., Zak, D.R., and Upchurch, R.A. (2012) Dispersal limitation and the assembly of soil *actinobacteria* communities in a long-term chronosequence. *Ecol Evol* **2:** 538–549.

El-Nakeeb, M.A., and Lechevalier, H.A. (1963) Selective isolation of aerobic Actinomycetes. *Appl Microbiol* **11:** 75–77.

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

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

Fierer, N., Carney, K.M., Horner-Devine, M.C., and Megonigal, J.P. (2009) The biogeography of ammonia-oxidizing bacterial communities in soil. *Microb Ecol* **58:** 435–445.

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

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

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

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

Goslee, S.C., and Urban, D.L. (2007) The ecodist package for dissimilarity-based analysis of ecological data. *J Stat Softw* **22:** 1–19.

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

Guo, Y., Zheng, W., Rong, X., and 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, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10:** 497–506.

Hatosy, S.M., Martiny, J.B.H., Sachdeva, R., Steele, J., Fuhrman, J.A., and Martiny, A.C. (2013) Beta diversity of marine bacteria depends on temporal scale. *Ecology* **94:** 1898–1904.

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

Hewitt, G.M. (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, G.M. (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., and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res* **9:** 868–677.

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

Jolley, K.A., Chan, M.S., and Maiden, M.C.J. (2004) mlstdbNet - distributed multi-locus sequence typing (MLST) databases. *BMC Bioinform* **5:** 86.

Kâmpfer, P. (2006) The family *Streptomycetaceae* Part I: taxonomy. In *The Prokaryotes*, Vol 3. Dworkin, M., Falkow, S., Rosenberg, E., Schliefer, K., and Stackebrandt, E. (eds). New York: Springer, pp. 538–604.

Kenefic, L.J., Pearson, T., Okinaka, R.T., Schupp, J.M., Wagner, D.M., Hoffmaster, A.R., *et al.* (2009) Pre-columbian origins for north american anthrax. *PLoS One* **4:** e4813.

Kieser, T., Bibb, M., Buttner, M., Charter, K., and Hopwood, D. (2000) *Practical Streptomyces Genetics*. Norwich, UK: John Innes Foundation.

Kim, B.J., Kim, C.J., Chun, J., Koh, Y.H., Lee, S.H., Hyun, J.W., *et al.* (2004) Phylogenetic analysis of the genera *streptomyces* and *kitasatospora* based on partial RNA polymerase beta-subunit gene (*rpoB*) sequences. *Int J Syst Evol Microbiol* **54:** 593–598.

Knowles, L., and Maddison, W. (2002) Statistical phylogeography. *Mol Ecol* **11:** 2623–2635.

Labeda, D.P. (2011) Multilocus sequence analysis of phytopathogenic species of the genus *Streptomyces*. *Int J Syst Evol Microbiol* **61:** 2525–2531.

Lauber, C.L., Hamady, M., Knight, R., and Fierer, N. (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* **75:** 5111–5120.

Librado, P., and Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451–1452.

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

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

Maggs, C.A., Castilho, R., Foltz, D., Henzler, C., Jolly, M.T., 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 ağeneralized regression approach. *Cancer Res* **27:** 209–220.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4:** 102–112.

Martiny, J.B.H., Eisen, J.A., Penn, K., Allison, S.D., and Horner-Devine, M.C. (2011) Drivers of bacterial β -diversity depend on spatial scale. *Proc Natl Acad Sci U S A* **108:** 7850–7854.

McCarthy, A.J., and Williams, S.T. (1992) Actinomycetes as agents of biodegradation in the environment - a review. *Gene* **115:** 189–192.

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

Mikheyev, A.S., Vo, T., and Mueller, U.G. (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., and Wayne, R. (2006) Postglacial population expansion drives the evolution of long-distance migration in a songbird. *Evolution* **60:** 2403–2409.

Miller, S.R., Strong, A.L., Jones, K.L., and Ungerer, M.C. (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, G.G., Schemske, D.W., Cornell, H.V., Allen, A.P., Brown, J.M., Bush, M.B., *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.

Nekola, J.C., and White, P.S. (1999) The distance decay similarity in biogeography and ecology. *J Biogeogr* **26:** 867–878.

Neufeld, J.D., and Mohn, W.W. (2005) Unexpectedly high bacterial diversity in arctic tundra relative to boreal forest soils, revealed by serial analysis of ribosomal sequence tags. *Appl Environ Microbiol* **71:** 5710–5718.

Nielsen, R., and Beaumont, M.A. (2009) Statistical inferences in phylogeography. *Mol Ecol* **18:** 1034–1047.

Ochman, H., Elwyn, S., and Moran, N.A. (1999) Calibrating bacterial evolution. *Proc Natl Acad Sci U S A* **96:** 12638–12643.

Ottow, J.C.G. (2013) Rose bengal as a selective aid in the isolation of fungi and actinomycetes from natural sources. *Mycologia* **64:** 304–315.

Panchal, M. (2006) The automation of nested clade phylogeographic Analysis. *Bioinformatics* **23:** 509–510.

Papke, R.T., Ramsing, N.B., Bateson, M.M., and Ward, D.M. (2003) Geographical isolation in hot spring cyanobacteria. *Environ Microbiol* **5:** 650–659.

Pattengale, N.D., Alipour, M., Bininda-Emonds, O.R.P., Moret, B.M.E., and Stamatakis, A. (2010) How many bootstrap replicates are necessary? *J Comput Biol* **17:** 337–354.

Pearson, T., Giffard, P., Beckstrom-Sternberg, S., Auerbach, R., Hornstra, H., Tuanyok, A., *et al.* (2009) Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. *BMC Biol* **7:** 78.

Pommier, T., Canback, B., Lundberg, P., Hagstrom, A., and Tunlid, A. (2009) RAMI: a tool for identification and characterization of phylogenetic clusters in microbial communities. *Bioinformatics* **25:** 736–742.

Posada, D., Crandall, K.A., and Templeton, A.R. (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol Ecol* **9:** 487–488.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Roberts, M.A., and Crawford, D.L. (2000) Use of randomly amplified polymorphic DNA as a means of developing genus- and strain-specific *streptomyces* DNA probes. *Appl Environ Microbiol* **66:** 2555–2564.

Rong, X., and Huang, Y. (2010) Taxonomic evaluation of the *streptomyces griseus* clade using multilocus sequence analysis and DNA-DNA hybridization, with proposal to combine 29 species and three subspecies as 11 genomic species. *Int J Syst Evol Microbiol* **60:** 696–703.

Rousk, J., Bååth, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G., *et al.* (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* **4:** 1340–1351.

Schliep, K.P. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics* **27:** 592–593.

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

Shimodaira, H., and Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* **16:** 1114–1116.

Smouse, P.E., Long, J.C., and Sokal, R.R. (1986) Multiple regression and correlation extensions of the mantel test of matrix correspondence. *Syst Zool* **35:** 627–632.

Soltis, D., Gitzendanner, M., Strenge, D., and Soltis, P. (1997) Chloroplast DNA intraspecific phylogeography of plants from the pacific northwest of north America. *Plant Syst Evol* **206:** 353–373.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

Stevens, R.D. (2006) Historical processes enhance patterns of diversity along latitudinal gradients. *Proc R Soc B Biol Sci* **273:** 2283–2289.

Sul, W.J., Oliver, T.A., Ducklow, H.W., Amaral-Zettler, L.A., and Sogin, M.L. (2012) Marine bacteria exhibit a bipolar distribution. *Proc Natl Acad Sci U S A* **110:** 2342–2347.

Swan, B.K., Tupper, B., Sczyrba, A., Lauro, F.M., Martinez-Garcia, M., González, J.M., *et al.* (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci U S A* **110:** 11463–11468.

Takasuka, T.E., Book, A.J., Lewin, G.R., Currie, C.R., and Fox, B.G. (2013) Aerobic deconstruction of cellulosic biomass by an insect-associated *Streptomyces*. *Sci Rep* **3:** 1–10.

Tavaré, S. (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. *Am Math Soc Lect Math Life Sci* **17:** 57–86.

Telford, R.J., Vandvik, V., and Birks, H.J.B. (2006) Dispersal limitations matter for microbial morphospecies. *Science* **312:** 1015.

Templeton, A.R., Boerwinkle, E., and Sing, C.F. (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in Drosophila. *Genetics* **117:** 343–351.

Templeton, A.R., Routman, E., and Phillips, C.A. (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *ambystoma tigrinum*. *Genetics* **140:** 767–782.

Vos, M., and Velicer, G.J. (2008) Isolation by distance in the spore-forming soil bacterium *myxococcus xanthus*. *Curr Biol* **18:** 386–391.

Wallace, A. (1878) *Tropical Nature, and Other Essays*. London: Macmillan and Co.

Watve, M., Tickoo, R., Jog, M., and Bhole, B. (2001) How many antibiotics are produced by the genus *streptomyces*? *Arch Microbiol* **176:** 386–390.

Wawrik, B., Kutliev, D., Abdivasievna, U.A., Kukor, J.J., Zylstra, G.J., and Kerkhof, L. (2007) Biogeography of actinomycete communities and type ii polyketide synthase genes in soils collected in new jersey and central Asia. *Appl Environ Microbiol* **73:** 2982–2989.

Welch, R.A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., *et al.* (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* **99:** 17020–17024.

Whitaker, R.J., Grogan, D.W., and Taylor, J.W. (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301:** 976–978.

Wiens, J.J., and Donoghue, M.J. (2004) Historical biogeography, ecology and species richness. *Trends Ecol Evol* **19:** 639–644.

Williams, D., Dunkerley, D., DeDecker, P., Kershaw, P., and Chappell, M. (1998) *Quaternary Environments*. London: Arnold.

Wilson, A.B., and Veraguth, I.E. (2010) The impact of pleistocene glaciation across the range of a widespread european coastal species. *Mol Ecol* **19:** 4535–4553.

Supporting information

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

Table S1. Good's coverage estimation for concatenated MLSA sequences and individual loci. Values are calculated for unique sequences or for operational taxonomic units defined at a 0.01 nucleotide dissimilarity cutoff.

Table S2. Reported values are *p*-values for the Shimodaira-Hasegawa test for phylogenetic incongruence. Rows represent maximum likelihood topologies and columns the nucleotide alignments used to evaluate incongruence. Corresponding phylogenic trees were constructed as defined in the methods and are depicted in Supporting Information Fig. S4.

Table S3. MLSA PCR conditions and primer sets. PCR amplifications were performed using AmpliTaq Gold reagents (Applied Biosystems, Foster City, CA, USA) and the following 25 µl reaction: 11.75 µl H_2O, 2.5 µl 10X buffer, 3 µl 25 mM $MgCl_2$, 2 µl dNTP mixture (2.5 mM each dNTP, 10 mM total dNTPs, Promega, Madison, WI, USA), 1 µl forward primer from 10 µM stock, 1 µl reverse primer from 10 µM stock, 2.5 µl DMSO, 0.25 µl AmpliTaq Gold (5 U/µl), and 1 µl DNA template. For all MLSA primer sets, the following reaction conditions were used: 95°C for 10 min; 35 cycles of 95°C for 20 s, 65°C for 30 s, 72°C for 45 s; 72°C for 10 min. For the 16S locus, reactions conditions were the same expect for 35 cycles of annealing at 55°C for 90 s.

Fig. S1. The maximum likelihood tree was constructed with a GTRGAMMA evolution model using partial SSU rRNA gene sequences from three representatives for each phylogroup. Scale bar represents nucleotide substitutions per site. Branches with bootstrap support values greater than 75% are labelled, and the outgroup is *Mycobacterium smegmatis*. Strains are coloured to reveal phylogroup membership as indicated in the colour legend provided.

Fig. S2. Structure analysis was performed on all sequences, irrespective of phylogroup membership. (This approach differs from that depicted in Fig. 1 which shows the results of independent Structure analyses performed on each separate phylogroup.) The coloured bars in outer rings indicate genetic contributions from different ancestral populations as indicated in the legend (a). Since WA1063 and MS200 share an ancestral population, we performed an additional Structure analysis that included only sequences from these phylogroups, and this analysis reveals clear subdivision between these phylogroups (b). Taxa names in the inner ring are coloured to indicate the phylogroup designation of each strain as defined in the legend. The tree was constructed from concatenated MLSA loci nucleotide sequences using maximum likelihood with a GTRGAMMA evolution model. Scale bar represents nucleotide substitutions per site. The root was defined by *Mycobacterium smegmatis*. Nodes with bootstrap confidences > 80 are indicated with grey circles, and precise bootstrap values are found in Supporting Information Fig. S4A.

Fig. S3. The number of unique MLSA haplotypes is plotted against the percent of total isolates sampled for MLSA (Table 2). The rarefaction curves for phylogroups MS200 and F34 are identical and are depicted with dashed lines.

Good's coverage estimations for each phylogroup can be found in Table S1.

Fig. S4. Phylogenetic trees were constructed from nucleotide sequences of concatenated MLSA loci (A), *atpD* (B), *gryB* (C), *recA* (D), *rpoB* (E), *trpB* (F) by maximum likelihood with a GTRGAMMA evolution model. The scale bar represents nucleotide substitutions per site. Branches with bootstrap support values > 75 are labeled. For Supporting Information Fig. S4A, representative isolates from PubMLST were included to orient phylogroup taxonomy to existing *Streptomyces* type strains. The root was defined by *Mycobacterium smegmatis* in each case. The outer colour strip depicts phylogroup membership consistent with Fig. 2 [green (MAN125), yellow (MAN196), teal (WA1063), blue (MS200), pink (MS152), purple (F34)]. The key for phylogroup isolate names can be found in Table 1.

Fig. S5. Circles represent pairwise comparisons of Bray-Curtis dissimilarity and geographic distance between sites for all phylogroups. Bray-Curtis dissimilarity was calculated from allele composition across sites. The Mantel coefficient (r) and *p*-value under the null hypothesis of no correlation between the matrixes are reported.