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

1

1   **Summary**

2   We examined the biogeography of *Streptomyces* at regional spatial scales to identify factors that

3   govern patterns of microbial diversity within the genus. *Streptomyces* are spore forming

4   filamentous bacteria which are widespread in soil and a major source of antibiotics. We applied

5   population genetic approaches to analyze geographic and genetic population structure in six

6   phylogroups of *Streptomyces* identified from a geographically explicit culture collection.

7   *Streptomyces* strains were isolated from soils associated with perennial grass habitats sampled

8   across a spatial scale of more than 6,000 km. We observed that *Streptomyces* population structure

9   varies at regional spatial scales and that allelic diversity correlates with geographic distance. In

10   addition, patterns of allelic diversity and gene flow are defined by latitude. Finally, we found that

11   nucleotide diversity within phylogroups is negatively correlated with latitude. These results

12   indicate that phylogroup diversification is constrained by dispersal limitation at regional spatial

13   scales. We hypothesize that historical demographic processes have influenced the contemporary

14   biogeography of *Streptomyces*.

15

16   **Introduction**

17   Patterns of microbial biogeography have been widely documented (Whitaker *et al*., 2003; Vos and

18   Velicer, 2008; Bissett *et al*., 2010; Martiny *et al*., 2011; Gilbert *et al*., 2012; Hatosy *et al*., 2013),

19   and yet we are only beginning to understand the evolutionary forces that generate and maintain

20   these patterns. Explorations of biogeography are valuable because biogeographical patterns

21   illustrate fundamental principles of evolution and ecology. Biogeographical patterns are ultimately

22   governed by rates of dispersal and diversification (Martiny *et al*., 2006; Hanson *et al*., 2012). Since

23   microbial dispersal cannot be observed directly, rates of dispersal are typically inferred from extant

24   patterns of genetic diversity. It has been hypothesized that microbes disperse ubiquitously due to

25   their small cell size and massive population numbers (Finaly, 2002; Finlay and Fenchel, 2004).

26   Yet endemism and dispersal limitation have been observed for a range of microbes (Cho and

27   Tiedge, 2000; Green and Bohannan, 2006; Telford *et al*., 2006; Boucher *et al*., 2011), and microbial

28   dispersal limitation has been verified experimentally (Bell, 2010). Evidence in the literature

29   pertaining to microbial dispersal limitation, which initially appears to be contradictory, can be

30   explained by at least two factors: first, dispersal constraints are likely to vary with respect to

31   different species and habitats; and second, methods used to define microbial diversity vary

32   dramatically in their taxonomic and phylogenetic resolution. Each of these factors has been

33   discussed previously (Hanson *et al*., 2012; Choudoir *et al*., 2012), and we will only consider them

34   briefly here.

35

36   Patterns of microbial dispersal and gene flow appear to differ between habitats and species. At one

37   end of the spectrum, globally widespread microbes such as *Prochlorococcus* and *Pelagibacter*

38   show little variation in gene content between the Atlantic and Pacific Oceans suggesting that

39   dispersal can homogenize genetic diversity in pelagic systems (Coleman and Chisholm, 2010). At

40   the other end of the spectrum are extremophiles such as *Sulfolobus* and thermophilic

41   *Synechococcus*, which live in island-like volcanic habitats and exhibit strong patterns of allopatric

42   divergence resulting from dispersal limitation (Papke *et al*., 2003; Whitaker *et al*., 2003; Cadillo-

43   Quiroz *et al*., 2012). Terrestrial microbes fall somewhere between these extremes. For example,

44   soil dwelling microbes such as *Burkholderia pseudomallei*, *Burkholderia mallei*, and *Bacillus*

45   *anthracis* exhibit biogeographical patterns governed by dispersal limitation at regional spatial

46   scales (Kenefic *et al*., 2009; Pearson *et al*., 2009).

3

47

48   The phylogenetic resolution at which microbial diversity is defined can have a profound impact

49   on our ability to discern patterns of microbial biogeography (as reviewed Hanson *et al.*, 2012).

50   Surveys of SSU rRNA genes in terrestrial habitats indicate that environmental variables including

51   temperature (Fierer *et al.*, 2009; Miller *et al.*, 2009), pH (Fierer and Jackson, 2006; Lauber *et al.*,

52   2009; Rousk *et al.*, 2010), and salinity (Lozupone and Knight, 2007) are more important than

53   geographic distance or latitude in determining spatial patterns of microbial diversity. However,

54   SSU rRNA gene sequences have an extremely low rate of nucleotide substitution (Ochman *et al.*,

55   1999), and microbes with similar or even identical SSU rRNA genes can have extensive genomic

56   and ecological diversity (Welch *et al.*, 2002; Jaspers and Overmann; 2004). Thus, this marker has

57   low sensitivity for detecting neutral processes that drive patterns of biogeography, such as dispersal

58   limitation and genetic drift (Green and Bohannan, 2006; Hanson *et al.*, 2012). However, these

59   neutral processes are readily explored using geographically and ecologically explicit culture

60   collections characterized at high genetic resolution.

61

62   *Streptomyces* are ubiquitous across soil habitats, and many species are easily cultured, making this

63   genus an excellent candidate for making a taxon-specific survey of biogeography. Furthermore,

64   *Streptomyces* species have high rates of gene exchange both within and between species

65   (Doroghazi and Buckley, 2010). Hence, this genus is an ideal model to explore dispersal limitations

66   and gene flow in terrestrial systems. *Streptomyces* are gram-positive *Actinobacteria* (Kâmpfer,

67   2006) known for their complex developmental cycle, which entails filamentous growth and the

68   formation of desiccation resistant spores which are readily dispersed (Keiser *et al.*, 2000).

69   *Streptomyces* play a significant role in the terrestrial carbon cycle (McCarthy and Williams, 1992;

4

70   Takasuka *et al.*, 2013), represent important agricultural pathogens (Loria *et al.*, 2006; Labeda,

71   2011), and are prolific producers of antibiotics (Watve *et al.*, 2001). Despite their importance, we

72   lack an evolutionary framework to understand *Streptomyces* biogeography.

73

74   *Streptomyces* diversity varies spatially, though the influence of geographic distance and ecological

75   variation remains poorly resolved. Ecological adaptation constrains the environmental distribution

76   of *Streptomyces* and their genetic and phenotypic diversity can vary in relation to soil

77   characteristics at small spatial scale (1 m – 60 m) in prairie soils (Davelos *et al.*, 2004a and 2004b),

78   and in dune habitats (Antony-Babu *et al.*, 2008). There is also evidence for dispersal limitation at

79   very large (continental) spatial scales with endemic species observed in North America and Central

80   Asia (Wawrik *et al.*, 2007). Remarkably, genetic analysis of *Streptomyces pratensis* has revealed

81   that strains of this species are in linkage equilibrium across a range that spans 1,000 km (Doroghazi

82   and Buckley, 2010 and 2014). One interpretation of this finding is that *Streptomyces* are subject to

83   widespread dispersal and unlimited gene flow at regional spatial scales. However, linkage

84   equilibrium can also be observed for dispersal limited species that have undergone a recent

85   demographic range expansion (Doroghazi and Buckley, 2010 and 2014). Demographic range

86   expansion has previously been implicated as a factor that can explain ancestral patterns of gene

87   exchange in *Streptomyces* clades (Andam *et al.*, 2016a). Combined, these data suggest a role for

88   both adaptive and neutral processes in governing the diversity and biogeography of *Streptomyces*,

89   but the evolutionary interpretation of these data depends on the degree to which dispersal limitation

90   drives patterns of diversity in *Streptomyces*.

91

5

92   We evaluated *Streptomyces* biogeography to explore the ecological and evolutionary forces that

93   govern diversification within this genus. The most powerful approach for detecting evolutionary

94   patterns that result from dispersal limitation is to examine taxon-specific patterns of biogeography

95   across ecologically similar sites. This approach controls for the effects of selection and provides a

96   powerful test of neutral evolutionary processes (Hanson *et al.*, 2012). We constructed a taxon-

97   specific isolate collection of *Streptomyces* found in ecologically similar grassland sites spanning

98   the United States of America. In a study of *Streptomyces* species-level diversity, which evaluated

99   polymorphism at a single locus (*rpoB*), we observed evidence for dispersal limitation manifesting

100 in a latitudinal diversity gradient (Andam *et al.*, 2016b). The latitudinal diversity gradient is one

101 of the earliest and most well documented patterns of biogeography (Wallace, 1878; Hillebrand,

102 2004), but it is essentially undocumented in terrestrial bacterial systems. The hypothesis of a

103 latitudinal gradient in species diversity makes the prediction that latitudinal patterns of gene flow

104 and genetic diversity should also vary *within* populations. Here we evaluate the hypothesis of a

105 latitudinal diversity gradient in terrestrial *Streptomyces* by employing multilocus sequence analysis

106 (MLSA) to characterize population-level patterns of gene flow and genetic diversity for six

107 phylogroups of *Streptomyces* which are distributed widely across the United States.

108

109 **Results**

110 *Characterization of Streptomyces phylogroups*

111 We identified a total of 308 isolates representing the six targeted phylogroups, and these isolates

112 spanned 13 sites (Table 1). Strains within phylogroups share 99.4-100% SSU rRNA gene sequence

113 identity, and all strains share greater than 97% nucleotide identity in their SSU rRNA gene

114 sequences (Figure S1). Strains within phylogroups also share 97.6%-99.5% average nucleotide

115   identity (ANI) across concatenated MLSA loci, and all strains share greater than 88% ANI (Figure

116   1, Table 2). Phylogroup genetic ancestry is congruent with boundaries defined by nucleotide

117   cutoffs (Figure S2), and this indicates that these phylogroups approximate biological populations.

118   We observed 122 unique MLSA haplotypes, with each phylogroup represented by 13-26

119   haplotypes (Figure 1, Table 2). Good's coverage for each of the six phylogroups ranges from 0.88-

120   1.0 for individual loci and 0.94-1.0 for concatenated MLSA loci clustered at 99% nucleotide

121   identity (Table S1). Hence, allelic diversity and nucleotide diversity are well sampled (Table S1),

122   though unique haplotypes remain under sampled (Figure S3). The per site nucleotide diversity (π)

123   of each phylogroup ranges from 0.0026 to 0.011 (Table 2).

124

125   Four of the phylogroups do not match any *Streptomyces* species described in PubMLST (less than

126   97% ANI across MLSA loci, Figure S4A) (Jolley *et al.*, 2004). While the other two, phylogroups

127   WA1063 and MS152 belong to the *S. griseus* species cluster (Rong and Huang, 2010). Isolates in

128   phylogroup WA1063 share greater than 99% MLSA identity with *S. anulatus* and *S. praecox* and

129   form a monophyletic clade with these type strains (Figure S4A). Isolates in phylogroup MS152

130   share greater than 98% MLSA identity with *S. mediolani*, *S. albovinaceus*, and *S. griseinus* and

131   form a paraphyletic clade that includes these type strains (Figure S4A).

132

133   We find evidence of horizontal gene transfer consistent with previous observations of *Streptomyces*

134   species (Doroghazi and Buckley, 2010; Andam *et al.*, 2016a). There is significant phylogenetic

135   incongruence between MLSA loci (Figure S4, Table S2), suggesting that horizontal gene transfer

136   has shaped the phylogeny of these groups. The six phylogroups exhibit evidence of population

137   structure and admixture, with each phylogroup composed of 3.2 ± 0.8 (mean $\pm$ s.d.) subpopulations

7

138   (Figure 1). Evidence of admixture suggests horizontal gene transfer within phylogroups and is

139   consistent with previous evidence of gene exchange in *Streptomyces* (Doroghazi and Buckley,

140   2010). Evidence of recombination within phylogroups MAN196, MS200, MS152, and F34 (PHI

141   test, $p < 0.05$, Table 2) further supports the conclusion of gene exchange within populations. In

142   addition, phylogroups WA1063 and MS152 share two identical *atpD* alleles (Figure S4B), but it is

143   not clear whether these alleles are shared as a result of contemporary horizontal gene transfer or

144   vertical inheritance from the most recent common ancestor of the two clades. The latter

145   explanation is more parsimonious given the low level of polymorphism between the phylogroups

146   WA1063 and MS152.

147

148   *Evidence for dispersal limitation*

149   Strains of the six phylogroups were obtained from soil samples from 13 sites of diverse geographic

150   origin (Table 1). Each phylogroup was detected in 4.2 ± 0.4 sites, and this distribution differs

151   significantly from expectations for a random distribution of strains across sites (permutation test,

152    $p < 0.0005$), thereby rejecting the hypothesis of panmixia. Each phylogroup subpopulation was

153   observed in 2.2 ± 0.9 sites (mean $\pm$ s.d.), and this value is lower than expected if subpopulations

154   are randomly distributed across the sites occupied by each phylogroup (permutation test, $p <$

155   0.001). These results indicate that phylogroup distribution is constrained geographically and that

156   phylogroups have subpopulation structure that is also geographically explicit.

157

158   The geographic distribution of *Streptomyces* allelic diversity indicates dispersal limitation.

159   Identical alleles are shared among phylogroup members across each phylogroup's geographic

160   range, which can exceed 5,000 km (Figure 2, Table 1). However, dissimilarity in allele composition

161   increases with geographic distance, and this result is significant (Bray-Curtis dissimilarity, Mantel

162   r = 0.29, *p* = 0.005) (Figure S5). Hence, alleles are more likely shared between geographically

163   similar sites indicating dispersal limitation with potential for long range dispersal. This result is

164   significant for all individual loci except *recA* (Bray-Curtis dissimilarity, *aptD* Mantel r = 0.31, *p* =

165   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,

166   p = 0.047). In addition, all MLSA haplotypes (Figure 1, Figure 3) are unique to a single site, with

167   the sole exception being a haplotype from phylogroup MS200 which is observed in both MS and

168   WI (Sun Prairie).

169

170   Analysis of haplotype distribution is consistent with diversification due to dispersal limitation.

171   Nested clade analysis (NCA) establishes significant phylogeographic inferences for phylogroups

172   MAN196, MAN125, WA1063, and MS152 (Figure 3) but not for MS200 and F34. Nested clade

173   phylogeographic inference postulates potential evolutionary and historical demographic processes

174   that support extant patterns of diversity and biogeography. For instance, population subdivision of

175   MAN125 across the Pacific Northwest (Figure 3A) and MAN196 between Maine and the Pacific

176   Northwest (Figure 3B) is consistent with restricted gene flow due to historical long distance

177   dispersal events. Likewise, population subdivision of MS152 and between the Southeast (MS and

178   FL) and CA (Figure 3C) and WA1063 between WI (Brookfield) and OR (Figure 3D) is consistent

179   with allopatric fragmentation.

180

181   *Latitudinal diversity gradient*

182   The distribution and diversity of the phylogroups reveals a latitudinal diversity gradient. Strains

183   from MAN125, MAN196, WA1063 occur mostly north of 40°N latitude, while strains from

9

184 MS200, MS152, F34 occur mostly south of this latitude (Table 1, Figure 2). This pattern of

185 North/South partitioning is significant for each phylogroup when evaluated against the expectation

186 of a random distribution across sites (permutation test, $p < 0.01$ for each phylogroup after

187 Bonferroni correction). Furthermore, partial Mantel tests were performed to evaluate the latitudinal

188 and longitudinal vector components of geographic distance in relation to the allele composition of

189 sites. There remains a significant relationship between allele composition and geographic distance

190 when we control for longitude, (Mantel r = 0.23, $p = 0.022$), but this relationship is no longer

191 significant when we control for latitude (Mantel r = 0.15, $p = 0.12$). This result indicates that allele

192 composition changes more across latitude than it does across longitude. The latitudinal partitioning

193 of alleles can be readily observed in the pattern of allele sharing between sites (Figure 2). Finally,

194 we also observed a significant relationship between per site nucleotide diversity of phylogroup

195 MLSA loci and the average latitude of sites in which they are found (R = -0.91, $p = 0.012$; Figure

196 4). This result indicates phylogroups recovered from lower latitudes have higher genetic diversity

197 than those recovered from higher latitudes.

198

Discussion

200 We used population genetic approaches to analyze spatial patterns of genetic diversity for six

201 *Streptomyces* phylogroups isolated from geographically disparate but ecologically similar sites

202 across the United States (Table 1). The distribution of phylogroups is nonrandom and dispersal

203 limited (Figure S5) with phylogroups inhabiting geographic ranges defined by latitude (Table 1,

204 Figure 2). In addition, the genetic diversity of phylogroups is inversely proportional to the latitude

205 from which they were isolated (Figure 4). These findings suggest that there are latitudinal barriers

10

206   to dispersal, and that patterns of *Streptomyces* biogeography result from dispersal limitation and

207   regional diversification due to genetic drift.

208

209   The phylogroups we describe are coherent phylogenetic groups that approximate biological

210   populations (Figure S2). The members of each phylogroup share a distinguishable geographic

211   range (Table 1, Figure 2), a recent common ancestor (Figure 1), and greater than 97% ANI across

212   MLSA loci (Table 2). Despite geographic and genetic subpopulation structure (Figure 1),

213   phylogroup members frequently share identical alleles across demes. It is worth noting that since

214   all of the strains examined share > 97% SSU rRNA gene identity, these geographic patterns would

215   not be detected through analysis of rRNA genes (Figure S1). We infer the presence of

216   recombination within all phylogroups using both nucleotide polymorphism and phylogenetic

217   methods (Table 2, Table S2).

218

219   Regional patterns of biogeography among phylogroups are consistent with limitations to dispersal

220   and gene flow. Allopatric processes like genetic drift can drive diversification between populations

221   that are geographically isolated. The geographic distribution of our phylogroups is nonrandom,

222   and we find regional subpopulation structure within phylogroups (Figure 1, Figure 3). Although

223   we find identical alleles in sites thousands of kilometers apart (Figure 2, Figure S4), MLSA

224   haplotypes are not shared across sites (with the single exception of a haplotype shared between

225   MS and WI) (Figure 3). We also observe a significant distance decay relationship for MLSA allele

226   composition and geographic distance between sites (Figure S5). This data implies that while gene

227   flow is moderate across the geographic range of a phylogroup, dispersal limitation and genetic

228   drift create appreciable regional population structure.

11

229

230   *Streptomyces* phylogroup diversity is consistent with a latitudinal diversity gradient. We find that

231   latitude is a significant predictor of gene flow (Figure 2). Furthermore, intra-phylogroup nucleotide

232   diversity has a significant negative relationship with average latitude (Figure 4), which is

233   congruous with the latitudinal diversity gradient observed for diverse macroorganisms (Hillebrand,

234   2004). There is conflicting evidence for latitudinal diversity gradients among microorganisms.

235   Evidence for microbial latitudinal diversity gradients comes from marine systems (Fuhrman *et al*.,

236   2008; Sul *et al*., 2013; Swan *et al*., 2013), with contrary evidence obtained in terrestrial systems

237   (Neufeld and Mohn, 2005; Chu *et al*., 2010). However, most analyses of terrestrial bacterial

238   biogeography are derived from analyses of SSU rRNA genes, and we show that analyses of SSU

239   rRNA genes lack the sensitivity needed to detect the biogeographic patterns that we observe for

240   *Streptomyces* (Figure S1).

241

242   Several hypotheses have been advanced to explain the formation of latitudinal diversity gradients

243   (Wiens and Donoghue, 2004; Mittelbach *et al*., 2007). Ecological hypotheses posit that factors

244   such as carrying capacity, productivity, and niche availability vary across latitude and that these

245   factors impose constraints on biodiversity (Currie *et al.*, 2004; Mouchet *et al.*, 2015). Evolutionary

246   hypotheses invoke the positive relationship between temperature and the kinetics of metabolism

247   to predict that evolutionary rates and cladogenesis correspond with temperature (Allen *et al*.,

248   2002). Historical hypotheses propose that the latitudinal diversity gradient is the product of

249   historical geological, ecological, or demographic events that have influenced dispersal and

250   diversification (Wiens and Donoghue, 2004; Stevens, 2006). For example, the influence of

251   Pleistocene glacial events on the biogeography of diverse species of terrestrial and aquatic plants

252   and animals is well documented (Soltis *et al.*, 1997; Bernatch and Wilson, 1998; Conroy and Cook,

253   2000; Milá *et al.*, 2006; Maggs *et al.*, 2008; Wilson and Veraguth, 2010). There is also evidence

254   that Pleistocene glaciation events have impacted both microbial communities (Eisenlord *et al.*,

255   2012) and populations (Kenefic *et al.*, 2009; Mikheyev *et al.*, 2008).

256

257   The biogeography of our *Streptomyces* phylogroups is consistent with historical demography and

258   dispersal limitation resulting in a contemporary latitudinal diversity gradient. For instance,

259   MAN125 is nearly in linkage equilibrium ($I_A = 0.09$) though its members span a geographic range

260   of over 2,000 km across the Pacific Northwest. Similar patterns of recombination have also been

261   observed in a *S. pratensis* population which spanned 1,000 km across sites present in North

262   Carolina and northern New York (Doroghazi and Buckley, 2010 and 2014). The observation of

263   linkage equilibrium indicates that either the population lacks contemporary barriers to gene flow

264   or the population has experienced a recent demographic expansion. Coupled with evidence of

265   limitations to gene flow at regional scales (Figure 2, Figure S5), the most parsimonious explanation

266   for these conflicting observations is a recent demographic range expansion. Conversely, latitudinal

267   gradients in marine bacterial diversity have been attributed to the increased kinetics of metabolism

268   which increase the evolutionary tempo in tropical latitudes (Fuhrman *et al.*, 2008). Since latitude

269   is correlated with temperature, we cannot dismiss the influence of kinetics as a cause of the

270   latitudinal diversity gradient in terrestrial *Streptomyces*. Furthermore, it is possible that

271   unappreciated ecological variables, such as the species composition of perennial grass

272   communities, could shape the diversity gradient. Ultimately, it is likely that latitudinal diversity

273   gradients can arise from a combination of ecological, evolutionary, and historical processes that

274   vary in their relative influence with respect to different species and different habitats.

13

275

276   Evidence for a latitudinal diversity gradient coupled with evidence of demographic range

277   expansion, while not conclusive, is consistent with predictions of phylogenetic niche conservatism

278   (Wiens and Donoghue, 2004; Stevens, 2006) which posits that these biogeographical patterns

279   result from the historical effects of changes in paleoclimate (Hewitt *et al*., 1996, 2000, and 2004).

280   The genetic consequences of glaciation events are described in depth by Hewitt (1996, 2000, and

281   2004), and the population structure we observe in our *Streptomyces* phylogroups is consistent with

282   post-glacial demographic range expansion. One of the expectations of post-glacial expansion is

283   "southern richness versus northern purity" (Hewitt, 2004). This is evident in the negative

284   correlation we observe between latitude and the nucleotide diversity of phylogroups (Figure 4).

285   Williams *et al*. (1998) justifies 40°N latitude as approximating late Pleistocene glacial and non-

286   glacial regions with respect to species distributions in North America. Hence, the latitudinal

287   delineation of allele distributions for *Streptomyces* phylogroups roughly corresponds to the extent

288   of ice coverage during the late Pleistocene (Figure 2), which suggests historical population

289   expansion from lower to higher latitudes.

290

291   Haplotype distributions of phylogroups MAN125, MAN196, WA1063, and MS152 are consistent

292   with allopatric diversification resulting from dispersal limitation (Figure 3). Haplotype nested

293   clade analysis (NCA) predicts that historical dispersal events and range expansion across northern

294   regions resulted in limits on dispersal during intermediate timescales, allowing genetic drift to

295   create the phylogroup population structures observed today. There is moderate criticism of NCA

296   (Knowles and Maddison, 2002; Nielsen and Beumont, 2009) due to the subjective nature of

297   inferring historical processes and the wide potential for stochastic processes creating similar

14

298   patterns of biogeography. Yet these tools can provide useful hypotheses. Northern phylogroups

299   MAN125 and MAN196 share a common ancestor with southern phylogroups MS200 and F34

300   (Figure 1). The contemporary population structure of MAN125 and MAN196 is consistent with a

301   historical range expansion from a common ancestor shared by both clades (Figure 1, Figure 3).

302   Further population structure within each phylogroup likely resulted from barriers to gene flow and

303   historical dispersal events across the Pacific Northwest for MAN125 (Figure 3A) and between

304   Maine and the Pacific Northwest for MAN196 (Figure 3B). Analysis of haplotype distribution

305   within WA1063 and MS152 is also consistent with diversification of populations as a result of

306   gene flow limitation between the Midwest and Pacific Northwest for WA1063 (Figure 3D) and

307   between the Southeast and West for MS152 (Figure 3C).

308

309   Phylogroups WA1063 and MS152 share a recent common ancestor (Figure 1, Figure S2) and also

310   share two identical alleles at the *atpD* locus (Figure S4B). These phylogroups have distinct, non-

311   overlapping geographic ranges with WA1063 found in higher latitudes and MS152 in lower

312   latitudes (Table 1, Figure 2). WA1063 and MS152 have 0.0205 net nucleotide substitutions per

313   site across concatenated MLSA loci. We evaluate the possible time range for divergence between

314   WA1063 and MS152 by extrapolating very roughly from the nucleotide substitution rate (μ =

315   4.5×10^{-9}) and generation time (100-300 generations per year) for *E. coli* (Ochman *et al.*, 1999),

316   since corresponding values are not available for *Streptomyces* or their relatives. Based upon these

317   gross approximations, we would estimate that WA1063 and MS152 diverged 15,000-50,000 years

318   ago, corresponding to events in the late Pleistocene (Clayton *et al.*, 2006). Hence, it is likely that

319   the identical *atpD* alleles found in both WA1063 and MS152 were inherited from a shared

320   ancestral population (Figure S3B) as opposed to inheritance from contemporary gene exchange.

15

321

322   Through population genetic analysis of six *Streptomyces* phylogroups, we find evidence for

323   dispersal limitation associated with geographically explicit patterns of gene flow which manifest

324   in a latitudinal gradient of nucleotide diversity. Furthermore, these data support the hypothesis that

325   historical demographic processes influence the contemporary biogeography of *Streptomyces*. Due

326   to their spore forming capabilities and potential for long range dispersal, *Streptomyces* are an ideal

327   system for assessing limits on gene flow among terrestrial bacteria. Future research should seek to

328   determine the degree to which dispersal limitation is due to limits on spore mobility, density

329   dependent blocking, or a result of adaptive constraints. A better understanding of *Streptomyces*

330   biogeography and the evolutionary forces that govern *Streptomyces* diversification may ultimately

331   assist in the discovery of novel genetic diversity and possibly novel antibiotics within this genus.

332

333   **Experimental Procedures**

334   *Strain isolation and DNA extraction*

335   We assembled a culture collection of more than 1,000 *Streptomyces* from 15 sites across the United

336   States with soil sampled at 0-5 cm depth (Andam *et al.,* 2016b). Sites were selected to represent a

337   narrow range of ecological characteristics including meadow, pasture, or native grasslands

338   dominated by perennial grasses and having moderately acidic soil (pH: 6.0 ± 1.0; ave. $\pm$ s.d.).

339   Strains were isolated using uniform conditions and this will select for strains having similar

340   physiological characteristics. The analysis of physiologically similar strains from ecologically

341   similar sites improves our ability to detect biogeographical patterns that result from drift by

342   minimizing the importance of selection (as reviewed by Hanson *et al.*, 2012). Soil was air dried,

343   and *Streptomyces* strains were isolated on glycerol-arginine agar plates of pH 8.7 containing

16

344   cycloheximide and Rose Bengal (El-Nakeeb and Lechevalier, 1963; Ottow, 1972) as previously

345   described (Doroghazi and Buckley, 2010). Genomic DNA was extracted for each isolate from

346   purified cultures, which were grown by shaking at 30˚C in liquid yeast extract-malt extract medium

347   (YEME) containing 0.5% glycine (Kieser *et al.*, 2000), by using a standard

348   phenol/chloroform/isoamyl alcohol protocol (Roberts and Crawford, 2000).

349

350   The gene encoding the RNA polymerase beta-subunit (*rpoB*) provides a robust, advantageous

351   alternative to the SSU rRNA locus for phylogenetic analyses of the genera *Streptomyces* (Kim *et*

352   *al.*, 2004). We previously assessed genetic diversity of our culture collection using partial *rpoB*

353   sequences clustered at 0.01 patristic distances with RAMI (Pommier *et al.*, 2009), and using this

354   approach we identified 107 species-like phylogenetic clusters, or phylogroups (Andam *et al.*,

355   2016b). We selected six of these phylogroups that represented the geographic and phylogenetic

356   diversity of our culture collection for further analysis. MLSA was performed on 17-47 isolates

357   from each phylogroup. Phylogroup names are capitalized and based on a representative isolate; for

358   example, phylogroup WA1063 is named for *Streptomyces* sp. wa1063. Isolates are identified with

359   a lowercase letter code indicating site of origin followed by a strain number; for example, isolate

360   wa1063 is strain 1063 isolated from Washington (WA) state.

361

362   *Multilocus sequence analysis (MLSA)*

363   We adapted the MLSA scheme developed for *Streptomyces* by Guo *et al.* (2008), which targets the

364   five housekeeping loci *atpD*, *gyrB*, *recA*, *rpoB*, and *trpB* as described in Doroghazi and Buckley

365   (2010). The V2 and V2 regions of SSU rRNA sequences were amplified using the universal primer

366   8F: AGAGTTTGATCCTGGCTCAG. Reactions for Sanger sequencing were performed using

17

367   forward primers for all loci except *rpoB*, for which the reverse primer was used. Trace files were

368   uniformly screened using CAP3 (Huang and Madan, 1999) interfaced with a Perl script. Sequences

369   were aligned using MUSCLE (Edgar, 2004), trimmed to 431 bp, 415 bp, 446 bp, 557 bp, and 489

370   bp, for genes *atpD*, *gyrB*, *recA*, *rpoB*, *trpB*, respectively, and concatenated consistent with the

371   genomic order in *Streptomyces coelicolor* A3(2). SSU rRNA sequences were trimmed to 357 bp,

372   creating an alignment spanning the V1 and V2 regions. Gene sequences are available on GenBank

373   with accession numbers XXXX-XXXX.

374

375   Good's coverage estimation and haplotype rarefaction was determined using mothur (Schloss *et*

376   *al.*, 2009). DNA polymorphism statistics including: number of segregating sites, nucleotide

377   diversity, and Tajima's D, were determined with DnaSP v5 (Librado and Rozas, 2009) and LDhat

378   (McVean *et al.*, 2002). Population scaled mutation rates (Watterson's theta; $\theta_w = 2N_e\mu$),

379   recombination or gene conversion rates ($\rho = 2N_e r/2$), and relative rates of recombination (ρ/θ_w) were

380   estimated using LDhat (McVean *et al.*, 2002) and are expressed per nucleotide site. The standard

381   index of association (I_A) was calculated from allelic data with LIAN v3.5 (Haubold and Hudson,

382   2000) using the Monte-Carlo test and 100 iterations. The pairwise homoplasy index (PHI) statistic

383   was determined using PhiPack (Bruen *et al.*, 2006), and statistical significance was evaluated

384   under a null hypothesis of no recombination. Sequence identity across phylogroups was calculated

385   with mothur (Shloss *et al.*, 2009)

386

387   *Phylogengy reconstruction*

388   Maximum likelihood (ML) trees were constructed from the nucleotide sequences of individual and

389   concatenated MLSA loci using the generalized time reversible nucleotide substitution model

390 (Tavaré, 1986) with gamma distributed rate heterogeneity among sites (GTRGAMMA) supported

391 in RAxML v7.3.0 (Stamatakis, 2006). Bootstrap support was determined for the highest-scoring

392 ML tree of 20 iterations, and the number of bootstrap replicates was determined using the extended

393 majority rule (autoMRE) convergence criteria (Pattengale *et al.*, 2010). Root placement is defined

394 by *Mycobacterium smegmatis*. Significant phylogenetic incongruence between loci was

395 determined using the Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999) implemented

396 in the R package phangorn (Schliep, 2011).

397

398 *Population structure*

399 Concatenated MLSA sequences were analyzed using Structure v2.3.3 (Pritchard *et al.*, 2000) to

400 examine population affiliation, subdivision, and admixture within and between phylogroups.

401 Structure was run using an admixture model with a burn-in length of $1.0E^6$ and $3.0E^6$ replicates.

402 The most probable number of sub-populations (k) was evaluated with 10 independent runs and

403 chosen using the Evanno method (Evanno *et al.*, 2005), with k = 1 through k = 6 within

404 phylogroups and through k = 8 between phylogroups, implemented by Structure Harvester (Earl

405 and vonHoldt, 2012). After choosing the most probable k-value, the program Clumpp was used to

406 permute outputs of the independent runs (Jakobsson and Rosenberg, 2007).

407

408 *Patterns of dispersal and gene flow*

409 We used permutation tests to evaluate whether phylogroup distribution across sites could be

410 explained by panmixia. We compared the observed distributions to those expected under a random

411 distribution model using 1,000 permutations to assess significance. In addition, correlations

412 between geographic distance and allele composition between sites were assessed using Mantel and

19

413   partial Mantel tests (Mantel, 1967; Smouse *et al*., 1986). These tests were performed with the R

414   package ecodist (Goslee and Urban, 2007) using the Pearson correlation method and 1,000

415   permutations. Bray-Curtis dissimilarity was calculated from unique allele composition across sites.

416

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

418   Templeton *et al*., 1995) implemented in TCS v1.18 (Clement *et al*., 2000). Nested clade

419   information was used to infer processes that could explain the geographic and genetic distribution

420   of sequences using the program GeoDis v2.2 (Posada *et al*., 2000). Both TCS v1.18 and GeoDis

421   v2.2 were performed in ANeCA (Panchal, 2007).

422

423   **Acknowledgements**

427

428   **References**

429   Allen AP, Brown JH, Gillooly JF. (2002) Global biodiversity, biochemical kinetics, and the

430   energetic-equivalence rule. Science **297**: 1545–1548.

431

432   Andam CP, Choudoir MJ, Nguygen AVT, Park HS, Buckley DH. (2016a) Contributions of

433   ancestral inter-species recombination to the genetic diversity of extant *Streptomyces* lineages.

434   ISMEJ In Press. doi: 10.1038/ismej.2015.230.

435

436 Andam CP, Doroghazi JR, Campbell AN, Kelly PJ, Choudoir MJ, Buckley DH. (2016b). A

437 latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*. mBio In Press.

438

439 Antony-Babu S, Stach JEM, Goodfellow M. (2008) Genetic and phenotypic evidence for

440 *Streptomyces griseus* ecovars isolated from a beach and dune sand system. Antonie Van

441 Leeuwenhoek **94**: 63–74.

442

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

444 1365.

445

446 Bernatchez L, Wilson CC. (1998) Comparative phylogeography of Nearctic and Palearctic

447 fishes. Mol Ecol **7**: 431–452.

448

449 Bissett A, Richardson AE, Baker G, Wakelin S, Thrall PH. (2010) Life history determines

450 biogeographical patterns of soil bacterial communities over multiple spatial scales. Mol Ecol **19**:

451 4315–4327.

452

453 Boucher Y, Cordero OX, Takemura A, Hunt DE, Schliep K, Bapteste E, *et al*. (2011) Local

454 mobile gene pools rapidly cross species boundaries to create endemicity within global *Vibrio*

455 *cholerae* populations. mBio **2**: e00335–10.

456

457 Bruen TC, Philippe H, Bryant D. (2005) A simple and robust statistical test for detecting the

458 presence of recombination. Genetics **172**: 2665–2681.

21

459

460 Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, *et al.* (2012) Patterns of

461 gene flow define species of thermophilic Archaea. PLoS Biol **10**: e1001265.

462

463 Cho J, Tiedje JM. (2000) Biogeography and degree of endemicity of fluorescent *Pseudomonas*

464 strains in soil. Appl Environ Microbiol **66**: 5448-5456.

465

466 Choudoir MJ, Campbell AN, Buckley DH. (2012) Grappling with Proteus: population level

467 approaches to understanding microbial diversity. Front Microbiol **3**: 336.

468

469 Chu H, Fierer N, Lauber CL, Caporaso JG, Knight R, Grogan P. (2010) Soil bacterial diversity in

470 the Arctic is not fundamentally different from that found in other biomes. Environ Microbiol **12**:

471 2998–3006.

472

473 Clayton L, Attig JW, Mickelson DM, Johnson MD, Syverson KM. (2006) Glaciation of

474 Wisconsin: Wisconsin Geological and Natural History Survey Educational Series 36. Wisconsin

475 Geological and Natural History Survey. Madison, WI.

476

477 Clement M, Posada D, Crandall KA. (2000) TCS: A computer program to estimate gene

478 genealogies. Mol Ecol **9**: 1657–1659.

479

480 Coleman ML, Chisholm SW. (2010) Ecosystem-specific selection pressures revealed through

481 comparative population genomics. Proc Natl Acad Sci U S A **107**: 18634–18639.

482

483 Conroy CJ, Cook JA. (2000) Phylogeography of a post-glacial colonizer: *Microtus longicaudus*

484 (Rodentia: Muridae). Mol Ecol **9**: 165–175.

485

486 Currie DJ, Mittelbach GG, Cornell HV, Field R, Guégan JF, Hawkins BA*, et al.* (2004)

487 Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness.

488 Ecol Lett **7**: 1121–1134.

489

490 Davelos AL, Xiao K, Samac DA, Martin AP, Kinkel LL. (2004a) Spatial variation in

491 *Streptomyces* genetic composition and diversity in a prairie soil. Microb Ecol **48**: 601–612.

492

493 Davelos AL, Kinkel LL, Samac DA. (2004b) Spatial variation in frequency and intensity of

494 antibiotic interactions among *Streptomycetes* from prairie soil. Appl Environ Microbiol **70**:

495 1051–1058.

496

497 Doroghazi JR, Buckley DH. (2014) Intraspecies comparison of *Streptomyces pratensis* genomes

498 reveals high levels of recombination and gene conservation between strains of disparate

499 geographic origin. BMC Genomics **15**: 970.

500

501 Doroghazi JR, Buckley DH. (2010) Widespread homologous recombination within and between

502 *Streptomyces* species. ISME J **4**: 1136–1143.

503

504   Earl DA, vonHoldt BM. (2012) STRUCTURE HARVESTER: a website and program for

505   visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour

506   **4**: 359–361.

507

508   Edgar RC. (2004) MUSCLE: multiple sequence alignment with high accuracy and high

509   throughput. Nucleic Acid Res **32**: 1792–1797.

510

511   Eisenlord SD, Zak DR, Upchurch RA. (2012) Dispersal limitation and the assembly of soil

512   *Actinobacteria* communities in a long-term chronosequence. Ecol Evol **2**: 538–549.

513

514   El-Nakeeb MA, Lechevalier HA. (1963) Selective isolation of aerobic Actinomycetes. Appl

515   Microbiol 11: 75–77.

516

517   Evanno G, Regnaut S, Goudet J. (2005) Detecting the number of clusters of individuals using the

518   software STRUCTURE: a simulation study. Mol Ecol **14**: 2611–2620.

519

520   Fierer N, Carney KM, Horner-Devine MC, Megonigal JP. (2009) The biogeography of

521   ammonia-oxidizing bacterial communities in soil. Microb Ecol **58**: 435–445.

522

523   Fierer N, Jackson RB. (2006) The diversity and biogeography of soil bacterial communities. Proc

524   Natl Acad Sci U S A **103**: 626–631.

525

24

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

527   1061–1063.

528

529   Finlay BJ, Fenchel T. (2004) Cosmopolitan metapopulations of free-living microbial eukaryotes.

530   Protist **155**: 237–244.

531

532   Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown M V, Green JL, *et al.* (2008) A

533   latitudinal diversity gradient in planktonic marine bacteria. Proc Natl Acad Sci U S A **105**: 7774–

534   7778.

535

536   Gilbert JA, Steele JA, Caporaso JG, Steinbrück L, Reeder J, Temperton B, *et al.* (2012) Defining

537   seasonal marine microbial community dynamics. ISME J **6**: 298–308.

538

539   Goslee SC, Urban DL. (2007) The ecodist package for dissimilarity-based analysis of ecological

540   data. J Stat Softw **22**: 1–19.

541

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

543   501–507.

544

545   Guo Y, Zheng W, Rong X, Huang Y. (2008) A multilocus phylogeny of the *Streptomyces*

546   *griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete

547   systematics. Int J Syst Evol Microbiol **58**: 149–159.

548

549   Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. (2012) Beyond biogeographic

550   patterns: processes shaping the microbial landscape. Nat Rev Microbiol **10**: 497–506.

551

552   Hatosy SM, Martiny JBH, Sachdeva R, Steele J, Fuhrman JA, Martiny AC. (2013) Beta diversity

553   of marine bacteria depends on temporal scale. Ecology **94**: 1898–1904.

554

555   Haubold B, Hudson RR. (2000) LIAN 3.0: detecting linkage disequilibrium in multilocus data.

556   Bioinformatics **16**: 847–848.

557

558   Hewitt GM. (1996) Some genetic consequences of ice ages, and their role in divergence and

559   speciation. Biol J Linnean Soc **58**: 247–276.

560

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

562

563   Hewitt GM. (2004) Genetic consequences of climatic oscillations in the Quaternary. Philos Trans

564   R Soc Lond B Biol Sci **359**: 183–195.

565

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

567   211.

568

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

570   677.

571

26

572   Jakobsson M, Rosenberg NA. (2007) CLUMPP: a cluster matching and permutation program for

573   dealing with label switching and multimodality in analysis of population structure.

574   Bioinformatics **23**: 1801–1806.

575

576   Jaspers E, Overmann J. (2004) Ecological significance of microdiversity: identical 16S rRNA

577   gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies.

578   Appl Environ Microbiol **70**: 4831–4839.

579

580   Jolley KA, Chan MS, Maiden MCJ. (2004) mlstdbNet - distributed multi-locus sequence typing

581   (MLST) databases. BMC Bioinformatics **5**: 86.

582

583   Kâmpfer P. (2006) The family *Streptomycetaceae* Part I: Taxonomy. In: The Prokaryotes,

584   Volume 3. Dworkin, M, Falkow, S, Rosenberg, E, Schliefer, K, & Stackebrandt, E. (eds). New

585   York: Springer, pp. 538–604.

586

587   Kenefic LJ, Pearson T, Okinaka RT, Schupp JM, Wagner DM, Hoffmaster AR, *et al.* (2009) Pre-

588   Columbian origins for North American anthrax. PLoS One **4**: e4813.

589

590   Kieser T, Bibb M, Buttner M, Charter K, Hopwood D. (2000) Practical *Streptomyces* Genetics.

591   Norwich, UK: John Innes Foundation.

592

593   Kim BJ, Kim CJ, Chun J, Koh YH, Lee SH, Hyun JW, *et al.*(2004) Phylogenetic analysis of the

594   genera *Streptomyces* and *Kitasatospora* based on partial RNA polymerase beta-subunit gene

595   (*rpoB*) sequences. Int J Syst Evol Microbiol. **54**: 593–598.

596

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

598

599   Labeda DP. (2011) Multilocus sequence analysis of phytopathogenic species of the genus

600   *Streptomyces*. Int J Syst Evol Microbiol **61**: 2525–2531.

601

602   Lauber CL, Hamady M, Knight R, Fierer N. (2009) Pyrosequencing-based assessment of soil pH

603   as a predictor of soil bacterial community structure at the continental scale. Appl Environ

604   Microbiol **75**: 5111–5120.

605

606   Librado P, Rozas J. (2009) DnaSP v5: a software for comprehensive analysis of DNA

607   polymorphism data. Bioinformatics **25**: 1451–1452.

608

609   Loria R, Kers J, Joshi M. (2006) Evolution of plant pathogenicity in *Streptomyces*. Annu Rev

610   Phytopathol **44**: 469–487.

611

612   Lozupone CA, Knight R. (2007) Global patterns in bacterial diversity. Proc Natl Acad Sci U S A

613   **104**: 11436–11440.

614

615   Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, *et al.* (2008) Evaluating

616   signatures of glacial refugia for North Atlantic benthic marine taxa. Ecology **89**: S108–S122.

617

618   Mantel N. (1967) The detection of disease clustering and a generalized regression approach.

619   Cancer Research **27**: 209–220.

620

621   Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, *et al.* (2006)

622   Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol **4**: 102–112.

623

624   Martiny JBH, Eisen JA, Penn K, Allison SD, Horner-Devine MC. (2011) Drivers of bacterial β -

625   diversity depend on spatial scale. Proc Natl Acad Sci U S A **108**: 7850–7854.

626

627   McCarthy AJ, Williams ST. (1992) Actinomycetes as agents of biodegradation in the

628   environment - a review. Gene **115**: 189–192.

629

630   McVean G, Awadalla P, Fearnhead P. (2002) A coalescent-based method for detecting and

631   estimating recombination from gene sequences. Genetics **160**: 1231–1241.

632

633   Mikheyev AS, Vo T, Mueller UG. (2008) Phylogeography of post-Pleistocene population

634   expansion in a fungus-gardening ant and its microbial mutualists. Mol Ecol **17**: 4480–4488.

635

636   Mila B, Smith T, Wayne R. (2006) Postglacial population expansion drives the evolution of

637   long-distance migration in a songbird. Evolution **60**: 2403–2409.

29

638

639   Miller SR, Strong AL, Jones KL, Ungerer MC. (2009) Bar-coded pyrosequencing reveals shared

640   bacterial community properties along the temperature gradients of two alkaline hot springs in

641   Yellowstone National Park. Appl Environ Microbiol **75**: 4565–4572.

642

643   Mittelbach GG, Schemske DW, Cornell H V., Allen AP, Brown JM, Bush MB, *et al.* (2007)

644   Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecol

645   Lett **10**: 315–331.

646

647   Mouchet M, Levers C, Zupan L, Kuemmerle T, Plutzar C, Erb K, *et al.* (2015) Testing the

648   effectiveness of environmental variables to explain European terrestrial vertebrate species

649   richness across biogeographical scales. PLoS One **10**: e0131924.

650

651   Neufeld JD, Mohn WW. (2005) Unexpectedly high bacterial diversity in arctic tundra relative to

652   boreal forest soils, revealed by serial analysis of ribosomal sequence tags. Appl Environ

653   Microbiol **71**: 5710–5718.

654

655   Nielsen R, Beaumont MA. (2009) Statistical inferences in phylogeography. Mol Ecol **18**: 1034–

656   1047.

657

658   Ochman H, Elwyn S, Moran NA. (1999) Calibrating bacterial evolution. Proc Natl Acad Sci U S

659   A **96**: 12638–12643.

660

661 Ottow JCG. (2013) Rose bengal as a selective aid in the isolation of fungi and Actinomycetes

662 from natural sources. Mycologia **64**: 304–315.

663

664 Panchal M. (2006) The automation of Nested Clade Phylogeographic Analysis. Bioinformatics

665 **23**: 509–510.

666

667 Papke RT, Ramsing NB, Bateson MM, Ward DM. (2003) Geographical isolation in hot spring

668 cyanobacteria. Environ Microbiol **5**: 650–659.

669

670 Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. (2010) How

671 many bootstrap replicates are necessary? J Comput Biol **17**: 337–354.

672

673 Pearson T, Giffard P, Beckstrom-Sternberg S, Auerbach R, Hornstra H, Tuanyok A*, et al*. (2009)

674 Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer.

675 BMC Biol **7**: 78.

676

677 Pommier T, Canback B, Lundberg P, Hagstrom A, Tunlid A. (2009) RAMI: a tool for

678 identification and characterization of phylogenetic clusters in microbial communities.

679 Bioinformatics **25**: 736–742.

680

681 Posada D, Crandall KA, Templeton AR. (2000) GeoDis: A program for the cladistic nested

682 analysis of the geographical distribution of genetic haplotypes. Mol Ecol **9**: 487–488.

683

684   Pritchard JK, Stephens M, Donnelly P. (2000) Inference of population structure using multilocus

685   genotype data. Genetics **155**: 945–959.

686

687   Roberts MA and Crawford DL. (2000) Use of randomly amplified polymorphic DNA as a means

688   of developing genus- and strain-specific *Streptomyces* DNA probes. Appl Environ Microbiol **66:**

689   2555-2564.

690

691   Rong X, Huang Y. (2010) Taxonomic evaluation of the *Streptomyces griseus* clade using

692   multilocus sequence analysis and DNA-DNA hybridization, with proposal to combine 29 species

693   and three subspecies as 11 genomic species. Int J Syst Evol Microbiol **60**: 696–703.

694

695   Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, *et al*. (2010) Soil

696   bacterial and fungal communities across a pH gradient in an arable soil. ISME J **4**: 1340–1351.

697

698   Schliep KP. (2011) phangorn: phylogenetic analysis in R. Bioinformatics **27**: 592–593.

699

700   Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, *et al*. (2009)

701   Introducing mothur: open-source, platform-independent, community-supported software for

702   describing and comparing microbial communities. Appl Environ Microbiol **75**: 7537–7541.

703

704   Shimodaira H, Hasegawa M. (1999) Multiple comparisons of log-likelihoods with applications

705   to phylogenetic inference. Mol Biol Evol **16**: 1114–1116.

706

707   Smouse PE, Long JC, Sokal RR. (1986) Multiple regression and correlation extensions of the

708   Mantel test of matrix correspondence. Syst Zool **35**: 627–632.

709

710   Soltis D, Gitzendanner M, Strenge D, Soltis P. (1997) Chloroplast DNA intraspecific

711   phylogeography of plants from the Pacific Northwest of North America. Plant Syst Evol **206**:

712   353–373.

713

714   Stamatakis A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with

715   thousands of taxa and mixed models. Bioinformatics **22**: 2688–2690.

716

717   Stevens RD. (2006) Historical processes enhance patterns of diversity along latitudinal gradients.

718   *Proc R Soc B Biol Sci* **273**: 2283–2289.

719

720   Sul WJ, Oliver TA, Ducklow HW, Amaral-zettler LA, Sogin ML. (2012) Marine bacteria exhibit

721   a bipolar distribution. Proc Natl Acad Sci U S A **110**: 2342–2347.

722

723   Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM, *et al*. (2013)

724   Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface

725   ocean. Proc Natl Acad Sci U S A **110**: 11463–11468.

726

727   Takasuka TE, Book AJ, Lewin GR, Currie CR, Fox BG. (2013) Aerobic deconstruction of

728   cellulosic biomass by an insect-associated *Streptomyces*. Sci Rep **3**: 1–10.

729

730   Tavaré S. (1986) Some probabilistic and statistical problems in the analysis of DNA sequences.

731   Am Math Soc Lect Math Life Sci **17**: 57–86.

732

733   Telford RJ, Vandvik V, Birks HJB. (2006) Dispersal limitations matter for microbial

734   morphospecies. Science **312**: 1015.

735

736   Templeton AR, Boerwinkle E, Sing CF. (1987) A cladistic analysis of phenotypic associations

737   with haplotypes inferred from restriction endonuclease mapping. i. basic theory and an analysis

738   of alcohol dehydrogenase activity in Drosophila. Genetics **117**: 343–351.

739

740   Templeton AR, Routman E, Phillips CA. (1995) Separating population structure from population

741   history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in

742   the tiger salamander, *Ambystoma tigrinum*. Genetics **140**: 767–782.

743

744   Vos M, Velicer GJ. (2008) Isolation by distance in the spore-forming soil bacterium

745   *Myxococcus xanthus*. Curr Biol **18**: 386–391.

746

747   Wallace A. (1878) Tropical nature, and other essays. London: Macmillan and Co.

748

749   Watve M, Tickoo R, Jog M, Bhole B. (2001) How many antibiotics are produced by the genus

750   *Streptomyces*? Arch Microbiol **176**: 386–390.

751

752   Wawrik B, Kutliev D, Abdivasievna UA, Kukor JJ, Zylstra GJ, Kerkhof L. (2007) Biogeography

753   of Actinomycete communities and type ii polyketide synthase genes in soils collected in New

754   Jersey and Central Asia. Appl Environ Microbiol **73**: 2982–2989.

755

756   Welch RA, Burland V, Plunkett G, Redford P, Roesch P, Rasko D, *et al.* (2002) Extensive

757   mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*.

758   Proc Natl Acad Sci U S A **99**: 17020–17024.

759

760   Whitaker RJ, Grogan DW, Taylor JW. (2003) Geographic barriers isolate endemic populations

761   of hyperthermophilic archaea. Science **301**: 976–978.

762

763   Wiens JJ, Donoghue MJ. (2004) Historical biogeography, ecology and species richness. Trends

764   Ecol Evol **19**: 639–644.

765

766   Williams D, Dunkerley D, DeDecker P, Kershaw P, Chappell M. (1998) Quaternary

767   environments. London: Arnold.

768

769   Wilson AB, Veraguth IE. (2010) The impact of Pleistocene glaciation across the range of a

770   widespread European coastal species. Mol Ecol **19**: 4535–53.

771

772

773

774

775 **Table and Figure Legends**

776

777 **Table 1.** A total of 755 *Streptomyces* strains were isolated from 13 sites, and 308 of these strains

778 were found to represent the six targeted phylogroups. The numbers of isolates per site that

779 belonged to each of our target phylogroups is indicated. Isolate names begin with a letter code

780 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 (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

781

782

783

784

785

786

36

787   **Table 2.** MLSA was performed on a total of 17-47 strains from each phylogroup. Summary

788   statistics for concatenated MLSA nucleotide sequences were determined as described in

789   methods. Phylogroup average latitude (Ave Lat) was determined using the total number of

790   isolates per phylogroup provided in Table 1. The standard index of association is zero for a

791   population in linkage equilibrium.

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**

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
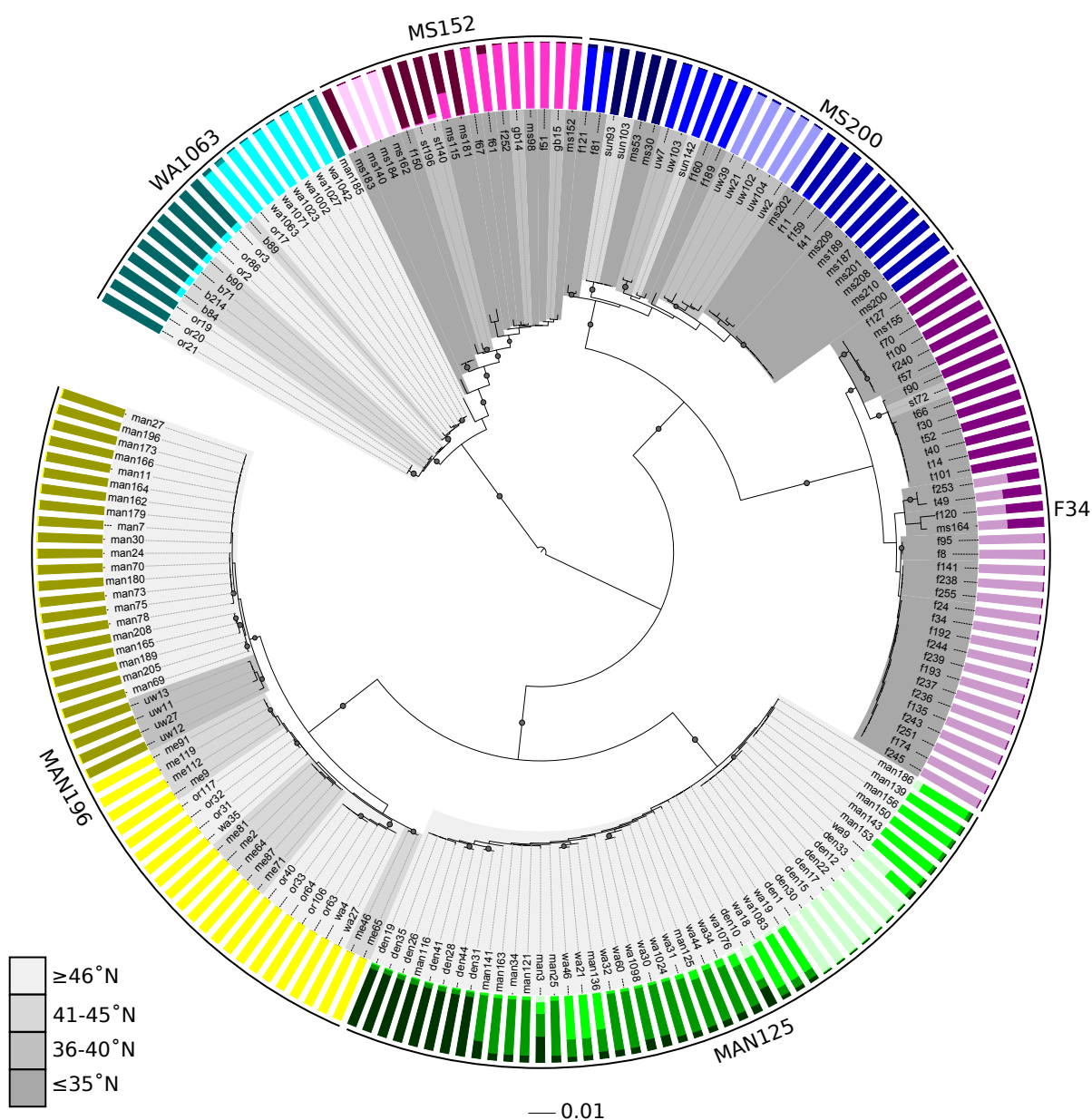π: per site nucleotide diversity
θ_W: per site Watterson's theta ($2N_e\mu$)
ρ: per site rate of recombination/gene conversion ($2N_e r/2$)
ρ/θ_W: relative rate of recombination to mutation
I_A: standard index of association
Φ_W: pairwise homoplasy index (Phi) statistic with *p*-value < 0.05*, 0.01** under the null hypothesis of no recombination

792

793

794

795

796

797

798

799

800

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

38

804   confidences > 80 are indicated with gray circles, and precise bootstrap values are found in Figure

805   S4A. The colored bars in the outer ring indicate genetic contributions from different ancestral

806   populations as inferred by Structure analysis. The shading of the inner ring indicates the latitude

807   from which each strain was isolated according to the scale provided. The isolation site for each

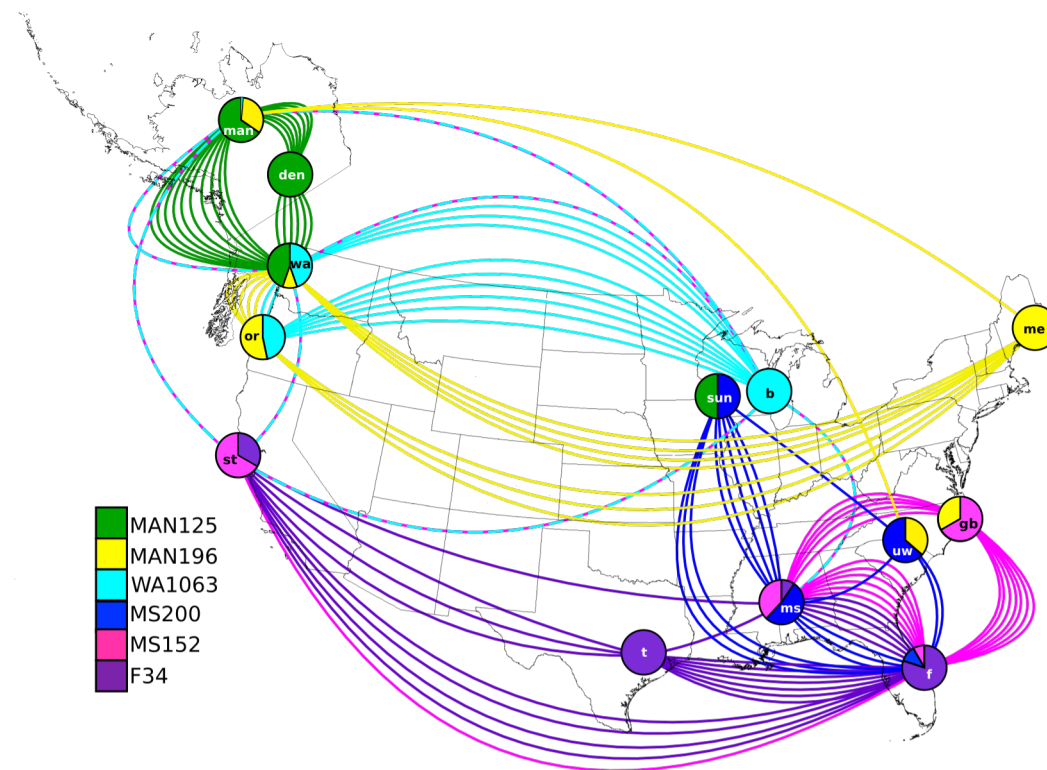808   strain can be determined by isolate names as indicated in Table 1.

809

810   **Figure 2.** Circles depict sample sites and are labeled according to site code (Table 1). The relative

811   abundance of each phylogroup at each site is indicated by color according to the legend, with raw

812   counts provided in Table 1. Solid colored lines represent identical alleles shared by phylogroup

813   members subsampled for MLSA across sites. Dashed multicolored lines depict identical *atpD*

814   alleles shared by strains in phylogroups WA1063 and MS152 (Figure S4B).
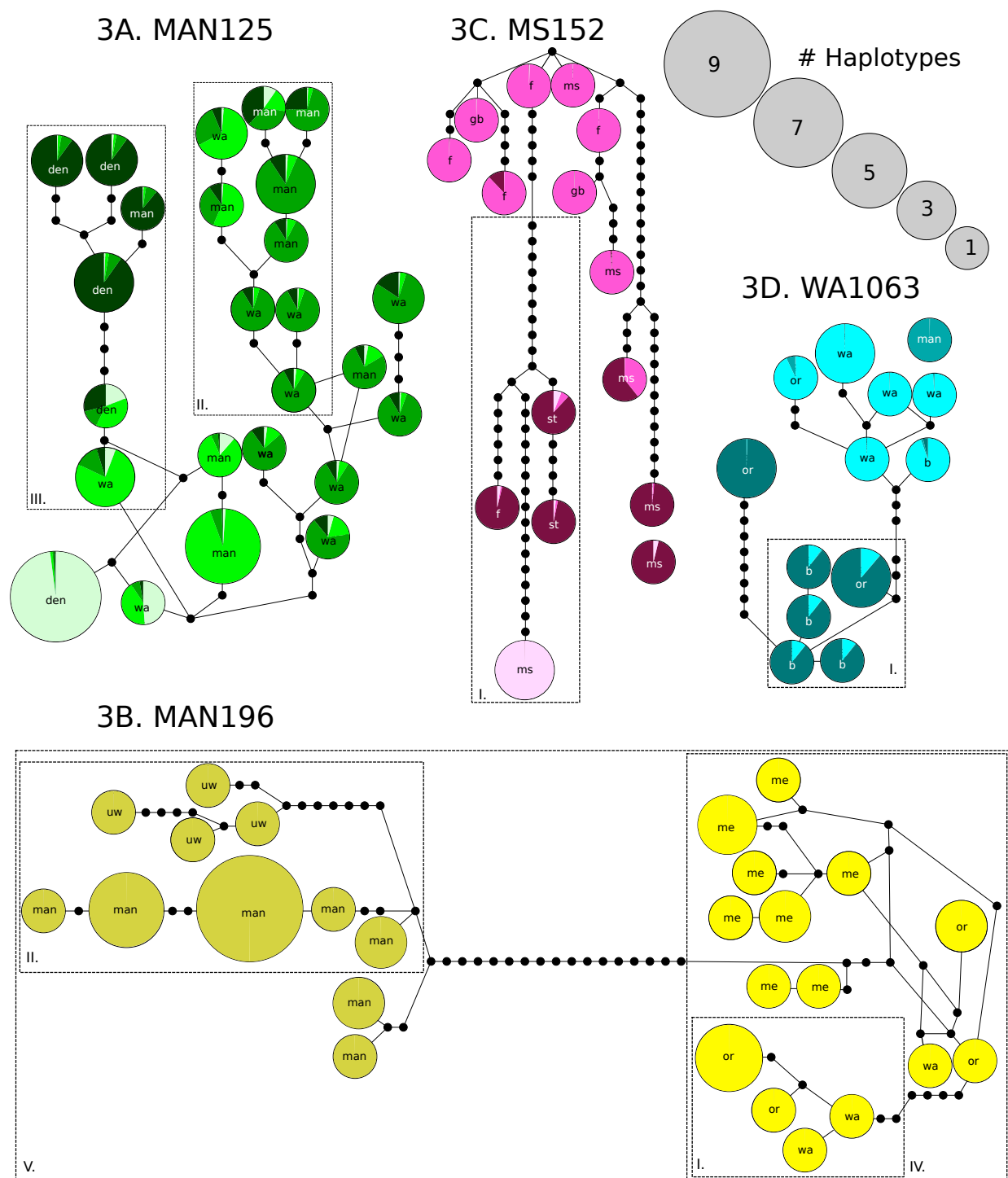
815

39

Figure 3. Haplotype networks for phylogroups MAN125 (3A), MAN196 (3B), MS152 (3C), WA1063 (3D). Circles represent unique MLSA haplotypes whose radius is proportional to the number of isolates sharing that haplotype, and colors correspond to strain ancestry and

820　　subpopulation affiliation as determined by Structure analysis and defined in Figure 1. Haplotypes

821　　are labeled with a letter code referring to sample site location as indicated in Table 1. Black circles

822　　represent un-sampled, inferred haplotypes with each circle designating a single nucleotide

823　　polymorphism. The length of edges between nodes is uninformative. Dashed rectangles encompass

824　　clades that have significant phylogeographic inferences from nested clade analysis, as described

825　　in methods. Roman numerals correspond to the following inferences: I. Allopatric fragmentation;

826　　II. Long distance colonization and/or past fragmentation; III. Restricted gene flow with isolation

827　　by distance; IV. Restricted gene flow but with some long-distance gene flow over intermediate

828　　ranges not occupied by the species; or past gene flow followed by extinction of intermediate

829　　populations; V. Contiguous range expansion.
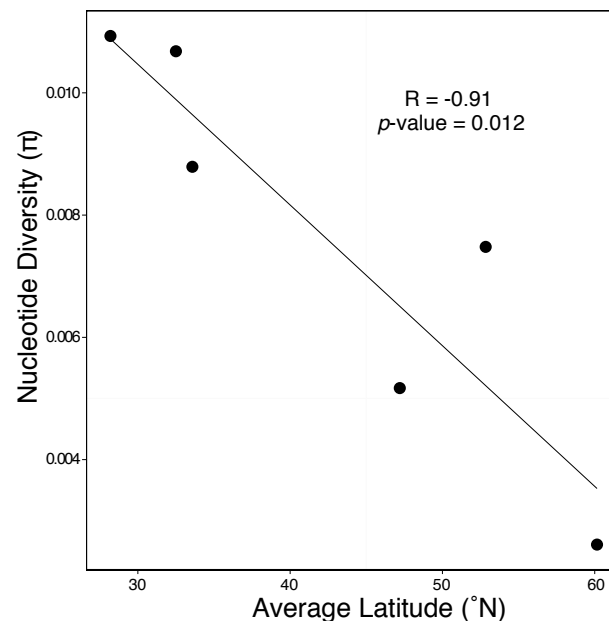

830

831　　**Figure 4.** Average latitude was determined using a weighted average based on the total number of

832　　isolates per site as indicated in Table 1. Nucleotide diversity was calculated using concatenated

833　　MLSA loci and is expressed per site.