The biogeography of *Streptomyces* in New Zealand enabled by high-throughput sequencing of
genus-specific *rpoB* amplicons

Higgins, S.A.^{1,a,*}, Panke-Buisse, K.^{1,b}, and Daniel H. Buckley¹

¹School of Integrative Plant Science, Cornell University, Ithaca, NY 14853 USA

*Correspondence to:

Steven A. Higgins

Boyce Thompson Institute

Room 104

Ithaca, NY 14853

sah382@cornell.edu

Running title: Amplicon sequencing of *Streptomyces rpoB* in New Zealand

^a Current address is Boyce Thompson Institute, Ithaca, NY, USA

^b Current address is USDA Agricultural Research Service, Madison, WI, USA

Summary

We evaluated *Streptomyces* biogeography in soils along a 1,200 km latitudinal transect across New Zealand (NZ). *Streptomyces* diversity was examined using high-throughput sequencing of *rpoB* amplicons generated with a *Streptomyces* specific primer set. We detected 1,287 *Streptomyces rpoB* operational taxonomic units (OTUs) with 159 ± 92 (average $\pm$ s.d.) *rpoB* OTUs per site. Only 12% ($n = 149$) of these OTUs matched *rpoB* sequences from cultured specimens (99% nucleotide identity cutoff). *Streptomyces* phylogenetic diversity (Faith's PD) was correlated with soil pH, mean annual temperature, and plant community richness (Spearman's r : 0.77, 0.64, and -0.79, respectively; $p < 0.05$), but not with latitude. In addition, soil pH and plant community richness both explained significant variation in *Streptomyces* beta diversity. *Streptomyces* communities exhibited both high dissimilarity and strong dominance of one or a few species at each site. Taken together, these results suggest that dispersal limitation due to competitive interactions limits the colonization success of spores that relocate to new sites. Cultivated *Streptomyces* isolates represent a major source of clinically useful antibiotics, but only a small fraction of extant diversity within the genus have been identified and most species of *Streptomyces* have yet to be described.

Introduction

The genus *Streptomyces* consists of a cosmopolitan group of spore-forming microorganisms whose ease of cultivation has facilitated their investigation for over a century (Waksman *et al.*, 1948; Kämpfer *et al.*, 2014). They are aerobic, filamentous members of the gram-positive bacterial phylum *Actinobacteria*, possessing characteristic developmental, biosynthetic, and catabolic genes which have contributed to their success in a diversity of terrestrial and aquatic habitats for approximately 350 myr (Seipke *et al.*, 2012; Kämpfer *et al.*, 2014; McDonald and Currie, 2017). Furthermore, their utility as model organisms in investigations of natural product synthesis, cell biology, and evolution has transformed microbiology and biotechnology research (Kämpfer *et al.*, 2014; Chater, 2016). While their capacity for carbon mineralization (Ruddick and Williams, 1972; Barder and Crawford, 1981; Spiker *et al.*, 1992; Dari *et al.*, 1995; Schlatter *et al.*, 2009) and secondary metabolism (Chater, 2016; Niu *et al.*, 2016; Gallagher *et al.*, 2017; Choudoir *et al.*, 2018) have been well-documented, much less is known regarding their diversity, ecology, and biogeography in terrestrial habitats. To date, biogeographical studies of the *Streptomyces* have largely focused on genetic variation within cultivated representatives to inform population level analyses (Davelos *et al.*, 2004; Antony-Babu *et al.*, 2008; Andam *et al.*, 2016). While instructive, these investigations are limited to the cultivable minority of *Streptomyces* species and are likely insufficient to capture the breadth of localized, ecological processes shaping *Streptomyces* biodiversity and community structure *in situ*. Nevertheless, cultivation-based approaches have enabled valuable physiological investigations of diverse traits within members of the *Streptomyces* (Schlatter *et al.*, 2013; Choudoir and Buckley, 2018), yet cultivation-independent explorations of the structure of *Streptomyces* communities are lacking and represent an underexplored area of research regarding the ecology of this diverse genus.

The universality of PCR primers targeting the 16S rRNA (16S) gene across bacterial taxa has facilitated key scientific investigations related to expanding our understanding of microbial

diversity (Baker *et al.*, 2003; Lauber *et al.*, 2009). However, the significant conservation within the 16S gene is indicative of a slow evolutionary rate (1% nucleotide divergence $\propto$ 50 myr) permitting coarse diversity profiling, but is insufficient to detect fine-grained population structure over recent evolutionary time scales (i.e., tens to hundreds of thousands of years) (Choudoir *et al.*; Ochman and Wilson, 1987; Ochman *et al.*, 1999; Adékambi *et al.*, 2009). For diversity assessments within some congeneric microbial lineages, the 16S gene is inadequate to delineate “species” or fine-grain operational taxonomic units (OTUs), and alternative genes have been identified that improve discrimination where fine scale diversity trends are of interest (Case *et al.*, 2007; Antony-Babu *et al.*, 2017). The gene encoding the beta subunit of the DNA-directed RNA polymerase (*rpoB*) has been shown to differentiate closely-related *Streptomyces* species when the 16S gene could not (Kim *et al.*, 2004; Case *et al.*, 2007). Hence, protein-coding genes, such as *rpoB*, are preferred for biogeographic explorations of microorganisms and may facilitate identification of underlying community assembly mechanisms (Koeppel and Wu, 2013).

While cultivation-dependent investigations of *Streptomyces* species have uncovered novel biogeographic patterns related to *Streptomyces* species diversity (Choudoir *et al.*, 2016; Choudoir and Buckley, 2018), these isolation-based studies are laborious and do not scale to the large number of samples required for investigations of broad-scale biogeographic phenomena. Hence, we set out to develop and apply *Streptomyces*-specific *rpoB* PCR primers compatible with high-throughput Illumina sequencing to investigate *Streptomyces* community structure in environmental samples. To demonstrate the utility of the *Streptomyces rpoB* primers, we amplified and sequenced *Streptomyces rpoB* genes from soils spanning 1,200 km along a latitudinal transect in New Zealand (NZ). Biogeographical studies in NZ have uncovered distinctive patterns of floral and faunal diversity trends owing to the continental island’s unique paleoclimatic and geologic history (Fleming, 1975; Heads, 1997; Wallis and Trewick, 2009; McGlone *et al.*, 2010; Tonkin *et al.*, 2018), and information regarding the diversity and structure of *Streptomyces* communities across this temperate island region are lacking. Historical phenomena (e.g., glaciation, climatic anomalies) are correlated with both *Streptomyces* phylogenetic diversity and total bacterial community richness (Eisenlord *et al.*, 2012; Choudoir *et al.*, 2016; Delgado-Baquerizo *et al.*, 2017), but the degree to which paleo and contemporary climate shape *Streptomyces* diversity across NZ is unknown and may advance our understanding of bacterial diversity trends within NZ soils. Therefore, we performed high-throughput sequencing of *Streptomyces rpoB* amplicons along a latitudinal transect in NZ to investigate ecological processes (selection, drift, dispersal, diversification) shaping *Streptomyces* communities in soil. Furthermore, we test the hypothesis that phylogenetic clustering of *Streptomyces* communities will occur poleward, consistent with a legacy of widespread glaciation on NZ’s south island as has previously been observed in North America (Andam *et al.*, 2016).

Results and Discussion

Determining Streptomyces diversity by rpoB analysis

Patterns of rpoB alpha diversity

Using *Streptomyces* specific *rpoB* primers we observed 159 ± 92 *Streptomyces rpoB* OTUs per site (mean $\pm$ s.d., range: 53 to 395) across the 1,200 km NZ latitudinal transect. When using 16S gene data to identify *Streptomyces* OTUs, only 9 ± 2 OTUs and 16 ± 5 ESVs were observed per site. Accordingly, the *rpoB* primers demonstrated higher resolution than 16S gene primers for detecting diversity within the genus *Streptomyces* (Fig. S1). This finding corroborates previous observations that the 16S gene lacks sufficient genetic resolution to discriminate between *Streptomyces* species (Labeda et al. 2017, Cheng et al. 2016, Rong et al. 2013), and that *rpoB* is a preferred taxonomic marker for members of the *Streptomyces* (Andam 2016). The rate of new species discovery using *rpoB* sequences declined rapidly as read counts approached 1,000 (Fig. S2). At a sequence depth of 5,000 counts per sample, *rpoB* OTU coverage met or exceeded 99% for all but two samples (Table S1). Sequencing coverage for *rpoB* OTUs in each site ranged between 0.95 and 0.99 at a read count of 922 sequences (Table S1), the lowest read count observed in this study. Thus, amplicon pools generated with the *Streptomyces rpoB* primers can be multiplexed to determine near complete *Streptomyces* species inventories across large numbers of soil samples.

Of the four abiotic variables examined, pH and MAT were significantly correlated with *Streptomyces* phylogenetic diversity (PD) (Fig. 1A, Table 1), whereas MAR and MAWS were non-significant and negatively correlated with PD (Fig. 1B, Table 1). No significant correlation was detected between *Streptomyces* PD and elevation. In addition to abiotic variables, we observed a significant negative correlation between *Streptomyces* PD and NZ plant species richness (Table 1). While clear links exist between *Streptomyces* community structure and plant growth promotion or plant host genotype (Seipke et al., 2012; Bakker et al., 2013; Viaene et al., 2016; Essarioui et al., 2017), additional experiments are needed to support a causal influence of plant diversity on *Streptomyces* diversity. Aside from their beneficial effects on plants, a minority of *Streptomyces* species also represent harmful plant pathogens that cause scabs lesions on root storage organs (e.g., potato) (Loria et al., 2006; Anderson et al., 2014). Of the 1,287 *rpoB* OTUs, only 10 OTUs matched *rpoB* sequences classified as canonical scab-causing *Streptomyces* species. However, these OTUs were detected in 6 out of 15 NZ soil samples, and ranged in abundance from 0 to 3.8% (0.5 ± 1.1 %; mean $\pm$ s.d.) of sequences in a given sample. Although additional evidence is required to confirm the presence of pathogenic *Streptomyces* species in the NZ soils, our results suggest that high-throughput *rpoB* sequencing could provide an alternative, cultivation-independent method to surveil putatively pathogenic populations of *Streptomyces* spp. in soils.

Patterns of *rpoB* evenness

Estimates of *Streptomyces* community evenness were low (0.19 ± 0.06 , mean $\pm$ s.d.) and only MAR demonstrated a significant negative correlation with E_{var} (Table 1). The dominant *rpoB* OTU comprised up to 60% (31 ± 15 %, mean $\pm$ s.d.) of *Streptomyces* sequences observed within each site, indicating strong dominance within *Streptomyces* communities in NZ soils. *Streptomyces* spp. are ubiquitous and easily cultivated (Kämpfer et al., 2014), but additional studies are required to identify if a cultivation bias exists that selects for numerically dominant *Streptomyces* spp. in soils. Although changes in community composition during storage could

explain the strong dominance we observed, electrophoresis analysis of *Streptomyces* 16S gene profiles from fresh soils also display dominance (Inbar *et al.*, 2005). In addition, when we compared soils in long term dry storage to those frozen rapidly we did not observe a systematic bias towards dominance (Fig. S3).

***Streptomyces* beta diversity based on rpoB**

Spatial, historical, and climatic correlates of Streptomyces beta diversity

Streptomyces community composition was highly dissimilar between sites and did not correlate with geographic distance (Fig. 2; Mantel $r < 0.05$, $P > 0.4$; Procrustes Gower statistic, $m_{12} = 0.81 - 0.86$, $P > 0.36$). Furthermore, no difference in *Streptomyces* community composition was detected between NZ's north and south islands when assessed using either *rpoB* ($F_{1,13} = 1.17$, $R^2 = 0.08$, $p > 0.1$) or 16S gene analyses ($F_{1,13} = 1.15$, $R^2 = 0.08$, $p > 0.2$) (Fig. 3). Instead, soil pH was correlated with both total bacterial and *Streptomyces* community composition across NZ (Fig. 3), as previously observed for soil bacterial communities (Lauber *et al.*, 2009). Hence, habitat filtering via pH regimes might play a strong role in structuring soil *Streptomyces* communities as has been observed generally for soil *Actinobacteria* (Lauber *et al.*, 2009).

With respect to *Streptomyces* community composition, pairwise dissimilarity between sites saturated within 250 km (Fig. 2). Saturation of pairwise dissimilarity was less rapid for UniFrac metrics that incorporate phylogenetic distance (Fig. 2), suggesting phylogenetic conservation of community composition at regional spatial scales (< 250 km). The abrupt saturation in pairwise dissimilarity observed between sites is surprising considering *Streptomyces* form desiccation resistant aerial spores that facilitate dispersal (Sarmiento-Vizcaíno *et al.*, 2016). However, dispersal depends on both displacement and colonization. The pattern of high dominance and high community dissimilarity that we observe suggests that priority effects and/or antagonistic interactions likely constrain colonization success despite high rates of spore displacement (Fukami, 2015; Wright and Vetsigian, 2016). Antagonism mediated by secondary metabolite production is common among divergent *Streptomyces* spp. (Traxler and Kolter, 2015), and such antagonistic interactions can promote coexistence of closely related species (Vetsigian *et al.*, 2011; Cordero *et al.*, 2012).

In addition to signatures of dispersal limitation, we detected a positive association between *Streptomyces* community net relatedness index (NRI) and latitude ($F_{1,13} = 3.59$; Adj. $R^2 = 0.16$; $p = 0.08$; Fig. 4A-B), with NRI values increasing towards the equator. The NRI is a standardized measure of phylogenetic clustering and increases or decreases when phylogenetic diversity is lower (clustered) or higher (over-dispersed) than expected considering the species richness (Webb *et al.*, 2002; Swenson, 2009). Although this relationship was not significant, removal of three suspected outliers (points 2, 14, 15; Fig. 4A) revealed a significant relationship ($F_{1,10} = 9.03$, $P = 0.013$, Adj. $R^2 = 0.42$) between NRI and latitude. A significant NRI-latitude relationship was observed for *Streptomyces* communities in North America, with phylogenetic clustering increasing poleward (Andam *et al.*, 2016). However, the NZ NRI-latitude association is opposite to the previously described trend, with phylogenetic clustering increasing towards the equator (Fig. 4A-B).

The *Streptomyces* NRI-latitude relationship described by Andam et al. is consistent with niche conservatism following retreat of continental glaciers in North America after the last glacial maximum (LGM) (Andam *et al.*, 2016). However, the LGM in NZ (~25-10 kya) was characterized by montane glaciers that were largely restricted to discrete regions of NZ's south island (McGlone, 1985; Wallis and Trewick, 2009; Newnham *et al.*, 2013). Glacial dynamics in New Zealand during the LGM differed from those of continental regions. Recolonization dynamics in NZ at the end of the LGM are hypothesized to have resulted from species expansion from scattered 'micro-refugia' rather than northward expansion from southern refugia as hypothesized for North America and Europe (Provan and Bennett, 2008; Wallis and Trewick, 2009). Furthermore, the predicted vegetation cover of ice-free regions on NZ's South Island during the LGM was predominantly dry grassland and shrubland habitat in which *Streptomyces* species thrive (Newnham *et al.*, 2013; Kämpfer *et al.*, 2014), whereas forested habitat, within which *Streptomyces* are not favored, was more common on the North Island (McGlone, 1985; Newnham *et al.*, 2013). The climatic stability of habitat favorable to *Streptomyces* on the South Island (De Wever *et al.*, 2009; Keppel *et al.*, 2012) may explain the high phylogenetic diversity (lower NRI values) of sites in southern NZ (Fig. 4A). In contrast, the persistence of beech and podocarp forests on the North Island may have selected specifically for *Streptomyces* species capable of degrading plant woody components (Bontemps *et al.*, 2013), resulting in phylogenetically clustered communities (positive NRI values, Fig. 4A). These results are consistent with the hypothesis that historical changes in paleoclimate have influenced extant patterns of *Streptomyces* biogeography.

Patterns of *rpoB* OTU sharing among sites revealed greater *rpoB* OTU sharing between North and South Island sites than within comparisons of South Island sites alone (Fig. 4C), which is consistent with a legacy of glaciation on NZ's South Island. Some South Island sites exhibited low *rpoB* OTU sharing (sites 5, 8, 10, 12; Fig. 4C), and three of these four sites have roughly double the estimated MAR of other sites (Table 2), suggesting environmental filtering could be contributing to the observed co-occurrence patterns. We also investigated *rpoB* OTU ranges within NZ, which were observed to be highly constrained for most OTUs (Fig. 4D, observed panel). To assess whether the observed species ranges were an artifact of the sampling design, we randomly shuffled *rpoB* community data using a null model algorithm that constrains site richness to the observed richness but enables *rpoB* OTU occupancy to vary freely between sites (Fig. 4D, richness panel). Under this null model scenario, the average *rpoB* OTU range (408 km) approaches the average pairwise distance among sites (496 km), which deviates considerably from the observed *rpoB* OTU range distribution and indicates species distributions are not random. While a significant fraction of *Streptomyces* OTUs possess a wide distribution (13% of *Streptomyces* OTUs have sample ranges > 1,000 km; Fig. 4D), the majority appear to be constrained by biotic or abiotic filters that impair their recruitment to only one or a few local sites across NZ. Furthermore, we observed a positive relationship between *Streptomyces* OTU mean relative abundance and their geographic range size (Fig. 4E). This relationship has been documented for bacteria, plants, and animals (Brown, 1984; Gaston *et al.*, 1997; Harte *et al.*, 2001; Nemergut *et al.*, 2011), but information regarding range size of discrete bacterial lineages using high-resolution genetic markers are lacking. Multiple mechanisms to explain this relationship have been hypothesized, such as neutral processes, differential dispersal capabilities

(e.g., spore morphology differences among *Streptomyces* OTUs), differential birth and death rates, habitat selection, and differential resource utilization (Gaston *et al.*, 1997; Nemergut *et al.*, 2011), but further investigation is required to identify the underlying causes of the abundance-range relationship for *Streptomyces* OTUs.

The variables pH, MAR, and plant species richness were all significantly associated with *Streptomyces* community composition (jaccard-based PCoA; envfit $R^2 = 0.49, 0.44, 0.64$ and $p = 0.016, 0.029, \text{ and } 0.002$, respectively; Table S2). Similar results were observed using an unweighted UniFrac PCoA ordination (Table S2). Accordingly, pH was significantly associated with Jaccard-based PCoA axis 1, whereas PCoA axis 2 was significantly associated with NZ plant species richness (Fig 5A, 5B, respectively). Although the association between NZ plant species richness and the Jaccard-based ordination remained significant when controlling for both MAT, MAR, and latitude ($t_{3,11} = -2.71, p = 0.022$), our sample size is limited and our results should be viewed as preliminary. Regardless, this finding warrants further study due to the plant-associated lifestyle of many members of the *Streptomyces* and prior reports of links between *Streptomyces* community composition and plant diversity (Bakker *et al.*, 2013; Viaene *et al.*, 2016; Rey and Dumas, 2017).

Community assembly processes governing *Streptomyces* community turnover

Variable selection was identified as the greatest driver of community assembly across the NZ sites (Table 3), while the contribution of homogeneous selection to community turnover was low (7% of all comparisons). These results are not surprising considering the geographic scale of our analysis, which encompasses sites with widely varying environmental characteristics (Barker, 2005). Homogenizing dispersal contributed to approximately 30% of community assembly processes identified (Table 3). Most pairwise site comparisons within 500 km had β_{RC} values < -0.95 , which indicates greater community similarity than expected when compared with null modelling (Fig. S4A). However, some β_{RC} values < -0.95 were also detected between sites located $>1,000$ km apart, which may be indicative of long range dispersal of *Streptomyces* spp. (Fig. S4A). The potential for long-range dispersal of *Streptomyces* has been observed previously (Andam *et al.*, 2016; Choudoir *et al.*, 2016), and 13% ($n = 171$) of the *rpoB* OTUs were identified in New Zealand sites separated by $>1,000$ km. While dispersal limitation accounted for only 6% of community assembly patterns, values of β_{RC} increased with increasing pairwise distance and suggests an underestimation of dispersal limitation in our analysis (Fig. S4B). One possible explanation is that non abundance weighted metrics may not achieve the same sensitivity as their abundance weighted counterparts (Stegen *et al.*, 2012, 2015). However, abundance weighting of β_{RC} as RC_{bray} (Stegen *et al.*, 2013) is inappropriate for dry-stored soils for which sample abundance might not represent *in situ* abundance (Clark and Hirsch, 2008; Ivanova *et al.*, 2017). While the β_{RC} metric is also influenced by the size of the regional species pool selected (Chase *et al.*, 2011), *Streptomyces* spores can be readily dispersed by insects or wind (Lloyd, 1969; Ruddick and Williams, 1972), and the choice of the totality of NZ as a regional *Streptomyces* species pool is not unrealistic. Regardless, the approach that we have applied, combining null modeling with multiple lines of evidence, has been shown to provide

robust insight into assembly processes governing microbial communities (Chase *et al.*, 2011; Stegen *et al.*, 2015).

Comparison of environmental rpoB to cultivated representatives

Streptomyces rpoB OTUs and uncultivated rpoB sequence diversity

Of the *rpoB* OTUs classified as *Streptomyces*, only 12% (n = 149) matched *rpoB* sequences from cultivated *Streptomyces* strains (n = 1808) described in the ARS Microbial Genomic Sequence Database (99% nucleotide identity cutoff; see Materials and Methods for details). Therefore, despite their cultivability (Kämpfer *et al.*, 2014), only a small fraction of *Streptomyces rpoB* OTUs were linked to cultivated representatives. However, the 149 OTUs linked to cultivated strains represented one quarter of all sequence counts (108,068/428,122), and four of these OTUs (OTU_3, OTU_5, OTU_6, and OTU_8) represented 65% (n = 70,652) of this fraction. In addition, the percentage of *rpoB* OTUs matching *rpoB* sequences extracted from 998 *Streptomyces* genomes was only 10% (n = 126) of the 1,287 OTUs observed. Of the four dominant cultivated OTUs, OTU_6 failed to match an *rpoB* sequence from the 998 genomes examined. The *rpoB* sequence of OTU_6 was most similar to ARS type specimen *Streptomyces sanglieri* strain NRRL B-24279 (99.7% nucleotide identity), but was only 98.857% similar to the genome of *Streptomyces* sp. DvalAA-43 (NCBI accession FMBW01000037.1), suggesting a representative genome for this strain may be lacking. Thus, data derived from the *rpoB* primer set may assist in prioritizing future culture collection sequencing efforts of *Streptomyces* isolates. The fact that *rpoB* OTUs matching cultivated strains represented 12% of taxa observed but 25% of sequence counts suggests that most *Streptomyces* communities are dominated by few taxa and contain many rare taxa. Since isolation of many pure cultures is laborious, isolation efforts are likely to under sample *Streptomyces* diversity. In addition, the low representation of *rpoB* OTUs among cultivated strains and sequenced genomes suggests that the genus *Streptomyces* contains an enormous genomic diversity that remains largely uncharacterized. Future cultivation efforts might combine high-throughput *Streptomyces rpoB* amplicon sequencing and single cell (or spore) cultivation to capture rare taxa present in *Streptomyces* communities.

Conclusions

Biogeographical surveys of the *Streptomyces* have long benefitted from cultivation-dependent methodologies, but the development of efficient, cultivation-independent methodologies are preferred for large scale biogeographic studies. We introduce a high-throughput sequencing primer set that targets the DNA-directed RNA polymerase beta subunit (*rpoB*) within the bacterial genus *Streptomyces* and examined the biogeography of *Streptomyces* in soils across a 1,200 km transect in New Zealand (NZ). The *rpoB* PCR primers provided a near complete inventory of *Streptomyces* OTUs in NZ soils and provided far greater coverage of *Streptomyces* spp. than could be achieved by cultivation efforts or by 16S rRNA gene sequencing data from the same sites. Specific abiotic (pH, MAR), biotic (plant richness), and historical (glaciation) features of NZ were found to significantly correlate with *Streptomyces* PD. Phylogenetic clustering of *Streptomyces* communities was opposite to that observed in North America, with clustering increasing towards the equator. Presumably, the presence of favorable grassland and shrubland refugia afforded by restricted, montane glaciers in NZ contributed to the lack of

phylogenetic clustering on NZ's South Island. These results, taken together, suggest that extant *Streptomyces* communities are shaped by dispersal of spores from species pools determined by historical biogeographical processes, with community structure of individual sites shaped by competitive interactions that govern colonization dynamics and which are constrained by local abiotic (soil pH, temperature) and biotic (plant communities) conditions. Overall, the *rpoB* primer set enabled the detection of a heretofore unobserved *Streptomyces* diversity, greatly exceeding those currently documented in culture and genome collections, and hinting at an untapped reservoir of novel *Streptomyces* antibiotics and biologics.

Experimental Procedures

Soil sampling, storage, and data acquisition

Soils (0-10 cm depth) were sampled from 15 locations up to ~1,200 km apart throughout NZ between February and April of 2012 (Table 2). Soils were sieved (2 mm mesh), air-dried, and stored in sterile, plastic whirl-pak bags (Nasco, Fort Atkinson, WI, USA) at room temperature until DNA extraction (~ 3 yr). Air-dried, archived soils have been extensively studied and represent a valuable resource for soil microbiologists due to the cost and space required to maintain frozen soil samples (Cary and Fierer, 2014; Dolfing and Feng, 2015). However, the community composition of archived, air-dried soils is significantly different from frozen/fresh soils and careful consideration is required when analyzing air-dried soil that has been stored for long periods (years) (Dolfing *et al.*, 2004; Clark and Hirsch, 2008; Tzeneva *et al.*, 2009; Ivanova *et al.*, 2017). *Streptomyces* produce desiccation resistant spores that can survive dry conditions for many years (Ensign, 1978). While we found that storage method did alter *Streptomyces* relative abundance, analysis of air dried relative to frozen soils indicated significant correlation in the relative abundance of *Streptomyces* OTUs between the two storage methods (see SI and Fig. S3 for additional details).

The pH of the soil was measured by pH probe using 1:2 (w/v) soil dilutions in 10 mM CaCl₂ (Staff, 2014). The mean annual air temperature (MAT) (°C), rainfall (MAR) (mm), and wind speed (MAWS) (m/sec) for 2012 were retrieved from the National Institute of Water and Atmospheric Research (NIWA) CliFlo (<https://cliflo.niwa.co.nz/>) database weather stations nearest (35 ± 21 km) to soil sampling sites. Plant species survey data (NZ National Vegetation Survey databank; <https://landcareresearch.co.nz/>) were accessed from the Global Biodiversity Information Facility (<https://www.gbif.org/>) in December 2018. Only species presence/absence data from vegetation plots surveyed after the year 1960 and within 100 km of the soil sampling sites were used as input for downstream analyses. To control for sampling effort, plant species presence/absence data were generated by rarefying to the minimum number of observations (9,790) per site and resampling with replacement 999 times. The overall presence/absence of each plant species was scored at a site by iterating over all species identified in the dataset and scoring their presence/absence within all 999 draws of 9,790 observations for each site (a link to a figshare repository containing the raw data is provided in the *Data availability* section).

Microbial strains for PCR amplification

Genomic DNA stocks prepared from *Streptomyces* isolates cultivated from a variety of habitats were used as positive controls for *rpoB* amplification (Choudoir *et al.*, 2016). Additional isolates of *Streptomyces*, *Kitasatospora*, *Streptacidiphilus*, and *Catenulispora* were received by request from the United States Department of Agriculture Agricultural Research Service culture collection (<https://nrrl.ncaur.usda.gov>). These species were cultivated as indicated by documentation accompanying the isolate and DNA was extracted by one minute of bead-beating (0.1 mm silica beads) using an established chloroform-phenol extraction method (Griffiths *et al.*, 2000).

PCR primer design and analysis, DNA isolation, and Illumina sequencing

Streptomyces rpoB gene sequences were downloaded from the PubMLST database (Jolley *et al.*, 2018), nucleotide alignment performed with MAFFT, and PCR primers designed using DECIPHER software (Wright, 2016) to amplify a region of sufficient length for successful Illumina high-throughput amplicon sequencing projects (usually ~200-500 nt). DECIPHER was also provided *rpoB* sequences from members of the *Kitasatospora* (N = 10) and *Streptacidiphilus* (N = 9) to enable DECIPHER to discriminate against these sequences during primer design and enhance specificity of the primers for *Streptomyces rpoB* gene sequences (see SI Results and Discussion for additional details). The *rpoB* primers were named Smyces_rpoB1563F (5'- GGAGGACCGCTTCGTCATC-3') and Smyces_rpoB1968R (5'- GTACGTGGTGTACGTGCCG-3') and amplify a 406 bp region within the *rpoB* gene. The numbers within the primer names correspond to the primer alignment position within the *rpoB* sequence from *Streptomyces coelicolor* strain A3(2) (NCBI accession NP_628815.1). The *rpoB* primers containing modifications for Illumina sequencing (as per (Kozich *et al.*, 2013)) are provided in Supporting Information (SI) (Table S3).

Forward and reverse primers targeting the V4 region of the 16S rRNA (16S) gene in bacteria (5'- GTGCCAGCMGCCGCGGTAA-3' and 5'- GGACTACHVGGGTWTCTAAT-3', respectively) were used for overall bacterial community analyses in NZ soils as previously described (Kozich *et al.*, 2013). DNA was extracted from NZ soils using a PowerSoil DNA isolation kit (Qiagen, Hilden, Germany) following manufacturer's protocols and subsequently stored at -20 °C until use. Briefly, *rpoB* and 16S gene end-point PCR reactions consisted of 1 µl of DNA (~0.5-19.5 ng DNA total), 12.5 µl of Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA, USA), 0.625 µl of 4X Quant-iT PicoGreen dsDNA assay reagent (Thermo Fisher Scientific, Waltham, MA, USA), 1.25 µl each of 10 µM dual-barcode forward and reverse primers, 1.25 µl of 20,000 ng/µl bovine serum albumin (New England Biolabs), and 7.125 µl of molecular grade water. Thermocycler conditions were 98 °C for 30 sec, 30 cycles of 98 °C for 10 sec, 55 °C for 30 sec, and 72 °C for 1 min, followed by extension for 5 min and melt curve analysis at 0.5 °C/5 sec from 65 to 95 °C. PCR Amplification of *rpoB* gene amplicons from a variety of *Streptomyces* isolates revealed clean, consistent amplicons of the expected size (Fig. S5). Although *Streptomyces rpoB* gene copies per g soil were not quantified in the present study, the *rpoB* primer pair reported can also be used in quantitative PCR assays. The details of these methods are reported in SI (Table S4, Fig. S6).

After PCR amplification, all end-point *rpoB* and 16S gene amplicons were pooled at equimolar concentrations (~1-2 ng/μl) using a SequalPrep normalization kit (Thermo Fisher Scientific, Waltham, MA, USA) and purified by gel extraction using the Wizard SV gel and PCR clean-up system (Promega, Fitchburg, WI, USA) following manufacturer's instructions. The resulting sequencing libraries were provided to the Cornell University Biotechnology Resource Center for paired-end Illumina MiSeq sequencing (2 x 250 bp) (Illumina, Inc., San Diego, CA). A 20 pM concentration of phiX spike-in at 10-20 % [v/v] was added to 16S and *rpoB* gene sequencing libraries to improve the quality of the low diversity amplicon libraries.

Sequence processing and analysis

For simplicity of reporting our 16S gene processing, we report both OTU and exact sequence variant (ESV) identification methods for the *rpoB* gene. However, we only present OTU data for the *rpoB* gene in the manuscript. While ESVs are recommended for analyses of microbial communities using the 16S gene (Knight *et al.*, 2018), their utility for protein-coding nucleotide sequences is uncertain considering distinct ESVs could comprise synonymous or non-synonymous substitutions occurring within a single species. Furthermore, empirical analysis demonstrated that 99% nucleotide identity OTUs are a reasonable proxy for species when compared with a species definition for *Streptomyces* using multi-locus sequence analysis (Fig. S7) (Rong and Huang, 2012). See SI for details regarding the analysis.

Raw sequences were demultiplexed using dual-barcode FASTQ sequence information (Kozich *et al.*, 2013) with the tool deML (Renaud *et al.*, 2015), followed by paired-end read merging using PEAR v0.9.10 (with parameters -p 0.001 and -q 30) (Zhang *et al.*, 2014). Residual Illumina adapters and *rpoB* primer sequences were removed with Cutadapt v1.14 (Martin, 2011). The merged *rpoB* reads were either clustered *de novo* at 99% nucleotide identity or identified as exact sequence variants (ESVs) with VSEARCH v2.8.1 (Rognes *et al.*, 2016) or USEARCH (UNOISE3 function) v10.0.240 (Edgar, 2016), respectively. Prior to 99% *rpoB* OTU clustering with VSEARCH, *rpoB* sequences were filtered using a per base expected error of 0.25 (--fastq_maxee 0.25) and trimmed to 350 nt in length (--fastq_truncLen 350). The output sequences were dereplicated (singletons removed), pre-clustered at 99.5% nucleotide identity, and chimeric sequences identified and removed by UCHIME (--uchime_denovo option in VSEARCH) (Edgar *et al.*, 2011). For ESV identification, UNOISE3 was run using default settings except the minimum abundance was set to 9 (-minsize 9) to filter out ESVs that may represent sequencing artifacts. Taxonomic assignments were made using the BLCA tool (Gao *et al.*, 2017) (using parameters -n 100 -j 50 -d 0.1 -e 0.01 -b 100 -c 0.9 --iset 90) and a custom *rpoB* sequence database (see below). The classification of 16S gene OTUs and ESVs followed a similar methodology, except that a 97% nucleotide identity cutoff was provided to VSEARCH for 16S gene OTU identification and taxonomy assigned using BLCA and NCBI's 16S rRNA gene collection (Gao *et al.*, 2017). The number of *rpoB* reads per sample varied between 925 and 90,276 (25,658 ± 26,732, mean ± s.d.). The number of reads per sample for 16S gene OTUs and ESVs ranged between 12,574 and 45,532 (21,529 ± 8,803, mean ± s.d.) and 16,974 and 55,920 (27,328 ± 10,640), respectively. Of the 3,322 *rpoB* OTUs detected in the present study, only 1,297 (39%) were classified as members of the *Streptomycetaceae* (Table 4). However, the

abundance of the 2,025 non-target OTUs comprised only 9% (39,866 / 467,988) of the total read counts (Table 4; see SI for additional details).

A phylogeny of 16S gene and *rpoB* OTU sequences was generated by sequence alignment with MAFFT v7.407 (Katoh *et al.*, 2009) followed by tree construction using FastTree v 2.1.8 (Price MN, Dehal PS, 2010) (with parameters -gtr -gamma -spr 4 -mlacc 2 -slownni -bionj -nt). The resultant phylogeny was midpoint-rooted with the R package phangorn (Schliep, 2010). Visual inspection of the *rpoB* gene phylogeny revealed three OTUs classified as *Streptomyces* (OTU_3178, OTU_3273, and OTU_3090) possessing long branch lengths of 6.0611, 0.1380, and 6.0579, respectively, and low sequence counts (≤ 5 counts). As a reference, the interquartile branch length range was between 0.0033 and 0.0134. Hence, the three *rpoB* OTUs were removed prior to analysis since they may represent sequencing artifacts.

Available *Streptomyces* genomes were accessed from NCBI (as of 9/11/2017; n = 998) and prodigal v2.6.2 (Hyatt *et al.*, 2010) was used to predict protein-coding gene and amino acid sequences for consistent gene prediction among genomes. Then, predicted protein-coding sequences were provided to CheckM v1.0.7 (Parks *et al.*, 2015) via the analyze function and the output was parsed for sequences containing an N-terminal protein domain diagnostic of bacterial RpoB (RNA_pol_Rpb2_1, pfam id: PF40563). In total, 990 full-length *rpoB* sequences were used as a reference genome database. An additional set of 1,834 partial *rpoB* gene sequences from cultivated *Streptomyces* species were accessed from the ARS Microbial Genomic Sequence Database server (<http://199.133.98.43>) (Labeda *et al.*, 2017). The nucleotide version of the basic local alignment search tool (BLAST) (Altschul *et al.*, 1990) was used to align *rpoB* OTUs to reference *rpoB* sequences from both genomes and type culture collections. An *rpoB* OTU was considered as a ‘match’ to a reference *rpoB* sequence when alignments satisfied the criteria of $\geq 99\%$ nucleotide identity and $\geq 99\%$ query coverage. Of the 1,834 partial *rpoB* database sequences, 85 were identified as *Streptomyces* species associated with scab-causing disease (Zhang *et al.*, 2016). Any *rpoB* OTUs demonstrating significant alignment ($\geq 99\%$ nucleotide identity, $\geq 99\%$ query coverage) to these sequences were classified as putative pathogens.

Statistics

The R package phyloseq was used to pre-process or manipulate 16S and *rpoB* community data (McMurdie and Holmes, 2013). The iNEXT package was used for calculating richness estimates and rarefaction of the *rpoB* community data (Hsieh *et al.*, 2016). The *rpoB* OTU richness estimates were normalized to a consistent sample coverage among samples (95% in the present study; Chao *et al.* 2014; Mateo-Tomás *et al.* 2015) when comparing *rpoB* OTU richness to spatial (latitude), environmental (pH, temperature, rainfall, wind speed), and biological (plant species diversity) correlates of *Streptomyces rpoB* OTU richness in NZ. Indirect gradient analysis of the *rpoB* dissimilarity data was performed using multidimensional scaling (PCoA) with the cmdscale command in R (R Core Team, 2018). The adonis and envfit functions (Oksanen *et al.*, 2019) were used to evaluate whole community associations with environmental variables using 9,999 permutations. Ordinary least squares regression was performed with the R ‘stats’ package and regression tables produced using the R package ‘stargazer’ (Hlavac, 2015; R Core Team, 2018). Community evenness was estimated using the codyn package in R and the E_{var} evenness metric, which ranges between 0 (complete dominance) and 1 (complete evenness).

(Smith *et al.*, 1996; Hallett *et al.*, 2016). We estimated species contributions to beta diversity (SCBD) from the Hellinger-transformed *Streptomyces* OTU table using the beta.div function in the adespatial R package (Dray *et al.*, 2017). The net relatedness index (NRI), an estimate of phylogenetic diversity compared to a null model of species richness, was estimated using the picante package and multiplied by -1 for comparability with previous phylocom NRI values using *Streptomyces* community data (Kembel *et al.*, 2010; Andam *et al.*, 2016). Missing NIWA climate values (rainfall and temperature for Hope Saddle Lookout and rainfall for Kaikoura Peninsula) were imputed with the R package ‘mice’ using the ‘pmm’, or predictive mean matching, method (van Buuren and Groothuis-Oudshoorn, 2011). We investigated the contributions of selection, dispersal, drift using a null modeling approach (*sensu* Stegen) by comparing non-abundance weighted (presence-absence) beta mean nearest taxon index (β NTI) and the beta Raup-Crick metrics (β _{RC}) (Chase *et al.*, 2011; Stegen *et al.*, 2013, 2015). The original β _{RC} metric was preferred over the RC_{bray} metric since the latter requires relative abundance information that may not be appropriate for air-dried, archived soils. Null modeling of β NTI was performed by subtracting from the observed beta mean nearest taxon distance (β MNTD) the mean of the null β MNTD and dividing by the s.d. of the null β MNTD (999 replicates, taxon labels shuffled using tipShuffle in the R picante package) (Kembel *et al.*, 2010; Stegen *et al.*, 2013; Swenson, 2014). Values of β _{RC} were calculated using 9,999 replicates with the raup_crick function reported by Chase *et al.* (2011). The simultaneous analysis of both β NTI and β _{RC} metrics enables the partitioning of community variation into components of selection (variable and homogeneous, β NTI > 2 and < -2, respectively), dispersal (limited and homogenizing dispersal, $|\beta$ NTI| < 2 and β _{RC} > 0.95 or $|\beta$ NTI| < 2 and β _{RC} < -0.95, respectively) and drift (neither selection nor dispersal dominated or “undominated”, $|\beta$ NTI| < 2 and $|\beta$ _{RC} < 0.95) (Stegen *et al.*, 2015). Additional null modeling was performed with 1,000 iterations using the randomizeMatrix function in the picante package in R (Kembel *et al.*, 2010).

Data availability

Both 16S and *rpoB* gene sequences amplified from the NZ soil DNA were submitted to the Sequence Read Archive (Leinonen *et al.*, 2010) under study no. SRP194926. The National Center for Biotechnology Information BioProject and BioSample accession numbers are PRJNA540985 and SAMN11569796, respectively. A figshare repository containing all scripts, data, and reference *rpoB* sequences is publicly accessible at (<https://doi.org/10.6084/m9.figshare.c.4573250>).

Acknowledgements

S.A.H. is grateful to Dr. Ashely M. Frank for her assistance with discussions regarding manuscript writing, organization, and revision. D.H.B. is grateful to Dr. Paul B. Rainey and The Institute for Advanced Study, Massey University for providing research support, laboratory access, and helpful advice. This material is based upon work supported by the National Science Foundation under Grant No. DEB-1456821 awarded to D.H.B.

Author Contributions

519 D.H.B collected samples and designed the experiment, S.A.H. and K.P. performed analyses, and
520 S.A.H. and D.H.B. wrote the manuscript.

521 **Competing Interests**

522 The authors declare no conflict of interest with the work reported.

523 **References**

- 524 Adékambi, T., Drancourt, M., and Raoult, D. (2009) The *rpoB* gene as a tool for clinical
525 microbiologists. *Trends Microbiol* **17**: 37–45.
- 526 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local
527 alignment search tool. *J Mol Biol* **215**: 403–410.
- 528 Andam, C.P., Doroghazi, J.R., Campbell, A.N., Kelly, P.J., Choudoir, M.J., and Buckley, D.H.
529 (2016) A latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*.
530 *MBio* **7**: e02200-15.
- 531 Anderson, J., Schlatter, D., Otto-Hanson, L., and Kinkel, L. (2014) Variation in *Streptomyces*
532 densities and inhibitory capacities from diverse natural habitats in New Zealand. In
533 *Phytopathology*. Scientific Societies, pp. S3.1-S3.138.
- 534 Antony-Babu, S., Stach, J.E.M., and Goodfellow, M. (2008) Genetic and phenotypic evidence
535 for *Streptomyces griseus* ecovars isolated from a beach and dune sand system. *Antonie Van*
536 *Leeuwenhoek* **94**: 63–74.
- 537 Antony-Babu, S., Stien, D., Eparvier, V., Parrot, D., Tomasi, S., and Suzuki, M.T. (2017)
538 Multiple *Streptomyces* species with distinct secondary metabolomes have identical 16S
539 rRNA gene sequences. *Sci Rep* **7**: 11089.
- 540 Baker, G.C., Smith, J.J., and Cowan, D.A. (2003) Review and re-analysis of domain-specific 16S
541 primers. *J Microbiol Methods* **55**: 541–555.
- 542 Bakker, M.G., Bradeen, J.M., and Kinkel, L.L. (2013) Effects of plant host species and plant
543 community richness on streptomycete community structure. *FEMS Microbiol Ecol* **83**: 596–
544 606.
- 545 Barder, M.J. and Crawford, D.L. (1981) Effects of carbon and nitrogen supplementation on
546 lignin and cellulose decomposition by a *Streptomyces*. *Can J Microbiol* **27**: 859–863.
- 547 Barker, G. (2005) The character of the New Zealand landsnail fauna and communities: some
548 evolutionary and ecological perspectives. *Rec West Aust Museum, Suppl* **68**: 53–102.
- 549 Bontemps, C., Toussaint, M., Revol, P.-V.V., Hotel, L., Jeanbille, M., Uroz, S., et al. (2013)
550 Taxonomic and functional diversity of *Streptomyces* in a forest soil. *FEMS Microbiol Lett*
551 **342**: 157–167.
- 552 Brown, J.H. (1984) On the relationship between abundance and distribution of species. *Am Nat*
553 **124**: 255–279.
- 554 van Buuren, S. and Groothuis-Oudshoorn, K. (2011) mice: Multivariate Imputation by Chained
555 Equations in R. *J Stat Softw* **045**: 1–67.

- 556 Cary, S.C. and Fierer, N. (2014) The importance of sample archiving in microbial ecology. *Nat*
557 *Rev Microbiol* **12**: 789-790.
- 558 Case, R.J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W.F., and Kjelleberg, S. (2007)
559 Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. *Appl*
560 *Environ Microbiol* **73**: 278–288.
- 561 Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., and Ellison, A.M.
562 (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and
563 estimation in species diversity studies. *Ecol Monogr* **84**: 45–67.
- 564 Chase, J.M., Kraft, N.J.B., Smith, K.G., Vellend, M., and Inouye, B.D. (2011) Using null models
565 to disentangle variation in community dissimilarity from variation in α -diversity. *Ecosphere*
566 **2**: art24.
- 567 Chater, K.F. (2016) Recent advances in understanding *Streptomyces*. *F1000Research* **5**: 2795.
- 568 Choudoir, J.M., Pepe-Ranne, C., and Buckley, H.D. (2018) Diversification of secondary
569 metabolite biosynthetic gene clusters coincides with lineage divergence in *Streptomyces*.
570 *Antibiotics* **7**: 12.
- 571 Choudoir, M.J. and Buckley, D.H. (2018) Phylogenetic conservatism of thermal traits explains
572 dispersal limitation and genomic differentiation of *Streptomyces* sister-taxa. *ISME J* **12**:
573 2176–2186.
- 574 Choudoir, M.J., Doroghazi, J.R., and Buckley, D.H. (2016) Latitude delineates patterns of
575 biogeography in terrestrial *Streptomyces*. *Environ Microbiol* **18**: 4931–4945.
- 576 Choudoir, M.J., Panke-Buisse, K., Andam, C.P., and Buckley, D.H. (2017) Genome surfing as
577 driver of microbial genomic diversity. *Trends Microbiol.* **25**: 624-636.
- 578 Clark, I.M. and Hirsch, P.R. (2008) Survival of bacterial DNA and culturable bacteria in
579 archived soils from the Rothamsted Broadbalk experiment. *Soil Biol Biochem* **40**: 1090–
580 1102.
- 581 Cordero, O.X., Wildschutte, H., Kirkup, B., Proehl, S., Ngo, L., Hussain, F., et al. (2012)
582 Ecological populations of bacteria act as socially cohesive units of antibiotic production and
583 resistance. *Science* **337**: 1228–1231.
- 584 Dari, K., Béchet, M., and Blondeau, R. (1995) Isolation of soil *Streptomyces* strains capable of
585 degrading humic acids and analysis of their peroxidase activity. *FEMS Microbiol Ecol* **16**:
586 115–121.
- 587 Davelos, A.L., Xiao, K., Samac, D.A., Martin, A.P., and Kinkel, L.L. (2004) Spatial variation in
588 *Streptomyces* genetic composition and diversity in a prairie soil. *Microb Ecol* **48**: 601–612.
- 589 Delgado-Baquerizo, M., Bissett, A., Eldridge, D.J., Maestre, F.T., He, J.-Z.Z., Wang, J.-T.T., et
590 al. (2017) Palaeoclimate explains a unique proportion of the global variation in soil bacterial
591 communities. *Nat Ecol Evol* **1**: 1339–1347.
- 592 Dolfing, J. and Feng, Y. (2015) The importance of soil archives for microbial ecology. *Nat Rev*
593 *Microbiol* **13**: 1.

- 594 Dolfing, J., Vos, A., Bloem, J., Ehler, P.A.I., Naumova, N.B., and Kuikman, P.J. (2004)
595 Microbial diversity in archived soils. *Science* **306**: 813–813.
- 596 Dray, S., Blanchet, G., Borcard, D., Guenard, G., Jombart, T., Larocque, G., et al. (2019)
597 adespatial: Multivariate multiscale spatial analysis. [WWW document]. URL
598 <https://CRAN.R-project.org/package=adespatial>.
- 599 Edgar, R.C. (2016) UNOISE2: improved error-correction for Illumina 16S and ITS amplicon
600 sequencing. *bioRxiv* 81257.
- 601 Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves
602 sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.
- 603 Eisenlord, S.D., Zak, D.R., and Upchurch, R.A. (2012) Dispersal limitation and the assembly of
604 soil *Actinobacteria* communities in a long-term chronosequence. *Ecol Evol* **2**: 538–549.
- 605 Ensign, J.C. (1978) Formation, properties, and germination of actinomycete spores. *Annu Rev*
606 *Microbiol* **32**: 185–219.
- 607 Essarioui, A., LeBlanc, N., Kistler, H.C., and Kinkel, L.L. (2017) Plant community richness
608 mediates inhibitory interactions and resource competition between *Streptomyces* and
609 *Fusarium* populations in the rhizosphere. *Microb Ecol* **74**: 157–167.
- 610 Fleming, C.A. (1975) The geological history of New Zealand and its biota. In *Biogeography and*
611 *Ecology in New Zealand*. Kuschel, G. (ed). Dordrecht: Springer Netherlands, pp. 1–86.
- 612 Fukami, T. (2015) Historical contingency in community assembly: Integrating niches, species
613 pools, and priority effects. *Annu Rev Ecol Evol Syst* **46**: 1–23.
- 614 Gallagher, K.A., Wanger, G., Henderson, J., Llorente, M., Hughes, C.C., and Jensen, P.R. (2017)
615 Ecological implications of hypoxia-triggered shifts in secondary metabolism. *Environ*
616 *Microbiol* **16**: 2182–2191.
- 617 Gao, X., Lin, H., Revanna, K., and Dong, Q. (2017) A Bayesian taxonomic classification method
618 for 16S rRNA gene sequences with improved species-level accuracy. *BMC Bioinformatics*
619 **18**: 1–10.
- 620 Gaston, K.J., Blackburn, T.M., and Lawton, J.H. (1997) Interspecific abundance-range size
621 relationships: An appraisal of mechanisms. *J Anim Ecol* **66**: 579–601.
- 622 Griffiths, R.I., Whiteley, A.S., O'Donnell, A.G., and Bailey, M.J. (2000) Rapid method for
623 coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA-
624 and rRNA-based microbial community composition. *Appl Environ Microbiol* **66**: 5488–
625 5491.
- 626 Hallett, L.M., Jones, S.K., MacDonald, A.A.M., Jones, M.B., Flynn, D.F.B., Ripplinger, J., et al.
627 (2016) codyn: An R package of community dynamics metrics. *Methods Ecol Evol* **7**: 1146–
628 1151.
- 629 Harte, J., Blackburn, T., and Ostling, A. (2001) Self-similarity and the relationship between
630 abundance and range size. *Am Nat* **157**: 374–386.
- 631 Heads, M. (1997) Region patterns of biodiversity in New Zealand: One degree analysis of plant

- 632 and animal distributions. *J R Soc New Zeal* **27**: 337–354.
- 633 Hlavac, M. (2015) stargazer: well-formatted regression and summary statistics tables. [WWW
634 document]. URL <https://CRAN.R-project.org/package=stargazer>.
- 635 Hsieh, T.C., Ma, K.H., and Chao, A. (2016) iNEXT: an R package for rarefaction and
636 extrapolation of species diversity (Hill numbers). *Methods Ecol Evol* **7**: 1451–1456.
- 637 Hyatt, D., Chen, G.-L., LoCascio, P., Land, M., Larimer, F., and Hauser, L. (2010) Prodigal:
638 prokaryotic gene recognition and translation initiation site identification. *BMC*
639 *Bioinformatics* **11**: 1–11.
- 640 Inbar, E., Green, S.J., Hadar, Y., and Minz, D. (2005) Competing factors of compost
641 concentration and proximity to root affect the distribution of streptomycetes. *Microb Ecol*
642 **50**: 73–81.
- 643 Ivanova, E.A., Korvigo, I.O., Aparin, B.F., Chirak, E.L., Pershina, E. V, Romaschenko, N.S., et
644 al. (2017) The preservation of microbial DNA in archived soils of various genetic types.
645 *PLoS One* **12**: e0173901.
- 646 Jolley, K.A., Bray, J.E., and Maiden, M.C.J.J. (2018) Open-access bacterial population
647 genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome*
648 *open Res* **3**: 124.
- 649 Kämpfer, P., Glaeser, S.P., Parkes, L., van Keulen, G., and Dyson, P. (2014) The family
650 *Streptomycetaceae*. In *The Prokaryotes: Actinobacteria*. Rosenberg, E., DeLong, E.F.,
651 Lory, S., Stackebrandt, E., and Thompson, F. (eds). Berlin, Heidelberg: Springer Berlin
652 Heidelberg, pp. 889–1010.
- 653 Katoh, K., Asimenos, G., and Toh, H. (2009) Multiple alignment of DNA sequences with
654 MAFFT. *Methods Mol Biol* **537**: 39–64.
- 655 Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., et al.
656 (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–
657 1464.
- 658 Keppel, G., Van Niel, K.P., Wardell-Johnson, G.W., Yates, C.J., Byrne, M., Mucina, L., et al.
659 (2012) Refugia: identifying and understanding safe havens for biodiversity under climate
660 change. *Glob Ecol Biogeogr* **21**: 393–404.
- 661 Kim, B.-J., Kim, C.-J., Chun, J., Koh, Y.-H., Lee, S.-H., Hyun, J.-W., et al. (2004) Phylogenetic
662 analysis of the genera *Streptomyces* and *Kitasatospora* based on partial RNA polymerase β -
663 subunit gene (*rpoB*) sequences. *Int J Syst Evol Microbiol* **54**: 593–598.
- 664 Knight, R., Vrbanc, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J., et al. (2018)
665 Best practices for analysing microbiomes. *Nat Rev Microbiol* **16**: 410–422.
- 666 Koeppel, A.F. and Wu, M. (2013) Species matter: the role of competition in the assembly of
667 congeneric bacteria. *Isme J* **8**: 531–540.
- 668 Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013)
669 Development of a dual-index sequencing strategy and curation pipeline for analyzing

670 amplicon sequence data on the miseq illumina sequencing platform. *Appl Environ*
671 *Microbiol* **79**: 5112–5120.

672 Labeda, D.P., Dunlap, C.A., Rong, X., Huang, Y., Doroghazi, J.R., Ju, K.-S., and Metcalf, W.W.
673 (2017) Phylogenetic relationships in the family *Streptomycetaceae* using multi-locus
674 sequence analysis. *Antonie Van Leeuwenhoek* **110**: 563–583.

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

678 Leinonen, R., Sugawara, H., and Shumway, M. (2010) The sequence read archive. *Nucleic Acids*
679 *Res* **39**: D19–D21.

680 Lloyd, A.B. (1969) Dispersal of streptomycetes in air. *Microbiology* **57**: 35–40.

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

683 Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
684 *EMBnet.journal* **17**: 10–12.

685 Mateo-Tomás, P., Olea, P.P., Moleón, M., Vicente, J., Botella, F., Selva, N., et al. (2015) From
686 regional to global patterns in vertebrate scavenger communities subsidized by big game
687 hunting. *Divers Distrib* **21**: 913–924.

688 McDonald, B.R. and Currie, C.R. (2017) Lateral gene transfer dynamics in the ancient bacterial
689 genus *Streptomyces*. *MBio* **8**: 1–12.

690 McGlone, M.S. (1985) Plant biogeography and the late Cenozoic history of New Zealand. *New*
691 *Zeal J Bot* **23**: 723–749.

692 McGlone, M.S., Richardson, S.J., and Jordon, G.J. (2010) Comparative biogeography of New
693 Zealand tree species richness, height, leaf traits and range sizes. *N Z J Ecol* **34**: 1–15.

694 McMurdie, P.J. and Holmes, S. (2013) phyloseq: An R package for reproducible interactive
695 analysis and graphics of microbiome census data. *PLoS One* **8**: e61217.

696 Nemergut, D.R., Costello, E.K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S.K., et al.
697 (2011) Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* **13**: 135–
698 144.

699 Newnham, R., McGlone, M., Moar, N., Wilmshurst, J., and Vandergoes, M. (2013) The
700 vegetation cover of New Zealand at the Last Glacial Maximum. *Quat Sci Rev* **74**: 202–214.

701 Niu, G., Chater, K.F., Tian, Y., Zhang, J., and Tan, H. (2016) Specialised metabolites regulating
702 antibiotic biosynthesis in *Streptomyces* spp. *FEMS Microbiol Rev* **40**: 554–573.

703 Ochman, H., Elwyn, S., and Moran, N.A. (1999) Calibrating bacterial evolution. *PNAS* **96**:
704 12638–12643.

705 Ochman, H. and Wilson, A.C. (1987) Evolution in bacteria: Evidence for a universal substitution
706 rate in cellular genomes. *J Mol Evol* **26**: 74–86.

707 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2019)
708 vegan: Community Ecology Package. [WWW document]. URL [https://CRAN.R-](https://CRAN.R-project.org/package=vegan)
709 [project.org/package=vegan](https://CRAN.R-project.org/package=vegan).

710 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM:
711 assessing the quality of microbial genomes recovered from isolates, single cells, and
712 metagenomes. *Genome Res* **25**: 1043–1055.

713 Price MN, Dehal PS, A.A. (2010) FastTree 2 - Approximately maximum-likelihood trees for
714 large alignments. *PLoS One* **5**: e9490.

715 Provan, J. and Bennett, K.D. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends*
716 *Ecol Evol* **23**: 564–571.

717 R Core Team (2018) R: A language and environment for statistical computing. [WWW
718 document]. URL <https://www.R-project.org/>.

719 Renaud, G., Stenzel, U., Maricic, T., Wiebe, V., Kelso, J., Cox, M., et al. (2015) deML: robust
720 demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* **11**:
721 770–772.

722 Rey, T. and Dumas, B. (2017) Plenty is no plague: *Streptomyces* symbiosis with crops. *Trends*
723 *Plant Sci* **22**: 30–37.

724 Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016) VSEARCH: a versatile open
725 source tool for metagenomics. *PeerJ* **4**: e2584.

726 Rong, X. and Huang, Y. (2012) Taxonomic evaluation of the *Streptomyces hygroscopicus* clade
727 using multilocus sequence analysis and DNA–DNA hybridization, validating the MLSA
728 scheme for systematics of the whole genus. *Syst Appl Microbiol* **35**: 7–18.

729 Ruddick, S.M. and Williams, S.T. (1972) Studies on the ecology of actinomycetes in soil V.
730 Some factors influencing the dispersal and adsorption of spores in soil. *Soil Biol Biochem* **4**:
731 93–103.

732 Sarmiento-Vizcaíno, A., Braña, A.F., González, V., Nava, H., Molina, A., Llera, E., et al. (2016)
733 Atmospheric dispersal of bioactive *Streptomyces albidoflavus* strains among terrestrial and
734 marine environments. *Microb Ecol* **71**: 375–386.

735 Schlatter, D., Fubuh, A., Xiao, K., Hernandez, D., Hobbie, S., and Kinkel, L. (2009) Resource
736 amendments influence density and competitive phenotypes of *Streptomyces* in soil. *Microb*
737 *Ecol* **57**: 413–420.

738 Schlatter, D.C., DavelosBaines, A.L., Xiao, K., and Kinkel, L.L. (2013) Resource use of
739 soilborne *Streptomyces* varies with location, phylogeny, and nitrogen amendment. *Microb*
740 *Ecol* **66**: 961–971.

741 Schliep, K.P. (2010) phangorn: phylogenetic analysis in R. *Bioinformatics* **27**: 592–593.

742 Seipke, R.F., Kaltenpoth, M., and Hutchings, M.I. (2012) *Streptomyces* as symbionts: an
743 emerging and widespread theme? *FEMS Microbiol Rev* **36**: 862–876.

744 Smith, B., Wilson, J.B., and A, J.B. (1996) A consumer's guide to evenness indices. *Oikos* **76**:

- 745 70–82.
- 746 Spiker, J.K., Crawford, D.L., and Thiel, E.C. (1992) Oxidation of phenolic and non-phenolic
747 substrates by the lignin peroxidase of *Streptomyces viridosporus* T7A. *Appl Microbiol*
748 *Biotechnol* **37**: 518–523.
- 749 Soil Survey Staff. (2014) Soil survey laboratory methods manual. Soil survey investigations
750 report no. 42, Version 5.0. R. Burt and Soil Survey Staff (eds.). Lincoln, NE: U.S.
751 Department of Agriculture, Natural Resources Conservation Service.
- 752 Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W., Murray, C.J., et al. (2013)
753 Quantifying community assembly processes and identifying features that impose them. *Isme*
754 *J* **7**: 2069–2079.
- 755 Stegen, J.C., Lin, X., Fredrickson, J.K., and Konopka, A.E. (2015) Estimating and mapping
756 ecological processes influencing microbial community assembly. *Front Microbiol* **6**: 1–15.
- 757 Stegen, J.C., Lin, X., Konopka, A.E., and Fredrickson, J.K. (2012) Stochastic and deterministic
758 assembly processes in subsurface microbial communities. *Isme J* **6**: 1653–1664.
- 759 Swenson, N.G. (2014) Functional and phylogenetic ecology in R. New York: Springer-Verlag.
- 760 Swenson, N.G. (2009) Phylogenetic resolution and quantifying the phylogenetic diversity and
761 dispersion of communities. *PLoS One* **4**: e4390.
- 762 Tonkin, J.D., Death, R.G., Muotka, T., Astorga, A., and Lytle, D.A. (2018) Do latitudinal
763 gradients exist in New Zealand stream invertebrate metacommunities? *PeerJ* **6**: e4898.
- 764 Traxler, M.F. and Kolter, R. (2015) Natural products in soil microbe interactions and evolution.
765 *Nat Prod Rep* **32**: 956–970.
- 766 Tzeneva, V.A., Salles, J.F., Naumova, N., de Vos, W.M., Kuikman, P.J., Dolfing, J., and Smidt,
767 H. (2009) Effect of soil sample preservation, compared to the effect of other environmental
768 variables, on bacterial and eukaryotic diversity. *Res Microbiol* **160**: 89–98.
- 769 Vetsigian, K., Jajoo, R., and Kishony, R. (2011) Structure and evolution of *Streptomyces*
770 interaction networks in soil and in silico. *PLoS Biol* **9**: e1001184.
- 771 Viaene, T., Langendries, S., Beirinckx, S., Maes, M., and Goormachtig, S. (2016) *Streptomyces*
772 as a plant’s best friend? *FEMS Microbiol Ecol* **92**: fiw119.
- 773 Waksman, S.A., Reilly, H.C., and Harris, D.A. (1948) *Streptomyces griseus* (Krainsky)
774 Waksman and Henrici. *J Bacteriol* **56**: 259–269.
- 775 Wallis, G.P. and Trewick, S.A. (2009) New Zealand phylogeography: evolution on a small
776 continent. *Mol Ecol* **18**: 3548–3580.
- 777 Webb, C.O., Ackerly, D.D., McPeck, M.A., and Donoghue, M.J. (2002) Phylogenies and
778 community ecology. *Annu Rev Ecol Syst* **33**: 475–505.
- 779 De Wever, A., Leliaert, F., Verleyen, E., Vanormelingen, P., Van Der Gucht, K., Hodgson, D.A.,
780 et al. (2009) Hidden levels of phylodiversity in Antarctic green algae: Further evidence for
781 the existence of glacial refugia. *Proc R Soc B Biol Sci* **276**: 3591–3599.

782 Wong, R.G., Wu, J.R., and Gloor, G.B. (2016) Expanding the UniFrac toolbox. *PLoS One* **11**:
783 e0161196.

784 Wright, E.S. (2016) Using DECIPHER v2.0 to analyze big biological sequence data in R. *R J* **8**:
785 352–359.

786 Wright, E.S. and Vetsigian, K.H. (2016) Inhibitory interactions promote frequent bistability
787 among competing bacteria. *Nat Commun* **7**: 1–7.

788 Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014) PEAR: a fast and accurate Illumina
789 Paired-End reAd mergeR. *Bioinformatics* **30**: 614–620.

790 Zhang, Y., Bignell, D.R.D., Zuo, R., Fan, Q., Huguet-Tapia, J.C., Ding, Y., and Loria, R. (2016)
791 Promiscuous pathogenicity islands and phylogeny of pathogenic *Streptomyces* spp. *Mol*
792 *Plant-Microbe Interact* **29**: 640–650.

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

Table 1. Spearman rank correlations between *Streptomyces* diversity and select variables estimated for the NZ soil samples ($n = 15$). Diversity measurements reported are $\log_{10}$ of observed *rpoB* OTUs (OTU_{obs}), Faith's phylogenetic diversity (PD), community evenness (E_{var}), and net relatedness index (NRI). Significant results ($p < 0.05$) are indicated by bold font.

	pH	MAR	MAT	MAWS	Latitude	Plant richness
OTU_{obs}	0.42	-0.59	0.08	-0.41	-0.08	-0.44
PD	0.77	-0.32	0.64	-0.03	0.40	-0.79
E_{var}	0.38	-0.67	0.15	-0.28	-0.25	-0.34
NRI	0.30	-0.15	0.32	0.11	0.30	-0.73

MAR = mean annual rainfall (mm), MAT = mean annual temperature ($^{\circ}\text{C}$), MAWS = mean annual wind speed (m/s).

Table 2. Names and characteristics of soil sites sampled across New Zealand.

Number	Code	Site	Latitude	Longitude	pH	MAR (mm) ^a	MAT (°C) ^a	MAWS (m/s) ^a	Elevation (m) ^b
1	LB4N	Great Barrier Island 4N	-36.33	175.49	5.2	1294.8	14.3	4	541
2	LB1B	Great Barrier Island 1B	-36.33	175.49	5.3	1294.8	14.3	4	541
3	ACOR	Oteha Rowe	-36.73	174.69	5.5	1067.1	13.8	2.6	11
4	NZOH	Ohinewai	-37.98	175.76	5	1018.1	14	1.8	130
5	NZRD	Rangipo Desert	-39.4	175.8	4.8	2843.2	6.7	4.2	970
6	NZMW	Mangaweka	-39.8	175.8	5.2	1134.7	11.1	2.3	357
7	NZPA	Pangatotara	-41.18	172.9	5	1069.2	12.6	1.2	47
8	NZHS	Hope Saddle Lookout	-41.63	172.72	4.3	2906.2 ^c	12.2 ^c	3.5 ^c	603
9	NZKP	Kaikoura Peninsula	-42.42	173.7	5.6	1031.2 ^c	12.2	4.3	70
10	NZWS	Woodstock Road	-42.82	170.9	4.5	2197.2	12.2	3.8	22
11	NZHV	Haast Valley	-43.94	169.15	4.6	2906.2	11.3	2.8	29
12	NZCF	Cameron Flats	-44.16	169.29	4.4	727.6	9.2	2.7	346
13	NZKG	Kawarau Gorge	-44.7	169.13	6.1	454.6	10.6	2.8	287
14	NZKR	Kingsland Rd	-45.15	168.75	5.5	735.3	9.7	3.5	310
15	NZRT	Red Tussock Conservation Area	-45.56	168.04	4.4	1031.2	9.5	2.6	486

^aMean annual rainfall (MAR), temperature (MAT), and wind speed (MAWS) data were accessed from the National Institute of Water and Atmospheric Research (NIWA) CliFlo (<https://cliflo.niwa.co.nz/>) database weather station locations nearest (35 ± 21 km) to sites where soils were sampled.

^bElevation was accessed from <https://www.maps.ie/coordinates.html>.

^cValues were imputed using the R package mice (see Materials and Methods for details).

Table 3. List of community assembly processes and their overall contributions to *Streptomyces* community dynamics.

Process	No. pairwise comparisons	Percentage of total
Variable selection	72	34
Homogeneous selection	14	7
Dispersal limitation	12	6
Homogenizing dispersal	62	29
Undominated ^a	50	24

^aNeither selection nor dispersal dominates.

Table 4. Proportion of *rpoB* OTUs classified as *Streptomyces* spp. or other taxa.

Taxon	No. OTUs	Percentage of reads
<i>Streptomyces</i>	1,287	~91%
Non- <i>Streptomyces</i> within <i>Streptomycetaceae</i>	10	<1%
Non-target ^a	2,025	~9%

^aIncludes *rpoB* and *rpoC* gene sequences.

Figure 1. Association between *Streptomyces* phylogenetic diversity (PD) and environmental parameters. MAR = mean annual rainfall; MAT = mean annual temperature; r = spearman rank correlation.

Figure 2. *Streptomyces* community pairwise dissimilarity metrics regressed onto pairwise geographic distance. The presence/absence formula for Jaccard distance (A) and abundance-weighted Bray-Curtis (B) metrics were calculated between pairs of sites and plotted against pairwise geographic distance (km). We also calculated dissimilarity metrics that incorporate phylogenetic information between samples using the original weighted UniFrac (wUniFrac) (C) and GUniFrac method (wGUniFrac) (D) that accounts for overweighting of proportional abundances of the original weighted UniFrac method (Wong *et al.*, 2016). Grey dashed lines in panels C and D represent generalized additive models of the formula $y \sim \ln(x + 1)$ and transparent grey shading surrounding the regression line represents upper and lower 95 % confidence intervals.

Figure 3. Multidimensional scaling (PCoA) ordinations of *Streptomyces rpoB* and 16S gene pairwise community dissimilarity among New Zealand sites. The proportion of variation explained by each coordinate axis is presented in parentheses on each axis. Sites located within New Zealand's North or South Island are represented by circles and triangles, respectively. The green fill color represents pH, which ranges from 4.08 (light green) to 6.09 (dark green). Numbers inside symbols indicate the site number reported in Table 2 and Figure 4. The metric used to calculate pairwise dissimilarity is reported in the plot title. uUniFrac = unweighted UniFrac metric.

Figure 4. Distribution and diversity of *Streptomyces rpoB* OTUs in soils along a latitudinal transect across New Zealand. Net relatedness index (NRI) regressed on latitude (A) across the 15 study sites in New Zealand (B). The pairwise number of shared *rpoB* OTUs among sites (C). The distribution of range (km) values for each *rpoB* OTU for the observed and null model data sets (D). The relationship between natural log transformed mean *Streptomyces* OTU relative abundance and proportion of occupied sites (E). OTU sample range in the richness null model panel represents the average of 1,000 null model iterations.

Figure 5. The association of *Streptomyces* community beta diversity with pH and NZ plant species richness within 100 km of the NZ sites. The first and second ordination axes of a Jaccard-based *Streptomyces* community PCoA were regressed against pH (A) and plant species richness (B).