

Article

Diversification of Secondary Metabolite Biosynthetic Gene Clusters Coincides with Lineage Divergence in *Streptomyces*

Mallory J. Choudoir , Charles Pepe-Ranney and Daniel H. Buckley *

School of Integrative Plant Science, Bradfield Hall 705, Cornell University, Ithaca, NY 14853, USA; mjchoudoir@gmail.com (M.J.C.); chuck.peperanney@gmail.com (C.P.-R.)

* Correspondence: dbuckley@cornell.edu; Tel.: +1-607-255-1716

Received: 11 January 2018; Accepted: 7 February 2018; Published: 13 February 2018

Abstract: We have identified *Streptomyces* sister-taxa which share a recent common ancestor and nearly identical small subunit (SSU) rRNA gene sequences, but inhabit distinct geographic ranges demarcated by latitude and have sufficient genomic divergence to represent distinct species. Here, we explore the evolutionary dynamics of secondary metabolite biosynthetic gene clusters (SMGCs) following lineage divergence of these sister-taxa. These sister-taxa strains contained 310 distinct SMGCs belonging to 22 different gene cluster classes. While there was broad conservation of these 22 gene cluster classes among the genomes analyzed, each individual genome harbored a different number of gene clusters within each class. A total of nine SMGCs were conserved across nearly all strains, but the majority (57%) of SMGCs were strain-specific. We show that while each individual genome has a unique combination of SMGCs, this diversity displays lineage-level modularity. Overall, the northern-derived (NDR) clade had more SMGCs than the southern-derived (SDR) clade (40.7 ± 3.9 and 33.8 ± 3.9 , mean and S.D., respectively). This difference in SMGC content corresponded with differences in the number of predicted open reading frames (ORFs) per genome (7775 ± 196 and 7093 ± 205 , mean and S.D., respectively) such that the ratio of SMGC:ORF did not differ between sister-taxa genomes. We show that changes in SMGC diversity between the sister-taxa were driven primarily by gene acquisition and deletion events, and these changes were associated with an overall change in genome size which accompanied lineage divergence.

Keywords: *Streptomyces*; biogeography; comparative genomics; diversification; secondary metabolite biosynthetic gene clusters; SMGC; natural products

1. Introduction

Microbial secondary metabolism encapsulates a remarkable diversity of natural products with an extensive range of biological activities. Secondary metabolites differ from primary metabolites in that they are not involved in essential catabolic and anabolic activities required for normal growth and reproduction, but may contribute significantly to an individual's fitness [1]. While primary metabolic pathways are often conserved deeply within a phylogeny, secondary metabolic pathways are more divergent, often being species or strain-specific, with conservation sometimes observed among closely related species and genera [2]. This phylogenetic pattern suggests an adaptive role for secondary metabolites, and if secondary metabolism pathways provide adaptive benefits, their evolution might drive or reinforce evolutionary processes that result in microbial diversification and speciation [3].

The values of natural products to humanity are widely recognized, yet because most research has focused on their discovery and human-centric relevance, we are still far from understanding their biological role in natural systems. The discovery and application of antibiotics revolutionized medicine in the 1940's, sparking the "golden age" of antibiotics between 1950 and 1960, during

which time approximately half of the microbial-derived drugs we use today were discovered [4]. Presently, thousands of bioactive compounds with antibacterial, antifungal, and antitumor activities are cataloged [5,6], and yet these represent only a fraction of actual natural product diversity [7]. In addition, microbial populations in situ are exposed to natural products at concentrations far below the lethal clinical dose, and hence these compounds may serve different functions in the environment from those observed during therapeutic application. We know that secondary metabolites can mediate diverse biotic interactions including mutualistic interactions, competition for nutrients, metal scavenging, and plant-microbe and insect-microbe symbioses [8–10], which can all have profound impacts on microbial fitness. It is clear that natural products must have considerable impacts on microbial ecology and evolution and that understanding the biology and evolutionary history of natural products will enhance our ability to use these agents therapeutically.

Soil-dwelling actinomycetes are the predominant source of microbial-derived therapeutic natural products, and the majority of described bioactive compounds originate from the genus *Streptomyces* [6,7]. The *Streptomyces* life cycle resembles that of many fungi, consisting of filamentous growth, formation of mycelia, and production of aerial hyphae and spores. Indeed, *Streptomyces* were thought to be an intermediary between bacteria and fungi until as recently as the 1950's [11]. However, *Streptomyces* are Gram-positive *Actinobacteria* with long linear chromosomes that have a high G+C content [12]. Traditionally, *Streptomyces* species are often known to produce several secondary metabolites when grown in culture. Genome sequencing, however, reveals that *Streptomyces* contain an enormous reservoir of “cryptic secondary metabolites” which are not expressed under standard laboratory conditions [13]. For instance, while *Streptomyces coelicolor* A3(2) was known to produce several well characterized secondary metabolites, genome sequencing discovered that it actually contained >20 biosynthetic gene clusters not expressed when grown in culture [14]. Genes within secondary metabolite biosynthetic gene clusters (SMGCs) are co-localized as operons within discrete genomic regions. SMGCs have recognizable functional domains, so SMGCs are readily predicted using bioinformatics [15]. Phylogenetic conservation of SMGCs between closely related microbes suggests that these secondary metabolites may have ecological roles which facilitate microbial diversification [16–18].

The evolutionary and ecological processes that govern SMGC diversity remain largely unexplored. The richness of SMGCs within soils is linked to both edaphic and biotic factors [19,20]. For example, the production of antibiotics by *Streptomyces* isolated from prairie soils is highly variable between strains and correlates poorly with 16S rRNA gene phylogeny, suggesting a role of selection acting at small spatial scales [21,22]. Conversely, at larger spatial scales, *Streptomyces* SMGC composition varies in relation to both spatial distance and environmental dissimilarity [23]. Furthermore, evidence within *Streptomyces* for endemism at inter-continental and regional geographic scales [16,24,25] suggests limits to dispersal at large spatial scales. These data indicate that both adaptive and neutral processes contribute to patterns of SMGC biogeography.

Microbial biogeography is readily explored with geographically explicit microbial culture collections (reviewed in [26]), and the genus *Streptomyces* is an ideal model system to evaluate the influence of SMGC dynamics on patterns of diversification. We previously assembled a culture collection of *Streptomyces* from sites spanning the United States, and we observed evidence for dispersal limitation, as well as a latitudinal gradient of species richness and intraspecific nucleotide diversity [27,28]. From this culture collection, we have identified *Streptomyces* sister-taxa that have geographic ranges delimited by latitude and have patterns of gene flow and genomic diversity consistent with their diversification from a recent common ancestor [28,29]. Here, we evaluate changes in SMGC diversity between these *Streptomyces* sister-taxa to explore SMGC evolutionary dynamics during the divergence of *Streptomyces* species.

2. Results and Discussion

2.1. Genomic Divergence between *Streptomyces* Sister-Taxa

We used comparative genomics to analyze patterns of genomic diversity and SMGC content in 24 *Streptomyces* representing sister-taxa and related strains. These strains were identified through a phylogenetic analysis of a *Streptomyces* culture collection [28] generated from soils of ecologically similar grassland sites spanning 6000 km across the continental United States (Figure 1, Table S1). The sister-taxa, which we have designated the northern-derived (NDR) and southern-derived (SDR) clades, were defined by their geographic range and genomic similarity (Figure 1). Each clade contains ten isolates, and an additional four genomes represent intermediate (INT) taxa.

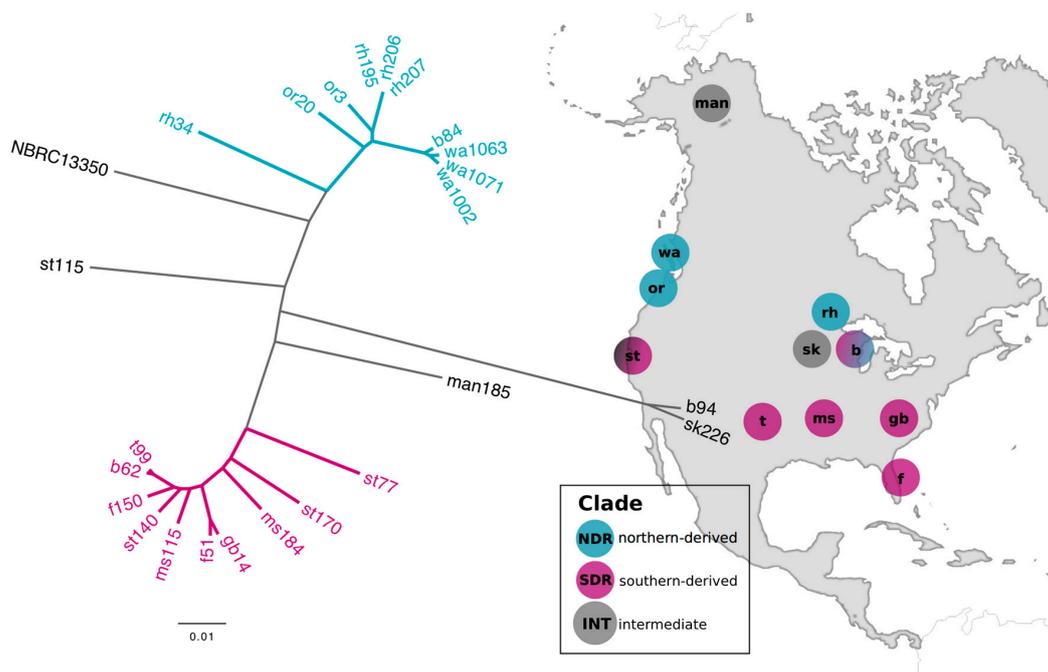


Figure 1. The northern-derived (NDR) and southern-derived (SDR) clades are closely related sister-taxa and yet were isolated from soils of different latitude. The un-rooted tree was constructed from multiple whole genome alignments with maximum likelihood and a GTRGAMMA model of evolution. Scale bar represents nucleotide substitutions per site. Colored branches depict the northern-derived (NDR) and southern-derived (SDR) clades. Strain names reflect the sample site they were isolated from (Table S1). Genome NBRC 13350 is the publically available type strain *Streptomyces griseus* subsp. *griseus* NBRC 13350. Sample locations are shown in the right panel and labeled with the site code. Circles are colored to reflect the geographic distribution of clades. (Figure modified from [29]).

Assembled genomes are 7.5–9.1 Mb with a G+C content of 71.4–72.5% and 6776–8078 predicted open reading frames (ORFs) (Table S2). The core gene content across all 24 strains is comprised of 3234 orthologous genes (representing 2778 single-copy genes), with a total of 22,054 genes in the overall pan-genome. All isolates affiliate taxonomically with the *Streptomyces griseus* species cluster [30] and share >90% average nucleotide identity (ANI) with the type strain *Streptomyces griseus* subsp. *griseus* NBRC 13350 (Figure 1).

The NDR core genome is comprised of 4234 genes, and the SDR core genome is comprised of 4400 genes. The NDR and SDR clades share a recent phylogenetic ancestor and have nearly identical 16S rRNA genes (inter-lineage nucleotide dissimilarity of 0–0.21% between strains). Strains within each clade have a whole genome ANI value ranging from 95.6% to 99.9%, while the ANI between strains of NDR and SDR range from 92.6% to 93.3% (Figure 1). Distinct microbial species are

typically distinguished by ANI in the range of 95–96% [31]. Comparative population genomics reveals signatures of genomic differentiation and gene flow limitation between NDR and SDR consistent with expectations of allopatric diversification [29]. Collectively, these results indicate that NDR and SDR clades represent distinct microbial species which have recently diverged from a common ancestor.

2.2. Secondary Metabolite Biosynthetic Gene Cluster (SMGC) Identification and Classification

We used antiSMASH [32] to identify SMGCs in the genomes of our *Streptomyces* sister-taxa. To assess the novelty of these SMGCs, we utilized antiSMASH's downstream annotation pipeline, which annotates SMGCs based on similarity to genes and pathways present within the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database. The antiSMASH pipeline annotated 120 SMGCs across the 24 strains (Table S3). Each genome had between 28 and 47 SMGCs which ranged in size from 1 to 137 Kb (20.9 ± 15.7 Kb, mean \pm S.D., respectively) (Figure 2). This range in SMGC content is consistent with the results obtained from previous genomic surveys of *Streptomyces* [14,33–36]. The NDR clade has a greater number of SMGCs per genome than the SDR clade (40.7 ± 3.9 , 33.8 ± 3.9 , mean \pm S.D., respectively; t -test, $p < 0.001$; Figure 2a). The NDR clade also has a greater number of ORFs per genome than the SDR clade (7775 ± 196 and 7093 ± 205 , mean and S.D., respectively; t -test, $p < 0.001$; Table S2). Correspondingly, NDR strains also have larger genomes than SDR strains (8.7 ± 0.25 Mb and 7.9 ± 0.21 Mb, mean \pm S.D., respectively; t -test, $p < 0.001$; Table S2). We observed a strong positive correlation between genome size and number of SMGCs across all genomes examined (Pearson's $r = 0.66$, $p < 0.001$).

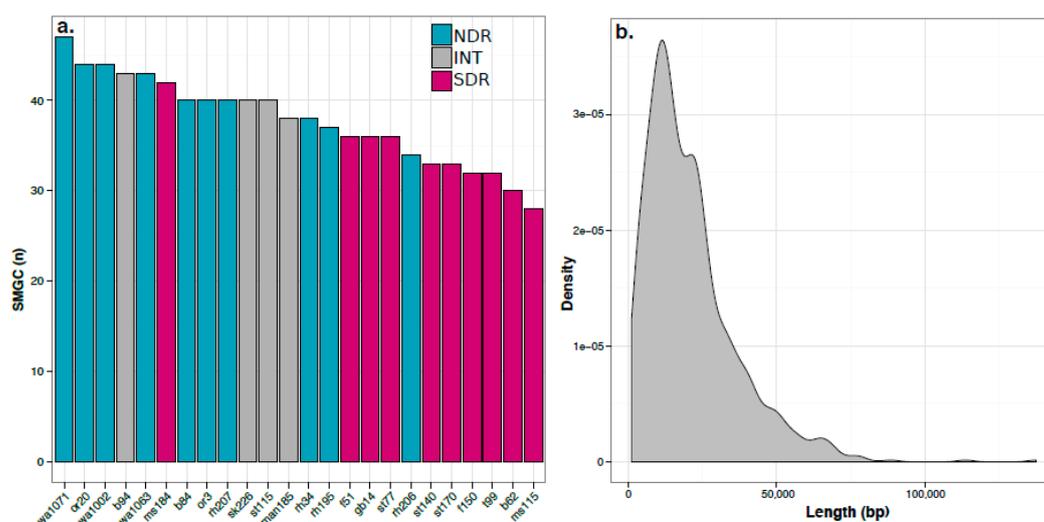


Figure 2. NDR strains have more secondary metabolite biosynthetic gene clusters (SMGCs) than SDR strains (t -test, $p < 0.001$). (a). Bars indicate the number of SMGCs identified in each genome and are colored according to clade affiliation, and genome names reflect the site of isolation as identified in Table S1; (b). Kernel density plot shows the distribution of SMGC length (bp).

Only 21% ($n = 25$) of the MIBiG-annotated SMGCs represent well-characterized biosynthetic gene clusters (in which $\geq 70\%$ of the genes in a SMGC show similarity to genes within the most similar known cluster from the MIBiG database) (Table S3). In addition, each genome harbors five to 25 potentially novel SMGCs with low similarity to biosynthetic pathways within the MIBiG database. These findings indicate that the diversity of *Streptomyces* SMGCs found within public databases remains low and that a vast reservoir of *Streptomyces* SMGC diversity remains to be characterized within natural populations.

The SMGCs predicted by antiSMASH within our *Streptomyces* sister-taxa encompass 22 classes of natural products. Most of these classes, including bacteriocin, butyrolactones, ectoine, lantipeptide,

melanin, non-ribosomal peptide synthases (NRPS), siderophore, polyketide synthases (PKS), and terpene gene clusters, are widely conserved at the genus level [2]. The most abundant SMGC classes in our genomes are NRPS and terpene clusters (Figure 3, Table S3). Many of the predicted gene clusters are NRPS-PKS hybrids (Table S3). Given the similar structure and activity between NRPS and PKS [37], it is unsurprising that hybrid NRPS-PKS clusters are commonly detected in *Streptomyces* genomes [38,39]. Most SMGC classes are present in both NDR and SDR clades, but the relative abundance of each class differs between genomes, as well as between clades (Figure 3). We observe the significant enrichment of melanin and ladderane gene clusters in NDR compared to SDR (*t*-test with Bonferroni correction, $p < 0.002$). Additionally, NDR genomes harbor linaridin gene clusters, which are entirely absent from SDR genomes (Figure 3) but are found in the type strain *Streptomyces griseus* NBRC subsp. *griseus* 13350 [40]. Interestingly, antiSMASH did not identify aminoglycoside biosynthetic clusters in our *Streptomyces* isolates, and all of these genomes presumably lack genes for streptomycin biosynthesis (Figure 3). Schatz and Waksman reported the isolation of streptomycin from *Streptomyces griseus* in 1944, and this was the first antibiotic used to successfully combat tuberculosis [41]. However, not all *Streptomyces griseus* isolates produce streptomycin [42,43].

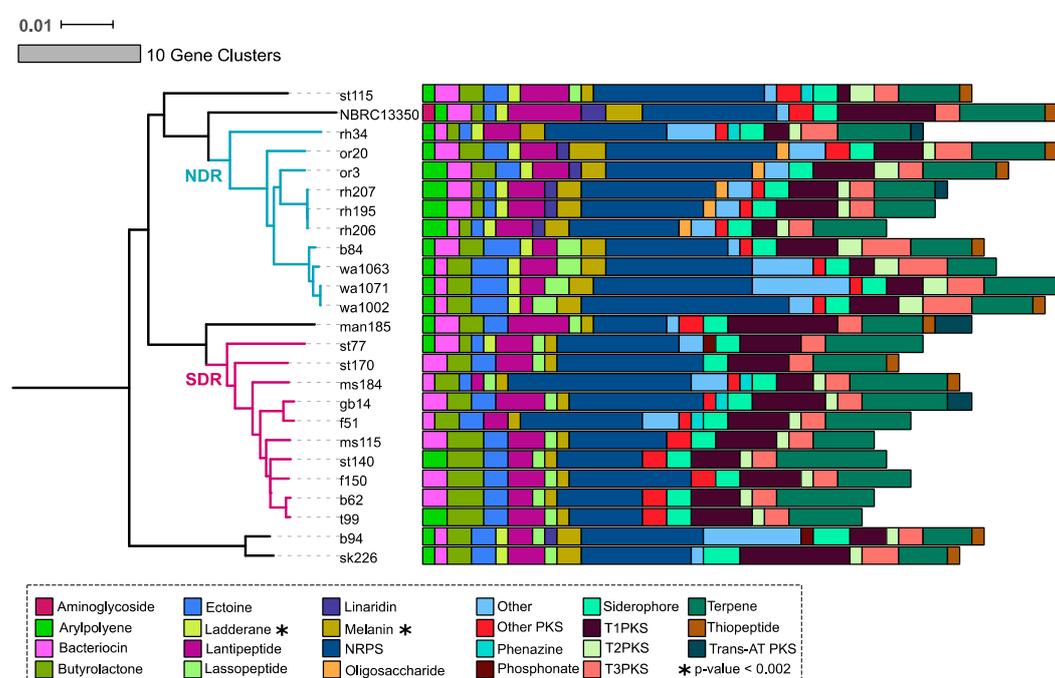


Figure 3. A total of 22 SMGC classes were observed in NDR and SDR genomes by antiSMASH [32]. The tree reflects phylogenetic relationships between *Streptomyces* sister-taxa genomes and was constructed from multiple whole genome alignments (see Figure 1). Scale bar represents nucleotide substitutions per site. Tree branches are colored according to clade affiliation. Bars depict the number of gene clusters belonging to each class for each genome. Colors illustrate gene cluster class as provided by the legend. Asterisks note gene cluster classes that are significantly enriched between clades (*t*-test and Bonferroni correction for multiple comparisons, $p < 0.002$).

2.3. Core and Accessory SMGCs of *Streptomyces* Sister-Taxa

Comparative population genomics and pan-genome analyses can offer powerful insights into the processes underlying species divergence [44,45]. Given that many of our SMGCs have low similarity to biosynthetic pathways in public databases, we determined shared orthologous SMGCs within our genomes using an annotation-independent approach that compares SMGCs based on similarity in nucleotide composition and gene content (see Materials and Methods). This approach identified 310 non-redundant SMGCs within the pan-genome of all 24 strains (Figures 4 and 5); this number

is greater than the number of MIBiG-annotated SMGCs because it classified both known and unknown pathways into distinct non-redundant gene clusters. Only two SMGCs are conserved in all 24 genomes, an ectoine gene cluster and the siderophore desferrioxamine B (Figure 6). Desferrioxamine siderophores are commonly observed in other species of *Streptomyces* and acintomycetes [46,47].

We observed that core SMGC content increased with phylogenetic similarity, but that more than half of the SMGCs were strain-specific (Figures 4 and 5). NDR and SDR shared nine core SMGCs (present in $\geq 80\%$ of genomes), while NDR strains shared 11 core SMGCs (nine in the conserved core and two in the NDR-specific core), and SDR strains shared 15 core SMGCs (nine in the conserved core and six in the SDR-specific core) (Figure 6). In addition, there were 158 accessory SMGCs (present in $< 80\%$ genomes) in NDR and 114 accessory SMGCs in SDR (Figure 4). Most SMGCs were observed at low to intermediate frequencies (Figures 4 and 5), and 177 SMGCs were strain-specific, with each *Streptomyces* genome harboring one to 19 exclusive SMGCs. These estimates are generally consistent with previous observations that indicate each different *Streptomyces* species will harbor a distinct repertoire of natural product pathways [17]. For example, Seipke [36] estimated 18 core SMGCs for six *Streptomyces albus* isolates. However, despite the phylogenetic conservation of core SMGC content, even *Streptomyces* with identical 16S rRNA gene sequences can have distinct secondary metabolite profiles [48], indicating that SMGC content exhibits significant strain to strain variability within a species. Thus, we propose that core SMGCs reflect the shared evolutionary history of *Streptomyces* genomes, while patterns of the accessory SMGC carriage suggest lineage and strain-specific processes across more recent evolutionary time scales.

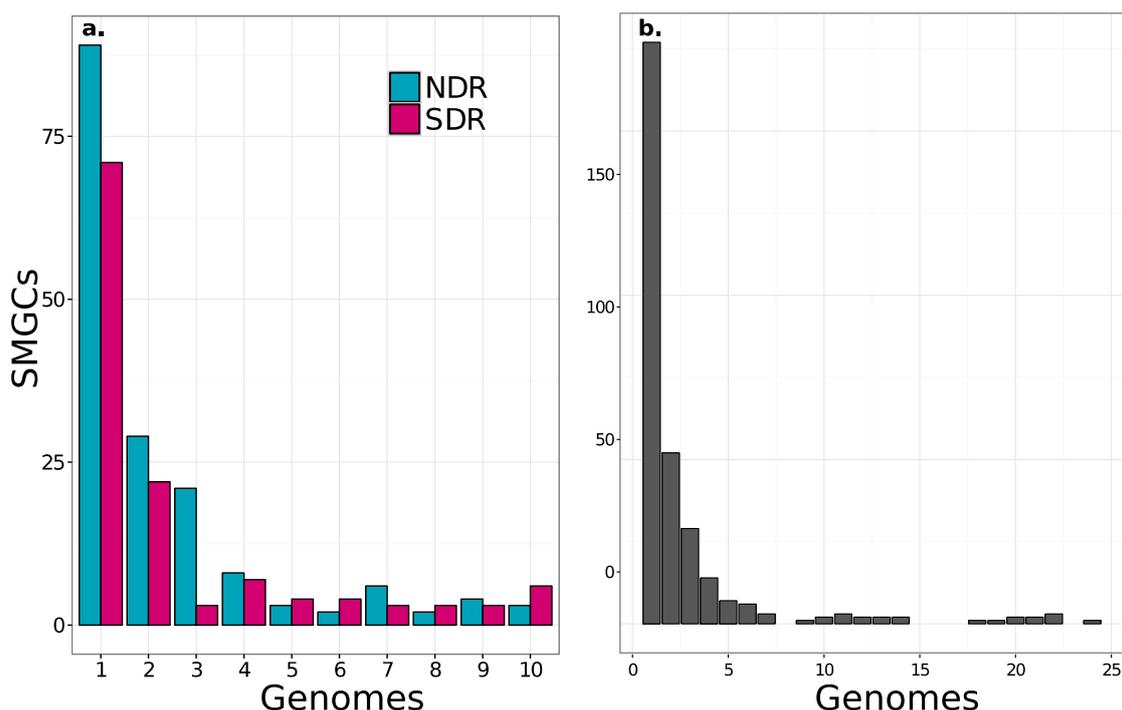


Figure 4. The frequency distribution of SMGCs across strains shows that most SMGCs are strain-specific and fewer are species-specific. Results are shown both for NDR and SDR. (a) and for all 24 genomes; (b). Non-redundant orthologous SMGCs were defined using our annotation-independent approach (see Materials and Methods).

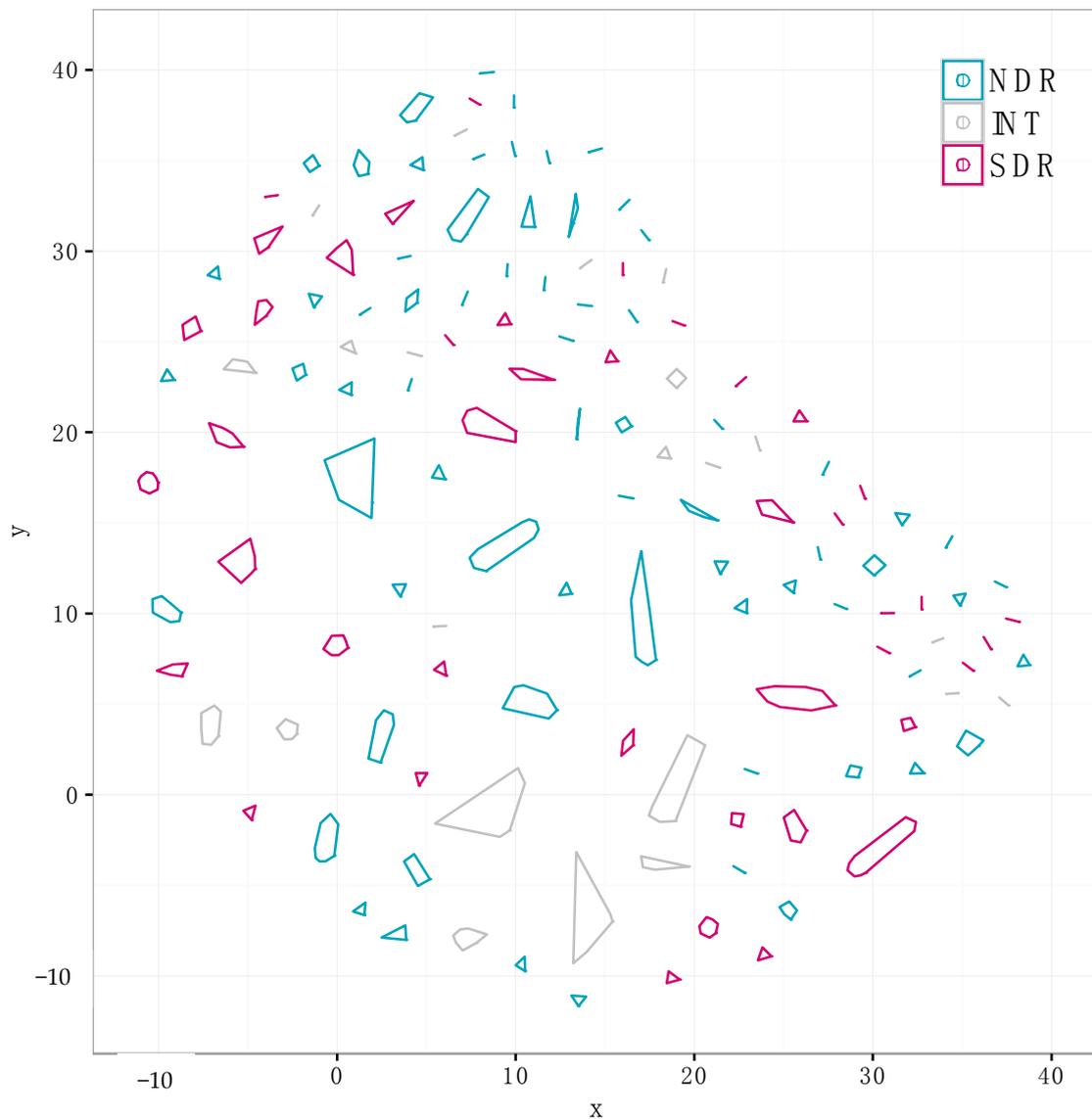


Figure 5. We identified 310 non-redundant distinct SMGCs using our annotation-independent gene clustering approach (see Materials and Methods). Each point represents a unique SMGC from a single genome, and colors correspond to clade affiliation. SMGCs with a similar gene composition are clustered spatially, and cluster membership is depicted with polygons. The same data is presented in a different network diagram in Figure S1.

2.4. Evolutionary Dynamics of Core and Accessory SMGCs

To address potential lineage-specific mechanisms of divergence, we next evaluated the evolutionary dynamics of SMGCs. Most shared SMGCs occur within rather than between clades (Figures 5 and S1). A total of 78 SMGCs are shared among two or more NDR genomes, and 55 are shared among SDR genomes, but only 37 SMGCs are shared across clade boundaries (i.e., found in both NDR and SDR genomes). Furthermore, network analysis reveals unique patterns of SMGC sharing that manifests as nodes of connectivity within clades (Figure S1). This network indicates that there is a core set of SMGC content which links NDR and SDR and which must be ancestral, that there is a clade-specific core set of SMGCs which link the strains of each clade together based on shared SMGC content, and that there are a large number of strain-specific SMGCs (Figures 5 and S1).

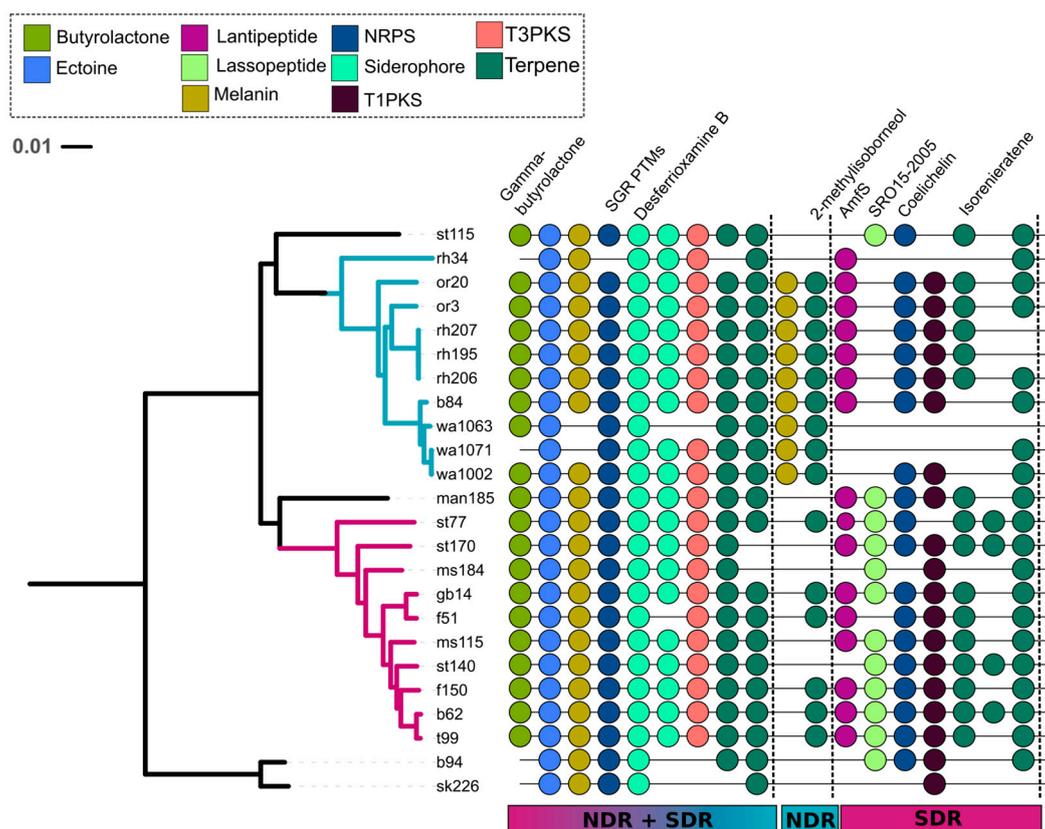


Figure 6. A total of nine core (i.e., conserved in $\geq 80\%$ of genomes) SMGCs were found in both NDR and SDR. The NDR clade had 11 core SMGCs and the SDR clade had 15 core SMGCs. The tree reflects phylogenetic relationships between *Streptomyces* sister-taxa genomes and was constructed from multiple whole genome alignments (see Figure 1). Scale bar represents nucleotide substitutions per site. Tree branches are colored according to clade affiliation. Core orthologous SMGCs (depicted by colored circles) were determined using the antiSMASH [32] MIBiG annotation pipeline or were defined using our annotation-independent approach (see Materials and Method). Colors correspond to SMGC class (see legend), and natural product annotations are labeled if available.

Differences in gene content between closely related microbes ultimately result from gene gain and loss events [49–51]. Although deletion bias is strong in bacterial genomes [52], gene acquisitions can drive rapid genome innovation and evolution [53]. Gene clusters are often acquired through horizontal gene transfer leading to the formation of new operons in bacterial genomes [54], and many SMGCs in actinomycetes are believed to be the result of horizontal gene transfer [16,18,33,55]. Parsimony predicts that low frequency and strain-specific genes are likely the result of a recent acquisition, while high frequency “near core” genes are the likely result of recent deletion events [56]. Hence, we are able to infer SMGC gain and loss dynamics in our *Streptomyces* sister-taxa from SMGC frequency distributions (Figure 4).

The majority of SMGCs observed within the sister-clades occurred in only one or a few strains, and this suggests that gene acquisition is a major force that drives the diversity of SMGC pathways in *Streptomyces*. However, each clade has a distinct set of core and accessory SMGCs (Figures 3, 5, 6 and S1), and this suggests that SMGC composition (Figure 7) may underlie ecological traits that promote or reinforce lineage divergence. For example, nearly all genomes within the NDR clade (with the exception of rh34) harbor a melanin gene cluster which is absent from both the intermediate (INT) and SDR genomes, suggesting that horizontal gene transfer of the melanin gene cluster into the immediate ancestor of NDR accompanied lineage divergence (Figures 6 and 7). Overall, NDR has more low

frequency SMGCs (present in one to three strains) than SDR (139 and 96, respectively) (Figure 4). This result suggests a greater rate of gene acquisition in NDR than in SDR and is consistent with the observation that NDR has more SMGCs (Figure 2) and larger genomes overall than SDR. While this difference in gene content is potentially adaptive, it could also be explained as a consequence of neutral demographic processes such as genome surfing (reviewed in [57]). However, the distribution of SMGC frequencies does not differ significantly between clades (Kolmogorov-Smirnov test, $p = 0.4$). Hence, while it seems clear that gene acquisition is a major driver of SMGC biodiversity, the role of gene acquisition in driving lineage divergence remains unclear.

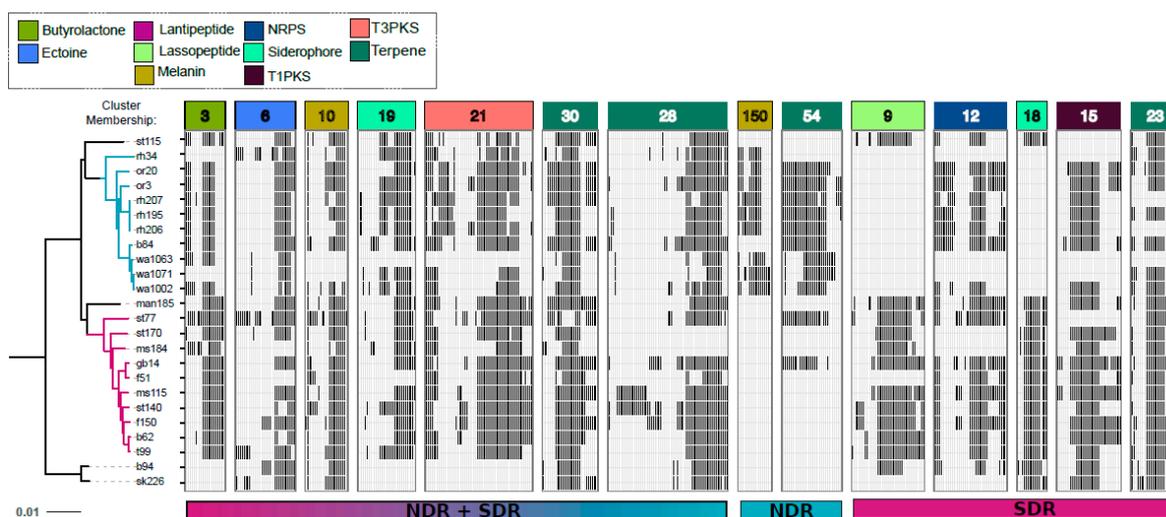


Figure 7. Gene content of core SMGCs vary within and between clades as a result of gene acquisition and deletion events. Panels depict the gene content (i.e., genetic architecture) of core SMGCs (i.e., conserved in $\geq 80\%$ of genomes), the NDR-specific SMGC core, and the SDR-specific SMGC core. Black bars within the panels represent orthologous genes. The tree reflects phylogenetic relationships between *Streptomyces* sister-taxa genomes and was constructed from multiple whole genome alignments (see Figure 1). Scale bar represents nucleotide substitutions per site. Panel colors correspond to SMGC class (see legend). Panels are labeled with the SMGC cluster membership (see Table S3) defined using our annotation-independent approach (see Materials and Methods).

We also see evidence that NDR has undergone the deletion of SMGC-associated genes inherited from the common ancestor of NDR and SDR. For example, the SRO15-2005 lassopeptide gene cluster is conserved in SDR and found in INT but absent from NDR, suggesting that deletion of this lassopeptide accompanied NDR divergence (Figures 6 and 7). We also find that core SMGC gene loss is more common in NDR than SDR (strain-level deletions occur in six out of nine core gene clusters within NDR and two out of nine core gene clusters in SDR) (Figure 6). Similarly, we can observe SDR species-specific core gene clusters (AmfS, coelichelin, a T1PKS, and a terpene) that are found in only 70% (i.e., near core) of NDR strains (Figure 6). This pattern suggests that these SMGCs were present in the common ancestor of the two clades and subsequently deleted from NDR isolates. In addition, the butyrolactone operon (cluster 3) is comprised of more genes in SDR than in NDR, and this likely indicates active gene loss within this pathway for NDR strains (Figure 7).

Taken together, these results suggest that the sister-clades are under different evolutionary pressures which drive dissimilarity in SMGC composition. NDR genomes have increased in size relative to their ancestors suggesting an overall increase in the rate of gene acquisition via horizontal gene exchange, and this increase in gene acquisition has resulted in an increase in strain-specific SMGC content in NDR. In addition, the presence of NDR-specific core SMGCs (e.g., melanin gene cluster) indicates that some horizontally acquired SMGC have gone to fixation within NDR. At the same time, deletion events in NDR have pruned away SMGCs inherited from ancestral lineages (i.e., those clusters

present in SDR and INT). We hypothesize that these changes in SMGC content are likely to have effects on fitness which should act to reinforce lineage divergence either as a result of antagonism or niche differentiation.

3. Materials and Methods

3.1. *Streptomyces* Isolation and DNA Extraction

We built a culture collection of >1000 *Streptomyces* isolated from grassland soils (pH 3.9–7.3) sampled at 0–5 cm from sites across the United States [27]. Pure *Streptomyces* cultures were obtained from air-dried soils on glycerol-arginine agar (pH 8.7) containing antifungals as previously described [58]. Genomic DNA was extracted using a standard phenol/chloroform/isoamyl alcohol protocol from liquid cultures grown in yeast extract-malt extract medium (YEME) with 0.5% glycine [5] for 72 h shaking at 30 °C.

3.2. Whole Genome Sequencing, Assembly, and Annotation

Streptomyces genomic sequencing libraries were prepped with the Nextera DNA Library Preparation Kit (Illumina, San Diego, CA, USA), and draft genomes were generated using the Illumina HiSeq2500 platform (Illumina, San Diego, CA, USA) and paired-end 2×100 bp reads at the Cornell University Biotechnology Resource Center (BRC). Quality control and assembly was performed with the A5 pipeline [59], and genomes were annotated using the online RAST Server [60]. Multiple whole genome alignments were obtained with Mugsy [61], and trimAL v1.2 removed poorly aligned regions [62]. Orthologous genes were identified using ITEP [63] with MCL clustering parameters as follows: inflation value = 2.0, cutoff = 0.04, maxbit score. Average nucleotide identity (ANI) was determined using mother [64]. Genome sequences are available at NCBI under BioProject ID PRJNA401484 accession numbers SAMN07606143–SAMN07606166.

3.3. Phylogenetic Reconstruction

The phylogenetic relationship between genomes was reconstructed from DNA sequences of multiple whole genome alignments using maximum likelihood (ML) with the generalized time reversible nucleotide substitution model [65] with gamma distributed rate heterogeneity among sites (GTRGAMMA) in RAxML v7.3.0 [66]. Bootstrap support was determined using the RAxML rapid bootstrapping algorithm [67].

3.4. Secondary Metabolite Biosynthetic Gene Cluster (SMGC) Identification

Secondary metabolite biosynthetic gene clusters (SMGC) were predicted and annotated using the online server antiSMASH 3.0 [32]. We also used an annotation-independent approach to identify SMGCs shared between genomes. For each SMGC identified by antiSMASH, we used Prodigal [68] to call open reading frames (ORFs) and Parasail with default parameters to identify orthologous genes and orthologous gene groups [69]. We used the R package igraph [70] to cluster similar SMGCs, define cluster membership, and thus determine which SMGCs are shared between genomes. Cluster membership was determined based on gene content using a binary (i.e., Jaccard) dissimilarity distance of ≤ 4.0 generated from an orthologous group presence/absence table. Dissimilarity distances of >4.0 did not result in an appreciable gain in the number of total clusters. The SMGC network was visualized and analyzed with Cytoscape 3.3.0 [71].

4. Conclusions

We used comparative genomics to examine SMGC diversity within strains of two closely related *Streptomyces* species that recently diverged from a common ancestor. Our objective was to observe and explore the evolutionary dynamics of SMGCs that accompany evolutionary diversification and to assess SMGC conservation within and between closely related species. It is clear that gene gain and loss events drive major differences in SMGC composition, both within and between species. While both

species share conserved core SMGCs, each clade has its own species-specific SMGC core, and the majority of SMGCs were strain-specific. This pattern indicates that these SMGCs, not present in shared ancestors, were acquired recently due to horizontal gene exchange.

In addition, we observe that SMGCs that have been inherited from a shared ancestor can vary considerably in gene content, both due to the acquisition and deletion of individual genes within each gene cluster. We observe SMGC gain and loss dynamics that differ between clades and identify SMGC acquisition and deletion events that correspond to ancestral diversification events. These findings show that SMGC modification is associated with lineage divergence, though whether these changes cause or reinforce divergence directly or are an indirect product of evolutionary divergence remains to be seen. A limitation of the comparative genomics approach is that we cannot assess the ecological activity of a pathway from genome sequence data. It is possible that some (or all) of the strain-specific pathways, if acquired by recent horizontal exchange, may be non-functional. It is also possible that changes in SMGC architecture and gene content could alter pathway functionality and that pathways deemed orthologous on the basis of genetic similarity may have different functions in different strains.

Finally, we can conclude that, while strains within a species will share a core set of SMGCs, the number of accessory SMGC within a given species can be quite large, with each strain having its own repertoire of strain-specific SMGCs. Furthermore, the majority of these strain-specific SMGCs remain uncharacterized and lack similarity to SMGCs documented in public databases.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2079-6382/7/1/12/s1>, Figure S1: Each clade has a distinct SMGC network. The network illustrates inter- and intra-clade sharing of SMGC content. Large circles represent the genomes of *Streptomyces* strains and are labeled with isolate names and colored according to clade affiliation. Smaller circles represent non-redundant distinct SMGCs identified using our annotation-independent approach (see Materials and Methods). Lines connect each SMGC to the strains in which they are found. Network nodes and edges are scaled in proportion to the number of connections and colored according to gene cluster class (see legend). Network is arranged in the organic layout using Cytoscape 3.3.0 [71]. Core SMGCs can be observed as larger central nodes while strain specific and low frequency SMGCs occur around the edges of the graph; Table S1: The 24 *Streptomyces* genomes were isolated from 11 sites. Isolate names begin with the site code from which they were isolated from followed by strain number; Table S2: Genome and assembly characteristics for 24 *Streptomyces* genomes. The clade affiliations include the northern-derived (NDR), southern-derive (SDR), and intermediate (INT). Sample site of each isolate can be found in Table S1. Values report assembled draft genome size, genome-wide G+C content, the number of predicted open reading frames (ORFs), and the number of predicted secondary metabolite biosynthetic gene clusters (SMGCs) per genome; Table S3: SMGCs are predicted by antiSMASH [32] in our 24 *Streptomyces* genomes. For each SMGC, columns report the affiliated genome, clade, gene cluster class (hybrids are indicated by hyphens), gene cluster length (bp), natural product annotation provided by antiSMASH, cluster membership (Clust Memb), MIBiG database identification, the portion of genes with similarity to genes within the most similar known cluster from the MIBiG database (% Genes w/Similarity). Cluster membership was determined using our annotation-independent approach (see Materials and Methods). NA indicates information is not available.

Acknowledgments: This material is based upon work supported by the National Science Foundation under Grant No. DEB-1456821 awarded to Daniel H. Buckley.

Author Contributions: Mallory J. Choudoir and Daniel H. Buckley conceived and designed the study; Mallory J. Choudoir performed the analyses and analyzed the data; Charles Pepe-Ranney contributed to the analyses; Mallory J. Choudoir and Daniel H. Buckley wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Kossel, A. Ueber die chemische zusammensetzung der zelle. *Arch. Physiol.* **1891**, *4*, 181–186.
2. Doroghazi, J.R.; Metcalf, W.W. Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genom.* **2013**, *14*, 611. [[CrossRef](#)] [[PubMed](#)]
3. Karlovsky, P. *Secondary Metabolites in Soil Ecology*; Springer: New York, NY, USA, 2008; pp. 1–19.
4. Davies, J. Where have all the antibiotics gone? *Can. J. Infect. Dis. Med. Microbiol.* **2006**, *17*, 287–290. [[CrossRef](#)] [[PubMed](#)]

5. Kieser, T.; Bibb, M.J.; Buttner, M.J.; Charter, K.F.; Hopwood, D.A. *Practical Streptomyces Genetics*; John Innes Foundation: Norwich, UK, 2000.
6. Bérdy, J. Bioactive microbial metabolites. *J. Antibiot.* **2005**, *58*, 1–26. [[CrossRef](#)] [[PubMed](#)]
7. Watve, M.; Tickoo, R.; Jog, M.; Bhole, B. How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* **2001**, *176*, 386–390. [[CrossRef](#)] [[PubMed](#)]
8. O'Brien, J.; Wright, G.D. An ecological perspective of microbial secondary metabolism. *Curr. Opin. Biotechnol.* **2011**, *22*, 552–558. [[CrossRef](#)] [[PubMed](#)]
9. Vaz Jauri, P.; Bakker, M.G.; Salomon, C.E.; Kinkel, L.L. Subinhibitory antibiotic concentrations mediate nutrient use and competition among soil *Streptomyces*. *PLoS ONE* **2013**, *8*, e81064. [[CrossRef](#)] [[PubMed](#)]
10. Traxler, M.F.; Kolter, R. Natural products in soil microbe interactions and evolution. *Nat. Prod. Rep.* **2015**, *32*, 956–970. [[CrossRef](#)] [[PubMed](#)]
11. Hopwood, D.A. Soil to genomics: The *Streptomyces* chromosome. *Annu. Rev. Genet.* **2006**, *40*, 1–23. [[CrossRef](#)] [[PubMed](#)]
12. Kämpfer, P. The family Streptomycetaceae Part I: Taxonomy. In *The Prokaryotes*; Springer: New York, NY, USA, 2006; pp. 538–604.
13. Van der Meij, A.; Worsley, S.F.; Hutchings, M.I.; van Wezel, G.P. Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol. Rev.* **2017**, *41*, 392–416. [[CrossRef](#)] [[PubMed](#)]
14. Bentley, S.D.; Chater, K.F.; Cerdeño-Tarraga, A.M.; Challis, G.L.; Thomson, N.R.; James, K.D.; Harris, D.E.; Quail, M.A.; Kieser, H.; Harper, D.; et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **2002**, *417*, 141–147. [[CrossRef](#)] [[PubMed](#)]
15. Blin, K.; Medema, M.H.; Kazempour, D.; Fischbach, M.A.; Breitling, R.; Takano, E.; Weber, T. AntiSMASH 2.0—A versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* **2013**, *41*, W204–W212. [[CrossRef](#)] [[PubMed](#)]
16. Jensen, P.R.; Williams, P.G.; Oh, D.-C.; Zeigler, L.; Fenical, W. Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. *Appl. Environ. Microbiol.* **2007**, *73*, 1146–1152. [[CrossRef](#)] [[PubMed](#)]
17. Doroghazi, J.R.; Albright, J.C.; Goering, A.W.; Ju, K.S.; Haines, R.R.; Tchalukov, K.A.; Labeda, D.P.; Kelleher, N.L.; Metcalf, W.W. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **2014**, *10*, 963–968. [[CrossRef](#)] [[PubMed](#)]
18. Ziemert, N.; Lechner, A.; Wietz, M.; Millán-Aguiñaga, N.; Chavarria, K.L.; Jensen, P.R. Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E1130–E1139. [[CrossRef](#)] [[PubMed](#)]
19. Wawrik, B.; Kerkhof, L.; Zylstra, G.J.; Kukor, J.J. Identification of unique type II polyketide synthase genes in soil. *Appl. Environ. Microbiol.* **2005**, *71*, 2232–2238. [[CrossRef](#)] [[PubMed](#)]
20. Charlop-Powers, Z.; Owen, J.G.; Reddy, B.V.B.; Ternei, M.A.; Brady, S.F. Chemical-biogeographic survey of secondary metabolism in soil. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3757–3762. [[CrossRef](#)] [[PubMed](#)]
21. Davelos, A.L.; Kinkel, L.L.; Samac, D.A. Spatial variation in frequency and intensity of antibiotic interactions among *Streptomyces* from prairie soil. *Appl. Environ. Microbiol.* **2004**, *70*, 1051–1058. [[CrossRef](#)] [[PubMed](#)]
22. Davelos Baines, A.L.; Xiao, K.; Kinkel, L.L. Lack of correspondence between genetic and phenotypic groups amongst soil-borne *Streptomyces*. *FEMS Microbiol. Ecol.* **2007**, *59*, 564–575. [[CrossRef](#)] [[PubMed](#)]
23. Charlop-Powers, Z.; Owen, J.G.; Reddy, B.V.B.; Ternei, M.A.; Guimarães, D.O.; de Frias, U.A.; Pupo, M.T.; Seepe, P.; Feng, Z.; Brady, S.F. Global biogeographic sampling of bacterial secondary metabolism. *Elife* **2015**, *4*, e05048. [[CrossRef](#)] [[PubMed](#)]
24. Wawrik, B.; Kutliev, D.; Abdivasievna, U.A.; Kukor, J.J.; Zylstra, G.J.; Kerkhof, L. Biogeography of actinomycete communities and type II polyketide synthase genes in soils collected in New Jersey and Central Asia. *Appl. Environ. Microbiol.* **2007**, *73*, 2982–2989. [[CrossRef](#)] [[PubMed](#)]
25. Reddy, B.V.B.; Kallifidas, D.; Kim, J.H.; Charlop-Powers, Z.; Feng, Z.; Brady, S.F. Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. *Appl. Environ. Microbiol.* **2012**, *78*, 3744–3752. [[CrossRef](#)] [[PubMed](#)]
26. Choudoir, M.J.; Campbell, A.N.; Buckley, D.H. Grappling with Proteus: Population level approaches to understanding microbial diversity. *Front. Microbiol.* **2012**, *3*. [[CrossRef](#)] [[PubMed](#)]

27. Andam, C.P.; Doroghazi, J.R.; Campbell, A.N.; Kelly, P.J.; Choudoir, M.J.; Buckley, D.H. A Latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*. *MBio* **2016**, *7*. [[CrossRef](#)] [[PubMed](#)]
28. Choudoir, M.J.; Doroghazi, J.R.; Buckley, D.H. Latitude delineates patterns of biogeography in terrestrial *Streptomyces*. *Environ. Microbiol.* **2016**, *18*, 4931–4945. [[CrossRef](#)] [[PubMed](#)]
29. Choudoir, M.J.; Buckley, D.H. Phylogenetic conservatism of thermal traits explains dispersal limitation and genomic differentiation of *Streptomyces* sister-taxa. *ISME J.* **2018**, under review.
30. Rong, X.; Huang, Y. Taxonomic evaluation of the *Streptomyces griseus* clade using multilocus sequence analysis and DNA-DNA hybridization, with proposal to combine 29 species and three subspecies as 11 genomic species. *Int. J. Syst. Evol. Microbiol.* **2010**, *60*, 696–703. [[CrossRef](#)] [[PubMed](#)]
31. Konstantinidis, K.T.; Ramette, A.; Tiedje, J.M. The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2006**, *361*, 1929–1940. [[CrossRef](#)] [[PubMed](#)]
32. Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H.U.; Bruccoleri, R.; Lee, S.Y.; Fischbach, M.A.; Müller, R.; Wohlleben, W.; et al. AntiSMASH 3.0—A comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **2015**, *43*, W237–W243. [[CrossRef](#)] [[PubMed](#)]
33. Omura, S.; Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Takahashi, C.; Shinose, M.; Takahashi, Y.; Horikawa, H.; Nakazawa, H.; Osonoe, T.; et al. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: Deducing the ability of producing secondary metabolites. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 12215–12220. [[CrossRef](#)] [[PubMed](#)]
34. Ohnishi, Y.; Ishikawa, J.; Hara, H.; Suzuki, H.; Ikenoya, M.; Ikeda, H.; Yamashita, A.; Hattori, M.; Horinouchi, S. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J. Bacteriol.* **2008**, *190*, 4050–4060. [[CrossRef](#)] [[PubMed](#)]
35. Aigle, B.; Lautru, S.; Spitteller, D.; Dickschat, J.S.; Challis, G.L.; Leblond, P.; Pernodet, J.L. Genome mining of *Streptomyces ambofaciens*. *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 251–263. [[CrossRef](#)] [[PubMed](#)]
36. Seipke, R.F. Strain-level diversity of secondary metabolism in *Streptomyces albus*. *PLoS ONE* **2015**, *10*, e0116457. [[CrossRef](#)] [[PubMed](#)]
37. Du, L.; Sánchez, C.; Shen, B. Hybrid peptide–polyketide natural products: Biosynthesis and prospects toward engineering novel molecules. *Metab. Eng.* **2001**, *3*, 78–95. [[CrossRef](#)] [[PubMed](#)]
38. Medema, M.H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M.A.; Weber, T.; Takano, E.; Breitling, R. AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **2011**, *39*, W339–W346. [[CrossRef](#)] [[PubMed](#)]
39. Olano, C.; García, I.; González, A.; Rodríguez, M.; Rozas, D.; Rubio, J.; Sánchez-Hidalgo, M.; Braña, A.F.; Méndez, C.; Salas, J.A. Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microb. Biotechnol.* **2014**, *7*, 242–256. [[CrossRef](#)] [[PubMed](#)]
40. Claesen, J.; Bibb, M.J. Biosynthesis and regulation of grisemycin, a new member of the linaridin family of ribosomally synthesized peptides produced by *Streptomyces griseus* IFO 13350. *J. Bacteriol.* **2011**, *193*, 2510–2516. [[CrossRef](#)] [[PubMed](#)]
41. Zetterström, R. Selman, A. Waksman (1888–1973) Nobel Prize in 1952 for the discovery of streptomycin, the first antibiotic effective against tuberculosis. *Acta Paediatr.* **2007**, *96*, 317–319. [[CrossRef](#)] [[PubMed](#)]
42. Waksman, S.A.; Reilly, H.C.; Johnstone, D.B. Isolation of streptomycin-producing strains of *Streptomyces griseus*. *J. Bacteriol.* **1946**, *52*, 393–397. [[PubMed](#)]
43. Hotta, K.; Ishikawa, J. Strain- and species-specific distribution of the streptomycin gene cluster and kan-related sequences in *Streptomyces griseus*. *J. Antibiot.* **1988**, *41*, 1116–1123. [[CrossRef](#)] [[PubMed](#)]
44. Tettelin, H.; Riley, D.; Cattuto, C.; Medini, D. Comparative genomics: The bacterial pan-genome. *Curr. Opin. Microbiol.* **2008**, *11*, 472–477. [[CrossRef](#)] [[PubMed](#)]
45. Lefebure, T.; Bitar, P.D.P.; Suzuki, H.; Stanhope, M.J. Evolutionary dynamics of complete *Campylobacter* pan-genomes and the bacterial species concept. *Genome Biol. Evol.* **2010**, *2*, 646–655. [[CrossRef](#)] [[PubMed](#)]
46. Imbert, M.; Béchet, M.; Blondeau, R. Comparison of the main siderophores produced by some species of *Streptomyces*. *Curr. Microbiol.* **1995**, *31*, 129–133. [[CrossRef](#)]
47. Roberts, A.A.; Schultz, A.W.; Kersten, R.D.; Dorrestein, P.C.; Moore, B.S. Iron acquisition in the marine actinomycete genus *Salinispora* is controlled by the desferrioxamine family of siderophores. *FEMS Microbiol. Lett.* **2012**, *335*, 95–103. [[CrossRef](#)] [[PubMed](#)]

48. Antony-Babu, S.; Stien, D.; Eparvier, V.; Parrot, D.; Tomasi, S.; Suzuki, M.T. Multiple *Streptomyces* species with distinct secondary metabolomes have identical 16S rRNA gene sequences. *Sci. Rep.* **2017**, *7*, 11089. [[CrossRef](#)] [[PubMed](#)]
49. Gogarten, J.P.; Townsend, J.P. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* **2005**, *3*, 679–687. [[CrossRef](#)] [[PubMed](#)]
50. Marri, P.R.; Hao, W.; Golding, G.B. Gene gain and gene loss in *Streptococcus*: Is it driven by habitat? *Mol. Biol. Evol.* **2006**, *23*, 2379–2391. [[CrossRef](#)] [[PubMed](#)]
51. Reno, M.L.; Held, N.L.; Fields, C.J.; Burke, P.V.; Whitaker, R.J. Biogeography of the *Sulfolobus islandicus* pan-genome. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 8605–8610. [[CrossRef](#)] [[PubMed](#)]
52. Mira, A.; Ochman, H.; Moran, N.A. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **2001**, *17*, 589–596. [[CrossRef](#)]
53. Ochman, H.; Lawrence, J.G.; Groisman, E.A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **2000**, *405*, 299–304. [[CrossRef](#)] [[PubMed](#)]
54. Homma, K.; Fukuchi, S.; Nakamura, Y.; Gojobori, T.; Nishikawa, K. Gene cluster analysis method identifies horizontally transferred genes with high reliability and indicates that they provide the main mechanism of operon gain in 8 species of γ -Proteobacteria. *Mol. Biol. Evol.* **2006**, *24*, 805–813. [[CrossRef](#)] [[PubMed](#)]
55. Penn, K.; Jenkins, C.; Nett, M.; Udworthy, D.W.; Gontang, E.A.; McGlinchey, R.P.; Foster, B.; Lapidus, A.; Podell, S.; Allen, E.E.; et al. Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J.* **2009**, *3*, 1193–1203. [[CrossRef](#)] [[PubMed](#)]
56. Bolotin, E.; Hershberg, R. Gene loss dominates as a source of genetic variation within clonal pathogenic bacterial species. *Genome Biol. Evol.* **2015**, *7*, 2173–2187. [[CrossRef](#)] [[PubMed](#)]
57. Choudoir, M.J.; Panke-Buisse, K.; Andam, C.P.; Buckley, D.H. Genome surfing as driver of microbial genomic diversity. *Trends Microbiol.* **2017**, *25*, 624–636. [[CrossRef](#)] [[PubMed](#)]
58. Doroghazi, J.R.; Buckley, D.H. Widespread homologous recombination within and between *Streptomyces* species. *ISME J.* **2010**, *4*, 1136–1143. [[CrossRef](#)] [[PubMed](#)]
59. Tritt, A.; Eisen, J.A.; Facciotti, M.T.; Darling, A.E. An integrated pipeline for de novo assembly of microbial genomes. *PLoS ONE* **2012**, *7*, e42304. [[CrossRef](#)] [[PubMed](#)]
60. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST server: Rapid annotations using subsystems technology. *BMC Genom.* **2008**, *9*, 75. [[CrossRef](#)] [[PubMed](#)]
61. Angiuoli, S.V.; Salzberg, S.L. Mugsy: Fast multiple alignment of closely related whole genomes. *Bioinformatics* **2011**, *27*, 334–342. [[CrossRef](#)] [[PubMed](#)]
62. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. TrimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [[CrossRef](#)] [[PubMed](#)]
63. Benedict, M.N.; Henriksen, J.R.; Metcalf, W.W.; Whitaker, R.J.; Price, N.D. ITEP: An integrated toolkit for exploration of microbial pan-genomes. *BMC Genom.* **2014**, *15*, 8. [[CrossRef](#)] [[PubMed](#)]
64. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [[CrossRef](#)] [[PubMed](#)]
65. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. In *Lectures on Mathematics in the Life Sciences*; American Mathematical Society: Providence, RI, USA, 1986; Volume 17, pp. 57–86.
66. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [[CrossRef](#)] [[PubMed](#)]
67. Stamatakis, A.; Hoover, P.; Rougemont, J.; Renner, S. A rapid bootstrap algorithm for the RAxML Web Servers. *Syst. Biol.* **2008**, *57*, 758–771. [[CrossRef](#)] [[PubMed](#)]
68. Hyatt, D.; LoCascio, P.F.; Hauser, L.J.; Uberbacher, E.C. Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **2012**, *28*, 2223–2230. [[CrossRef](#)] [[PubMed](#)]
69. Daily, J. Parasail: SIMD C library for global, semi-global, and local pairwise sequence alignments. *BMC Bioinform.* **2016**, *17*, 81. [[CrossRef](#)] [[PubMed](#)]

70. Csardi, G.; Nepusz, T. The Igraph Software Package for Complex Network Research. 2006. Available online: http://www.interjournal.org/manuscript_abstract.php?361100992 (accessed on 6 February 2018).
71. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).