AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

Habitat Adaptation Drives Speciation of a *Streptomyces* Species with Distinct Habitats and Disparate Geographic Origins

Jiao Wang,[a,c] Yisong Li,[a,c] Adrián A. Pinto-Tomás,[b] Kun Cheng,[a*] Ying Huang[a]

aState Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China
bCenter for Research in Cell and Molecular Biology, Center for Research in Microscopic Structures and Biochemistry Department, School of Medicine, University of Costa Rica, San José, Costa Rica
cCollege of Life Sciences, University of Chinese Academy of Sciences, Beijing, China

ABSTRACT Microbial diversification is driven by geographic and ecological factors, but how the relative importance of these factors varies among species, geographic scales, and habitats remains unclear. *Streptomyces*, a genus of antibiotic-producing, spore-forming, and widespread bacteria, offers a robust model for identifying the processes underlying population differentiation. We examined the population structure of 37 *Streptomyces olivaceus* strains isolated from various sources, showing that they diverged into two habitat-associated (free-living and insect-associated) and geographically disparate lineages. More frequent gene flow within than between the lineages confirmed genetic isolation in *S. olivaceus*. Geographic isolation could not explain the genetic isolation; instead, habitat type was a strong predictor of genetic distance when controlling for geographic distance. The identification of habitat-specific genetic variations, including genes involved in regulation, resource use, and secondary metabolism, suggested a significant role of habitat adaptation in the diversification process. Physiological assays revealed fitness trade-offs under different environmental conditions in the two lineages. Notably, insect-associated isolates could outcompete free-living isolates in a free-iron-deficient environment. Furthermore, substrate (e.g., sialic acid and glycogen) utilization but not thermal traits differentiated the two lineages. Overall, our results argue that adaptive processes drove ecological divergence among closely related streptomycetes, eventually leading to dispersal limitation and gene flow barriers between the lineages. *S. olivaceus* may best be considered a species complex consisting of two cryptic species.

IMPORTANCE Both isolation by distance and isolation by environment occur in bacteria, and different diversification patterns may apply to different species. *Streptomyces* species, typified by producing useful natural products, are widespread in nature and possess high genetic diversity. However, the ecological processes and evolutionary mechanisms that shape their distribution are not well understood. Here, we show that the population structure of a ubiquitous *Streptomyces* species complex matches its habitat distribution and can be defined by gene flow discontinuities. Using comparative genomics and physiological assays, we reveal that gains and losses of specific genomic traits play a significant role in the transition between free-living and host-associated lifestyles, driving speciation of the species. These results provide new insights into the evolutionary trajectory of *Streptomyces* and the notion of species.

KEYWORDS *Streptomyces*, gene flow, habitat adaptation, lifestyle, population genetics, speciation

Delineating microbial populations is vitally important in understanding microbial evolution and diversity (1). However, the concept of microbial species and the processes contributing to microbial speciation remain controversial (2, 3). In the last

decade, several operational approaches have been proposed to address this issue (4–6). These advances, combined with the rapid proliferation of genomic data, have greatly facilitated research on the processes that shape population genetic patterns.

Traditionally, microorganisms are considered cosmopolitan due to their large population sizes and high dispersal potential (7), but this view has recently been challenged (8). As in larger organisms, isolation by distance (IBD) has been reported in bacteria and archaea (9–11), which means that the genetic differentiation of strains increases with geographic distance (12). This pattern could be generated by drift or past selection and subsequently be maintained by spatial dispersal limitation (10, 13). Straightforwardly, due to restricted dissemination and thus reduced encounters, gene flow barriers could arise from geographic isolation, eventually resulting in allopatric speciation (14). Alternatively, barriers to gene flow could emerge via ecological differentiation independent of geographic distance (15, 16), which is common among microbial populations and is also known as isolation by environment (IBE) (17). Several studies have revealed that closely related strains coexisting at a small spatial scale can be clustered into ecologically and genetically divergent groups where recombination barriers exist without physical isolation (18, 19). Furthermore, given that environmental differences may be correlated with geographic distance, the effect of geographic isolation is probably overestimated (10). In summary, microbial speciation is driven by multiple interacting processes, yet it is unclear how the relative importance of each process varies among spatial scales and habitat types.

In addition, the effects of these processes might depend on the physiological traits of microbes (10). For example, bacteria of the genera *Bacillus* and *Streptomyces* can produce stress-resistant spores, which enhance their dispersal and survival. Therefore, these spore-forming microorganisms are globally distributed with little dispersal limitation among ecologically similar habitats (20–23). Their distribution patterns are governed by either environmental selection (20, 24) or historical demographic processes driven by paleoclimate dynamics (21–23). Nevertheless, previous studies analyzed microorganisms either at a relatively low genetic resolution or with a quite high level of genetic differentiation. A more interesting issue is that most published works on microbial speciation focused on populations either at local scales with distinct habitats (6, 25) or across large spatial scales with the same ecological niche (11, 23). However, many microbial species are widely distributed in terms of both habitat and geography. Further surveys concentrating on highly related strains collected from various habitats and geographic regions are necessary to explore how environmental selection and geographic isolation contribute to population differentiation.

Streptomycetes, well known for their prolific secondary metabolism, are filamentous bacteria with complex lifestyles (26, 27). Due to sporulation, *Streptomyces* strains are distributed worldwide and are often found in various ecological settings (28, 29). These traits make streptomycetes robust models for examining how geography and ecology influence population structure and evolution. In addition to their well-known edaphic and aquatic habitats, host-associated systems, notably insect-associated habitats, are also major habitats for *Streptomyces* (30). Stimulated by the discovery of new natural products, many previous studies investigated the defensive mutualism between *Streptomyces* and insects (31, 32). *Streptomyces* can evolve to yield novel antibiotics in the niches of insect-*Streptomyces* symbioses (33, 34). Another study indicated that insect hosts feeding on plant biomass likely select symbiotic *Streptomyces* for increased cellulose-degrading activity (35). These findings suggest that interactions with insects might facilitate the diversification of *Streptomyces*. However, the evolutionary mechanisms underlying the transition between free-living and insect-associated lifestyles, especially how symbiotic interactions drive the population differentiation and speciation of streptomycetes, remain poorly understood.

We have been collecting conspecific *Streptomyces* strains all over the world for several years. Our previous studies of a collection of *Streptomyces albidoflavus* strains (with a genome-wide average nucleotide identity [ANI] of >98.15%) from diverse habitats

have indicated that ecological adaptation plays an important role in shaping the population structure of this species (36, 37). Here, we studied another collection of 37 more divergent strains (with a genome-wide ANI of >96.93%), residing in both free-living (FL) and insect-associated (IA) habitats encompassing multiple spatial scales. This collection includes 35 conspecific isolates previously assigned to *Streptomyces olivaceus* (including the type strain *S. olivaceus* CGMCC 4.1369) based on multilocus sequence analysis and phenotypic characteristics (38) and two *S. olivaceus* isolates with ANI values of 98.59% and 98.60% to the type strain. We investigated the population genetic structure and gene flow pattern and whether the differentiation within *S. olivaceus* is driven by geographic isolation, ecological isolation, or perhaps both. We hypothesized that ecological factors play a dominant role in driving the genetic divergence of *S. olivaceus* and that ecological dispersal limitation exists between FL and IA habitat types. Furthermore, using comparative genomics and physiological characterization, we identified genes that may account for habitat adaptation and links between genomic and ecological variations.

RESULTS

S. olivaceus strains segregate into two habitat-associated and geographically disparate lineages. We collected 37 *S. olivaceus* strains that were isolated from diverse habitats and geographic locations (Table S1), comprising 29 strains from various soil and marine sources considered free-living (FL) habitats and 8 strains isolated from insect-associated (IA) habitats. The general genomic features of all strains are shown in Table S1. The genome size ranged from 8.11 to 8.66 Mb, and the number of coding sequences (CDSs) per genome was 6,966 to 7,512. Gene orthology analysis showed that the pan-genome consisted of 12,949 genes, including 2,973 strain-specific genes (Fig. 1). The core genome contained 5,591 genes (5,007 single-copy genes) and was predicted to encode 27 core secondary metabolite biosynthetic pathways (Table S2). The mean pairwise genome-wide ANI among the strains was 98.63%, with the lowest ANI of 96.93% observed between strains CR8 and KLBMP 5084 (Fig. S1A). The type strain *S. olivaceus* CGMCC 4.1369 shared an ANI of >97.16% with the other 36 strains. These ANI values were all above the typical prokaryotic species cutoff of 95 to 96% (4, 5). Furthermore, these strains formed a stable single-monophyletic cluster in the phylogenomic tree of the genus *Streptomyces* (Fig. S1B and C), sharing an ANI of <88.4% with other streptomycetes. These results support that the 37 strains belong to a cohesive unit, descending from a last common ancestor.

Despite the high genetic relatedness of *S. olivaceus* strains, two robust clades were clearly identified by both hierarchical clustering analysis based on the pan-genome profile (Fig. 1) and phylogenetic analysis based on concatenated single-copy core genes (containing 4,948,371 nucleotides [nt]) (Fig. 2A). The two clades were clustered strictly according to habitat types (FL and IA), which overlapped with two roughly divided geographic regions, Costa Rica in the Western Hemisphere and China and the southern/southwest Indian Ocean in the Eastern Hemisphere (Fig. 2B). In particular, one clade consisted of 8 IA strains (clade IA) isolated from Costa Rica, sharing an intraclade ANI of 97.45 to 99.99%, and the other clade included the remaining 29 FL strains (clade FL) isolated from China and the southern/southwest Indian Ocean, sharing an intraclade ANI of 98.23 to 99.998%. In addition, fastSTRUCTURE (39) analysis inferred two ancestral populations (best K = 2, where K denotes the number of populations) consistent with the two clades (Fig. 2C). All these data indicate that *S. olivaceus* consists of two closely related but distinct sister lineages, an FL lineage and an IA lineage.

Gene flow barriers between lineages. To reveal the processes generating the differentiation between the FL and IA lineages, we investigated core-genome gene flow among *S. olivaceus* strains. First, we employed ClonalFrameML v1.12 (40) to quantify the degree of homologous recombination (Table 1). In the species and FL clade, recombination contributed only slightly more to the diversity than mutation (relative impact of recombination versus mutation on the per-site substitution rate [r/m] = 1.32 and 1.48, respectively). In contrast, we observed a strong recombination effect in clade
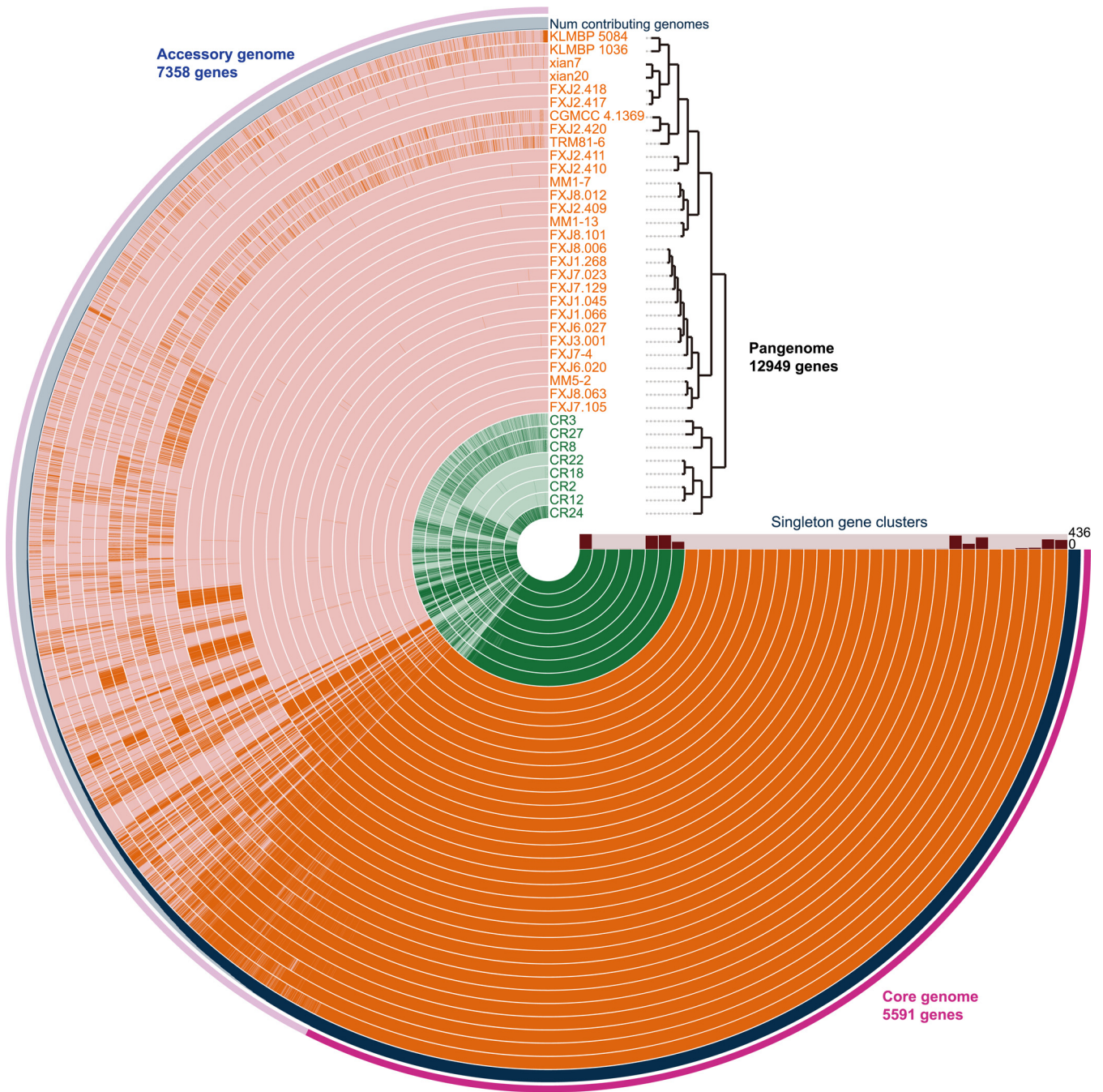
FIG 1 Pan-genome analysis of *S. olivaceus*. The tree on the top right displays the hierarchical clustering of 37 genomes based on the presence/absence patterns of 12,949 pan-genomic genes. The inner layers represent individual genomes with habitat sources colored green (insect-associated, IA) and orange (free-living, FL). In the layers, bars indicate the presence of genes in a given genome. In the outermost layer, the core (5,591 genes) and accessory (7,358 genes) genomes are indicated in dark and light pink, respectively. The second blue layer shows the number of genomes harboring each gene.

IA (r/m = 3.95). Given that the impact of recombination might be underestimated among highly similar strains (41, 42), we repeated the analysis without the very closely related strains. As a result, the effect of recombination on diversity increased markedly in clade FL (r/m = 2.46) but remained relatively low in the species overall (r/m = 1.58), indicating higher recombination levels within than between the clades (Table 1). Furthermore, 1,252 (65.3%) and 853 (86.6%) recombination events in clades FL and IA, respectively, likely had within-clade and external (outside the clades) origins, representing the main sources of gene exchange in *S. olivaceus* (Fig. S2). Only 3.0% and
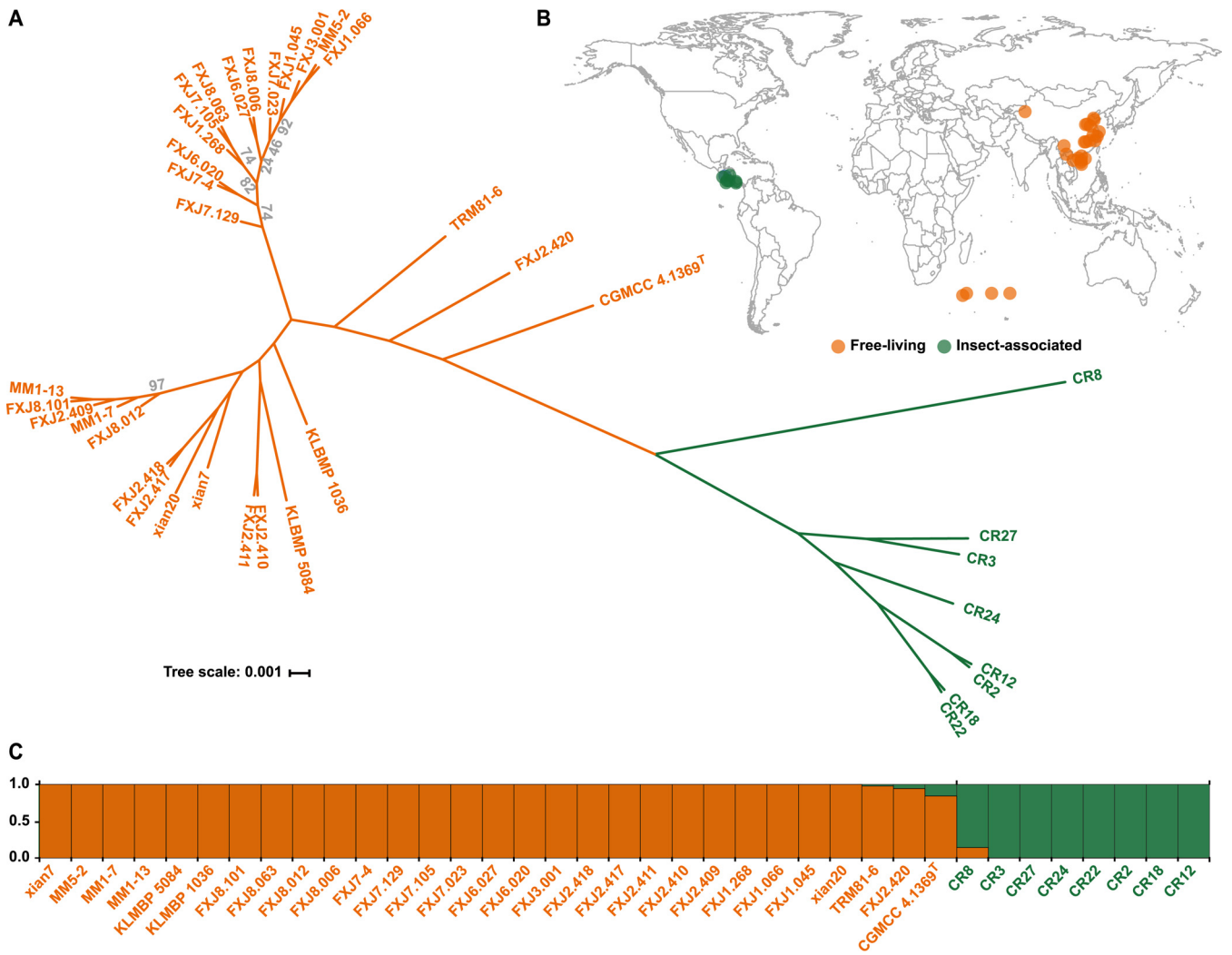
FIG 2 (A) Maximum likelihood phylogenetic tree generated from concatenated single-copy core genes of 37 *S. olivaceus* strains. The scale bar represents nucleotide substitutions per site. Bootstrap values less than 100% are shown in gray at the nodes. Clades are indicated by colors: clade FL (orange) and clade IA (green). The colors of strain numbers reflect their isolation sources: free-living (orange) and insect-associated (green). (B) The geographic distribution of isolates. Strain CGMCC 4.1369ᵀ is not presented here because information on its geographic origin was unavailable. (C) Ancestral population structure of *S. olivaceus* reconstructed using fastSTRUCTURE. Bar plots show the proportion of each strain originating from two estimated ancestral populations.

7.4% of recombination events in clades FL and IA, respectively, were explicitly predicted to have originated from the other clade. The evidently reduced interclade recombination events suggested restricted gene flow between the two clades.

To confirm the gene flow pattern, we employed PopCOGenT (6) to predict the contemporary recombination pattern and fine-scale population structure of *S. olivaceus*. This approach identified three separate populations, I to III (Fig. 3A). Population I corresponded well to clade FL and exclusively included most FL strains (except for two FL strains that were assigned as singletons). Interestingly, the IA strains were assigned to populations II and III, although they were isolated from a single insect species collected in the same area in Costa Rica. This result indicated limited recent gene flow between the FL and IA lineages and even within the IA lineage. The singleton nodes might imply underrepresented diversity due to limited sampling. In addition, as a complement to the overall and contemporary recombination analyses, we used fastGEAR to infer non-ancestral (affected a subset of recipient lineages) and ancestral (shared by all strains in lineages) recombination (43). Notably, most recombination events were identified as nonancestral, including those between the clades; only a dozen events within clade IA

TABLE 1 Recombination parameters of *S. olivaceus* quantified using ClonalFrameML v1.12

Clade	Events	R/θ^a	δ^b	ν^c	r/m^d
Clade IA ($n = 8$)	985	1.3693	88.1951	0.0327	3.9478
Clade FL ($n = 29$)	1,917	0.4951	83.9948	0.0357	1.4843
Clade FL subset[e] ($n = 11$)	1,921	0.7650	96.1770	0.0334	2.4583
Total subset[f] ($n = 17$)	2,761	0.5786	82.7363	0.0330	1.5817
Total ($n = 37$)	2,766	0.4752	75.8956	0.0365	1.3179

[a]Per-site rate of initiation of recombination relative to mutation.
[b]Mean length of DNA imported by homologous recombination.
[c]Divergence rate, per site, of DNA imported by homologous recombination.
[d]The relative impact of recombination versus mutation on the per-site substitution rate. Equal to $(R/\theta) \times \delta \times \nu$.
[e]A set of clade FL strains without 18 FL strains that are collapsed into single nodes in Fig. 3A.
[f]A set of *S. olivaceus* strains without 18 FL strains and 2 IA strains that are collapsed into single nodes in Fig. 3A.

were ancestral (Fig. 3B). Once again, the result that recombination was much more frequent within than between the clades confirmed the gene flow pattern (Fig. 3B).

Moreover, we detected relatively high genome-wide interclade genetic differentiation (F_{ST}) values (0.61 ± 0.06, mean $\pm$ standard deviation [SD]) (Fig. 4A), signifying strong genetic differentiation between the lineages. In addition, among 168,047 single nucleotide polymorphisms (SNPs), 11,994 were clade-specific SNPs with some alleles present in clade IA but others in clade FL. The clade-specific SNPs were distributed throughout the core genome, with only four small peaks (Table S3) containing four times the average SNPs (11.3 SNPs) in a 5-kb sliding window (Fig. 4B). This SNP distribution pattern provided additional evidence for genome differentiation and restricted gene flow between the lineages.

Geographic isolation versus ecological isolation. Both ecological differentiation and limited gene flow caused by physical factors can generate population genetic patterns in bacteria (44). To examine the relative contributions of these processes in *S. olivaceus*, we considered both geographic and ecological information of the strains. FL strains were isolated from habitats spanning 0- to 107-m geographic distances, indicating an ambiguous biogeographical pattern. Additionally, the Mantel test for the 28 FL strains (without strain CGMCC 4.1681^T due to a lack of geographical location information) indicated no significant correlation between phylogenetic and geographic
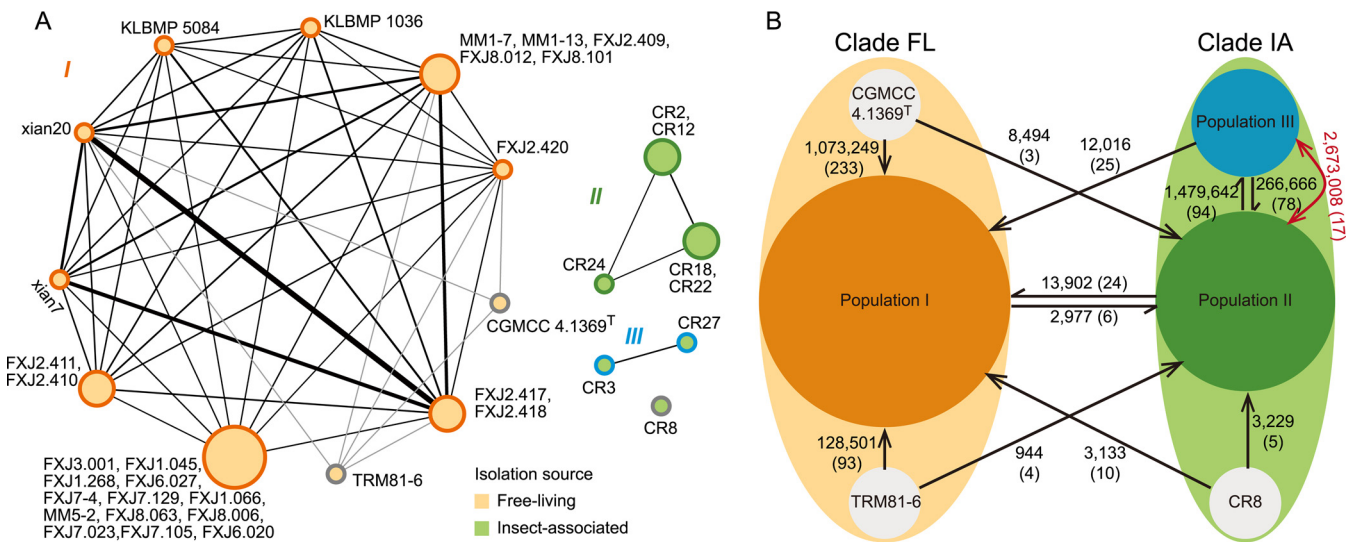

FIG 3 (A) Gene flow network and population structure identified with PopCOGenT. Edges represent the gene flow between genomes, with thicknesses corresponding to the amount of inferred gene flow and colors representing gene flow within (black) and between (gray) populations. Nodes are colored by isolation source, and node outlines are colored by population assignment: population I (orange), population II (green), population III (blue), and singletons (gray). Node size represents the number of clonal clusters (isolates that are too closely related are collapsed into a single node). (B) Diagram of recombination in the core genomes of *S. olivaceus* identified with fastGEAR. Population assignments were obtained from PopCOGenT. Black arrows indicate recent recombination events from donor to recipient populations. The red arrows indicate ancestral recombination events. The total length of recombinant regions (base pairs) and the numbers of recombination events are labeled on the vectors.
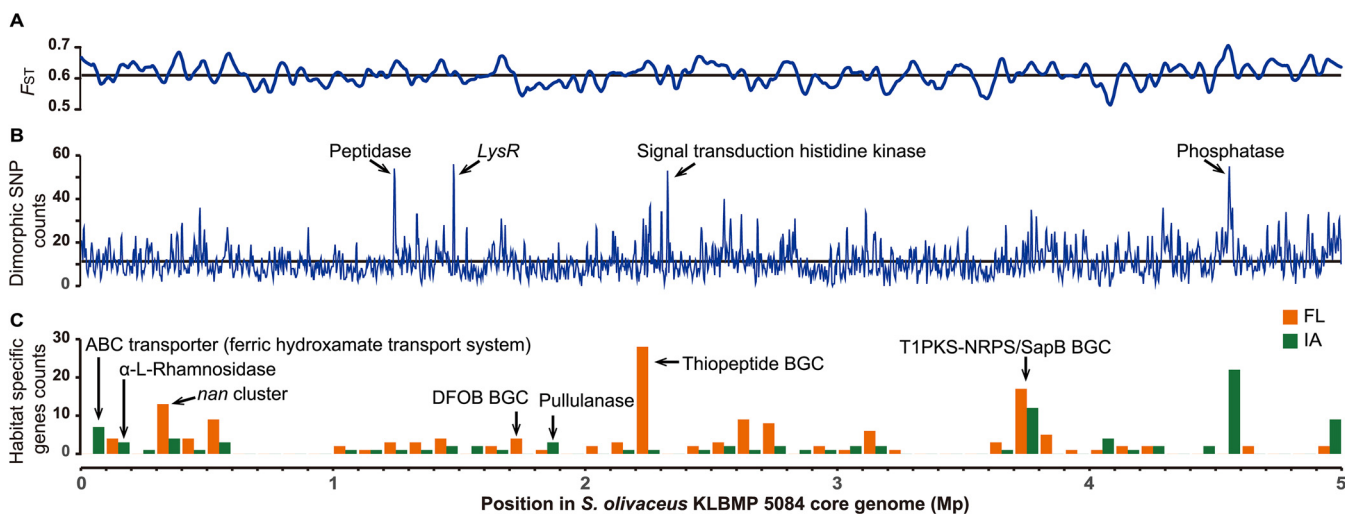
FIG 4 Genome differentiation of *S. olivaceus*. (A) Values of genetic differentiation (F_{ST}) for 5-kb sliding windows moving in 2.5-kb steps are shown. This indicates genome-wide differentiation between clades FL and IA. The black line shows the mean F_{ST} value per 5 kb. (B) The numbers of clade-specific dimorphic SNPs within a 5-kb sliding window with a 2.5-kb overlap plotted along the core genome. The black line shows the average number of clade-specific SNPs per 5 kb. (C) Bar graph showing the distribution of clade-specific genes along the core genome. Each bar represents the number of genes within a 100-kb sliding window. BGC, biosynthetic gene cluster; DFOB, desferrioxamine B; *nan*, sialic acid catabolic gene cluster.

distances ($r = 0.072$, $P = 0.621$). These findings suggest that the movement of *S. olivaceus* is not limited among FL habitats on intercontinental scales. For example, soil isolates FXJ1.045 and FXJ1.066 were geographically separated from deep-sea isolate FXJ8.063 over 10^7 m, while they were clustered into a single node in PopCOGenT analysis (Fig. 3A), indicating a recent dispersal event or rampant gene flow. Because geography was confounded with habitat type in our study, we performed partial Mantel tests for all strains, which revealed a significant effect of habitat type ($r = 0.935$, $P < 0.001$) but not geographic distance ($r = -0.237$, $P = 0.999$) on phylogenetic distance.

Because the two clades reside in distinct ecological habitats, environmental selection may play an important role in genetic divergence, which could be proven by ecological specialization to different habitat types. To test this hypothesis, we performed principal-component analysis (PCA) on clusters of orthologous groups (COG) annotation profiles of the genomes to explore the functional differentiation in *S. olivaceus*. Two separate functional groups were identified, which displayed high congruency with the genetic structure (Fig. S3A), suggesting functional divergence between clades FL and IA. Moreover, the intraclade strains shared more accessory genes than the interclade strains (Fig. S3B), indicating that the accessory genome might carry functionally important genes that facilitate habitat adaptation and drive population differentiation.

Functional divergence in the accessory genome between lineages. To further investigate functional differentiation in the accessory genomes, we explored the clade-specific genes that are universal in one clade but absent in the other, which might contain candidate habitat-adaptive genes. We found 150 and 98 orthologous genes (OGs) that were specific to clades FL and IA, respectively (Table S4), and distributed along the genome (Fig. 4C). Based on COG annotation, 175 clade-specific OGs (70.6%) were functionally categorized (Fig. S3C). Among these, 42 OGs, the highest number of OGs, were assigned to the COG class transcription (COG K), suggesting that transcriptional regulation facilitates *S. olivaceus* survival in specific habitats.

More detailed analyses of the functions of the clade-specific OGs revealed that many of them were involved in resource exploitation and secondary metabolism. These OGs were often physically adjacent and clustered into genomic regions. Particularly interesting was the distribution of the sialic acid catabolic gene cluster (*nan*) (from KLBMP 5084_605 to KLBMP 5084_615) (Fig. 5A), which was reported as an FL habitat-associated genomic region in our previous study of *S. albidoflavus* (37). Two other interesting regions were involved in the uptake of iron, encoding an ABC transporter used for hydroxamate-type
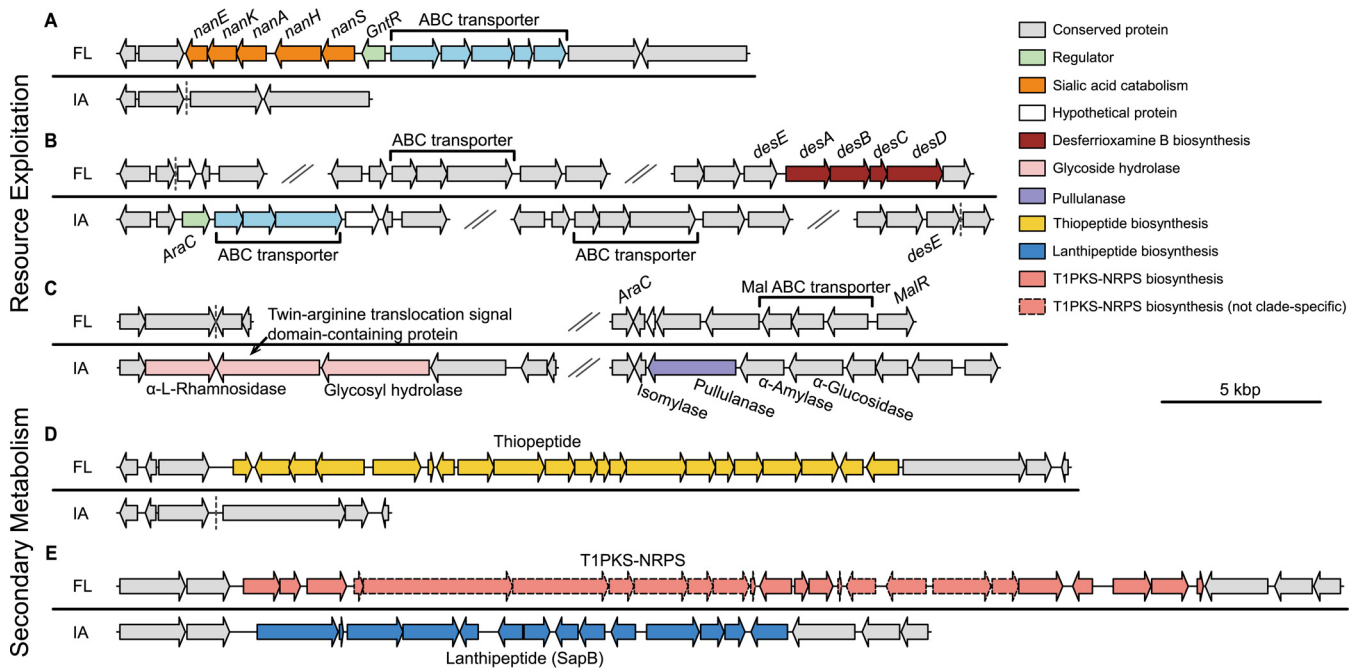
FIG 5 Genetic organizations of the regions containing clade-specific genes involved in resource exploitation and secondary metabolism. Conserved genomic regions flanking these genes are colored in gray. (A) The *nan* gene cluster for sialic acid catabolism. (B) The transport system for the hydroxamate-type iron complex and the biosynthesis gene cluster of desferrioxamine B. (C) The region containing a gene annotated as *α*-L-rhamnosidase and the region containing a putative pullulanase gene in a starch catabolic gene cluster. (D) The biosynthesis gene cluster of a thiopeptide. (E) The biosynthesis gene clusters of a T1PKS-NRPS and SapB.

iron complex import (from CR12_85 to CR12_88) and a biosynthetic pathway for desferrioxamine B (DFOB) (from KLBMP 5084_2430 to KLBMP 5084_2433) (Fig. 5B). Although DFOB is a hydroxamate-type siderophore widespread in *Streptomyces* (45), all IA strains retained the complete transport system of DFOB but had lost its biosynthetic genes (46). In addition to the transport system that has high homology with its counterpart in FL strains, IA strains also had another hydroxamate transport system that was clade-specific (from CR12_85 to CR12_88). Besides these gene clusters, several clade-specific genes may also confer benefits to the populations that harbor them. For example, two genes (CR12_143 and CR12_1679) specific to IA strains were annotated as *α*-L-rhamnosidase and pullulanase, respectively (Fig. 5C), which might help break down complex carbohydrates such as naringin, glycogen, and amylopectin (47, 48). Further analysis of the genomic regions flanking these genes showed that the pullulanase gene was in a starch catabolic gene cluster that was conserved across all strains (Fig. 5C), indicating a gene loss event in FL strains.

Moreover, there were two clade-specific regions harboring secondary metabolite biosynthetic gene clusters (BGCs). One region was specific to FL strains (from KLBMP 5084_3249 to KLBMP 5084_3269) and predicted to encode a thiopeptide product (Fig. 5D), which shared little homology with known BGCs in MIBiG (49). The other was a versatile region with four genotypes (Fig. 5E). One genotype (from CR12_5271 to CR12_5284) was specific to IA strains, encoding a BGC displaying high homology with SapB (78%), a lanthipeptide that functions as a biological surfactant and facilitates the formation of aerial mycelium (50). The other three genotypes existed in FL strains, including several FL-specific genes (from KLBMP 5084_5460 to KLBMP 5084_5462, from KLBMP 5084_5471 to KLBMP 5084_5473, and from KLBMP 5084_5479 to KLBMP 5084_5483). These genes were part of a T1PKS-NRPS-type BGC, which had no significant hits in MIBiG. Overall, the clade-specific genes tend to be involved in habitat adaptation, indicating different evolutionary trajectories between the FL and IA lineages.

Evidence for habitat-associated physiological differentiation. Given that little free iron is available in host environments (51), the ability to scavenge this element
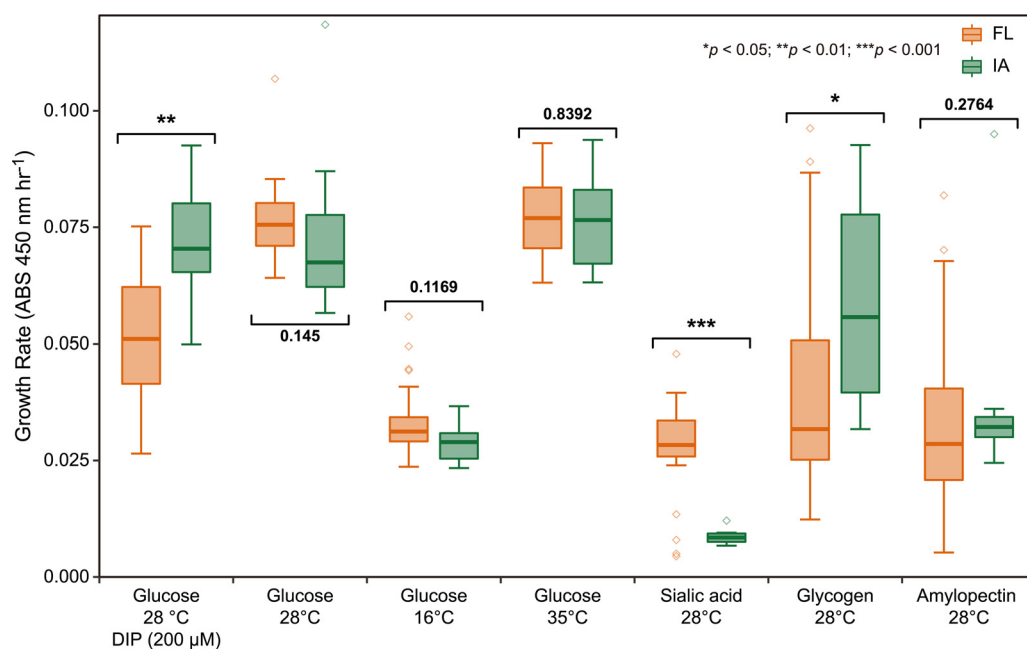
FIG 6 Growth rates (ABS 450 nm h^{-1}) of the two lineages under different iron availability, temperature, and sole carbon source conditions, in which three replicates were performed for each strain. The free-iron-limited condition was simulated by adding the iron chelator DIP. The box shows the upper and lower quartiles, and the short line within the box represents the median. The error bar depicts the 1.5 interquartile range above the upper box boundary and below the lower box boundary. Outliers are plotted as individual points. *P* values were determined using a Mann-Whitney *U* test.

from free-iron-deficient environments is essential for the survival of host-associated microorganisms. We performed an absorbance-based assay to quantify the growth rates of *S. olivaceus* strains under different iron availability conditions. Low iron availability was simulated by adding the iron chelator 2,2′-bipyridyl (DIP) to the medium. FL and IA strains had nearly equal growth rates on the control medium without DIP, whereas IA strains grew significantly faster than FL strains on the medium with DIP (Mann-Whitney *U* test, $P = 0.002$) (Fig. 6, Table S5). This result indicates that IA strains have greater average fitness than FL strains under free-iron-limiting conditions, which might be the result of adaptation to insect-associated habitats.

The growth of *S. olivaceus* under different temperature conditions was tested using the same approach. We found no statistically significant difference in growth rates between clades FL and IA at 16, 28, or 35°C (Fig. 6, Table S5). Considering that the strains were isolated from different latitudes, we separated them into tropical (16 strains) and subtropical (20 strains) groups. However, the two groups showed no significant differences in growth rates at different temperatures (Fig. S4). Motivated by the prediction of clade-specific genes involved in carbohydrate utilization, we compared the growth rates of FL and IA strains on media containing different sole carbon sources. As expected, sialic acid and glycogen differentiated the two clades very well (Fig. 6, Table S5). Sialic acid could be utilized only by FL strains (Mann-Whitney *U* test, $P < 0.001$), while glycogen supported faster growth of IA strains than of FL strains (Mann-Whitney *U* test, $P = 0.037$) (Fig. 6, Table S5). However, amylopectin did not differentiate the two clades (Mann-Whitney *U* test, $P = 0.276$). Neither carbohydrate utilization nor iron acquisition capacity differentiated the tropical and subtropical groups (Fig. S4).

DISCUSSION

Different evolutionary forces interact in nature. Here, we aimed to identify the processes responsible for the differentiation of two closely related but genetically distinct *Streptomyces* lineages (FL and IA), considering both geographic and ecological factors. One possible explanation for this genetic pattern is geographic isolation (12), which

has been observed in the hyperthermophilic archaeon *Sulfolobus* (9), the bacterial pathogen *Salmonella* (52), and the free-living soil bacterium *Curtobacterium* (11). This scenario, however, did not explain the dispersal limitation observed here. The long-distance dispersal and gene flow of *S. olivaceus* were not limited at intercontinental scales (over 10^7 m). Notably, although the two clades were isolated at different latitudes, growth under different temperatures did not differentiate them.

A promising interpretation for the differentiation of the two lineages is ecological divergence, which has been proven to be a driver of bacterial speciation (11, 25, 53). In contrast to previous studies of ecological speciation, ours did not reveal obvious genomic "islands" or even "continents" in the core genomes, which is a signature of ecological differentiation (19, 44, 53). This is because these genomic patterns persist for only a short time after the population splits (19, 54). The two lineages of *S. olivaceus* diverged approximately 6.91 million years (my) ago (confidence interval, 4.74 to 10.09 my ago; Fig. S5), which is comparable to the time span (8 ± 2.5 my) required for *Streptomyces* species to accumulate 1% core gene amino acid divergence (55). Combined with the relatively high number of clade-specific dimorphic SNPs and the less evident dimorphic SNP peaks in *S. olivaceus* genomes (Fig. 4B), these results suggest that after a long period of ecological isolation and dispersal limitation, the initial recombination barriers have spread throughout the genomes, resulting in two young sister species.

As massive gene flux may occur within natural *Streptomyces* populations (56), the speciation of *S. olivaceus* was likely initiated by gene gain or loss in the ancestral population, which allowed the mutants to colonize a novel habitat (53). The observation that most of the clade-specific regions were flanked by conserved core genes, combined with the absence of mobile genetic elements around these regions (Fig. 5), suggests an important role of homologous recombination in the gene gain or loss. Subsequent homologous recombination, possibly combined with vertical inheritance, might have facilitated the sweep of habitat-specific genes among new populations, leading to dispersal limitation across habitats due to fitness trade-offs. Thus, the reduced encounters and accumulated sequence divergence between populations could gradually enhance ecological barriers to gene flow and eventually result in genetic isolation across the genome. The identification of habitat-specific genes in accessory genomes and the observation of fitness trade-offs under different environmental conditions provide strong evidence for this explanation. Although we could not determine which genes initiated ecological speciation, our analysis identified a set of habitat-adaptive genes.

Most noteworthy is the *nan* gene cluster for sialic acid catabolism. Sialic acids, a family of nine carbon monosaccharides, are basic components of some cell surface glycoproteins and glycolipids (57). Sialic acids play pivotal roles in the development and function of the nervous system in insects (58–60). Therefore, sialic acid catabolism seems to be harmful to insect hosts, which is in line with the distribution pattern of the *nan* cluster and corresponding functional corroboration in *S. olivaceus* (Fig. 5 and 6). Strains without the *nan* cluster could easily colonize and adapt to insect-associated niches, while strains with the ability to utilize sialic acids may perform better within the complex and varying free-living environment. As a consequence, the dispersal from FL to IA habitats might be limited, and vice versa. These results confirmed our previous identification of the *nan* cluster as a habitat-adaptive region in *Streptomyces* species (37), providing compelling evidence of convergent evolution toward this successful ecological strategy (61, 62). Most likely, as compensation for the inability of IA strains to utilize sialic acids, they possess pullulanase genes and utilize glycogen (not amylopectin), the major carbohydrate reserve in insects (63), more efficiently than FL strains (Fig. 6). The preference for different carbon sources illustrates habitat partitioning between the two lineages.

The different growth rates of the two lineages under free-iron-limited conditions were also linked to fitness differences, providing additional evidence for ecological adaptation. Competition for iron resources is an important aspect of microbial interactions (64, 65), especially within host environments, where free iron is generally deficient (51). Hence, the loss of DFOB biosynthetic genes but retention of the DFOB

transport system and the acquisition of another transporter for the hydroxamate-type iron complex might enable IA strains to pirate siderophores for growth (16, 66, 67). This is an energy-efficient strategy consistent with the Black Queen Hypothesis (68). Based on the fact that DFOB is a common hydroxamate-type siderophore and can be accessed from the surrounding community, this strategy confers IA strains selective advantages in environments with fierce iron competition. Furthermore, we identified four core siderophore BGCs in *S. olivaceus* genomes, two of which are hydroxamate-type siderophores (aerobactin-like and coelichelin) (Table S2). How these BGCs are regulated is unclear, but the observation that IA strains have significantly higher growth rates than FL strains under iron-limiting conditions indicates that the former may produce more siderophores and transporters and consequently have greater fitness than the latter in free-iron-deficient environments. Together, the results suggest that the different iron acquisition abilities between the two clades also contribute to ecological divergence and dispersal limitation.

In addition to adaptation to insect-associated environments, IA strains, in turn, might protect their insect hosts by producing detoxifying enzymes. For example, the clade IA-specific gene CR12_143 encodes putative α-L-rhamnosidase, an enzyme that cleaves terminal α-L-rhamnose from a large number of natural glycosides, e.g., naringin, rutin, and diosmin. These flavonoids mediate the defense response in plants, and the removal of rhamnose from their skeleton would suppress this process (47). Thus, the IA strains might provide their insect hosts counterdefenses against plant toxins. Moreover, secondary metabolites, which can serve as weapons and signal molecules in microbial interactions (69), may provide producers selective advantages. For instance, the FL-specific BGCs thiopeptide and T1PKS-NRPS might contribute to the survival of FL strains by mediating local ecological interactions in the complex community.

Although the genome-wide ANI (96.93 to 99.998%) of *S. olivaceus* meets the standard threshold (95 to 96%) for prokaryotic species delineation (4, 5), the population structure, gene flow pattern, and ecological distribution and divergence results strongly indicate that the two clades represent independent lineages. These results further lead to an important question about whether the two lineages have diverged to the extent that they can represent two distinct species. When a species designation is questioned, multiple approaches should be taken into consideration. According to the unified species concept, two newly divergent lineages could be considered two species if they are evolving separately (70). In our case, despite the markedly restricted recent gene flow, low-level nonancestral recombination was still detected between the two lineages (Fig. 3), suggesting that they probably represent two incompletely separated species (71). Moreover, the two lineages are indistinguishable in the vast majority of phenotypic properties (38; unpublished data) and are far more distant from other streptomycetes than from each other in the genome-based *Streptomyces* phylogeny (Fig. S1B and C). All these findings, combined with the genetic and ecological distinction of the two lineages, support that they are recently diverged species that may still be in the gray zone of speciation described by de Queiroz (70). Although a universal ANI cutoff is taken as part of the criterion for segregating prokaryotic taxa, this approach is not suitable for the most recent products of speciation (72). Consequently, we propose to use the Open Nomenclature qualifier "species complex" (73) to accommodate the *S. olivaceus* strains. The FL lineage turns out to represent *S. olivaceus sensu stricto*, while the IA lineage may represent a cryptic species (74, 75) that needs further taxonomic investigation for a formal description.

Consistent with our previous findings in *S. albidoflavus*, the population structure of the *S. olivaceus* species complex was also associated with the change between FL and IA lifestyles. Although *S. albidoflavus* and *S. olivaceus* have been separated for more than 100 my (55), their similar evolutionary patterns highlight the potential role of lifestyle transitions in the divergence of streptomycetes. This is strengthened by the facts that *S. albidoflavus* is still at an advanced stage of speciation (37), while *S. olivaceus* has undergone recent speciation. Ecological interactions between microorganisms and insects, including nutritional and defensive interactions, have been described

extensively (76, 77). The best-studied example is the insect-fungus agricultural symbiosis, which has undergone coevolution over tens of million of years (78). The domestication of fungi has involved transitions from loose cultivation to irreversible domestication (78), which is combined with the loss of some carbohydrate-degrading enzyme genes and restricted gene flow between fungal cultivars and their FL relatives (79). Similar evolutionary trajectories were found in our studies (i.e., gene loss and reduced gene flow). Intriguingly, the convergent strategy of sialic acid catabolism loss in both the *S. albidoflavus* and *S. olivaceus* clades associated with insects might indicate coevolution between *Streptomyces* species and their insect hosts. In addition to this gene cluster, a series of habitat-adaptive genes, such as genes involved in glycogen and iron utilization, collectively contributed to the transition between FL and IA lifestyles. Taken together, our results indicate that interactions with insects facilitate diversification and ecological divergence in *Streptomyces*. Hence, there is a great need to consider insect-associated habitats, one of the major sources of *Streptomyces*, when studying *Streptomyces* speciation.

Our study has some limitations. First, clade IA included a relatively small number of isolates, and the diversity of this clade might be underestimated, as observed in the inference of recombination origins. More isolates from IA sources should be included in future studies. Another potential limitation is that clades FL and IA are not sympatric. Although we reasoned that free-living *S. olivaceus* is not geographically restricted at intercontinental scales, there is no direct evidence to exclude the effect of geographic isolation. Adding sympatric but ecologically diverged conspecific isolates to the study will be required to clarify this question, although that previous isolation did not yield *S. olivaceus* strains from Chinese insects or Costa Rican soils. Other abiotic factors and biotic variables such as interactions with local communities might also impose selection pressure on the strains. Nevertheless, our findings highlight the roles of habitat adaptation and divergence in the origin and maintenance of population structure in *Streptomyces*. Further studies focusing on closely related isolates with detailed environmental metadata are likely to provide a better understanding of ecological adaptation and its contribution to the speciation and diversity of bacteria.

MATERIALS AND METHODS

Isolate collection. A total of 37 *S. olivaceus* strains were used in this study, which were isolated from diverse habitats at multiple spatial scales (Table S1) and categorized into two habitat-associated groups, FL and IA. The FL group encompassed 19 strains isolated from soil and soil-associated habitats in China and 10 marine isolates derived from Chinese seas and the Indian Ocean. The IA group comprised 8 strains isolated from imperial moths (*Eacles imperialis*, Saturniidae) collected in Costa Rica.

Genome sequencing, assembly, and annotation. One complete reference genome (KLBMP 5084) was downloaded from the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) database. For the remaining 36 strains, short-insert paired-end libraries were constructed and sequenced by Novogene Bioinformatics Technology Co., Ltd. (Beijing, China), using the Illumina HiSeq 2500 system. All good-quality paired reads were assembled using SOAPdenovo (80). All genome accession numbers are listed in Table S1.

All genomes were processed using Prodigal v2.6.2 with default parameters to predict coding sequences (CDSs). The CDSs were annotated with the COG 2003–2014 (81), Pfam 31.0 (82), and NCBI nonredundant (NR; accession date, 10 April 2019) (83) databases. Secondary metabolite biosynthetic gene clusters (BGCs) in each genome were predicted via antiSMASH v5.0 (84). The pairwise genome-wide ANI was calculated using pyANI (85).

Pan-genome analysis and phylogenetic inference. Orthologous groups (OGs) were identified using OrthoMCL algorithms (86) implemented in the Anvi'o pan-genomic workflow v6.2 (87) with BLAST 2.7.1+ (88), and the pan-genome results were visualized using "anvi-display-pan." Single-copy core genes, which were defined as OGs with exactly one copy in each genome, were extracted from the "anvi-summarize" results. Each single-copy core OG was aligned with MUSCLE v3.8.1551 (89) at the amino acid level and then converted to nucleotide alignment through PAL2NAL v14 (90). Poorly aligned regions were trimmed using Gblocks v0.91b (91) with −t = c. The resulting alignments were concatenated with a custom Python script to obtain core genome sequences. A phylogenetic tree based on the single-copy core genomes was reconstructed using FastTree v2.1.10 (92) under the general time reversal model and visualized using iTOL (93). The population structure was predicted from the single nucleotide polymorphism (SNP) data set of single-copy core genomes using fastSTRUCTURE (39) with runs for population numbers (K values) of 2 to 10. The approximate divergence times for two clades was estimated using RelTime (94) as previously described (55). In addition, the phylogenomic tree of the genus *Streptomyces* was constructed to examine the phylogenetic status of the 37 strains using the pipelines

described by Li et al. (95). Briefly, all genomes assigned to *Streptomyces* by the EzBioCloud database (96) (accession date, 24 September 2021) were downloaded from the NCBI genome FTP site (https://www.ncbi.nlm.nih.gov/genomes). After quality control, 996 genomes, together with the 37 *S. olivaceus* genomes, were subjected to the GToTree (97) workflow to identify the set of 138 actinobacterial marker genes, which was then used to generate the *Streptomyces* phylogeny.

Recombination and gene flow analyses. ClonalFrameML v1.12 (40) was used to detect and quantify homologous recombination. Specifically, recombination events among all 37 isolates and among the phylogenetic clades were reconstructed using core genome alignments. For each recombination event, genetic distances between the recombinant fragments and their homologous sequences from all nodes and leaves were calculated with the R package APE (98) to infer the likely origin of recombination, as described in a previous study (99). The clustering algorithm PopCOGenT (6) was used to detect recent gene flow and the contemporary boundaries of gene flow units that were defined as ecologically relevant populations. FastGEAR (43) was used to identify nonancestral and ancestral recombination events, using the core genome alignments of all 37 isolates and the population assignment result of PopCOGenT as input. Additionally, F_{ST} values of the core genome in 5-kb sliding windows (2.5-kb overlap) were calculated using the R package PopGenome (100).

The contribution of geography and ecology to population differentiation. The pairwise geographic distances (if available) and phylogenetic distances based on the core genome between isolates were calculated using the R packages geosphere (101) and APE, respectively. In the matrix of habitat differences, all within-habitat-type pairwise comparisons were assigned a value of 0, whereas all FL versus IA pairwise comparisons were assigned a value of 1. The effects of geography and ecology on population structure were examined using Mantel and partial Mantel tests in the R package VEGAN (102) with 10,000 permutations. Mantel tests were performed between pairwise phylogenetic distances and geographic distances for the FL group only. Partial Mantel tests were performed for the whole species to test the effect of one factor while controlling for the other factor. This analysis was necessary given that the habitat types and sampling locations were confounded in our study.

Identification and analysis of ecologically relevant variants. Principal-component analysis (PCA) of COG annotation profiles of *S. olivaceus* genomes was performed using the prcomp function in R. The Bray-Curtis distance matrix, used as an input file for PCA, was calculated from COG annotation profiles after Hellinger transformation with the R package VEGAN. To estimate the similarity of flexible genomes, the Jaccard distances (*d*) based on the dispensable gene 0/1 matrix were calculated with vegdist in the R package VEGAN. These distances were then used as input to PHYLIP v3.698 (http://evolution.gs.washington.edu/phylip/) to generate a neighbor-joining tree. A heatmap of Jaccard similarity (1-*d*) was visualized with the R package pheatmap (https://github.com/raivokolde/pheatmap). Habitat-specific accessory genes and SNPs were identified using a custom Python script, with the pan-genome presence/absence matrix and the SNP file extracted from core genomes used as inputs, respectively.

Evaluation of growth under different conditions. To evaluate the growth of strains under different conditions, an absorbance-based approach described in a previous study (23) was used with a few modifications. In brief, a 5-μL spore suspension (10^5 CFU/mL) of each strain was plated as a lawn onto 150 μL solid medium within 96-well plates, and the absorbance at 450 nm was measured every half an hour with a CLARIOstar Plus plate reader (BMG Labtech, Ortenberg, Germany) for 3 to 5 days. The change in absorbance was plotted against time to create a growth curve, and the growth rate was determined as the slope of the curve in the linear range. To test for differences in carbon source utilization among strains, we used basal mineral salt agar medium (MM) (103) supplemented with a final concentration of 1% sole carbon sources, including *N*-acetylneuraminic acid, glycogen, and amylopectin. Growth on MM was used as a negative control, whereas growth on MM supplemented with glucose was used as a positive control. To simulate free-iron-limited conditions, the iron chelator 2,2′-bipyridyl (DIP) was added to MM agar containing 1% glucose to a final concentration of 200 μM. Control experiments were carried out using the same medium without the addition of DIP. In addition, the growth rates at different temperatures (16 to 35°C) were measured using MM agar containing 1% glucose. All growth assays were performed in triplicate.

Data availability. The genome accession numbers for the strains in this study are listed in Table S1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, TIF file, 2.2 MB.
FIG S2, TIF file, 1.1 MB.
FIG S3, TIF file, 2.4 MB.
FIG S4, TIF file, 2.6 MB.
FIG S5, TIF file, 1.4 MB.
TABLE S1, DOCX file, 0.05 MB.
TABLE S2, DOCX file, 0.03 MB.
TABLE S3, DOCX file, 0.03 MB.
TABLE S4, DOCX file, 0.1 MB.
TABLE S5, DOCX file, 0.03 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Achtman M, Wagner M. 2008. Microbial diversity and the genetic nature of microbial species. Nat Rev Microbiol 6:431–440. https://doi.org/10.1038/nrmicro1872.

2. Rocha EPC. 2018. Neutral theory, microbial practice: challenges in bacterial population genetics. Mol Biol Evol 35:1338–1347. https://doi.org/10.1093/molbev/msy078.

3. Shapiro BJ, Polz MF. 2015. Microbial speciation. Cold Spring Harb Perspect Biol 7:a018143. https://doi.org/10.1101/cshperspect.a018143.

4. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91. https://doi.org/10.1099/ijs.0.64483-0.

5. Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106: 19126–19131. https://doi.org/10.1073/pnas.0906412106.

6. Arevalo P, VanInsberghe D, Elsherbini J, Gore J, Polz MF. 2019. A reverse ecology approach based on a biological definition of microbial populations. Cell 178:820–834. https://doi.org/10.1016/j.cell.2019.06.033.

7. Finlay BJ, Clarke KJ. 1999. Ubiquitous dispersal of microbial species. Nature 400:828–828. https://doi.org/10.1038/23616.

8. Martiny JB, Bohannan BJ, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreas L, Reysenbach AL, Smith VH, Staley JT. 2006. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 4:102–112. https://doi.org/10.1038/nrmicro1341.

9. Whitaker RJ, Grogan DW, Taylor JW. 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science 301:976–978. https://doi.org/10.1126/science.1086909.

10. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JB. 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol 10:497–506. https://doi.org/10.1038/nrmicro2795.

11. Chase AB, Arevalo P, Brodie EL, Polz MF, Karaoz U, Martiny JBH. 2019. Maintenance of sympatric and allopatric populations in free-living terrestrial bacteria. mBio 10:e02361-19. https://doi.org/10.1128/mBio.02361-19.

12. Wright S. 1943. Isolation by distance. Genetics 28:114–138. https://doi.org/10.1093/genetics/28.2.114.

13. Jones DS, Schaperdoth I, Macalady JL. 2016. Biogeography of sulfur-oxidizing Acidithiobacillus populations in extremely acidic cave biofilms. ISME J 10:2879–2891. https://doi.org/10.1038/ismej.2016.74.

14. Whitaker RJ. 2006. Allopatric origins of microbial species. Philos Trans R Soc Lond B Biol Sci 361:1975–1984. https://doi.org/10.1098/rstb.2006.1927.

15. Lassalle F, Muller D, Nesme X. 2015. Ecological speciation in bacteria: reverse ecology approaches reveal the adaptive part of bacterial cladogenesis. Res Microbiol 166:729–741. https://doi.org/10.1016/j.resmic.2015.06.008.

16. Cordero OX, Polz MF. 2014. Explaining microbial genomic diversity in light of evolutionary ecology. Nat Rev Microbiol 12:263–273. https://doi.org/10.1038/nrmicro3218.

17. Wang IJ, Bradburd GS. 2014. Isolation by environment. Mol Ecol 23:5649–5662. https://doi.org/10.1111/mec.12938.

18. Polz MF, Alm EJ, Hanage WP. 2013. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. Trends Genet 29:170–175. https://doi.org/10.1016/j.tig.2012.12.006.

19. Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, Krause DJ, Whitaker RJ. 2012. Patterns of gene flow define species of thermophilic Archaea. PLoS Biol 10:e1001265. https://doi.org/10.1371/journal.pbio.1001265.

20. Polonca S. 2020. Environment shapes the intra-species diversity of Bacillus subtilis isolates. Microb Ecol 79:853–864. https://doi.org/10.1007/s00248-019-01455-y.

21. Andam CP, Doroghazi JR, Campbell AN, Kelly PJ, Choudoir MJ, Buckley DH. 2016. A latitudinal diversity gradient in terrestrial bacteria of the genus Streptomyces. mBio 7:e02200-15–e02215. https://doi.org/10.1128/mBio.02200-15.

22. Choudoir MJ, Doroghazi JR, Buckley DH. 2016. Latitude delineates patterns of biogeography in terrestrial Streptomyces. Environ Microbiol 18:4931–4945. https://doi.org/10.1111/1462-2920.13420.

23. Choudoir MJ, Buckley DH. 2018. Phylogenetic conservatism of thermal traits explains dispersal limitation and genomic differentiation of Streptomyces sister-taxa. ISME J 12:2176–2186. https://doi.org/10.1038/s41396-018-0180-3.

24. Connor N, Sikorski J, Rooney AP, Kopac S, Koeppel AF, Burger A, Cole SG, Perry EB, Krizanc D, Field NC, Slaton M, Cohan FM. 2010. Ecology of speciation in the genus Bacillus. Appl Environ Microbiol 76:1349–1358. https://doi.org/10.1128/AEM.01988-09.

25. Wang XJ, Zhang Y, Ren ML, Xia TY, Chu X, Liu C, Lin XQ, Huang YJ, Chen ZY, Yan AX, Luo HW. 2020. Cryptic speciation of a pelagic Roseobacter population varying at a few thousand nucleotide sites. ISME J 14:3106–3119. https://doi.org/10.1038/s41396-020-00743-7.

26. Flardh K, Buttner MJ. 2009. Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium. Nat Rev Microbiol 7:36–49. https://doi.org/10.1038/nrmicro1968.

27. Chater KF, Biro S, Lee KJ, Palmer T, Schrempf H. 2010. The complex extracellular biology of Streptomyces. FEMS Microbiol Rev 34:171–198. https://doi.org/10.1111/j.1574-6976.2009.00206.x.

28. van der Meij A, Worsley SF, Hutchings MI, van Wezel GP. 2017. Chemical ecology of antibiotic production by actinomycetes. FEMS Microbiol Rev 41:392–416. https://doi.org/10.1093/femsre/fux005.

29. Seipke RF, Kaltenpoth M, Hutchings MI. 2012. Streptomyces as symbionts: an emerging and widespread theme? FEMS Microbiol Rev 36:862–876. https://doi.org/10.1111/j.1574-6976.2011.00313.x.

30. Chevrette MG, Carlson CM, Ortega HE, Thomas C, Ananiev GE, Barns KJ, Book AJ, Cagnazzo J, Carlos C, Flanigan W, Grubbs KJ, Horn HA, Hoffmann FM, Klassen JL, Knack JJ, Lewin GR, McDonald BR, Muller L, Melo WGP, Pinto-Tomas AA, Schmitz A, Wendt-Pienkowski E, Wildman S, Zhao M, Zhang F, Bugni TS, Andes DR, Pupo MT, Currie CR. 2019. The antimicrobial potential of Streptomyces from insect microbiomes. Nat Commun 10:516. https://doi.org/10.1038/s41467-019-08438-0.

31. Kaltenpoth M. 2009. Actinobacteria as mutualists: general healthcare for insects? Trends Microbiol 17:529–535. https://doi.org/10.1016/j.tim.2009.09.006.

32. Van Arnam EB, Currie CR, Clardy J. 2018. Defense contracts: molecular protection in insect-microbe symbioses. Chem Soc Rev 47:1638–1651. https://doi.org/10.1039/c7cs00340d.

33. Haeder S, Wirth R, Herz H, Spiteller D. 2009. Candicidin-producing Streptomyces support leaf-cutting ants to protect their fungus garden against

the pathogenic fungus *Escovopsis*. Proc Natl Acad Sci U S A 106: 4742–4746. https://doi.org/10.1073/pnas.0812082106.

34. Kim KH, Ramadhar TR, Beemelmanns C, Cao SG, Poulsen M, Currie CR, Clardy J. 2014. Natalamycin A, an ansamycin from a termite-associated *Streptomyces* sp. Chem Sci 5:4333–4338. https://doi.org/10.1039/C4SC01136H.

35. Book AJ, Lewin GR, McDonald BR, Takasuka TE, Wendt-Pienkowski E, Doering DT, Suh S, Raffa KF, Fox BG, Currie CR. 2016. Evolution of high cellulolytic activity in symbiotic *Streptomyces* through selection of expanded gene content and coordinated gene expression. PLoS Biol 14:e1002475. https://doi.org/10.1371/journal.pbio.1002475.

36. Cheng K, Rong X, Pinto-Tomas AA, Fernandez-Villalobos M, Murillo-Cruz C, Huang Y. 2015. Population genetic analysis of *Streptomyces albidoflavus* reveals habitat barriers to homologous recombination in the diversification of streptomycetes. Appl Environ Microbiol 81:966–975. https://doi.org/10.1128/AEM.02925-14.

37. Li YS, Pinto-Tomas AA, Rong XY, Cheng K, Liu MH, Huang Y. 2019. Population genomics insights into adaptive evolution and ecological differentiation in streptomycetes. Appl Environ Microbiol 85:e02555-18. https://doi.org/10.1128/AEM.02555-18.

38. Cheng K. 2016. The roles of homologous recombination and habitat barriers to it in the evolution of streptomycetes. PhD thesis. University of Chinese Academy of Sciences, Beijing, China.

39. Raj A, Stephens M, Pritchard JK. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197: 573–U207. https://doi.org/10.1534/genetics.114.164350.

40. Didelot X, Wilson DJ. 2015. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLoS Comput Biol 11:e1004041. https://doi.org/10.1371/journal.pcbi.1004041.

41. Ravenhall M, Skunca N, Lassalle F, Dessimoz C. 2015. Inferring horizontal gene transfer. PLoS Comput Biol 11:e1004095. https://doi.org/10.1371/journal.pcbi.1004095.

42. Shapiro BJ, Polz MF. 2014. Ordering microbial diversity into ecologically and genetically cohesive units. Trends Microbiol 22:235–247. https://doi.org/10.1016/j.tim.2014.02.006.

43. Mostowy R, Croucher NJ, Andam CP, Corander J, Hanage WP, Marttinen P. 2017. Efficient inference of recent and ancestral recombination within bacterial populations. Mol Biol Evol 34:1167–1182. https://doi.org/10.1093/molbev/msx066.

44. Krause DJ, Whitaker RJ. 2015. Inferring speciation processes from patterns of natural variation in microbial genomes. Syst Biol 64:926–935. https://doi.org/10.1093/sysbio/syv050.

45. Challis GL. 2005. A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. Chembiochem 6:601–611. https://doi.org/10.1002/cbic.200400283.

46. Barona-Gomez F, Lautru S, Francou FX, Leblond P, Pernodet JL, Challis GL. 2006. Multiple biosynthetic and uptake systems mediate siderophore-dependent iron acquisition in *Streptomyces coelicolor* A3(2) and *Streptomyces ambofaciens* ATCC 23877. Microbiology (Reading) 152: 3355–3366. https://doi.org/10.1099/mic.0.29161-0.

47. Eliades LA, Rojas NL, Cabello MN, Voget CE, Saparrat MC. 2011. Alpha-L-rhamnosidase and beta-D-glucosidase activities in fungal strains isolated from alkaline soils and their potential in naringin hydrolysis. J Basic Microbiol 51:659–665. https://doi.org/10.1002/jobm.201100163.

48. Hii SL, Tan JS, Ling TC, Ariff AB. 2012. Pullulanase: role in starch hydrolysis and potential industrial applications. Enzyme Res 2012:921362. https://doi.org/10.1155/2012/921362.

49. Kautsar SA, Blin K, Shaw S, Navarro-Munoz JC, Terlouw BR, van der Hooft JJJ, van Santen JA, Tracanna V, Duran HGS, Andreu VP, Selem-Mojica N, Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian L, Collemare J, Linington RG, Weber T, Medema MH. 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res 48:D454–D458.

50. Kodani S, Hudson ME, Durrant MC, Buttner MJ, Nodwell JR, Willey JM. 2004. The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene *ramS* in *Streptomyces coelicolor*. Proc Natl Acad Sci U S A 101:11448–11453. https://doi.org/10.1073/pnas.0404220101.

51. Schaible UE, Kaufmann SH. 2004. Iron and microbial infection. Nat Rev Microbiol 2:946–953. https://doi.org/10.1038/nrmicro1046.

52. Liao J, Orsi RH, Carroll LM, Wiedmann M. 2020. Comparative genomics reveals different population structures associated with host and geographic origin in antimicrobial-resistant *Salmonella enterica*. Environ Microbiol 22:2811–2828. https://doi.org/10.1111/1462-2920.15014.

53. Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabo G, Polz MF, Alm EJ. 2012. Population genomics of early events in the ecological differentiation of bacteria. Science 336:48–51. https://doi.org/10.1126/science.1218198.

54. Wolf JB, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of speciation. Nat Rev Genet 18:87–100. https://doi.org/10.1038/nrg.2016.133.

55. McDonald BR, Currie CR. 2017. Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. mBio 8:e00644-17. https://doi.org/10.1128/mBio.00644-17.

56. Tidjani AR, Lorenzi JN, Toussaint M, van Dijk E, Naquin D, Lespinet O, Bontemps C, Leblond P. 2019. Massive gene flux drives genome diversity between sympatric *Streptomyces* conspecifics. mBio 10:e01533-19. https://doi.org/10.1128/mBio.01533-19.

57. Angata T, Varki A. 2002. Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. Chem Rev 102:439–469. https://doi.org/10.1021/cr000407m.

58. Repnikova E, Koles K, Nakamura M, Pitts J, Li H, Ambavane A, Zoran MJ, Panin VM. 2010. Sialyltransferase regulates nervous system function in *Drosophila*. J Neurosci 30:6466–6476. https://doi.org/10.1523/JNEUROSCI.5253-09.2010.

59. Soya S, Şahar U, Yıkılmaz M, Karaçalı S. 2017. Determination of sialic acids in the nervous system of silkworm (*Bombyx mori* L.): effects of aging and development. Arch Biol Sci (Beogr) 69:369–378. https://doi.org/10.2298/ABS160401117S.

60. Ghosh S. 2018. Sialylation and sialyltransferase in insects. Glycoconj J 35: 433–441. https://doi.org/10.1007/s10719-018-9835-6.

61. Stern DL. 2013. The genetic causes of convergent evolution. Nat Rev Genet 14:751–764. https://doi.org/10.1038/nrg3483.

62. Sackton TB, Clark N. 2019. Convergent evolution in the genomics era: new insights and directions. Philos Trans R Soc Lond B Biol Sci 374: 20190102. https://doi.org/10.1098/rstb.2019.0102.

63. Tolmasky DS, Rabossi A, Quesada-Allue LA. 2001. Synthesis and mobilization of glycogen during metamorphosis of the medfly *Ceratitis capitata*. Arch Biochem Biophys 392:38–47. https://doi.org/10.1006/abbi.2001.2394.

64. Barber MF, Elde NC. 2015. Buried treasure: evolutionary perspectives on microbial iron piracy. Trends Genet 31:627–636. https://doi.org/10.1016/j.tig.2015.09.001.

65. Gu S, Wei Z, Shao Z, Friman VP, Cao K, Yang T, Kramer J, Wang X, Li M, Mei X, Xu Y, Shen Q, Kummerli R, Jousset A. 2020. Competition for iron drives phytopathogen control by natural rhizosphere microbiomes. Nat Microbiol 5:1002–1010. https://doi.org/10.1038/s41564-020-0719-8.

66. West SA, Buckling A. 2003. Cooperation, virulence and siderophore production in bacterial parasites. Proc Biol Sci 270:37–44. https://doi.org/10.1098/rspb.2002.2209.

67. Cordero OX, Ventouras LA, DeLong EF, Polz MF. 2012. Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. Proc Natl Acad Sci U S A 109:20059–20064. https://doi.org/10.1073/pnas.1213344109.

68. Morris JJ, Lenski RE, Zinser ER. 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. mBio 3:e00036-12. https://doi.org/10.1128/mBio.00036-12.

69. Tyc O, Song C, Dickschat JS, Vos M, Garbeva P. 2017. The ecological role of volatile and soluble secondary metabolites produced by soil bacteria. Trends Microbiol 25:280–292. https://doi.org/10.1016/j.tim.2016.12.002.

70. de Queiroz K. 2007. Species concepts and species delimitation. Syst Biol 56:879–886. https://doi.org/10.1080/10635150701701083.

71. de Queiroz K. 2020. An updated concept of subspecies resolves a dispute about the taxonomy of incompletely separated lineages. Herpetol Rev 51:459–461.

72. Cohan FM. 2019. Systematics: the cohesive nature of bacterial species taxa. Curr Biol 29:R169–R172. https://doi.org/10.1016/j.cub.2019.01.033.

73. Sigovini M, Keppel E, Tagliapietra D, Isaac N. 2016. Open Nomenclature in the biodiversity era. Methods Ecol Evol 7:1217–1225. https://doi.org/10.1111/2041-210X.12594.

74. Bickford D, Lohman DJ, Sodhi NS, Ng PK, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. Trends Ecol Evol 22:148–155. https://doi.org/10.1016/j.tree.2006.11.004.

75. Struck TH, Feder JL, Bendiksby M, Birkeland S, Cerca J, Gusarov VI, Kistenich S, Larsson KH, Liow LH, Nowak MD, Stedje B, Bachmann L, Dimitrov D. 2018. Finding evolutionary processes hidden in cryptic species. Trends Ecol Evol 33:153–163. https://doi.org/10.1016/j.tree.2017.11.007.

76. Steinert M, Hentschel U, Hacker J. 2000. Symbiosis and pathogenesis: evolution of the microbe-host interaction. Naturwissenschaften 87:1–11. https://doi.org/10.1007/s001140050001.

77. Brownlie JC, Johnson KN. 2009. Symbiont-mediated protection in insect hosts. Trends Microbiol 17:348–354. https://doi.org/10.1016/j.tim.2009.05.005.

78. Schultz TR, Brady SG. 2008. Major evolutionary transitions in ant agriculture. Proc Natl Acad Sci U S A 105:5435–5440. https://doi.org/10.1073/pnas.0711024105.

79. Mueller UG, Gerardo NM, Aanen DK, Six DL, Schultz TR. 2005. The evolution of agriculture in insects. Annu Rev Ecol Evol Syst 36:563–595. https://doi.org/10.1146/annurev.ecolsys.36.102003.152626.

80. Luo RB, Liu BH, Xie YL, Li ZY, Huang WH, Yuan JY, He GZ, Chen YX, Pan Q, Liu YJ, Tang JB, Wu GX, Zhang H, Shi YJ, Liu Y, Yu C, Wang B, Lu Y, Han CL, Cheung DW, Yiu SM, Peng SL, Zhu XQ, Liu GM, Liao XK, Li YR, Yang HM, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18. https://doi.org/10.1186/2047-217X-1-18.

81. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43:D261–D269. https://doi.org/10.1093/nar/gku1223.

82. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. Nucleic Acids Res 40:D290–D301. https://doi.org/10.1093/nar/gkr1065.

83. Li WZ, Jaroszewski L, Godzik A. 2002. Tolerating some redundancy significantly speeds up clustering of large protein databases. Bioinformatics 18:77–82. https://doi.org/10.1093/bioinformatics/18.1.77.

84. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310.

85. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods 8:12–24. https://doi.org/10.1039/C5AY02550H.

86. van Dongen S, Abreu-Goodger C. 2012. Using MCL to extract clusters from networks. Methods Mol Biol 804:281–295. https://doi.org/10.1007/978-1-61779-361-5_15.

87. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319. https://doi.org/10.7717/peerj.1319.

88. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

89. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.

90. Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–W612. https://doi.org/10.1093/nar/gkl315.

91. Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

92. Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26:1641–1650. https://doi.org/10.1093/molbev/msp077.

93. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242–W245. https://doi.org/10.1093/nar/gkw290.

94. Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S. 2012. Estimating divergence times in large molecular phylogenies. Proc Natl Acad Sci U S A 109:19333–19338. https://doi.org/10.1073/pnas.1213199109.

95. Li Y, Wang M, Sun ZZ, Xie BB. 2021. Comparative genomic insights into the taxonomic classification, diversity, and secondary metabolic potentials of *Kitasatospora*, a genus closely related to *Streptomyces*. Front Microbiol 12:683814. https://doi.org/10.3389/fmicb.2021.683814.

96. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755.

97. Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics 35:4162–4164. https://doi.org/10.1093/bioinformatics/btz188.

98. Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290. https://doi.org/10.1093/bioinformatics/btg412.

99. Ozer EA, Nnah E, Didelot X, Whitaker RJ, Hauser AR. 2019. The population structure of *Pseudomonas aeruginosa* is characterized by genetic isolation of *exoU+* and *exoS+* lineages. Genome Biol Evol 11:1780–1796. https://doi.org/10.1093/gbe/evz119.

100. Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol 31:1929–1936. https://doi.org/10.1093/molbev/msu136.

101. Karney CFF. 2013. Algorithms for geodesics. J Geod 87:43–55. https://doi.org/10.1007/s00190-012-0578-z.

102. Dixon P. 2003. VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.

103. Pridham TG, Gottlieb D. 1948. The utilization of carbon compounds by some *Actinomycetales* as an aid for species determination. J Bacteriol 56:107–114. https://doi.org/10.1128/jb.56.1.107-114.1948.