Whole genome–based comparative analysis of the genus *Streptomyces* reveals many misclassifications

Marieke Mispelaere[1] · Anne-Sofie De Rop[2] · Cedric Hermans[1] · Sofie L. De Maeseneire[2] · Wim K. Soetaert[2] · Maarten L. De Mol[2] · Paco Hulpiau[1]

Abstract

Streptomyces species are experts in the production of bioactive secondary metabolites; however, their taxonomy has fallen victim of the tremendous interest shown by the scientific community, evident in the discovery of numerous synonymous in public repositories. Based on genomic data from NCBI Datasets and nomenclature from the LPSN database, we compiled a dataset of 600 *Streptomyces* species along with their annotations and metadata. To pinpoint the most suitable taxonomic classification method, we conducted a comprehensive assessment comparing multiple methodologies, including analysis of 16S rRNA, individual housekeeping genes, multilocus sequence analysis (MLSA), and Fast Average Nucleotide Identity (FastANI) on a subset of 409 species with complete data. Due to insufficient resolution of 16S rRNA and inconsistency observed in individual housekeeping genes, we performed a more in-depth analysis, comparing only FastANI and MLSA, which expanded our dataset to include 502 species. With FastANI validated as the preferred method, we conducted pairwise analysis on the entire dataset identifying 59 non-unique species among the 600, and subsequently refined the dataset to 541 unique species. Additionally, we collected data on 724 uncharacterized *Streptomyces* strains to investigate the uniqueness potential of the unannotated fraction of the *Streptomyces* genus. Utilizing FastANI, 289 strains could be successfully classified into one of the 541 *Streptomyces* species.

Key points

- *Evaluation of taxonomic classification methods for Streptomyces species.*
- *Whole genome analysis, specifically FastANI, has been chosen as preferred method.*
- *Various reclassifications are proposed within the Streptomyces genus.*

✉ Maarten L. De Mol
maarten.demol@ugent.be

✉ Paco Hulpiau
paco.hulpiau@howest.be

Marieke Mispelaere
marieke.mispelaere@howest.be

Anne-Sofie De Rop
annesofie.derop@ugent.be

Cedric Hermans
cedric.hermans@howest.be

Sofie L. De Maeseneire
sofie.demaeseneire@ugent.be

Wim K. Soetaert
wim.soetaert@ugent.be

1   Bioinformatics Knowledge Center (BiKC), Cluster Life Sciences, Campus Brugge Station, Howest University of Applied Sciences, Rijselstraat 5, 8200 Brugge, Belgium

2   Centre for Industrial Biotechnology and Biocatalysis (InBio.Be), Department of Biotechnology, Faculty of Bio-Science Engineering, Ghent University, Coupure Links 653, 9000 Ghent, Belgium

Introduction

Taxonomy plays a prominent role in understanding biology, evolution, and the interplay between environmental conditions and biological populations. Historically, the classification of life forms was based on shared characteristics such as anatomy, developmental processes, and behaviours (phenotype) while today genotypes form the basis for a taxonomic framework due to their uniqueness and the ever-increasing deposit of genetic information (Hugenholtz et al. 2021). Prokaryotes are single cell organisms that have dominated most of life's history on Earth and have survived many different environmental conditions, endowing extremely diverse genotypes (Nealson and Conrad 1999; Preiner et al. 2020). Among these prokaryotes, the phylum *Actinomycetota*, of which *Streptomyces* species represent the largest genus, has proven a treasure trove for natural product discovery and the pharmaceutical industry. Specifically, *Streptomyces* species are responsible for the production of around 100,000 antibiotic compounds, accounting for 70–80% of all natural bioactive products with applications in the pharmaceutical and agrochemical industry (Alam et al. 2022). However, their secondary metabolites are not limited to antibiotic properties but also possess a huge potential as herbicides, antifungal agents, and antiviral and anticancer drugs (Alam et al. 2022; Del Carratore et al. 2022). Due to the perpetual accumulation of new genomic information and the application potential of *Streptomyces* secondary metabolites (Lacey and Rutledge 2022), proper classification is of paramount importance.

Conventionally, 16S rRNA gene analysis was the gold standard for bacterial and thus *Streptomyces* phylogenetics, species delineation, and the identification of new species (Komaki 2022; Hassler et al. 2022). However, it suffers from poor taxonomic resolution at the species level (O'Connell et al. 2022), intragenomic heterogeneity, horizontal gene transfer (HGT), and recombination reports (Hassler et al. 2022). There has been an extensive discussion on the 16S rRNA threshold value for species discrimination, a parameter subject to regular updates. Currently, the threshold used in the literature ranges between 98.2 and 99% (Meier-Kolthoff et al. 2013). However, even when the similarity exceeds the threshold of 99%, the definitive assignment of whether they are of the same species remains uncertain. For example, some species have identical 16S rRNA gene sequences, introducing complexity into taxonomic assignments (Tamura et al. 2008). As databases hinged almost exclusively on 16S rRNA gene analysis, the reliability of previous classifications reported in the literature could be questioned. This has led to the development of more representative genotyping methods such as multilocus sequence analysis (MLSA) which comprises the analysis of several housekeeping genes and thus significantly increases species resolution in comparison to 16S rRNA analysis. While MLSA has led to reclassifications (Rong et al. 2009), it did encounter inconsistencies due to the lack of a common MLSA scheme that would permit overall comparability, and it could not account for HGT biases (Glaeser and Kämpfer 2015).

Decreasing DNA sequencing costs and increasing computational power have in recent years allowed whole genome sequence analysis. This analysis is typically accomplished through the evaluation of the overall genome-related index (OGRI), with the most commonly used parameters being the Average Nucleotide Identity (ANI) and digital DNA-DNA Hybridization (dDDH) (Chun et al. 2018). The latter offers great convenience due to its clear and objective threshold of 70% as species boundary while species delineation thresholds for ANI are species dependent (Tindall et al. 2010). By including as much genetic information as possible in phylogenetic analyses, a higher resolution is obtained, consequently leading to further reclassification of *Streptomyces* species. Due to their increased power, these whole genome sequence methods are significantly impacting taxonomy today (Zuo et al. 2018). The lack of uniformity in phylogenetic analysis for classification of *Streptomyces* strains, with for example 16S rRNA analysis or MLSA still occasionally happening without whole genome sequence analysis (Sáez-Nieto et al. 2021; Kurnijasanti and Sudjarwo 2022; Salehghamari et al. 2023; Wang et al. 2023a), complicates taxonomy and introduces uncertainties in public repositories today. Notably, culture collections such as the ATCC heavily depend on the classification accuracy of the associated publications, making misclassifications impactful on studies focused on culture collection data. To illustrate, only 22 *Streptomyces* species are reported in the ATCC Genome Portal, while there are 3311 *Streptomyces* species available in their culture collection. In combination with the vastly expanding available genomic information, this might prove to be a recipe for disaster as current misclassifications will be incorporated in related studies and new phylogenetic analyses. The issue of misclassification of *Streptomyces* strains within current databases is substantiated by the review of Komaki (2023) as over the past decade, approximately 34 *Streptomyces* species have undergone reclassification and were transferred to other genera (Komaki 2023). Additionally, 14 subspecies were reclassified to only four subspecies. Furthermore, a total of 63 species were reclassified as later heterotypic synonyms. Also, many unclassified *Streptomyces* genomes have been deposited in the NCBI database. To date, little effort has been made to investigate whether they are new species. This leads to loss of information that could be beneficial to for example natural product discovery, but also complicates future classification. This paper therefore assesses the current taxonomy of known

Streptomyces strains as well as attempts to classify unclassified *Streptomyces* species.

Material and methods

Dataset collection

On June 26, 2023, an Excel file was created using a custom-made Python script that retrieved *Streptomyces* metadata from both NCBI Genome (now NCBI Datasets) and LPSN (List of Prokaryotic names with Standing in Nomenclature) databases (Supplementary Table S1). The script used a comprehensive list of all species obtained from the LPSN database and then selectively filtered out the *Streptomyces* species. Additional information retrieved from the LPSN database, such as LPSN status and LPSN address, is documented in columns K to O within Supplementary Table S1. Following this, we examined the availability of genomes for these species on NCBI. Atypical genomes, which include those identified as contaminated, and genomes of subspecies were excluded (based on NCBI filters). Often, multiple genome assemblies were available from NCBI per species. To determine the best assembly for our taxonomical analysis, the initial step involved verifying the presence of an NCBI reference genome. In case no reference genome was present, the available genome assemblies were ranked based on their assembly level (i.e., degree of completeness, categorized as complete genomes, chromosomes, scaffolds, or contigs). If several options for the optimal assembly remained available, preference was given to the most recent. If RefSeq assemblies were available, we selected these over GenBank assemblies. Additional genome and assembly information, including accession number and sequence length, is available in columns C to J within Supplementary Table S1. In total, 604 *Streptomyces* species and their metadata were collected. Out of these, four were manually not retained: three unclassified *Streptomyces* species catalogued as classified and one *Kitasatospora* species, bringing the total number of *Streptomyces* species for data analysis to 600 (Supplementary Table S2). For these species, whole genome sequence FASTA files, coding sequence (CDS) FASTA files with annotated genes, and RNA FASTA files were downloaded from the NCBI genome FTP site using their assembly ID with a custom-made Python script.

Taxonomic analysis of characterized *Streptomyces* species

For 16S rRNA analysis, a Python script was used to extract these sequences from the previously acquired RNA FASTA files, specifically targeting and isolating the 16S rRNA sequences within the RNA sequence. There were six genomes (GCA_003719395.1, GCA_016464675.1, GCA_000715685.1, GCA_018114805.1, GCA_000715605.1, GCA_000715635.1) for which RNA FASTA files were unavailable (Supplementary Table S2). Only sequences with a 95% completeness were considered, with a length ranging between 1451 and 1604 nucleotides. Sequences that were partial and had low completeness scores lack the necessary resolving power, leading to incorrect identification results, were excluded, further reducing the 16S rRNA dataset with 111 species to 483 (Supplementary Table S2). Furthermore, in cases where multiple copies of the 16S rRNA sequence were available, the longest 16S rRNA sequence was selected. ClustalO (Sievers et al. 2011) with default settings and the –percent-id parameter was used to generate a multiple sequence alignment (MSA) for computing the identity matrix among the retained 16S rRNA sequences (Supplementary Table S3). Species with an identity exceeding 99.93% were considered to be part of the same species (Komaki 2022).

MLSA was conducted using five housekeeping genes specifically proposed for the demarcation of *Streptomyces* species: *gyrB*, *atpD*, *recA*, *trpB*, and *rpoB* (Komaki 2022). These gene sequences were extracted from the coding sequence (CDS) FASTA files, based on the annotation headers. However, similar to the RNA FASTA files, the CDS FASTA files were not available in the same six cases (Supplementary Table S2). Only sequences that were not annotated as partial were considered. Among these genes, several were missing or only partially present: *atpD* in six species, *gyrB* in ten species, *recA* in 72 species, *trpB* in nine species, and *rpoB* in ten (Supplementary Table S2). Initially, each housekeeping gene except *atpD* was individually assessed and an MSA and identity matrix was constructed for each gene using ClustalO with the same settings as the 16S rRNA analysis (Supplementary Tables S4 to S7). The thresholds for species delineation were set at 99% (*gyrB*), 99.5% (*recA*), 99.6% (*rpoB*), and 97.9% (*trpB*) (Komaki 2022). Due to insufficient resolution, a species threshold could not be established for *atpD*. Subsequently, the sequences of the remaining genes were concatenated together to form a single comprehensive sequence. Only 502 species that possessed all five housekeeping genes were taken into consideration (Supplementary Table S2). Again, ClustalO was used without the –percent-id parameter for the construction of an MSA and subsequently creating a distance matrix (Supplementary Table S8). The threshold used for the MLSA was 0.007, where species with an evolutionary distance less than 0.007 belong to the same species (Komaki 2022).

ANI values were calculated on the whole genome FASTA sequences for the original dataset of 600 species, with all species being compared pairwise. FastANI v1.33 was the preferred ANI algorithm being 50–4608× faster than Average Nucleotide Identity based on BLAST (ANIb), without compromising accuracy within the 80 to 100% ANI identity

range (Jain et al. 2018). A matrix file was generated as output by FastANI and was subsequently imported into Excel (Supplementary Table S9). To delineate species, a threshold of 96.7% was used (Hu et al. 2022). *Streptomyces caldifontis* (GCA_016464675.1) exhibited values around the lower exclusion threshold of 80%, leading to multiple "Not Applicable" (NA) results in the ANI analysis.

In total, 409 *Streptomyces* species for which FastANI, 16S rRNA, and MLSA results were available, underwent pairwise analysis (Supplementary Table S2). Exceeded thresholds for each method were listed to indicate species belonging to the same taxonomic group (Supplementary Tables S10 to S17). While at first 16S rRNA was included, its resolution quickly proved insufficient (see "Results" and "Discussion"), thus warranting its exclusion from further analysis. Additionally, a digital DNA-DNA hybridization was conducted with the Genome-to-Genome Distance Calculator (Meier-Kolthoff et al. 2022). This dDDH analysis was only performed for species exceeding the threshold for FastANI and/or MLSA due to their labor-intensive nature (Supplementary Tables S18, S19).

Strain reclassification

When the FastANI values exceeded their threshold, the classification of the investigated strains was analyzed in detail and a reclassification based on the Bacterial Code was proposed. To decide between which name had to be taken when two species were identified as synonyms, the guidelines of the International Code of Nomenclature of Prokaryotes (the Bacterial Code) were used (Oren et al. 2023).

Classification of uncharacterized *Streptomyces* species

Uncharacterized *Streptomyces* strains were acquired using the command-line tool NCBI datasets on October 19, 2023. Data from these strains were gathered at three hierarchical levels: complete, chromosome, and scaffold. For these levels, there were respectively 27, 204, and 493 genomes gathered (Supplementary Table S20). At these three levels, a FastANI analysis was performed, exclusively comparing uncharacterized species with the 600 previously acquired ones. The threshold value remained at 96.7%. Uncharacterized *Streptomyces* species that surpassed this threshold and therefore indicating a match with a known species can be found in Supplementary Table S21. When no hit with a characterized species was found, the strain was transferred to the TYGS server for further identification attempts, encompassing known species beyond the *Streptomyces* genus (Supplementary Table S22) (Meier-Kolthoff et al. 2022).

Results

Species delineation method analysis

To ensure correct classification of uncharacterized *Streptomyces* species, the most appropriate and convenient classification method was investigated. Hence, the *Streptomyces* genus was first analyzed by 16S rRNA, individual housekeeping genes, MLSA and FastANI analysis for the 409 *Streptomyces* species for which all necessary information was available (Supplementary Tables S3 to S9). As evidenced, several thresholds for these parameters were exceeded during pairwise comparison of species, suggesting that both strains are actually the same species. Such synonyms were uncovered in 60 instances for 16S rRNA analysis, 56 instances for *gyrB*, 78 instances for *trpB*, 32 instances for *recA* and 43 instances for *rpoB* housekeeping gene analysis, 38 instances for MLSA, and 44 instances for FastANI analysis (Supplementary Tables S10 to S17). The results revealed a substantial discrepancy between 16S rRNA, MLSA, and FastANI. In 24 out of 60 cases, the threshold was exclusively exceeded for 16S rRNA and not by any other method. This suggested that the 16S rRNA gene did not provide sufficient resolution. As a result, it was excluded from further analysis in this study. However, this outcome was expected, given that the 16S rRNA gene is composed of approximately 1526 nucleotides and a single nucleotide variation would still result in a 99.9345% match (Fig. 1). This value surpasses the threshold of species delineation (99.93%) and has been observed in multiple occasions, as were 100% matches. Similar to the 16S rRNA gene, individual housekeeping genes make it challenging to delineate species as they represent only a snapshot of the genomes.

Due to the inconsistency in results obtained with single gene taxonomy parameters like 16S rRNA and single housekeeping genes, a more in-depth analysis of MLSA and FastANI results was done to pinpoint the most appropriate method. In total, scores for both approaches were calculated for 502 species. For pairwise comparisons where the threshold was exceeded by at least one of the approaches, an additional dDDH analysis was conducted (Supplementary Table S18). Between MLSA and FastANI, only 20 differences in exceeding the threshold were observed (Supplementary Table S19). The dDDH analysis yielded 14 differences with MLSA and 6 differences with FastANI, suggesting that FastANI is the best method to use. This result was unsurprising as both FastANI and dDDH utilize the entire genome for comparison whereas MLSA relies solely on a concatenated sequence of five single housekeeping genes. Differences between FastANI and dDDH are given in Table 1.
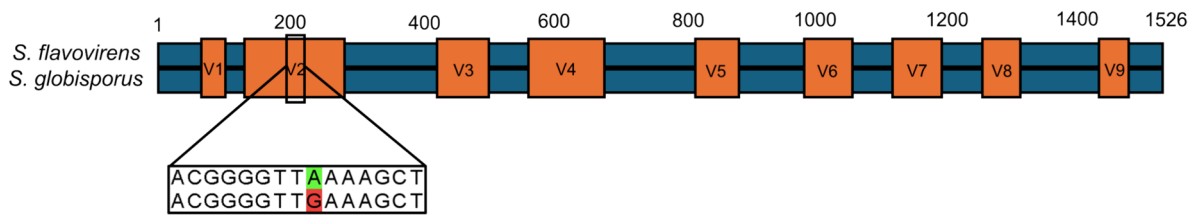
Fig. 1 Difference in 16S rRNA sequences between *Streptomyces globisporus* (GCF_003147545.1) and *Streptomyces flavovirens* (GCF_000720395.1). ClustalO alignment of both sequences identified a single mismatch at position 200 of 1526, represent-ing a 99.9345% similarity which was above the threshold for species demarcation. Conversely, the FastANI value for these species is 84.67%, considerably below the threshold. The figure also denotes the hypervariable regions of the 16S rRNA gene

Taxonomic analysis of previously characterized *Streptomyces* species

As illustrated above, FastANI has been validated as the preferred method for demarcation on species level regarding the *Streptomyces* genus. Its ability to compare whole genomes in a high-throughput manner allowed us to investigate *Streptomyces* species currently delineated in public repositories like LPSN and NCBI datasets prior to classifying unknown species in these databases. With a threshold of 96.7% for *Streptomyces* species (Hu et al. 2022), 59 out of 600 analyzed species did not seem to be unique (Fig. 2). This adjustment refines our dataset of 600 *Streptomyces* species back to 541 unique species. An overview of the pairwise alignments exceeding the species delineation threshold, suggesting these species are actually synonyms, is given in Table 2, together with the proposed reclassification based on the guidelines of the International Code of Nomenclature of Prokaryotes (the Bacterial Code) (Oren et al. 2023). While several of these findings have been previously reported by various authors, supporting our results, many new synonyms have been uncovered in our study. We thus propose to amend these errors in public repositories by classifying these species as synonyms to stop the taxonomy sprawl of *Streptomyces* species, leading to further misclassification and complexities.

Classification of uncharacterized *Streptomyces* species

As FastANI appeared to be the most reliable and appropriate method for *Streptomyces* species demarcation, this method was exclusively used for the classification of uncharacterized *Streptomyces* species. At the chromosome level, nine out of the 27 species were assigned to one of the 600 characterized *Streptomyces* species in our previously acquired dataset, surpassing the 96.7% threshold value (Supplementary Table S21). Among these nine species, six were identified to match with *S. mirabilis*, while the three remaining uncharacterized *Streptomyces* species (GCF_900104935.1, GCF_003352805.1, GCF_001040905.1) exhibited matches with three distinct species at chromosome level. At the complete level, we successfully classified 99 out of the 204 species with in total 62 already characterized *Streptomyces* species. Notably, several uncharacterized strains were found to match with the same known *Streptomyces* species, for instance, seven were identified as *S. albidoflavus*. At the scaffold level, 181 out of 493 were matched with 77

Table 1 Discrepancies between the FastANI and dDDH pairwise alignment of a subset of *Streptomyces* strains. Values exceeding the parameter threshold are highlighted in bold and point to edge cases close to the threshold. Organisms are considered to belong to the same species above thresholds of 96.7% for FastANI and 70% for dDDH

| Strain 1	Strain 2	FastANI (%)	dDDH (%)
Streptomyces albiflaviniger GCA_016103485.1	*Streptomyces asiaticus* GCF_018138715.1	95.99	**71.7**
Streptomyces albiflaviniger GCA_016103485.1	*Streptomyces cangkringensis* GCF_019059395.1	96.00	**71.7**
Streptomyces albiflaviniger GCA_016103485.1	*Streptomyces indonesiensis* GCF_018138705.1	95.99	**71.8**
Streptomyces albiflaviniger GCA_016103485.1	*Streptomyces rhizosphaericus* GCF_017942185.1	95.76	**71.6**
Streptomyces anthocyanicus GCF_014650795.1	*Streptomyces rubrogriseus* GCF_003112595.1	**96.76**	69.8
Streptomyces barkulensis GCF_002843305.1	*Streptomyces pini* GCF_900114215.1	**96.75**	67.3
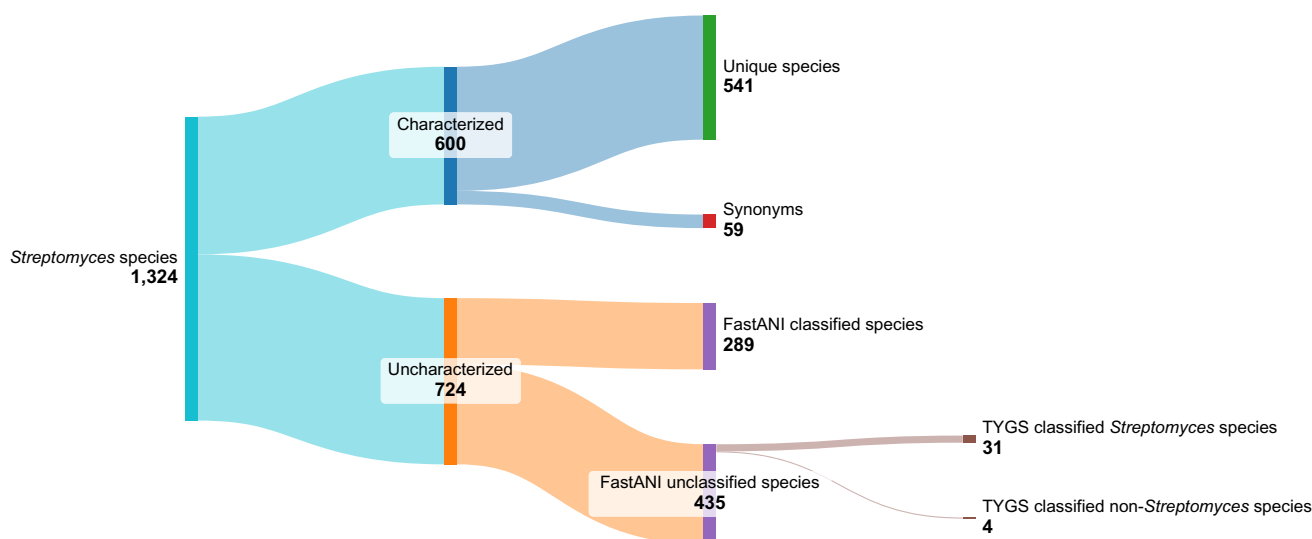
Fig. 2 Overview of the *Streptomyces* species investigated in this study. The genomes of 600 *Streptomyces* species on public repositories (characterized) appear to be not as unique as advocated, with almost 10% synonyms and 541 unique strains. Of the 724 unclassified *Streptomyces* species (uncharacterized), 289 could be classified with FastANI, while TYGS proposed classification of 35 strains and the uniqueness of 435 new species

characterized *Streptomyces* species, implicating that again several uncharacterized species were assigned to the same known *Streptomyces* species. Hence, in total, 289 uncharacterized species were classified as 106 already known species. Regarding the remaining species that could not be classified solely based on the *Streptomyces* genomes available in our dataset, we attempted to identify the species with the aid of the TYGS server (Supplementary Table S22). We were able to classify an additional 35 species, with 31 belonging to the *Streptomyces* genus and 4 belonging to other genera: *Microbacterium resistens*, *Kitasatospora cheerisanensis*, *Embleya scabrispora*, and *Cytobacillus praedii* (Fig. 2). The remaining ones were indicated as "potential new species" by TYGS.

Discussion

Method discussion

Due to missing RNA and CDS FASTA files, housekeeping genes, and insufficient completeness of 16S rRNA, taxonomic assessment based on 16S rRNA, MLSA, and FastANI was only performed for 409 out of 600 *Streptomyces* genomes available on NCBI Datasets (Supplementary Table S2). Based on this study, methods relying on single gene taxonomy parameters such as 16S rRNA and housekeeping genes did not provide sufficient resolution and showed divergent results. To conduct 16S rRNA taxonomic analysis, it is essential to have the RNA FASTA files accessible on NCBI Datasets with a satisfactory level of

completeness ($>95\%$) to avoid insufficient resolving power (Kim et al. 2012b). These conditions were not met for 117 out of the 600 available *Streptomyces* genomes because the 16S rRNA sequence was either absent or incomplete, i.e., falling short of the required length ($<95\%$) or exceeding it ($>105\%$). The repetitive nature of specific regions within the 16S rRNA gene poses a challenge for short-read sequencing technologies in achieving robust assemblies, often leading to incomplete 16S rRNA gene sequences. In that regard, long-read methods hold a distinct advantage over short-read methods. These technologies can capture the entire length of the 16S rRNA gene, in contrast to being limited to hypervariable regions alone. For example, in Fig. 1, the only mismatch between *S. globisporus* and *S. flavovirens* was located in the hypervariable V2 region. While the majority of 16S rRNA phylogenetic studies rely on the V3–V4 region (Klindworth et al. 2013), the V3–V5 has proven to perform poorly for the phylum *Actinobacteria* (Meier-Kolthoff et al. 2013) which can lead to biased results. Notably, a rigorous threshold of 99.93% was applied in the assessment of the 16S rRNA analysis, surpassing the threshold commonly found in literature, which typically ranges between 98.2 and 99% (Meier-Kolthoff et al. 2013). Despite this high threshold value, the threshold was exceeded for 60 out of 409 instances of pairwise comparisons in 16S rRNA whereas for FastANI this was for 44 instances (Supplementary Tables S10 to S12). The observation that in 40% (24/60) of the cases, only the threshold for 16S rRNA was surpassed and not for MLSA or whole genome–based methods highlights the low resolution of 16S rRNA and raises questions about the ongoing use of

Table 2 Suggested reclassification of *Streptomyces* strains obtained from NCBI Datasets. FastANI values are given for pairwise alignments exceeding the species demarcation threshold of 96.7%, suggesting these to be the same species. The suggested species name is highlighted in gray. For example, *S. sampsonii* and *S. wadaymensis* should be reclassified as *S. albidoflavus* according to our analysis. In case there was no reference genome used at the time of analysis, the species name is followed by *. Strains marked with † might possibly yield less reliable results. Please refer to the discussion section for further details. The taxonomic status indicates whether the suggested name has currently been approached according to the Bacterial Code, or whether the name has been validated. Additionally, references to the literature for the proposed reclassification are given

Strain 1	GenBank ID1	Strain 2	GenBank ID2	FastANI (%)	Taxonomic status	Reference
S. albidoflavus	GCF_000359525.1	*S. sampsonii**	GCF_001704195.1	98.79		
S. albidoflavus	GCF_000359525.1	*S. wadayamensis**	GCF_000698945.1	99.00	Approved	(Rong et al. 2009)
*S. sampsonii**	GCF_001704195.1	*S. wadayamensis**	GCF_000698945.1	98.74		
S. alboverticillatus	GCF_002150845.1	*S. triculaminicus*	GCF_017349075.1	99.61	Validated but unapproved	This study
S. albovinaceus	GCF_002154385.1	*S. globisporus**	GCF_003147545.1	99.22		
S. albovinaceus	GCF_002154385.1	*S. mediolani**	GCF_000721685.1	99.22	Approved	(Rong and Huang 2010; Kim et al. 2012a)
*S. mediolani**	GCF_000721685.1	*S. globisporus**	GCF_003147545.1	99.43		
S. pluricolorescens	GCF_014650395.1	*S. rubiginosohelvolus*	GCF_014649875.1	99.86	Approved	(Li et al. 2022)
*S. alboviridis**	GCF_000716935.1	*S. microflavus*	GCF_013364315.1	97.46		
*S. alboviridis**	GCF_000716935.1	*S. luridiscabiei**	GCF_001418625.1	97.26	Approved	(Rong and Huang 2010)
*S. luridiscabiei**	GCF_001418625.1	*S. microflavus*	GCF_013364315.1	99.04		
S. albus	GCF_024752535.1	*S. graminis**	GCF_000716525.1	99.57	Approved	(Labeda et al. 2014)
S. anthocyanicus	GCF_014650795.1	*S. coelicolor**	GCF_013317105.1†	99.35		
S. anthocyanicus	GCF_014650795.1	*S. rubrogriseus*	GCF_003112595.1	96.76	Approved	(Rong et al. 2009; Komaki 2022)
*S. coelicolor**	GCF_013317105.1†	*S. rubrogriseus*	GCF_003112595.1	96.81		
S. antimycoticus	GCF_005405925.1	*S. melanosporofaciens*	GCF_900105695.1	96.88	Approved	(Li et al. 2022)
S. ardesiacus	GCF_003112575.1†	*S. diastaticus**	GCF_010548605.1†	99.28	Approved	(Komaki and Tamura 2020)
S. asiaticus	GCF_018138715.1	*S. cangkringensis*	GCF_019059395.1	99.87		
S. asiaticus	GCF_018138715.1	*S. indonesiensis*	GCF_018138705.1	99.91		
S. asiaticus	GCF_018138715.1	*S. rhizosphaericus*	GCF_017942185.1	99.74	Validated but unapproved	This study
S. cangkringensis	GCF_019059395.1	*S. indonesiensis*	GCF_018138705.1	99.87		
S. cangkringensis	GCF_019059395.1	*S. rhizosphaericus*	GCF_017942185.1	99.74		
S. indonesiensis	GCF_018138705.1	*S. rhizosphaericus*	GCF_017942185.1	99.74		
*S. aureoverticillatus**	GCF_010669165.1	*S. lasiicapitis*	GCF_014646335.1	98.85	Approved	This study
*S. autolyticus**	GCF_001983975.1	*S. malaysiensis**	GCF_006974005.1	98.63		
*S. autolyticus**	GCF_001983975.1	*S. samsunensis*	GCF_024519315.1	97.67		
*S. autolyticus**	GCF_001983975.1	*S. solisilvae*	GCF_015689475.1	98.86		
*S. malaysiensis**	GCF_006974005.1	*S. samsunensis*	GCF_024519315.1	97.59	Not validated or approved	This study
*S. malaysiensis**	GCF_006974005.1	*S. solisilvae*	GCF_015689475.1	98.70		
S. samsunensis	GCF_024519315.1	*S. solisilvae*	GCF_015689475.1	97.62		
*S. baarnensis**	GCF_000716025.1	*S. griseus**	GCF_014655335.1	99.14	Approved	(Rong and Huang 2010; Kim et al. 2012a)
S. barkulensis	GCF_002843305.1	*S. pini*	GCF_900114215.1	96.75	Not validated or approved	This study
S. bellus	GCF_014649195.1	*S. coeruleorubidus*	GCF_008705135.1	97.99	Approved	This study
S. bobili	GCF_026343715.1	*S. galilaeus*	GCF_008704575.1	97.56	Approved	This study
S. brasiliscabiei	GCF_018927715.1	*S. griseiscabiei*	GCF_020010925.1	98.99	Validated but unapproved	This study
S. californicus	GCF_016906185.1	*S. purpeochromogenes**	GCF_000717965.1	99.97		
S. californicus	GCF_016906185.1	*S. violaceoruber*	GCF_002082175.1	98.86	Approved	(Komaki 2022)
*S. purpeochromogenes**	GCF_000717965.1	*S. violaceoruber*	GCF_002082175.1	98.85		
S. canarius	GCF_014650735.1	*S. olivaceoviridis*	GCF_014650115.1	97.30		
S. corchorusii	GCF_001514055.1	*S. olivaceoviridis*	GCF_014650115.1	96.71	Approved	(Li et al. 2022)
*S. capuensis**	GCF_000718295.1	*S. rimosus**	GCF_008704655.1	99.95		
*S. capuensis**	GCF_000718295.1	*S. peucetius*	GCF_000725565.1	97.78	Approved	This study
S. peucetius	GCF_000725565.1	*S. rimosus**	GCF_008704655.1	97.83		
S. cellulosae	GCF_026339705.1†	*S. thermocarboxydus*	GCF_024760485.1†	99.02	Approved	This study
S. daghestanicus	GCF_020521275.1	*S. niveoruber*	GCF_014649135.1	99.46	Approved	(Li et al. 2022)
S. durhamensis	GCF_000725475.1	*S. filipinensis*	GCF_014649495.1	96.76	Approved	(Li et al. 2022)
S. eurythermus	GCF_014649115.1	*S. lavenduligriseus*	GCF_000718625.1	96.91	Approved	This study
S. flaveolus	GCF_014648815.1	*S. solaniscabiei*	GCF_020312215.1	97.21	Approved	This study
S. flavovariabilis	GCF_000725785.1	*S. variegatus*	GCF_000955965.1	99.99	Validated but unapproved	This study
*S. fradiae**	GCF_023376055.1	*S. rubrolavendulae**	GCF_001750785.1	99.41	Approved	This study
*S. geysiriensis**	GCF_019641555.1	*S. plicatus*	GCF_014650135.1	98.58		
*S. geysiriensis**	GCF_019641555.1	*S. rochei*	GCF_020881015.1	98.58		
*S. geysiriensis**	GCF_019641555.1	*S. vinaceusdrappus*	GCF_014650215.1	98.55		
S. plicatus	GCF_014650135.1	*S. rochei*	GCF_020881015.1	99.06	Approved	(Komaki 2021)
S. plicatus	GCF_014650135.1	*S. vinaceusdrappus*	GCF_014650215.1	99.30		
S. rochei	GCF_020881015.1	*S. vinaceusdrappus*	GCF_014650215.1	99.13		
S. ginkgonis	GCF_024761905.1	*S. harbinensis*	GCF_013364095.1	99.25	Not validated or approved	This study
S. glebosus	GCF_009811575.1	*S. platensis*	GCF_008704855.1	97.65	Approved	This study
S. gougerotii	GCF_014648955.1	*S. rutgersensis*	GCF_014216335.1	99.63	Approved	This study
*S. griseoaurantiacus**	GCF_013913525.1	*S. jietaisiensis*	GCF_900102095.1	99.05	Approved	(Li et al. 2022)
S. griseofuscus	GCF_014673495.1	*S. murinus*	GCF_016467295.1	96.89	Approved	(Madhaiyan et al. 2020)
*S. griseoincarnatus**	GCF_023702435.1	*S. variabilis*	GCF_014649935.1	98.26	Approved	(Komaki 2021)
S. griseomycini	GCF_014649595.1	*S. griseostramineus†*	GCF_014649635.1	99.75	Approved	(Wang et al. 2023b)
S. janthinus	GCF_014649675.1	*S. violaceus**	GCA_000715635.1	97.24		
S. janthinus	GCF_014649675.1	*S. violarus*	GCF_014650255.1	96.95	Approved	(Li et al. 2022)
S. jumonjinensis	GCF_009600885.1	*S. katsurahamanus*	GCF_009600895.1	97.97	Not validated or approved	This study
*S. libani**	GCF_026344195.1	*S. nigrescens*	GCF_027626975.1	99.12	Approved	(Komaki and Tamura 2020)
S. luteocolor	GCF_001748305.1	*S. silaceus*	GCF_001419745.1	99.51	Not validated or approved	This study
S. nashvillensis	GCF_014650095.1†	*S. tanashiensis*	GCF_026427415.1	99.72	Approved	(Li et al. 2022)
S. noursei	GCF_016917755.1	*S. yunnanensis*	GCF_029010865.1	97.26	Approved	This study
S. novaecaesareae	GCF_000716875.1	*S. pyridomyceticus**	GCF_000717725.1	99.96	Approved	This study
S. plumbidurans	GCF_019857225.1	*S. spinosirectus**	GCF_021441825.1	99.87	Not validated or approved	This study
*S. pseudogriseolus**	GCF_020312235.1	*S. rubiginosus*	GCF_014649895.1	97.00	Approved	(Li et al. 2022)
*S. silvensis**	GCF_001482415.1	*S. typhae*	GCF_009755605.1	98.66	Not validated or approved	This study

16S rRNA. Therefore, we excluded 16S rRNA taxonomic analysis from further analysis.

For the single housekeeping gene analysis and MLSA, a key issue was the lack of these genes in the *Streptomyces* genomes or their partial availability. Remarkably, the *recA* gene was in 72 out of the 600 cases only partially present or even lacking (Supplementary Table S2). This could be due to incomplete assembly or annotation mistakes, as *recA* is a core gene encoding recombinase A which has been described as indispensable for viability (Muth et al. 1997). In total, 502 *Streptomyces* genomes contained all five housekeeping genes (Supplementary Table S2). The gene *atpD* was not included in our analysis, as recent literature proved its insufficient resolution for species delineation (Komaki 2022). In contrast, whole genome–based methods are more robust towards the abovementioned issues as for example they are not influenced by annotation mistakes; however, they can also suffer from poor assembly. In 38 out of 409 cases, the taxonomic threshold was exceeded for MLSA (Supplementary Table S17), while this was 44 times with FastANI analysis (Supplementary Table S11). As these numbers were close to each other, we performed for these cases a dDDH analysis to emphasize a more pronounced distinction between gene-based (MLSA) and whole genome–based methods (FastANI and dDDH). As a result, dDDH analysis showed 14 differences in exceeding the threshold with MLSA whereas there were only six differences between dDDH and FastANI results (Supplementary Table S19). Hence, this indicated consistency regarding the whole genome–based methods. Table 1 shows that four of these six discrepancies include *Streptomyces albiflaviniger* (GCA_016103485.1) while the other two are close to the threshold (69.8% for dDDH and 96.75% for FastANI). Notably, the genome of *Streptomyces barkulensis* (GCF_002843305.1) consisted of 119 scaffolds while that of *S. albiflaviniger* (GCA_016103485.1) consisted of 5265 contigs. Both had a relatively low coverage (23× and 9.9×, respectively). While the quality of the genomes can have an effect on species delineation, these discrepancies can also be the result of the type of ANI tool used (i.e., ANIb, OrthoANI, ANIm, FastANI, and gANI). Palmer and colleagues have compared 7 ANI methods and noted slight differences in outcome (Palmer et al. 2020). Their conclusions also support our use of FastANI as it is the fastest and thus the most suitable approach to investigate large numbers of pairwise genome comparisons. However, as evidenced by the results in Table 1, for analyses near the inferred ANI threshold of 96.7% for *Streptomyces* sp., the ANI results should be carefully considered. Such findings should be examined more closely by for example comparing multiple ANI methods against dDDH and/or analyzing phenotypic and chemotaxonomic characteristics for additional insights. Also, species delineation thresholds for ANI are species dependent, and we feel confident that the threshold value used in our assessment is correct as it is often used and recognized for *Streptomyces* species (Hu et al. 2022; Wang et al. 2023b). Analyzing ANI values of over 90,000 bacterial and archaeal genomes has set a species threshold at ≥ 95% (Murray et al. 2021), meaning that strains with such a score are synonyms. However, in the case of *Streptomyces* strains in particular, an ANI score of ≥ 96.7% has been suggested as a more stringent species demarcation (Hu et al. 2022). In literature, threshold values ranging from 95 to 96% are frequently employed for *Streptomyces* taxonomic analyses (Jain et al. 2018; Park et al. 2021; Nikolaidis et al. 2023; Song et al. 2023).

Despite these couple differences between both methods, FastANI has proven to be a reliable and convenient method for the classification of *Streptomyces* species. In addition, due to the labor-intensive character of dDDH, as there is a limit on the number of simultaneous analyses and due to the higher amount of computational power required, FastANI is the preferred method. Indeed, taxonomic analysis in the past solely relied on single gene taxonomy parameters as whole genome sequencing was very expensive and overvalued at that time; however, today there is a remarkable trend in the decrease of next- and third-generation sequencing cost. Therefore, taxonomic analysis based on whole genomes should become the golden standard in order to get rid of the misclassification challenge (Komaki 2023).

Emended classification of previously characterized *Streptomyces* species

To provide clarity in the classification of currently uncharacterized *Streptomyces* species, we first investigated the existing taxonomy of this genus. While we used a stringent threshold for species delineation with FastANI, this study uncovered multiple synonyms for the first time. With a FastANI threshold of 96.7%, 59 out of 600 analyzed *Streptomyces* species did not seem to be unique according to our analysis. Table 2 provides a summary of our suggestions for reclassification. We utilized genomic data primarily comprised of reference genomes whenever available. However, certain species lacked a reference genome. In such cases, the genomes of these species may contain errors or contamination. For example, between downloading the data and writing this manuscript, we observed an alteration in NCBI datasets: the *S. coelicolor* genome used in the analysis, GCF_013317105.1, is now flagged as contaminated. To ensure accurate conclusions, we reran FastANI comparing *S. rubrogriseus* and *S. anthocyanicus* with *S. coelicolor* GCF_000203835.1. Nonetheless, the FastANI values remained unchanged, indicating both strains are later heterotypic synonyms of *S. coelicolor*. Furthermore *S. cellulosae* GCF_026339705.1 and *S. thermocarboxydus* GCF_024760485.1 are now also marked as

contaminated. Unfortunately, as there are no uncontaminated genomes available for *S. thermocarboxydus*, a reanalysis for this comparison was impossible. Additionally, *S. diastaticus* GCF 010548605.1 is now marked as "unverified source organism," only *S. diastaticus* subsp. *diastaticus* GCF_014648935.1 available as uncontaminated genome. Reanalyzing the species demarcation with *S. ardesiacus* yielded a FastANI value of only 81.74%, suggesting indeed a potential discrepancy with the original genome of *S. diastaticus* GCF 010548605.1. In the meantime, *S. griseostramineus* GCF_014649635.1 and *S. nashvillensis* GCF_014650095.1 underwent a reclassification on NCBI, respectively, to *S. griseomycini* and *S. tanashiensis*, which is in line with our FastANI analysis. These changes happened over the course of several months, proving the importance of depositing high-quality data and the curation of public repositories.

As another remark, we expected that *S. canarius* and *S. corchorusii* would also be identified as synonyms as *S. canarius* and *S. olivaceoviridis* are synonyms as well as *S. olivaceoviridis* and *S. corchorusii*. However, this was not the case (FastANI value of 96.4%). Similarly, *S. violaceus* and *S. violarus* were expected to be synonyms but the FastANI value was 96.5%.

The number of newly discovered synonyms is quite surprising as many other studies on *Streptomyces* reclassification have been published in recent years. In general, they align perfectly with this study, although some differences are present. Rong et al. (2009) investigated the taxonomy of the *S. albidoflavus* clade based on MLSA with the concatenated sequences of the *aptD*, *gyrB*, *recA*, *rpoB*, and *trpB* housekeeping genes and with DNA-DNA Hybridization (DDH), and proposed the reclassification of *S. sampsonii* and *S. coelicolor* to *S. albidoflavus* (Rong et al. 2009). However, we did not observe the same result for *S. coelicolor* in our FastANI approach. The identification of *S. wadayamensis* as a later synonym of *S. albidoflavus* was not reported to date and is unique to this study. In 2010, Rong and Huang proposed several other reclassifications based on MLSA with *atpD*, *gyrB*, *recA*, *rpoB*, and *trpB* as well as DDH analysis. They suggested that *S. baarnensis* is a synonym of *S. fimicarius* while our data proposes to classify *S. baarnensis* within the *S. griseus* clade (Rong and Huang 2010). Various reclassifications of *S. baarnensis* have been proposed since 2010, also resulting in the classification of this strain as *S. griseus*, which is in line with our findings. The lack of transparency on these alterations and publications associated with these changes is another example of the complexity of *Streptomyces* taxonomy. The authors also proposed to classify *S. californicus* as a synonym of *S. puniceus* while our FastANI analysis is the first to suggest that *S. californicus*, *S. purpeochromogenes*, and *S. violaceoruber* are all synonyms. This result might hold major implications as Komaki (2022) proposed that

S. anthocyanicus is a synonym of *S. violaceoruber* while we did not come to the same conclusion. Additionally, Rong and Huang suggested reclassifying *S. mediolani* as a later synonym of *S. albovinaceus* while no association with *S. globisporus* had been made. Finally, their proposed reclassification of *S. alboviridis* and *S. luridiscabiei* as synonyms of *S. microflavus* is in line with our findings.

In 2012, Kim et al. (2012a) suggested that *S. fimicarius* belongs to the *S. setonii* clade based on 16S rRNA and *gyrB* analysis while *S. setonii* is now a heterotypic synonym of *S. griseus*. They also uncovered that *S. albovinaceus* is a later synonym of *S. globisporus*. In 2020, Madhaiyan and colleagues suggested that *S. griseofuscus* should be renamed to *S. murinus* based on genome-based analyses with dDDH, ANI, and Average Amino Acid Identity (AAI), which is in line with our results (Madhaiyan et al. 2020). The same year, Komaki and Tamura (2020) proposed to classify *S. diastaticus* subsp. *ardesiacus* as an independent species (*S. ardesiacus* sp. nov.) based on 16S rRNA and dDDH analysis; however, our FastANI results suggest that *S. diastaticus* and *S. ardesiacus* are synonyms. The next year, Komaki (2021) proposed that *S. plicatus*, *S. geysiriensis*, and *S. vinaceusdrappus* are later heterotypic synonyms of *S. rochei*, that *S. variabilis* is a later synonym of *S. griseoincarnatus*, and that *S. libani* subsp. *libani* is the same species as *S. nigrescens* based on MLSA and dDDH analysis. These findings are in line with our own results.

An extensive study by Li and colleagues in 2022 proposed various reclassifications based on a genome-based study with ANI and dDDH (Li et al. 2022). Most of these findings were in line with our own results, such as the proposed reclassification of *S. canarius* and *S. olivaceoviridis* as synonyms of *S. corchorusii*, *S. durhamensis* as a synonym of *S. filipinensis*, *S. jietaisiensis* as a synonym of *S. griseoaurantiacus*, *S. janthinus* and *S. violaceus* as a synonym of *S. violarus*, *S. nashvillensis* as a synonym of *S. tanashiensis*, *S. rubiginosus* as a synonym of *S. pseudogriseolus*, *S. pluricolorescens* as a synonym of *S. rubiginosohelvolus*, and *S. melanosporofaciens* as a synonym of *S. antimycoticus*. However, the observed similarity between *S. niveoruber* and *S. griseoviridis* diverged from our own conclusions. It was however noted that *S. niveoruber* is a synonym of *S. daghestanicus*, which was published 1 year prior to *S. niveoruber*, warranting the use of *S. daghestanicus* as species name. Finally, Wang et al. (2023b) indicated that *S. griseomycini* and *S. griseostramineus* are the same species, with the latter being the later heterotypic synonym of the former, which is also in line with our findings.

As evidenced by these proposed reclassifications, it is urgent time to clean up the taxonomy of the *Streptomyces* genus. Several of these synonyms were published a while ago and have not been amended in public repositories, leading to erroneous classification of new strains and

overcomplicating taxonomy. Furthermore, as our assessment was focused exclusively on *Streptomyces* species with available genomes on NCBI, we anticipate a need for additional reclassifications. The LPSN database lists 869 child taxa under the species *Streptomyces*, yet our FastANI analysis covered only a subset of 600 species whose genomes were publicly available.

Classification of uncharacterized *Streptomyces* species

Based on our findings, we used FastANI to try and classify publicly available uncharacterized *Streptomyces* species sourced from NCBI Datasets. Despite our stringent threshold, we were able to classify 289 uncharacterized species out of 724 (Supplementary Table S21) (Fig. 2). Notably, some of these uncharacterized species were not unique. For instance, 20 were classified as *S. albidoflavus*, 13 as *S. parvus*, 11 as *S. mirabilis*, and 10 as *S. rochei*. In total, the number of uncharacterized species that can be classified amounts to 106 characterized species. To further assess unclassified species, we first used the TYGS server which searches for synonyms within and beyond the *Streptomyces* genus (Supplementary Table S22). Out of the remaining 435 uncharacterized strains, 35 were associated with existing species (Fig. 2). Interestingly, four of these were not found to be *Streptomyces* species at all: *M. resistens*, *K. cheerisanensis*, *E. scabrispora*, and *C. praedii*. While *K. cheerisanensis* and *E. scabrispora* belong to the family *Streptomycetaceae*, *M. resistant* belongs to the class *Actinomycetia*. On the other hand, *C. praedii* belongs to *Bacillota* phylum, which differs from *Streptomyces* species who are classified under the phylum *Actinomycetota*. These misclassifications evidence the need for improved curation and classification of genomes deposited on public repositories.

The discrepancy between TYGS and our FastANI results, leading to the remaining 31 species classified by TYGS, was due to the absence of certain genomes in our own dataset. *S. koelreuteriae* and *S. chengmaiensis* were added after our data retrieval from NCBI Datasets, and *S. regensis* was marked as contaminated. In other cases, the genome match from TYGS was not found on NCBI datasets (i.e., *S. chrysomallus*, *S. cinnamonensis*, *S. sporocinereus*, *S. fluorescens*), or a different strain isolate was used on TYGS (*S. scabiei*, *S. griseus*, *S. anulatus*, *S. xinghaiensis*). Comparison of various isolates of these strains or loosening the stringent FastANI threshold a little could bridge the discrepancies between TYGS and FastANI. To further classify uncharacterized *Streptomyces* strains, a viable strategy might involve initially filtering out all unique species. Various whole genome sequence techniques like dDDH and AAI might help uncover new synonyms. While ANI was specifically developed to delineate species, AAI is more powerful in associating strains that are more distantly related. Once a more suitable genus has been identified for uncharacterized, unique strains, a FastANI analysis of that genus can be done analogously to our study.

Finally, our findings emphasize the need for action, urging adjustments in the NCBI database and thorough reexamination of species within the *Streptomyces* genus. In the meantime, dubious entries should be flagged in public repositories and more stringent data deposition conditions should be enforced. Improving future classifications could be achieved by implementing more comprehensive curation in these databases, thereby avoiding the submission of incomplete data.

Declarations

References

Alam K, Mazumder A, Sikdar S, Zhao YM, Hao J, Song C, Wang Y, Sarkar R, Islam S, Zhang Y, Li A (2022) Streptomyces: the

biofactory of secondary metabolites. Front Microbiol 13:968053. https://doi.org/10.3389/FMICB.2022.968053

Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, Rooney AP, Yi H, Xu XW, De Meyer S, Trujillo ME (2018) Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol 68:461–466. https://doi.org/10.1099/ijsem.0.002516

Del Carratore F, Hanko EK, Breitling R, Takano E (2022) Biotechnological application of *Streptomyces* for the production of clinical drugs and other bioactive molecules. Curr Opin Biotechnol 77:102762. https://doi.org/10.1016/J.COPBIO.2022.102762

Glaeser SP, Kämpfer P (2015) Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. Syst Appl Microbiol 38:237–245. https://doi.org/10.1016/J.SYAPM.2015.03.007

Hassler HB, Probert B, Moore C, Lawson E, Jackson RW, Russell BT, Richards VP (2022) Phylogenies of the 16S rRNA gene and its hypervariable regions lack concordance with core genome phylogenies. Microbiome 10:104. https://doi.org/10.1186/s40168-022-01295-y

Hu S, Li K, Zhang Y, Wang Y, Fu L, Xiao Y, Tang X, Gao J (2022) New insights into the threshold values of multi-locus sequence analysis, average nucleotide identity and digital DNA–DNA hybridization in delineating *Streptomyces* species. Front Microbiol 13:910277. https://doi.org/10.3389/fmicb.2022.910277

Hugenholtz P, Chuvochina M, Oren A, Parks DH, Soo RM (2021) Prokaryotic taxonomy and nomenclature in the age of big sequence data. ISME J 15(7):1879–1892. https://doi.org/10.1038/s41396-021-00941-x. (15)

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9

Kim K-O, Shin K-S, Kim MN, Shin K-S, Labeda DP, Han J-H, Kim SB (2012a) Reassessment of the status of *Streptomyces setonii* and reclassification of *Streptomyces fimicarius* as a later synonym of *Streptomyces setonii* and *Streptomyces albovinaceus* as a later synonym of *Streptomyces globisporus* based on combined 16S rRNA/*gyrB* gene sequence analysis. Int J Syst Evol Microbiol 62:2978–2985. https://doi.org/10.1099/ijs.0.040287-0

Kim O-S, Cho Y-J, Lee K, Yoon S-H, Kim M, Na H, Park S-C, Jeon YS, Lee J-H, Yi H, Won S, Chun J (2012b) Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int J Syst Evol Microbiol 62:716–721. https://doi.org/10.1099/ijs.0.038075-0

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41:e1. https://doi.org/10.1093/nar/gks808

Komaki H (2021) Reclassification of 15 *Streptomyces* species as synonyms of *Streptomyces albogriseolus*, *Streptomyces althioticus*, *Streptomyces anthocyanicus*, *Streptomyces calvus*, *Streptomyces griseoincarnatus*, *Streptomyces mutabilis*, *Streptomyces pilosus* or *Streptomyces rochei*. Int J Syst Evol Microbiol 71:004718. https://doi.org/10.1099/ijsem.0.004718

Komaki H (2022) Resolution of housekeeping gene sequences used in MLSA for the genus *Streptomyces* and reclassification of *Streptomyces anthocyanicus* and *Streptomyces tricolor* as heterotypic synonyms of *Streptomyces violaceoruber*. Int J Syst Evol Microbiol 72:005370. https://doi.org/10.1099/IJSEM.0.005370

Komaki H (2023) Recent progress of reclassification of the genus *Streptomyces*. Microorganisms 11:831. https://doi.org/10.3390/microorganisms11040831

Komaki H, Tamura T (2020) Reclassification of *Streptomyces diastaticus* subsp. *ardesiacus*, *Streptomyces gougerotii* and *Streptomyces*

rutgersensis. Int J Syst Evol Microbiol 70:4291–4297. https://doi.org/10.1099/ijsem.0.004287

Kurnijasanti R, Sudjarwo SA (2022) Identification of Streptomyces spp. from garbage dump soils in Surabaya, Indonesia. Vet World 15:634. https://doi.org/10.14202/VETWORLD.2022.634-639

Labeda DP, Doroghazi JR, Ju K-S, Metcalf WW (2014) Taxonomic evaluation of *Streptomyces albus* and related species using multilocus sequence analysis and proposals to emend the description of *Streptomyces albus* and describe *Streptomyces pathocidini* sp. nov. Int J Syst Evol Microbiol 64:894–900. https://doi.org/10.1099/ijs.0.058107-0

Lacey HJ, Rutledge PJ (2022) Recently discovered secondary metabolites from *Streptomyces* species. Molecules 27:887. https://doi.org/10.3390/MOLECULES27030887

Li C-J, Wang M, Zhang Z, Zhi X-Y, Yang L-L (2022) Genome-based analyses reveal heterotypic synonyms of *Streptomyces* species and associated subspecies. Arch Microbiol 204:581. https://doi.org/10.1007/s00203-022-03204-1

Madhaiyan M, Saravanan VS, See-Too W-S (2020) Genome-based analyses reveal the presence of 12 heterotypic synonyms in the genus *Streptomyces* and emended descriptions of *Streptomyces bottropensis*, *Streptomyces celluloflavus*, *Streptomyces fulvissimus*, *Streptomyces glaucescens*, *Streptomyces murinus*, and *Streptomyces variegatus*. Int J Syst Evol Microbiol 70:3924–3929. https://doi.org/10.1099/ijsem.0.004217

Meier-Kolthoff JP, Göker M, Spröer C, Klenk H-P (2013) When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 195:413–418. https://doi.org/10.1007/s00203-013-0888-4

Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M (2022) TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res 50:D801–D807. https://doi.org/10.1093/nar/gkab902

Murray CS, Gao Y, Wu M (2021) Re-evaluating the evidence for a universal genetic boundary among microbial species. Nat Commun 12:4059. https://doi.org/10.1038/S41467-021-24128-2

Muth G, Frese D, Kleber A, Wohlleben W (1997) Mutational analysis of the *Streptomyces lividans recA* gene suggests that only mutants with residual activity remain viable. Mol Gen Genet 255:420–428. https://doi.org/10.1007/s004380050514

Nealson KH, Conrad PG (1999) Life: past, present and future. Philos Trans R Soc Lond B Biol Sci 354:1923–1939. https://doi.org/10.1098/RSTB.1999.0532

Nikolaidis M, Hesketh A, Frangou N, Mossialos D, Van de Peer Y, Oliver SG, Amoutzias GD (2023) A panoramic view of the genomic landscape of the genus *Streptomyces*. Microb Genom 9:001028. https://doi.org/10.1099/mgen.0.001028

O'Connell LM, Blouin T, Soule A, Burne RA, Nascimento MM, Richards VP (2022) Optimization and evaluation of the 30S–S11 rRNA gene for taxonomic profiling of oral streptococci. Appl Environ Microbiol 88:e0045322. https://doi.org/10.1128/AEM.00453-22

Oren A, Arahal DR, Göker M, Moore ERB, Rossello-Mora R, Sutcliffe IC (2023) International Code of Nomenclature of Prokaryotes. Prokaryotic Code (2022 Revision). Int J Syst Evol Microbiol 73:005585. https://doi.org/10.1099/ijsem.0.005585

Palmer M, Steenkamp ET, Blom J, Hedlund BP, Venter SN (2020) All ANIs are not created equal: implications for prokaryotic species boundaries and integration of ANIs into polyphasic taxonomy. Int J Syst Evol Microbiol 70:2937–2948. https://doi.org/10.1099/ijsem.0.004124

Park CJ, Caimi NA, Buecher DC, Valdez EW, Northup DE, Andam CP (2021) Unexpected genomic, biosynthetic and species diversity of *Streptomyces* bacteria from bats in Arizona and New

Mexico, USA. BMC Genomics 22:247. https://doi.org/10.1186/s12864-021-07546-w

Preiner M, Asche S, Becker S, Betts HC, Boniface A, Camprubi E, Chandru K, Erastova V, Garg SG, Khawaja N, Kostyrka G, Machné R, Moggioli G, Muchowska KB, Neukirchen S, Peter B, Pichlhöfer E, Radványi Á, Rossetto D, Salditt A, Schmelling NM, Sousa FL, Tria FDK, Vörös D, Xavier JC (2020) The future of origin of life research: bridging decades-old divisions. Life 10:20. https://doi.org/10.3390/LIFE10030020

Rong X, Huang Y (2010) Taxonomic evaluation of the *Streptomyces griseus* clade using multilocus sequence analysis and DNA-DNA hybridization, with proposal to combine 29 species and three subspecies as 11 genomic species. Int J Syst Evol Microbiol 60:696–703. https://doi.org/10.1099/ijs.0.012419-0

Rong X, Guo Y, Huang Y (2009) Proposal to reclassify the *Streptomyces albidoflavus* clade on the basis of multilocus sequence analysis and DNA-DNA hybridization, and taxonomic elucidation of *Streptomyces griseus* subsp. *solvifaciens*. Syst Appl Microbiol 32:314–322. https://doi.org/10.1016/j.syapm.2009.05.003

Sáez-Nieto JA, Carrasco G, del Pino S, Medina-Pascual MJ, Garrido N, Villalón P, Valdezate S (2021) Identification and antimicrobial susceptibility of *Streptomyces* and other unusual Actinobacteria clinical isolates in Spain. New Microbes New Infect 44:100946. https://doi.org/10.1016/J.NMNI.2021.100946

Salehghamari E, Moradi M, Sardabi M, Etesami S, Hassani G, Hosseini M, Taheri F, Eshrati Z, Tahmaseb M, Irian S, Amoozegar M (2023) Diversity of actinomycete and their metabolites isolated from Howz Soltan Lake. Iran Arch Microbiol 205:1–13. https://doi.org/10.1007/S00203-022-03364-0/TABLES/2

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75

Song L, Wang F, Liu C, Guan Z, Wang M, Zhong R, Xi H, Zhao Y, Wen C (2023) Isolation and evaluation of *Streptomyces melanogenes* YBS22 with potential application for biocontrol of rice blast disease. Microorganisms 11:2988. https://doi.org/10.3390/microorganisms11122988

Tamura T, Ishida Y, Otoguro M, Hatano K, Labeda D, Price NP, Suzuki K-I (2008) Reclassification of *Streptomyces caeruleus* as a synonym of *Actinoalloteichus cyanogriseus* and reclassification of *Streptomyces spheroides* and *Streptomyces laceyi* as later synonyms of *Streptomyces niveus*. Int J Syst Evol Microbiol 58:2812–2814. https://doi.org/10.1099/ijs.0.65560-0

Tindall BJ, Rosselló-Móra R, Busse H-J, Ludwig W, Kämpfer P (2010) Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol 60:249–266. https://doi.org/10.1099/ijs.0.016949-0

Wang J, Wang R, Gao J, Liu H, Tang W, Liu Z, Li X (2023a) Identification of three *Streptomyces* strains and their antifungal activity against the rubber anthracnose fungus *Colletotrichum siamense*. J Gen Plant Pathol 89:67–76. https://doi.org/10.1007/S10327-022-01109-X

Wang Y, Long P, Fu L, Xiao Y, Wang N, Gao J (2023b) Reclassification of *Streptomyces griseostramineus* as a synonym of *Streptomyces griseomycini* based on a polyphasic taxonomic approach. Int J Syst Evol Microbiol 73:006021. https://doi.org/10.1099/ijsem.0.006021

Zuo G, Qi J, Hao B (2018) Polyphyly in 16S rRNA-based LVTree versus monophyly in whole-genome-based CVTree. Genom Proteomics Bioinforma 16:310–319. https://doi.org/10.1016/J.GPB.2018.06.005