1

Characterization and Engineering of *Streptomyces griseofuscus* DSM 40191 as a Potential Host for Heterologous Expression of Biosynthetic Gene Clusters

Tetiana Gren[a], Christopher M. Whitford[a], Omkar S. Mohite[a], Tue S. Jørgensen[a], Eftychia E. Kontou[a], Julie B. Nielsen[a], Sang Yup Lee[a,b], Tilmann Weber[a#]

[a]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Denmark

[b]Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering, Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology, Republic of Korea

[#] correspondence address:
Prof. Dr. Tilmann Weber
The Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark
Kemitorvet, bygning 220
2800 Kgs. Lyngby
Denmark
Email: tiwe@biosustain.dtu.dk
Phone: +45 24 89 61 32

Running Head: Characterization of *Streptomyces griseofuscus* DSM 40191

Word content of the Abstract and Importance section is 350 words.

Word content of the main text is 5694 words

Abstract

Streptomyces griseofuscus DSM 40191 is a fast growing *Streptomyces* strain that remains largely underexplored as a heterologous host. Here, we report the genome mining of *S. griseofuscus*, followed by the detailed exploration of its phenotype, including production of native secondary metabolites and ability to utilise carbon, nitrogen, sulphur and phosphorus sources. Furthermore, several routes for genetic engineering of *S. griseofuscus* were explored, including use of GusA-based vectors, CRISPR-Cas9 and CRISPR-cBEST-mediated knockouts. Using CRISPR-BEST technology, core genes of 4 biosynthetic gene clusters (BGCs) that are situated on the chromosome arms were inactivated and the outcomes of the inactivations were tested. Two out of the three native plasmids were cured using CRISPR-Cas9 technology, leading to the generation of strain *S. griseofuscus* DEL1. DEL1 was further modified by full deletion of a pentamycin BGC and an unknown NRPS BGC, leading to the generation of strain DEL2, lacking approx. 500 kbp of the genome, which corresponds to a 5,19% genome reduction. Sequencing confirmed that DEL2 does not bear any crucial off-target effects or rearrangements in its genome. It can be characterized by faster growth and inability to produce three main native metabolites of *S. griseofuscus*: lankacidin, lankamycin, pentamycin and their derivatives. To test the ability of DEL2 to heterologously produce secondary metabolites, the actinorhodin BGC was used. We were able to confirm the production of actinorhodin by both *S. griseofuscus* wild type and DEL2. We believe that this strain will serve as a good chassis for heterologous expression of BGCs.

Importance

The rise of antibacterial resistance calls on the development of the next generation of antibiotics, majority of which are derived from natural compounds, produced by actinomycetes. The manipulation, refactoring and expression of BGCs coding for such natural products is a promising approach in secondary metabolite discovery. Thus, the development of a versatile panel of heterologous hosts for the expression of BGCs is essential. We believe that first-to-date systematic, detailed characterisation of *S. griseofuscus*, a highly promising chassis strain, will not only facilitate the further development of this particular strain, but also will set a blueprint for characterisation of other potential hosts.

Introduction

The vast majority of clinically used antibiotics, antifungals, anticancer and immunosuppressive drugs are derived from natural products (1), originating from soil-inhabiting Actinobacteria (2). Genes, necessary for the production of a particular compound, are typically arranged in biosynthetic gene clusters (BGCs) that can be detected using genome mining tools like antiSMASH (3). With the revolution in whole genome sequencing it became clear that the majority of actinobacterial strains possess genomes of large sizes, with 20 to 40 of BGCs on average (4). However, typically these bacteria only produce few compounds under laboratory conditions (5). Other antibiotics are not produced or produced only in small amounts and require induction by various, frequently unknown, environmental cues (5, 6). In order to unravel the biosynthetic richness of Actinobacteria, multiple genetic engineering methods, e.g. induction of BGCs by co-expression of regulatory genes (7, 8), use of transcription factor decoys (9), refactoring of the BGCs (10, 11) or knockouts of the core genes of "constitutively" produced compounds (12), have been explored. Recently, broader attention has been paid to development of heterologous expression hosts for discovery and characterization of novel metabolites. Heterologous hosts are typically well-characterized strains, that possess plurality of needed characteristics, e.g. fast and disperse growth, amenability to genetic engineering, high yields of produced secondary metabolites (11, 13, 14). The use of such chassis strains is highly advantageous, when the native producer strain can not be genetically engineered or is slow growing (13). At the moment, several of such *Streptomyces* hosts are available, such as derivatives of *S. coelicolor*, *S. lividans*, *S. avermitilis, S. chattanoogensis, S. albus, S. venezuelae* (15–20). Recently, several groups have reported the generation of *S. albus* J1074 derivatives that harbor knockouts of multiple native clusters (21, 22) and were successfully used for expression of BGCs (23, 24). The reported success rate of BGCs expression, however, remains low, reaching 30% (14). We believe that the further extension of the panel of heterologous hosts will be highly beneficial to solve this challenge.

One of the promising potential heterologous hosts, *S. griseofuscus,* was first isolated in Japan and used as a host for isolation and propagation of bacteriophages and plasmids (25–28). Some of the strains were used for industrial production of ε-poly-L-lysine (29–31) and puromycin (32), and reported as a source of novel compounds (9, 33). Even though several publications have stipulated multiple positive characteristics *of S. griseofuscus*, e.g. fast growth, ease of

88 transformation and genetic manipulation, it was never methodically studied regarding its

89 qualities as a potential platform for expression of BGCs.

90 In this paper we describe a comprehensive genotypic and phenotypic characterization of *S.*

91 *griseofuscus* DSM40191 as a potential heterologous production host. The *De novo* sequenced

92 genome of *S. griseofuscus* was used for mining of natively encoded BGCs. For the first time,

93 we have explored use of various genetic tools for engineering of *S. griseofuscus*, ranging from

94 the use of integrative and replicative vectors, CRISPR-Cas9 and CRISPR-BEST based

95 knockouts to multiplexed-targeting of several BGCs using one CRISPR-BEST construct. All

96 of the generated knockouts were tested regarding their growth and production of metabolites.

97

Results

1. Genome mining and comparative analysis

100 *Streptomyces griseofuscus* DSM 40191 (=NRRL B-5429) was received from the German

101 Collection of Microorganisms and Cell Cultures. We re-sequenced and assembled the genome

102 of *S. griseofuscus de novo* (34). It consists of a linear 8,721,740 bp chromosome and three

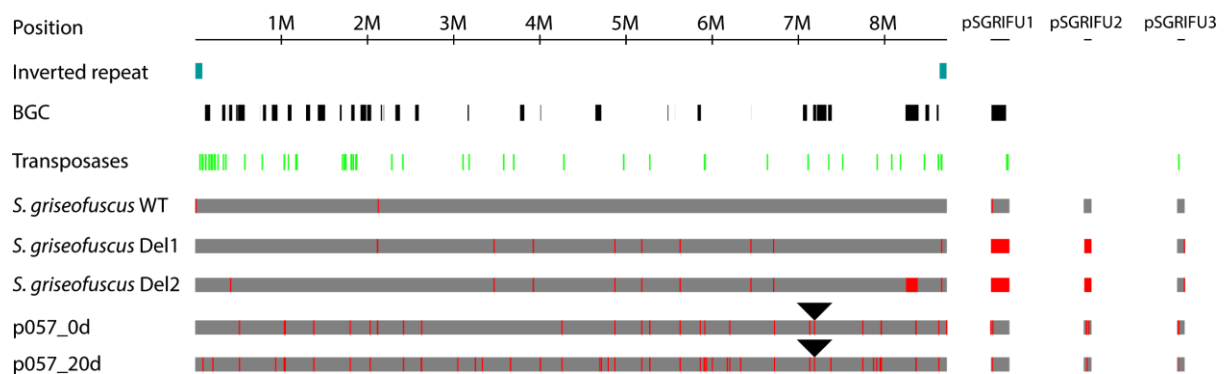103 plasmids: pSGRIFU1 (220kb), pSGRIFU2 (88kb), and pSGRIFU3 (86kb) (Fig. 1).

104

107 **Fig. 1. Overview of *S. griseofuscus* strains and mutations.** The positions of genomic inverted repeats are

108 highlighted in blue, BGCs in black and transposases in green. The positions of identified mutations are highlighted

109 in red. Mutations detected in the wild type Illumina dataset can be considered technical noise. Position of the

110 CRISPR-cBEST introduced STOP-codon is indicated with a black triangle. Strains p057_0D and p057_20D relate

111 to the long term cultivation experiment, in which CRISPR-cBEST generated strain *S. griseofuscus* IHEP81_06602

112 (p057_0D), that contains an introduced STOP-codon in BGC 30, was transferred 20 consecutive times in liquid

113 ISP2 media without selective pressure, thus generating strain p057_20D.

114

Whole genome comparison between *S. griseofuscus* DSM40191, *S. coelicolor* A3(2), *S. venezuelae* ATCC 10712

Here, we analyzed how much genomic content is shared between *S. griseofuscus* DSM40191 and other well-studied model *Streptomyces* strains *S. coelicolor* A3(2) and *S. venezuelae* ATCC 10712. Their genomes were downloaded from NCBI (accession IDs NC_003888 and NZ_CP029197) and compared to *S. griseofuscus* DSM 40191 by calculating bidirectional best blastp hits between the genes. We found that 3918 genes were shared across all three genomes (core genome), whereas *S. griseofuscus* shared additional 937 and 522 genes with *S. coelicolor* and *S. venezuelae* respectively (Fig. 2A, Dataset 1). The net total of genes present across the three strains (pangenome set) was 13415. In order to understand the biological functions of the shared genes, we annotated all three genomes with KEGG biological subsystems (35). For the genome of *S. griseofuscus*, we found 2842 ortholog genes in KEGG, which are involved in 1498 KEGG reactions. Whereas the genomes of *S coelicolor* and *S. venezuelae* contained 8152 and 7112 genes, which map to 2900 and 2670 KEGG gene IDs, and 1475 and 1452 KEGG reaction IDs, respectively (Dataset 1). After comparing KEGG genes and reactions content, we found that the three genomes shared 1488 common KEGG gene IDs and 1280 reactions. Next, we compared the distribution of number of genes and number reactions belonging to different KEGG pathways across three organisms (Fig. 2C & D). We found that the number of genes involved in subsystems such as membrane transport, signaling and cellular processes were significantly lower in *S. griseofuscus* than others. Whereas, the number of genes involved in the subsystems like metabolism of terpenoids and polyketides, genetic information processing, amino acid metabolism, xenobiotics biodegradation and metabolism were significantly larger in *S. griseofuscus* (Fig. 2C). Comparison of number of reactions across different metabolic pathways showed that the three strains shared very similar metabolism. However, the number of reactions involved in pathways such as amino acid metabolism was still larger in *S. griseofuscus* (Fig. 2D). Overall, we found that *S. griseofuscus* and *S. coelicolor* had larger genomic and metabolic content than *S. venezuelae*. Additionally, higher genomic and metabolic contents were shared between *S. griseofuscus* and *S. coelicolor*. In the later section, we compared the phenotype microarray data of these strains to get experimental understanding of their metabolic growth capabilities.
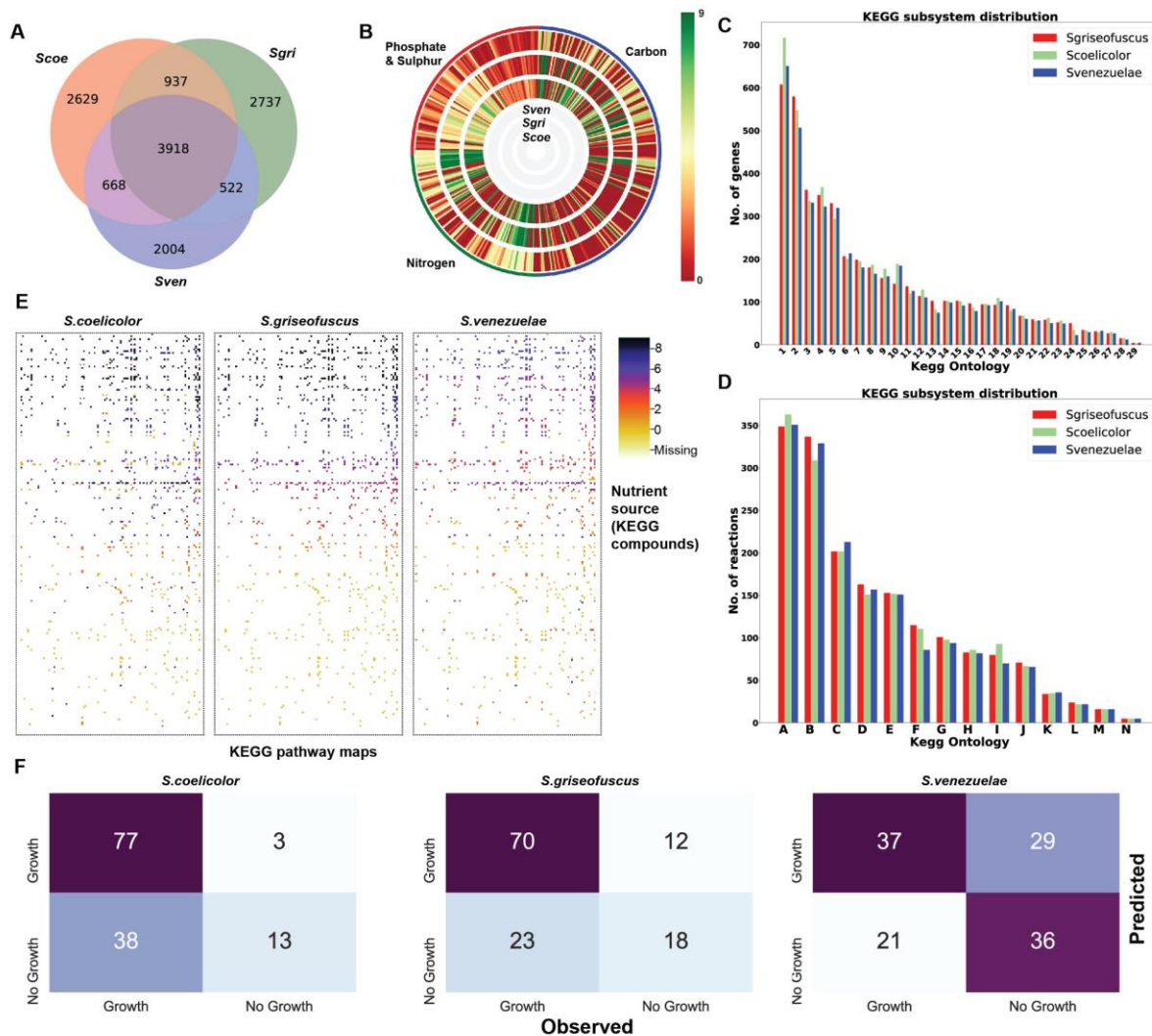
Fig. 2. Genome and phenotype microarray comparison of three strains. A. Number of genes shared among the three organisms - *S. griseofuscus* (sgri), *S. coelicolor* (scoe), *S. venezuelae* (sven). B. Phenotype microarray data represented by activity index rings generated using DuctApe across 4 biolog plates consisting 379 carbon, nitrogen, phosphate and sulphate nutrient sources across three organisms. C. Distribution of number of genes per KEGG subsystem across three organisms. For details on this dataset, please see Dataset 1. D. Distribution of number of reactions per KEGG metabolic pathway across three organisms. A- Carbohydrate metabolism; B- Amino acid metabolism; C- Metabolism of cofactors and vitamins, D- Lipid metabolism, E- Nucleotide metabolism, F- Metabolism of terpenoids and polyketides, G - Energy metabolism, H- Biosynthesis of other secondary metabolites, I- Xenobiotics biodegradation and metabolism, J- Metabolism of other amino acids, K- Glycan biosynthesis and metabolism, L- Translation, M- Not included in regular maps, N- Signal transduction.E. Heatmap with activity index of different KEGG nutrients (y-axis) against KEGG pathway maps (x-axis). F. Confusion matrix representing genome-scale model based prediction and observed growth phenotypes of three organisms.

2. Phenotype characterization

Growth and sporulation on different media

To systematically study a particular strain, it is important to choose an appropriate media that will support fast growth and sporulation in liquid and solid cultures. Therefore, we tested several solid media and have chosen MS as the most accommodating for sporulation (Fig. S1, Table S1). To confirm disperse growth in liquid media, optical density at 600 nm (OD_{600}) based growth curve was built in liquid ISP2 (Fig. S2, Table S1).

Comparison of physiological features of *S. griseofuscus*, *S. coelicolor* and *S. venezuelae* using Biolog microarrays

In order to characterize the phenotype of *S. griseofuscus* and its ability to utilize different substrates, we have conducted a multiple parallel cultivation using BioLog microarrays. This technology is not easily applicable for studying actinobacterial strains due to the formation of "clumps" of mycelia. However, in the case of *S. griseofuscus*, its uncomplicated growth enables such studies. As a direct comparison, we have used well studied heterologous hosts *S. coelicolor* and *S. venezuelae*. Previously, parallel micro-scale cultivations were used for characterization of an industrially important *S. lividans* TK24 (36).

We tested a total of 379 substrates, including 190 different carbon sources (PM1 and PM2), 95 nitrogen sources (PM3), 94 phosphate and sulphur sources (PM4). The kinetic growth data from Biolog was analyzed together with the genomes using DuctApe software (37) that correlated genomic and phenomic data based on KEGG metabolic pathways. An activity index between 0 to 9 was used to represent the growth on each substrate, where activity index higher than 3 was used as a cutoff to define growth. We found that 171, 172 and 117 of the 379 substrates were utilized by *S. griseofuscus, S. coelicolor* and *S. venezuelae*, respectively (Dataset 2). Comparing the growth of three different strains, we found that 90 substrates were commonly utilized by all three strains, whereas 14, 19 and 7 substrates were utilized uniquely by *S. griseofuscus, S. coelicolor* and *S. venezuelae*, respectively. Some of the substrates uniquely utilized by *S. griseofuscus* include ethanolamine, 2-aminoethanol, cytidine, thymidine, D-serine and D-threonine. Additionally, we found that *S. griseofuscus* shared a total of 145 common growth substrates with *S. coelicolor*, signalling higher mutual metabolic similarity.

190   Next, we analyzed growth on substrates with different nutrient source categories. We observed

191   that a total of 72 carbon sources were utilized by *S. coelicolor*, which was higher than the

192   number of carbon sources used by *S. griseofuscus* (64) and *S. venezuelae* (61). In particular, *S.*

193   *coelicolor* could utilize more substrates involved in carbohydrate metabolism. Carbon sources

194   uniquely utilized by *S. griseofuscus* include 2-aminoethanol, alpha-keto-valeric acid, D-malic

195   acid. On the contrary, the number of nitrogen sources utilized were higher in *S. griseofuscus*

196   (60) as compared to *S. coelicolor* (52) and *S. venezuelae* (49). These could be primarily

197   attributed to the categories amino acid metabolism and other non-defined classes of

198   metabolism. Unique nitrogen sources utilized by *S. griseofuscus* include L-phenylalanine, D-

199   serine, ethanolamine. We found that *S. venezuelae* could only use 6 of the phosphate sources

200   which was substantially lower than both *S. griseofuscus* (47) and *S. coelicolor* (46). Uniquely

201   utilized phosphate sources by *S. griseofuscus* included 2-aminoethyl phosphonic acid and

202   dithiophosphate. In general, we observe that the capability of *S. griseofuscus* to utilize different

203   nutrient source categories is much higher than *S. venezuelae,* and is similar or even higher than

204   *S. coelicolor*. Comparison of the ability of *S. griseofuscus* to grow on different nutrient sources

205   can guide the design of growth media and, thus, leads to optimal growth and metabolite

206   production.

207   To investigate the connection between these growth activity profiles and the genomic diversity

208   of the strains, a matrix was generated using the *dape* module of DuctApe, where the activity

209   on different nutrients (rows) that are part of different KEGG pathways (columns) is highlighted

210   (Fig. 2E, Dataset 3). For example, the average growth activity indices of all the nitrogen source

211   nutrients belonging to KEGG pathway of biosynthesis of amino acids (map:01230) were 6.51,

212   5.96 and 4.18 for *S. griseofuscus, S. coelicolor* and *S. venezuelae*. In particular, thiamine

213   metabolism pathway (map:00730) showed higher average growth index on nitrogen nutrients

214   in *S. griseofuscus* (7.11) as compared to *S. coelicolor* (5.89) and *S. griseofuscus* (5.78). Overall,

215   the growth activity heatmaps of nutrients vs KEGG pathways were more similar in *S.*

216   *griseofuscus* and *S. coelicolor*. Whereas, *S. venezuelae* was found to have lower growth activity

217   across nutrients from different pathways. The higher genomic similarity between *S.*

218   *grisoefuscus* and *S. coelicolor* that was observed in the previous section further corroborates

219   with this phenomic similarity. In addition to this genome to phenome comparison based on

220   KEGG pathways, we used genome-scale metabolic models to compare the *in-silico* predicted

221   growth against observed phenotypes across different substrates (Fig. 2F). We reconstructed

222   draft genome-scale metabolic models for *S. griseofuscus* and *S. venezuelae* based on homology

223   comparison against the genome scale model of *S. coelicolor* (Dataset 4). The models also

224 predicted growth on more number of nutrients in case of *S. griseofuscus* and *S. coelicolor* as

225 compared to *S. venezuelae* (Dataset 5). Thus, we conclude that *S. griseofuscus* possesses very

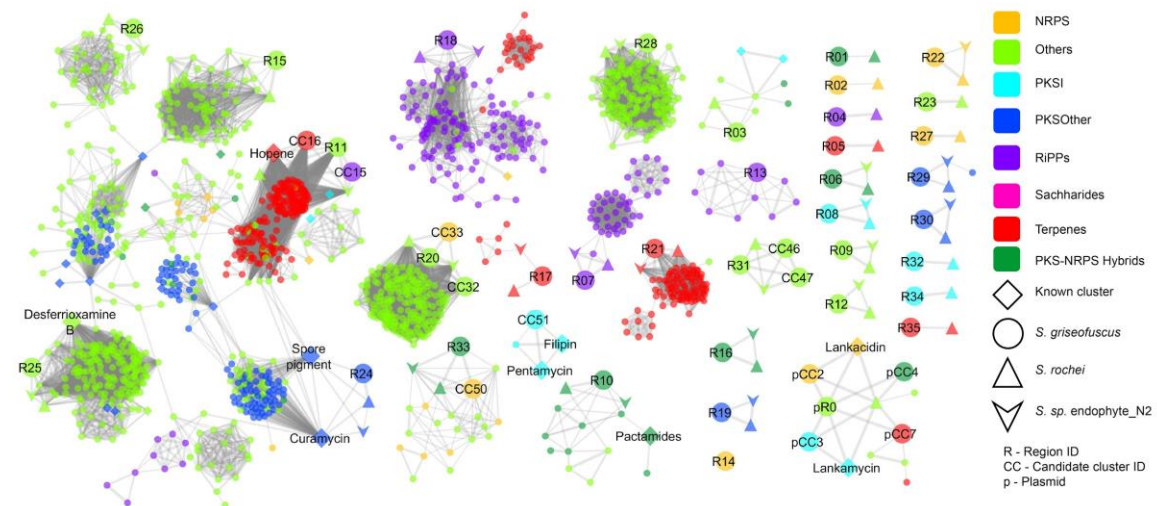226 similar or even superior, metabolic capabilities while compared to well studied *Streptomyces*

227 strains.

228

229

230

231 **Fig. 3. Similarity network of BGCs in *S.griseofuscus* against 212 public genomes and MIBIG database of**

232 **known BGCs**. Comparison of all the regions and few selected candidate clusters against BGCs of known

233 compounds from MIBIG database and 212 public genomes. Different colors denote different types of BGCs as

234 shown in legends. BGCs from *S. griseofuscus, S. rochei* 7434AN4*, Streptomyces. sp. endophyte_N2* and MIBIG

235 database are shown with different shapes and sizes. All regions, selected candidate clusters and known BGCs are

236 annotated by text. Detailed comparison of selected BGCs using CORASON can be found in Fig. S3.

237

3. Secondary metabolite potential of *Streptomyces griseofuscus*

Analysis of the genome using antiSMASH and BiG-SCAPE

240 In order to estimate the capabilities of the strain to synthesize secondary metabolites, it is

241 important to characterize the BGCs present in the genome. We therefore carried out a genome

242 mining analysis using antiSMASH (3). We detected 35 regions of BGCs encoding for different

243 types of secondary metabolites on the chromosome. These regions can be split into 53

244 candidate clusters. The megaplasmid pSGRIFU1 (CP051007) includes one antiSMASH-

245 predicted region with seven candidate clusters. No BGCs were detected on pSGRIFU2 and

246 pSGRIFU3. We observed that the genome of *S. griseofuscus* harbored 4 NRPS, 3 PKSI, 5 PKS-

247 NRPS hybrids, 4 other PKS types, 4 terpenes, 4 RiPPs and 11 other types of BGCs as defined

248 by antiSMASH. Some of these BGCs putatively coded for hopene, geosmin, spore pigment,

9

249   desferrioxamine B, ectoine and pentamycin. Two of the candidate clusters from the plasmid

250   showed similarity to known BGC encoding lankamycin and lankacidin C. We have manually

251   selected some of the candidate clusters for further analysis.

252   In order to investigate if the BGCs in *S. griseofuscus* are also detected in other *Streptomyces*

253   genomes, we carried out a BGC similarity analysis involving a dataset of 212 publically

254   available complete high-quality *Streptomyces* genomes. In total, 6380 BGCs of different types

255   were detected across this dataset of genomes. We generated a similarity network of 35 regions,

256   12 manually selected candidate clusters from *S. griseofuscus*, 6380 BGCs from public genomes

257   and 1808 known BGCs from MIBIG database (38) using BiG-SCAPE . The network with the

258   cutoff of 0.3 raw_distance metric was further analyzed using Cytoscape (Fig. 3). All BGC

259   families that did not include one of the BGCs from *S. griseofuscus* were ignored for the

260   subsequent analyses (Dataset 6). We found that only one BGC (region 14) of type NRPS-like

261   was a singleton in the network, uniquely observed in *S. griseofuscus*. We observed that 8 of the

262   BGCs were exclusively present in one other genome, namely *S. rochei* 7434AN4. In addition,

263   further 9 BGCs are also present in *Streptomyces. sp.* endophyte_N2 (Genbank Accn.:

264   CP028719) in addition to *S. rochei* 7434AN4 (Genbank Accn.: AP018517). This suggests that

265   these 17 BGCs from *S. grieseofuscus* are also very rarely observed across streptomycetes.

266   Among BGCs that are relatively common across the dataset, we found that candidate cluster

267   50 of region 33 had similar BGCs across 18 other *Streptomycetes,* including *S. collinus Tu 365*

268   (Fig. S3), whereas the candidate cluster 51 was similar to known cluster encoding for

269   pentamycin. Overall, we have established that most of the BGCs (33) were also present in *S.*

270   *rochei* 7434AN4, indicating two genomes with highly similar content.

271   This similarity has led us to examine the relation of *S. griseofuscus* to other strains of its species

272   and to *S. rochei* strains. Currently, there are 4 complete assemblies *S. griseofuscus* and 3 of *S.*

273   *rochei* genomes available in NCBI database. Among these are aforementioned *S. rochei*

274   7434AN4 and type strain *S. rochei* NRRL B-2410. By calculating pairwise Average Nucleotide

275   Identity (ANI) between all genomes (39), we have identified ANI between *S. griseofuscus*

276   DSM40191 and *S. rochei* 7434AN4 to be at 99.54%, while similarity of *S. rochei* 7434AN4 to

277   the type strain *S. rochei* NRRL B-2410 is at 84.03%, highly similar to the one between *S.*

278   *rochei* NRRL B-2410 and *S. griseofuscus* DSM 40191 of 83.95%. This clearly signals that *S.*

279   *rochei* 7434AN4 strain was probably misclassified and is indeed a *S. griseofuscus* strain. This

280   may explain the large number of similar BGCs, shared between *S. griseofuscus* DSM 40191

281   and *S. rochei* 7434AN4 and high level of similarity between two of the largest plasmids in both

282   strains pSGRIFU1 and pSLA2-L.

283

Characterization of secondary metabolites, produced by the *S. griseofuscus*

In the genome mining study, we have identified several known BGC, studied in other strains. Among these are lankacidin and lankamycin BGCs, encoded on plasmid pSGRIFU1, previously studied in *S. rochei* (40, 41) and pentamycin BGC33 (42). Due to the good preservation of these BGCs in the genome we expected to detect the production of lankamycin, lankacidin-family compounds and pentamycin. In addition, it was reported previously, that some strains of *S. griseofuscus* are able to produce azinomycins A and B (43), acetylcholine esterase inhibitor physostigmine (44), ϵ-poly-L-lysine (29, 30), lankacidin C and A (25). In order to check the production in *S. griseofuscus* DSM 40191, we have performed exploratory cultivation in 5 different liquid media (ISP2, MAM and CDMZ medium (44), minimal medium (MM) (45), and medium 65) (46), that were described in literature for the production of respective compounds and have attempted to identify them in extracts. Lankacidin A, C, and lankamycin were tentatively identified by HR-MS. The production of pentamycin was identified by HR-MS and confirmed using the pentamycin standard (Fig. S4). Physostigmine was not detected in any of the conditions (Table S2).

We have noted the production of hydrophobic extracellular vesicles by *S. griseofuscus*, a widely spread, but poorly studied phenomenon among Actinobacteria (47). It is known that the extracellular vesicles might contain secondary metabolites (48, 49). To study the profile of the extracellular vesicles in *S. griseofuscus*, they were collected and directly injected for LC-MS measurements. Among many compounds we have tentatively identified lankamycin , that was previously detected in the cultivation extractions and is likely released as a part of extracellular vesicles.

4. Development of genetic engineering methods

Even though the transformation, conjugation, protoplast generation for *S. griseofuscus* were established before, including attempts of genetic engineering (32, 33, 50) it was never systematically tested with different vectors and engineering methods. While generating a heterologous host strain, it is important to have access to the fastest knockout-leading techniques that lead to the least off-target modifications.

Transfer of integrative and replicative GusA-based vectors

As the first step we tested whether *S. griseofuscus* is compatible with *gusA* reporter system plasmids (51): pSETGUS, an integrative phiC31- based plasmid, and pKG1139, a replicative plasmid. Both plasmids were successfully conjugated into *S. griseofuscus* and allowed for visual screening of the exconjugant colonies (Fig. S5). To determine the position of the pSETGUS integration site, which is of importance to rationally utilize it for the integration of desirable elements, we randomly picked three independent *S. griseofuscus* pSETGUS colonies and sequenced them using Oxford Nanopore sequencing, similarly to (52). The exact location of the integration site is at 4,242,328 bp in the HEP81_03793 gene, coding for putative chromosome condensation protein. The determined *attB* site of *S. griseofuscus* contains the conserved core "TT" sequence (53).

CRISPR-Cas9 mediated gene knockout

CRISPR-Cas9-based molecular tools offer precision and ease in handling in comparison to other techniques. Over the recent years CRISPR tools have been adapted for use in streptomycetes (54). As the introduction of double strand breaks can lead to rearrangements and off-target effects in the genome, we validated various CRISPR-Cas9-based engineering methods by targeting genes on the chromosome and on one of the plasmids. For this purpose we used pGM1190-based CRISPR-Cas9 plasmid (55), based on a temperature sensitive replicon, shown to be functional in *S. griseofuscus* by using GusA-based pKG1139.

As the first target, we wanted to eliminate plasmid pSGRIFU1 that harbours 4 BGCs, among them lankacidin, lankamycin, cryptic polyketide and carotenoid BGCs. This plasmid has a very high similarity to the plasmid pSLA2-L of *S. rochei,* where these clusters were characterized (41). A sgRNA was designed to target the DNA primase/helicase-coding region, which is essential for plasmid replication. Three random colonies were selected after the CRISPR procedure and sequenced via Illumina whole genome sequencing. Surprisingly, in all clones, both the targeted pSGRIFU1, but also pSGRIFU2, were lost, leaving only plasmid pSGRIFU3 present in the genome. To estimate the amount of changes in the plasmid-cured strains in comparison to the wild type genome, the WGS data was analyzed with breseq, which identified 11 mutations (six SNVs, three insertions, and two deletions). One of the colonies was selected for further work and named DEL1.

In parallel, we attempted to knockout chromosomally located BGC region number 33, which encodes a putative pentamycin BGC and an uncharacterized NRPS BGC (Fig. S3). The

344   conjugation of the knockout plasmid resulted in less than 10 colonies, 2 of which were selected

345   for Illumina MiSEQ sequencing. It revealed that even though both clones accumulated several

346   SNPs, they did not contain the intended mutation (data not shown). Even after the experiment

347   was repeated, we were not able to select knockout-carrying colonies.

348   In order to verify whether the deletion of pentamycin-NRPS clusters is possible in the plasmid-

349   cured conditions, knockout plasmid was transferred to DEL1. In contrast to the experiments

350   with the wild type, a large number of exconjugants was received. After the plasmid curing,

351   three of the independently received colonies were sequenced with Illumina NextSeq and one

352   of them was additionally sequenced using Nanopore technology. This clone, further referred

353   to as DEL2, was confirmed to contain full deletion of pentamycin-NRPS cluster region and

354   contained a comparatively small amount of SNPs (Fig. 1).

355   In the strain *S. rochei* 7434AN4, which is closely related to *S. griseofuscus*, curing of all three

356   plasmids has been reported to change the topology of the chromosome from linear to circular

357   (41, 56). It is believed that the *tap-tpg* gene pair, which encodes for telomere-associated protein

358   and a terminal protein for end patching, located on both pSLA2-L and pSLA2-M plasmids, is

359   responsible for maintaining the linear architecture of chromosome. Because both the genomes

360   and the associated plasmids in *S. rochei* and *S. griseofuscus* are similar, we investigated if the

361   chromosome of *S. griseofuscus* had circularized during the plasmid curing. We therefore

362   sequenced the strain DEL2 with the Nanopore technology. The assembly graph clearly showed

363   a chromosome with inverted repeat consistent with a linear chromosome. In order to verify

364   presence of the *tap-tpg* homologues in the genome of *S. griseofuscus*, a BLAST search was

365   performed against each gene pair from pSLA2-L and pSLA2-M. The homologues of *tapR1-*

366   *tpgR1* and *tapRM-tpgRM* were found on all three plasmids of *S. griseofuscus* (Table S3). This

367   fact could explain the preserved linear topology of the DEL2 chromosome. The removal of

368   putative pSGRIFU1 and 2 *tap*/*tpg* homologues, does not lead to the circularisation of the

369   chromosome, because of the remaining homologous genes, present on pSGRIFU3.

370   Both of the DEL1 and DEL2 strains did not show any significant changes in their morphology,

371   growth or sporulation (Fig. S1, S2, Table S1). In order to verify the influence of genetic

372   manipulations on the metabolites produced by DEL1 and DEL2, parallel cultivations in ISP2

373   media were made. In comparison to the wild type, strain DEL2 lost the possibility to produce

374   pentamycin, lankacidins and lankamycin, as it was expected (Fig. S4).

375   In order to test whether *S. griseofuscus* is suitable for the expression of heterologous BGCs,

376   we have expressed *S. coelicolor* actinorhodin BGC in the wild type and DEL2. As evident from

13

377 the formation of dark-blue halo, wild type and DEL2 strains are both able to produce

378 actinorhodin in heterologous conditions (Fig. S6).

CRISPR-cBEST mediated knockouts

380 The recent extension of our CRISPR-Cas9 toolkit, CRISPR-cBEST system (57), utilizes

381 cytidine deaminase, fused to dCas9. It allows for introduction of STOP-codons in the positions

382 of interest by converting CG base pairs to AT. Recently, we have reported the use of this system

383 in *S. griseofuscus* (57). In order to test the usability of CRISPR-cBEST for engineering of *S.*

384 *griseofuscus*, the targeted BGCs were selected on the so-called "arm" regions of the

385 chromosome (57). It is well known that the introduction of double strand breaks of the DNA

386 by Cas9 might lead to multiple unwanted consequences. It is particularly dangerous in case of

387 the ends of the chromosome, that by itself frequently undergo rearrangements (58). Therefore,

388 these BGC regions are particularly difficult to engineer. In order to verify whether CRISPR-

389 cBEST system would help to omit these limitations, the targets were selected in 4 different

390 BGCs-containing regions number 4, 30, 31, 34. on the right and left arms of the chromosome.

391 The pCRISPR-CBE plasmids were constructed according to the protocol (59) sequenced and

392 transferred to *S. griseofuscus* via conjugation. Correct clones with the STOP-codons in BGC

393 regions 4, 30, 31 and 34 were confirmed via Sanger sequencing of the region of interest (57).

394 In order to determine the outcomes of each mutation, the morphology, growth and metabolite

395 production was assessed and individually described (Fig. S1, S2 Table S1). We have grown all

396 of the CRISPR-cBEST generated mutants in ISP2 liquid media and compared their production

397 profiles to the wild type (data is not shown). In these initial tests, we were not able to identify

398 specific metabolites, produced from each of these BGCs, possibly because production

399 conditions for these metabolites were not met, or they are cryptic.

400 It was shown that by using the multiplexed CRISPR-cBEST plasmids it is possible to target

401 multiple genes from different BGCs in *S. coelicolor* (57). Therefore, it was our next target to

402 verify such a possibility in *S. griseofuscus*. For this purpose, a multiplex plasmid was

403 constructed, targeting 4 BGCs on the left arm of the chromosome. The sgRNA guides, selected

404 earlier, were used, yielding plasmid pCRISPR-MCBE-1-2-4-6, targeting BGCs 1, 2, 4 and 6.

405 The plasmid was verified via Sanger sequencing and transferred to *S. griseofuscus* via

406 conjugation. Up to 24 exconjugant colonies were tested via PCR. Each of the targeted regions

407 was amplified using a selected set of primers, the fragments were purified and sequenced by

408 Sanger sequencing. As the result of the screening, for each of the targeted regions at least one

409 successful editing event was detected. We were able to select a colony of *S. griseofuscus*

14

410   pCRISPR-MCBE-1-2-4-6 with a total of three edited targets (E3I2) (Table S4). Strain E3I2 has

411   exhibited signs of sporulation deficiencies and changes in morphology, that might be related

412   to the specific combination of the mutations that were introduced (Fig. S1). However, growth

413   of this strain was clearly not inhibited in liquid cultures (Fig. S2, Table S1). In addition, the

414   metabolite biosynthesis profile of E3I2 was verified in ISP2 liquid media (data is not shown).

415   We were not able to identify specific metabolites, linked to the inactivated BGC, probably

416   because the conditions for the production of these metabolites were not met or these particular

417   BGCs were not expressed.

418   One of the significant problems for CRISPR-Cas9-mediated targeting are unwanted off-target

419   effects. Similarly, such problems exist while using CRISPR-BEST systems. It was shown that

420   while using the CRISPR-BEST for the generation of knockouts in *S. coelicolor*, a relatively

421   small amount of SNPs can be observed (57). However, the influence of the presence of

422   CRISPR-BEST plasmid on the accumulation of the SNPs over the time during continuous

423   cultivation was never studied.

424

425

WT

1		G	C	T	A
	A	0	1	0	0
From	T	0	0	0	0
	C	0	0	0	0
	G	0	2	0	0

To

p057_0D

2		G	C	T	A
	A	6	1	1	0
From	T	2	4	0	1
	C	7	0	3	0
	G	0	6	0	2

To

p057_20D

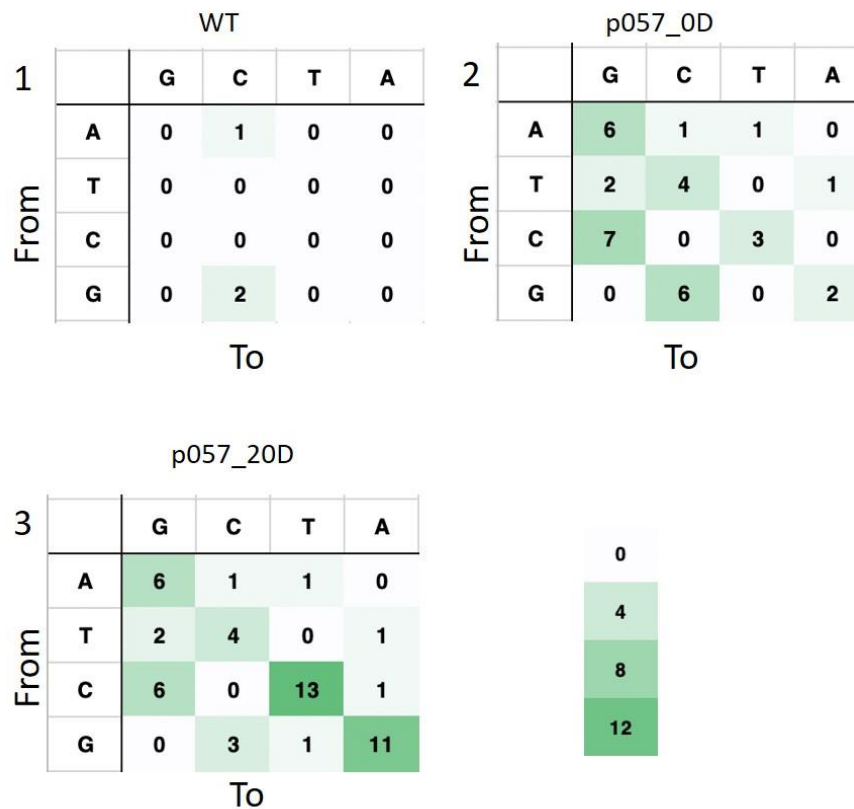| 3		G	C	T	A
	A	6	1	1	0
From	T	2	4	0	1
	C	6	0	13	1
	G	0	3	1	11

To

0
4
8
12

426

427 **Fig. 4. Genome wide off-target evaluation of CRISPR-BEST mediated mutations in the strain preserved**
428 **and sequenced after introduction of Stop-codon (p057_0D; 2) in comparison to the same strain that was**
429 **passed consecutively 20 times in liquid cultures (p057_20D; 3)**. Mutations, predicted for the WT Illumina
430 dataset (1), used to produce the reference, the level of which thus can be considered a technical noise.

431

432 In order to study these effects we have performed a long term cultivation experiment with
433 CRISPR-cBEST generated mutant strain *S. griseofuscus* HEP81_06602 (p057). The initial and
434 resulting strains were sequenced using Illumina NextSEQ and compared to the wild type strain,
435 using breseq analysis (Fig. 1 and 4). Notably, the introduced Trp221Stop mutation in the
436 putative gene HEP81_06602 (BGC 30) was maintained even after 20 transfers without the
437 antibiotic pressure (Fig. 1).

438 The breseq analysis of the Illumina-generated reads of the wild type genome had detected 3
439 SNPs, which can be considered a technical baseline. In case of strain p057_0D this number
440 increased to 33 SNPs altogether, with a majority of them being C to T exchanges, which
441 putatively can be attributed to nonspecific activity of the CRISPR-BEST cytidine aminase.
442 After 20 consecutive transfers this number increased to 50 SNPs with a majority of them being
443 C to T and A to G exchanges. Both numbers are falling in the range of the previously reported
444 for *S. coelicolor* (57) and are promising for the engineering of *S. griseofuscus*.

16

Discussion

445 Actinobacterial genomes usually code for 20-30 secondary metabolites (4). However, under

446 laboratory conditions a limited number of compounds are produced, possibly due to the

447 absence of appropriate cues for the BGC expression. One strategy to address this challenge is

448 the expression of BGCs in heterologous hosts. Even though there are several of such strains

449 available, the current rate of successful expression of BGCs is low (14). Here, we present a

450 further diversification of the heterologous strain panel with a potentially highly adaptable,

451 flexible and easy-to-handle heterologous host, *S. griseofuscus*. The literature-provided

452 information regarding *S. griseofuscus* is comparably dated. Therefore, we have decided to start

453 our work with the wild type strain *S. griseofuscus* DSM 40191 and test its qualities first hand.

454 While exploring its phylogenetic position, we have identified an ANI to *S. rochei* 7434AN4 of

455 99.54%, while the ANI of *S. rochei* 7434AN4 to the type strain *S. rochei* NRRL B-2410 is only

456 at 84.03%. This indicates that *S. rochei* 7434AN4 may be misclassified and rather should be

457 included in the *griseofuscus* group. *S. rochei* 7434AN4 is described in detail as a producer of

458 lankamycins, lankacidins and pentamycin (40, 42). However, it was never systematically

459 characterised regarding its genotype, phenotype or possibility of use of different genetic

460 engineering methods. Therefore, detecting the possible similarity between *S. rochei* 7434AN4

461 and *S. griseofuscus* DSM40191 has been helpful in order to ease characterisation of *S.*

462 *griseofuscus* metabolic profile.

463 The comparison of *S. griseofuscus* genome to genomes of the model streptomycetes *S.*

464 *coelicolor* and *S. venezuelae* has revealed a close genetic similarity, retained in the ability to

465 process sources of carbon, nitrogen and sulphur. Notably, *S. griseofuscus* is genetically, and

466 phenotypically closer related to *S. coelicolor* than to *S. venezuelae*. This enables the use of the

467 methods developed for *S. coelicolor*. *S. griseofuscus* utilizes various media sources for its

468 growth, displaying apparent metabolic flexibility, needed for heterologous hosts. In the growth

469 tests, the strain had displayed fast uncomplicated growth in the liquid media, compatible with

470 OD measurements, and abundant sporulation on solid media. In addition, we demonstrated that

471 *S. griseofuscus* produces native metabolites lankacidin, lankamycin and pentamycin, as it was

472 expected from its phylogenetic "proximity" to *S. rochei* 7434AN4. This information was later

473 used in the selection of target BGCs for genetic engineering.

474 Heterologous host strains have to be frequently manipulated genetically in order to be molded

475 as producers of specific secondary metabolites. Therefore, it is important to have access to a

17

477   set of efficient genetic engineering techniques. We have tested various methods, such as

478   integrative and replicative vectors, CRISPR-Cas9 based knockouts and CRISPR-cBEST base

479   editing. *S. griseofuscus* is accessible to all these methods, implemented for single or

480   multiplexed knockouts, which generate minimal off-target effects. Even though CRISPR-based

481   systems were shown to be applicable in Actinobacteria, their use is still largely limited to model

482   strains and typically requires further adjustments (60). It is, therefore, highly beneficial that *S.*

483   *griseofuscus* does not require any of such additional efforts.

484   Our first target for the genome reduction of *S. griseofuscus* has been the curing of the largest

485   plasmid pSGRIFU1, highly similar to pSLA2-L of *S. rochei* (42). It carries 4 BGCs, among

486   these the lankacidin and lankamycin BGCs, products which were detected in the wild type

487   supernatant. As it was expected (40), CRISPR-Cas9 mediated curation of pSGRIFU1 led to the

488   disappearance of these metabolites in the supernatants of the mutant strain DEL1. To generate

489   strain DEL1 we have targeted the DNA primase/helicase-coding region of the pSGRIFU1

490   plasmid, which also has caused the curing of plasmid pSGRIFU2. This could be explained by

491   the presence of several putative off target sites in pSGRIFU2. Curation of pSGRIFU2 was

492   unexpected, but a positive outcome for the genome minimization and did not lead to

493   chromosome cyclisation of DEL1, as it was shown for *S. rochei* (56). We believe that the

494   homologues of genes *tap*/*tpg*, that were detected on plasmid pSGRIFU3, are needed to sustain

495   the architecture of the chromosome.

496   Furthermore, we have successfully removed pentamycin and NRPS BGC from the genome of

497   DEL1, generating strain DEL2. The metabolomic studies have shown that strain DEL2 does

498   not produce pentamycin (Fig. S5).

499   DEL2 can be considered as the first genome reduced strain of *S. griseofuscus* with 5,19% of

500   its genomes removed without any negative effects on its growth, morphology or sporulation

501   levels. In addition, actinorhodin BGC was successfully expressed in DEL2 (Fig. S6).

502   Therefore, we believe that strain DEL2 can be considered as the first step towards developing

503   heterologous production host, suitable for expression of polyketide BGCs.

504

Conclusions

505

506

507 For the last several decades, *S. griseofuscus* has remained largely unexplored. In this paper, a

508 detailed characterization of *S. griseofuscus* DSM 40191 is presented, including studies of its

509 genotype, phenotype and metabolic capabilities. We believe that the results, described in this

510 manuscript clearly demonstrate the high potential of *S. griseofuscus* strain as to its further

511 development as heterologous host strain. We plan to continue our work on the genome

512 reduction of *S. griseofuscus* DEL2, using step-by-step removal of BGCs with simultaneous

513 characterisation of mutants growth and metabolite production. In the following next phase of

514 development, we plan to adapt the resulting strain to expression of different types of BGCs, by

515 introducing additional integration sites and by modifying the primary metabolism to suit the

516 precursor needs.

517

518

519

520

521

Materials and Methods

1. Strains, used or constructed in this study

Strains *S. griseofuscus* DSM 40191 and *S. venezuelae* DSM 40230 were received from DSMZ strain collection as freeze-dried pellets. Strain *S. coelicolor* M145 was received from Y. Tong (NBC group, DTU, Denmark). *E. coli* Top10 and Mach1 (both from Thermo Fisher Scientific) were used as general cloning hosts in this study. Other strains that were used in this study are described in the Table S4.

2. Growth and cultivation conditions

All of the Streptomyces strains were cultivated using 3 different solid media: MS (45), ISP2 and ISP4 (International Streptomyces Project medium 2 and 4; premixed Difco^{TM} ISP2 dehydrated medium (Fisher Scientific, #DF0770-17-9) and premixed Difco^{TM} ISP4 dehydrated medium (Fisher Scientific, #DF0772-17-7)). DSMZ medium 65 (GYM *Streptomyces* medium) is the medium recommended by the German Collection of Microorganisms and Cell Cultures (DSMZ) for cultivation of *S. griseofuscus* DSM 40191 and prepared as described by DSMZ (46). CDMZ, a chemically defined media for detection of physostigmine, was prepared as described in (44). Otherwise, *Streptomyces* minimal media (45), MAM, molicidin A medium, and version of ISP2 without addition of agar, were used for liquid media cultivations. For cultivations in liquid medium 250 ml shake flasks were used in combination with approximately 30 3 mm glass beads, incubated at 30 °C and 200 rpm. For the cultivations in liquid media, a two stage cultivation principle was used. At first, pre-cultures were directly inoculated with spore suspensions, that were collected from fully grown MS plates and incubated overnight. The OD600 was measured for each pre-culture and re-inoculated in the main culture up to the final OD600 of 0.1. The OD600 measurements were carried out in the 1.5 mL disposable polystyrene spectrophotometer cuvettes. Before each measurement, the cuvettes were thoroughly mixed using a vortex-mixer and each cuvette was measured three times. Therefore, each presented data point corresponds to a total of nine measurements. The maximum specific growth rate μ_{max} was determined by plotting the natural logarithm of all OD_{600} values against the cultivation time. Further on, a linear fit for data points corresponding

552 to the exponential phase was calculated and the slope of that fit corresponds to μ_{max}. Using

553 μ_{max}, the minimal time needed for the OD_{600} value to double (td) was calculated as described

554 in (61).

3. Cloning, primers and constructed plasmids

556 Spacer sequences, primers, constructed and used plasmids are listed in Table S5. All spacer

557 sequences were selected with the help of CRISPy-web (62, 63).

558 The procedure of ssDNA oligo bridging was used for integration of spacers into CRISPR

559 plasmids. The CRISPR plasmid of interest was digested with NcoI at 37 °C for 30 min and

560 dephosphorylated using FastAP at 37 °C for another 30 min. The reaction was inactivated by

561 incubation at 65 °C for 10 min. The 20 nt spacers were ordered as oligos from IDT with 20 nt

562 overlaps to the backbone on both sides (Table S5). After dilution to 100 µM, the oligos were

563 diluted to a concentration of 0.2 µM using 1x NEBuffer 2. A 10 µl reaction mix was prepared

564 containing 30 ng of the linearized backbone, 5 µl of the 0.2 µM oligo, and ddH$_2$O to 10 µl. 10

565 µl of NEBuilder® HiFi DNA Assembly Master Mix were added and the reaction was incubated

566 for 1 h at 50 °C. Up to 5 µl were transformed into electrocompetent *E. coli* Mach1. Positive

567 colonies were identified by colony PCR and running the samples on a 2 % agarose gel. Putative

568 positive colonies were further analyzed by the in-house Sanger sequencing. The construction

569 of CRISPR-Cas9 plasmids and their transfer was carried out as described in (59), All *E.coli-*

570 *Streptomyces* conjugation experiments were conducted according to the modified protocol

571 from (45) in which the heat shock step was completely omitted. The MS media with addition

572 of magnesium chloride solution (45) was used for plating of conjugation mixes.

4. Genome comparison of *S. griseofuscus, S. coelicolor and S. venezuelae*

574 To compare genetic features of *S. griseofuscus* with other well studied *Streptomyces* strains,

575 we downloaded the genomes of *S. coelicolor* A(3)2 (NCBI accession.: NC_003888) and *S.*

576 *venezuelae* ATCC 10712 (NCBI accession: NZ_CP029197). All three genomes were annotated

577 using the KEGG annotation server that describes the biological subsystem of each gene. The

578 number of genes in different biological subsystems were counted for each of the three genomes.

579 Similarly, to compare the metabolic properties, the number of unique KEGG reactions were

580 counted for different metabolic pathways across the three genomes. The genomic and

581 phenotypic microarray data was analyzed together using the *dape* module of the DuctApe

582 software that is used to correlate genomic data with phenomic data for multiple strains. As part

583 of this software, bidirectional best blast hits (E-value threshold – 1e-10) were calculated to
584 generate the pangenome of the three strains. Based on this pangenome, the number of shared
585 genes, unique KEGG gene IDs, and unique KEGG reaction IDs across the three genomes were
586 calculated.

587

588 5. BioLog phenotypic microarrays and data analysis

589 Phenotype MicroArrays were ordered from BioLog, which in addition had provided
590 customized protocol for the cell inoculation and measurements. In this study testing plates PM
591 1 to 4 were used, out of which PM1 and 2 contain Carbon sources, PM2 - Nitrogen sources,
592 PM4 - Phosphorus and Sulphur sources. All the measurements for all strains were performed
593 in three technical and two biological replicates. *S. coelicolor* M145, *S. venezuelae* DSM 40230
594 and *S. griseofuscus* DSM 40191 were all grown on MS plates for 6 days until clear sporulation
595 signs appeared. The spores were collected in sterile water and diluted to 80% of cell density
596 using BioLog-supplied turbidimeter (catalogue number: 3531). BioLog redox dye mix G was
597 used for all measurements. The recipe of minimal media and other supplementing resources
598 were received from BioLog. The cells were mixed with the prepared media and inoculated in
599 the PM plates using multichannel pipettes and immediately loaded into the OmniLog
600 instrument for measurements (catalogue number: 91171).

601 The kinetic data for all testing plates PM1 to PM4 generated by BioLog was further analyzed
602 using DuctApe software (37) along with the genomic data. Activity index between 0 to 9 was
603 assigned for each nutrient source to represent the growth activity on each nutrient source.
604 Average activity index was further calculated for the two replicas that were used to generate
605 phenotype data. The average activity index across 379 nutrient sources across 4 PM plates was
606 visualized using activity index rings for the three different strains (Dataset 2). Sources with
607 differential growth activities were further analyzed. To compare the growth activity with
608 genomic features, a matrix with activity on different KEGG nutrients (rows) against KEGG
609 pathways with the nutrient (columns) was calculated. These matrices were compared using
610 heatmaps where rows and columns are ordered as per *S. griseofuscus* growth activity index and
611 number of reactions per KEGG pathway respectively (Dataset 3).

612

613 **Genome scale model reconstruction**

614   We used Python-based genome-scale model reconstruction tool 'Genome-scale Modelling

615   with Secondary Metabolism' (GMSM) that implements bidirectional blastp hits-based

616   homology modelling(unpublished) to automatically reconstruct models for both *S. venezuelae*

617   and *S. griseofuscus*. We used a model of *S. coelicolor* A3(2), iMK120848, as a template for the

618   homology modelling. The in-silico growth predictions were made using flux balance analysis

619   on different simulated mediums using COBRApy. Confusion matrix was created from

620   predicted and observed growth phenotypes to compare the model prediction against Biolog

621   data.

6. DNA isolation, sequencing and assembly

623   As described in (34), DNA from *S. griseofuscus* was purified using the Qiagen Genomic Tip

624   100 (Venlo, Netherlands) kit. Illumina (San Diego, California, USA) WGS libraries were

625   constructed using the KAPA (St. Louis, Missouri, USA) HYPRplus kit and sequenced on an

626   Illumina MiSeq machine with a 2x150nt sequencing kit, except for the library for strain DEL2,

627   which was sequenced on a Illumina NextSeq 500. Macrogen Inc. (Seoul, South Korea)

628   generated the PacBio data on an RSII machine. The Illumina data was adaptertrimmed using

629   Adapterremoval2 (v. 2.2.2) (64) with the switches --trimns –trimqualities. As described in (34),

630   the genome of *S. griseofuscus* was assembled using PacBio data with the assembly program

631   Flye (v. Flye 2.4.1-geb89c9e) (65) with the switches --genome-size 8m --iterations 5 for five

632   consecutive rounds of polishing using the PacBio data. The assembly was then polished with

633   the Illumina data using the polishing module of Unicycler (v. 0.4.8-beta) (66). After this

634   assembly process, the inverted repeat (IR) chromosome ends were detached from the rest of

635   the chromosome on a separate contig. We manually added the IR to both ends of the

636   chromosome, repeated the illumina polishing with unicycler-polish, and used Minimap2 (v.

637   2.16-r922) (67) and Bowtie2 (v. 2.3.5) (68) to map pacbio and illumina reads to the manually

638   curated genome assembly to confirm that the IR ends of the chromosome were correctly

639   attached. We used artemis genome viewer (v .18.0.2) (69) to visualize the mappings. BUSCO

640   (v. 4.0.5) (70) was used for estimating the quality of the genome assembly and Bandage (v.

641   0.8.1) (71) was used to view and evaluate the assembly graph. The assembled genome was

642   gene annotated using Prodigal (v.2.6.3) (72) and the identified genes were functionally

643   annotated using Prokka (v. 1.14.0) (73) with the PFAM-A (v. 32.0) (74) database and six

644   publically available manually annotated actinobacterial genomes of high quality using the

645   prokka --proteins switch (see Table SI6). These databases were used in addition to the default

646 databases, not instead of them. RNAmmer (75) was used for rRNA gene prediction and

647 Aragorn was used for tRNA prediction (76). Nanopore data from the strain DEL2 was obtained

648 to check for circularization of the genome. The DNA of the DEL2 strain was extracted as

649 described above, and a nanopore library was constructed with the rapid (SQK-RBK004) kit

650 from Oxford Nanopore Technologies (Oxford, United Kingdom) and the barcode RB05. The

651 data was generated on a MinION machine with a 9.4.1 flowcell. The raw data was

652 demultiplexed using Deepbinner (v. 0.2.0) (77) and basecalled using Guppy (v. 3.2.2+9fe0a78,

653 Oxford Nanopore Technologies, Oxford, United Kingdom), before the technical sequence was

654 removed using Porechop (v. 0.2.4; https://github.com/rrwick/Porechop). The Nanopore data

655 was assembled using flye assembler as described above for PacBio data.

656

7. Comparison of BGCs of *S. griseofuscus* with BGCs detected across *Streptomyces sp.*

659 Genome sequences with complete assembly annotation of *Streptomycetaceae* group are

660 collected from the PATRIC database (78) Among 214 complete genome sequences, one was

661 annotated of poor quality owing to its low fine consistency score (PATRIC ID: 2588708.3).

662 One other genome had 22 contigs (PATRIC ID: 1969.5). Remaining 212 genomes of high

663 quality are processed with antiSMASH v5.1 to detect 6380 BGCs in total. The antiSMASH

664 predicted 34 BGCs on chromosome and one BGC on plasmid of *S. griseofuscus*. Some of these

665 BGCs were further split into candidate clusters manually. We used BiG-SCAPE software (79)

666 to generate similarity network of all 6380 BGCs detected in 212 public genomes, 1808 known

667 BGCs from MIBIG database v1.3, 35 BGCs from *S. griseofuscus* genomes and manually

668 selected 14 candidate clusters. The cutoffs of 0.3, 0.5 and 0.7 were used on raw_index similarity

669 metric of BiG-SCAPE analysis. The similarity network was visualized using Cytoscape v3.7.

670 Further some of the candidate clusters of interest (e.g. candidate cluster number 5) were

671 analyzed using the CORASON feature of BiG-SCAPE to generate gene cluster alignments

672 based on similarity.

673

24

8. Extractions and evaluation of metabolite production

For targeted and untargeted metabolomics analysis, 50 ml liquid cultures with the medium of interest (ISP2 and CDMS) were prepared and cultivated in 250 ml shake flasks for 5 days. After 5 days, the cultures were harvested and centrifuged for 20 min at 10,000 x g. The supernatants were collected in 250 ml glass bottles and extracted using equal volumes (1:1) of ethyl acetate (liquid-liquid extraction). The organic phase was separated from the aqueous phase using a 500 ml separatory funnel, after shortly shaking, and collected in a 250 ml glass bottle. The process was repeated 3 times using fresh solvent. The extracts were combined for each sample and evaporated using a Büchi Rotavapor R-300 in combination with a Büchi Heating Bath B-300 Base, a Büchi Interface I-300, a Büchi Vacuum Pump V-300, and a Julaba Recirculating Cooler FL601. The temperature of the water bath was set to 38 °C, the rotation speed to 140 rpm, and the pressure to 150 mbar, which was gradually decreased to 50 mbar to complete dryness. Next, the analytes were dissolved in 1 ml of 50% v/v methanol and transferred to 2.0 mL Eppendorf tubes. The samples were centrifuged for 10 minutes in 12,000 CFU and the supernatants were transferred in new eppendorf tubes. The samples were again evaporated using an Eppendorf Concentrator Plus Speedvac without heating, dissolved in 100 µl of 50% v/v methanol and centrifuged for 10 min in 12,000 CFU. The supernatants were transferred to autosampler vials with inserts, and analyzed using a LC-MS system. Analysis was performed using a Thermo Dionex Ultimate 3000 UHPLC system with a diode array detector (DAD) interfaced with an Orbitrap Fusion Tribid mass spectrometer (Thermo Scientific, San Jose, USA), using an EASY-IC ESI source. Separation conditions were as follows: Column, Agilent Zorbax Eclipse Plus C18, 100 x 2.1 mm i.d., 1.8 µm particles. The mobile phase used was (A) purified water with 0.1 % formic acid and (B) acetonitrile with 0.1 % formic acid. The flowrate was set to 0.35 ml/min and column temperature was set to 35 °C, while the injection volume was set to 1 µl. The following gradient was used: 0-0.5 min. 5 % B, increasing to 100 % B at 13 min., and holding until 15 min., returning to 5 % B at 15.1 min., and equilibrating for 1.9 min., for a total run time of 17 min. Full-scan mass spectrometric detection was performed in positive and negative ESI mode with the following parameters: source voltage, 3500 V (positive mode) and 2700 V (negative mode); sheath gas flow rate (N2), 50 arbitrary units; auxiliary gas, 10 arbitrary units; sweep gas, 1 arbitrary unit; Ion transfer tube temperature, 325 °C; Vaporizer temperature, 350 °C. MS full-scan analysis was performed using the orbitrap with the following settings: Orbitrap resolution, 120,000; scan range 100-1,000 Da; RF lens, 50 %. Before analysis, the MS was calibrated using ESI Positive Ion

707 Calibration Solution (P/N 88323) and ESI Negative on Calibration Solution (P/N 88324,

708 Thermo Scientific, San Jose, USA). Fragmentation data for compound annotation were

709 obtained using data-dependent MS/MS analysis by selecting the top four most intense ions per

710 cycle. Dynamic exclusion was used to exclude ions for 3 s after two measurements within 4 s.

711 Fragmentation was performed using an assisted fragmentation HCD 15, 30, 45, and 60 % at a

712 resolution of 30,000 with an AGC target of 1×10^5 and a maximum injection time of 64 ms.

713 Data analysis was performed using Thermo Scientific Compound Discoverer 3.0.0.294. Using

714 the software, data from the LC-MS analysis were aligned, and compound annotation was

715 performed by matching against molecular formulas from StreptomeDB2 and AntiBase, as well

716 as fragmentation spectra from mzCloud. In the case of pentamycin detection, a pure

717 pentamycin standard was used, diluted in 10% v/v methanol to 10^{-3} mg/ml concentration. In

718 regards to the other compounds, no standards were measured, due to their absence in the

719 market, so their detection is putative.

Data availability

721 The Genbank accession numbers for *S. griseofuscus* genome are CP051006 (chromosome),

722 CP051007 (pSGRIFU1), CP051008 (pSGRIFU2), and CP051009 (pSGRIFU3). All

723 sequencing data can be found in BioProject PRJNA622435.

724 Following data files including Excel tables and XML format model files are available at link :

725 (https://figshare.com/s/ad8ff9010782c1bc1032).

726

727 Dataset 1. Genome comparison data

728 Excel tables with gene annotations from NCBI genbank and KEGG subsystems. Gene

729 presence-absence table for genes across the pangenome of S. griseofuscus, S. coelicolor, S.

730 venezuelae. KEGG subsystem gene and reaction counts per organism.

731

732 Dataset 2. Phenotype microarray data analysis with DuctApe

733 Excel tables with raw kinetic growth data across plates PM1 to PM4 with 2 replicates for S.

734 griseofuscus, S. coelicolor, S. venezuelae. Biolog activity index (0 to 9) across 384 nutrients in

735 each strain and comparison of number of active nutrients across the three strains.

736

737 Dataset 3. Genotype phenotype mapping using KEGG pathways

738   Excel tables with matrix of biolog activity index across different KEGG nutrients mapped on

739   to the KEGG pathways. Combined matrix with mean activity differences across three strains.

740

741   Dataset 4. Genome scale metabolic models

742   The SMBL formatted genome-scale metabolic models of the three *Streptomyces* strains used

743   in the study

744

745   Dataset 5: Comparison of genome scale model based growth predictions and BioLog data

746   Excel table with list of reactions considered in the three genome scale models and comparison

747   of in-silico growth prediction on different nutrients against BioLog data.

748

749   Dataset 6. BGC comparison across *Streptomyces* using BiGSCAPE

750   Information on BGCs detected in *S. griseofuscus* and information on the public genomes

751   selected along with the BGCs selected for comparison. Similarity network data used for

752   generation of Fig. 3 with information on nodes and edges.

753

754

Legends of Supplementary Information files

Fig. S1. Phenotype of *S. griseofuscus* strains on solid media. 1 - ISP2 media; 2 - MS media; 3 - ISP4 media. A- wild type; B - *S. griseofuscus* IppsD_1 (p054); C - *S. griseofuscus* ItycC_2 (p056); D -*S. griseofuscus* HEP81_06602 (p057); E - *S. griseofuscus* IspkC (p059); F - *S. griseofuscus* DEL1; G - *S. griseofuscus* DEL2; I - *S. griseofuscus* E3I2

Fig. S2. OD600 measurements of *S. griseofuscus*-derived strains, while grown in ISP2 liquid media. A - characterization of DEL1, DEL2, E3I2 strains in comparison to the wild type; B - characterization of slow growing strain p056 in comparison to the wild type; C - characterization of *S. griseofuscus* IppsD_1 (p054), *S. griseofuscus* HEP81_06602 (p057) and *S. griseofuscus* IspkC (p059) strains in comparison to the wild type; D - comparison of all 3 independant cultivations of wild type strains

Fig. S3. Alignment of selected families of BGCs. Here selected regions and candidate clusters of *S. griseofuscus* are aligned against similar BGCs detected across the dataset. Note: Pentamycin BGC was manually added in the above analysis as default BiGSCAPE uses an older MIBIG version lacking pentamycin BGC.

Fig. S4. Results of metabolite production analysis of cultivations and extractions of wild type, DEL1 and DEL2 strains, grown in liquid ISP2 media. A) Extracted ion chromatograms of pentamycin ($C_{35}H_{58}O_{12}$) production of strains WT, DEL1 and DEL2 cultivated in ISP2 media in comparison to the extraction of pure ISP2 media and pentamycin standard. All cultivations were done in triplicates; B) the extracted EICs for kujimycin A ($C_{40}H_{70}O_{15}$) are plotted for comparative metabolomics between blank ISP2 medium, three replicates of *S. griseofuscus* WT in ISP2 and three replicates of *S. griseofuscus* DEL1 in ISP2. The peak at 8.70 min is a putative match for the molecule and it is clearly absent in case of DEL1 strain

Fig. S5. A and B - three independent colonies of *S. griseofuscus* pKG1139 (I), *S. griseofuscus* pSETGUS (II) and wild type (III), grown on ISP2 solid media after the 24 hours incubation at 40°C to verify possibility of plasmid curing, A- colonies before the addition of X-Gluc; B - colonies after the addition of X-Gluc

787   Fig. S6. Strains *S. griseofuscus* wild type, DEL1 and DEL2, that contain integrative plasmid

788   construct, carrying complete actinorhodin gene cluster, cloned from *S. coelicolor* M145. All

789   strains were grown on MS media, supplemented with apramycin for 6 days before the photo

790   was taken

791

792   Table S1. OD600 measurements of *S. griseofuscus*-derived strains, while grown in ISP2 liquid

793   media. The growth curve was built in liquid ISP2 culture based on three independently

794   performed cultivations, each of which consisted of three cultivation flasks and each data point

795   was measured three times. Please, review this table in connection to Fig. S3. The μ_{max} for the

796   wild type was calculated to be 0.43 ± 0.08 h^{-1} on average for three cultivations with td of 1.68

797   ± 0.29 h. Sporulation of the DEL1 and DEL2 strains was estimated for their growth on MS

798   media, where the wild type sporulation was accounted for at $3,43\pm3,05\times10^8$ CFU/mL. Wild

799   type sporulation was observed only on MS and ISP4 media, but not ISP2, with an average

800   amount of spores on ISP4 - $2,60\pm4,58\times10^8$ CFU/mL, therefore MS media was chosen as the

801   most suitable for growth experiments.

802

803   Table S2. Compounds detected in ethyl acetate extractions of supernatants of *S. griseofuscus*

804   wild type cultivations in different liquid media. RT - retention time, MM - minimal medium,

805   M65 - medium 65. Detection of pentamycin was confirmed via comparison with a pure

806   pentamycin standard

807

808   Table S3. BLAST hits of genes *tapR1/tpgR1* from *S. rochei* 7434AN4 genome to genes in *S.*

809   *griseofuscus* genome

810

811   Table S4. Summary of all primers, genetic constructs and strains used in this study

Acknowledgements

References

1. Hutchings MI, Truman AW, Wilkinson B. 2019. Antibiotics: past, present and future. Curr Opin Microbiol 51:72–80.

2. Genilloud O. 2017. Actinomycetes: still a source of novel antibiotics. Nat Prod Rep 34:1203–1232.

3. 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.

4. Belknap KC, Park CJ, Barth BM, Andam CP. 2020. Genome mining of biosynthetic and chemotherapeutic gene clusters in Streptomyces bacteria. 1. Sci Rep 10:1–9.

5. Hoskisson PA, Seipke RF. 2020. Cryptic or Silent? The Known Unknowns, Unknown Knowns, and Unknown Unknowns of Secondary Metabolism. mBio 11.

6. Tomm HA, Ucciferri L, Ross AC. 2019. Advances in microbial culturing conditions to activate silent biosynthetic gene clusters for novel metabolite production. J Ind Microbiol Biotechnol 46:1381–1400.

7. Chen C, Zhao X, Chen L, Jin Y, Zhao ZK, Suh J-W. 2015. Effect of overexpression of endogenous and exogenous Streptomyces antibiotic regulatory proteins on tacrolimus (FK506) production in Streptomyces sp. KCCM11116P. RSC Adv 5:15756–15762.

8. Wei Q, Aung A, Liu B, Ma J, Shi L, Zhang K, Ge B. 2020. Overexpression of wysR gene enhances wuyiencin production in ΔwysR3 mutant strain of Streptomyces albulus var. wuyiensis strain CK-15. J Appl Microbiol 129:565–574.

9. Wang B, Guo F, Dong S-H, Zhao H. 2019. Activation of silent biosynthetic gene clusters using transcription factor decoys. 2. Nat Chem Biol 15:111–114.

10. Horbal L, Marques F, Nadmid S, Mendes MV, Luzhetskyy A. 2018. Secondary metabolites overproduction through transcriptional gene cluster refactoring. Metab Eng 49:299–315.

11. Palazzotto E, Tong Y, Lee SY, Weber T. 2019. Synthetic biology and metabolic engineering of actinomycetes for natural product discovery. Biotechnol Adv 37:107366.

12. Culp EJ, Yim G, Waglechner N, Wang W, Pawlowski AC, Wright GD. 2019. Hidden antibiotics in actinomycetes can be identified by inactivation of gene clusters for common antibiotics. Nat Biotechnol 37:1149–1154.

13. Hug JJ, Krug D, Müller R. 2020. Bacteria as genetically programmable producers of bioactive natural products. Nat Rev Chem 1–22.

14. Myronovskyi M, Luzhetskyy A. 2019. Heterologous production of small molecules in the optimized *Streptomyces* hosts. Nat Prod Rep 36:1281–1294.

15. Ahmed Y, Rebets Y, Estévez MR, Zapp J, Myronovskyi M, Luzhetskyy A. 2020. Engineering

855   of Streptomyces lividans for heterologous expression of secondary metabolite gene clusters.

856   Microb Cell Factories 19:5.

857 16. Bu Q-T, Yu P, Wang J, Li Z-Y, Chen X-A, Mao X-M, Li Y-Q. 2019. Rational construction of

858   genome-reduced and high-efficient industrial Streptomyces chassis based on multiple

859   comparative genomic approaches. Microb Cell Factories 18:16.

860 17. Ikeda H, Kazuo S, Omura S. 2014. Genome mining of the Streptomyces avermitilis genome and

861   development of genome-minimized hosts for heterologous expression of biosynthetic gene

862   clusters. J Ind Microbiol Biotechnol 41:233–250.

863 18. Komatsu M, Komatsu K, Koiwai H, Yamada Y, Kozone I, Izumikawa M, Hashimoto J, Takagi

864   M, Omura S, Shin-ya K, Cane DE, Ikeda H. 2013. Engineered Streptomyces avermitilis host for

865   heterologous expression of biosynthetic gene cluster for secondary metabolites. ACS Synth Biol

866   2:384–396.

867 19. Kim EJ, Yang I, Yoon YJ. 2015. Developing Streptomyces venezuelae as a cell factory for the

868   production of small molecules used in drug discovery. Arch Pharm Res 38:1606–1616.

869 20. Yin S, Li Z, Wang X, Wang H, Jia X, Ai G, Bai Z, Shi M, Yuan F, Liu T, Wang W, Yang K.

870   2016. Heterologous expression of oxytetracycline biosynthetic gene cluster in Streptomyces

871   venezuelae WVR2006 to improve production level and to alter fermentation process. Appl

872   Microbiol Biotechnol 100:10563–10572.

873 21. Myronovskyi M, Rosenkränzer B, Nadmid S, Pujic P, Normand P, Luzhetskyy A. 2018.

874   Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous

875   expression of secondary metabolite clusters. Metab Eng 49:316–324.

876 22. Fazal A, Thankachan D, Harris E, Seipke RF. 2020. A chromatogram-simplified Streptomyces

877   albus host for heterologous production of natural products. Antonie Van Leeuwenhoek

878   113:511–520.

879 23. Myronovskyi M, Rosenkränzer B, Stierhof M, Petzke L, Seiser T, Luzhetskyy A. 2020.

880   Identification and Heterologous Expression of the Albucidin Gene Cluster from the Marine

881   Strain Streptomyces Albus Subsp. Chlorinus NRRL B-24108. Microorganisms 8.

882 24. Rodríguez Estévez M, Gummerlich N, Myronovskyi M, Zapp J, Luzhetskyy A. 2019.

883   Benzanthric Acid, a Novel Metabolite From Streptomyces albus Del14 Expressing the

884   Nybomycin Gene Cluster. Front Chem 7:896.

885 25. Sakamoto JM, Kondo SI, Yumoto H, Arishima M. 1962. Bundlins A and B, two antibiotics

886   produced by Strepto. J Antibiot (Tokyo) 15:98–102.

887 26. Larson JL, Hershberger CL. 1984. Shuttle vectors for cloning recombinant DNA in Escherichia

888   coli and Streptomyces griseofuscus C581. J Bacteriol 157:314–317.

889 27. Larson JL, Hershberger CL. 1986. The minimal replicon of a streptomycete plasmid produces

890   an ultrahigh level of plasmid DNA. Plasmid 15:199–209.

891 28. McHenney MA, Baltz RH. 1988. Transduction of plasmid DNA in Streptomyces spp. and
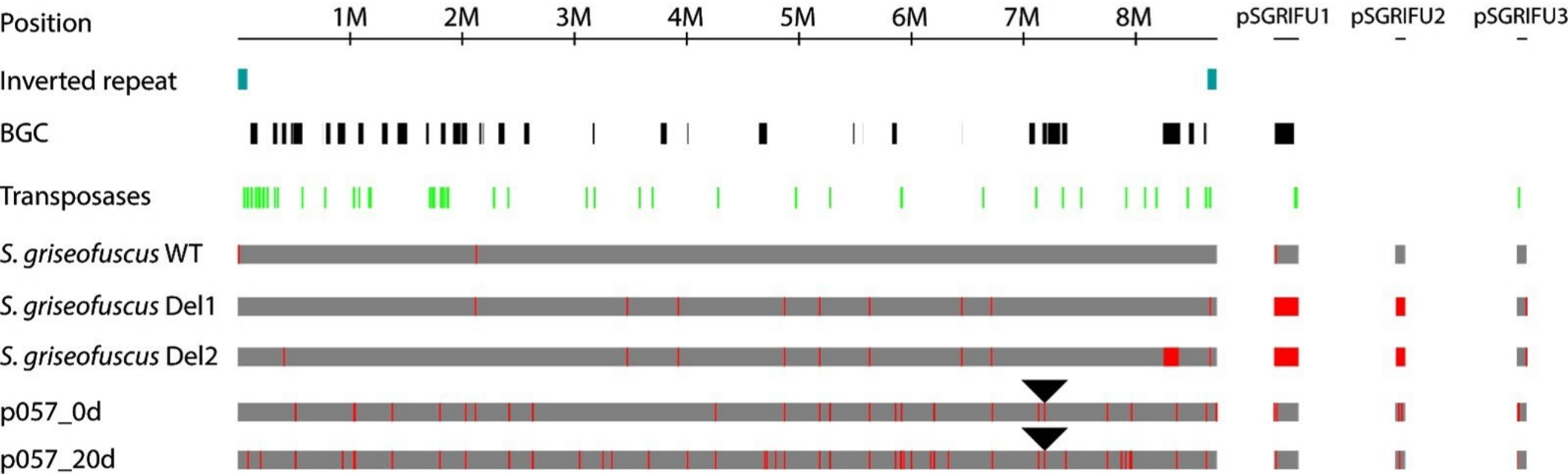
892 related genera by bacteriophage FP43. J Bacteriol 170:2276–2282.

893 29. Li S, Tang L, Chen X, Liao L, Li F, Mao Z. 2011. Isolation and characterization of a novel ε-
894 poly-l-lysine producing strain: Streptomyces griseofuscus. J Ind Microbiol Biotechnol 38:557–
895 563.

896 30. Li S, Chen X, Dong C, Zhao F, Tang L, Mao Z. 2013. Combining Genome Shuffling and
897 Interspecific Hybridization Among Streptomyces Improved ε-Poly-l-Lysine Production. Appl
898 Biochem Biotechnol 169:338–350.

899 31. Li S, Ji J, Hu S, Chen G. 2020. Enhancement of ε-poly-L-lysine production in Streptomyces
900 griseofuscus by addition of exogenous astaxanthin. Bioprocess Biosyst Eng 43:1813–1821.

901 32. Lacalle RA, Tercero JA, Jiménez A. 1992. Cloning of the complete biosynthetic gene cluster for
902 an aminonucleoside antibiotic, puromycin, and its regulated expression in heterologous hosts.
903 EMBO J 11:785–792.

904 33. Wang B, Guo F, Huang C, Zhao H. 2020. Unraveling the iterative type I polyketide synthases
905 hidden in Streptomyces. Proc Natl Acad Sci U S A 117:8449–8454.

906 34. Gren T, Jørgensen T, Whitford C, Weber T. 2020. High-quality sequencing, assembly and
907 annotation of the Streptomyces griseofuscus DSM 40191 genome. Microbiol Resour Announc
908 Press.

909 35. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome
910 annotation and pathway reconstruction server. Nucleic Acids Res 35:W182–W185.

911 36. Koepff J, Keller M, Tsolis KC, Busche T, Rückert C, Hamed MB, Anné J, Kalinowski J,
912 Wiechert W, Economou A, Oldiges M. 2017. Fast and reliable strain characterization of
913 Streptomyces lividans through micro-scale cultivation. Biotechnol Bioeng 114:2011–2022.

914 37. Galardini M, Mengoni A, Biondi EG, Semeraro R, Florio A, Bazzicalupo M, Benedetti A,
915 Mocali S. 2014. DuctApe: A suite for the analysis and correlation of genomic and OmniLog^{TM}
916 Phenotype Microarray data. Genomics 103:1–10.

917 38. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi
918 YH, Claesen J, Coates RC, Cruz-Morales P, Duddela S, Düsterhus S, Edwards DJ, Fewer DP,
919 Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJN,
920 Hillwig ML, Ishida K, Jones AC, Jones CS, Jungmann K, Kegler C, Kim HU, Kötter P, Krug D,
921 Masschelein J, Melnik AV, Mantovani SM, Monroe EA, Moore M, Moss N, Nützmann H-W,
922 Pan G, Pati A, Petras D, Reen FJ, Rosconi F, Rui Z, Tian Z, Tobias NJ, Tsunematsu Y,
923 Wiemann P, Wyckoff E, Yan X, Yim G, Yu F, Xie Y, Aigle B, Apel AK, Balibar CJ, Balskus
924 EP, Barona-Gómez F, Bechthold A, Bode HB, Borriss R, Brady SF, Brakhage AA, Caffrey P,
925 Cheng Y-Q, Clardy J, Cox RJ, De Mot R, Donadio S, Donia MS, van der Donk WA, Dorrestein
926 PC, Doyle S, Driessen AJM, Ehling-Schulz M, Entian K-D, Fischbach MA, Gerwick L,
927 Gerwick WH, Gross H, Gust B, Hertweck C, Höfte M, Jensen SE, Ju J, Katz L, Kaysser L,
928 Klassen JL, Keller NP, Kormanec J, Kuipers OP, Kuzuyama T, Kyrpides NC, Kwon H-J, Lautru
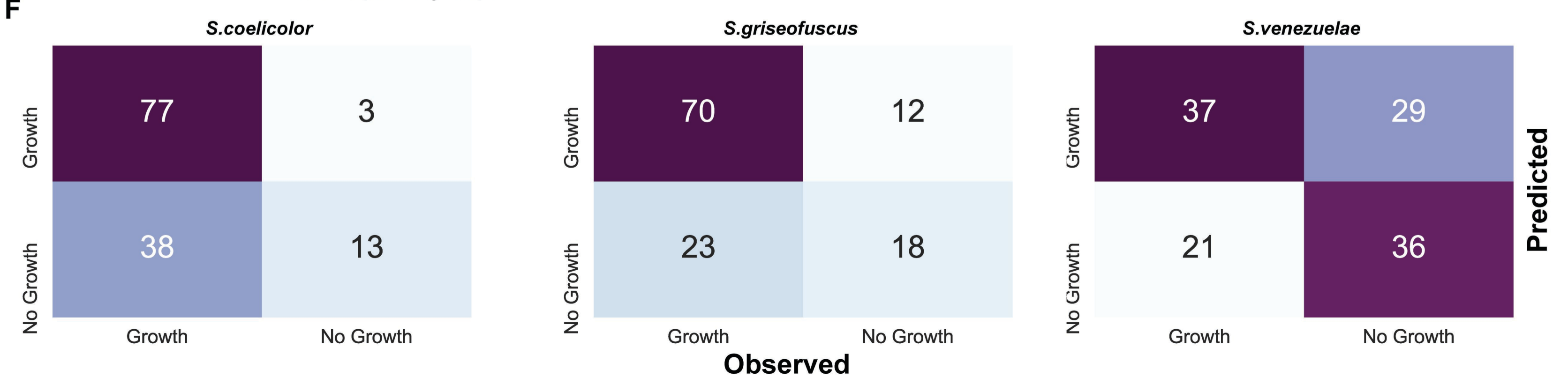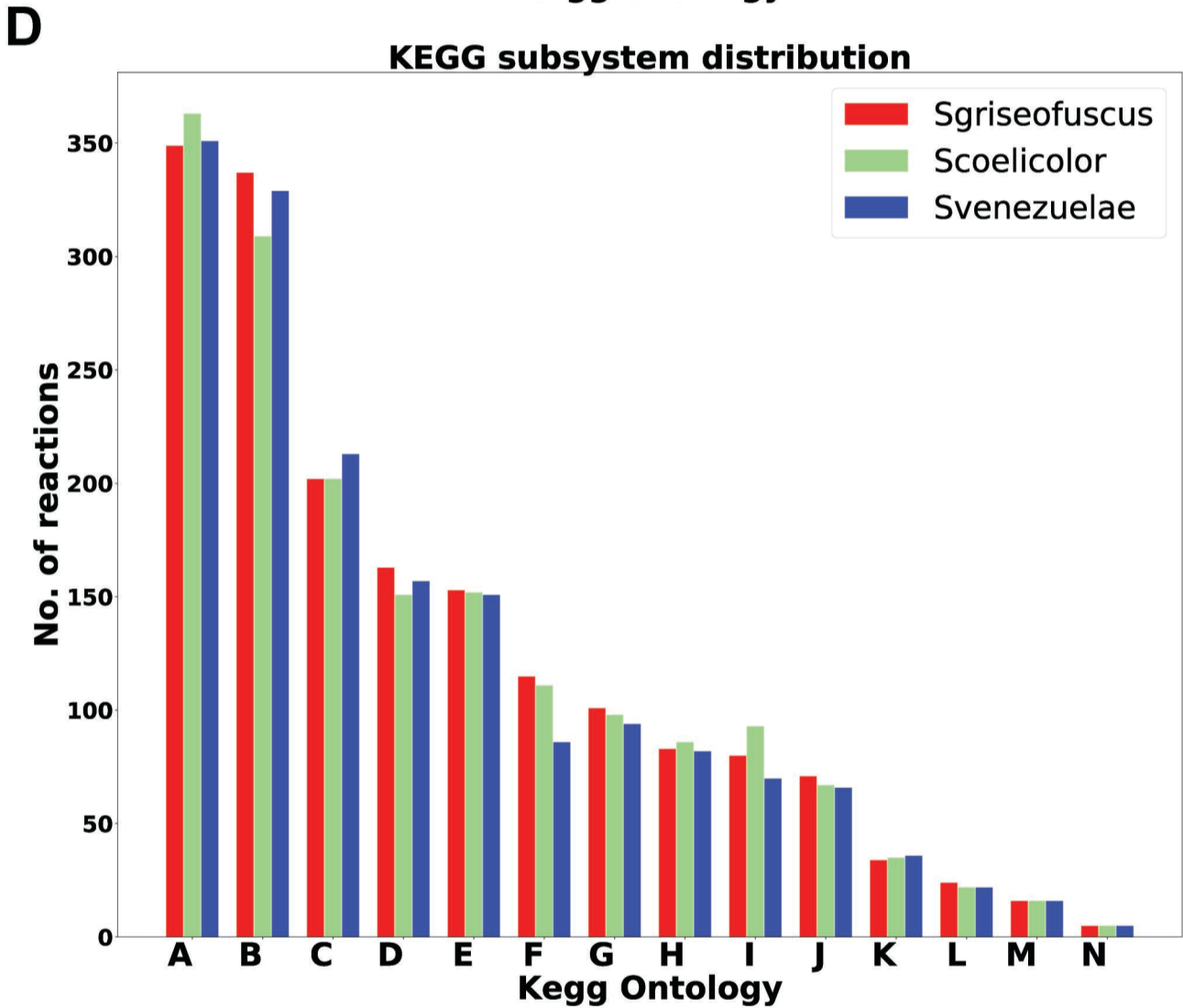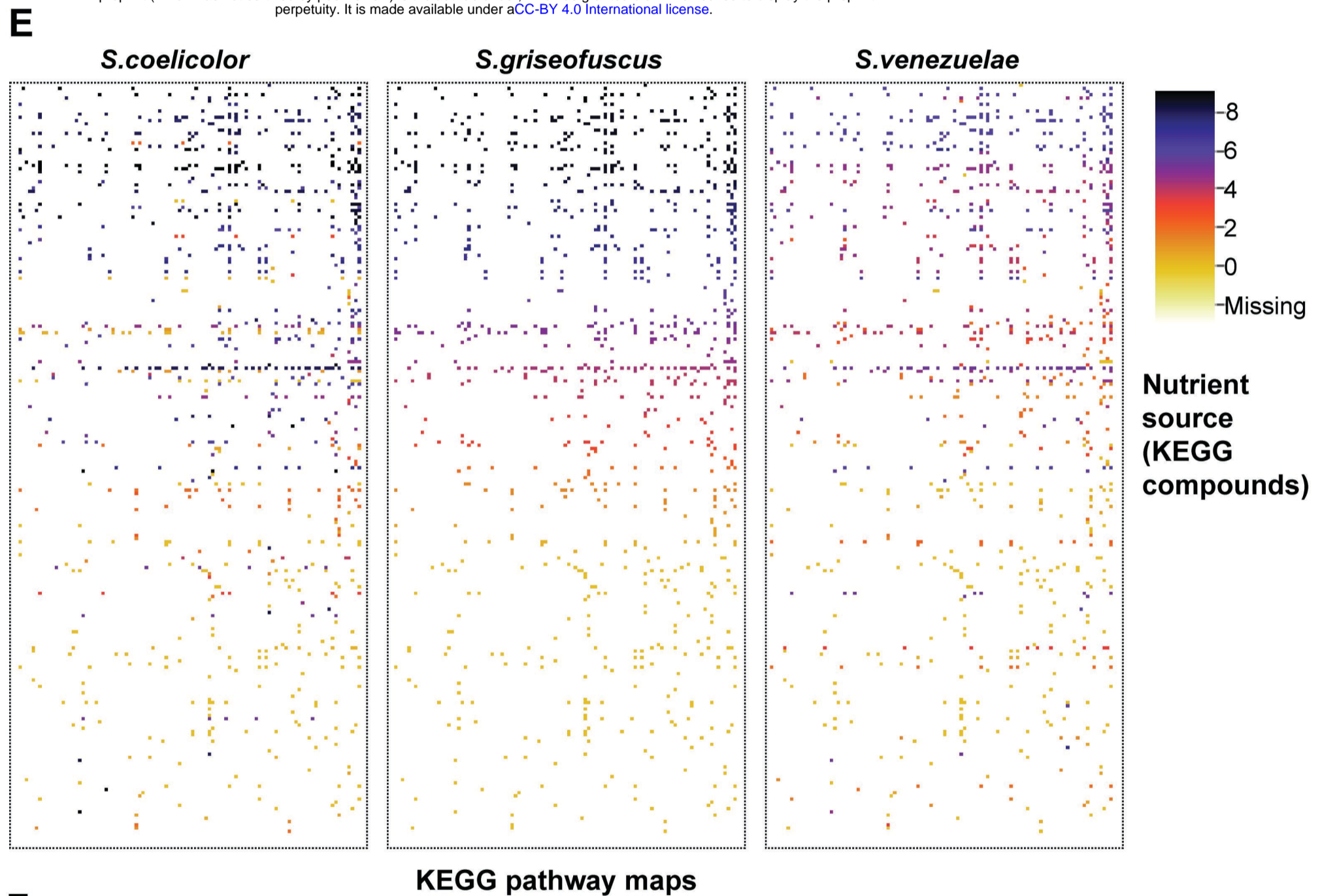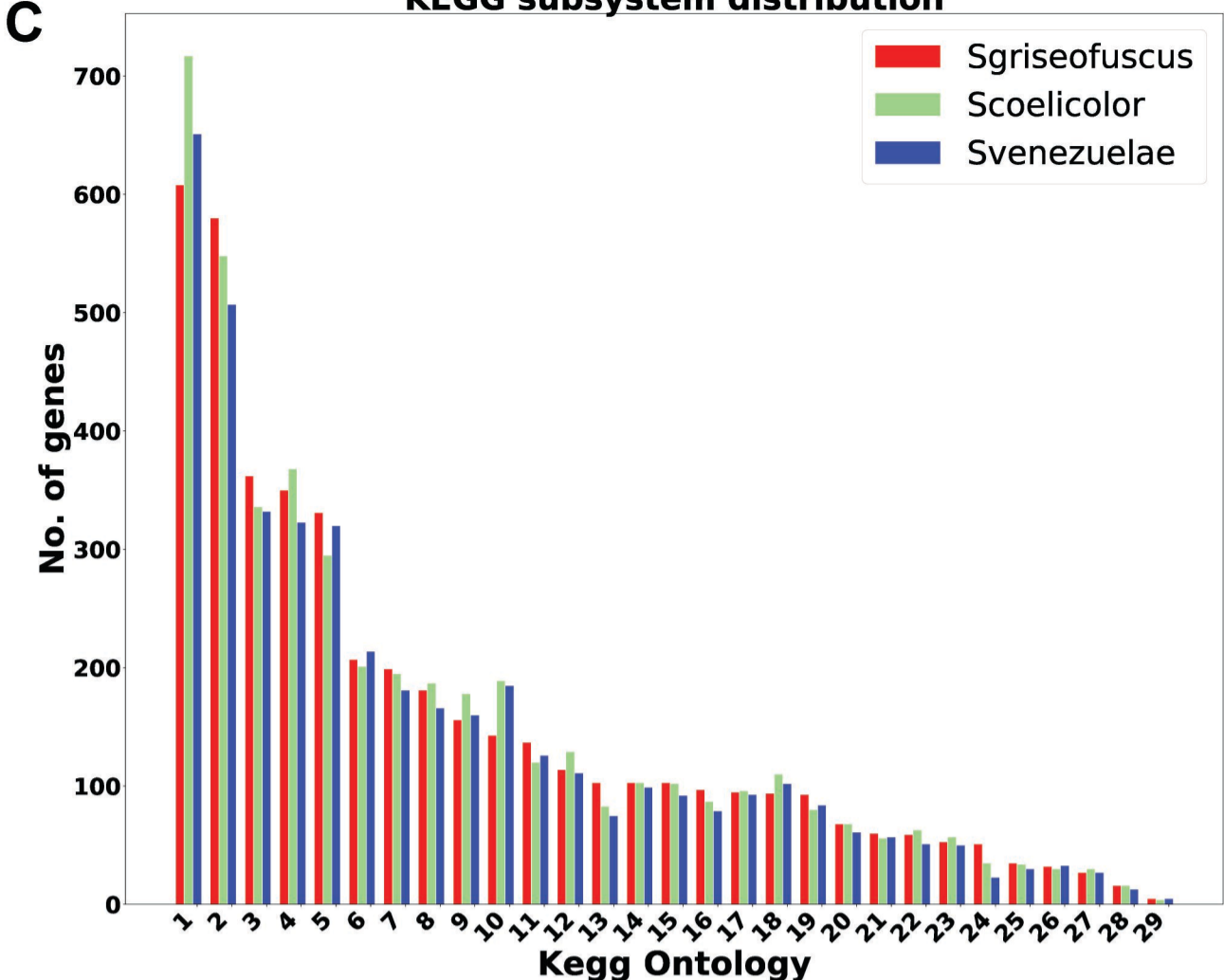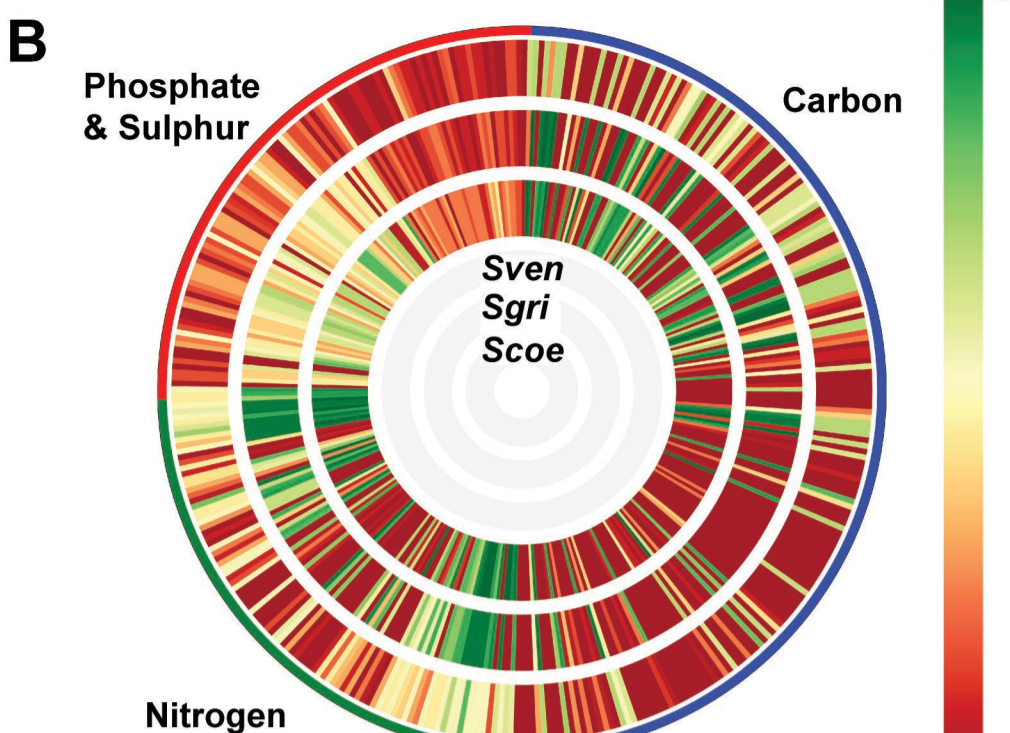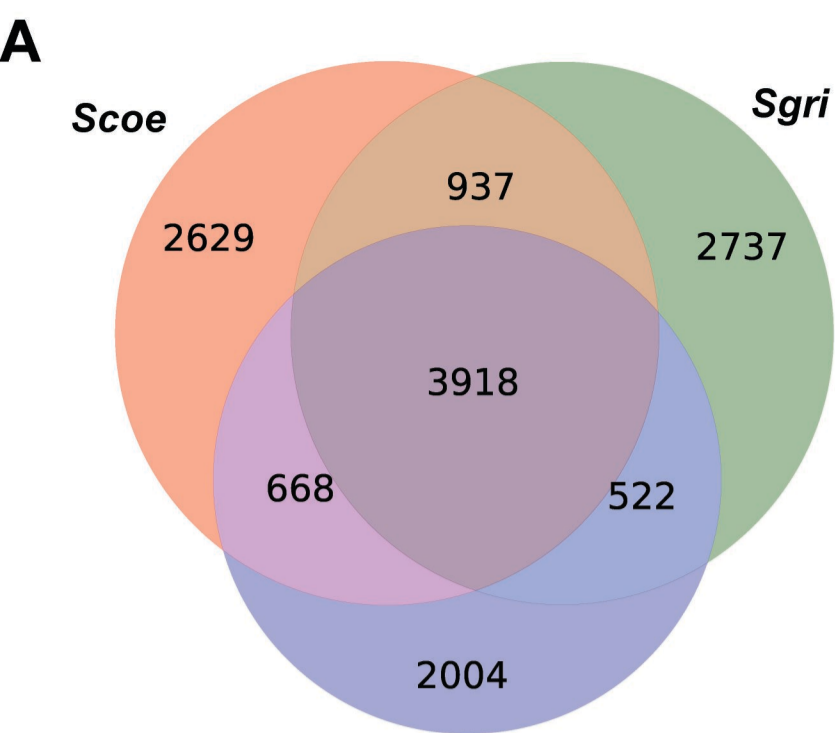
929   S, Lavigne R, Lee CY, Linquan B, Liu X, Liu W, Luzhetskyy A, Mahmud T, Mast Y, Méndez

930   C, Metsä-Ketelä M, Micklefield J, Mitchell DA, Moore BS, Moreira LM, Müller R, Neilan BA,

931   Nett M, Nielsen J, O'Gara F, Oikawa H, Osbourn A, Osburne MS, Ostash B, Payne SM,

932   Pernodet J-L, Petricek M, Piel J, Ploux O, Raaijmakers JM, Salas JA, Schmitt EK, Scott B,

933   Seipke RF, Shen B, Sherman DH, Sivonen K, Smanski MJ, Sosio M, Stegmann E, Süssmuth

934   RD, Tahlan K, Thomas CM, Tang Y, Truman AW, Viaud M, Walton JD, Walsh CT, Weber T,

935   van Wezel GP, Wilkinson B, Willey JM, Wohlleben W, Wright GD, Ziemert N, Zhang C,

936   Zotchev SB, Breitling R, Takano E, Glöckner FO. 2015. Minimum Information about a

937   Biosynthetic Gene cluster. 9. Nat Chem Biol 11:625–631.

938 39. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput

939   ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. 1. Nat Commun

940   9:5114.

941 40. Cao Z, Yoshida R, Kinashi H, Arakawa K. 2015. Blockage of the early step of lankacidin

942   biosynthesis caused a large production of pentamycin, citreodiol and epi -citreodiol in

943   Streptomyces rochei. 5. J Antibiot (Tokyo) 68:328–333.

944 41. Nindita Y, Cao Z, Fauzi AA, Teshima A, Misaki Y, Muslimin R, Yang Y, Shiwa Y, Yoshikawa

945   H, Tagami M, Lezhava A, Ishikawa J, Kuroda M, Sekizuka T, Inada K, Kinashi H, Arakawa K.

946   2019. The genome sequence of Streptomyces rochei 7434AN4, which carries a linear

947   chromosome and three characteristic linear plasmids. Sci Rep 9:10973.

948 42. Arakawa K. 2014. Genetic and biochemical analysis of the antibiotic biosynthetic gene clusters

949   on the *Streptomyces* linear plasmid. Biosci Biotechnol Biochem 78:183–189.

950 43. Nagaoka K, Matsumoto M, Oono J, Yokoi K, Ishizeki S, Nakashima T. 1986. Azinomycins A

951   and B, new antitumor antibiotics. I. Producing organism, fermentation, isolation, and

952   characterization. J Antibiot (Tokyo) 39:1527–1532.

953 44. Zhang J, Marcin C, Shifflet MA, Salmon P, Brix T, Greasham R, Buckland B, Chartrain M.

954   1996. Development of a defined medium fermentation process for physostigmine production by

955   Streptomyces griseofuscus. Appl Microbiol Biotechnol 44:568–575.

956 45. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. Practical streptomyces

957   genetics. John Innes Foundation Norwich.

958 46.   Gym Streptomyces medium.

959 47. Fröjd MJ, Flärdh K. 2019. Extrusion of extracellular membrane vesicles from hyphal tips of

960   Streptomyces venezuelae coupled to cell-wall stress. Microbiology 165:1295–1305.

961 48. Hoefler BC, Stubbendieck RM, Josyula NK, Moisan SM, Schulze EM, Straight PD. 2017. A

962   Link between Linearmycin Biosynthesis and Extracellular Vesicle Genesis Connects

963   Specialized Metabolism and Bacterial Membrane Physiology. Cell Chem Biol 24:1238-1249.e7.

964 49. Schrempf H, Merling P. 2015. Extracellular Streptomyces lividans vesicles: composition,

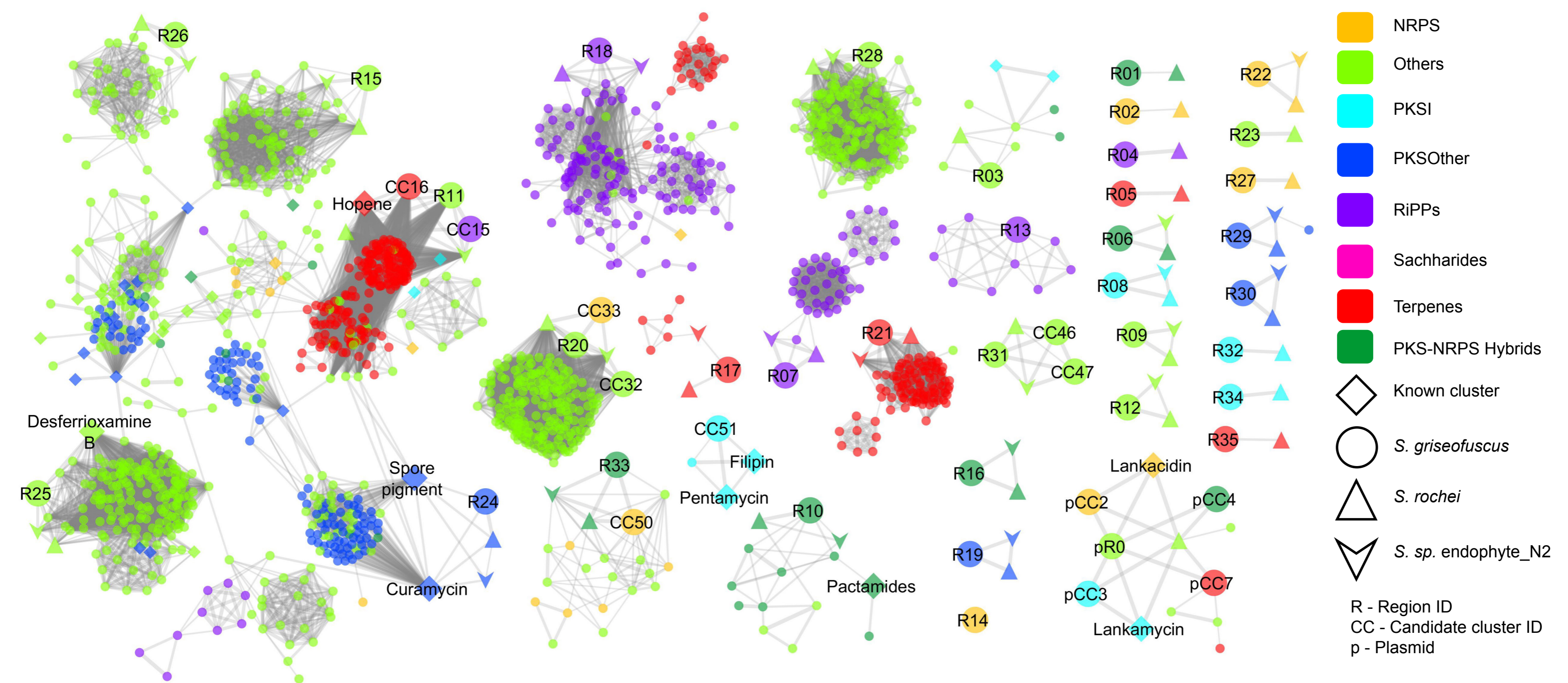965   biogenesis and antimicrobial activity. Microb Biotechnol 8:644–658.

966 50. Tercero JA, Lacalle RA, Jiménez A. 1992. Cosmid pJAR4, a novel Streptomyces-Escherichia
967   coli shuttle vector for the cloning of Streptomyces operons. FEMS Microbiol Lett 75:203–206.
968 51. Myronovskyi M, Welle E, Fedorenko V, Luzhetskyy A. 2011. Beta-glucuronidase as a sensitive
969   and versatile reporter in actinomycetes. Appl Environ Microbiol 77:5370–5383.
970 52. Gren T, Ortseifen V, Wibberg D, Schneiker-Bekel S, Bednarz H, Niehaus K, Zemke T, Persicke
971   M, Pühler A, Kalinowski J. 2016. Genetic engineering in Actinoplanes sp. SE50/110 -
972   development of an intergeneric conjugation system for the introduction of actinophage-based
973   integrative vectors. J Biotechnol 232:79–88.
974 53. Baltz RH. 2012. Streptomyces temperate bacteriophage integration systems for stable genetic
975   engineering of actinomycetes (and other organisms). J Ind Microbiol Biotechnol 39:661–672.
976 54. Tong Y, Weber T, Lee SY. 2019. CRISPR/Cas-based genome engineering in natural product
977   discovery. Nat Prod Rep 36:1262–1280.
978 55. Tong Y, Charusanti P, Zhang L, Weber T, Lee SY. 2015. CRISPR-Cas9 Based Engineering of
979   Actinomycetal Genomes. ACS Synth Biol 4:1020–1029.
980 56. Nindita Y, Cao Z, Yang Y, Arakawa K, Shiwa Y, Yoshikawa H, Tagami M, Lezhava A,
981   Kinashi H. 2015. The tap-tpg gene pair on the linear plasmid functions to maintain a linear
982   topology of the chromosome in Streptomyces rochei. Mol Microbiol 95:846–858.
983 57. Tong Y, Whitford CM, Robertsen HL, Blin K, Jørgensen TS, Klitgaard AK, Gren T, Jiang X,
984   Weber T, Lee SY. 2019. Highly efficient DSB-free base editing for streptomycetes with
985   CRISPR-BEST. Proc Natl Acad Sci U S A 116:20366–20375.
986 58. Hoff G, Bertrand C, Piotrowski E, Thibessard A, Leblond P. 2018. Genome plasticity is
987   governed by double strand break DNA repair in Streptomyces. Sci Rep 8:5272.
988 59. Tong Y, Whitford CM, Blin K, Jørgensen TS, Weber T, Lee SY. 2020. CRISPR-Cas9, CRISPRi
989   and CRISPR-BEST-mediated genetic manipulation in streptomycetes. Nat Protoc 15:2470–
990   2502.
991 60. Ye S, Enghiad B, Zhao H, Takano E. 2020. Fine-tuning the regulation of Cas9 expression levels
992   for efficient CRISPR-Cas9 mediated recombination in Streptomyces. J Ind Microbiol
993   Biotechnol 47:413–423.
994 61.   2013. Copyright, p. ii. In Doran, PM (ed.), Bioprocess Engineering Principles (Second Edition).
995   Academic Press, London.
996 62. Blin K, Pedersen LE, Weber T, Lee SY. 2016. CRISPy-web: An online resource to design
997   sgRNAs for CRISPR applications. Synth Syst Biotechnol 1:118–121.
998 63. Blin K, Shaw S, Tong Y, Weber T. 2020. Designing sgRNAs for CRISPR-BEST base editing
999   applications with CRISPy-web 2.0. Synth Syst Biotechnol 5:99–102.
1000 64. Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming,
1001   identification, and read merging. BMC Res Notes 9:88.
1002 65. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using

1003   repeat graphs. Nat Biotechnol 37:540–546.

1004   66.   Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome

1005   assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.

1006   67.   Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–

1007   3100.

1008   68.   Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. 4. Nat Methods

1009   9:357–359.

1010   69.   Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated

1011   platform for visualization and analysis of high-throughput sequence-based experimental data.

1012   Bioinformatics 28:464–469.

1013   70.   Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and

1014   Annotation Completeness, p. 227–245. *In* Kollmar, M (ed.), Gene Prediction: Methods and

1015   Protocols. Springer, New York, NY.

1016   71.   Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo

1017   genome assemblies. Bioinforma Oxf Engl 31:3350–3352.

1018   72.   Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal:

1019   prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics

1020   11:119.

1021   73.   Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl

1022   30:2068–2069.

1023   74.   El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ,

1024   Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn

1025   RD. 2019. The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432.

1026   75.   Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer:

1027   consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108.

1028   76.   Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in

1029   nucleotide sequences. Nucleic Acids Res 32:11–16.

1030   77.   Wick RR, Judd LM, Holt KE. 2018. Deepbinner: Demultiplexing barcoded Oxford Nanopore

1031   reads with deep convolutional neural networks. PLOS Comput Biol 14:e1006583.

1032   78.   Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N,

1033   Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, Machi D, Mao C,

1034   Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC, Overbeek R,

1035   Parrello B, Pusch GD, Shukla M, Thomas C, VanOeffelen M, Vonstein V, Warren AS, Xia F,

1036   Xie D, Yoo H, Stevens R. 2020. The PATRIC Bioinformatics Resource Center: expanding data

1037   and analysis capabilities. Nucleic Acids Res 48:D606–D612.

1038   79.   Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI,

1039   De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, Roeters A, Lokhorst W,

1040   Fernandez-Guerra A, Cappelini LTD, Goering AW, Thomson RJ, Metcalf WW, Kelleher NL,

1041   Barona-Gomez F, Medema MH. 2020. A computational framework to explore large-scale

1042   biosynthetic diversity. Nat Chem Biol 16:60–68.

| Position | 1M | 2M | 3M | 4M | 5M | 6M | 7M | 8M | pSGRIFU1 | pSGRIFU2 | pSGRIFU3 |

Inverted repeat

BGC

Transposases

S. griseofuscus WT

S. griseofuscus Del1

S. griseofuscus Del2

p057_0d

p057_20d

A

B

Phosphate & Sulphur

Carbon

Nitrogen

Sven
Sgri
Scoe

C

KEGG subsystem distribution

No. of genes

Kegg Ontology

Sgriseofuscus
Scoelicolor
Svenezuelae

D

KEGG subsystem distribution

No. of reactions

Kegg Ontology

Sgriseofuscus
Scoelicolor
Svenezuelae

E

S.coelicolor

S.griseofuscus

S.venezuelae

Nutrient source (KEGG compounds)

KEGG pathway maps

F

S.coelicolor

Predicted

	Growth	No Growth
Growth	77	3
No Growth	38	13

S.griseofuscus

	Growth	No Growth
Growth	70	12
No Growth	23	18

S.venezuelae

	Growth	No Growth
Growth	37	29
No Growth	21	36

Observed

	NRPS
	Others
	PKSI
	PKSOther
	RiPPs
	Sachharides
	Terpenes
	PKS-NRPS Hybrids
◇	Known cluster
○	*S. griseofuscus*
△	*S. rochei*
∨	*S. sp.* endophyte_N2

R - Region ID
CC - Candidate cluster ID
p - Plasmid

WT

1

From \ To	G	C	T	A
A	0	1	0	0
T	0	0	0	0
C	0	0	0	0
G	0	2	0	0

p057_0D

2

From \ To	G	C	T	A
A	6	1	1	0
T	2	4	0	1
C	7	0	3	0
G	0	6	0	2

p057_20D

3

From \ To	G	C	T	A
A	6	1	1	0
T	2	4	0	1
C	6	0	13	1
G	0	3	1	11

0
4
8
12