bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TITLE

Genomic adaptation in the CAZyome and specialised metabolism of the plant-associated *Streptomyces violaceusniger* clade

AUTHORS

Gayrard D.[1,2], Veyssière M.[1], Adam K[1]., Martinez Y.[3], Vandecasteele C.[4], Vidal M.[4], Dumas B.[1*], Rey T.[1,2*]

CORRESPONDING AUTHOR'S EMAIL

reyt@desangosse.com and dumas@lrsv.ups-tlse.fr

AFFILIATIONS

[1]Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, Toulouse INP, Auzeville-Tolosane, France

[2]De Sangosse, Bonnel, Pont-Du-Casse, France

[3]CNRS, Plateforme Imagerie-Microscopie, Fédération de Recherche FR3450, Castanet-Tolosan, France

[4]INRAE, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France

*corresponding authors

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ABSTRACT

Streptomycetes are Gram-positive actinobacteria largely represented in the plant root microbiota. The genetic determinants involved in the presence of *Streptomyces* in the rhizosphere are largely unknown and can rely on the ability to degrade plant-derived compounds such as cell-wall polysaccharides and on the production of specialised metabolites. To address whether *Streptomyces* strains recruited into root microbiota share genomic specificities related to these two functions, we engaged a comparative genomic analysis using a newly sequenced rhizospheric strain, *Streptomyces* sp. AgN23 and strains from the phylogenetically related *S. violaceusniger* clade. This analysis enlightens a shared prominent CAZyome potentially involved in plant polysaccharides degradation and a strong conservation of antimicrobials biosynthetic clusters (rustmicin, mediomycin, niphimycin, nigericin) as well as plant bioactive compounds (nigericin, echosides, elaiophylin). Taken together, our work supports the hypothesis that specific hydrolytic enzymes and specialised metabolites repertoires may play important roles in the development of *Streptomyces* strains in the rhizosphere.

KEYWORDS

Streptomyces, Long-read sequencing, Biosynthetic Gene Clusters, Phylogenomic, Specialised Metabolites, Transcriptome

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Streptomycetes are aerobic and Gram-positive actinobacteria forming branched vegetative mycelium before developing aerial hyphae bearing spores [1]. These bacteria received considerable attention from a biotechnological point of view, notably regarding their enzymatic repertoire [2] and in the drug discovery field, leading to the structure elucidation of more than 6000 specialised metabolites [3,4]. Their tremendous ability to produce antimicrobial compounds relies on the wealth and diversity of Biosynthetic Gene Clusters (BGCs) encoded in their genomes [5]. The recent soaring of microbial metabarcoding approaches have highlighted *Streptomyces* spp. prominent abundance in plant roots microbiota [6,7]. Since several of these metabolites display strong antimicrobial activities, *Streptomyces* recruitment in the rhizosphere and inside root tissues presumably protect the host from pathogens. Thus, several *Streptomyces* based products have been developed for agriculture, but these strains only cover a subset of plant colonising *Streptomyces* species [8,9]. Notwithstanding, little is known regarding the biological function of these bacteria in the plant environment and the gene families involved in their adaptation to this ecological niche [10,11]. More than 100 hundred fully assembled genome sequences of *Streptomyces* strains are currently available, paving the way to whole genome-based phylogeny and comparisons of BGCs content across Streptomyces clades [12-14]. This knowledge open inroads to rationalise and investigate the potential use of *Streptomyces* strains in agriculture.

Here we report a gapless assembly of *Streptomyces* sp. AgN23 (AgN23), previously isolated from grapevine rhizosphere and identified as a strong inducer of plant defences [15]. We used this high-quality assembly to position AgN23 in the *Streptomyces violaceusniger* genomospecies. We then dissected the original features in the CAZyome and BGC content of AgN23 and other *Streptomyces violaceusniger* strains and shed light on specificities in the specialised metabolism and carbohydrate degradation capabilities of this lineage. Our data reflect the high potential of *Streptomyces* species of the *S. violaceusniger* clade to interact with plants and protect them from fungal pathogens.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RESULTS AND DISCUSSION

Chromosome scale assembly of AgN23 and genome based taxonomic assignation to the *S. violaceusniger* clade

We performed a PacBio® RSII long-read sequencing and obtained a linear chromosome of 10,86 Mb for AgN23. This genome sequence was polished using Illumina® MiSeq sequences to produce the final assembly. The final assembly was uploaded on the MicroScope Platform[16]. The genome annotation retrieved 10,514 protein coding sequences, 10,458 of them being supported by RNA-seq (Supplementary Information 1). checkM analysis was performed using 455 genomes and 315 lineage-specific markers and validated the completeness of the assembly and the annotation [17]. Moreover, its completeness was also supported by BUSCO analysis [18].

Table 1: Summary of the assembly and the annotation of *Streptomyces* sp. AgN23 complete chromosome obtained by PacBio® and Illumina® sequencing.

Assembly Statistics	***Streptomyces* sp. AgN23**
Genome size (bp)	10 858 739
Coverage (PacBio®)	77X
G+C content (%)	70.9
Number of protein coding genes	10,514
CheckM Completeness	100%
Contamination	1,84%
BUSCO	451/452
CAZymes	278
Number BGCs (antiSMASH 5.0)	45

More than 500 *Streptomyces* species have been described based on their 16S rRNA sequence. Previous sequencing of AgN23 16S rRNA showed strongest conservation with representatives of the *S. violaceusniger* clade, most notably *S. castelarensis* [15]. However, lack of variation in 16S rRNA may confound strains belonging to different species [19]. Recent development of long-read technologies and massive sequencing of *Streptomyces* has leveraged whole genome based phylogenies [5]. Thus, we decided to consolidate the taxonomic affiliation of AgN23 with the gapless assembly of AgN23 chromosome. We performed an autoMLST approach based on publicly available *Streptomyces* sequences. In brief, 63 conserved housekeeping genes showing neutral dN/dS were concatenated and

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

aligned as a basis for tree building (Figure 1)[20]. As a result, using AgN23 as query, we affiliated the strain to a clade containing three other stains showing Average Nucleotide Identity higher than 95% with AgN23 which can thus be considered as belonging to the same species, namely *S. melanosporofaciens*, *S. antimycoticus* [21,22] and *S. violaceusniger* NRRL F-8817. In total, 24 isolates harboured ANI>85% (Supplementary Information 2).

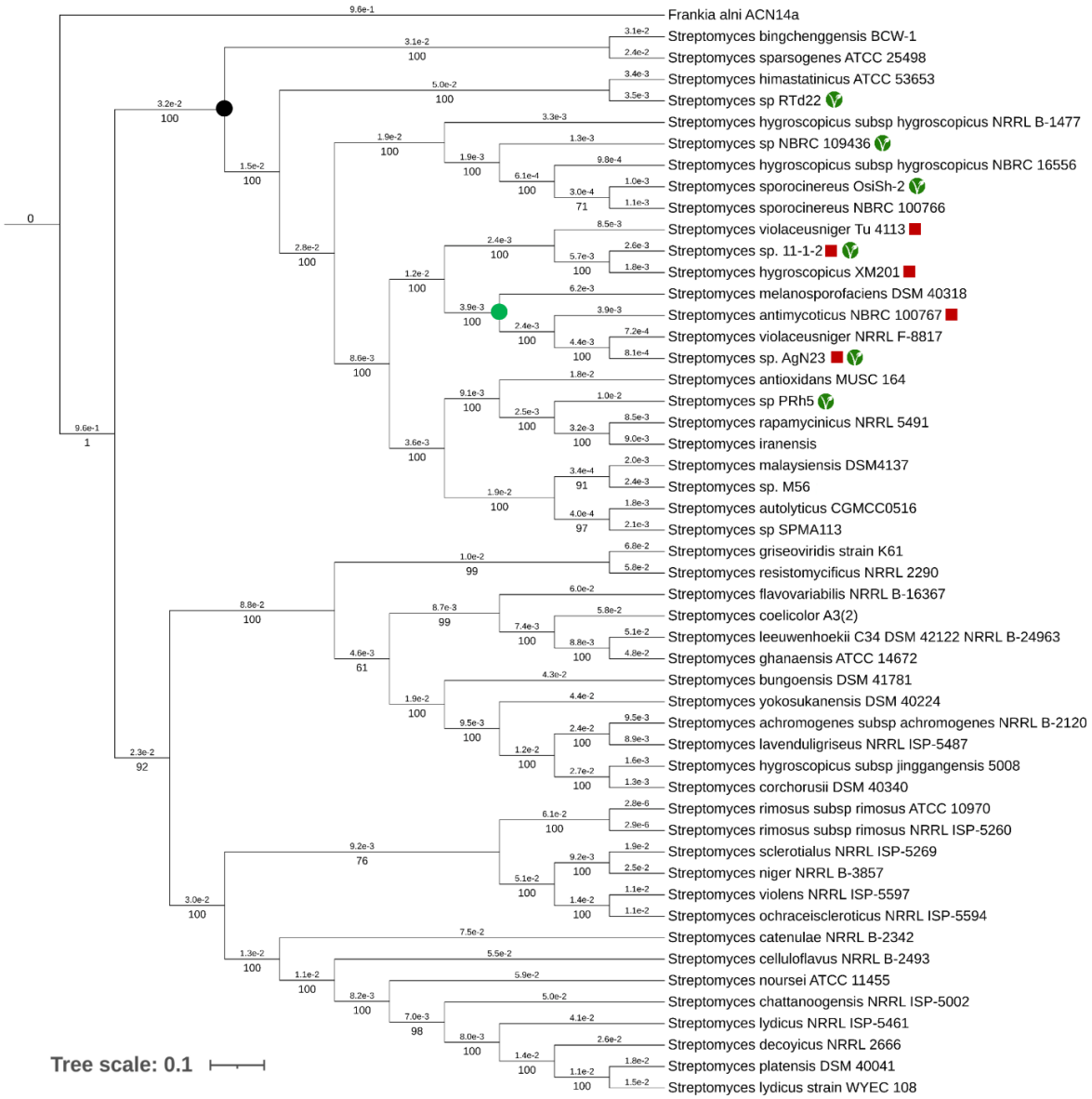
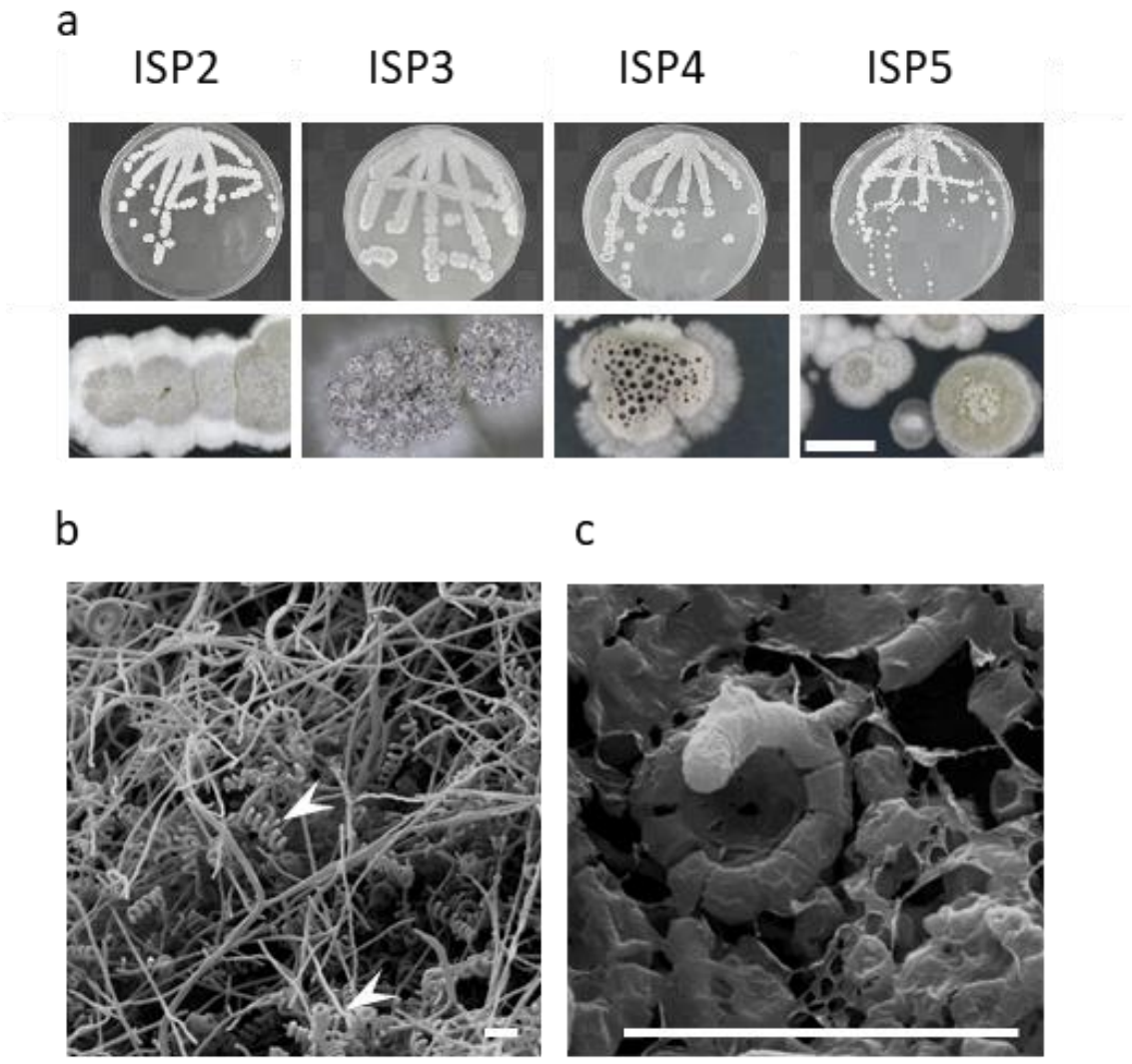


Figure 1: Multi Locus Sequence Typing assigned AgN23 to the *S. violaceusniger* clade. Phylogenetic tree based on genomic sequences with AutoMLST. The green node highlights the isolates considered to be from the same species (ANI>95%). The black node highlights the clade formed by isolates with ANI>85% as compared to AgN23. The red squares highlighted the five strains that were used for the BGC conservation study. The green logo indicates plant-isolated strains. *Frankia alni* ACN14a was used as outgroup, bootstrap=100.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

This group of AgN23-related strains contains hallmark representative of the *S. violaceusniger* clade, such as, *S. hygroscopicus*, *S. sparsogenes*, *S. malaysiensis* [23], *S. himastatinicus, S. rapamycinicus,* and other close species *S. autolyticus* [24], *S. antioxidans* [25] and *S. iranensis* [26]. Interestingly, *Streptomyces* strains RT-d22 [27], *Streptomyces* sp. Strain PRh5 [28], *Streptomyces* sp. 11-1-2 [29], *S. hygroscopicus* Osish-2 [30-34] and *Streptomyces* sp. NBRC 109436 [35] were isolated from the rhizosphere of diverse plants across the world. It suggests that interaction with plants is widespread among the strains belonging to the *S. violaceusniger* clade. Finally, cultivation of AgN23 on a range of ISP media confirmed that AgN23 display typical phenotypes of S. *violaceusniger* clade with whitish colony then turning grey during the sporulation process and resulting in the formation of spiralled chains of rugose-ornamented spores (Figure 2) [36-38].



bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2: AgN23 harbours typical phenotypic features the *S. violaceusniger* clade. a) AgN23 forms a white mycelium turning grey at the onset of sporulation is observed on a range of ISP media, the Petri plate are 9 cm in diameter and the scale bar is 1mm. b) Scanning Electron Microscopy observation of spiralled chains of spores (white arrow) are formed by AgN23 (scale bar = 6µm). c) Scanning Electron Microscopy observation of AgN23 spores showing a rugose-ornamented surface (scale bar = 6µm).

AgN23 exhibits a wide repertoire of CAZymes related to plant and fungi cell wall degradation

Streptomycetes display extensive abilities to degrade polysaccharides based on their rich repertoires of Carbohydrate-Active Enzymes (CAZymes) [2,39]. CAZymes may represent a major advantage to get access to carbon sources derived from plant cell wall in the rhizosphere. We predicted AgN23 and four others *S. violaceusnige*r sp. CAZymes with dbCAN2 to detect potential CAZyome adaptation to plant cell walls. The CAZyome of the hydrolytic strain *Streptomyces* sp. SirexAA-E [2], the model strain *S. coelicolor* A3(2) and four strains belonging to the *S. violaceusniger* clade, namely *S. antimycoticus* NBRC 100767, *Streptomyces* sp. 11-1-2, *S. hygroscopicus* XM201 and *S. violaceusniger* Tü 4113 were annotated for comparison with AgN23. Noteworthy, the selection of *S. violaceusniger* strains was based on two criterions, the availability of a chromosome scale assembly and an ANI score higher than 90% with AgN23. Around 2.8% of AgN23 coding sequences were annotated by dbCAN2 as CAZymes (Supplementary Information 3). Such large repertoire of CAZymes has been previously described in the *Streptomyces* sp. SirexAA-E as a marker for high potential hydrolytic *Streptomyces* strains, notably cellulose [2]. Most of these enzymes are predicted to degrade major plant cell wall components, cellulose, hemicellulose and pectin, as well as cutin, a waxy polymer of the plant cuticle (Figure 3). We found that AgN23 and the other *S. violaceusniger* sp. possess Polysaccharide Lyase PL1 and PL9 involved in pectin degradation which are absent from *S. coelicolor* or *Streptomyces* sp. SirexAA-E. Endoglucanases related to cellulose degradation, namely AA3 (Auxilliary Activity) and GH6 (Glycoside Hydrolase) families are expanded in AgN23 and its related strains as compared to the two outgroup. A similar observation was drawn for the CE1 (Carbohydrate Esterase) and GH43 involved in Hemicellulose catabolism. Strikingly, AgN23 genome encode 9 GH43 whilst the other *S. violaceusniger* sp. all possess a single copy gene and the 2 outgroups none. Finally, cutins degradation mediated by

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

AA6 CAZymes is shared by *S. violaceusniger* sp. but not found in *S. coelicolor*. The expression level of all these CAZymes vary between Log2=8 and Log2=17 read counts in pure culture, suggesting that AgN23 constitutively express this prominent CAZyome (Figure 3). Noteworthy, the most expressed CAZymes families in AgN23 are AA3, CE1 both families being expanded in *S. violaceuniger* sp. Taken together, these data illustrate the strong potential of AgN23 to use polysaccharides derived from plant biomass as nutrient source in the rhizosphere niche.

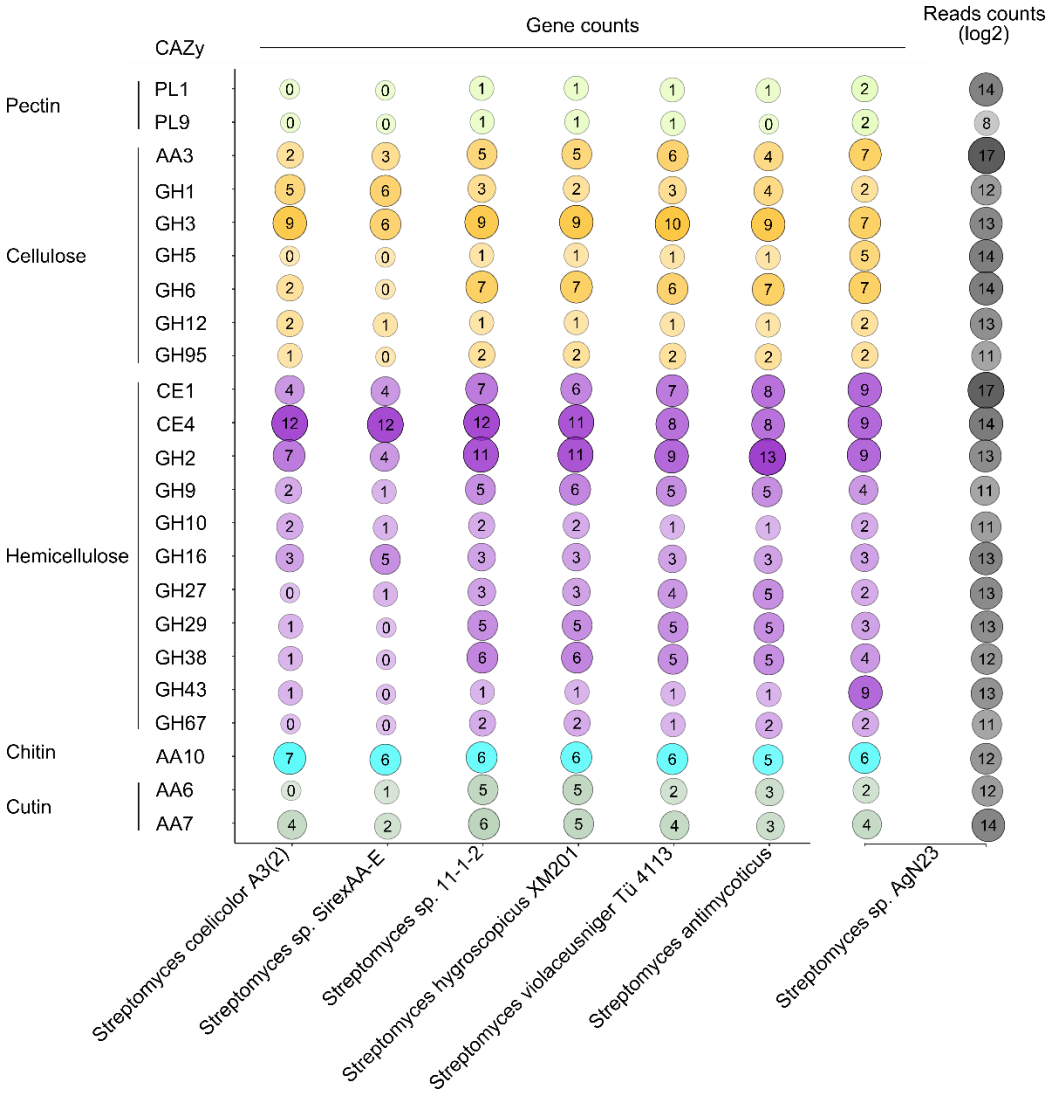


Figure 3: Plant cell wall associated CAZyome is expanded in *S. violaceusniger* sp. as compared to *S. coelicolor* and *S.* SirexAA-E. CAZymes families with at least 2 representatives in AgN23 are ordered according to their target polymers and the type of enzymatic activity: Glycoside Hydrolases (GHs), Polysaccharide Lyases (PLs), Carbohydrate Esterases (CEs) and Auxiliary Activities (AAs). The transcription of AgN23 CAZyome is displayed as sums of genes mean read counts for each family on a Log2 base (n=3).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Specialised metabolism of AgN23 may interfere with plants physiology and their microbiota

Specialised metabolites are presumed to play key roles in adaption of streptomycetes to their environment. We used antiSMASH 5.0 to detect and annotate AgN23 BGCs according to their similarity to reference clusters deposited in the MiBIG database [40,41]. Forty-Five BGCs were detected in the AgN23 genome, all of them being expressed during AgN23 cultivation according to RNA-seq analysis (Table 2, Supplementary information 4). Their expression levels range from Log2=7 up to Log2=15 reads counts for their central biosynthetic genes as defined by antiSMASH. Twenty BGCs of AgN23 showed at least 50% of similar gene content with MiBIG reference BGCs (Table 2). These candidates BGCs are notably involved in the biosynthesis of volatiles terpenes including geosmin (region 3, 10, 23), indoles (region 21), bacteriocins (region 5 and 20) and a bicyclomycin-like antibacterial (region 43). Other BGCs are potentially involved in the production of siderophores including coelichelin and desferrioxamin (region 24, 27, 42) or stress protectant (ectoin, region 16). Additional BGCs are likely involved in the regulation of the bacteria lifecycle such as spore pigment (region 30), hopene (region 31), and butyrolactone (region 36). BGCs encoded by regions 2, 18, 39 and 40 are similar to the biosynthesis pathways of the antifungal compounds rustmicin, also known as galbonolide A, mediomycin A, nigericin and niphimycins C-E respectively [42-45]. In addition, echoside (region 26), elaiophylin (region 38) and nigericin (region 39) are structural analogues of terfestatin, pteridic acid and monensin respectively, three compounds affecting plant immunity and development [46-50]. Taken together, these data reveal that AgN23 likely secrete specialised metabolites with a potential to regulate host plant biology along with its rhizosphere microbiota.

Table 2: antiSMASH annotation of AgN23 chromosomal regions coding for Biosynthetic Gene Clusters. The functional category of each BGCs was determined by antiSMASH. The BGC type, best hit in the MiBIG database as its percentage of similarity to the query are indicated along with the bacterial strain from whom the cluster was described. The expression level of each BGCs was determined by doing the mean of the reads count of the core biosynthetic genes of each BGC from the RNA-seq data, the result is expressed on a Log2 base.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Region	BGC type	Putative compound (MiBIG similarity score)	MiBIG accession	Species	Mean reads counts (Log2)
1	NRPS	atratumycin (7%)	BGC0001975	*S. atratus*	10
2	PKS-like,terpene	rustmicin (20%)	BGC0000065	*S. galbus KCCM 41354*	10
3	terpene	tiancilactone (17%)	BGC0002019	*Streptomyces sp. CB03234*	9
4	NRPS	dechlorocuracomycin (8%)	BGC0001569	*S. noursei ATCC 11455*	12
5	bacteriocin				9
6	T1PKS, hglE-KS	leinamycin (2%)	BGC0001101	*S. atroolivaceus*	11
7	NRPS				8
8	T1PKS, NRPS	meridamycin (52%)	BGC0001011	*Streptomyces sp.* NRRL	9
9	terpene	2-methylisoborneol (100%)	BGC0000658	*S. griseus* NBRC 13350	7
10	terpene	pristinol (100%)	BGC0001746	*S. pristinaespiralis* ATCC	8
11	NRPS-like, T1PKS	amipurimycin (90%)	BGC0001957	*S. novoguineensis*	12
12	T1PKS, T3PKS, NRPS, betalactone	totopotensamide (43%)	BGC0001807	*Streptomyces pactum* SCSIO 02999	9
13	lanthipeptide	lipopolysaccharide (8%)	BGC0000774	*Xanthomonas campestris*	9
14	NRPS	mildiomycin (11%)	BGC0000882	*S. rimofaciens strain*	9
15	T1PKS, siderophore	apoptolidin (23%)	BGC0000021	*Nocardiopsis sp.* FU 40	9
16	ectoine	ectoine (100%)	BGC0000853	*Streptomyces anulatus*	10
17	terpene				13
18	T1PKS	mediomycin A (88%)	BGC0001932	*Kitasatospora mediocidica*	12
19	ladderane	atratumycin (31%)	BGC0001975	*S. atratus*	11
20	NRPS	ochronotic pigment (75%)	BGC0000918	*S. avermitilis*	11
21	indole	5-isoprenylindole-3-carboxylate β-D-glycosyl	BGC0001483	*Streptomyces sp.* RM-5-8	10
22	NRPS, arylpolyene, ladderane	RP-1776 (46%)	BGC0000429	*Streptomyces sp.* Acta 2897	11
23	terpene	geosmin (100%)	BGC0001181	*S. coelicolor* A3(2)	12
24	siderophore	desferrioxamin B (100%)	BGC0000941	*S. griseus* NBRC 13350	10
25	terpene	carotenoid (63%)	BGC0000633	*S. avermitilis*	15
26	NRPS-like	echosides (100%)	BGC0000340	*Streptomyces sp.* LZ35	12
27	siderophore				8
28	bacteriocin				12
29	NRPS	formicamycins A-M (20%)	BGC0001590	*Streptomyces sp.* KY5	13
30	T2PKS	spore pigment (83%)	BGC0000271	*S. avermitilis*	13
31	terpene	hopene (76%)	BGC0000663	*S. coelicolor* A3(2)	10
32	lanthipeptide	steffimycin D (16%)	BGC0000273	*S. steffisburgensis*	10
33	other	mitomycin (18%)	BGC0000915	*S. lavendulae*	10
34	NRPS	cadaside (14%)	BGC0001968	*Uncultured bacterium*	9
35	T1PKS				10
36	butyrolactone				11
37	hserlactone	daptomycin (3%)	BGC0000336	*S. filamentosus* NRRL 11379	11
38	T1PKS	elaiophylin (87%)	BGC0000053	*Unknown*	11
39	T1PKS	nigericin (100%)	BGC0000114	*S. violaceusniger*	10
40	T1PKS	niphimycins C-E (87%)	BGC0001700	*Streptomyces sp.* IMB7-145	10
41	NRPS-like	BD-12 (67%)	BGC0001379	*S. luteocolor*	10
42	NRPS	coelichelin (90%)	BGC0000325	*S. coelicolor A3(2)*	10
43	CDPS	bicyclomycin (100%)	BGC0001468	*S. cinnamoneus*	10
44	betalactone				9
45	NRPS	acyldepsipeptide (10%)	BGC0001967	*S. hawaiiensis*	9

178

179

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A core set of biosynthetic gene clusters conserved across *S. violaceusniger* species

The *S. violaceusniger* clade hosts several strains possessing BGCs involved in the synthesis of fungicidal polyene macrolides such as nigericin, elaiophylin and geldanamycin [36,38,51-56]. However, strains sharing phylogenetic vicinity may differ in their specialised metabolism due to variation in their content of BGCs [57,58]. To assess the conservation of BGCs during species radiation among the *S. violaceusniger* clade, we set up a comparative genomic approach on fully assembled genomes belonging to *S. violaceusniger* clade.

We decided to dissect the BGCs content of AgN23 to uncover biosynthesis pathways that are shared with other *S. violaceusniger* representative, found outside of this clade or unique to the strain. We predicted BGC-containing regions from genomes of the four *S. violaceusniger* strains previously selected for CAZy comparisons. The antiSMASH outputs were introduced into BiG-SCAPE to group the 235 detected BGCs into 93 gene clusters families and to visualise their relationships by sequence similarity networks (Figure 4a).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

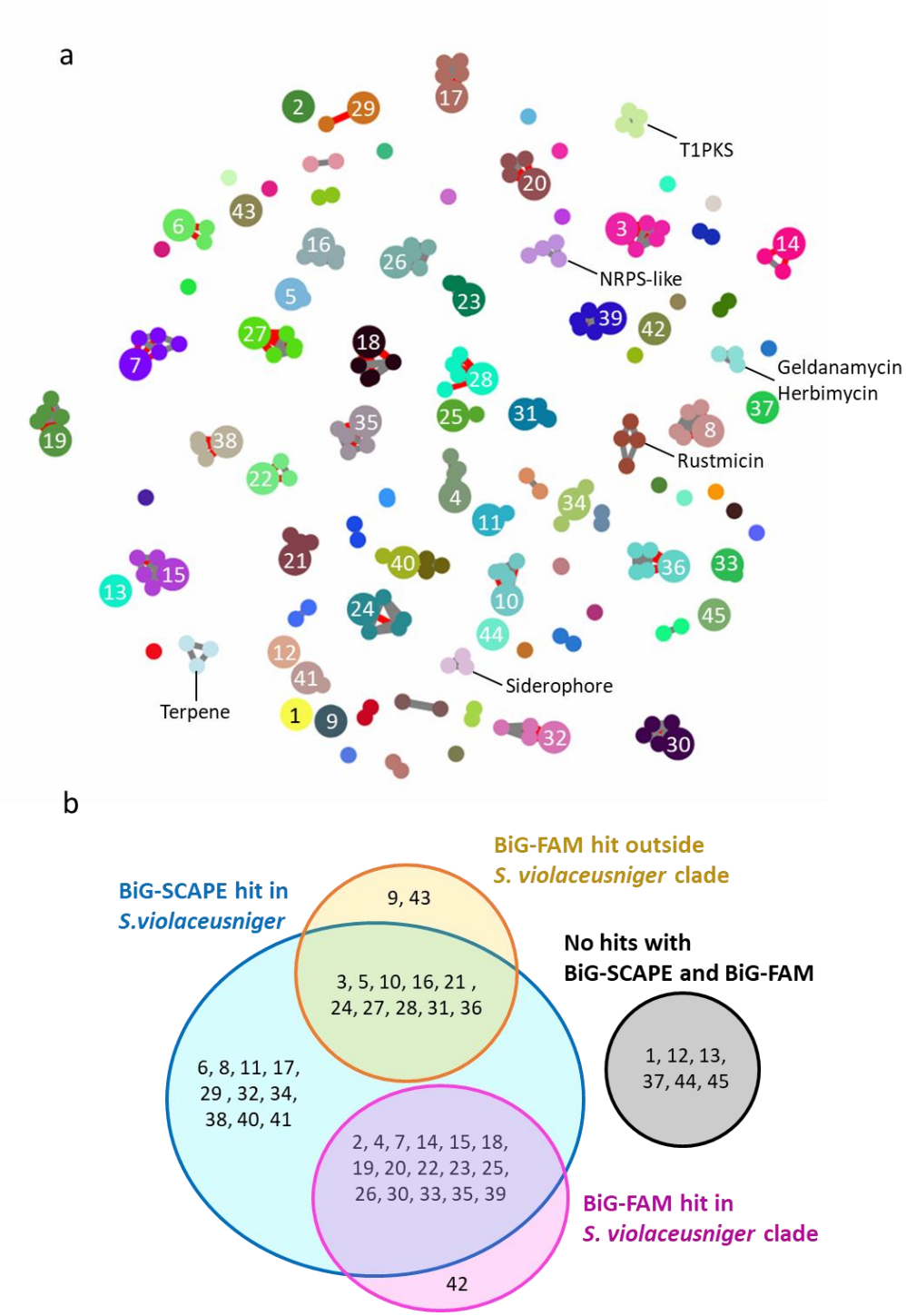


Figure 4: Twenty-six BGC-containing regions in *Streptomyces* sp. AgN23 genome are specifically found within *S. violaceusniger* strains. a) BiG-SCAPE similarity network analysis of AgN23 and four close isolates reveals that 22 BGC-containing regions are found by the given members of the *S. violaceusniger* clade. Each node represents a BGC-containing region from one of the 5 compared genomes. They are coloured according one of the 93 gene cluster families and are clustered within regions displaying similar organisation. AgN23 BGC are displayed as larger nodes harbouring its region number. b) Venn Diagram showing the combination of BiG-SCAPE and BiG-FAM analyses to distribute hits of BGC-containing region of *Streptomyces* sp. AgN23 genome in or out *S. violaceusniger* clade.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

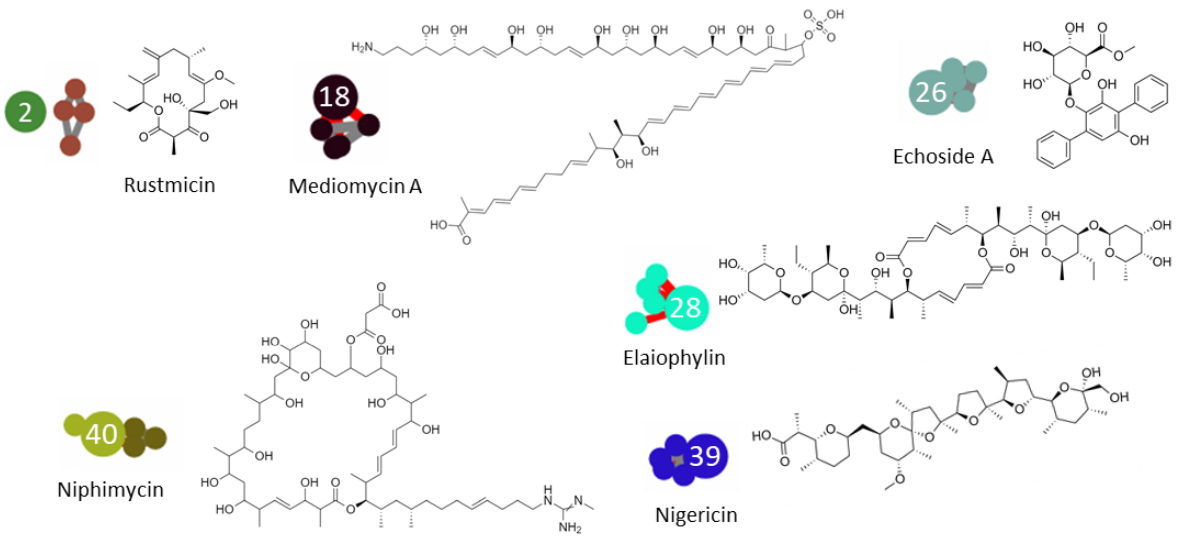
Twenty-two BGCs from the AgN23 genome showed conservation among all other close isolates while 13 others BGCs had at least one homologue in another isolate and 10 did not show homology with any of the other four strains selected. We then wondered which of these BGCs were specific to the *S. violaceusniger* clade or shared with another *Streptomyces* lineage. To address this question, the sequences of the 45 AgN23 BGCs were compared to the 1,225,071 BGCs of the BiG-FAM database using the BiG-SLICE algorithm. Twenty-eight BGCs obtained a BGC-to-GCFs (Gene Cluster Families) pairing distance lower than 900 in BiG-FAM, meaning that they share similar domain architectures with previously described BGCs (Figure 4b). The combination of the two analyses allowed us to conclude on orthologues distribution across *Streptomyces* phylogeny for each BGC-containing region detected in *Streptomyces* AgN23 genome.

Ten BGCs belonged to gene cluster families encountered both inside and outside of the *S. violaceusniger* clade (Figure 4b). This suggests that they are widely conserved among the *Streptomyces* genus. This group contains BGCs similar to the BGCs of terpene pristinol (region 10), ectoin (region 16), desferrioxamin B (region 24) and hopene (region 31) but also unknown BGCs annotated as terpene (Region 3), bacteriocins (region 5 and 28), indole (region 21), siderophore (region 27) and butyrolactone (region 36).

We then found that 26 of AgN23 BGCs shared orthologues exclusively with other S*. violaceusniger* representatives. Among them, we found the regions 20 and 30 predicted to encode enzymes for the biosynthesis of pigments that might be responsible of for the characteristic grey coloration of spores among members of the clade. Aside from such regions involved in the strain life cycle, the regions 18, 39 and 40 are predicted to code for the biosynthesis of the antifungals mediomycin A, nigericin and niphimycins C-E respectively and the regions 26 and 38 may be involved in synthesis of analogues of the plant bioactive echosides and elaiophylin (Figure 5). Notably, the BGC involved in the biosynthesis of geldanamycin, a hallmark phytotoxic, antifungal and antibacterial compound of *S. violaceusniger*

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

226 species was detected in 3 strains but not AgN23 and its closest neighbour *S. antimycoticus* (Figure 4a)
227 [59].



228

229 **Figure 5: Candidate BGCs for antimicrobials and plant bioactive compounds of *Streptomyces* sp.**
230 **AgN23 are conserved in the *S. violaceusniger* clade.** Connected nodes are part of the sequence
231 similarity network produced by the BiG-SCAPE comparison of AgN23 genome with four other *S.*
232 *violaceusniger* isolates. Each node represents a BGC, they are connected when sharing sequence
233 similarity, colours representing BiG-SCAPE gene clusters families. AgN23 BGCs are displayed as larger
234 nodes harbouring the number of their corresponding region. The structures on the right correspond
235 to putative compounds produced by highly similar BGCs according to antiSMASH.

236

237 Intriguingly, the BIG-SCAPE analysis clustered regions containing rustmicin-like BGCs from the four *S.*
238 *violaceusniger* strains we selected for comparative studies but did not add to this cluster the region 2
239 of AgN23 also annotated as rustmicin by antiSMASH (Table 2). Since the organisation of the five BGCs
240 are very similar to the subcluster GalA-E of *S. galbus* responsible for the biosynthesis of the
241 macrolactone ring of the compound, we propose that rustmicin biosynthesis is among core BGCs
242 functions of the *S. violaceusniger* clade [60] (Figure 6). Given *S. violaceusniger* isolates are frequently
243 isolated in interactions with plant roots, these data suggest that these specialised metabolites support
244 their accommodation into the plant rhizosphere and root microbiota [36,61].

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

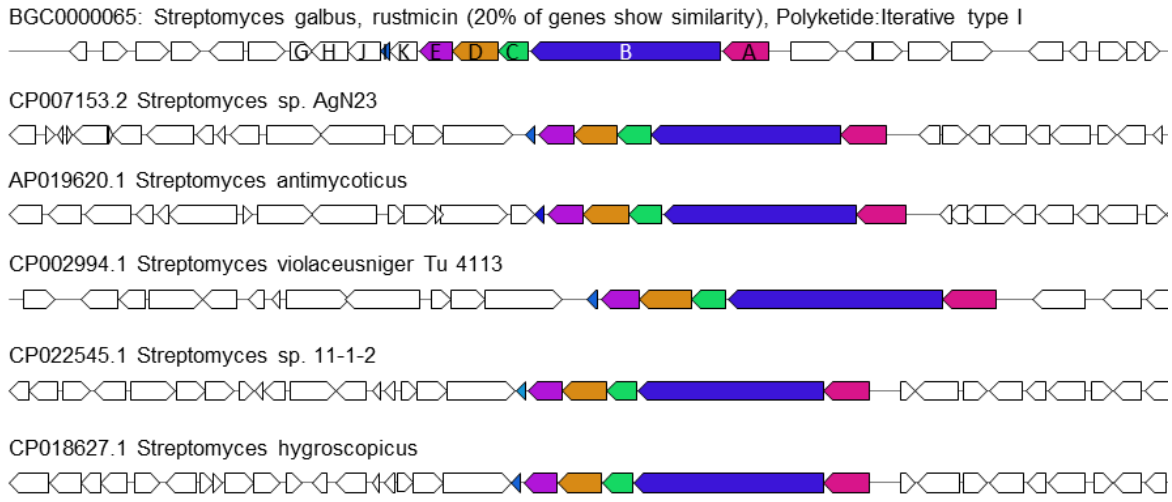


Figure 6: Rustmicin ABCDE biosynthetic gene cluster is conserved across *S. violaceusniger* clade. Comparison of the rustmicin BGC sequence from MiBIG (BGC0000065) with genomic regions that are similar in AgN23 and four others *S. violaceusniger* genomes. Coloured genes are putative homologs according to antiSMASH with ClusterBLAST *e* -value < 1E-05, an identity > 30% and a query cover > 25%.

AgN23 chromosome terminal regions harbour putatively unique biosynthetic gene clusters

Whilst we found a large core set of 26 BGCs is shared by AgN23 and four other *S. violaceusniger* isolates, six BGC-containing regions predicted in AgN23 genome show no similarity to any BGCs found within or outside the *violaceusniger* clade according to BiG-FAM and BiG-SCAPE analysis (Figure 4b). However, some of these regions received a MiBIG hit attributed by antiSMASH. This is notably the case of the region 12 which show similarity with the BGC responsible of the biosynthesis of totopotensamide in *S. pactum* a strain being outside of the *S. violaceusniger* clade (Table 2). The remaining regions showed no or <10% MiBIG hit and may constitute uncharacterised biosynthetic pathways for novel compounds.

Streptomyces chromosome organisation consists in a central conserved genome whilst terminal sequences contain more variable gene content described as accessory genome [62-64]. This accessory genome undergoes frequent amplification and deletion events as well as interspecies homologous recombination [65-70]. Regarding terminal sequences, 3 out of 5 regions located in the first 700 kb of the AgN23 chromosome left arm showed neither hit in other *S. violaceusniger* strains nor significant

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

alignment with known MiBIG reference (Supplementary Information 3, Figure 7). The same observation was drawn for the region 42 to 45 located in the last 700 kb of the chromosome right extremity. Interestingly, region 42 have 90% similarity to the one encoding coelichelin in *S. coelicolor* and is found in only one other *S. violaceusniger* stain, namely NRRL F-8817. The neighbouring region 43, contains a BGC with 100% similarity to the biosynthetic pathway of the antimicrobial bicyclomycin, which is absent from other *S. violaceusniger* strains but is found outside of the clade (Table 2). Thus, regions 42 and 43 may have been acquired by AgN23 through horizontal gene transfer from non-S. *violaceusniger* strains. Taken together these data suggest that AgN23 chromosome fulfils the classical organisation of *Streptomyces* genome with variable genetic sequences located in terminal regions.

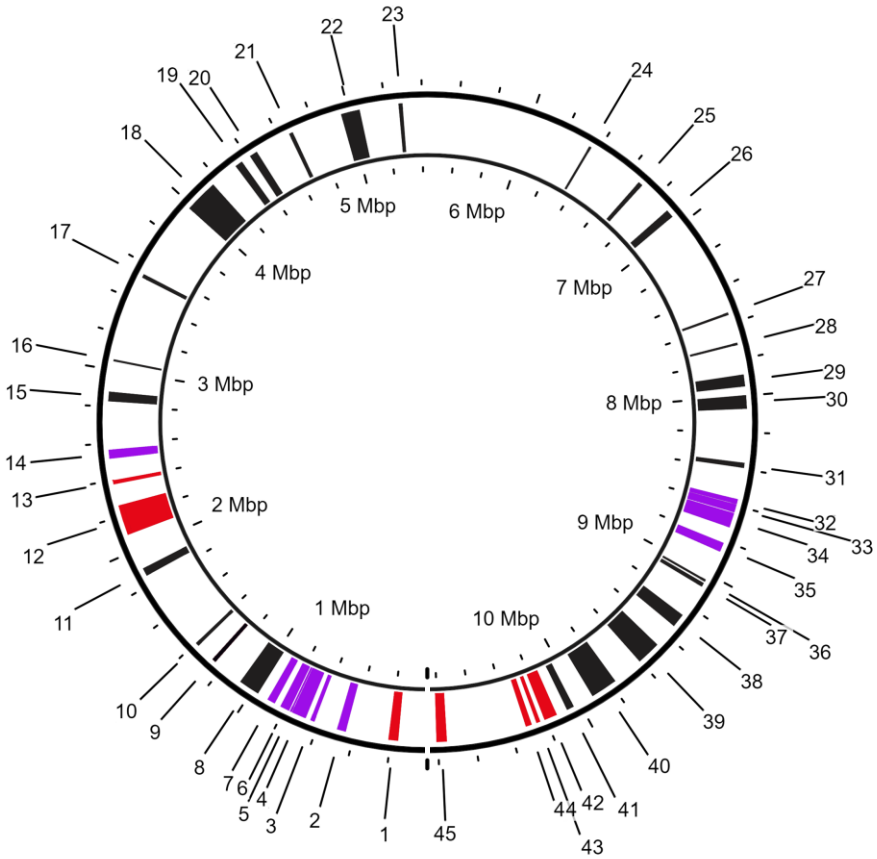


Figure 7: Schematic representation of antiSMASH-predicted BGCs-containing region along *Streptomyces* sp. AgN23 linear chromosome. Regions are numbered from 1 (left arm) to 45 (right arm) on the chromosome. The regions in red are non-conserved into *S. violaceusniger* clade according to both BiG-SCAPE and BiG-FAM analysis. The regions coloured in purple had no hit or a hit <=20% of similarity with a BGC from the MiBIG database.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

CONCLUSION

Plant roots recruit abundant and diverse consortia of microorganisms whilst exploring the soil. The relevance of this root microbiota in plant nutrition and resistance to stresses is currently unveiled through metabarcoding studies of plant microbiota. *Streptomyces* genus constitutes one of the most prominent bacterial genera colonising plant roots [6]. However, such studies do not inform on the species diversity of *Streptomyces* and the genomic features enabling their colonisation of plant roots. Therefore, genome sequencing of root associated *Streptomyces* is an important step toward the description of their molecular interaction with host plant. Here we produced a complete chromosome sequence of *Streptomyces* sp. AgN23, a strain isolated from grapevine rhizosphere and shown to elicit plant defence responses [15]. A whole genome-based phylogeny of AgN23 showed that it belongs to the *S. violaceusniger* clade from whom several other genome-sequenced strains have been isolated in rhizosphere of diverse plants around the world. We found that AgN23 genome, as well as other representative strains of the S. violaceusniger clade is rich in CAZymes able to degrade plant-derived carbohydrates and BGCs producing specialised metabolite potentially involved in the interaction with the plant. In addition, this phylogenetic lineage is considered as having a high potential in terms of specialised metabolism as its members have a large genome between 10.7 and 12.7 Mb and possess from 45 to 55 BGCs [71]. The wealth of genomic data available in this clade allowed us to unveil common trends in the BGCs specifically found in *S. violaceusniger* isolates. The ability of these strains to produce antifungal compounds (nigericin, niphimycin, mediomycin, rustmicin) as well as elicitors of plant defence (nigericin) strongly suggests that they play an important role in protection from pathogens. In addition, echosides and elaiophylin from whom structural analogues have been shown to interfere with auxin like responses suggest they may interfere with plant development [47,49]. Designing reverse genetic approach to inactivate candidates BGCs will allow dissecting the involvement of AgN23 specialised metabolism in the interaction with the plant and its microbiota.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

MATERIALS AND METHODS

AgN23 cultivation and HMW DNA extraction

AgN23 strain was cultivated as described previously [15,72]. In brief the strain was grown on solid modified Bennet medium (D-Glucose 10 g/l; Soybean peptones 2.5 g/l; Yeast Extract 1.5 g/l; Agarose 16 g/l) or International Streptomyces Project media [73,74] . To produce spore inoculum, we incubated Bennet plates for two weeks at 22°C in the darkness before filling them with 10 ml of sterile water. The mycelium was scraped with a spreader and the resulting solution was filtered in 50 ml syringe filled with non-absorbent cotton wool. For DNA extraction the AgN23 mycelium was grown at 28°C and under 250 rpm in 250 ml Erlenmeyer flasks containing 50 ml of liquid Bennet. Approximately 100 mg of AgN23 pellets were collected by centrifugation at 11 000 g and flash frozen in liquid nitrogen. Genomic DNA was isolated using the Macherey-Nagel Nucleobond RNA/DNA kit according to the manufacturer's instructions.

Library preparation for genome sequencing

Library preparation and sequencing were performed at the GeT-PlaGe core facility (Castanet-Tolosan), according to the manufacturer's instructions "Shared protocol-20kb Template Preparation Using BluePippin Size Selection system (15kb size cut-off)". At each step, DNA was quantified using the Qubit dsDNA HS Assay Kit (Life Technologies). DNA purity was tested using the NanoDrop (Thermo Fisher) and size distribution and degradation assessed using High Sensitivity Large Fragment 50kb Analysis Kit used with a Fragment analyser (AATI). Purification steps were performed using 0.45X AMPure PB beads (PacBio). A total of 10 µg of DNA was purified then sheared at 40kb using the Megaruptor system (Diagenode). Using SMRTBell template Prep Kit 1.0 (PacBio), a DNA and END damage repair step was performed on 5 µg of sample. Then, blunt hairpin adapters were ligated to the library. The library was treated with an exonuclease cocktail to digest unligated DNA fragments. A size selection step using a 10kb cut-off was performed on the BluePippin Size Selection system (Sage Science) with 0.75% agarose cassettes, Marker S1 high Pass 15-20 kb. Conditioned Sequencing Primer V2 was

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

annealed to the size selected SMRTbell. The annealed library was then bound to the P6-C4 polymerase using a ratio of polymerase to SMRTbell at 10:1. Then after a magnetic bead-loading step (OCPW), SMRTcell libraries were sequenced on 2 SMRTcells on RSII instrument at 0.18 to 0.23 nM with a 360 min movie. The initially generated raw sequencing reads were evaluated in terms of the average quality score at each position, GC content distribution, quality distribution, base composition, and other metrics. Sequencing reads with low quality were also filtered out before the genome assembly and annotation of gene structure. Finally, microbial DNA potential contamination was excluded after comparison by blast of the draft assembly of the first SMRT cell against a 16S ribosomal RNA sequences data bank (Bacteria and Archaea).

Genome assembly, annotation, and comparative genomics

The subreads were assembled with the Pacific Biosciences software SMRTanalysis version 2.3.0 using default settings with a minimum subreads length of 3 kb to exclude smaller sequenced reads and a read score of better than 0.8 to enrich in reads with a low error rate. The single Unitig obtained by long-read sequencing was corrected with Mi-Seq (Illumina®) data using Pilon (version 1.21), resulting in 165 substitutions and two deletions of 44 and 5 bases. This final genome assembly was retained for subsequent analysis. The gene annotation was performed on the MicroScope Microbial Genome Annotation & Analysis Platform [16]. CAZy genes annotation was obtained running the HMMER tool (E-Value < 1e-15, coverage > 0.35) on the dbCAN2 meta server [39]. The antiSMASH 5.0 server was used to detect BGC-containg regions in the AgN23 chromosome and annotate detected sequences based on the MiBIG 2.0 repository [40,41]. The comparative genomics was performed using combination of two tools. The first was the Antibiotic Resistant Target Seeker (ARTS) webservice to introduce the perform antiSMASH analyses of the 5 compared genomes with the BIG-SCAPE software [75,76]. The second was the BiG-FAM server that we use to find similar BGCs that the one predicted by AgN23 antiSMASH results among the 1,225,071 BGCs stored in the BiG-FAM database using the BiG-SLiCE software (1.0.0) with an arbitrary clustering threshold (T=900.0) [77,78]

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

356 **Phylogenomic analysis**

357 Genome of *Streptomyces* sp. AgN23 and related strains and outgroups were selected to perform
358 genome-scale phylogeny of AgN23 with autoMLST [20]. *Streptomyces* sp. M56, *Streptomyces* sp. 11-1-2,
359 *S. rapamycinicus* NRRL 5491, *S. malaysiensis* DSM4137, S. antimycoticus NBRC 100767 were chosen
360 based on their belonging to the *S. violaceusniger* clade according to the bibliography as well as the
361 genomes of the model strain *S. coelicolor* A(3)2 and of the two biocontrol strains, *S. lydicus* WYEC108
362 and *S. griseoveridis* K61, were added. *Frankia alni* ACN14a were used as outgroup (Supplementary
363 Table 1). Phylogenetic tree has a bootstrap value of 100.

364 **Library preparation for transcriptome sequencing and expression analysis**

365 AgN23 was cultivated on Bennett medium at 28°C for 48h. Total RNA was isolated using the RNeasy
366 Plant Mini Kit Qiagen according to manufacturer's instructions. Three replicates were prepared for the
367 libraries construction. RNA-seq libraries have been prepared according to Illumina's protocols using
368 the Illumina TruSeq Stranded mRNA sample prep kit to analyse mRNA. Briefly, ribodepletion was
369 carried out to clean up mRNAs using poly-T beads. Then, RNAs were fragmented to generate double
370 stranded cDNA and adaptors were ligated to be sequenced. Eleven cycles of PCR were applied to
371 amplify libraries. Library quality was assessed using a Fragment Analyzer and libraries were quantified
372 by QPCR using the Kapa Library Quantification Kit. RNA-Seq experiments have been performed on an
373 Illumina HiSeq4000 using a paired-end read length of 2x150 bp with the Illumina HiSeq4000
374 sequencing kits. The analysis was performed using the bioinformatics pipeline implemented in the
375 MicroScope Platform [16]. In a first step, raw reads of each sample (R1 fastq files from the paired-end
376 run) were mapped onto *Streptomyces* sp. AgN23 reference genome with BWA-MEM (v.0.7.4) [79]. An
377 alignment score equal to at least half of the read was required for a hit to be retained. SAMtools
378 (v.0.1.8) was then used to extract reliable alignments with a Mapping Quality (MAPQ)>=1 from SAM
379 formatted files [80]. The number of reads matching each genomic object of the reference sequence was
380 subsequently counted with the toolset BEDTools (v.2.10.1) [81]. The mean read counts of each gene was

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

381 calculated from three independent samples. The BGCs expression levels have been determined by
382 calculating the mean expression of biosynthetic genes for each BGC as defined by antiSMASH. A mean
383 expression for each BGC across the three biological repetitions was then determined.

384 **Scanning electronic microscopy**

385 The observation of AgN23 mycelium and spore development by Scanning Electronic microscopy was
386 performed on a FEG FEI Quanta 250. Agar plugs of AgN23 two weeks old culture were placed on
387 micrometric platen, frozen in liquid nitrogen and finally metallized with platinum. The samples were
388 observed microscopically at an accelerating voltage of 5.00 kV.

389 **SUPPLEMENTARY INFORMATION**

390 **Supplementary Information 1**: Annotation of AgN23 full chromosome. For each gene the frame of
391 translation, sequence length and position on the chromosome are indicated. All genes were annotated
392 according to the Microscope platform, see materials and methods. In addition, the mean raw reads
393 count found in RNA-seq data is indicated for each gene.

394 **Supplementary Information 2:** Genomes having a Mash-based estimated ANI (Average Nucleotide
395 Identity) superior or egal to 80% according to autoMLST.

396 **Supplementary Information 3:** Prediction of the CAZyme encoding genes using HMMER dbCAN2.
397 Mean raw reads count from the RNA-seq is indicated for each gene as well as the predicted targets of
398 the putative enzymes.

399 **Supplementary Information 4:** Gene identified by antiSMASH in the region containing a biosynthetic
400 gene cluster. Mean raw reads count from the RNA -eq is indicated for each gene. Annotated central
401 biosynthetic genes are indicated as Y. Those are the ones used for the calculation of mean BGC
402 expression in Table 2.

403

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DATA AVAILIBILITY

The raw reads sequences of AgN23 genome are available at NCBI on the Sequence Read Archive portal for PacBio and MiSeq data (SRR13990229 and SRR14028548 respectively). The Genome assembly is available on the NCBI nucleotide portal under the accession NZ_CP007153.1. This genome sequence was uploaded on the MicroScope platform for genome annotation and analysis (https://mage.genoscope.cns.fr/microscope/home/index.php) [16]. The RNA-seq raw reads are archived in the NCBI Bioproject PRJNA745930. The genome sequence used in BiG-SCAPE analysis can be retrieved from the NCBI GenBank portal for *S. antimycoticus* NBRC 100767, *Streptomyces* sp. 11-1-2, *S. hygroscopicus* XM201, *S. violaceusniger* Tü 4113, Streptomyces coelicolor A3(2) and Streptomyces sp. SirexAA-E with the accession number AP019620.1, CP022545.1, CP018627.1, CP002994.1, AL645882 and CP002993 respectively.

ACKNOWLEDGEMENT

This work was funded by the Fond Unique Interministériels (NEOPROTEC project), the Fonds Européen de Développement Économique et Régional (FEDER), the Agence Nationale de la Recherche (LabCom BioPlantProtec ANR-14-LAB7-0001 and STREPTOCONTROL ANR-17-CE20-0030) and the Région Occitanie (projet GRAINE-BioPlantProducts). The Laboratoire de Recherche en Sciences Végétales (LRSV) belongs to the TULIP Laboratoire d'Excellence (ANR-10-LABX-41). Work performed in the GeT core facility, Toulouse, France (https://get.genotoul.fr) was supported by the France Génomique National infrastructure, funded as part of the "Investissement d'Avenir" program managed by the Agence Nationale de la Recherche (contract ANR-10-INBS-09) and by the GET-PACBIO program (FEDER Programme opérationnel FEDER-FSE MIDI-PYRENEES ET GARONNE 2014-2020). D. Gayrard was funded by Agence Nationale de la Recherche Technique, with the Convention Industrielle de Formation par la Recherche and Association Nationale de la Recherche et de la Technologie (grant number 2016/1297). We warmly thanks Olivier Bouchez for helpful discussions on Next Generation Sequencing strategy throughout the project. We are grateful to Sylvie Lautru and Jean-Luc Pernodet

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(I2BC, Paris) for helpful discussions. We thank Clément Nicolle, Thierry Grollier and Valérie Arnal for helpful comments on earlier version of the manuscript.

AUTHOR CONTRIBUTIONS

Gayrard D. performed the research and wrote the manuscript. Veyssière M. and Adam K., Martinez Y., Vandecasteele C. and Vidal M. performed the research. Dumas B. and Rey T. designed the research and wrote the manuscript.

COMPETING INTERESTS

The following information may be foreseen as competing interest. B. Dumas is one of inventors of the patent WO2015044585A1 relating the use of AgN23 in Agriculture. T. Rey and D. Gayrard are full-time researchers in the AgChem company De Sangosse (Pont-Du-Casse, France), which registers and markets crop-protection products.

LITERATURE CITED

1 Bush, M. J., Tschowri, N., Schlimpert, S., Flärdh, K. & Buttner, M. J. c-di-GMP signalling and the regulation of developmental transitions in streptomycetes. *Nat Rev Microbiol* **13**, 749-760, doi:10.1038/nrmicro3546 (2015).

2 Book, A. J. *et al.* Evolution of High Cellulolytic Activity in Symbiotic Streptomyces through Selection of Expanded Gene Content and Coordinated Gene Expression. *PLOS Biology* **14**, e1002475, doi:10.1371/journal.pbio.1002475 (2016).

3 Moumbock, A. F. A. *et al.* StreptomeDB 3.0: an updated compendium of streptomycetes natural products. *Nucleic Acids Res* **49**, D600-D604, doi:10.1093/nar/gkaa868 (2021).

4 Barka, E. A. *et al.* Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol Mol Biol Rev* **80**, 1-43, doi:10.1128/MMBR.00019-15 (2016).

5 Lee, N. *et al.* Thirty complete Streptomyces genome sequences for mining novel secondary metabolite biosynthetic gene clusters. *Sci Data* **7**, 55, doi:10.1038/s41597-020-0395-9 (2020).

6 Lundberg, D. S. *et al.* Defining the core Arabidopsis thaliana root microbiome. *Nature* **488**, 86-90, doi:10.1038/nature11237 (2012).

7 Fitzpatrick, C. R. *et al.* Assembly and ecological function of the root microbiome across angiosperm plant species. *Proc Natl Acad Sci U S A* **115**, E1157-E1165, doi:10.1073/pnas.1717617115 (2018).

8 Vurukonda, S. S. K. P., Giovanardi, D. & Stefani, E. Plant Growth Promoting and Biocontrol Activity of Streptomyces spp. as Endophytes. *Int J Mol Sci* **19**, doi:10.3390/ijms19040952 (2018).

9 Hamedi, J. & Mohammadipanah, F. Biotechnological application and taxonomical distribution of plant growth promoting actinobacteria. *Journal of industrial microbiology & biotechnology* **42**, 157-171 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10 Rey, T. & Dumas, B. Plenty Is No Plague: Streptomyces Symbiosis with Crops. *Trends Plant Sci* **22**, 30-37, doi:10.1016/j.tplants.2016.10.008 (2017).

11 Viaene, T., Langendries, S., Beirinckx, S., Maes, M. & Goormachtig, S. Streptomyces as a plant's best friend? *FEMS Microbiol Ecol* **92**, doi:10.1093/femsec/fiw119 (2016).

12 Medema, M. H. The year 2020 in natural product bioinformatics: an overview of the latest tools and databases. *Nat Prod Rep* **38**, 301-306, doi:10.1039/d0np00090f (2021).

13 Tracanna, V., de Jong, A., Medema, M. H. & Kuipers, O. P. Mining prokaryotes for antimicrobial compounds: from diversity to function. *FEMS Microbiol Rev* **41**, 417-429, doi:10.1093/femsre/fux014 (2017).

14 Kautsar, S. A., Blin, K., Shaw, S., Weber, T. & Medema, M. H. BiG-FAM: the biosynthetic gene cluster families database. *Nucleic Acids Res* **49**, D490-D497, doi:10.1093/nar/gkaa812 (2021).

15 Vergnes, S. *et al.* Phyllosphere colonisation by a soil Streptomyces sp. promotes plant defense responses against fungal infection. *Mol Plant Microbe Interact*, doi:10.1094/MPMI-05-19-0142-R (2019).

16 Vallenet, D. *et al.* MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res* **48**, D579-D589, doi:10.1093/nar/gkz926 (2020).

17 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).

18 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212, doi:10.1093/bioinformatics/btv351 (2015).

19 Labeda, D. P. *et al.* Phylogenetic study of the species within the family Streptomycetaceae. *Antonie Van Leeuwenhoek* **101**, 73-104, doi:10.1007/s10482-011-9656-0 (2012).

20 Alanjary, M., Steinke, K. & Ziemert, N. AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential. *Nucleic Acids Res* **47**, W276-W282, doi:10.1093/nar/gkz282 (2019).

21 Kumar, Y. & Goodfellow, M. Five new members of the Streptomyces violaceusniger 16S rRNA gene clade: Streptomyces castelarensis sp. nov., comb. nov., Streptomyces himastatinicus sp. nov., Streptomyces mordarskii sp. nov., Streptomyces rapamycinicus sp. nov. and Streptomyces ruanii s. *International Journal of Systematic and Evolutionary Microbiology* **58**, 1369-1378, doi:10.1099/ijs.0.65408-0 (2008).

22 Komaki, H. & Tamura, T. Reclassification of Streptomyces castelarensis and Streptomyces sporoclivatus as later heterotypic synonyms of Streptomyces antimycoticus. *International Journal of Systematic and Evolutionary Microbiology*, doi:10.1099/ijsem.0.003882 (2020).

23 Goodfellow, M., Kumar, Y., Labeda, D. P. & Sembiring, L. The Streptomyces violaceusniger clade: a home for streptomycetes with rugose ornamented spores. *Antonie van Leeuwenhoek* **92**, 173-199, doi:10.1007/s10482-007-9146-6 (2007).

24 Yin, M., Jiang, M., Ren, Z., Dong, Y. & Lu, T. The complete genome sequence of Streptomyces autolyticus CGMCC 0516, the producer of geldanamycin, autolytimycin, reblastatin and elaiophylin. *J Biotechnol* **252**, 27-31, doi:10.1016/j.jbiotec.2017.04.037 (2017).

25 Ser, H.-L. *et al.* Streptomyces antioxidans sp. nov., a Novel Mangrove Soil Actinobacterium with Antioxidative and Neuroprotective Potentials. *Frontiers in Microbiology* **7**, doi:10.3389/fmicb.2016.00899 (2016).

26 Hamedi, J. *et al.* Streptomyces iranensis sp. nov., isolated from soil. *Int J Syst Evol Microbiol* **60**, 1504-1509, doi:10.1099/ijs.0.015339-0 (2010).

27 Chagas, F. O. *et al.* Genome Sequence of Streptomyces sp. Strain RTd22, an Endophyte of the Mexican Sunflower. *Genome Announc* **4**, doi:10.1128/genomeA.00693-16 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

28 Yang, H., Zhang, Z., Yan, R., Wang, Y. & Zhu, D. Draft Genome Sequence of Streptomyces sp. Strain PRh5, a Novel Endophytic Actinomycete Isolated from Dongxiang Wild Rice Root. *Genome Announc* **2**, doi:10.1128/genomeA.00012-14 (2014).

29 Bown, L. & Bignell, D. R. D. Draft Genome Sequence of the Plant Pathogen. *Genome Announc* **5**, doi:10.1128/genomeA.00968-17 (2017).

30 Zeng, J. *et al.* The Role of Iron Competition in the Antagonistic Action of the Rice Endophyte Streptomyces sporocinereus OsiSh-2 Against the Pathogen Magnaporthe oryzae. *Microb Ecol* **76**, 1021-1029, doi:10.1007/s00248-018-1189-x (2018).

31 Xu, T. *et al.* Isolation and evaluation of endophytic Streptomyces endus OsiSh-2 with potential application for biocontrol of rice blast disease. *J Sci Food Agric* **97**, 1149-1157, doi:10.1002/jsfa.7841 (2017).

32 Cao, L. *et al.* Streptomyces hygroscopicus OsiSh-2-induced mitigation of Fe deficiency in rice plants. *Plant Physiol Biochem* **158**, 275-283, doi:10.1016/j.plaphy.2020.11.013 (2021).

33 Gao, Y. *et al.* Endophytic. *mBio*, e0156621, doi:10.1128/mBio.01566-21 (2021).

34 Xu, T. *et al.* The antifungal action mode of the rice endophyte Streptomyces hygroscopicus OsiSh-2 as a potential biocontrol agent against the rice blast pathogen. *Pestic Biochem Physiol* **160**, 58-69, doi:10.1016/j.pestbp.2019.06.015 (2019).

35 Komaki, H. *et al.* Draft genome sequence of Streptomyces sp. TP-A0867, an alchivemycin producer. *Stand Genomic Sci* **11**, 85, doi:10.1186/s40793-016-0207-1 (2016).

36 Goodfellow, M., Kumar, Y., Labeda, D. P. & Sembiring, L. The Streptomyces violaceusniger clade: a home for Streptomycetes with rugose ornamented spores. *Antonie Van Leeuwenhoek* **92**, 173-199, doi:10.1007/s10482-007-9146-6 (2007).

37 Kumar, Y., Aiemsum-Ang, P., Ward, A. C. & Goodfellow, M. Diversity and geographical distribution of members of the Streptomyces violaceusniger 16S rRNA gene clade detected by clade-specific PCR primers. *FEMS Microbiol Ecol* **62**, 54-63, doi:10.1111/j.1574-6941.2007.00374.x (2007).

38 Kumar, Y. & Goodfellow, M. Five new members of the Streptomyces violaceusniger 16S rRNA gene clade: Streptomyces castelarensis sp. nov., comb. nov., Streptomyces himastatinicus sp. nov., Streptomyces mordarskii sp. nov., Streptomyces rapamycinicus sp. nov. and Streptomyces ruanii sp. nov. *Int J Syst Evol Microbiol* **58**, 1369-1378, doi:10.1099/ijs.0.65408-0 (2008).

39 Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**, W445-W451, doi:10.1093/nar/gks479 (2012).

40 Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**, W81-W87, doi:10.1093/nar/gkz310 (2019).

41 Epstein, S. C., Charkoudian, L. K. & Medema, M. H. A standardized workflow for submitting data to the Minimum Information about a Biosynthetic Gene cluster (MIBiG) repository: prospects for research-based educational experiences. *Stand Genomic Sci* **13**, 16, doi:10.1186/s40793-018-0318-y (2018).

42 Hu, Y. *et al.* Identification and Proposed Relative and Absolute Configurations of Niphimycins C-E from the Marine-Derived Streptomyces sp. IMB7-145 by Genomic Analysis. *J Nat Prod* **81**, 178-187, doi:10.1021/acs.jnatprod.7b00859 (2018).

43 Harvey, B. M. *et al.* Insights into polyether biosynthesis from analysis of the nigericin biosynthetic gene cluster in Streptomyces sp. DSM4137. *Chem Biol* **14**, 703-714, doi:10.1016/j.chembiol.2007.05.011 (2007).

44 Cai, P. *et al.* Polyene Antibiotics from Streptomyces mediocidicus. *Journal of Natural Products* **70**, 215-219, doi:10.1021/np060542f (2007).

45 Kim, H. J. *et al.* A single module type I polyketide synthase directs de novo macrolactone biogenesis during galbonolide biosynthesis in Streptomyces galbus. *J Biol Chem* **289**, 34557-34568, doi:10.1074/jbc.M114.602334 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

563 46 Hayashi, K. *et al.* Active core structure of terfestatin A, a new specific inhibitor of auxin
564 signaling. *Bioorg Med Chem* **16**, 5331-5344, doi:10.1016/j.bmc.2008.02.085 (2008).
565 47 Wang, X. *et al.* Terfestatins B and C, New p-Terphenyl Glycosides Produced by Streptomyces
566 sp. RM-5-8. *Org Lett* **17**, 2796-2799, doi:10.1021/acs.orglett.5b01203 (2015).
567 48 Yamazoe, A., Hayashi, K., Kepinski, S., Leyser, O. & Nozaki, H. Characterization of terfestatin
568 A, a new specific inhibitor for auxin signaling. *Plant Physiol* **139**, 779-789,
569 doi:10.1104/pp.105.068924 (2005).
570 49 Igarashi, Y., Iida, T., Yoshida, R. & Furumai, T. Pteridic acids A and B, novel plant growth
571 promoters with auxin-like activity from Streptomyces hygroscopicus TP-A0451. *J Antibiot*
572 *(Tokyo)* **55**, 764-767, doi:10.7164/antibiotics.55.764 (2002).
573 50 Maintz, J. *et al.* Chemical Activation of EDS1/PAD4 Signaling Leading to Pathogen Resistance
574 in Arabidopsis. *Plant and Cell Physiology* **59**, 1592-1607, doi:10.1093/pcp/pcy106 (2018).
575 51 Nagpure, A., Choudhary, B. & Gupta, R. K. Mycolytic enzymes produced by Streptomyces
576 violaceusniger and their role in antagonism towards wood-rotting fungi. *J Basic Microbiol* **54**,
577 397-407, doi:10.1002/jobm.201200474 (2014).
578 52 Riclea, R., Citron, C. A., Rinkel, J. & Dickschat, J. S. Identification of isoafricanol and its terpene
579 cyclase in Streptomyces violaceusniger using CLSA-NMR. *Chem Commun (Camb)* **50**, 4228-
580 4230, doi:10.1039/c4cc00177j (2014).
581 53 Chen, X. *et al.* Genome Sequence of Streptomyces violaceusniger Strain SPC6, a Halotolerant
582 Streptomycete That Exhibits Rapid Growth and Development. *Genome Announc* **1**,
583 doi:10.1128/genomeA.00494-13 (2013).
584 54 Kang, M. J., Strap, J. L. & Crawford, D. L. Isolation and characterization of potent antifungal
585 strains of the Streptomyces violaceusniger clade active against Candida albicans. *J Ind*
586 *Microbiol Biotechnol* **37**, 35-41, doi:10.1007/s10295-009-0641-9 (2010).
587 55 Hayakawa, M., Yoshida, Y. & Iimura, Y. Selective isolation of bioactive soil actinomycetes
588 belonging to the Streptomyces violaceusniger phenotypic cluster. *J Appl Microbiol* **96**, 973-
589 981, doi:10.1111/j.1365-2672.2004.02230.x (2004).
590 56 Hayashi, K. *et al.* WS9326A, a novel tachykinin antagonist isolated from Streptomyces
591 violaceusniger no. 9326. I. Taxonomy, fermentation, isolation, physico-chemical properties
592 and biological activities. *J Antibiot (Tokyo)* **45**, 1055-1063, doi:10.7164/antibiotics.45.1055
593 (1992).
594 57 Martinet, L. *et al.* On the Risks of Phylogeny-Based Strain Prioritization for Drug Discovery:.
595 *Biomolecules* **10**, doi:10.3390/biom10071027 (2020).
596 58 Choudoir, M. J., Pepe-Ranney, C. & Buckley, D. H. Diversification of Secondary Metabolite
597 Biosynthetic Gene Clusters Coincides with Lineage Divergence in Streptomyces. *Antibiotics*
598 *(Basel)* **7**, doi:10.3390/antibiotics7010012 (2018).
599 59 Heisey, R. M. & Putnam, A. R. Herbicidal effects of geldanamycin and nigericin, antibiotics
600 from Streptomyces hygroscopicus. *J Nat Prod* **49**, 859-865 (1986).
601 60 Kim, H.-J. J. *et al.* A single module type I polyketide synthase directs de Novo macrolactone
602 biogenesis during galbonolide biosynthesis in Streptomyces galbus. *Journal of Biological*
603 *Chemistry* **289**, 34557-34568, doi:10.1074/jbc.M114.602334 (2014).
604 61 Sembiring, L., Ward, A. C. & Goodfellow, M. Selective isolation and characterisation of
605 members of the Streptomyces violaceusniger clade associated with the roots of
606 Paraserianthes falcataria. *Antonie Van Leeuwenhoek* **78**, 353-366,
607 doi:10.1023/a:1010226515202 (2000).
608 62 Kim, J. N. *et al.* Comparative Genomics Reveals the Core and Accessory Genomes of
609 Streptomyces Species. *J Microbiol Biotechnol* **25**, 1599-1605, doi:10.4014/jmb.1504.04008
610 (2015).
611 63 Lioy, V. S. *et al.* Dynamics of the compartmentalized Streptomyces chromosome during
612 metabolic differentiation. *Nat Commun* **12**, 5221, doi:10.1038/s41467-021-25462-1 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465742; this version posted October 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

64 Szafran, M. J. *et al.* Spatial rearrangement of the Streptomyces venezuelae linear chromosome during sporogenic development. *Nat Commun* **12**, 5222, doi:10.1038/s41467-021-25461-2 (2021).

65 Choudoir, M. J. & Buckley, D. H. Phylogenetic conservatism of thermal traits explains dispersal limitation and genomic differentiation of Streptomyces sister-taxa. *ISME J* **12**, 2176-2186, doi:10.1038/s41396-018-0180-3 (2018).

66 Andam, C. P., Choudoir, M. J., Vinh Nguyen, A., Sol Park, H. & Buckley, D. H. Contributions of ancestral inter-species recombination to the genetic diversity of extant Streptomyces lineages. *ISME J* **10**, 1731-1741, doi:10.1038/ismej.2015.230 (2016).

67 Tidjani, A. R. *et al.* Massive Gene Flux Drives Genome Diversity between Sympatric. *MBio* **10**, doi:10.1128/mBio.01533-19 (2019).

68 McDonald, B. R. & Currie, C. R. Lateral Gene Transfer Dynamics in the Ancient Bacterial Genus. *MBio* **8**, doi:10.1128/mBio.00644-17 (2017).

69 Doroghazi, J. R. & Buckley, D. H. Widespread homologous recombination within and between Streptomyces species. *ISME J* **4**, 1136-1143, doi:10.1038/ismej.2010.45 (2010).

70 Kun Cheng, X. R., Ying Huang,. Vol. 102 (Molecular Phylogenetics and Evolution, 2016).

71 Chung, Y.-H. *et al.* Comparative Genomics Reveals a Remarkable Biosynthetic Potential of the Streptomyces Phylogenetic Lineage Associated with Rugose-Ornamented Spores. *Msystems* **6**, e00489-00421 (2021).

72 Errakhi, R., Bouteau, F., Lebrihi, A. & Barakate, M. Evidences of biological control capacities of Streptomyces spp. against Sclerotium rolfsii responsible for damping-off disease in sugar beet (Beta vulgaris L.). *World Journal of Microbiology and Biotechnology* **23**, 1503-1509, doi:10.1007/s11274-007-9394-7 (2007).

73 Shirling, E. B. & Gottlieb, D. in *International Journal of Systematic Bacteriology* Vol. 16 313-340 (1966).

74 Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical streptomyces genetics*. Vol. 291 (John Innes Foundation Norwich, 2000).

75 Mungan, M. D. *et al.* ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. *Nucleic Acids Res* **48**, W546-W552, doi:10.1093/nar/gkaa374 (2020).

76 Navarro-Muñoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat Chem Biol* **16**, 60-68, doi:10.1038/s41589-019-0400-9 (2020).

77 Kautsar, S. A., van der Hooft, J. J. J., de Ridder, D. & Medema, M. H. BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. *Gigascience* **10**, doi:10.1093/gigascience/giaa154 (2021).

78 Kautsar, S. A., Blin, K., Shaw, S., Weber, T. & Medema, M. H. BiG-FAM: the biosynthetic gene cluster families database. *Nucleic Acids Res* **49**, D490–D497, doi:10.1093/nar/gkaa812 (2021b).

79 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* (2013).

80 Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

81 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).