Impacts of horizontal gene transfer on the compact genome of the clavulanic acid-producing *Streptomyces* strain F613-1

Jun Li^{1,2} · Zhilong Zhao³ · Weihong Zhong² · Chuanqing Zhong⁴ · Gongli Zong¹ · Jiafang Fu¹ · Guangxiang Cao¹

Received: 25 September 2018 / Accepted: 1 November 2018 / Published online: 8 November 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Mobile genetic elements involved in mediating horizontal transfer events contribute to bacterial evolution, and bacterial genomic plasticity and instability result in variation in functional genetic information in *Streptomyces* secondary metabolism. In a previous study, we reported the complete genome sequence of the industrial *Streptomyces* strain F613-1, which produces high yields of clavulanic acid. In this study, we used comparative genomics and bioinformatics to investigate the unique genomic features of this strain. Taken together, comparative genomics were used to systematically investigate secondary metabolism capabilities and indicated that frequent exchange of genetic materials between *Streptomyces* replicons may shape the remarkable diversities in their secondary metabolite repertoires. Moreover, a 136.9-kb giant region of plasticity (RGP) was found in the F613-1 chromosome, and the chromosome and plasmid pSCL4 are densely packed with an exceptionally large variety of potential secondary metabolic gene clusters, involving several determinants putatively accounting for antibiotic production. In addition, the differences in the architecture and size of plasmid pSCL4 between F613-1 and ATCC 27064 suggest that the pSCL4 plasmid could evolve from pSCL4-like and pSCL2-like extrachromosomal replicons. Furthermore, the genomic analyses revealed that strain F613-1 has developed specific genomic architectures and genetic patterns that are well suited to meet the requirements of industrial innovation processes.

Keywords Comparative genomics · Genetic exchanges · Horizontal transfer · *Streptomyces clavuligerus* · Compact genome · Clavulanic acid · Secondary metabolism

Introduction

Streptomyces species are Gram-positive bacteria with genomes of high G + C%. A range of previous studies have focused on their ability to produce a variety of antimicrobial and valuable bio-based chemicals. Based on this scenario, the complex secondary metabolism and genetic regulation

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-018-1498-2>) contains supplementary material, which is available to authorized users.

✉ Guangxiang Cao
caozhong0402@163.com

Jun Li
lijunbio@zjut.edu.cn

Zhilong Zhao
zhaozhilong@lyu.edu.cn

Weihong Zhong
whzhong@zjut.edu.cn

Chuanqing Zhong
zhongchq@163.com

Gongli Zong
mls_zonggl@ujn.edu.cn

Jiafang Fu
mefujiafang@163.com

¹ Shandong Medicinal Biotechnology Center, Shandong Academy of Medical Sciences, Jinan, China

² College of Biotechnology and Bioengineering, Zhejiang University of Technology, Hangzhou, China

³ School of Pharmacy, Linyi University, Linyi, China

⁴ Shandong Jianzhu University, Jinan, China

in *Streptomyces* have been extensively studied, to understand the biosynthesis of natural products. With an emphasis on the commercial value of secondary metabolites, the biotechnological potential associated with secondary metabolism has been systematically investigated with the advancement and development of high-throughput sequencing techniques and up-to-date bioinformatics approaches.

As one of the most widely studied genera, hundreds of draft genomes of *Streptomyces* have presently been demonstrated and archived in GenBank. However, the limited number of completely assembled sequences reflects the challenges in acquiring whole genomes and closing the sequence gaps for this genus. The entire genome sequences of *Streptomyces lividans* TK24 (Ruckert et al. 2015) and *S. lividans* 66 (Cruz-Morales et al. 2013) were released, approximately one decade after the first published complete genome of the closely related strain *S. coelicolor* A3(2) (Bentley et al. 2002). Based on the feasibility and significance of combining the PacBio SMRT and Illumina MiSeq platforms, the *S. leeuwenhoekii* genome was assembled recently, previously consisting of a linear scaffold and two additional replicons (Gomez-Escribano et al. 2015). The formicamycin producer *Streptomyces formicae* KY5 (Holmes et al. 2018), streptolydigin producer *Streptomyces lydicus* 103 (Jia et al. 2017) and anticancer anthracyclines producer *Streptomyces peucetius* ATCC 27952 (Dhakal et al. 2018) have also been sequenced to acquire their metabolism and genomic diversity.

The replicons (chromosomes, plasmids) characterized in *Streptomyces* species are generally regarded as linear. In addition, spontaneous deletions, genomic amplification and sequence arrangements commonly occur in the chromosome, contributing to diversified genetic instability and metabolic capability of *Streptomyces* (Li et al. 2016; Lin et al. 1993; Volff and Altenbuchner 1998; Xu et al. 2016). The linear 6.84 Mb chromosome of *Streptomyces albus* J1074 could be considered as a ‘naturally minimized’ scaffold of the *Streptomyces* genomes (Ian et al. 2014); such ‘compact’ genomic architecture could assist in the definition of essential genes for environmental fitness and bacterial growth/replication, which is also involved in the natural product studies (Gonzalez et al. 2016).

Of note, *Streptomyces clavuligerus* can produce clavulanic acid, which is used clinically together with conventional β -lactam antibacterial therapy for the treatment of infectious diseases caused by pathogenic microorganisms that exhibit resistance to β -lactam antibiotics (Medema et al. 2010). The draft genome sequence of *S. clavuligerus* strain ATCC27064 revealed that it was composed of one linear chromosome and four linear plasmids, respectively, nominated pSCL1, pSCL2, pSCL3 and pSCL4 (Song et al. 2010). Genome analysis revealed the clavulanic acid gene cluster, the clavam gene cluster, and the paralog gene cluster, which are proposed, respectively, to account for the biosynthesis

of clavulanic acid, cefamycin C, and clavams. Interestingly, the clavulanic acid gene cluster and the clavam gene cluster are located on the chromosome, whereas the paralog gene cluster is located on plasmid pSCL4 (Medema et al. 2010; Song et al. 2010).

Streptomyces clavuligerus strain F613-1 is an important industrial producer with high clavulanic acid production, but poor ability to produce 5S clavam compounds, and its complete genome of F613-1 has been sequenced (Cao et al. 2016). In this study, we analyzed the compact genome sequence of F613-1, which contains a 6.78 Mb chromosome, considerably smaller than other reported *S. clavuligerus* genomes, and a 707 Kb pSCL4-like plasmid (smaller than 1.8 Mb pSCL4, however, carrying the intact paralog gene cluster related with biosynthesis of clavulanic acid). Frequent exchanges of genetic material between F613-1 plasmids and chromosomes may shape a rapidly varying repertoire of secondary metabolic capacities and evolving genetic features that aim at adapting to changing conditions and industrial requirement. Taken together, customized comparative genomics suggested that this ‘highly plastic’ genome could constitute a chassis for the biosynthesis and modification of natural products. In addition, analysis of several valuable metabolic genes suggested future investigations for optimizing the industrial use of this strain.

Methods and materials

Bacterial strain and culture condition

Streptomyces clavuligerus F613-1 is an industrial producer of clavulanic acid, which was developed and derived from strain ATCC 27064. ATCC 27064 was used as control in this work. *S. clavuligerus* was grown at 25 °C on BSCA plates (containing 1.5% malt extract, 0.3% tryptone, 0.4% glucose, 2.0% agar powder, pH 7.5) for 8 days to collect spores. To extract genomic DNA, spores were harvested from 1 cm² agar, inoculated into 50 ml of tryptic soy broth medium (Oxoid, UK) in 250-ml flasks and grown at 25 °C on a rotary shaker at 250 rpm for 60 h to harvest mycelium. To produce clavulanic acid, *S. clavuligerus* was grown on liquid seed medium (2% soybean meal, 2% glycerol, 1.2% cornstarch, 0.5% yeast extract, 0.08% potassium phosphate, pH 8.0) and then inoculated into a fermentation tank containing 3.0% soybean meal, 2.5% soy isolate, 3.0% maltedextrin, 1.0% glycerol trioleate, 0.25% potassium phosphate, 0.1% magnesium chloride, 0.02% sodium chloride, and 0.05% foamaster, pH 7.0; supplemented 1.5% glycerol was added during the fermentation process. The concentration of clavulanic acid was detected as described (Zhong et al. 2014), with a Thermo Scientific Hypersil ODS-3 (C18) 5 μ m column.

Genome annotation and analysis

The genome of the strain F613-1 has been sequenced and assembled (Cao et al. 2016). Genome annotation was previously performed as recommended by the NCBI (National Center for Biotechnology Information) online pipeline for prokaryotic species. In brief, putative protein-coding DNA sequences (CDS) were predicted using Glimmer 3.02 (Delcher et al. 2007), and the shorter genes (< 100-bp) were inferred using Prodigal V2.60 (Hyatt et al. 2010). Ori-Finder was used to investigate its genomic replication origin site (*oriC*) (Luo et al. 2014). Functions for the corresponding protein candidates were predicted using BLASTp-based assignment against all reference annotated information of *S. clavuligerus* ATCC 27064 and the non-redundant (nr) protein sequence database archived in NCBI. The assembled genome sequences were also annotated and examined using the RAST server (Overbeek et al. 2014) for CDS, ARAGORN (Laslett and Canback 2004) for the annotation of tRNA genes, and RNAmmer (Lagesen et al. 2007) for annotation of rRNA genes. Clusters of orthologous groups (COGs) (Galperin et al. 2015) for each gene product were predicted using standalone RPS-BLAST. Identification of biosynthetic gene clusters for antibiotics and other secondary metabolites was conducted by searching against the antiSMASH databases (Blin et al. 2017).

Comparative genome analysis

The F613-1 genome sequence was aligned to ATCC 27064 via the Mauve suite (Darling et al. 2004). Strain-specific CDS were identified with VR profile (Li et al. 2018) and mGenomeSubtractor with *H* value ≥ 0.42 (cutoffs: 60% identity, 70% length matching ratio) (Shao et al. 2010). Completely sequenced *Streptomyces* genomes were collected to access the phylogenetic status of F613-1, and whole genomic protein sequence-level phylogeny was calculated by the CVTree3 program (Zuo and Hao 2015).

Analysis of the mobile genome

A bacterial ‘mobile genomic region’ annotation strategy was designed to analyze the assembled replicons and to identify mobile genetic elements (e.g., ISs, integrons, prophages). Plasmids were previously identified and assembled after sequencing the genome, and we designed a mobilome annotation procedure to detect other mobile-related elements in sequenced bacterial genomes as follows. PlasmidFinder (Carattoli et al. 2014) was used to type replicons of plasmids that had been previously assembled and identified. Plasmids whose types could not be probed by the above tools were compared against the NCBI nucleotide database using the BLASTn program (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>)

to find plasmids that were highly similar at the nucleotide level. If this method failed to retrieve a match, plasmid replication proteins were compared against the GenBank nr and shotgun (WGS) databases to find homologous replicons for reference at the protein level.

ISs were mainly identified with ISSaga (Varani et al. 2011) and ISfinder (Siguier et al. 2006); ISCR elements were also identified (<http://medicine.cf.ac.uk/infect-immun/research/infection/antibacterial-agents/iscr-elements/>) and then manually curated for the corresponding intactness, terminal inverse repeats (IRs), and flanking direct repeats (DRs) defined with WebACT (Abbott et al. 2005). IS element orientation was defined according to the orientation of the embedded transposase gene (Roberts et al. 2008). VRprofile (Li et al. 2018) (available at <http://bioinfo-mml.sjtu.edu.cn/VRprofile>), Islander (Hudson et al. 2015), and IslandViewer (Dhillon et al. 2015) were used to search for other types of mobile genetic elements (MGEs) and bacterial secretion systems. In addition, prophages were predicted with PHAST (Arndt et al. 2016) and *Phage_Finder* (Fouts 2006), and boundaries were manually checked.

Data availability

The sequences of the main chromosome and auxiliary plasmid pSCL4 of strain F613-1 were submitted to GenBank under accession number CP016559 and CP016560, respectively.

Results

Clavulanic acid yield of *S. clavuligerus* F613-1

Streptomyces clavuligerus F613-1 is an industrial producer that can produce up to 4.87 g/l of clavulanic acid in 125-m³ fermentation tanks (Zhong et al. 2014). Similar to *S. clavuligerus* ATCC 27064, F613-1 showed productive sporulation on BSCA plates after 7–8 days (Table 1). Notably, no pigment could be detected on BSCA plates of F613-1, whereas ATCC 27064 produced a variety of burgundy pigments.

Mobile genome of *S. clavuligerus* F613-1.

The general characteristics of F613-1 genome are summarized in Table S1. F613-1 chromosomal *oriC* region seems to be located between 3,465,360 bp and 3,466,344 bp, closely adjacent to the *dnaA* and *dnaN* genes. We first compared the pan-genomes of strain F613-1 and ATCC 27064 and found a variety of F613-1-specific genes for which homologs were not identified in ATCC 27064 (Supplementary Table S2).

Table 1 General traits of the *S. clavuligerus* strain F613-1

Feature	F613-1	ATCC 27064
Time to spore formation on BSCA	7–8 days	7–8 days
Color of spores on BSCA	Gray	Gray
Pigment on BSCA	None	Burgundy
Clavulanic acid production in fermentation tanks	4.87 g/l	NA

Science China Life Sciences 56:591–600

NA not available

The aforesaid genes were observed to localize mainly onto a large RGP (genomic region of plasticity). Next, self-customized comparative genomics of the genome of *S.*

clavuligerus F613-1 (Figs. 1, 2, and Supplementary Figure S1–S2) revealed its remarkable potential to produce a large number of interesting novel secondary metabolite products. A few number of these secondary metabolite gene clusters were localized to the linear pSCL4-like plasmid. The ATCC 27064 mega-plasmid and homologous smaller plasmid in F613-1 were both densely packed with exchanged genetic material (Supplementary Table S3), suggesting that linear plasmids perform a dramatic role in genomic flux (Figs. 1, 3).

Subsequent comparative genomics suggest that significant transfer of genetic information among *Streptomyces* plasmids and chromosomes modulates their secondary metabolite repertoires. It is not unusual for the genetic material flow/exchange between different replicons, and we

Fig. 1 Unique genetic arrangement of the F613-1 large region of plasticity (RGP), likely originating from different backbones of *Streptomyces* species. **a** Left panel, RGP-carrying genes identified as accessory genes in the F613-1 genome. Histogram of BLASTn-based *H* values for all annotated protein-coding genes in the F613-1 chromosome compared with the ATCC 27064 and NRRL 3585 chromosomal sequences (color coded). The *H* value reflects the degree of similarity in terms of the length of match and the degree of identity at the nucleotide level between the matching gene in the subject genome and

the query gene. Genes were identified as conserved based on *H* values greater than 0.42 (black). Right panel, linear map of RGP (specific genes) is shown in white, whereas conserved genes are shown in black (100 Kb for each scale). Chromosome map of F613-1 with gene black/white-shade-coded based on the number of comparator genomes was identified as harboring a nucleotide sequence-conserved homolog. **b** Comparative schematics of the identified RGP and its homologs in other *Streptomyces* isolates indicate the RGP foreign origin and potential transferability. Homology genes are shown as color matched

Fig. 2 Genomic synteny analysis of *S. clavuligerus* F613-1, ATCC 27064, and NRRL 3585. Comparative analysis revealed a large genetic insertion/loss event in the left arm region of the F613-1

genome, and this RGP (region of genomic plasticity) was outlined since these three genomes were found nearly identical by the Mauve-based prediction

speculated that the megaplasmid originated from the backbone of a smaller replicon. Moreover, on the basis of complicated recombination of chromosomal arm regions and the smaller plasmid, the putative genetic exchange events observed in smaller size of the *S. clavuligerus* F613-1 genome are proposed to be attributed to the mega-plasmid pSCL4 in ATCC 27064. Additionally, homologous recombination between the original pSCL4- and pSCL2-like plasmid deduced degeneration of *S. clavuligerus* pSCL4 as described in Figs. 3a and 4a, suggesting its lower metabolic cost (almost identical gene clusters found in F613-1 and ATCC 27064, Fig. 4b and Supplementary Table S2). Prospectively, with the development and advancement of *Streptomyces* genomics, the proposed intermediate replicons could be captured, such as those indicated in Fig. 3b. Of note, we also identified integrase/transposase homologs in F613-1 (Supplementary Table S4); however, no intact IS elements were characterized *in silico*, denoting complex and frequent recombination in F613-1 leading to wide dissemination of a variety of IS remnants.

Diverse secondary metabolism potential of F613-1

A variety of secondary metabolite gene clusters were identified in the F613-1 genome and the cognate plasmid pSCL4 (Supplementary Table S3) via the antiSMASH 3.0 suite. Surprisingly, a large RGP (Supplementary Table S2) was found to carry a biosynthetic gene cluster for a non-ribosomal peptide synthetase (NRPS) that generates

teleocidin B. This RGP was found to range from positions 442,577 to 579,466 bp in the left arm of the F613-1 chromosome (Fig. 1), and a plasmid-borne paralog gene cluster related with biosynthesis of clavulanic acid was identified in the 707-Kb pSCL4-like plasmid (Fig. 4). Interestingly, all gene clusters associated with biosynthesis of clavulanic acid between F613-1 and ATCC 27064 are almost identical. Results obtained using the antiSMASH 3.0 server were evaluated, and 37 gene clusters (28 in chromosome and 9 in plasmid, respectively) predicted to be involved in secondary metabolism were catalogued into several major groups, including polyketide synthases (PKSs) and NRPSs listed in Supplementary Table S3. These gene clusters occupied 16.0% of the total genomic length, with a total length of 501 kb in the arms region and 230 Kb of biosynthetic gene clusters in the plasmid. It is clear that *S. clavuligerus* strain F613-1 has a more unique genomic arrangement and metabolic diversity in terms of plasmid than that of the NRRL 3585 and ATCC 27064, which have been identified to be homologs exhibiting dynamic architectures.

The DNA fragment located in the RGP shown in Fig. 1 is the largest region of genomic instability, carrying genes likely comprising a NRPS for teleocidin_B-like products (cluster 4, Supplementary Table S3). Moreover, the largest biosynthetic gene cluster, located in the right-arm region (cluster 27, 94.7-Kb) of the F613-1 chromosome, consists of bafilomycin biosynthesis-like type I PKS genes (Supplementary Table S3). The gene *BB341_RS27560* flanked

by this region was postulated to encode a regulator classified into the *Streptomyces* antibiotic regulatory protein (SARP) family, involved in the expression/activities of the adjacent PKS. In addition, a gene cluster (cluster 14) annotated as encoding a hybrid type I PKS and NRPS spanning a 55.2-Kb region showed very few limited sequence identities (only 9%) to the experimentally characterized 205.4-Kb kanamycin biosynthesis gene cluster (NCBI Accession No. AB254080).

Remnants of two type VII secretion systems in F613-1

Two potential type VII secretion system (T7SS) gene clusters (locus tags: *BB341_RS05795* – *BB341_RS05860*, and *BB341_RS25865*–*BB341_RS25885*) were identified in *S. clavuligerus* F613-1 (Supplementary Figure S3). Unlike the ESAT-6 secretion system (ESX) of *Mycobacteria*, which contributes to virulence (Houben et al. 2014), T7SS have been proposed to be involved in sporulation of *S. coelicolor*

Fig. 4 The genetic context of the genetic flux between pSCL2 and pSCL4 contributing to the genomic architecture of F613-1 pSCL4 plasmid. **a** The F613-1 pSCL4 plasmid shares a nearly identical architecture of previously reported sequences from pSCL4 of strain ATCC 27064; however, a genetic exchange between pSCL2 and pSCL4 could assist F613-1 high yield of clavulanic acid. **b** The colinear alignment of chromosome-carrying and pSCL4-borne gene clusters responding for clavulanic acid production. *cvmG*: putative Gra-orf36-like protein; *cvmP*: putative peptidylarginine deiminase; *cvmH*: putative hydrolase; *cvm13* and *cvm7*: putative asparaginase; *cvm12*: putative transcriptional regulator; *cvm11*: putative efflux protein; *cas*: clavamate synthase isozyme; *cvm3*: putative oxidoreductase; *cvm2*: putative ribulose-5-phosphate epimerase; *cvm1*: putative aldo/keto reductase family 2 enzyme; *cvm4*: putative deacetylcephalosporin C; *cvm5*: putative luciferase; *cvm6*: putative pyridoxal phosphate-dependent aminotransferase; *cvm9*: putative DNA binding protein; *cvm10*: putative serine threonine protein kinase; *cvm7par*: putative transcriptional regulator; *cvm6par*: putative pyridoxal phosphate-dependent aminotransferase; *oat*: putative pyridoxal phosphate-dependent aminotransferase; *pah*: proclavamate amidinohydrolase isoenzyme; *bls*: beta-lactam synthetase isoenzyme; *ceaS*: carboxyethyl arginine synthase isoenzyme; *orfA*: hydroxymethyltransferase; *orfB*: YjgF family regulator; *orfC*: aminotransferase; *orfD*: dehydratase; *orfE*, *orfF* and *orfH*: membrane protein; *orfG*, *orfJ* and *orfK*: hypothetical protein; *orfI*: LysR family regulator

Fig. 3 The genetic architecture and possible genetic exchanges mechanism of the F613-1 plasmid pSCL4. **a** The alignment map indicates the proposed horizontal transferability between different replicons. Homology genes are shown as color matched. **b** The working model for plasmid pSCL4 degeneration is postulated resulting from

the fragment insertion and excision of the transitional intermediates; however, a more expanded landscape of pSCL4-related *Streptomyces* genomics is still required to define its origin. It is feasible that candidate direct and reverse repeat fragments likely contributed to 'in cis' and 'in trans' mating and excision patterns

a **b**

and to modulate development (Akpe San Roman et al. 2010). A T7SS in *S. scabies* was also found to modulate development and had no detectable association with virulence (Fyans et al. 2013). Surprisingly, two co-existing T7SS gene remnants in the *S. clavuligerus* F613-1 genome were characterized *in silico* (Supplementary Figure S3). Furthermore, above loci coding for bacterial secretion systems could result from heavily horizontally transferred insertion and acquisition, leading to the presence of less-conserved and non-intact bacterial secretion systems (Abrusci et al. 2014; Bi et al. 2013; Fyans et al. 2013; Houben et al. 2014; Li et al. 2015). Because of the lack of conservation of some secretion systems and disruption of others, experimental validation should focus on a limited number of atypical secretion systems, which has complicated the assessment of genetic and functional diversities. Bacterial secretion systems are often encoded within one or a few neighboring operons (Abrusci et al. 2014; Bi et al. 2013; Fyans et al. 2013; Houben et al. 2014; Li et al. 2015). This finding could facilitate the identification of new bacterial secretion systems in regions adjacent to known loci.

Discussion

In this study, we have used the vast body of knowledge accumulated from experimental studies and computational models to describe the genomic composition and genetic organization of F613-1, to evaluate and improve its significant industrial prospects. Different from the genome ATCC 27064 strain NRRL3585, only one chromosome and one pSCL4-like plasmid were identified and sequenced in strain F613-1. It was reported that the ‘minimized’ genome and plasmid elimination could facilitate higher yield of secondary products (Lu et al. 2017; Wu et al. 2012; Xu et al. 2016). In addition, our previous study revealed genome-wide gene expression profiles in F613-1 during high-level production of clavulanic acid, including differential changes of genes for primary metabolism and upregulation of biosynthetic gene clusters involved in clavulanic acid production. Changes in primary metabolism (RNA-Seq GEO: GSE104738) seemed to divert metabolic fluxes toward biosynthetic precursors for clavulanic acid, and the holomycin biosynthesis cluster was silent (Qin et al. 2017). These findings are similar to the transcriptome and proteome data from another clavulanic acid-producing industrial strain (Medema et al. 2011; Unsaldi et al. 2017). Our analyses suggest that genetic losses and insertions may have facilitated the fitness of F613-1 as an industrial strain, including enhancing clavulanic acid yield.

The mobilization of an NRPS-like secondary metabolite gene cluster onto the F613-1 linear chromosome was identified and, overall, the genomic patterns in strain F613-1 seem to have developed to acquire efficient biosynthesis and

substrate tolerance. The pSCL4 plasmid likely evolved by heavy genetic material exchanges between a larger plasmid (like ATCC 27064 pSCL4) and other genomic replicons, such as the pSCL2 plasmid. In addition, the spread of secondary metabolite genes onto large linear plasmids such as pSCL4 indicates that these plasmids could be diversified vehicles for horizontal transfer and serve as intermediate chassis for studies on the minimal *Streptomyces* genome. Based on this scenario, strain F613-1 carrying a small genome and high biosynthetic capabilities may be suitable for further modification for other industrial processes.

We sequenced the genome of F613-1 for its high yield of clavulanic acid. Systematics and omics-guided investigations on industrial microbes have gradually developed efficient approaches for chemical drug discovery and production optimization (Xu et al. 2016). At least 16.0% of the compact genome of F613-1 has been found to synthesize secondary metabolites collated from the prediction of the antiSMASH 3.0 server. Using REBASE-based searches (Roberts et al. 2015), we also found that F613-1 encodes no known restriction enzymes, suggesting that this strain could be deficient in preventing genetic exchange. The complete genome sequence of F613-1 and our analyses should enhance understanding of the secondary metabolism and complicated regulation of clavulanic acid biosynthesis, and thereby facilitate in-depth research for improving industrial production of clavulanic acid. In view of these findings, we hypothesize that F613-1, with its naturally reduced genome, can be modified to be an exogenous expression vector for ectopic genes of interest, identified from the microbiome extracted from industrial and/or environmental samples.

Acknowledgements We thank Lunan Pharmaceutical Group Corporation for providing *S. clavuligerus* F613-1. This study was supported by the National Natural Science Foundation of China (81471921, 31800118, 31670115 and 31700065), Shandong Natural Science Foundation (ZR2017BC040), the Innovation Project of Shandong Academy of Medical Sciences, and the China Postdoctoral Science Foundation (2017M621965). We appreciated Professor Susan Howard for her critical reading of this manuscript.

Compliance with ethical standards

Conflict of interest All the authors declare no financial or non-financial conflict of interests.

References

- Abbott JC, Aanensen DM, Rutherford K, Butcher S, Spratt BG (2005) WebACT—an online companion for the Artemis. Compar Tool Bioinform 21:3665–3666. <https://doi.org/10.1093/bioinformatics/bti601>
- Abrusci P, McDowell MA, Lea SM, Johnson S (2014) Building a secreting nanomachine: a structural overview of the T3SS.

- Curr Opin Struct Biol 25:111–117 <https://doi.org/10.1016/j.sbi.2013.11.001>
- Akpe San Roman S, Facey PD, Fernandez-Martinez L, Rodriguez C, Vallin C, Del Sol R, Dyson P (2010) A heterodimer of EsxA and EsxB is involved in sporulation and is secreted by a type VII secretion system in *Streptomyces coelicolor*. Microbiology 156:1719–1729. <https://doi.org/10.1099/mic.0.037069-0>
- Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucl Acids Res 44:W16–21 <https://doi.org/10.1093/nar/gkw387>
- Bentley SD et al (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3. Nature 417(2):141–147. <https://doi.org/10.1038/417141a>
- Bi D, Liu L, Tai C, Deng Z, Rajakumar K, Ou HY (2013) SecReT4: a web-based bacterial type IV secretion system resource. Nucl Acids Res 41:D660–D665 <https://doi.org/10.1093/nar/gks1248>
- Blin K, Medema MH, Kottmann R, Lee SY, Weber T (2017) The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. Nucl Acids Res 45:D555–D559 <https://doi.org/10.1093/nar/gkw960>
- Cao G, Zhong C, Zong G, Fu J, Liu Z, Zhang G, Qin R (2016) Complete genome sequence of *Streptomyces clavuligerus* F613-1, an industrial producer of clavulanic acid genome announcements. <https://doi.org/10.1128/genomeA.01020-16>
- Carattoli A et al (2014) *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrobial Agents Chemother 58:3895–3903. <https://doi.org/10.1128/AAC.02412-14>
- Cruz-Morales P et al (2013) The genome sequence of *Streptomyces lividans* 66 reveals a novel tRNA-dependent peptide biosynthetic system within a metal-related genomic island Genome. Biol Evol 5:1165–1175. <https://doi.org/10.1093/gbe/evt082>
- Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. <https://doi.org/10.1101/gr.2289704>
- Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA. with glimmer. Bioinformatics 23:673–679. <https://doi.org/10.1093/bioinformatics/btm009>
- Dhakal D, Lim SK, Kim DH, Kim BG, Yamaguchi T, Sohng JK (2018) Complete genome sequence of *Streptomyces peucetius* ATCC 27952, the producer of anticancer anthracyclines and diverse secondary metabolites. J Biotechnol 267:50–54. <https://doi.org/10.1016/j.jbiotec.2017.12.024>
- Dhillon BK et al (2015) IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. Nucl Acids Res 43:W104–108 <https://doi.org/10.1093/nar/gkv401>
- Fouts DE (2006) *Phage_Finder*: automated identification and classification of prophage regions in complete bacterial genome sequences. Nucl Acids Res 34:5839–5851. <https://doi.org/10.1093/nar/gkl732>
- Fyans JK, Bignell D, Loria R, Toth I, Palmer T (2013) The ESX/type VII secretion system modulates development, but not virulence, of the plant pathogen. *Streptomyces scabies*. Mol Plant Pathol 14:119–130. <https://doi.org/10.1111/j.1364-3703.2012.00835.x>
- Galperin MY, Makarova KS, Wolf YI, Koonin EV (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucl Acids Res 43:D261–D269 <https://doi.org/10.1093/nar/gku1223>
- Gomez-Escribano JP, Castro JF, Razmilic V, Chandra G, Andrews B, Asenjo JA, Bibb MJ (2015) The *Streptomyces leeuwenhoekii* genome: *de novo* sequencing and assembly in single contigs of the chromosome, circular plasmid pSLE1 and linear plasmid pSLE2. BMC Genomics 16:485. <https://doi.org/10.1186/s12864-015-1652-8>
- Gonzalez A, Rodriguez M, Brana AF, Mendez C, Salas JA, Olano C (2016) New insights into paulomycin biosynthesis pathway in *Streptomyces albus* J1074 and generation of novel derivatives by combinatorial biosynthesis. Microbial Cell Fact 15:56. <https://doi.org/10.1186/s12934-016-0452-4>
- Holmes NA, Devine R, Qin Z, Seipke RF, Wilkinson B, Hutchings MI (2018) Complete genome sequence of *Streptomyces formicae* KY5, the formicamycin producer. J Biotechnol 265:116–118. <https://doi.org/10.1016/j.jbiotec.2017.11.011>
- Houben EN, Korotkov KV, Bitter W (2014) Take five—Type VII secretion systems of *Mycobacteria*. Biochimica et biophysica Acta 1843:1707–1716. <https://doi.org/10.1016/j.bbarmac.2013.11.003>
- Hudson CM, Lau BY, Williams KP (2015) Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. Nucl Acids Res 43:D48–53 <https://doi.org/10.1093/nar/gku1072>
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinf 11:119 <https://doi.org/10.1186/1471-2105-11-119>
- Ian E et al (2014) Genomics of sponge-associated *Streptomyces* spp. closely related to *Streptomyces albus* J1074: insights into marine adaptation and secondary metabolite biosynthesis potential. PloS One 9:e96719. <https://doi.org/10.1371/journal.pone.0096719>
- Jia N, Ding MZ, Luo H, Gao F, Yuan YJ (2017) Complete genome sequencing and antibiotics biosynthesis pathways analysis of *Streptomyces lydicus*. 103 Sci Rep 7:44786. <https://doi.org/10.1038/srep44786>
- Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW (2007) RNAMmer: consistent and rapid annotation of ribosomal RNA genes. Nucl Acids Res 35:3100–3108. <https://doi.org/10.1093/nar/gkm160>
- Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucl Acids Res 32:11–16. <https://doi.org/10.1093/nar/gkh152>
- Li J, Yao Y, Xu HH, Hao L, Deng Z, Rajakumar K, Ou HY (2015) SecReT6: a web-based resource for type VI secretion systems found in bacteria. Environ Microbiol 17:2196–2202. <https://doi.org/10.1111/1462-2920.12794>
- Li P, Tai C, Deng Z, Gan J, Oggioni MR, Ou HY (2016) Identification and characterization of chromosomal *relBE* toxin-antitoxin locus in *Streptomyces cattleya* DSM46488. Sci Rep 6:32047 <https://doi.org/10.1038/srep32047>
- Li J, Tai C, Deng Z, Zhong W, He Y, Ou HY (2018) VRprofile: gene-cluster-detection-based profiling of virulence and antibiotic resistance traits encoded within genome sequences of pathogenic bacteria Briefings. Bioinformatics 19:566–574. <https://doi.org/10.1093/bib/bbw141>
- Lin YS, Kieser HM, Hopwood DA, Chen CW (1993) The chromosomal DNA of *Streptomyces lividans* 66 is linear. Mol Microbiol 10:923–933
- Lu C, Wu H, Su X, Bai L (2017) Elimination of indigenous linear plasmids in *Streptomyces hygroscopicus* var. jinggangensis and *Streptomyces* sp. FR008 to increase validamycin A and candicidin productivities. Appl Microbiol Biotechnol 101:4247–4257. <https://doi.org/10.1007/s00253-017-8165-0>
- Luo H, Zhang CT, Gao F (2014) Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. Front Microbiol 5:482. <https://doi.org/10.3389/fmicb.2014.00482>
- Medema MH et al (2010) The sequence of a 1.8-mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. Genome Biol Evol 2:212–224. <https://doi.org/10.1093/gbe/evq013>
- Medema MH et al (2011) Genome-wide gene expression changes in an industrial clavulanic acid overproduction strain of *Streptomyces*

- clavuligerus*. Microbial Biotechnol 4:300–305. <https://doi.org/10.1111/j.1751-7915.2010.00226.x>
- Overbeek R et al (2014) The SEED and the Rapid Annotation of microbial genomes using subsystems technology (RAST). Nucl Acids Res 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>
- Qin R, Fu J, Zong G, Wang Z, Cao G (2017) Characterization of clavulanic acid high-yield *Streptomyces clavuligerus* industrial strain by transcriptome analysis (in chinese). Biotechnol Bull 33:235–243. <https://doi.org/10.13560/j.cnki.biotech.bull.1985.2017-0275>
- Roberts AP et al (2008) Revised nomenclature for transposable genetic elements. Plasmid 60:167–173. <https://doi.org/10.1016/j.plasmid.2008.08.001>
- Roberts RJ, Vincze T, Posfai J, Macelis D (2015) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucl Acids Res 43:D298–299 <https://doi.org/10.1093/nar/gku1046>
- Ruckert C et al (2015) Complete genome sequence of *Streptomyces lividans* TK24. J Biotechnol 199:21–22. <https://doi.org/10.1016/j.jbiotec.2015.02.004>
- Shao Y, He X, Harrison EM, Tai C, Ou HY, Rajakumar K, Deng Z (2010) mGenomeSubtractor: a web-based tool for parallel *in silico* subtractive hybridization analysis of multiple bacterial genomes. Nucl Acids Res 38:W194–W200. <https://doi.org/10.1093/nar/gkq326>
- Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. Nucl Acids Res 34:D32–D36. <https://doi.org/10.1093/nar/gkj014>
- Song JY et al (2010) Draft genome sequence of *Streptomyces clavuligerus* NRRL 3585, a producer of diverse secondary metabolites. J Bacteriol 192:6317–6318. <https://doi.org/10.1128/JB.00859-10>
- Unsalidi E, Kurt-Kizildogan A, Voigt B, Becher D, Ozcengiz G (2017) Proteome-wide alterations in an industrial clavulanic acid producing strain of *Streptomyces clavuligerus*. Synth Syst Biotechnol 2:39–48. <https://doi.org/10.1016/j.synbio.2016.10.003>
- Varani AM, Siguier P, Gournbeyre E, Charneau V, Chandler M (2011) ISSaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol 12:R30 <https://doi.org/10.1186/gb-2011-12-3-r30>
- Volff JN, Altenbuchner J (1998) Genetic instability of the *Streptomyces*. chromosome. Mol Microbiol 27:239–246
- Wu H, Qu S, Lu C, Zheng H, Zhou X, Bai L, Deng Z (2012) Genomic and transcriptomic insights into the thermo-regulated biosynthesis of validamycin in *Streptomyces hygroscopicus* 5008. BMC Genomics 13:337. <https://doi.org/10.1186/1471-2164-13-337>
- Xu MJ et al (2016) Deciphering the streamlined genome of *Streptomyces xiamenensis* 318 as the producer of the anti-fibrotic drug candidate xiamenmycin. Sci Rep 6:18977. <https://doi.org/10.1038/srep18977>
- Zhong CQ, Cao GX, Jin XQ, Wang FS (2014) Studies on the formation and forming mechanism of the related substance E in potassium clavulanate production by HPLC-MS/MS Brazilian. J Pharm Sci 50:391–399
- Zuo G, Hao B (2015) CVTree3 Web Server for Whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy genomics. Proteomics Bioinform 13:321–331. <https://doi.org/10.1016/j.gpb.2015.08.004>