Annals of Microbiology (2018) 68:677–688
https://doi.org/10.1007/s13213-018-1374-7

ORIGINAL ARTICLE

# Unraveling the cellulolytic and hemicellulolytic potential of two novel *Streptomyces* strains

Ricardo Rodrigues de Melo[1,2] · Geizecler Tomazetto[3] · Gabriela Felix Persinoti[1] · Hélia Harumi Sato[2] · Roberto Ruller[4] · Fabio Marcio Squina[3]

Received: 10 May 2018 / Accepted: 27 August 2018 / Published online: 8 September 2018


## Abstract

The *Streptomyces* spp. are notorious plant biomass decomposers in soil environments, but only few strains were biochemically and genetically characterized. Here, we employed functional screening along with genomic sequencing for identification of novel lignocellulolytic *Streptomyces* strains. *Streptomyces* strains isolated from soil were functional screened based on their cellulolytic and hemicellulolytic capacities by enzymatic plate assays containing carboxymethylcellulose (CMC) and beechwood xylan as sole carbon source. Subsequently, genomes of *Streptomyces* strains were sequenced, annotated, and interpreted to correlate their genetic contents with biochemical properties. Among the 80 bacterial isolates that were screened for enzymatic activity, two *Streptomyces* strains (named as F1 and F7) exhiting higher endoglucanase and endoxylanase activities were selected for biochemical and genomic characterization. After cultivation on steam-pretreated sugarcane bagasse-based medium, the supernatant of the strains F1 and F7 exhibited enzymatic activity against different substrates, such as arabinan, rye arabinoxylan, β-glucan, starch, CMC, xylan, and chitin. Furthermore, strain F7 was able to degrade pectin, mannan, and lichenan. The genomic analysis of both strains revealed a diversity of carbohydrate-active enzymes. The F1 and F7 genomes encode 33 and 44 different types of glycosyl hydrolases families, respectively. Moreover, the genomic analysis also identified genes related to degradation of lignin-derived aromatic compounds. Collectively, the study revealed two novel *Streptomyces* strains and further insights on the degradation capability of lignocellulolytic bacteria, from which a number of technologies can arise, such as saccharification processes.

**Keywords** *Streptomyces* · Sugarcane bagasse · Genomics · Enzymatic assays · Bioconversion

Ricardo Rodrigues de Melo and Geizecler Tomazetto contributed equally to this work.

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s13213-018-1374-7) contains supplementary material, which is available to authorized users.

✉ Fabio Marcio Squina
fabio.squina@prof.uniso.br

[1] Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, SP, Brazil

[2] Departamento de Ciência de Alimentos, Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brazil

[3] Programa de Processos Tecnológicos e Ambientais, Universidade de Sorocaba (UNISO), Sorocaba, SP, Brazil

[4] Instituto de Biociências, Universidade Federal de Mato Grosso do Sul (UFMS), Campo Grande, MS, Brazil

## Introduction

The depletion of fossil fuels and environmental concerns has driven global politics towards the development of renewable energy resources. Lignocellulosic biomass represents the largest reservoir for renewable energy and value-added chemical production (Gomez et al. 2008; Himmel et al. 2010; Van Dyk and Pletschke 2012). The term lignocellulosic biomass refers to nonfood materials available from plants, including rice straw, cotton straw, corn stover, sugarcane bagasse, wood, grass, and others (Gomez et al. 2008; Sanderson 2011; Van Dyk and Pletschke 2012).

The transformation of the lignocellulosic biomass into fermentable sugars and biofuel production is a challenging process. Generally, lignocellulosic biomasses are composed of variable contents of cellulose, hemicellulose, and lignin. These polymers are highly organized and interlinked among themselves into recalcitrant structure (Gomez et al. 2008;

Himmel et al. 2010). However, in nature ecosystems, lignocellulosic biomass is constantly converted by microorganisms, which efficiently degrade lignocellulosic material (Himmel et al. 2010; McGuire and Treseder 2010; Koeck et al. 2014). The biomass-degrading microorganisms are found in various environments, including hot spring pool (Hamilton-Brehm et al. 2010; Brumm et al. 2015), cow rumen (Hess et al. 2011), biogas reactor (Hanreich et al. 2013; Tomazetto et al. 2015), and soil (Zhou et al. 2014; Jiménez et al. 2016; López-Mondéjar et al. 2016). These environments represent resource for isolation of novel microorganisms and enzymes involved on plant biomass conversion. For instance, *Streptomyces* have been reported to play an important role in the carbon cycle and plant biomass deconstruction in soils (Chater et al. 2010; Bontemps et al. 2013; Větrovský et al. 2014; Book et al. 2016). Indeed, *Streptomyces* species genomes have revealed several carbohydrate-active enzymes (CAZymes), including cellulases, hemicellulases, and lytic polysaccharide monooxygenases (Book et al. 2014, 2016; Pinheiro et al. 2016). Moreover, a recent study described three *Streptomyces* strains (pl6, pl88, and pr55) isolated from soil capable to metabolize polysaccharides and poplar lignin (Větrovský et al. 2014).

Although several *Streptomyces* strains have been described as plant biomass decomposer, the molecular and biochemical understanding on how these microorganisms degrade lignocellulosic biomass remain rather fragmented and restricted to a few strains, e.g., I1.2 (Pinheiro et al. 2016), SirexAA-E, SDPB6 (Book et al. 2014), ATexAB-D23, and LaPpAH-95 (Book et al. 2016).

In this sense, soil samples from distinct locations were used as inoculum on *Streptomyces* selective-agar plates for isolation of their species. Isolates obtained were screened based on their cellulolytic and hemicellulolytic capacities by enzymatic plate assays using carboxymethylcellulose (CMC) and beechwood xylan as sole carbon source. Among the 80 bacterial isolates, two *Streptomyces* strains demonstrated ability to secrete a wide range of enzymes against plant-based polysaccharides. To further investigate the mechanisms of lignocellulosic biomass degradation, the genomes of both strains were sequenced and analyzed regarding their CAZymes profiles. Their genomes encode several CAZymes confirming the cellulolytic and hemicellulolytic phenotypes. In addition, genomic analysis indicated that both strains presented gene cluster related to lignin degradation.

## Material and methods

### *Streptomyces* strains isolation

Eighty bacterial isolates were isolated from soil samples collected from different places at Brazil (São Paulo, Brazil: 22°49′8.861″ S, 47°3′39.085″ W; Minas Gerais, Brazil: 19°57′08.1″ S, 44°12′55.7″ W; Minas Gerais, Brazil: 20°36′29.624″ S, 46°2′30.739″ W, Table S1). Briefly, soil samples corresponding to the upper 10 cm were collected and transported to the laboratory into sterile bags. These samples were mixed with calcium carbonate (1:1), ground with a pestle and air-dried (El-Nakeeb and Lechevalier 1963). For each sample collected, 4 g of air-dried soil was mixed with 40-mL sterilized water and stirred vigorously. Several dilutions ($10^{-1}$ to $10^{-4}$) were spread on isolation medium ISP 4 agar plates (Shirling and Gottlieb 1966) and incubated at 30 °C for 3 days. Single colonies were transferred to medium ISP 2 agar plates (Shirling and Gottlieb 1966) and grown at 30 °C for 6–8 days. The procedure was repeated until pure cultures were obtained. The isolated strains were stored at − 80 °C as mixtures of mycelial fragments in 20% (*v/v*) glycerol or in medium ISP 2.

### Qualitative and quantitative screening

To screen cellulolytic and hemicellulolytic microorganisms among the isolate ones, they were cultivated on Bushnell Haas Broth (BHB, Sigma Aldrich®) mineral salts medium agar plates (g/L: $K_2HPO_4$ 1, $KH_2PO_4$ 1, $CaCl_2{\cdot}2H_2O$ 0.02, $NH_4NO_3$ 1, $FeCl_3$ 0.05, $MgSO_4{\cdot}7H_2O$ 0.2, and agar 15, pH 7.0) supplemented with 0.5% (*w/v*) of beechwood xylan or carboxymethylcellulose (CMC). After incubation at 30 °C for 48–72 h, agar plates were stained with Congo red and destained with 1-M NaCl (Teather and Wood 1982). Colonies showing clear halos were selected as indicative for the CMC (endoglucanase activity) and xylan (endoxylanase) degradation. Of the 80 strains screened, 48 showed a visual enzymatic activity (Table S1). The 14 strains exhibiting the highest enzymatic activities index (E.A.I) were selected for further analysis described as follows.

These strains were selected and initially grown on medium ISP 2 agar plates at 30 °C for 7 days. Each culture was inoculated $10^6$ spores $mL^{-1}$ (1%, *v/v*) into 25 mL of BHB medium and supplemented with beechwood xylan or CMC (0.5%, *w/v*) and incubated at 30 °C for 6 days in a shaker (New Brunswick Scientific, New Jersey, USA) at 180 rpm. Supernatants were taken daily to monitor the corresponding enzymatic activity (Table S2) using the DNS method as described in the section below ("Enzymatic assays" section). All experiments were done in biological triplicates.

### Cultivation and enzymatic production

Based on endoglucanase and endoxylanase activity results, the two strains exhibiting the highest enzymatic activities, namely F1 and F7, were cultivated on steam-pretreated

sugarcane bagasse (SCB) as a sole carbon source for enzyme production. SCB was pretreated by steam explosion at 200 °C for 15 min, as described by Rocha et al. (2012). The strains were grown on medium ISP 2 agar plates at 30 °C for 7 days, and then $10^6$ spores $mL^{-1}$ (1%, *v*/*v*) was inoculated into 80 mL of BHB medium supplemented with 1.0% (*w*/*v*) steam-pretreated SCB. The flasks were incubated at 30 °C for 10 days in a rotary shaker at 180 rpm. Samples of the culture supernatants were taken daily to evaluate their enzymatic activity against plant-based polysaccharides as described below.

## Enzymatic assays

The enzymatic activity (International Units, U) was performed using distinct substrates. All the polysaccharides were purchased from Sigma-Aldrich or Megazyme [beechwood xylan, rye arabinoxylan, β-glucan (barley), sugar beet arabinan, debranched arabinan, carboxymethylcellulose, tamarind xyloglucan, starch, icelandic moss lichenan, chitin from shrimp shells, arabinogalactan, mannan (ivory nut), and citrus pectin]. The enzymatic reactions were carried out in a miniaturized fashion by mixing 20 μL of culture media supernatant, 50 μL of the distinct substrates (0.5%, *w*/*v*), and 30 μL of sodium phosphate buffer (0.1 M) at pH 6. Reactions were incubated at 50 °C in a Thermostat® (Eppendorf, Hamburg, Germany) for 30 min or 18 h and stopped with the addition of 100 μL of 3,5 dinitrosalicylic acid (DNS) following immediately by boiling for 5 min at 99 °C (Miller 1959). The solutions were analyzed at 540 nm using the Infinite M200®spectrophotometer (Tecan, Switzerland) to measure the release of reducing sugars. One unit (1 U) of enzymatic activity corresponded to the formation of 1 μmol of reducing sugar equivalent per minute under the assay conditions. Total protein was measured using microtiter plates with Bio-Rad protein assay reagent (Bio-Rad Laboratories, USA), employing a procedure based on Bradford's method (Bradford 1976). Bovine serum albumin was used as standard. All experiments were done in biological triplicates.

## Identification of *Streptomyces* strains

The strains F1 and F7 were grown in ISP 2 broth at 30 °C in flasks agitated at 180 rpm for 3 days, and their cells harvested by centrifugation at 10,000 rpm for 10 min. The cell pellets were washed twice with sterile water. Genomic DNA was isolated using FastDNA SPIN Kit for soil (MP Biomedicals, Irvine, CA) according to the manufacturer's instructions. The quality of the total DNA was assessed by electrophoresis in a 0.8% agarose gel stained with ethidium bromide and visualized using UV transilluminator. The DNA concentration was measured by fluorimetry (Qubit® 2.0 Fluorometer-Life Technologies. Carlsbad, California, EUA) using the BR Qubit® dsDNA Assay (Life Technologies, Carlsbad, California, EUA).

To determine the relatedness of strains with their closest described relative *Streptomyces*, 16S rRNA gene sequences were amplified using polymerase chain reaction (PCR) with primers 27f (5′ AGAGTTTGATCMTGGCTCAG 3′) and 1492r (5′ TACCTTGTTACGACTT 3′). All PCR reactions were performed in 50 μL containing 50 ng of individual genomic DNA, 1-mM dNTP, 10 pmoL each primer, 2-mM $MgCl_2$, and 1 U of Taq DNA polymerase (Thermo Fischer Scientific, Waltham, USA). The PCR conditions consisted as follows: an initial denaturation at 94 °C for 3 min, followed by 30 cycles of 94 °C for 60 s, 55 °C for 60 s, and 72 °C for 2 min, and a final extension at 72 °C for 4 min. PCR products were purified using Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare Bio-Sciences, Pittsburgh, USA) and sequenced using a DNA ABI PRISM 377 Genetic Analyzer system (Applied Biosystems, USA). BLASTn search analysis revealed that strains F1 and F7 shared more than 97% 16S rRNA gene sequences identity with members of the genus *Streptomyces*.

## Genome sequencing, assembly, and annotation

Genomic DNA of both strains was used for the construction of libraries using Nextera® DNA Library Preparation Kit (Illumina, San Diego, CA), according to the manufacturer's protocol. The genomic libraries were sequenced on Illumina HiSeq sequencing platform at NGS sequencing facility at CTBE, generating approximately 8 and 6 million 2 × 100-bp reads, respectively.

Raw fastq files were quality checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and processed with Trimmomatic version 0.32 (Bolger et al. 2014) to quality trimming, adaptor removal, and minimum length filtering. High-quality reads were further analyzed using ProDeGe (Tennessen et al. 2015) to remove possible contaminants and then assembled using Spades version 3.6.2 (Bankevich et al. 2012), applying several kmers ($k = 21$, 27, 33, 55, 77) and careful mode option as parameters.

The draft genome sequences of strains F1 and F7 were uploaded into the IMG/ER annotation pipeline (Markowitz et al. 2009) for automatic prediction of genes using Prodigal (Hyatt et al. 2010). Functional annotation of predicted proteins was obtained using Prokka automatic pipeline (Seemann 2014). Briefly, it uses hierarchical searches against different databases such as bacterial proteins from RefSeq genomes, UniProt, Pfam, and TIGRFAM models. The Aragorn tool (Laslett and Canback 2004) was used to predict tRNA genes whereas rRNA genes were identified using RNAmmer 1.2. Signal peptides, and transmembrane helices were predicted

by SignalP and TMHMM, respectively. To determine the carbohydrate-active enzymes (CAZYmes) profiles for strains F1 and F7, putative protein sequences were annotated using dbCAN database version 4 (Yin et al. 2012) using HMMER3.1b package. Results obtained were manually assessed based on their functional annotation prediction and conserved protein domains to identify the corresponding best matching CAZymes.

## Phylogenetic analysis

The taxonomic assignment of strains F1 and F7 within *Bacteria* domain was determined based on 16S rRNA sequence analysis and multilocus phylogenetic analysis. Phylogenetic tree was reconstructed based on 16S rRNA gene sequences of strains isolated and closest related species with validly published name. All sequences were

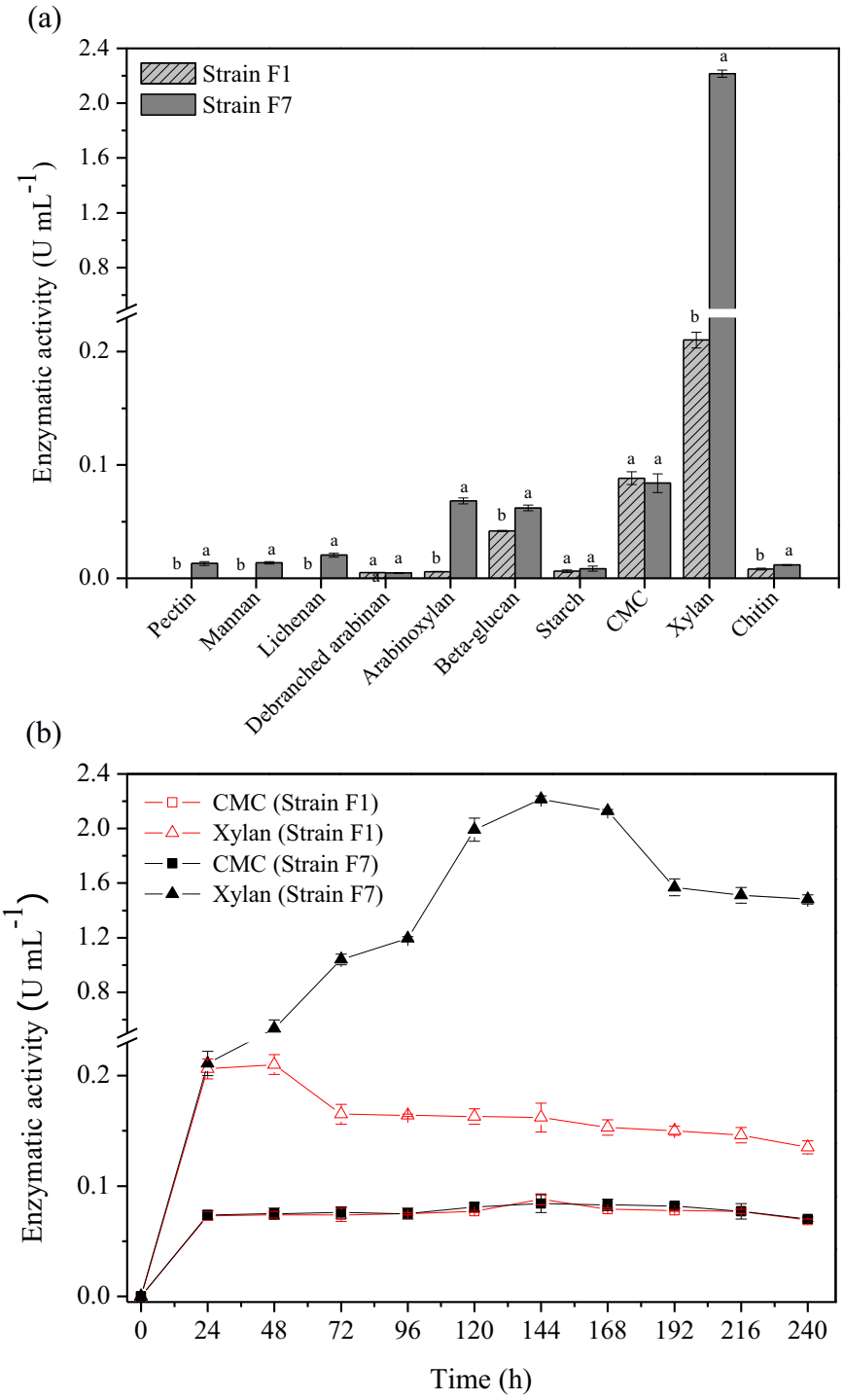


**Fig. 1** Enzymatic activities detected in the supernatants from strains F1 and F7 grown on steam-pretreated sugarcane bagasse (SCB) as the carbon source. **a** Bar graph showing the maximum enzymatic activity of each substrate tested. **b** Variation of the endoglucanase and endoxylanase activities in the supernatants from strains F1 and F7 cultivated for 10 days on SCB. The data represent the means ± standard errors from three technical replicates; lowercase letters indicate values statistically different between strains (Tukey's multiple comparisons test: $p < 0.01$). Abbreviations: CMC, carboxymethylcellulose

aligned using the ClustalX tool (Larkin et al. 2007) and manually refined. From those aligned, pairwise distances were calculated using the maximum composite likelihood approach, and phylogenetic tree was constructed applying the neighbor-joining method, as implemented in MEGA version 6 (Tamura et al. 2013). Bootstrap tests with 1000 replications/iterations were calculated to assess the node confidence level.

Multilocus phylogenetic analysis was done as previous described Book et al. (2014). Briefly, TIGRFAM models were used to search for single-copy genes conserved among 172 *Streptomyces* genome sequences available at RefSeq database, and the genome of *Kitasatospora setae* was considered as an out group. The sequences recovered from each genome were aligned using Mafft v7.299 (Katoh and Standley 2013) and the resulting alignment was concatenated using FASconCAT-G v1.02 (Kück and Longo 2014). The phylogenetic tree was generated using FastTree version 2.1.8 (Price et al. 2010) under WAG as a substitution model and the final tree was visualized using iTOL version 3 (Letunic and Bork 2007).

In addition, the relatedness among strains F1 and F7 and their closest relative *Streptomyces* strains was determined based on average nucleotide identity (gANI) and the fraction of orthologous genes (Aligment Fraction, AF) calculated with the MiSI (Microbial Species Identifier) method (Varghese et al. 2015).

## Statistical analysis

Statistical analysis of enzymatic activities was performed using parametric analysis of variance (ANOVA) with Tukey's multiple comparisons test ($\alpha = 0.05$). Calculation was made by Minitab v18 (Minitab Inc., PA, USA).

## Nucleotide sequence accession number

The draft genome sequences of strains F1 and F7 are available at the NCBI with the accession numbers FKJI03000000 and FKJH01000000, respectively.

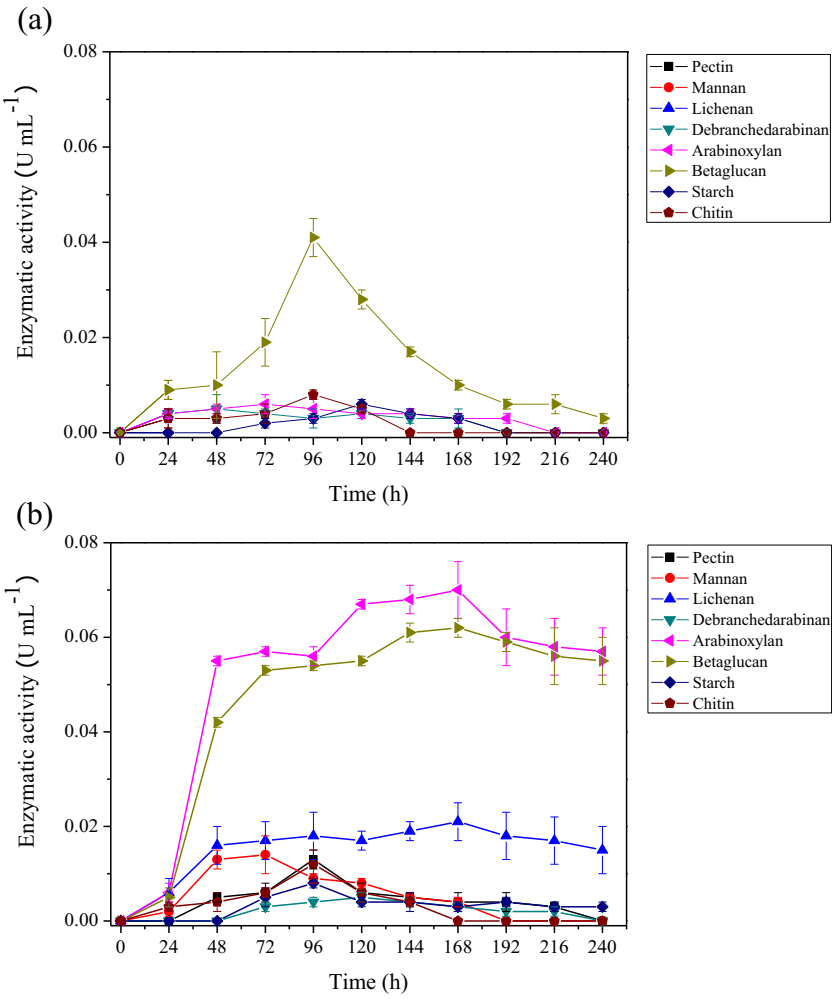


**Fig. 2** Variation of the enzymatic activities detected in the supernatants from strains F1 (**a**) and F7 (**b**) cultivated for 10 days on steam-pretreated sugarcane bagasse (SCB) as the carbon source. The data represent the means ± standard errors from three technical replicates

**Table 1** Genome statistics for strains F1 and F7

| Attribute | Strain | |
|---|---|---|
| | F1 | F7 |
| Genome size (bp) | 8,142,296 | 7,327,391 |
| DNA coding (bp) | 7,100,494 | 6,492,599 |
| DNA G + C (bp) | 5,915,378 | 5,326,280 |
| GC content (%) | 72.65 | 72.69 |
| DNA contigs | 69 | 66 |
| Total genes | 7355 | 6548 |
| Protein-coding genes | 7262 | 6463 |
| RNA genes | 93 | 86 |
| Genes with function prediction | 4093 | 3478 |
| Genes with Pfam domains | 5526 | 5172 |
| Genes with signal peptides | 447 | 439 |
| Genes with transmembrane helices | 1641 | 1577 |
| CRISPR repeats | 2 | 2 |

## Results and discussion

### Strain screening and enzymatic activity evaluation

Of the 80 bacterial isolates from soil samples grown on CMC—or beechwood xylan-agar plates, 48 exhibited degradation halos around them indicating endoglucanase and endoxylanase activities (Supplementary file 1: Table S1). Among them, 14 strains exhibiting the largest degradation halos were selected for further analysis using submerged fermentation.

Overall, the supernatant of the 14 *Streptomyces* isolates obtained from submerged fermentations containing CMC or beechwood xylan showed endoglucanase and endoxylanase activities (Supplementary file 1: Table S2). Endoglucanase activity varies between 0.18 to 0.40 U $mL^{-1}$ and endoxylanase 1.97 to 10.78 U $mL^{-1}$ using xylan as sole carbon source, and endoglucanase activity between 0.10 to 0.17 U $mL^{-1}$ and endoxylanase 0.12 to 0.28 U $mL^{-1}$ using CMC as carbon source. Considering the enzymatic activities (Table S2), two strains (F1 and F7) exhibited the highest endoglucanase and endoxylanase activities and they were selected for enzymatic production and genomic analysis.

Enzymatic assays of the strains F1 and F7 supernatants obtained from cultivation on SCB revealed that these microorganisms were capable to secrete a set of enzymes involved in the hydrolysis of distinct polysaccharides, including arabinan, rye arabinoxylan, β-glucan, starch, CMC, xylan, and chitin (Fig. 1a). The strain F7 was also able to secrete enzymes for pectin, mannan, and lichenan degradation. Overall, the supernatant of strain F7 exhibited higher enzymatic activities for degradation of plant-based polysaccharides than strain F1 during 10 days of the submerged fermentation on SCB (Figs. 1b and 2). No enzymatic activity was detected for degradation of arabinan sugar beet, xyloglucan, and arabinogalactan in both strains, even though the majority of genes encoding enzymes involved in these polysaccharides were predicted (see "CAZYme profiles" section), which could be because SCB was not capable to induce the expression of these enzymes.

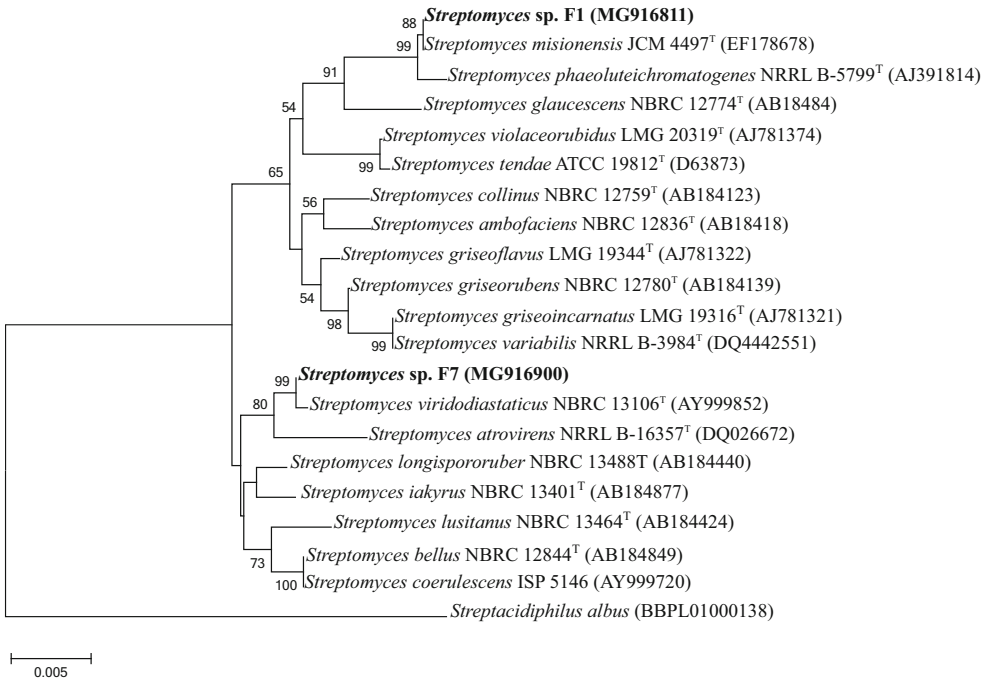


**Fig. 3** Phylogenetic analysis of 16S rRNA gene sequences of strains F1 and F7 (printed in bold) relative to the most closely related strains of the genus *Streptomyces*. The phylogenetic tree was constructed using neighbor-joining (NJ) algorithm. Bootstrap values higher than 60% are shown (1000 replications for bootstrapping were done). Genbank accession numbers are shown in parentheses and type strain with superscript "T". The sequence of *Streptacidiphilus albus* was used as an out group

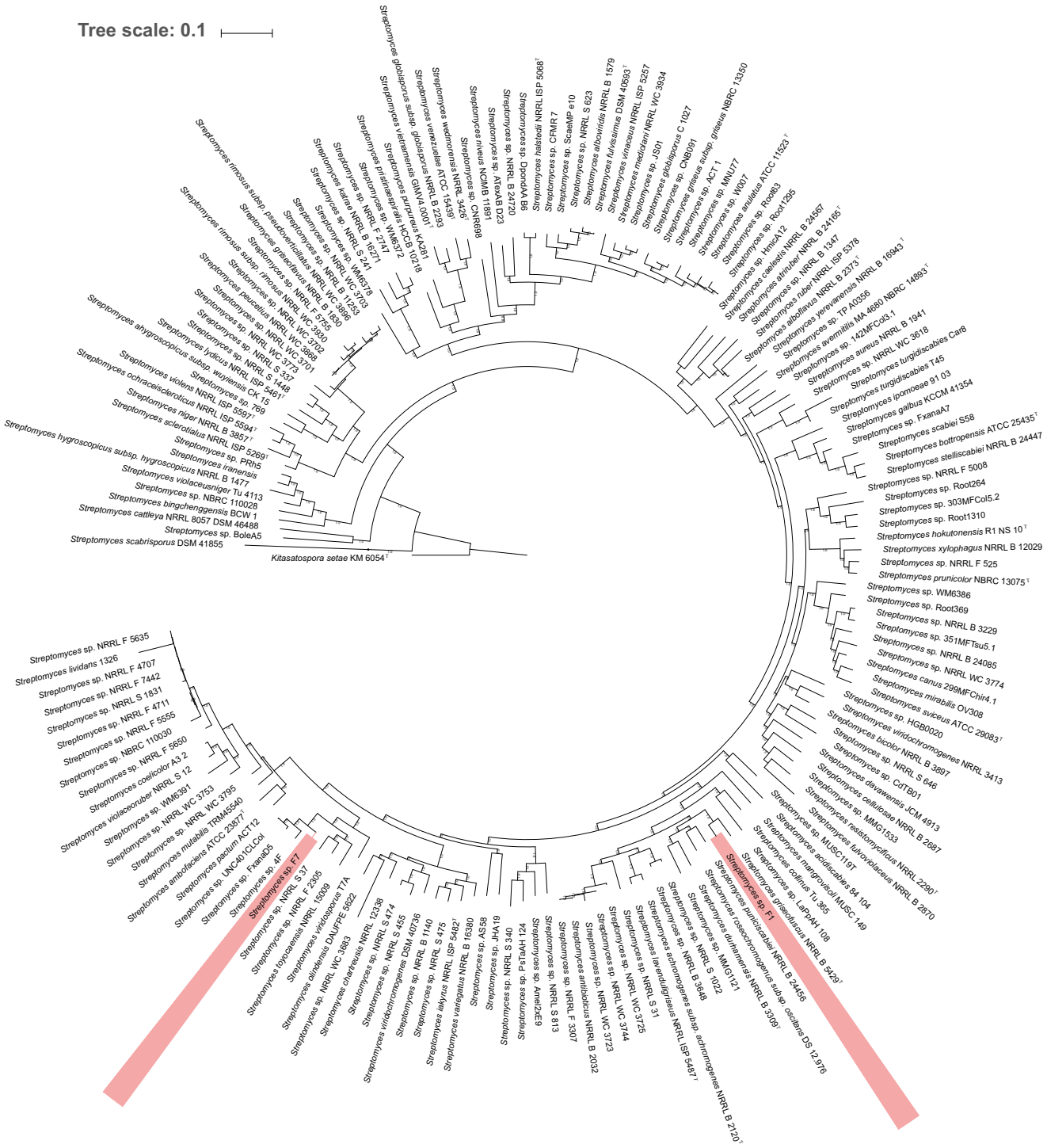


**Fig. 4** Multilocus phylogenetic analysis of *Streptomyces* genus. The multilocus phylogenetic tree was generated from a concatenated alignment of 288 conserved genes across all genomes of *Streptomyces*. Bootstrap values (expressed as percentages of 1000 replicates) higher than 0.8 are shown at branch nodes

Our findings corroborated with previous studies and the cellulolytic and hemicellulolytic potential of *Streptomyces* spp. (Chater et al. 2010; Bontemps et al. 2013; Book et al. 2014; Pinheiro et al. 2016). In a recently study using comparative genomic, transcriptomic, and biochemical analysis (Book et al. 2016), 29 *Streptomyces* strains have shown a relatively high rate of cellulose degrading activity. Pinheiro et al. (2016) reported that *Streptomyces* spp. were able to secrete a set of enzymes involved in degradation of distinct natural carbohydrates,

including cellulose, xylan, mannan, starch, chitin, pectin, and β-glucan.

## General features and phylogenetic analysis

Based on cellulolytic and hemicellulolytic profiles of strains F1 and F7, we decided to genetically characterize both strains to uncover their metabolic pathways dedicated to biomass degradation. The genomes of both strains were sequenced on the Illumina HiSeq system using the 2 × 100-pb sequencing strategy. The genomes of strains F1 and F7 resulted in 69 and 66 contigs, respectively, ranging from 10,212 to 760,841 bases (Table 1). Strain F1 consists of an 8.1 megabase pair (Mb) chromosome coding 7262 protein-coding sequences (CDS). The strain F7 genome is 7.33 Mb in size and contains 6463 CDSs. The genomes display a similar GC content of 72%.

The 16S rRNA gene sequence analyses placed strains F1 and F7 within the genera *Streptomyces* (Fig. 3). Pairwise analysis revealed that strain F1 was closely related to *Streptomyces misionensis* NRRL B-3230$^T$ sharing 99% 16S rRNA gene sequence identity. Whereas, strain F7 showed 100% 16S rRNA gene sequence identity with *Streptomyces viridodiastaticus* IFO 13106$^T$. It is important to underline that the percentage of identity obtained from 16S rRNA gene sequences alignment was calculated disregarding a gap (26 and 31 nucleotides) that are present only in strains F1 and F7.

Previous studies reported that 16S rRNA gene sequence analysis is not suitable to distinguish closely related *Streptomyces* species since several *Streptomyces* type strains share completely identical 16S rRNA gene sequences (Guo et al. 2008; Antony-Babu et al. 2017). Therefore, to refine the phylogenetic relationship of the strains isolated within *Streptomyces* genus, a multilocus phylogenetic tree was constructed based on 288 single-copy conserved genes found across strains F1, F7, and 172 *Streptomyces* genomes available at RefSeq database (Fig. 4). Indeed, multilocus analysis indicated a phylogenetic relationship different from the analysis based on 16S rRNA gene sequences. While, strain F1 forms a monophyletic clade with *Streptomyces griseofuscus* NRRL B 5429$^T$, which is a type species of this genus, strain F7 is closely related to species with currently not validly published names.

In addition, gANI and AF among strains F1 and F7 and their closest related species were calculated to support the clades. Strains F1 and F7 shared a maximum pairwise similarity of 90.28% and 94.30% ANI and AF of 0.73 and 0.82 across *Streptomyces* strains within their respectively clades (Supplementary file 1: Table S3). Considering the threshold at minimum of 95 or 96.5% of ANI, which corresponds to 70% DNA-DNA hybridization, and AF above 0.6 between two whole genome as strong evidence for same species (Goris et al. 2007; Varghese et al. 2015), values of ANI and AF for strains F1 and F7 alongside the multilocus tree confirmed that both strains represent novel species with genus *Streptomyces*.

## CAZYme profiles

The CAZymes (Cantarel et al. 2009) present in *Streptomyces* strains F1 and F7 were identified using dbCAN web resource (Yin et al. 2012). Both strains have similar profile of CAZymes genes in their genomes (Table 2). Strain F1 genome encodes 85 glycoside hydrolyses (GH), 31 glycosyltransferases (GT), 18 carbohydrate esterases (CE), one polysaccharide lyases (PL), 31 carbohydrate-binding motifs (CBM), and nine classified as auxiliary activities (AA), which are CAZymes with redox activities (Tables 3 and S4). Whereas, strain F7 genome possesses 100 GHs, 22 GTs, 26 CEs, five PLs, 26 CBMs, and nine AAs. A detailed CAZymes gene prediction indicated that strains F1 and F7 genomes encode 33 and 44 different types of GH families, respectively, which are potentially involved in degradation of starch, chitin, hemicellulose, and cellulose (Table 3 and Supplementary file 2: Table S1). However, only strain F7 genome encodes pectin-, arabinogalactan-, lichenan-, and mannan-degradation from families PL1, PL3, GH35, GH64, GH5, and GH2, respectively.

As expected, *Streptomyces* strains F1 and F7 do not encode cellulosomes, which are the enzymatic complex composed of cohesins and dockerins modules and several lignocellulolytic enzymes with CBMs domains (Smith and Bayer 2013). However, both strains possess GHs connected to CBMs, which is not surprising since lignocellulolytic enzymes

**Table 2** Comparison of CAZyme genomic profile between strains F1 and F7

| | F1 | F7 |
|---|---|---|
| Genome size (Mb) | 8.14 | 7.32 |
| Protein-coding genes | 7262 | 6463 |
| No. of CAZymes proteins | 175 | 193 |
| % CAZymes proteins [a] | 2.36 | 2.92 |
| Total GH [b] | 85 | 100 |
| Total PL [c] | 1 | 5 |
| Total GT [d] | 31 | 22 |
| Total CE [e] | 18 | 26 |
| Total AA [f] | 9 | 9 |
| Total CBM [g] | 31 | 31 |

[a] Carbohydrate-active enzymes
[b] Glycosyl hydrolases
[c] Polysaccharide lyases
[d] Glycosyltransferase
[e] Carbohydrate esterase
[f] Auxiliary activity
[g] Carbohydrate-binding modules

**Table 3** Summarizes of CAZyme profile encoded by strains F1 and F7

| Substrate | Main known activity | CAZYme family | EC number[1] | Number of genes predicted | |
|---|---|---|---|---|---|
| | | | | F1 | F7 |
| Pectin | Pectinesterase | | 3.1.1.11 | 0 | 1 |
| | Pectate lyase | PL1 | 4.2.2.2 | 0 | 2 |
| | Pectate lyase | PL3 | 4.2.2.2 | 0 | 1 |
| | Alpha-L-rhamnosidase | AA10 | 3.2.1.40 | 0 | 1 |
| Starch | Alpha-amylase | GH13 | 3.2.1.1 | 5 | 4 |
| | Alpha-glucosidase | GH13 | 3.2.1.20 | 3 | 4 |
| | Alpha-glucosidase | GH31 | 3.2.1.20 | 1 | 1 |
| | 4-Alpha-glucanotransferase | GH77 | 2.4.1.25 | 0 | 1 |
| | Starch phosphorylase | GT35 | 2.4.1.1 | 1 | 1 |
| Chitin | Chitinase | GH18 | 3.2.1.14 | 3 | 4 |
| | Chitinase | GH19 | 3.2.1.14 | 0 | 1 |
| | Beta-N-acetylhexosaminidase | GH3 | 3.2.1.52 | 2 | 2 |
| | Hexosaminidase | GH20 | 3.2.1.52 | 3 | 3 |
| | N-acetylglucosamine 6-phosphate deacetylase | CE9 | 3.5.1.25 | 0 | 1 |
| Arabinogalactan | Endo-beta-1,4-galactanase | GH35 | 3.2.1.89 | 0 | 1 |
| Lichenan | Endo-1,3-beta-glucanase | GH64 | 3.2.1.39 | 1 | 2 |
| Xyloglucan (hemicellulose) | Alpha-D-xyloside | GH31 | 3.2.1.177 | 1 | 1 |
| Arabinan (hemicellulose) | Alpha-L-arabinofuranosidase | GH62 | 3.2.1.55 | 1 | 0 |
| | Alpha-L-arabinofuranosidase | GH51 | 3.2.1.55 | 1 | 1 |
| Xylan (hemicellulose) | Endo-1,4-beta-xylanase | GH10 | 3.2.1.8 | 4 | 5 |
| | Endo-1,4-beta-xylanase | GH11 | 3.2.1.8 | 1 | 2 |
| | Xylan 1,4-beta-xylosidase | GH39 | 3.2.1.37 | 0 | 1 |
| | Beta-xylosidase | GH43 | 3.2.1.37 | 1 | 2 |
| | Alpha-N-arabinofuranosidase | GH51 | 3.2.1.55 | 1 | 1 |
| | Alpha-L-arabinofuranosidase | GH54 | 3.2.1.55 | 4 | 0 |
| | Alpha-glucuronidase | GH67 | 3.2.1.139 | 0 | 1 |
| Mannan (hemicellulose) | Endo-1,4-beta-mannosidase | GH5 | 3.2.1.78 | 0 | 1 |
| | Beta-mannosidase | GH2 | 3.2.1.25 | 0 | 1 |
| Cellulose | Beta-glucosidase | GH1 | 3.2.1.21 | 4 | 5 |
| | Beta-glucosidase | GH3 | 3.2.1.21 | 1 | 3 |
| | Endoglucanase | GH5 | 3.2.1.4 | 1 | 1 |
| | Endoglucanase | GH6 | 3.2.1.4 | 2 | 3 |
| | Cellulose 1,4-beta-cellobiosidase | GH6 | 3.2.1.91 | 0 | 1 |
| | Endoglucanase | GH9 | 3.2.1.4 | 0 | 1 |
| | Cellulose 1,4-beta-cellobiosidase | GH12 | 3.2.1.91 | 1 | 0 |

[1] Enzyme commission number

connected to CBMs are broadly found in *Bacteria* domain (Tomazetto et al. 2015; López-Mondéjar et al. 2016; Pinheiro et al. 2016). The F7 genome encodes different CAZymes associated with CBMs domains, including endoglucanases, chitinase, xylanases, and α-amylase (Supplementary file 2: Table S1). Whereas, strain F1 genome was found to encode CBMs associated with only chitinase and endoglucanase. CBMs domains bind to plant cell wall polysaccharides and facilitating GHs activity (Campos et al. 2016). In addition, both strains contain genes encoding lytic polysaccharide monooxygenases (LPMOs) belonging to family AA10 (Supplementary file 2: Table S1). Chitinolytic and cellulolytic AA10 genes are frequently found in *Streptomyces* genomes (Book et al. 2014). Nevertheless, strains F1 and F7 genomes have only AA10 genes related to chitin degradation.

## Aromatic compound degradation pathways

Several bacteria have been reported to metabolize lignin, releasing aromatics compounds that are imported into the cell

for catabolism (Bugg et al. 2011; Brown and Chang 2014). Among the lignin-degrading bacteria, few *Streptomyces* strains were described to degrade lignin or lignin-derived compounds, such as catechol, protocatechuate, and gentisate (Ishiyama et al. 2004; Gottschalk et al. 2008; Davis et al. 2013; Větrovský et al. 2014). In this context, the corresponding metabolic pathways in strains F1 and F7 were manually reconstructed from their genome sequence data.

The genomic analyses of strains F1 and F7 revealed several genes involved on protocatechuate and gentisate catabolism (Fig. 5 and Supplementary file 2: Table S2). For gentisate catabolism, identical gene clusters consisting of four genes were found in both strains genomes. Sequence analysis predicted four encoding genes, including DNA-binding transcriptional regulator (IclR family), gentisate 1,2 dioxygenase, fumarylacetoacetate hydrolase, and maleylpyruvate isomerase. For protocatechuate catabolism, all genetic determinants for this aromatic degradation were found only in strain F1. These seven genes are organized in a cluster and flaked by transcriptional regulators. It is important to mention the presence of transcriptional regulator in all these clusters. For instance, the protocatechuate cluster in *Streptomyces* F7 genome consists of four genes involved on aromatic degradation preceded by a transcriptional regulator. Members of IclR family frequently control genes whose products are involved in the degradation of aromatics (Molina-Henares et al. 2006). These findings suggest that the clusters could be expressed and allow the strains to use these aromatic compounds as carbon sources.

## Concluding remarks

This work disclose two novel *Streptomyces* strains isolated from soil, able to grow on CMC, xylan, and steam-pretreated SCB as sole carbon sources, and secrete a range of hydrolytic enzymes. The enzymatic assays indicated that strain F7 was more efficient in the degradation of natural

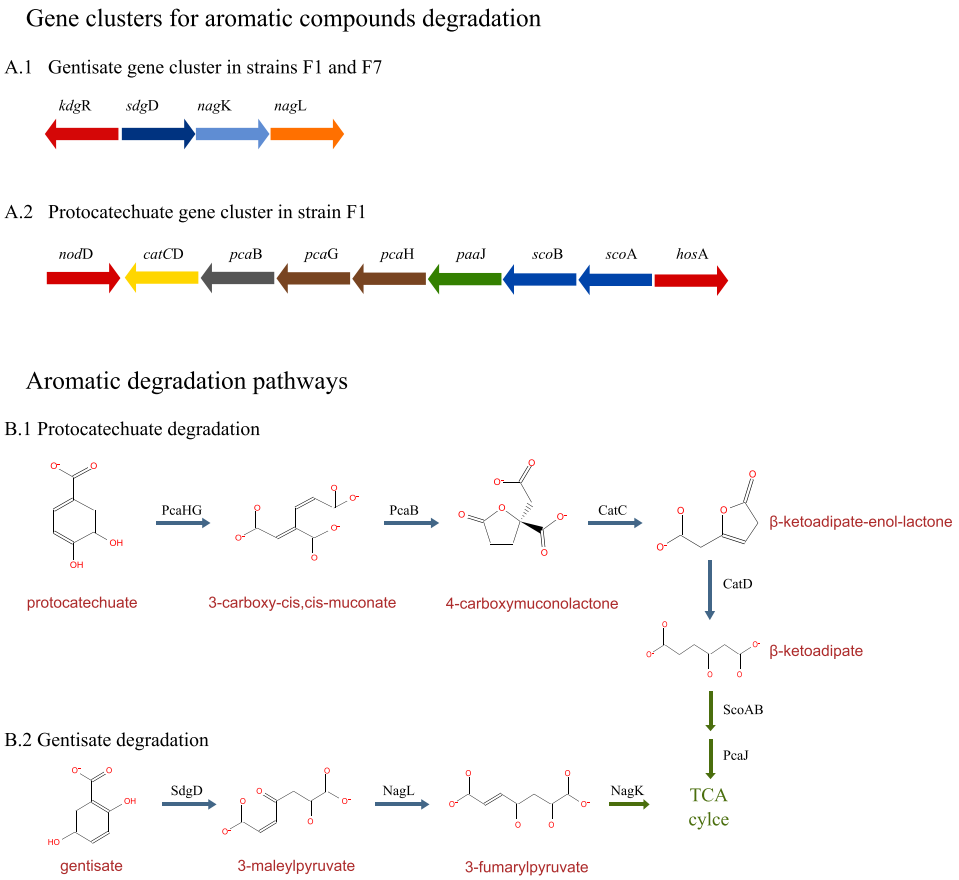


**Fig. 5** **a**) Schematic representation of the gene clusters encoding the gentisate and protocatechuate degradation and their corresponding metabolism found in F1 and F7 genomes. *A.1* and *A.2*, organization of the gentisate and protocatechuate clusters, respectively. The putative functions of the gene products are as follows: *kdg*R, transcriptional regulator KdgR (MarR family transcriptional regulator); *sdg*D, gentisate 1,2-dioxygenase; *nag*K, fumarylacetoacetate hydrolase family; *nag*L, maleylpyruvate isomerase; *nod*D, nodulation protein D (MarR Family); *cat*D, 3-oxoadipate enol-lactonase; *pca*B, β-carboxymuconatecycloisomerase; *pca*B,3-carboxy-cis,cis-muconate cycloisomerase; *pca*GH, protocatechuate 3,4-dioxygenase, α and β-subunits; *paa*J, 3-oxoadipyl-CoA thiolase; *sco*AB, 3-oxoacid CoA-transferase subunit A and B; *hos*A, transcriptional regulator (MarR family). **b**) Metabolic pathways of the protocatechuate and gentisate degradation (*B.1* and *B.2*, respectively). The enzyme names are shown above the arrows. Abbreviation: TCA cycle, tricarboxylic acid cycle

carbohydrates than strain F1. This observation was confirmed by genomic analysis: the genome of strain F1 encodes 85 glycoside hydrolases (GHs) which comprise 33 different types of GH families, and strain F7 harbors 100 GHs representing 44 GH families. Finally, the identification of gene clusters encoding enzymes dedicate to aromatic compound degradation suggests that the strains could also perform lignin degradation.

**Funding** This work was financially supported by grants from National Counsel of Technological and Scientific Development—CNPq (442333/2014-5 and 310186/2014-5) and São Paulo Research Foundation—FAPESP (14/50371-8 and 15/50590-4). R.R.M and G.T. were supported by FAPESP fellowship (2017/14253-9; 2015/23279-6, respectively). We thank the Brazilian Bioethanol Science and Technology Laboratory CTBE /NGS Sequencing Facility.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflicts of interest.

**Research involving human participants and/or animals** Not applicable.

**Informed consent** Not applicable.

## References

Antony-Babu S, Stien D, Eparvier V, Parrot D, Tomasi S, Suzuki MT (2017) Multiple Streptomyces species with distinct secondary metabolomes have identical 16S rRNA gene sequences. Sci Rep 7:1–8

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120

Bontemps C, Toussaint M, Revol PV, Hotel L, Jeanbille M, Uroz S, Turpault MP, Blaudez D, Leblond P (2013) Taxonomic and functional diversity of *Streptomyces* in a forest soil. FEMS Microbiol Lett 342:157–167

Book AJ, Lewin GR, McDonald BR, Takasuka TE, Doering DT, Adams AS, Blodgett JAV, Clardy J, Raffa KF, Fox BG, Currie CR (2014) Cellulolytic *Streptomyces* strains associated with herbivorous insects share a phylogenetically linked capacity to degrade lignocellulose. Appl Environ Microbiol 80:4692–4701

Book AJ, Lewin GR, McDonald BR, Takasuka TE, Wendt-Pienkowski E, Doering DT, Suh S, Raffa KF, Fox BG, Currie CR (2016) Evolution of high cellulolytic activity in symbiotic *Streptomyces* through selection of expanded gene content and coordinated gene expression. PLoS Biol 14:1–21

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Brown ME, Chang MCY (2014) Exploring bacterial lignin degradation. Curr Opin Chem Biol 19:1–7

Brumm P, Land ML, Hauser LJ, Jeffries CD, Chang YJ, Mead DA (2015) Complete genome sequence of *Geobacillus* strain Y4.1MC1, a novel CO-utilizing *Geobacillus thermoglucosidasius* strain isolated from bath hot spring in Yellowstone National Park. Bioenergy Res 8:1039–1045

Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R (2011) Pathways for degradation of lignin in bacteria and fungi. Nat Prod Rep 28:1883–1896

Campos BM, Liberato MV, Alvarez TM, Zanphorlin LM, Ematsu GC, Barud H, Polikarpov I, Ruller R, Gilbert HJ, De Mattos Zeri AC, Squina FM (2016) A novel carbohydrate-binding module from sugar cane soil metagenome featuring unique structural and carbohydrate affinity properties. J Biol Chem 291:23734–23743

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res 37:D233–D238

Chater KF, Biró S, Lee KJ, Palmer T, Schrempf H (2010) The complex extracellular biology of *Streptomyces*. FEMS Microbiol Rev 34:171–198

Davis JR, Goodwin L, Teshima H, Detter C, Tapia R, Han C, Huntemann M, Wei C-L, Han J, Chen A, Kyrpides N, Mavrommatis K, Szeto E, Markowitz V, Ivanova N, Mikhailova N, Ovchinnikova G, Pagani I, Pati A, Woyke T, Pitluck S, Peters L, Nolan M, Land M, Sello JK (2013) Genome sequence of *Streptomyces viridosporus* strain T7A ATCC 39115, a lignin-degrading *Actinomycete*. Genome Announc 1:e00416–e00413

El-Nakeeb MA, Lechevalier HA (1963) Selective isolation of aerobic *Actinomycetes*. Appl Microbiol 11:75–77

Gomez LD, Steele-king CG, Mcqueen-mason SJ (2008) Sustainable liquid biofuels from biomass: the writings on the walls. New Phytol 178(3):473–485

Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91

Gottschalk LMF, Bon EPS, Nobrega R (2008) Lignin peroxidase from *Streptomyces viridosporus* T7A: enzyme concentration using ultrafiltration. Appl Biochem Biotechnol 147:23–32

Guo YP, Zheng W, Rong XY, Huang Y (2008) A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics. Int J Syst Evol Microbiol 58:149–159

Hamilton-Brehm SD, Mosher JJ, Vishnivetskaya T, Podar M, Carroll S, Allman S, Phelps TJ, Keller M, Elkins JG (2010) *Caldicellulosiruptor obsidiansis* sp. nov., an anaerobic, extremely thermophilic, cellulolytic bacterium isolated from obsidian pool, Yellowstone National Park. Appl Environ Microbiol 76:1014–1020

Hanreich A, Schimpf U, Zakrzewski M, Schlüter A, Benndorf D, Heyer R, Rapp E, Pühler A, Reichl U, Klocke M (2013) Metagenome and metaproteome analyses of microbial communities in mesophilic biogas-producing anaerobic batch fermentations indicate concerted plant carbohydrate degradation. Syst Appl Microbiol 36:330–338

Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331:463–467

Himmel ME, Xu Q, Luo Y, Ding S-Y, Lamed R, Bayer EA (2010) Microbial enzyme systems for biomass conversion: emerging paradigms. Biofuels 1:323–341

Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119

Ishiyama D, Vujaklija D, Davies J (2004) Novel pathway of salicylate degradation by *Streptomyces* sp. strain WA46. Appl Environ Microbiol 70:1297–1306

Jiménez DJ, de Lima Brossi MJ, Schückel J, Kračun SK, Willats WGT, van Elsas JD (2016) Characterization of three plant

biomass-degrading microbial consortia by metagenomics- and metasecretomics-based approaches. Appl Microbiol Biotechnol:1–15

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780

Koeck DE, Pechtl A, Zverlov VV, Schwarz WH (2014) Genomics of cellulolytic bacteria. Curr Opin Biotechnol 29C:171–183

Kück P, Longo GC (2014) FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front Zool 11:81

Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16

Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23: 127–128

López-Mondéjar R, Zühlke D, Větrovský T, Becher D, Riedel K, Baldrian P (2016) Decoding the complete arsenal for cellulose and hemicellulose deconstruction in the highly efficient cellulose decomposer *Paenibacillus* O199. Biotechnol Biofuels 9:104

Markowitz VM, Mavromatis K, Ivanova NN, Chen IMA, Chu K, Kyrpides NC (2009) IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics 25:2271–2278

McGuire KL, Treseder KK (2010) Microbial communities and their relevance for ecosystem models: decomposition as a case study. Soil Biol Biochem 42:529–535

Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugars. Anal Chem:426–429

Molina-Henares AJ, Krell T, Eugenia Guazzaroni M, Segura A, Ramos JL (2006) Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors. FEMS Microbiol Rev 30:157–186

Pinheiro GL, de Azevedo-Martins AC, Albano RM, de Souza W, Frases S (2016) Comprehensive analysis of the cellulolytic system reveals its potential for deconstruction of lignocellulosic biomass in a novel *Streptomyces* sp. Appl Microbiol Biotechnol

Price MN, Dehal PS, Arkin AP (2010) FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS One:e9490

Rocha GJM, Gonçalves AR, Oliveira BR, Olivares EG, Rossell CEV (2012) Steam explosion pretreatment reproduction and alkaline delignification reactions performed on a pilot scale with sugarcane bagasse for bioethanol production. Ind Crop Prod 35:274–279

Sanderson K (2011) A chewy problem. Nature 474:S12–S14

Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069

Shirling EB, Gottlieb D (1966) Methods for caracterization of *Streptomyces* species. Int J Syst Bacteriol 16:313–340

Smith SP, Bayer EA (2013) Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. Curr Opin Struct Biol 23:686–694

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Teather RM, Wood PJ (1982) Use of Congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl Environ Microbiol 43:777–780

Tennessen K, Andersen E, Clingenpeel S, Rinke C, Lundberg DS, Han J, Dangl JL, Ivanova N, Woyke T, Kyrpides N, Pati A (2015) ProDeGe: a computational protocol for fully automated decontamination of genomes. ISME J 10:1–4

Tomazetto G, Hahnke S, Koeck DE, Wibberg D, Maus I, Pühler A, Klocke M, Schlüter A (2015) Complete genome analysis of *Clostridium bornimense* strain M2/40T: a new acidogenic *Clostridium* species isolated from a mesophilic two-phase laboratory-scale biogas reactor. J Biotechnol 232:38–49

Van Dyk JS, Pletschke BI (2012) A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-factors affecting enzymes, conversion and synergy. Biotechnol Adv 30:1458–1480

Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis T, Mavrommatis K, Kyrpides NC, Pati A (2015) Microbial species delineation using whole genome sequences. Nucleic Acids Res 43(14):6761–6711

Větrovský T, Steffen KT, Baldrian P (2014) Potential of cometabolic transformation of polysaccharides and lignin in lignocellulose by soil *Actinobacteria*. PLoS One 9(2):e89108

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445–W451

Zhou Y, Pope PB, Li S, Wen B, Tan F, Cheng S, Chen J, Yang J, Liu F, Lei X, Su Q, Zhou C, Zhao J, Dong X, Jin T, Zhou X, Yang S, Zhang G, Yang H, Wang J, Yang R, Eijsink VGH, Wang J (2014) Omics-based interpretation of synergism in a soil-derived cellulose-degrading microbial community. Sci Rep 4:1–6