Genome Sequences of Five Streptomyces Strains Isolated at Microscale
Matthieu Nicault, Abdoul-Razak Tidjani, Éric Gelhaye, Cyril Bontemps, Pierre Leblond

To cite this version:

HAL Id: hal-02801725
https://hal.univ-lorraine.fr/hal-02801725
Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Genome Sequences of Five *Streptomyces* Strains Isolated at Microscale

Matthieu Nicault,a,b Abdoul-Razak Tidjani,a Eric Gelhaye,b Cyril Bontemps,a Pierre Leblonda

aUniversité de Lorraine, INRA, DynAMic, Nancy, France
bUniversité de Lorraine, INRA, IAM, Nancy, France

ABSTRACT The genomes of five *Streptomyces* strains belonging to the same soil community were sequenced and assembled. The strains, which were isolated at microscale, belonged to different *Streptomyces* species. This sample provides access to understand the functioning of a *Streptomyces* community in an ecological context.

*Streptomyces* organisms are soil-dwelling bacteria that play a central role in soil homeostasis (biogeochemical cycle and biotic interactions) (1). Their large linear chromosomes code for a wide variety of specialized metabolites (e.g., antibiotics and antifungals) and enzymes of high biotechnological and medical interest (2, 3). Here, we report the genome sequences of five *Streptomyces* strains isolated from four grains of soil. These grains had a size on the order of cubic millimeters and were collected at a distance of a few centimeters from each other in a French forest soil. Each grain was diluted in sterile water and serially diluted on *Streptomyces* isolation medium (4). The five selected strains had mycelial and sporulating phenotypes. After isolation and purification, spores were inoculated in Hickey-Tresner broth medium (30°C for 30 h), from which genomic DNA was extracted and purified (5). This procedure included a proteinase K treatment and a chloroform extraction step in order to remove the covalently bound terminal protein (6). An assembly was generated from Oxford Nanopore single reads using Canu (v1.7.1). Sequencing was carried out using the sequencing kits SQK-LSK109 and EXP-NBD104, the flow cell FLO-MIN106 (R9.4.1 revD) on the GridION system (v3.1.8), and the base-calling program Guppy (v2.0.8). When strains were multiplexed, Porechop (v0.2.4) was used for demultiplexing (and adaptor trimming). The coverage ranged from 73 to 322. One large contig covering the whole genome of each strain was obtained (except for RPA4-5, for which two contigs were obtained). A circular extrachromosomal element of about 67 kb was predicted for strain RLA2-12. Illumina paired-end libraries (35 to 51 bp) were created from genomic DNA fragments using the NextSeq 500/550 high-output kit v2 (75 cycles) with the Westburg next-generation sequencing (NGS) DNA library preparation kit (Illumina TruSeq LT primers) (Table 1). Paired-end reads were generated using an Illumina NextSeq instrument (vNB552053). The minimum read size was set to 10 bp, and adaptor trimming was performed using Cutadapt (v1.15). The coverage of the paired-end reads varied between 50X and 90X. The final genome sequences were obtained by aligning the Illumina sequences to the Nanopore reference contigs using Pilon (v1.22) (7). Default parameters were used for all software. The total genome sizes, including extrachromosomal DNA when present, ranged from 9.04 Mb to 12.27 Mb, positioning the latter among the largest bacterial sequenced genomes (8) (Table 1). Although public annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (9), we performed an annotation step using the RAST tool kit (RASTik) (10, 11) and antiSMASH (12, 13) to predict the specialized metabolism repertoire. This showed that the sample had a large potential for synthesis, with 157 gene clusters. The strain RLA2-12 belonged...
to a previously isolated cluster of strains (identical 16S rRNA gene sequences) (8, 14) that was closely related to *Streptomyces olivochromogenes* DSM 40451 (99.9% identity in 16S rRNA gene sequences) (15). The strains RL81-33, RPA4-2, and S1D4-11 and to *Streptomyces mirabilis* CSSP107 exhibited identities of 99.9% to 99.2% in their 16S rRNA genes, indicating that they might belong to the same species or constitute sister species. While these strains belong to the large *Streptomyces* clade II described by McDonald and Currie (16), strain RPA4-5 was closely related to *Streptomyces platensis* ATCC 23948 (identical 16S rRNA genes), which belongs to the third, unstructured species coexist in the same microhabitat.

**Data availability.** Genome sequences and sequence reads (Illumina and Nanopore technologies) are available at GenBank and at the NCBI Sequence Read Archive (SRA) under the accession numbers listed in Table 1.

### ACKNOWLEDGMENTS

This work was funded by the French National Research Agency (ANR LABEX ARBRE grant ANR-11-LABX-0002-01), by the French National Institute for Agricultural Research (INRA), and by Région Lorraine (now called Région Grand Est).

### REFERENCES


### TABLE 1 Genome features, sequencing statistics, and accession numbers for the five *Streptomyces* isolates

<table>
<thead>
<tr>
<th>Strain (replicon)</th>
<th>Data for Illumina sequencing</th>
<th>Data for Oxford Nanopore sequencing</th>
<th>Replicon size (bp)</th>
<th>Genome size (bp)</th>
<th>Total no. of CDSs</th>
<th>GC content (%)</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>RL2-12</td>
<td>No. of reads (coverage [X])</td>
<td>SRA accession no.</td>
<td>No. of reads (coverage [X])</td>
<td>SRA accession no.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(pRLA2-12.1C)</td>
<td>21,079,784 (83) SRX8039999</td>
<td>269,118 (175) SRX8039129</td>
<td>10,825,588</td>
<td>10,892,946</td>
<td>10,116</td>
<td>70.3</td>
<td>JABAGQ0000000000</td>
</tr>
<tr>
<td>RL81-33</td>
<td>14,156,938 (50) SRX8040000</td>
<td>477,335 (222) SRX8039130</td>
<td>12,127,650</td>
<td>12,127,650</td>
<td>11,381</td>
<td>70.0</td>
<td>CPO50974</td>
</tr>
<tr>
<td>RPA4-2</td>
<td>19,808,024 (86) SRX8040001</td>
<td>231,311 (165) SRX8039131</td>
<td>9,856,149</td>
<td>9,856,149</td>
<td>9,287</td>
<td>70.9</td>
<td>CPO50975</td>
</tr>
<tr>
<td>RPA4-5</td>
<td>19,657,766 (93) SRX8040002</td>
<td>97,134 (73) SRX8039132</td>
<td>9,047,156</td>
<td>9,047,156</td>
<td>9,260</td>
<td>70.9</td>
<td>CPO50976</td>
</tr>
<tr>
<td>S1D4-11</td>
<td>15,509,220 (53) SRX8040003</td>
<td>180,498 (107) SRX8039133</td>
<td>12,276,515</td>
<td>12,276,515</td>
<td>12,065</td>
<td>69.9</td>
<td>CPO501010</td>
</tr>
</tbody>
</table>

* A Illumina libraries were generated using the universal TruSeq LT adapter 5′-AATGATACGGCGACACGTCTGAACTCCAGTCAC(index)ATCTCGTATGCCGTCTTCTGCTTG-3′ and the Nanopore adaptor sequences are not available and were automatically trimmed during base calling.

* B As determined through RAST automatic annotation. CDSs, coding sequences.

* C Indicates the circular configuration of the replicon, as predicted by the assembling.

