PATHWAYS IDENTIFICATION IN THE GRAM-POSITIVE BACTERIUM ANEURINIBACILLUS SPECIES H1

Jana Musilová
Doctoral Degree Programme (2), FEEC BUT
E-mail: musilovajana@feec.vutbr.cz

Supervised by: Karel Sedlář
E-mail: sedlar@feec.vutbr.cz

Abstract: Here, we present the first insight into the Gram-positive, promising polyhydroxyalkanoates producer Aneurinibacillus species H1. Both static and dynamic properties are described in this paper with the aim on identification of pathways occurring in the organism. The genome consists of a circular, 3,663,644 bp long chromosome and contains 4,654 protein-coding sequences and 129 RNAs in total. The GC content is 44.8%. Functional properties identification showed that CDS divide into 26 categories with the most prevalent Amino Acids and Derivatives group. Pathways inference revealed 201 pathways. The most represented group is metabolic pathways, which is further divided into 12 groups.

Keywords: PHA, genome assembly, pathway, protein-coding sequences

1 INTRODUCTION

With the increasing pollution of the planet by waste, such as plastics as one of the biggest sources of waste, new ways must be found to avoid it. The pollution may be significantly reduced by replacing these harmful materials with environmentally friendly ones. Polyhydroxyalkanoates (PHA) are polyesters, also labelled as biodegradable plastics, produced by microorganisms such as Aneurinibacillus species H1 [1]. This thermophilic, Gram-positive bacterium can synthetize PHA as intracellular granules to store unused energy.

Although Aneurinibacillus sp. H1 is a promising bacterium in a way of PHA overproduction, almost nothing is known about its genomic and phenotypic properties. The genome assembly data is not available; functional properties such as information about metabolic pathways is missing as well. Therefore, we present the first insight into the genome as well as its static and functional properties.

2 MATERIALS METHODS

2.1 GENOME ASSEMBLY AND ANNOTATION

The genome assembly was based on the sequencing data. The DNA sequences were obtained using the Oxford Nanopore technology, especially the MinION platform [2]. At first, reads were base-called using the Guppy basecaller v3.4.4 [3] and quality of the reads was checked by pycoQC v2.5.2 [4]. In the next step, the long-reads assembler Canu [5] was used to assemble individual reads. Finally, the sequence was polished using two tools. Firstly, the Racon [6] was used to create the initially polished assembly, and subsequently, the medaka [7] tool generated the final polished genome. The PAF files necessary for the polishing step were generated using minimap2 [8].

The assembled sequence was stored into a fasta file. Subsequently, the file and the RAST Server [9] were used for the genome annotation, such as identification of protein-coding sequences (CDS), RNAs, as well as genome length or GC content.
2.2 FUNCTIONAL ANNOTATION AND PATHWAYS IDENTIFICATION

Functional categorization of the genome features was done using the BlastKOALA [10] as well as the KO entries assignment. For visualization of the results, the SEED Viewer [11] was used. In a way to identify metabolic pathways in the bacterium, the KEGG Metabolic Analysis was done using the SEED Viewer and the KEGG Mapper [12].

3 RESULTS AND DISCUSSION

3.1 GENOME OVERVIEW

Genome of the Aneurinibacillus species H1 was reconstructed from 372,378 reads containing 2.7 billion bases in total. A median read length was 3,320 bases and a median Phred score corresponds to the value Q ≈ 9.6.

The whole assembly process resulted into the assembly consisting of one circular chromosome with a coverage exceeding 731×. The genome size is slightly above the average of gram-positive bacteria [13] with 3,663,644 bases but it is of the order of magnitude corresponding. The GC content with a value 44.8 corresponds to the gram-positive bacteria average. In the genome, a total number of 129 RNAs was discovered.

The number of CDS is 4,654 of which 1,977 functional parts were identified as identical to the model organism Bacillus subtilis subsp. 168. This information will help in next research of the Aneurinibacillus sp. H1, as almost half of processes can be studied on the well-described bacterium instead of conducting new laboratory experiments. Genomic properties are shown in the Figure 1.

![Figure 1: A chromosomal map of the Aneurinibacillus sp. H1 genome. The first (red) and second (blue) outermost circles represent CDSs on the forward and backward strands, respectively. The third circle (grey) represents pseudogenes. The inner circles represent the GC content and GC skew.](image-url)
3.2 FUNCTIONAL PROPERTIES AND IDENTIFIED PATHWAYS

Protein-coding sequences were classified into 26 functional systems. In total, 1,006 CDS were assigned to particular categories. The most prevalent group with 21% of assigned CDS is the “Amino Acids and Derivatives” category, followed by the groups “Carbohydrates” with 12%, “Protein Metabolism” corresponding to 11%, and “Cofactors, Vitamins, Prosthetic Groups, Pigments” group with 10% of all assigned protein-coding sequences. The Figure 2 displays all discovered groups together with the percentage.

Figure 2: A pie chart of the Aneurinibacillus sp. H1 functional properties.

Aneurinibacillus species H1 was associated to 201 pathways according to the KEGG encyclopedia. The most pathways belong to metabolism. In the metabolism category, 12 groups were discovered, each of them is divided into several subgroups. Individual groups together with a subgroup example (along with the KO entry and the number of associated processes) are shown in the Table 1. Map of bacterium’s metabolic pathways is displayed in the Figure 3.

Besides metabolic pathways, bacterial properties were also assigned to the pathways “Genetic Information Processing”; “Environmental Information Processing”; “Cellular Processes”; or “Organismal Systems”.

Table 1: Metabolic pathways associated with the Aneurinibacillus sp. H1 bacterium according to the KEGG encyclopedia.

<table>
<thead>
<tr>
<th>Global and overview maps</th>
<th>Metabolism of other amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>01100 Metabolic pathways (566)</td>
<td>00410 beta-Alanine metabolism (8)</td>
</tr>
<tr>
<td>01110 Biosynthesis of sec. mets. (243)</td>
<td>00430 Taurine and hypotaurine metabolism (5)</td>
</tr>
<tr>
<td>Carbohydrate metabolism</td>
<td>Glycan biosynthesis and metabolism</td>
</tr>
<tr>
<td>00010 Glycolysis / Gluconeogenesis (25)</td>
<td>00540 Lipopolysaccharide biosynthesis (1)</td>
</tr>
<tr>
<td>00020 Citrate cycle (TCA cycle) (21)</td>
<td>00541 O-Antigen and sugar biosynthesis (13)</td>
</tr>
<tr>
<td>Energy metabolism</td>
<td>Metabolism of cofactors and vitamins</td>
</tr>
<tr>
<td>00190 Oxidative phosphorylation (41)</td>
<td>00730 Thiamine metabolism (14)</td>
</tr>
<tr>
<td>00195 Photosynthesis (8)</td>
<td>00740 Riboflavin metabolism (7)</td>
</tr>
<tr>
<td>Lipid metabolism</td>
<td>Metabolism of terpenoids and polyketides</td>
</tr>
<tr>
<td>00061 Fatty acid biosynthesis (10)</td>
<td>00900 Terpenoid backbone biosynthesis (12)</td>
</tr>
<tr>
<td>00071 Fatty acid degradation (9)</td>
<td>00981 Insect hormone biosynthesis (1)</td>
</tr>
</tbody>
</table>
**CONCLUSIONS**

In this study, we revealed basic genomic and phenotypic properties of the polyhydroxyalkanoates producing bacterium *Aneurinibacillus* species H1 with the aim to discover pathways occurring in the bacterium.

In the first part, we described static bacterium’s properties. As the genome was not published yet, we assembled the complete genome in the first step, for which we used Oxford Nanopore, MinION sequencing data. The assembled genome consists of one circular chromosome with the size of 3,663,644 bases. The genome contains 4,654 protein-coding sequences in total, 129 RNAs and the GC content is 44.8. Comparing to the average of Gram-positive bacteria, the size is slightly above an average 2.53 Mbp, but GC content on average 49.4, which indicates that the genome has been assembled correctly. Nevertheless, average sequencing quality of Nanopore sequencing is around 90 % and polishing the sequence with Illumina sequencing data is needed before the genome can be reliably compared to other strains.

Description of functional properties of the *Aneurinibacillus* sp. H1 bacterium was the second part of this study. We classified in total 1,066 protein-coding sequences into 26 groups with most prevalent group Amino Acids and Derivatives. In the last step, we identified pathways occurring in the bacterium according to the KEGG encyclopedia. The pathways inference revealed 201 pathways with the most represented metabolic pathways divided into 12 groups.

---

**Figure 3:** A map of all metabolic pathways discovered in the *Aneurinibacillus* sp. H1.
ACKNOWLEDGEMENT

This work has been supported by grant project FCH/FEKT-J-21-7331. Study of Jana Musilová is supported by „Brno Ph.D. Talent Scholarship Holder – Funded by the Brno City Municipality“.

REFERENCES


[7] GitHub - nanoporetech/medaka: Sequence correction provided by ONT Research [online, accessed 2021-03-06]. Available at: https://github.com/nanoporetech/medaka


