DATA NOTE

**Genome of *Serratia plymuthica* UBCF_13, Insight into diverse unique traits [version 1; peer review: awaiting peer review]**

Raudhatul Fatiah¹, Irfan Suliansyah²,³, Djong Hon Tjong³,⁴, Lily Syukriani², Roza Yunita², Robi Trivano², Nurefni Azizah², Jamsari Jamsari¹,³

¹Agricultural Science, Universitas Andalas, Padang, West Sumatera, 25163, Indonesia
²Agrotechnology, Agricultural Faculty, Universitas Andalas, Padang, West Sumatera, 25163, Indonesia
³Biotechnology Magister Program, Universitas Andalas, Padang, West Sumatera, 25163, Indonesia
⁴Biology, Mathematics and Life Sciences Faculty, Universitas Andalas, Padang, West Sumatera, 25163, Indonesia

---

**Open Peer Review**

**Reviewer Status** AWAITING PEER REVIEW

Any reports and responses or comments on the article can be found at the end of the article.

---

**Abstract**

**Background:** Whole genome sequencing is become an essential tool to explore potential of microorganism and evolutionary study. The *Serratia plymuthica* UBCF_13 is one of phylloplane associated plant bacteria showing antifungal activity. For that reason, its complete genome information is necessary to enhance its potential as biocontrol against plant pathogenic fungal. Here, we report the genome sequence of *Serratia plymuthica* UBCF_13 to understand the molecular mechanism regarding its biocontrol ability.

**Methods:** Continuous short reads were attained from Illumina sequencing runs and reads 150 bp were merged into a single dataset. Pan-genome based method was used to identify core-genome of *S. plymuthica* species and unique gene in UBCF_13.

**Results:** Assambled Illumina reads of *S. plymuthica* strain UBCF_13 genome was produced a 5.46 Mb circular genome sequence. It was found 3321 genes belong to the core-genome sheared by the 18 strains evaluated. The UBCF_13 genome harbor 485 unique genes, where 300 of them only can be found in this strain

**Conclusions:** The sequence of UBCF_13 genome sequence data will contribute for further exploration of the potential of *S. plymuthica* UBCF_13 as bacteria producing antibiotic.

**Keywords**

Pan-genomic, core-genome, unique gene, Serratia plymuthica, genome sequencing, comparative genomic
Introduction

*Serratia plymuthica* bacteria have been isolated from many environmental sources and are found associated with diverse plants. Many strains of this species have been reported to have the ability to inhibit the growth of plant-pathogenic fungi and stimulate plant growth. UBCF_13 is one strain of this species. It has the ability to inhibit *Colletotrichum gloeosporioides*, a species of post-harvest pathogenic fungi that causes anthracnose disease in various plants.

Here, we report the complete genome sequence of this bacterium, constructed using Illumina sequencing technology. Our dataset may be useful as a comparative genome for evolutionary and speciation studies, as well as for the analysis of protein-coding RNA, biosynthetic gene clusters and may also be useful for further study such as the regulation of gene expression in relation to the antifungal activity of this bacterium.

Methods

Genomic DNA isolation and sequencing

*S. plymuthica* strain UBCF_13 was isolated from phylloplane of *Brassica juncea* L. in 2012 from District of Solok, Province of West Sumatera, Indonesia. The bacterium was cultivated in Luria–Bertani (LB) broth at 27°C for 16 hours with 150 rpm. The genomic DNA was extracted using the method of Chen and Kuo (1993), followed by degrading residual RNA by RNase. Library preparation and sequencing was done by Novogen (Hong Kong). Sequencing was performed using Illumina NovaSeq 6000 (Illumina NovaSeq 6000 Sequencing System, RRID:SCR_016387).

Genome assembly and annotation

Continuous short reads of 150 bp were merged into a single dataset. The dataset was obtained by combining a map-based gene references and de novo assembly that was performed in Geneious software (Geneious, RRID:SCR_010519). The annotation in genome submission was carried out using NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAP)11. The annotated genome sequence of UBCF_13 has been deposited in the NCBI GenBank under accession number CP068771.

Comparative genomics of *Serratia plymuthica* strains

Comparative genomics analysis was carried out using genome sequences of UBCF_13 from this research and 17 whole sequenced genomes of other *Serratia plymuthica* strains retrieved from NCBI's GenBank. The genomes were reannotated using the Prokka software tool (Prokka, Galaxy Version 1.14.6+galaxy0) (Prokka, RRID:SCR_014732) that is available from the NCBI. Identification of genes shared between the strains, and ‘presence-absence gene set’ was carried out using Roary; Galaxy Version 3.13.0+galaxy1, (Roary, RRID:SCR_018172) with a threshold similarity of 70%.

Genes that exist in all the strains are the core-genome. Phylogenetic trees were constructed using Maximum Likelihood based inference of large phylogenetic trees-RAXML, Galaxy Version 8.2.4+galaxy2 (RAXML, RRID:SCR_006086) based on mutialignment of concatenate core-genome. Phandango (Phandango, RRID:SCR_015243) was used to view the resulted output graphs.

Cluster of the orthologous groups of UBCF_13

The translated protein coding genes of UBCF_13 was used for identification of cluster of orthologous groups (COG). This was obtained from NCBI BLAST+psblast (Galaxy Version 2.10.1+galaxy0) and eggNOG Mapper (Galaxy Version 2.0.1+galaxy1) (eggNOG, RRID:SCR_002456). The result of COG identification was classified based on the categories in COG database NCBI.

Result and discussion

Comparative genomics of *Serratia plymuthica* strains

The whole genome sequencing reads of *Serratia plymuthica* UBCF_13 were assembled into a single circular 5.46 Mb chromosome with overall GC content of 56.2% (Table 1). *S. plymuthica* has a genome size in the range 5.40–5.70 Mb. The GC content percentage is 55.70–56.60. Based on genome reannotation by Prokka, it was found different number of CDS in each genome *S. plymuthica* (Table 1). All of the compared *S. plymuthica* genomes shared a highly conserved genomic architecture as inferred from synteny of protein coding orthologs.

Figure 1A shows the phylogenetic tree of 18 strains *S. plymuthica*. The phylogenetic tree shows that *S. plymuthica* UBCF_13 is in same cluster with strain AS9, PRI-2C, NCTC8015, and NCTC8900. The strain PRI-2C is reclassified and transferred to the species *S. inhibens*. The pan genome was performed together with other strains in order to obtain further insight into specific features in the UBCF_13. It was found 3315 belong to the core-genome shared by the 18 strains evaluated. The genome of the UBCF_13 harbors 488 unique genes, of which 300 genes are only contained by this strain. The presence-absence gene set was shown in file supplementary data 1.

The Cluster of Orthologous Groups of UBCF_13

Functional categories of the CDS in *S. plymuthica* UBCF_13 based on the Cluster of Orthologous Groups (COG) categories are shown in Table 2. The list of UBCF_13 COG and its functional classification based on COG database was shown in extended dataset 2.
Table 1. Genomic features comparison of *Serratia plymuthica* strains.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Accessions</th>
<th>Size (Mb)</th>
<th>GC%</th>
<th>CDSs</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>UBCF_13</td>
<td>CP068771.1</td>
<td>5.46</td>
<td>56.20</td>
<td>4920</td>
<td>Phylloplane of <em>Brassica juncea</em> L.</td>
</tr>
<tr>
<td>3Re4-18</td>
<td>CP012097.1</td>
<td>5.44</td>
<td>56.20</td>
<td>4937</td>
<td>Endorhiza of <em>Solanum tuberosum</em> L. cv. <em>Cilena</em></td>
</tr>
<tr>
<td>3Rp8</td>
<td>CP012096.1</td>
<td>5.55</td>
<td>56.10</td>
<td>5071</td>
<td><em>Brassica napus</em></td>
</tr>
<tr>
<td>4Rx13</td>
<td>CP006250.1</td>
<td>5.40</td>
<td>56.09</td>
<td>4831</td>
<td>The rhizosphere of <em>Brassica napus</em></td>
</tr>
<tr>
<td>AS9</td>
<td>CP002773.1</td>
<td>5.44</td>
<td>56.00</td>
<td>4970</td>
<td>Rapeseed roots grown in Sweden</td>
</tr>
<tr>
<td>C-1</td>
<td>CP053398.1</td>
<td>5.70</td>
<td>56.10</td>
<td>5316</td>
<td>Capsicum annuum</td>
</tr>
<tr>
<td>FDAARGOS_889</td>
<td>CP065699.1</td>
<td>5.44</td>
<td>56.20</td>
<td>4951</td>
<td></td>
</tr>
<tr>
<td>FDAARGOS_895</td>
<td>CP065688.1</td>
<td>5.35</td>
<td>55.90</td>
<td>4856</td>
<td></td>
</tr>
<tr>
<td>FDAARGOS_896</td>
<td>CP065747.1</td>
<td>5.53</td>
<td>55.80</td>
<td>5025</td>
<td></td>
</tr>
<tr>
<td>FDAARGOS_907</td>
<td>CP065673.1</td>
<td>5.47</td>
<td>55.90</td>
<td>4964</td>
<td></td>
</tr>
<tr>
<td>FDAARGOS_1138</td>
<td>CP068096.1</td>
<td>5.48</td>
<td>56.17</td>
<td>4960</td>
<td></td>
</tr>
<tr>
<td>NCTC12961</td>
<td>LS483469.1</td>
<td>5.35</td>
<td>55.90</td>
<td>5833</td>
<td></td>
</tr>
<tr>
<td>NCTC8015</td>
<td>LR134478.1</td>
<td>5.33</td>
<td>55.70</td>
<td>4901</td>
<td></td>
</tr>
<tr>
<td>NCTC8900</td>
<td>LR134151.1</td>
<td>5.32</td>
<td>55.80</td>
<td>5148</td>
<td></td>
</tr>
<tr>
<td>PRI-2c</td>
<td>CP015613.1</td>
<td>5.47</td>
<td>55.70</td>
<td>5009</td>
<td>Maize rhizosphere in the Netherlands</td>
</tr>
<tr>
<td>S13</td>
<td>CP006566.1</td>
<td>5.47</td>
<td>56.20</td>
<td>4959</td>
<td>Raw vegetable-processing line</td>
</tr>
<tr>
<td>RVH1</td>
<td>ARWD01000001.1</td>
<td>5.51</td>
<td>56.20</td>
<td>5023</td>
<td></td>
</tr>
<tr>
<td>V4</td>
<td>CP007439.1</td>
<td>5.51</td>
<td>56.20</td>
<td>5073</td>
<td>A biofilm of pasteurizer plates of milk processing machine</td>
</tr>
</tbody>
</table>

Figure 1. Maximum Likelihood tree and visualization of comparative genomic (pan genomic) between 18 strain *S. plymuthica*. (a) Phylogenetic of 18 strains *S. plymuthica* based on concatenate core-genome multialignment; and (b) visualization of presence (blue bar)-absence (white bar) gene in each of the strains.
Data availability
Data from *Serratia plymuthica* UBCF_13 is available at NCBI under Bio-Project PRJNA692765, including the complete genome with annotation at GenBank accession CP068771, and the read data in the Sequence Read Archive (SRA) database under the accession number SRR15012717.

Extended data
Dataset 1:
Dryad: Gene presence absence in *Serratia plymuthica* strains. [https://doi.org/10.5061/dryad.1zcrjdfsj](https://doi.org/10.5061/dryad.1zcrjdfsj)

This project contains the following extended data:
- Gene_presence_absence_in_Serratia_plymuthica_strains.csv (This data shows gene presence or absence across strain *Serratia plymuthica* species. It contains a list of coding protein gene (CDS) name and region where the gene exist in each genome (locus_tag). The region/locus where the CDS are existing has a specific tag for each genome, whereas the number after locus names represents the order of CDS in each genome. This data could be used to identify conserved and unique genes in *Serratia plymuthica*.)
- Readme_Gene_presence_absence_in_Serratia_plymuthica_strains.txt (This text file provides information about the above data)

Dataset 2:
Dryad: Classification of the UBCF_13 COG based on COG database in NCBI [https://doi.org/10.5061/dryad.sn02v6x4g](https://doi.org/10.5061/dryad.sn02v6x4g)
This project contains the following extended data:

- Classification_of_the_UBCF_13_COG_based_on_COG_database_in_NCBI.csv (This data contains information about the assignment of Clusters of Orthologous Gene (COG) for each coding protein gene in Serratia plymuthica UBCF_13 and their functional classification based on COG database in NCBI)

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

References

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com