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: Assembled 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
Corresponding author: Jamsari Jamsari (jamsari@agr.unand.ac.id)

Author roles: Fatiah R: Data Curation, Formal Analysis, Investigation, Methodology; Suliansyah I: Funding Acquisition, Supervision; Tjong DH: Supervision, Writing – Review & Editing; Syukriani L: Funding Acquisition, Project Administration; Yunita R: Funding Acquisition, Project Administration, Writing – Review & Editing; Trivano R: Funding Acquisition; Azizah N: Funding Acquisition; Jamsari J: Conceptualization, Funding Acquisition, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This study was fully funded by the General Directorate of Higher Education through PMDSU Research Grant Fiscal year 2020 contract number: T/10/UN.16.17/PT.01.03/AMD/ PMDSUPangan/2020 and 01/PL/SPK/PNP/ FAPERTA-Unand/2021

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Fatiah R et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Fatiah R, Suliansyah I, Tjong DH et al. Genome of Serratia plymuthica UBCF_13, Insight into diverse unique traits [version 1; peer review: awaiting peer review] F1000Research 2021, 10:826 https://doi.org/10.12688/f1000research.54402.1

First published: 18 Aug 2021, 10:826 https://doi.org/10.12688/f1000research.54402.1
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 gloeosporiodes, 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 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_006871).

Genome assembly and annotation
Continuous short reads of 150 bp were merged into a single dataset. The dataset was obtained by combination of 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). 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 multialignment 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+rpsblast (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 pangenome 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 function classification based on COG database was shown in extended dataset 2.
Table 1. Genomic features comparison of *Serratia plymuthica* strains.

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

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.1zcrjdfsj24

This project contains the following extended data;
- Gene_presence_abscence_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_abscence_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.sn02v6x4g23

| Function Class                                                                 | Number of genes |
|--------------------------------------------------------------------------------|-----------------|
| A RNA processing and modification                                              | 1               |
| C Energy production and conversion                                             | 259             |
| D Cell cycle control, cell division, chromosome partitioning                   | 61              |
| E Amino acid transport and metabolism                                          | 583             |
| F Nucleotide transport and metabolism                                          | 120             |
| G Carbohydrate transport and metabolism                                        | 476             |
| H Coenzyme transport and metabolism                                           | 255             |
| I Lipid transport and metabolism                                               | 215             |
| J Translation, ribosomal structure and biogenesis                              | 272             |
| K Transcription                                                                | 526             |
| L Replication, recombination and repair                                        | 153             |
| M Cell wall/membrane/envelope biogenesis                                       | 311             |
| N Cell motility                                                                | 96              |
| O Posttranslational modification, protein turnover, chaperones                  | 212             |
| P Inorganic ion transport and metabolism                                       | 367             |
| Q Secondary metabolites biosynthesis, transport and catabolism                 | 125             |
| R General function prediction only                                             | 476             |
| S Function unknown                                                             | 178             |
| T Signal transduction mechanisms                                               | 233             |
| U Intracellular trafficking, secretion, and vesicular transport                 | 82              |
| V Defense mechanisms                                                           | 131             |
| W Extracellular structures                                                     | 30              |
| X Mobilome: prophages, transposons                                             | 42              |
| Z Cytoskeleton                                                                | 2               |
| - No COG assignment                                                           | 518             |
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

1. Neupane S, Högbom N, Alström S, et al.: Complete genome sequence of the rapeseed plant-growth promoting Serratia plymuthica strain AS9. Stand Genomic Sci. 2012; 6(1): 54–62. PubMed Abstract | Publisher Full Text | Free Full Text
2. Adam E, Müller H, Erlacher A, et al.: Complete genome sequences of the Serratia plymuthica strains 3Rp8 and 3Re4-18, two rhizosphere bacteria with antagonistic activity towards fungal phytopathogens and plant growth promoting abilities. Stand Genomic Sci. 2016; 11(1): 61. PubMed Abstract | Publisher Full Text | Free Full Text
3. Cleto S, van der Auwera G, Almeida C, et al.: Genome Sequence of Serratia plymuthica V4. Genome Announc. 2014; 2(3): e00340-14. PubMed Abstract | Publisher Full Text | Free Full Text
4. Neupane S, Finlay RD, Kyprides NC, et al.: Complete genome sequence of the plant-associated Serratia plymuthica strain AS13. Stand Genomic Sci. 2012; 7(1): 22–30. PubMed Abstract | Publisher Full Text | Free Full Text
5. de Vleesschauwer D, Cherinin L, Höfte MM: Differential effectiveness of Serratia plymuthica IC1270-induced systemic resistance against hemibiotrophic and necrotrophic leaf pathogens in rice. BMC Plant Biol. 2009; 9(1): 9. PubMed Abstract | Publisher Full Text | Free Full Text
6. Müller H, Fürnkranz M, Grube M, et al.: Genome Sequence of Serratia plymuthica Strain 513, an Endophyte with Germination- and Plant-Growth-Promoting Activity from the Flower of Styrían Oil Pumpkin. Genome Announc. 2013; 1(4): e00594-13. PubMed Abstract | Publisher Full Text | Free Full Text
7. Kurze S, Bahl H, Dahl R, et al.: Biological Control of Fungal Strawberry Diseases by Serratia plymuthica HRO-C48. Plant Dis. 2001; 85(5): 529–534. PubMed Abstract | Publisher Full Text
8. Vleesschauwer D: Using Serratia plymuthica to control fungal pathogens of plants. CAB Rev Perspect Agric Vet Sci Nutr Nat Resour. 2007; 2(046): 1–9. PubMed Abstract | Publisher Full Text
9. Campos D, Cottet L, Castillo A: Antifungal activity of Serratia plymuthica CCGG2742 against a new wild isolate of the phytopathogenic fungus Alternaria tenuissima. Plant Dis. 2018; P05-05-17-0709-R.E. Reference Source
10. Aiyah SN, Sulastri S, Retmi R, et al.: Suppression of Colletotrichum gloeosporioides by Indigenous Phyllobacterium and its Compatibility with Rhizobacteria. Asian J Plant Pathol. 2017; 11(3): 139–147. Publisher Full Text
11. Chen WP, Kuo TT: A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. Nucleic Acids Res. 1993; 21(9): 2260. PubMed Abstract | Publisher Full Text | Free Full Text
12. Kearse M, Moir R, Wilson A, et al.: Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28(12): 1647–1649. PubMed Abstract | Publisher Full Text | Free Full Text
13. Li W, O’Neill KB, Haft DH, et al.: RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. Nucleic Acids Res. 2021; 49(01): D1020–D1028. PubMed Abstract | Publisher Full Text | Free Full Text
14. Cuccuru G, Orsini M, Pinna A, et al.: Orione, a web-based framework for NGS analysis in microbiology. Bioinformatics. 2014; 30(13): 1928–9. PubMed Abstract | Publisher Full Text | Free Full Text
15. Seemann T: Prokka: Rapid prokaryotic genome annotation. Bioinformatics. 2014; 30(14): 2068–9. PubMed Abstract | Publisher Full Text
16. Page AJ, Cummins CA, Hunt M, et al.: Roary: Rapid large-scale prokaryote pan genome analysis. bioRxiv. 2015; 019315. Publisher Full Text
17. Stamatakis A: RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30(9): 1312–1313. PubMed Abstract | Publisher Full Text | Free Full Text
18. Hadfield J, Croucher NJ, Goater RJ, et al.: Phandango: An interactive viewer for bacterial population genomics. Bioinformatics. 2018; 34(2): 292–293. PubMed Abstract | Publisher Full Text | Free Full Text
19. Cock PJA, Chilton JM, Grüning B, et al.: NCBI BLAST+ integrated into Galaxy. Gigascience. 2015; 4(1): 39. PubMed Abstract | Publisher Full Text | Free Full Text
20. Huerta-Cepas I, Forslund K, Coelho LP, et al.: Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol. 2017; 34(8): 2115–2122. PubMed Abstract | Publisher Full Text | Free Full Text
21. Galperin MY, Wolf YI, Makarova KS, et al.: COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res. 2012; 40(1): D129–D135. PubMed Abstract | Publisher Full Text | Free Full Text
22. Hennessy RC, Dichmann SI, Martens HJ, et al.: Complete genome sequence of the Serratia inhibens sp. nov., a new antifungal species isolated from potato (Solanum tuberosum). Int J Syst Evol Microbiol. 2020; 70(7): 4204–4211. PubMed Abstract | Publisher Full Text
23. Raudhatul F, Irfan S, Djong HT, et al.: Classification of the UBCF_13 COG based on COG database in NCBI, Dryad. Dataset. 2021. http://www.doi.org/10.5061/dryad.s2n02v6x4g. PubMed Abstract | Publisher Full Text
24. Raudhatul F, Irfan S, Djong HT, et al.: Gene presence absence in Serratia plymuthica strains, Dryad. Dataset. 2021. http://www.doi.org/10.5061/dryad.1zcrjdfsj
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