IOSR Journal of Biotechnology and Biochemistry (IOSR-JBB)
ISSN: 2455-264X, Volume 9, Issue 6 Series I (Nov. – Dec. 2023), PP 43-45
www.iosrjournals.org

Draft Genome Sequencing Of A Gastrodia Elata Pathogen Erwinia Billingiae LS-1

Xiangyang Li[1,2]
[1]*College of Environment and Life Sciences, Kaili University, 556011, Kaili, China*.
[2]*Bacterial Genome Data Mining & Bioinformatic Analysis Center, Kaili University, 556011, Kaili, China*.

ABSTRACT*:*
Most of the member in Erwinia are plant pathogens. We report here the draft genome sequence of strain LS-1, which was isolated from the decaying aerial stem of a Gastrodia elata plant. By using 16S rRNA gene, average nucleotide identity (ANI) and core-gene based phylogenetic tree analysis, strain LS-1 is assigned to Erwinia billingiae. E. billingiae LS-1's genome has a genome size of 5,123,268 bp, including 4,721 coding sequences, 86 RNA genes, and 174 pseudogenes. Only 35% genes can be assigned to RAST subsystem categories. Strain LS-1's genome harbored 89 potential genes implicated in bacterium-plant interactions that may aid its adaptation to plant environments.

KEYWORDS: *Gastrodia elata, Erwinia billingiae, Genomic sequence.*

Date of Submission: 29-11-2023 Date of Acceptance: 09-12-2023

I. INTRODUCTION

Recent *Gastrodia elata* is a perennial herbaceous plant in the Orchid family, and its bulbs have high medicinal and health value in China (Li et al., 2023). However, *G. elata* is affected by major diseases, such as stem rot (Li and Cheng, 2021). The aim of this study is to isolate and identity the strain LS-1 that causes stem rot by pure culture, 16S rRNA gene sequencing and sequence similar search; determining the whole genome sequence of Strain LS-1 to to further study the feature of strain LS-1 and uncover its pathogenic factors related to its pathogenic mechanism.

II. MATERIALS AND METHOD

We collected a diseased *G. elata* plant from an experimental base (107°53′38″E, 26°31′44.40″N) at Kaili University on May 05, 2018. A single clone was isolated by spreading the decaying aerial stem of the diseased plant onto Luria-Bertani (LB) plates and culturing it at 30°C for 18 h. The 16S rRNA gene was amplified by a colony PCR using primers 27 F (5′-AGAGTTTGATCMTGGCTCAG-3′) and 1492 R (5′-TACGGYTACCTTGTTACGACTT-3′) and subsequently sequenced with an Applied Biosystems (ABI) 3730xl DNA analyzer. The MiniBEST bacterial genomic DNA extraction kit (TaKaRa, Dalian, China) was used to extract the total DNA of strain LS-1 according to the manufacturer's instructions. Sequencing of the strain LS-1 genome was performed on an Illumina HiSeq 2500 platform using a paired-end 2 × 150-bp strategy. Raw reads were filtered by Trimmomatic (Bolger et al., 2014) and FastUniq (Xu et al., 2012). Clean reads were assembled using SOAPdenovo version 2.04 (Luo et al., 2012) with default parameters and a *k*-mer length of 77. We assessed the quality and accuracy of assembled genome using CheckM v1.1.3 (Parks et al., 2015) and QUAST 5.0.2. Genome was annotated via NCBI Prokaryotic Genome Annotation Pipeline (Tatusova et al., 2016)

III. RESULTS AND DISCUSSION

By performing a BLASTn analysis on the NCBI database, the 16S rRNA gene sequence has the highest possible similarity of 99.93% with that of *Erwinia* spp. Therefore, strain LS-1 is assigned to the genus *Erwinia*.

Members of the genus *Erwinia* are Gram-negative bacteria of the Enterobacteriaceae family, and lots of species are plant pathogens causing diseases in several economically important plants (Li et al., 2023). For instance, *E. amylovora* is the causal agent of fire blight, a devastating plant disease affecting a wide range of host species within Rosaceae and a major global threat to commercial apple and pear production (Pique et al., 2015).

In order to further study the feature of strain LS-1 and uncover its pathogenic factors related to its pathogenic mechanism. We determined its whole-genomic sequence. In total, we obtained 2,075,313,900 bp raw reads. After filter low-quelity sequences by removing reads containing adapters, reads containing poly-N (>10),

duplicated reads (caused by the PCR during library construction), and low-quality reads (more than 40% with $Q < 20$ bases) using Trimmomatic (Bolger et al., 2014) and FastUniq (Xu et al., 2012), 822,000,000 bp clean reads were remained for assembly analysis. These clean reads were assembled into 48 scaffolds (52 contigs) using SOAPdenovo version 2.04 (Luo et al., 2012) with default parameters and a *k*-mer length of 77. We assessed the quality and accuracy of assembled genome using CheckM v1.1.3 (Parks et al., 2015) and QUAST 5.0.2 (Gurevich et al., 2013). CheckM determines the estimated completeness of a genome and detects possible contamination based on lineage-specific sets of single-copy genes. Strain LS-1 have a completeness of > 90% and a low contamination (≤5% contamination) (Chen et al., 2021). QUAST is used to report assembly quality metrics, and the result indicated that the contig number is ≤ 500， and N50 ≥ 40 kb. Thus, the above-mentioned data showed a high assembly quality for LS-1's genomes. Summarily, this strain has a genome size of 5,123,268 bp with 160-fold coverage. Its average GC content is 54.81%. The N_{50} and N_{90} scaffold sizes are 460,386 bp and 106,275 bp, respectively.

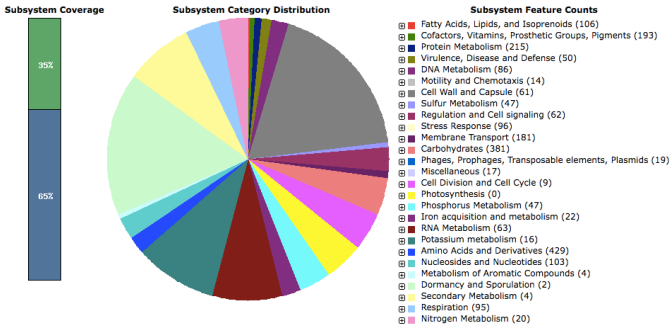

Figure No 1. Gene function analysis using RAST server. Color (green and blue) bar on left denoted genes can be or not assigned to subsystem categories, respectively.

We annotated the genome of strain LS-1 using the NCBI Prokaryotic Genome Annotation Pipeline (Tatusova et al., 2016). A total of 4807 genes were identified in the strain's genome, including 4,721 coding sequences, 86 RNA genes, and 174 pseudogenes. We analysis the gene function using RAST server, and the results showed that only 35% genes can be assigned to subsystem categories (Figure no 1), and more than half of genes for this strain have unknown function. It indicated that this species gets less attention and more studies should be given to this species. The genome of strain LS-1 harbored 89 potential virulence factors implicated in bacterium-plant interactions, such as flagella, lipopolysaccharide, the type V secretion system, type VI secretion system, and the type VII secretion system, and several effectors, like the hemolysin-coregulated protein (hcp) effectors of type VI secretion system (Liu et al., 2022).

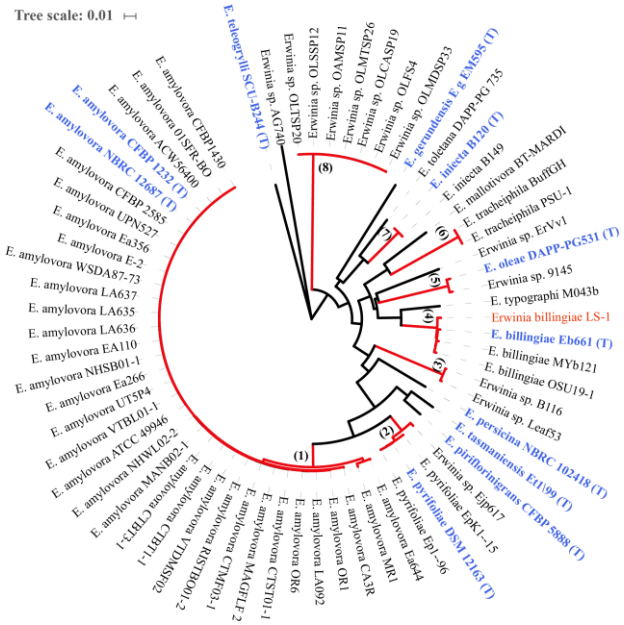

Figure No 2. Phylogenomic tree delineating the affiliations of 67 *Erwinia* strains. Core genes were used to build the NJ tree. Pairs of strains within each group has an ANIm >95%. Type strains were marked by blue bold font, all the branch boostrap > 50%.

As of Dec 05, 2023, there were genome sequences for 378 *Erwinia* strains, publicly available in the NCBI genome database. Based on average nucleotide identity (ANI) analysis using FastANI (Jain et al., 2018), Strain LS-1 has a maximum ANI value of 96.77% with strain Eb661, the type strain of *Erwinia billingiae* (GenBank accession number NC_014306.1). Thus, strain LS-1 could be assigned into *E. billingiae* according to a promising of 95% ANIm cutoff with the type strain E661. We infer the phylogeny of strain LS-1 within the genus *Erwinia*. A total of 67 genomes of *Erwinia* spp. were selected to obtain the core genes using OrthoFinder. The single-copy core genes that are shared by all genomes and present as only a single copy in each genome were used to infer the phylogenetic tree using MEGAX version 10.0.4 with the NJ algorithm (Kumar et al., 2018) (Figure no 2). Strains were divided into 18 subgroups which may represented 18 species base a 95% ANI cutoff, including 4 unknown species. Strains LS-1 and type strain E661 clustered tighter. It confirmed that Strains LS-1 belonged to *E. billingiae*. In addition, *Erwinia* sp. Ejp617 should be assigned into *E. pyrifoliae*.

This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number QXNN00000000. The version described in this paper is version QXNN01000000.

IV. CONCLUSION

To our best knowledge, *E. billingiae* LS-1 was the first reported pathogenic genome sequence isolated from *Gastrodia elata* (Yuan et al., 2018). The successful identification of the Strain LS-1 and determination of its reported genomes in this study would provide useful information for understanding the pathogenic mechanism, and control the occurrence of stem rot.

REFERENCES

[1]. Bolger, A.M., Lohse, M., And Usadel, B. (2014). Trimmomatic: A Flexible Trimmer For Illumina Sequence Data. Bioinformatics. 30, 2114-2120. Doi: 10.1093/Bioinformatics/Btu170

[2]. Chen, M.Y., Teng, W.K., Zhao, L., Hu, C.X., Zhou, Y.K., Han, B.P., Et Al. (2021). Comparative Genomics Reveals Insights Into Cyanobacterial Evolution And Habitat Adaptation. ISME J. 15, 211-227. Doi: 10.1038/S41396-020-00775-Z

[3]. Gurevich, A., Saveliev, V., Vyahhi, N., And Tesler, G. (2013). QUAST: Quality Assessment Tool For Genome Assemblies. Bioinformatics. 29, 1072-1075. Doi: 10.1093/Bioinformatics/Btt086

[4]. Jain, C., Rodriguez, R.L., Phillippy, A.M., Konstantinidis, K.T., And Aluru, S. (2018). High Throughput ANI Analysis Of 90K Prokaryotic Genomes Reveals Clear Species Boundaries. Nat Commun. 9, 5114. Doi: 10.1038/S41467-018-07641-9

[5]. Kumar, S., Stecher, G., Li, M., Knyaz, C., And Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis Across Computing Platforms. Mol Biol Evol. 35, 1547-1549. Doi: 10.1093/Molbev/Msy096

[6]. Li, J., And Cheng, L. (2021). Fusarium Solani Causing Root Rot Disease On Gastrodia Elata In Shaxi, China. Plant Dis. Doi: 10.1094/PDIS-04-21-0753-PDN

[7]. Li, J., He, K., Zhang, Q., Wu, X., Li, Z., Pan, X., Et Al. (2023). Draft Genome And Biological Characteristics Of Fusarium Solani And Fusarium Oxysporum Causing Black Rot In Gastrodia Elata. Int J Mol Sci. 24. Doi: 10.3390/Ijms24054545

[8]. Liu, B., Zheng, D., Zhou, S., Chen, L., And Yang, J. (2022). VFDB 2022: A General Classification Scheme For Bacterial Virulence Factors. Nucleic Acids Res. 50, D912-D917. Doi: 10.1093/Nar/Gkab1107

[9]. Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., Et Al. (2012). Soapdenovo2: An Empirically Improved Memory-Efficient Short-Read De Novo Assembler. Gigascience. 1, 18. Doi: 10.1186/2047-217X-1-18

[10]. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., And Tyson, G.W. (2015). Checkm: Assessing The Quality Of Microbial Genomes Recovered From Isolates, Single Cells, And Metagenomes. Genome Res. 25, 1043-1055. Doi: 10.1101/Gr.186072.114

[11]. Pique, N., Minana-Galbis, D., Merino, S., And Tomas, J.M. (2015). Virulence Factors Of Erwinia Amylovora: A Review. Int J Mol Sci. 16, 12836-12854. Doi: 10.3390/Ijms160612836

[12]. Tatusova, T., Dicuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Et Al. (2016). NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res. 44, 6614-6624. Doi: 10.1093/Nar/Gkw569

[13]. Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Et Al. (2012). Fastuniq: A Fast De Novo Duplicates Removal Tool For Paired Short Reads. Plos One. 7, E52249. Doi: 10.1371/Journal.Pone.0052249

[14]. Yuan, Y., Jin, X., Liu, J., Zhao, X., Zhou, J., Wang, X., Et Al. (2018). The Gastrodia Elata Genome Provides Insights Into Plant Adaptation To Heterotrophy. Nat Commun. 9, 1615. Doi: 10.1038/S41467-018-03423-5