

Draft genome sequence of a prodigiosin-hyperproducing *Serratia marcescens* strain isolated from Cairo, Egypt

Nora M. Elkenawy ^{1,*}, Noha H. Youssef,² Ramy K. Aziz,^{3,4,5} Magdy A. Amin,^{3,†} and Aymen S. Yassin^{3,4,*}

¹Drug Radiation Research Department, National Center for Radiation Research & Technology, 11787 Cairo, Egypt,

²Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078-5061, USA,

³Department of Microbiology and Immunology, Faculty of Pharmacy, Cairo University, 11562 Cairo, Egypt,

⁴The Center for Genome and Microbiome Research, Cairo University, 11562 Cairo, Egypt, and

⁵Microbiology and Immunology Research Program, Children's Cancer Hospital Egypt 57357, 11617 Cairo, Egypt

*Corresponding authors: Email: Nelkenawy@gmail.com (N.M.E.); aymen.yassin@pharma.cu.edu.eg (A.S.Y.)

†Professor Magdy A. Amin was deceased on November 30, 2019; all authors agree that he contributed to that manuscript with full authorship rights.

Abstract

Serratia marcescens is a Gram-negative bacterium with both environmental and host-associated strains. Pigmentation is reportedly inversely correlated with infection frequency, and prodigiosin is one of *Serratia* pigments that has medical and industrial applications. Here, we report the draft genome sequence of prodigiosin-hyperproducing *Serratia marcescens* strain N2, isolated from Cairo, Egypt. The sequence is assembled into 142 contigs, with a combined size of 5,570,793 bp. The assembled genome carries typical *S. marcescens* genes, with potential prodigiosin-biosynthesizing genes detected.

Keywords: *Serratia marcescens*; prodigiosin; bioinformatics; genome annotation; genome assembly

Introduction

Serratia marcescens is a Gram-negative rod-shaped bacterium, belonging to family *Enterobacteriaceae*. It has been isolated from various environmental and nosocomial sources. In the last decades, *S. marcescens* has been recognized as a significant opportunistic human pathogen, as it was found responsible for a variety of symptoms in hospitalized patients, including septicemia, meningitis, and infections of the urinary tract (Abreo and Altier 2019). Although some *S. marcescens* strains are associated with hospital infections, pigmented *S. marcescens* strains were shown to cause infections in a much lower frequency than nonpigmented strains. This observation implies that the infection risk is minimal during the mass production of pigment (Roy et al. 2014).

Prodigiosin is a red pigment produced as a secondary metabolite by *S. marcescens*, characterized by a distinctive tripyrrole structure, responsible for its reported multiple pharmacological effects as anti-cancer, anti-microbial, anti-oxidant, and immunosuppressant, as well as its unique application as a natural dye for olefins and textiles (Lin et al. 2019). Factors such as temperature, pH, dissolved oxygen levels, light and medium composition influence the production of prodigiosin (Aruldass et al. 2014). Here, we report the genome sequencing of *S. marcescens* strain N2, which was isolated from a local hospital in Cairo, Egypt. This strain is capable of producing 870 unit/cell of prodigiosin after 6 days of incubation (Elkenawy et al. 2017).

Materials and methods

DNA extraction

Genomic DNA was extracted by the PrepMan[®] ultra sample preparation reagent (Applied Biosystems, USA). Phylogenetic identity was confirmed by 16S rRNA gene sequencing, performed by the MicroSeq[®] 500 identification protocol (Applied Biosystems, USA), as previously described (Fontana et al. 2005).

Sequencing

The genome of *S. marcescens* strain N2 was sequenced on an Illumina MiSeq platform at Novogene (Beijing, China), following the standard Illumina protocols. A 300 × 2 paired-end chemistry was used, with an average library insert size of 700 bp. The Nextera XT DNA library prep kit (Illumina, San Diego, CA, USA) was used for the preparation of sequencing libraries from extracted DNA, as per the manufacturer's instructions.

Preprocessing and assembly

Reads were quality-filtered by Trimmomatic, version 0.36 (Bolger et al. 2014), with adaptor clipping option, a sliding window size of 4 and a sliding window minimum quality of 15. High-quality reads were assembled by three different tools: (1) IDBA assembler (v1.1.3) with minimum contig length of 2000 bp (Peng et al. 2012); (2) SPAdes assembler (v3.12.0), which is part of the Pathosystems Resource Integration Center (PATRIC) services (Antonopoulos

et al. 2019; Davis et al. 2020), with minimum contig length of 500 bp; and (3) Unicycler assembler (v0.4.8), within PATRIC services as well, with minimum contig length of 500 bp (Table 1).

Genome annotation

The assembled genome was annotated by PATRIC (version 3.6.9) annotation service (Antonopoulos et al. 2019; Davis et al. 2020), which uses the RASTtk algorithm (Aziz et al. 2008; Brettin et al. 2015).

Results and discussion

The sequencing yielded 13.3 Gb of paired-end reads, with an average read length of 150 bp.

All three genome assemblers produced high-quality contigs, ranging in number from 142 to 716, and an N₅₀ from 41,216 to 238,944 (Table 1). Based on the comparison between the three assemblers, Unicycler assembly was further considered for downstream annotation and analysis steps. The 142 contigs had a combined size 5,570,793 bp with an N₅₀ of 238,944 bp, an L₅₀ of 8, and a G + C content of 59.05%.

Genome annotation resulted in 5615 coding sequences (CDS), 86 tRNAs and 6 rRNA-encoding genes. The CDS included 1200 hypothetical proteins and 4415 proteins with functional assignments, of which 2064 proteins were assigned EC number, 1068 were assigned gene ontology (GO) terms, and 1471 proteins were assigned to known pathways. PATRIC assigned the proteins to 5024 genus-specific families (PLfams) and 5170 cross-genus families (PGfams). The assembled genome was also analyzed by NCBI prokaryotic genome annotation pipeline service (Tatusova et al. 2016), which is usually more conservative in gene calling and functional assignments. NCBI analysis predicted 5393 CDS, a number that is closest to the PATRIC genus-specific families. BLAST identified proteins that form clusters for prodigiosin production, including PigA, PigB, PigC, PigD, PigE, PigF, PigH, and PigI (Williamson et al. 2006). Similarly, transcriptional regulators for pigment production were also identified, e.g., PigS (Positive Transcription regulator), PigT, HexS (Direct downregulator), CRP (Indirect downregulator), PigP (Transcription activator), PigX (Transcription repressor), and Rap (Activation regulator of antibiotic and pigment) (Gristwood et al. 2011).

Finally, the use of PATRIC similar genome finder (Antonopoulos et al. 2019), which uses the Mash/MinHash algorithm (Ondov et al. 2016) predicted the closest genome neighbor to strain N2 to be *S. marcescens* strain ATCC 274 (NCBI accession number: AP021873, RefSeq ID: NZ_AP021873). Further analyses of the genome sequence, coupled with high-throughput techniques such as random transposon mutagenesis (*in vitro*) and genome-wide metabolic reconstruction (*in silico*), will enable the

identification of genes putatively involved in prodigiosin promotion and provide more insight into the biosynthesis and regulation of structurally diverse secondary metabolites.

Data availability

Raw reads are available under SRA accession SRX10696011. The whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under Bioproject ID PRJNA525074, Biosample ID SAMN11041520, and WGS accession SPSG00000000. The version described in this paper is version SPSG02000000. The annotated genome the PATRIC database is deposited under genome number 615.1488.

Acknowledgments

The center of Genome and Microbiome research was founded by a project from the Egyptian Academy for Scientific Research and Technology (ASRT) JESOR project #3046 in 2018 to A.S.Y., R.K.A., and N.H.Y. The project initiated the collaboration between Cairo University and Oklahoma State University, but the academy had no involvement in the publication process.

Conflicts of interest

The authors declare that there is no conflict of interest.

Literature cited

- Abreo E, Altier N. 2019. Pangenome of *Serratia marcescens* strains from nosocomial and environmental origins reveals different populations and the links between them. *Sci Rep*. 9:46.
- Antonopoulos DA, Assaf R, Aziz RK, Brettin T, Bun C, et al. 2019. PATRIC as a unique resource for studying antimicrobial resistance. *Brief Bioinform*. 20:1094–1102.
- Aruldass CA, Venil CK, Zakaria ZA, Ahmad WA. 2014. Brown sugar as a low-cost medium for the production of prodigiosin by locally isolated *Serratia marcescens* UTM1. *Int Biodeterior Biodegradation*. 95:19–24.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, et al. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics*. 9:75.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114–2120.
- Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, et al. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep*. 5.
- Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, et al. 2020. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res*. 48:D606–D612.
- Elkenawy NM, Yassin AS, Elhifnawy HN, Amin MA. 2017. Optimization of prodigiosin production by *Serratia marcescens* using crude glycerol and enhancing production using gamma radiation. *Biotechnol Rep (Amst)*. 14:47–53.
- Fontana C, Favaro M, Pelliccioni M, Pistoia ES, Favalli C. 2005. Use of the MicroSeq 500 16S rRNA gene-based sequencing for identification of bacterial isolates that commercial automated systems failed to identify correctly. *J Clin Microbiol*. 43:615–619.
- Gristwood T, McNeil MB, Clulow JS, Salmond GP, Fineran PC. 2011. PigS and PigP regulate prodigiosin biosynthesis in *Serratia* via differential control of divergent operons, which include predicted

Table 1 The reads assembly using three different tools

	IDBA	SPAdes	Unicycler
Total N contigs	271 ^a	716	142
N contigs >1,000	271	195	112
N contigs >10,000	127	55	44
Largest contig	201,798	405,373	532,301
N ₅₀	41,216	171,039	238,944
GC%	56.26	58.32	59.05
Total length	5,360,385	5,796,868	5,570,793

^a IDBA used a threshold of 2000 bp. as a minimum for assembled contigs to be included, while SPAdes and Unicycler used a threshold of 500 bp. This made the total number of contigs equal to the number of contigs >1,000 in IDBA assembly.

- transporters of sulfur-containing molecules. *J Bacteriol.* 193: 1076–1085.
- Lin C, Jia X, Fang Y, Chen L, Zhang H, et al. 2019. Enhanced production of prodigiosin by *Serratia marcescens* FZSF02 in the form of pigment pellets. *Electronic J Biotechnol.* 40:58–64.
- Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, et al. 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17:132.
- Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics.* 28:1420–1428.
- Roy P, Ahmed NH, Grover RK. 2014. Non-pigmented strain of *Serratia marcescens*: an unusual pathogen causing pulmonary infection in a patient with malignancy. *J Clin Diagn Res.* 8:DD05–DD06.
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, et al. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44:6614–6624.
- Williamson NR, Fineran PC, Leeper FJ, Salmond GP. 2006. The biosynthesis and regulation of bacterial prodiginines. *Nat Rev Microbiol.* 4:887–899.

Communicating editor: R. Kulathinal