



First draft genome sequencing of the *Salicola* genus

Lobna Daoud^{1,2,*}, Adel Hadj Brahim¹, Houda Hmani¹, Mouna Jlidi¹, Manel Ben Ali^{1,2}, Samir Bejar¹, Naser Aliye Feto³ and Mamdouh Ben Ali^{1,2}

¹ Laboratory of Microbial Biotechnology and Engineering Enzymes (LBMIE), Center of Biotechnology of Sfax (CBS), University of Sfax, Road of Sidi Mansour km 6, PO Box 1177 Sfax 3018, Tunisia; E-Mails: lobna.daoudm@gmail.com; adelhadjibrahim@gmail.com; houda_enis@yahoo.fr; asmahen.akremi@gmail.com; jlidimanno@yahoo.fr; manel.benali@gmail.com; samir.bejar@cbs.rnrt.tn; mamdouh.benali@cbs.rnrt.tn.

² Astrum Biotech, Business incubator, Center of Biotechnology of Sfax (CBS), University of Sfax, Road of Sidi Mansour km 6, PO Box 1177 Sfax 3018, Tunisia; E-Mails: lobna.daoudm@gmail.com; manel.benali@gmail.com; mamdouh.benali@cbs.rnrt.tn.

³ OMICS Research Group & Facility: Department of Biotechnology, Faculty of Applied & Computer Sciences, Vanderbijlpark Campus, Private Bag, X021 - Vanderbijlpark - 1911 - Andries Potgieter Blvd - South Africa; E-Mail: naserf@vut.a.za.

* Correspondence addressed to Lobna Daoud; E-Mail: lobna.daoudm@gmail.com; Tel.: +216 27 658 016; Fax: +216 74 875 818.

Received: / Accepted: / Published:

Abstract:

Salicola sp. strain SBJ9 is an extremely halophilic bacterium newly isolated from a hypersaline lake in Sfax (Tunisia). It is an aerobic Gram negative bacterium, bacilli shaped and motile. The strain SBJ9 grows optimally at 150 g/l (2.57 M) NaCl, 35 °C and pH 7 and it is able to hydrolyze proteins and lipids at high salinities. Actually, *Salicola* genus contains only two species (*S. marasensis* and *S. salis*). This research reported the first draft genome of a bacterium belonging to *Salicola* and the different obtained and annotated sequences and genes. *Salicola* sp. strain SBJ9 genomic sequence contains 4.643.441 bp with a G+C content of 60.52%. The annotation monitored by the RAST server reveals 65 RNA sequences and 3995 coding genes including 255 and 182 sub-systems of protein metabolism and lipids, respectively. The further analysis of SBJ9 genomic sequence will clarify the ideas regarding the adaptation of the *Salicola* strains and their proteins to the extremely high salt concentrations and will permit the development of new biocatalyzers for industrial applications.

Keywords: *Salicola*; extremely halophilic; next generation sequencing; draft genome; annotation.

1. Introduction

The industrial demand for new enzymes stable at harsh conditions is more and more increasing. In this context, we have focused on halophilic

microorganisms (halophiles) which are known for the production of enzymes with high activity

and stability at wide ranges of salinity and low water content media [1].

Salicola sp. SBJ9 is a newly isolated bacterium from a hypersaline lake in Tunisia (sebkha Bou Djemal in Sfax). In fact, *Salicola* genus is reported for the first time in 2006 and contains very few members with only two species: *S. marasensis* [2] and *S. salis* [3]. Strain SBJ9 showed an interesting protease activity which can be a good candidate for many applications suffering from the instability of their conventional enzymes. To isolate its gene, the genome of strain SBJ9 was sequenced and the proteases genes, along with other industrial enzymes, were isolated. This is the first genome sequencing of the *Salicola* genus and it was monitored by the newest and the best technology: the Illumina Next Generation Sequencing (NGS) technology.

The Illumina NGS workflows contain four basic steps [4]. First, a sequencing library is prepared by a random fragmentation of the DNA sample and the ligation of 5' and 3' adapters.

2. Results and Discussion

Salicola sp. SBJ9 is an extremely halophilic bacterium, bacilli shaped, Gram negative, aerobic and motile. It can grow from 100 to 350 g/l NaCl, pH 6 to 9 and 20 to 50 °C with an optimal growth at 150 g/l NaCl, pH 7 and 35 °C. The phylogenetic tree of the isolate showed that it is more related to *Salicola marasensis* 7mb 1 and *Salicola marasensis* 7 Sm6 rather than *Salicola salis* B 2 (**Figure 1**). Then, it is probably a member of *S. marasensis* species.

Second, the library is loaded into a flow cell where fragments are captured on a lawn of surface-bound oligos complementary to the library adapters. Each fragment is then amplified into distinct and clonal clusters through bridge amplification. This consists on the cluster generation step. Then, when cluster generation is complete, the sequencing step is ready. For that, Illumina uses a reversible terminator-based method that detects single bases as they are incorporated into DNA template strands. The result is highly accurate base-by-base sequencing that virtually eliminates sequence context-specific errors, even within repetitive sequence regions and homopolymers, feature that distinguish it from the other technologies. The final step is the data analysis. The newly identified sequence reads are aligned to a reference genome. Following alignment, many variations of analysis are possible, such as single nucleotide polymorphism (SNP) or insertion-deletion (indel) identification, read counting for RNA methods, phylogenetic or metagenomic analysis, and more.

The bacterium showed protease and lipase activities on agar plates. Protease activity was characterized and showed interesting characteristics that emphasize the importance of the isolate as a source of novel and interesting enzymes. For these reasons, the genome of *Salicola* sp. SBJ9 was sequenced and resulting sequences were annotated. This work reported the first draft genome of the genus *Salicola* and the first assembly and annotation of its sequenced sequences.

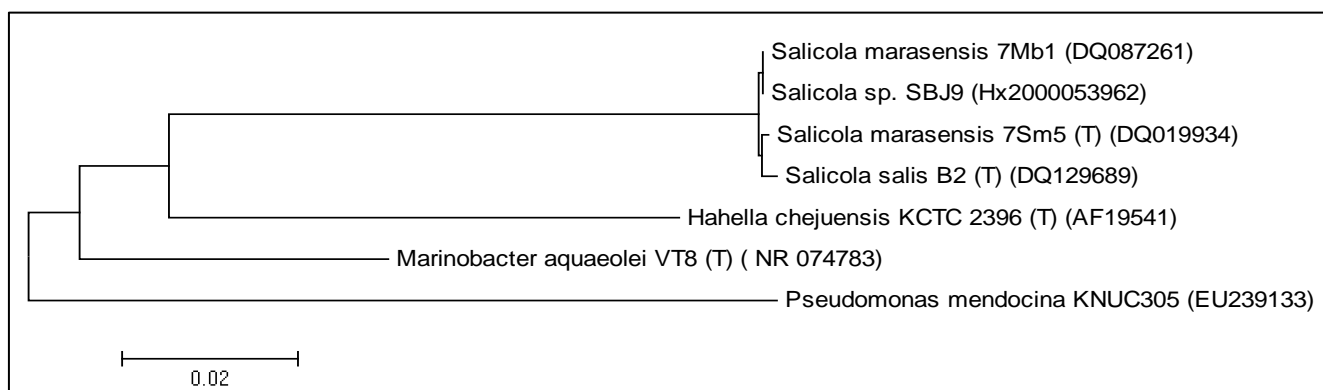


Figure 1: Phylogenetic tree of *Salicola* sp. SBJ9

The genome sequencing of *Salicola* sp. SBJ9, monitored by Illumina system, resulted in 3301046 reads assembled into 1490 contigs, re-assembled into 1303 scaffolds having a total length of 4643441 b with 138.868 b as the longer sequence and 60.52% as the content of G+C.

The assembled sequences were annotated, by the RAST server, and revealed 3995 coding sequences and 65 RNA sequences, distributed in 437 subsystems (Figure 2). Since *Salicola* sp. SBJ9 produced protease and lipase activities, as cited above, we were focused on these enzymes

in its genomic sequence. As a result, we found 33 genes for proteolytic enzymes, 3 genes for lipolytic enzymes, but no gene was detected for the di-, oligo- or polysaccharides hydrolytic enzymes.

The project of the draft genome sequencing of *Salicola* sp. SBJ9 was submitted in EMBL database under the accession number PRJEB15659.

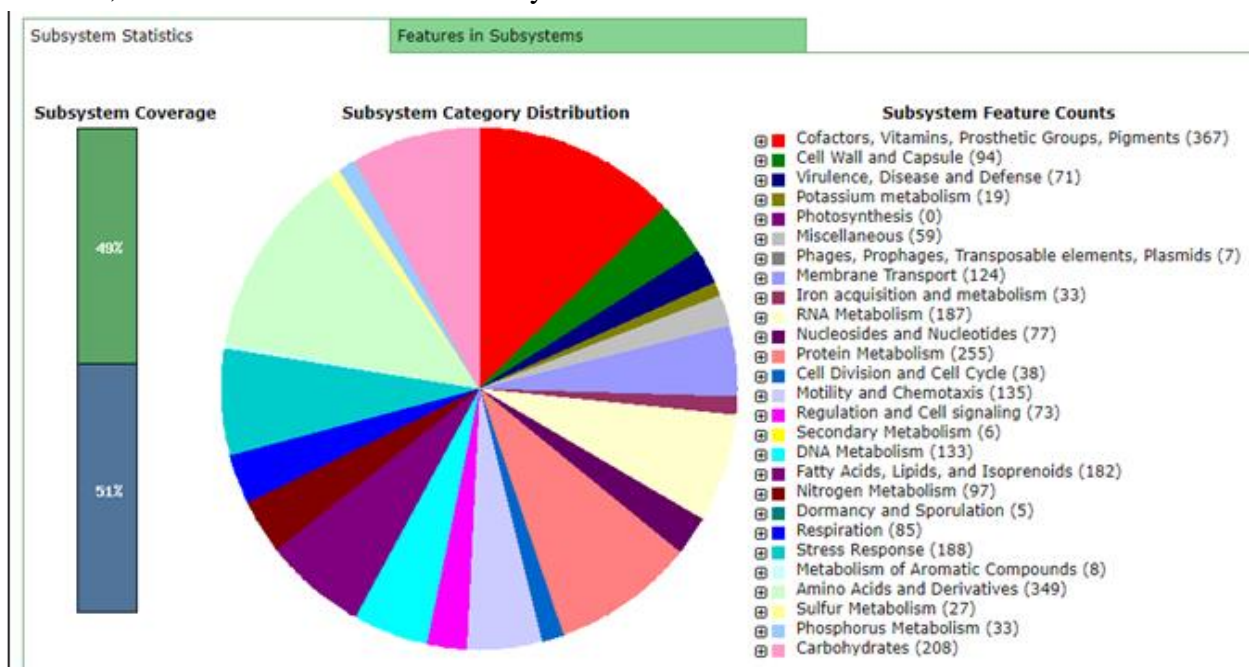


Figure 2. Distribution of the annotated sequences in subsystems. Their number in each subsystem is presented between ().

3. Materials and Methods

Morphological and biochemical characterization was achieved using classical methods [5]. Molecular identification was performed by the amplification of the 16S rDNA gene, using the universal primers S73 (5'-AGAGTTTGATCCTGGCTCAG) and S74 (5'-AAGGAGGTGATCCAGCC) as direct and reverse primers, respectively. Phylogenetic tree was constructed by the MEGA 6 software [6].

The *Salicola* sp. SBJ9 genomic DNA was sequenced according to the method provided by Illumina, a big American company specialized in development, manufacturing and commercialization of services and products for the sequencing, genotyping and genetic expression markets (<http://www.illumina.com>).

First, SBJ9 genomic DNA was extracted and quantified by NanoDrop spectrophotometer (Thermo Scientific™, France). Second, it was submitted to successive manipulations, detailed in the Illumina Guide [7] to generate a sample library. After that, the library mix was loaded onto the reagent cartridge in the designated reservoir and the Flow cell (washed and dried) was loaded. Finally, the reagent cartridge and the buffer bottle were loaded and the run, after a pre-run check, was started [8]. A post-run wash was performed [8].

De novo sequencing refers to sequencing a novel genome where there is no reference sequence available for alignment. Sequence reads are assembled as contigs and the coverage quality of de novo sequence data depends on the

size and continuity of the contigs (ie, the number of gaps in the data).

The annotation of the SBJ9 contigs and the analysis of annotated genes and sequences

4. Conclusions

This report presents the first draft genome sequencing of the genus *Salicola*. The strain template was *Salicola* sp. SBJ9, an extremely halophilic bacterium newly isolated from a hypersaline lake in Tunisia. The sequencing was monitored by the Illumina NGS technology using the MiSeq sequencing system. Results showed the presence of 3995 coding sequences and 65 RNA sequences on a total DNA length of 4643441 b with a G+C content of 60.52%.

were performed by the RAST server (<http://rast.theseed.org/FIG/rast.cgi>).

Among the coding sequences, 33 and 3 genes were detected for proteolytic and lipolytic enzymes, respectively, but no gene was detected for the di-, oligo- or polysaccharides hydrolytic activities. The draft genomic sequence of *Salicola* sp. SBJ9 will provide the genetic information to better understand the mechanisms of high salt adaptation and to isolate and develop novel stable enzymes for industrial applications.

Acknowledgments

This work was funded by the Tunisian Ministry of Higher Education and Scientific Research and Technology (contract program LMBEE-CBS, grant no. LR15CBS06).

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. **Daoud, L.**; Hmani, H.; Ben Ali, M.; Jlidi, M.; Ben Ali, M. An Original Halo-Alkaline Protease from *Bacillus halodurans* Strain US193: Biochemical Characterization and Potential Use as Bio-Additive in Detergents. *Journal of Polymers and the Environment* **2018**, 26, 23-32.
2. **Maturrano, L.**; Santos, F.; Rosselló-Mora, R.; Antón, J. Microbial diversity in Maras salterns, a hypersaline environment in the Peruvian Andes. *Applied and Environmental Microbiology* **2006**, 72 :6, 3887-3895.
3. **Kharroub, K.**; Aguilera, M.; Quesada, T.; Morillo, J.A.; Ramos-Cormenzana, A.; Boulharouf A.; Monteoliva-Sánchez, M. *Salicola salis* sp. nov., an extremely halophilic bacterium isolated from Ezzemoul sabkha in Algeria. *International Journal of Systematic and Evolutionary Microbiology* **2006**, 56: 11, 2647-2652.
4. An introduction to next-generation sequencing technology. Available online: https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf. (Accessed on 03 January 2019).
5. **DasSarma, P.**; Coker, J.A.; Huse, V.; DasSarma, S. Halophiles, Biotechnology. In: *Encyclopedia of Industrial Biotechnology, Bioprocess, Bioseparation, and Cell Technology*; Flickinger M.C., Hoboken N.J., Eds.; John Wiley and Sons, **2010**; pp 2769-2777.
6. **Tamura, K.**; Stecher, G.; Peterson, D.; Filipinski, A.; Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **2013**, 30, 2725-2729
7. Nextera XT Library Preparation Guide. Available online: http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-library-prep-guide-15031942-01.pdf (Accessed on 03 January 2019).
8. MiSeq System Guide. Available online: https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-system-guide-for-miseq-reporter-1000000061014-00.pdf (Accessed on 03 January 2019).