



ASSEMBLY, ANNOTATION AND COMPARISON OF A *PAENIBACILLUS POLYMYXA* GENOME SEQUENCE ISOLATED FROM “EL CHICO” NATIONAL PARK

Medina-Benitez David J; Flores-Bustamante Zoila R; Marsh-Moreno R; Ortega Lopez Jaime; Flores-Cotera Luis B. Department of Biotechnology and Bioengineering. CINVESTAV. Av. Instituto Politécnico Nacional # 2508 Col. San Pedro Zacatenco México, D.F. C.P. 07360.
dmedina@cinvestav.mx, lfcotera@cinvestav.mx

Key words: Paenibacillus, Genome, Assembly.

Introduction. *Paenibacillus polymyxa* is an important microorganism to study, in part because it can be a plant-growth promoter, produces multiple antibiotics, anti-fungal compounds and phytohormones¹. We isolated a *Paenibacillus polymyxa* (Pp) strain from “El Chico” National Park at Hidalgo State, Mexico². The genomic DNA (gDNA) was pyrosequenced using the 454 GS FLX system from Roche. Assemblies from sequencing data were obtained using Newbler and Mira assemblers, and then a comparison was made between the consensus assembly and the reported sequence of *Paenibacillus polymyxa* strain E681, isolated in South Korea³.

The objective of this work was to assemble and annotate the genome of *P. polymyxa* (Pp), and compare the sequence with that of strain E681 to identify similar and divergent genes and regions.

Methods. The gDNA of *Paenibacillus polymyxa* (Pp) was sequenced with a 454 GS FLX system. Two data sets were obtained from two sequencing runs, yielding a total of 996,134 reads with about 300 millions bp. The reads were independently assembled using 454 Newbler and MIRA assemblers, to produce two pre-assemblies with 88 and 121 contigs respectively. These assemblies were aligned and then merged. By alignment between the merged assembly and the E681 sequence, the contigs were sorted for further analysis. The RATT program was used to transfer the annotation from E681 to *Paenibacillus polymyxa* (Pp). In addition Markov model (Mkm) predictions identify open reading frames. The difference that occurs between the annotation transfer file and the generated by Mkm, provide genes belonging to *Paenibacillus polymyxa* (Pp).

Results. We obtained a genomic sequence of *P. polymyxa* (Pp) which comprise 6 contigs (Fig.1). Such sequence has 5,766,306 pb with 45.68% G+C content. The sequence has

13 rRNA operons, and 5,572 ORF according to Mkm predictions. *Paenibacillus polymyxa* (Pp) share 4,491 genes among the 4933 genes annotated for strain E681. The *P. polymyxa* (Pp) genomic sequence has the clusters encoding for the antibiotics polymyxin, bacitracin and bacillomycin, but no plasmids were found. In comparison, the genome of *P. polymyxa* Pp was 371 kb longer than the sequence of E681, distributed throughout the genome.

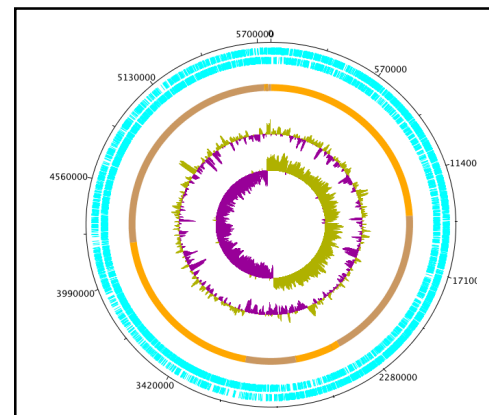


Fig.1 Genome of *Paenibacillus polymyxa* (Pp) comprising 6 contigs.

Conclusions. *P. polymyxa* (Pp) share the 91% of the annotated genes of strain E681. However the genome of stain Pp contains a region 150,000 pb long, that is not present in the E681 strain.

Acknowledgements. For financial support; ICyTDF through project PIFUTP08-90 and Fondo de Investigación Científica Básica Conacyt project 083684.

References.

1. Timmusk S, U. Granhall, E. Tillber. (1999). *SOIL BIOLOGY & BIOCHEMISTRY*, No. 31: 1847-1852
2. Flores-Bustamante Zoila R, Rivera-Orduna Flor N, Martinez-Cardenas Anahi, Flores-Cotera Luis B. (2010). *JOURNAL OF ANTIBIOTICS*, No. 8 Vol.63: 460-467
3. Jihyun F, Haeyoung J, Soo-Young P, Seong-Bin K Yon Kyoung P, Soo-Keun C, Choong-Min Ryu, Cheol-Goo H, Sa-Youl G, Tae Kwang O, Seung-Hwan P. (2010). *JOURNAL OF BACTERIOLOGY*, No. 22 Vol.192: 6103-6104.