

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE FARMACIA Y BIOQUÍMICA

**ESCUELA PROFESIONAL DE CIENCIA DE LOS
ALIMENTOS**

**Genoma completo de un patógeno del género
Clostridium aislado de conservas y análisis
bioinformático comparativo con secuencias de
importancia en inocuidad alimentaria**

TESIS

Para optar el Título Profesional de Licenciado en Ciencia y
Tecnología de los Alimentos

AUTOR

Daisy Maria OBISPO ACHALLMA

ASESOR

Oscar ACOSTA CONCHUCOS

Lima - Perú

2018

RESUMEN

La secuenciación del genoma completo (WGS) y las tecnologías de secuenciación de próxima generación (NGS), viene mostrando grandes avances en el abordaje de detección de patógenos alimentarios, proporcionando una resolución mejorada para caracterizar los brotes de enfermedades transmitidas por alimentos (ETA) y fortalecer el campo de la inocuidad alimentaria a nivel mundial. Por tal motivo, el objetivo de la investigación fue evaluar el genoma completo de un patógeno alimentario nativo del género *Clostridium* y el análisis bioinformático comparativo con secuencias de importancia en inocuidad alimentaria. Inicialmente, se aisló el ADN genómico a partir de conservas alimenticias y de una cepa nativa (cultivo positivo para *Clostridium sp*). En esta investigación, se aplicó la técnica PCR para la identificación molecular de las muestras, luego se secuenció el genoma de la cepa nativa de *Clostridium sp* con MiSeq Illumina, con una cobertura de 90X. Los *reads* fueron ensamblados con SPAdes y los *contigs* fueron filtrados con las herramientas MegaBLAST, MaxBin y AmphoraNet. La verificación de calidad final se realizó con Quast, la anotación del genoma con Prokka y PGAAP-NCBI y la construcción del mapa genómico circular con CGView. El tamaño del genoma fue de 3.966.781pb y 28,06% de contenido GC. La anotación del genoma con Prokka evidenció 3522 genes que codifican proteínas, 54 de ARNt y 8 ARNr, y con PGAAP-NCBI evidenció 3734 genes que codifican proteínas, 54 de ARNt y 8 ARNr. El genoma anotado posee un 99% de identidad para la especie de *Clostridium Botulinum*, con las secuencias nucleotídicas de las cepas de referencia CDC_67071 (GCA_001886775.1) y CDC_1632 (GCA_001889325.1). Además, se identificó al gen *bontB* (productor de neurotoxina tipo B y causante del botulismo alimentario), con un porcentaje de identidad de secuencias nucleotídicas en 68,73% con las cepas CDC_67071 y CDC_1632 de *C. botulinum*. En general, el genoma evaluado se identificó como *C. botulinum* y se determinó el gen *bontB* productor de toxina tipo B, lo que indica la gran utilidad de las tecnologías NGS y la bioinformática en el campo de la inocuidad alimentaria del país.

Palabras clave: *Clostridium*, WGS, NGS, bioinformática, inocuidad alimentaria.

ABSTRACT

Whole genome sequencing (WGS) and next generation sequencing (NGS) technologies have shown great advances in the approach of detection of food pathogens, providing an improved resolution to characterize the outbreaks of foodborne diseases (ETA) and strengthen the field of food safety worldwide. *For this reason*, the objective of the research was to evaluate the complete genome of a native food pathogen of the *Clostridium* genus and perform comparative bioinformatics analysis with sequences of importance in food safety. Initially, genomic DNA was isolated from canned food and of a native strain (positive culture for *Clostridium* sp.). In this investigation, the PCR technique was applied for the molecular identification of the samples, then the genome of the native *Clostridium* sp strain was sequenced using MiSeq Illumina, with a genomic coverage of 90X. The reads were assembled with SPAdes and the *contigs* were filtered with MegaBLAST, MaxBin and AmphoraNet tools. The final quality verification was performed with Quast, the annotation of the genome with Prokka and PGAAP-NCBI and the construction of the circular genomic map with CGView. The genome size is 3.966.781bp and 28,06% of GC contents. The annotation of the genome with Prokka showed 3522 genes encoding proteins, 54 of tRNA and 8 rRNA, and with PGAAP-NCBI evidenced 3734 genes that encode proteins, 54 of tRNA and 8 rRNA. The annotated genome has 99% identity for the *Clostridium Botulinum* species, with the nucleotide sequences of the reference strains CDC_67071 (GCA_001886775.1) and CDC_1632 (GCA_001889325.1). In addition, the gene *bontB* (producer of type B neurotoxin and cause of food botulism) was identified, with a percent identity of nucleotide sequences in 68.73% with strains CDC_67071 and CDC_1632 of *C. botulinum*. In general, the genome evaluated was identified as *C. botulinum* and the gene *bontB* that codes for the type B toxin was determined, which indicates the great utility of NGS technologies and bioinformatics in the field of food safety in the country.

Key words: *Clostridium*, WGS, NGS, bioinformatics, food safety.