
**ANÁLISIS COMPARATIVO DE METAGENOMAS MICROBIANOS DE
ECOSISTEMAS ACÍDICOS.**

**ANA ROSA MOYA BELTRÁN.
INGENIERO EN BIOINFORMÁTICA**

RESUMEN

El análisis genómico de comunidades microbianas representa una importante herramienta para investigar la historia evolutiva de la biodiversidad inherente y la interacción de funciones entre los microorganismos que las constituyen.

El método tradicional Sanger para la secuenciación de DNA, ampliamente utilizado en la secuenciación de genomas, resulta inadecuado para el procesamiento de comunidades microbianas complejas, las cuales a menudo contienen una mezcla de DNA de una gran cantidad de individuos y especies. Sin embargo, la secuenciación de última generación (NGS), ha permitido superar las limitaciones de las tecnologías convencionales, posibilitando la recuperación de información de secuencia directamente, sin necesidad de cultivar los microorganismos. Este enfoque ha sido denominado “metagenómica”.

El post-procesamiento de los datos metagenómicos es complejo y se diferencia del análisis genómico convencional en una serie de aspectos. Los más relevantes son la gran cantidad de información que se debe procesar (Gb vs. Mb) y el pequeño tamaño de fragmentos de DNA generados (100 pb vs. 800 pb). Ambos aspectos dificultan considerablemente la capacidad de ensamblar y reconstruir genomas individuales a partir de un metagenoma. Cuando se dispone de información genómica previa de los microorganismos, el análisis se ve considerablemente facilitado.

Esta tesis explora la constitución taxonómica y la composición génica de 5 metagenomas obtenidos de ambientes acídicos respecto de un conjunto de 56 genomas acidófilos individuales, utilizando métodos de análisis metagenómico y de redes.

Los resultados muestran: primero la dominancia del género *Acidithiobacillus*, revelando que este grupo de microorganismos juega un papel importante en ambientes acídicos, segundo la presencia de diversas familias de genes es develada en la composición génica de cada comunidad. Estos resultados basados

en el análisis de redes proveen una nueva oportunidad para estudiar y entender la genética y diversidad de muestras metagenómicas.

ABSTRACT

Genomic analysis of microbial communities represents an important approach for investigating the evolutionary history of inherent biodiversity and the interplay of functions between the microorganisms that constitute them.

The traditional Sanger method for DNA sequencing, widely used in the sequencing of individual genomes, is unsuitable for processing complex microbial communities, which often contain a mixture of DNA of several tens of thousands of individuals. However, next generation sequencing (NGS), has overcome the limitations of conventional technologies, allowing the recovery of sequence information directly, without the need to culture the microorganisms. This approach has been termed “metagenomics”.

The post-processing of metagenomic data is complex and is different from conventional genomic analysis in a number of respects. The most relevant are the large amount of information to be processed (Gb vs. Mb) and the small size of the generated DNA fragments (100 bp vs. 800 bp). Both aspects significantly hamper the ability to assemble and reconstruct individual genomes within the metagenome. When genomic information about the microorganism is available in advance, the analysis is greatly facilitated.

This thesis explored the taxonomic constitution, genetic composition and metabolic potential of 5 metagenomes obtained from acidic environments from a set of 56 individual acidophilic genomes using metagenomic and network analysis methods.

The results shows: first, the dominance of the *Acidithiobacillus* genus, revealing that this group of organisms plays an important role in acidic environments; second, the presence of various gene families were discovered by genetic composition in each community. These results, based in network analysis, provide a new opportunity to study and understand the genetic diversity and metagenomic samples.