

# **Ensamblaje metagenómico y predicción de rutas metabólicas por ShotGun en microbiomas del Lago Titicaca**

## **Introducción**

El lago Titicaca es considerado el lago navegable más alto del mundo. Posee una antigüedad aproximada de 3 millones de años. Además, se es consciente que dicho lago es empelado para el abastecimiento de agua potable de casi toda la población de Puno. A pesar de su inmensa importancia tanto histórica como ambiental, se observó que este ecosistema está constantemente expuesto a contaminantes provenientes de residuos sólidos vertidos por la población, relaves mineros, aguas servidas, entre otros.

Debido a estos sucesos que afectan al Titicaca, se infiere que los microbiomas presentes en el lago poseen diversas rutas metabólicas características a su ecosistema y factores externos. Siendo posible determinar rutas de degradación de celulosa, hemicelulosa, quitina, almidón, entre otros carbohidratos complejos. Además, determinar rutas de biosíntesis de antibióticos indicaría el gran potencial de las bacterias del lago Titicaca para la bioprospección de antibióticos naturales potencialmente novedosos.

Por ello, es de alta relevancia tener un perfil sólido del estado del Titicaca respecto a las rutas metabólicas que poseen las poblaciones bacterianas. El campo de la biología molecular brinda diversas herramientas que permiten realizar estos estudios. Este es el caso de la metagenómica, que analiza todos los genomas existentes en un consorcio de especímenes, en este caso, bacteriano. La técnica que se propone es la de Shotgun, prueba novel en el campo de la metagenómica que permite tener un perfil completo de todas las secuencias de nucleótidos extraídas previamente.

Finalmente, será requerido otras herramientas bioinformáticas y de programación. La realización de estos análisis a través de sistemas operativos como Windows es inviable. Por ello, se requiere el dominio de lenguajes de programación a través de la línea de comando en una arquitectura Unix como la del sistema operativo Linux.

## **Planteamiento del problema**

Los lagos constituyen el 50% de toda el agua de la superficie de la Tierra y aproximadamente el 49% de agua dulce líquida de manera superficial. Hoy en día, el turismo lacustre provoca un crecimiento urbano en la región de las tierras altas y drásticas actividades antropogénicas, que afectan negativamente sus recursos hídricos. Estas actividades incluyen la adición de efluentes, la eutrofización y la invasión de malas hierbas nocivas. Además, el aterramiento de la deposición y sedimentación de materia orgánica modifica el estado ecológico del lecho sedimentario. Los lagos de terreno montañoso están muy influenciados por la rápida tasa de descarga (Vreca and Muri, 2006).

En adición, el Titicaca es una frontera natural entre los países Bolivia y Perú. Respecto a la parte peruana, Flores & Quiñonez (2021) reportan que la contaminación de la bahía del interior del lago son en gran parte responsabilidad de las actividades antrópicas y que amerita proponer estrategias que permitan mitigar esta contaminación.

Precisamente, los lagos de gran altitud generalmente representan un ecosistema comparativamente único debido a su lejanía y al corto período de verano de aguas abiertas; sin embargo, están menos explorados en comparación con los lagos de tierras bajas. Siendo los sedimentos una rica fuente de alta diversidad microbiana en ambientes acuáticos (Bhat et al., 2011).

Entre las metodologías más importantes para un análisis exhaustivo en microbiomas está la de secuenciamiento de metagenomas, comparado a otras técnicas como PCR, qPCR, entre otros la metagenómica brinda un análisis global de los genes de resistencia a antibióticos en muestras ambientales (Li et al., 2015).

La mayoría de los microorganismos que se encuentran en el medio ambiente permanecen sin cultivar, los enfoques tradicionales basados en cultivos no logran determinar la diversidad taxonómica y funcional de la comunidad microbiana. Por este motivo, el enfoque metagenómico se ha aplicado con éxito a una variedad de ambientes acuáticos (Samson et al., 2019).

El enfoque metagenómico supera al cultivo microbiano al ayudar al descubrimiento de vías metabólicas y nuevas secuencias de genes de microorganismos no cultivables directamente del medio ambiente. Este es el principal beneficio de la metagenómica, ya que el cultivo de nuevos microorganismos es una técnica ardua. Estudios previos sobre microbios independientes no cultivados descubrieron que la diversidad de organismos en los lagos salobres es inusualmente alta con una riqueza de especies en comparación con los ecosistemas de agua dulce (Tang et al., 2015)

Uno de estos métodos, la metagenómica ShotGun, es la secuenciación no dirigida ("ShotGun") de todos los ("meta-") genomas microbianos "genómicos" presentes en una muestra. La secuenciación ShotGun se puede utilizar para realizar un perfil de la composición taxonómica y las rutas metabólicas de las comunidades microbianas y recuperar secuencias genómicas completas. Los enfoques como la secuenciación del gen 16S rRNA de alto rendimiento, que perfilan organismos seleccionados o genes marcadores únicos, a veces se denominan metagenómica, pero este es un nombre inapropiado, ya que no se enfocan en todo el contenido genómico de una muestra (Quince et al., 2015)

El ensamblaje metagenómico se construye mediante un gráfico de De Bruijn dividiendo cada lectura de secuenciación en subsecuencias superpuestas de una longitud fija  $k$ . Este conjunto de 'k-meros' superpuestos define los vértices y los bordes del gráfico de de Bruijn. La tarea del ensamblador es encontrar un camino a través del gráfico que reconstruya el o los genomas. Esta tarea se complica por los errores de secuenciación, que generan secuencias no genómicas, y secuencias repetitivas, que pueden provocar errores de ensamblaje y fragmentación del ensamblaje (Pevzner et al., 2001).

El ensamblaje del metagenoma presenta desafíos únicos. En primer lugar, cuando se ensambla un solo genoma, normalmente se supone que la cobertura de la secuencia a lo largo del genoma será aproximadamente uniforme. Un ensamblador puede usar la cobertura de secuencia para identificar copias repetidas, distinguir la secuencia verdadera de los errores de secuencia e identificar la variación alélica. El ensamblaje del metagenoma es más difícil, porque la cobertura de cada genoma constituyente depende de la abundancia de cada genoma en la comunidad (Simpson, 2014).

La predicción de rutas metabólicas puede obtenerse con un ensamblaje de metagenoma fragmentado, pero de alta calidad, utilizando herramientas de caracterización adaptadas de un solo genoma. Estos incluyen un paso de identificación de genes, especialmente relacionadas a rutas metabólicas y que poseen con una configuración de parámetros específicos de metagenómica, seguido de canalizaciones de anotación basadas en homología comúnmente utilizadas para caracterizar ensamblajes de genomas aislados puros (Zhu et al., 2010).

Para la identificación de las rutas metabólicas es necesario el uso de bases de datos que incluyen combinaciones de familias de proteínas anotadas manualmente y predichas computacionalmente, como KEGG o UniProt, se pueden usar para esta tarea y permiten la caracterización del potencial funcional del microbioma. Las rutas metabólicas altamente conservadas se detectan y cuantifican de manera más consistente en los metagenomas, lo que podría explicar por qué los rasgos funcionales a menudo son sorprendentemente consistentes en diferentes muestras y entornos (Quince et al., 2015).

Entonces, surge la siguiente interrogante ¿El secuenciamiento por Shotgun permitirá obtener un correcto ensamblaje metagenómico y la predicción de rutas metabólicas en microbiomas del lago Titicaca?

### **Antecedentes**

Rathour et al. (2020) realizaron un estudio metagenómico comparativo determinando la diversidad microbiana y su rol en el ciclo biogeoquímico del Lago Pangong. El lago Pangong es un lago de agua salobre de gran altitud del Himalaya situado en la parte oriental de Ladakh (Tíbet indio), a una altura de 4250 m sobre el nivel del mar. Se realizó la secuenciación metagenómica por Shotgun de los sedimentos del lago Pangong para examinar la diversidad taxonómica y las rutas metabólicas de las comunidades microbianas psicrófilas y psicrotolerantes residentes del lago. Reportando vías enzimáticas responsables del metabolismo del metano, el metabolismo del nitrógeno, la reducción del azufre, el benzoato y la degradación del xileno parecían estar completas en el conjunto de datos metagenómicos. Los genes de respuesta al estrés responsables de la adaptación al pH, el frío, la tolerancia a la sal, el estrés osmótico y el estrés oxidativo también se encontraron en abundancia en el metagenoma.

Adicionalmente, Warden et al. (2016) Caracterizaron los microbiomas del lago Clifton en Australia mediante Shotgun. Determinaron rutas metabólicas responsables de la síntesis de carbohidratos, la fotosíntesis, el metano y el metabolismo del nitrógeno, representaron del 46,5 al 47,9 % de las lecturas asignadas, mientras que el procesamiento de información genética (19,9 a 20,9 %) y el procesamiento de información ambiental (10,1 a 11,1 %) fueron los siguientes grupos más abundantes. Aproximadamente el 16 % del total de lecturas no se pudo clasificar según la función metabólica.

Finalmente, Rai & Bhattacharjee (2021) realizaron un perfil molecular a las comunidades bacterianas del lago Tsomgo en el Himalayas. Utilizaron el secuenciamiento por Shotgun. El análisis de las muestras revela que ambas muestras son muy ricas en proteínas transportadoras de aminoácidos y proteínas metabólicas de carbohidratos. Esto se ha manifestado por la presencia de una gran cantidad de CAZimas tanto en el suelo como en las muestras de agua.

## **Objetivos**

### **Objetivo General**

Determinar las rutas metabólicas en metagenomas de consorcios bacterianos del lago Titicaca.

### **Objetivos específicos**

Realizar el ensamblaje metagenómico de consorcios bacterianos del lago Titicaca secuenciados mediante ShotGun.

Obtener el perfil taxonómico de consorcios bacterianos del lago Titicaca.

## **Metodología**

### **Filtración de las muestras líquidas del lago Titicaca**

La muestra líquida obtenida se filtrará en una membrana de 0.2  $\mu\text{m}$  (Whatman Nucleopore – Sigma Aldrich) a través de un embudo de filtro desechable (Pall MicroFunnel – Thermo Fisher Scientific) conectado a un colector de vacío.

### **Extracción de ADN, cantidad, calidad e integridad**

Se utilizará el kit de purificación Purelink – Thermo Fisher Scientific; la cuantificación y calidad en un espectrofotómetro de microvolúmenes UV/Vis (Lambda – PCR max); la integridad de ADN extraído se determinará por medio de electroforesis de geles de agarosa a concentración del 1%.

### **Secuenciamiento por Shotgun**

Se llevará a cabo a una longitud de lectura PE 2x150 pb y 10 millones de lecturas por muestra (5M en cada dirección).

### **Análisis bioinformático**

Se programará a través de la línea de comando en una arquitectura Unix, usando los programas MEGAHIT para el ensamblado de genes y bases de datos de genes microbianos: KEGG. Para la obtención del perfil taxonómico se utilizará la línea de comando Unix y los códigos Metaphlan2 y Graphlan.

## **Referencias Bibliográficas**

Bhat, F. A., Yousuf, A. R., Aftab, A., Arshid, J., Mahdi, M. D., & Balkhi, M. H. (2011). Ecology and biodiversity in Pangong Tso (lake) and its inlet stream in Ladakh, India. *International Journal of Biodiversity and Conservation*, 3(10), 501-511.

- Flores, M. Y., & Quiñonez, J. (2021). Responsabilidad social y ambiental de las organizaciones sobre la contaminación de la bahía interior del lago Titicaca, Puno. *Ciencia Latina Revista Científica Multidisciplinar*, 5(5), 8270-8282.
- Li, B., Yang, Y., Ma, L., Ju, F., Guo, F., Tiedje, J. M., & Zhang, T. (2015). Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *The ISME journal*, 9(11), 2490-2502.
- Pevzner, P. A., Tang, H., & Waterman, M. S. (2001). An Eulerian path approach to DNA fragment assembly. *Proceedings of the national academy of sciences*, 98(17), 9748-9753.
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature biotechnology*, 35(9), 833-844.
- Rai, A., & Bhattacharjee, A. (2021). Molecular profiling of microbial community structure and their CAZymes via metagenomics, from Tsomgo lake in the Eastern Himalayas. *Archives of Microbiology*, 203, 3135-3146.
- Rathour, R., Gupta, J., Mishra, A., Rajeev, A. C., Dupont, C. L., & Thakur, I. S. (2020). A comparative metagenomic study reveals microbial diversity and their role in the biogeochemical cycling of Pangong lake. *Science of the Total Environment*, 731, 139074.
- Samson, R., Shah, M., Yadav, R., Sarode, P., Rajput, V., Dastager, S. G., ... & Khairnar, K. (2019). Metagenomic insights to understand transient influence of Yamuna River on taxonomic and functional aspects of bacterial and archaeal communities of River Ganges. *Science of the Total Environment*, 674, 288-299.
- Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J., & Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome research*, 19(6), 1117-1123.
- Tang, X., Xie, G., Shao, K., Dai, J., Chen, Y., Xu, Q., & Gao, G. (2015). Bacterial community composition in oligosaline lake Bosten: Low overlap of betaproteobacteria and bacteroidetes with freshwater ecosystems. *Microbes and environments*, 30(2), 180-188.
- Vreca, P., & Muri, G. (2006). Changes in accumulation of organic matter and stable carbon and nitrogen isotopes in sediments of two Slovenian mountain lakes (Lake Ledvica and Lake Planina), induced by eutrophication changes. *Limnology and Oceanography*, 51(1part2), 781-790.
- Warden, J. G., Casaburi, G., Omelon, C. R., Bennett, P. C., Breecker, D. O., & Foster, J. S. (2016). Characterization of microbial mat microbiomes in the modern thrombolite ecosystem of Lake Clifton, Western Australia using shotgun metagenomics. *Frontiers in Microbiology*, 7, 1064.
- Zhu, W., Lomsadze, A., & Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic acids research*, 38(12), e132-e132.