

## Investigar en Bioinformática. ¿Cómo evaluar la riqueza de una comunidad microbiológica?

**Elaborado por:** Cristóbal R. Santa María. Universidad Nacional de La Matanza, Departamento de Ingeniería e investigaciones Tecnológicas, San Justo, Bs. As. Argentina.

En el marco de la nueva disciplina metagenómica que analiza comunidades microbianas, las técnicas de secuenciación computacional de ADN permiten contar con información muestral de genes marcadores en cantidad suficiente como para intentar establecer los parámetros de riqueza y diversidad de poblaciones completas. Con los procedimientos de alineado, filtrado y agrupamiento de secuencias en “clusters” denominados Unidades Taxonómicas Operacionales (OTUs) se puede conocer la cantidad de especies, de familias o de géneros de una muestra e inferir el número de especies, familias o géneros presentes en la población total así como también su distribución de abundancia. Estas cantidades no representan otra cosa que la riqueza y la biodiversidad de la comunidad. Sin embargo dicha inferencia ofrece dificultades metodológicas pues las técnicas

estadísticas no paramétricas habitualmente utilizadas para realizarla subestiman la riqueza real de las comunidades. Esto es consecuencia del tamaño extremadamente pequeño de las muestras posibles frente al gran número de microorganismos presentes usualmente en las comunidades microbianas y plantea entonces un gran desafío estadístico y computacional.

Dentro del Proyecto de Investigación del Programa de Incentivos “Aplicaciones de Data Mining al Estudio de la Biodiversidad en Comunidades Microbiológicas”, llevado a cabo durante 2011 y 2014 en el Departamento de Ingeniería de la UNLAM, se desarrolló un Algoritmo de Recuento de Especies (ARE) basado en la ampliación simulada de la muestra a partir de una estimación de especie nueva diseñada por Alan Turing alrededor del año 1950. La efectiva mejora producida por tal

procedimiento fue testada construyendo una gran población simulada basada en la distribución de Fischer de las especies y seleccionando muestras de ella con las cuales se infirió la riqueza de la comunidad. Los resultados obtenidos mostraron la convergencia del algoritmo ARE hacia el valor de la riqueza poblacional conforme crecía el número de interacciones.

Lo propio ocurrió al considerar grandes muestras de comunidades reales y seleccionar pequeñas submuestras para inferir a partir de ellas la riqueza del total del conjunto por aplicación de ARE. La convergencia hacia el valor conocido de riqueza fue primero observada en forma empírica y luego confirmada mediante una demostración matemática a partir del Proceso Aleatorio de Cantidad de Especies utilizado como modelo. De esta manera se construyó un método de cálculo que mejora la estimación de la

riqueza de las comunidades microbianas y permite con ello una evaluación ecológica más precisa de la riqueza de suelos, de la degradación del medioambiente o de la flora bacteriana del tracto intestinal humano que son, entre otros, los campos de aplicación de la metagenómica. Esta ciencia estudia no ya el genoma de un individuo o de una especie sino que analiza en forma conjunta los genomas de todas las especies, familias o géneros que integran una comunidad biológica. Su gran desarrollo a partir del comienzo del presente siglo ha ido de la mano del aumento en las capacidades de almacenamiento y proceso de datos de las modernas computadoras así como del abaratamiento de la secuenciación de ADN que lo transforma en cadenas computacionales de símbolos dando pie a la investigación en el área denominada biología computacional o bioinformática

### **Fuente**

Santa María, C. y Soria, M. (2013). Inferencia de Parámetros de Biodiversidad por medio de Simulación. Buenos Aires. 4to Congreso de Matemática Aplicada, Computacional e Industrial.

Santa María, C. y Soria, M. (2012). Estimation of Species Richness in Microbial Communities. Oro Verde. Memorias del 3er Congreso Argentino de Bioinformática y Biología Computacional.

Santa María, C. y Soria, M. (2011). Estimación de Biodiversidad por Data Mining y Simulación. La Plata. XVII Congreso Argentino de Ciencias de la Computación.