

MetaEvoMining: una herramienta para la exploración de nuevas rutas biosintéticas aplicada a datos metagenómicos de un océano contaminado.

Anónimo

Abstract Encontrar rutas químicas que sintetizan nuevos compuestos como los antibióticos, o rutas que degradan contaminantes, es un desafío relevante para la humanidad. Los datos genómicos son un recurso útil para encontrar nuevos clústeres de genes biosintéticos (BGCs) que son genes relacionados con las capacidades metabólicas de los microorganismos. La recopilación de datos metagenómicos ha aumentado en los últimos 20 años y estudios recientes ofrecen como datos públicos más de 10,000 metagenomas de la Tierra. EvoMining es una herramienta de minería genómica que se ha usado como un método para identificar enzimas clave del metabolismo especializado. EvoMining no se ha utilizado en datos metagenómicos a pesar de la disponibilidad de gran cantidad de datos públicos, dejando mucha información de las rutas químicas de muchos microorganismos aún sin explorar. Para conectar EvoMining con los datos metagenómicos, desarrollamos un paquete de R que permite esta conexión.

Las aguas contaminadas son un entorno en el que los científicos están particularmente interesados por el peligro que representan para los servicios ecosistémicos. Este trabajo utiliza datos metagenómicos de unas muestras oceánicas de zonas contaminadas por hidrocarburos y busca explorar aquellas funciones metabólicas desconocidas de los microorganismos que lidian con este entorno. Para explorar nuevas enzimas, utilizamos EvoMining para los datos metagenómicos. Para encontrar familias de enzimas reclutadas en el metabolismo especializado, EvoMining utiliza distancias basadas en la evolución en genomas. En las familias de enzimas del metabolismo conservado, EvoMining encuentra aquellas familias que han ampliado su número de copias en los genomas en un linaje determinado. Después de la comparación con una base de datos de enzimas clave experimentalmente validadas, EvoMining predice novedades en las rutas metabólicas. Esta es una nueva estrategia alternativa para explorar rutas metabólicas en la información metagenómica.

MetaEvoMining consiste en tres módulos, uno de asignación funcional y taxonómica, otro de creación de bases de datos enzimáticas y el tercero de predicción de nuevas enzimas clave. La dificultad de usar EvoMining en datos metagenómicos es que ya no hay conocimiento sobre los números de copia en un genoma, y la clasificación taxonómica puede ser no tan precisa. Sin embargo, después de asignar taxonómicamente, los datos metagenómicos pueden ser complementados con datos genómicos, lo que permite hacer predicciones más fiables.

La metagenómica estudia el contenido genético de comunidades enteras de microorganismos que viven en un entorno. Estos microorganismos evolucionan para mejorar sus aptitudes en un entorno lleno de antagonismo. La información de estas comunidades enteras puede ayudarnos a descubrir nuevas rutas químicas. Nuestro paquete MetaEvoMining, es una herramienta que nos permitirá explorar la información de los microorganismos provenientes de un océano contaminado y nos ayudará a hacer predicciones acerca del metabolismo en datos metagenómicos de todo el mundo.

Palabras clave: Metagenómica - Minería genómica - Paquete de R - Biología - Bioinformática