



ID de aportación : 34

Tipo: **Oral**

## **refseqR : operaciones computacionales comunes con registros de la colección RefSeq (NCBI)**

La colección de secuencias de referencia del Centro Nacional de Información Biotecnológica (RefSeq, NCBI) mantiene un conjunto de secuencias completo, no redundante y bien anotado, que incluye genomas, transcritos y proteínas. En el momento de escribir estas líneas, el proyecto RefSeq contiene más de 60 millones de transcritos y 320 millones de secuencias de proteínas. En este trabajo describimos refseqR que proporciona un marco práctico para manejar secuencias biológicas alojadas en la colección RefSeq. refseqR simula el flujo de información genética dentro de un sistema biológico, permitiendo procesos direccionales desde loci genéticos recogidos como registros GenBank, a transcritos y de ahí a secuencias de proteínas curadas a partir de la base de datos RefSeq, así como otras combinaciones entre secuencias de estas moléculas. refseqR permite la interoperabilidad y la integración con varios objetos de Bioconductor proporcionando una conexión directa con otros proyectos. El paquete refseqR está implementado en R y se publica bajo la licencia MIT de código abierto.

**¿Presentas la comunicación a premio?**

### **Afiliación (del autor)**

Universidad de Córdoba

**Autor primario:** DIE, Jose V. (University of Cordoba)

**Clasificación de pistas:** Ciencias naturales