

El Instituto de Investigaciones Biológicas Avanzadas de Güejar Sierra dispone de una base de datos de ADN de diversas especies biológicas del Parque Nacional de Sierra Nevada codificadas como secuencias de caracteres (A, T, G y C) de distinta longitud según la especie. Esta base de datos se compone de 1000 registros donde cada registro contiene el nombre de la especie y la secuencia de su ADN, ambas representadas como cadenas de caracteres (cstring). Ejemplos de la información asociada con distintas especies son:

Nombre: “La mosca del Vinagre”, ADN: “ATAATGGACAAT”

Nombre: “La lombriz de tierra”, ADN: “GGATACT”

Nombre: “La ameba verde”, ADN: “AGAGAT”

El instituto necesita un programa de ordenador que les permita identificar a qué especie pertenece una toma de ADN que se ha recogido en el campo y analizado en el laboratorio. Teniendo en cuenta que las secuencias de ADN analizadas pueden comenzar en cualquier posición de la secuencia. Por ejemplo, las secuencias “GGATACT” y “ACTGGAT” pertenecen a la misma especie (“La lombriz de tierra”). Observad que la secuencia es la misma, aunque comienza en un punto distinto. Se pide:

- Diseñar las estructuras de datos necesarias para almacenar en memoria principal esta base de datos.
- Diseñar una función que, a partir de la base de datos y de una secuencia de ADN específica nos indique si está registrada en la base de datos y, en caso afirmativo, cual es el nombre de la especie. Se valorará especialmente la eficiencia de la solución, tanto en tiempo de respuesta como en uso de memoria.