



Diseño de Base de Datos
Grado en Ingeniería Informática en Sistemas de Información
MODELO E/R EXAMEN ENERO

La empresa GENE ONTOLOGY quiere diseñar una base de datos en la que almacenar la información de los genes, secuencias genéticas y pathways bioquímicos de un conjunto de organismos. Los requisitos son los siguientes:

- Parte1:

La base de datos almacenará información de múltiples organismos. De cada organismo se desea almacenar un código identificativo, el nombre científico, el nombre vulgar y el número de genes perteneciente a ese organismo de los cuales se guarda información en la BD. Hay que decir solo se almacenarán datos de organismos eucariotas.

De cada organismo se almacena información de un conjunto de genes. Hay que tener en cuenta que los organismos pueden compartir genes entre sí. Todos los genes tienen un código interno llamado GolD. Además, se almacena su nombre y su alias. Se da la circunstancia de que un gen puede tener varios nombres diferentes. Un gen está compuesto por varias secuencias de nucleótidos, numeradas de manera secuencial para cada gen. Además de este número identificativo, de las secuencias necesitamos conocer en que cromosoma la encontramos, la fecha en la que fue descubierta y el número de nucleótidos que la forman. Toda secuencia está formada por un número mínimo de nucleótidos al que denominaremos X. Las secuencias se relacionan entre sí mediante las alineaciones. Dos secuencias están alineadas si su lista de nucleótidos coincide en un determinado porcentaje. De cada relación entre secuencias habrá que almacenar ese porcentaje. Solo si el porcentaje es mayor al 70%, tendremos la alineación en cuenta.

Los nucleótidos son moléculas orgánicas formadas por la unión covalente de un monosacárido de cinco carbonos (pentosa), una base nitrogenada y un grupo fosfato. Cada nucleótido se identifica por su nombre (que puede ser: Adenina, Guanina, Timina, Citosina, Uracilo y Flavina) y la posición que ocupa en una secuencia concreta. Además, necesitamos conocer su tipo de base nitrogenada.

- Parte2:

Según la acción que desempeñen, las secuencias se pueden clasificar en: codificadores, aquellas que se utilizan para la fabricación de proteínas, y reguladores, que regulan la acción de otras secuencias. De las secuencias codificadoras necesitamos conocer el nombre de la proteína que codifica.

Los pathways son procesos bioquímicos que tienen lugar dentro de la célula y en los que intervienen las secuencias de nucleótidos. Una secuencia puede intervenir en un pathway, mediante la acción de la proteína que codifica. Sin embargo, si una secuencia interviene en un pathway y dicha secuencia es reguladora, necesitamos conocer qué tipo de regulación lleva a cabo en dicho proceso (Regulación positiva o regulación negativa). Los pathways están identificados por un nombre y además necesitamos conocer el número de fases que los componen y la función que dicho proceso desempeña.

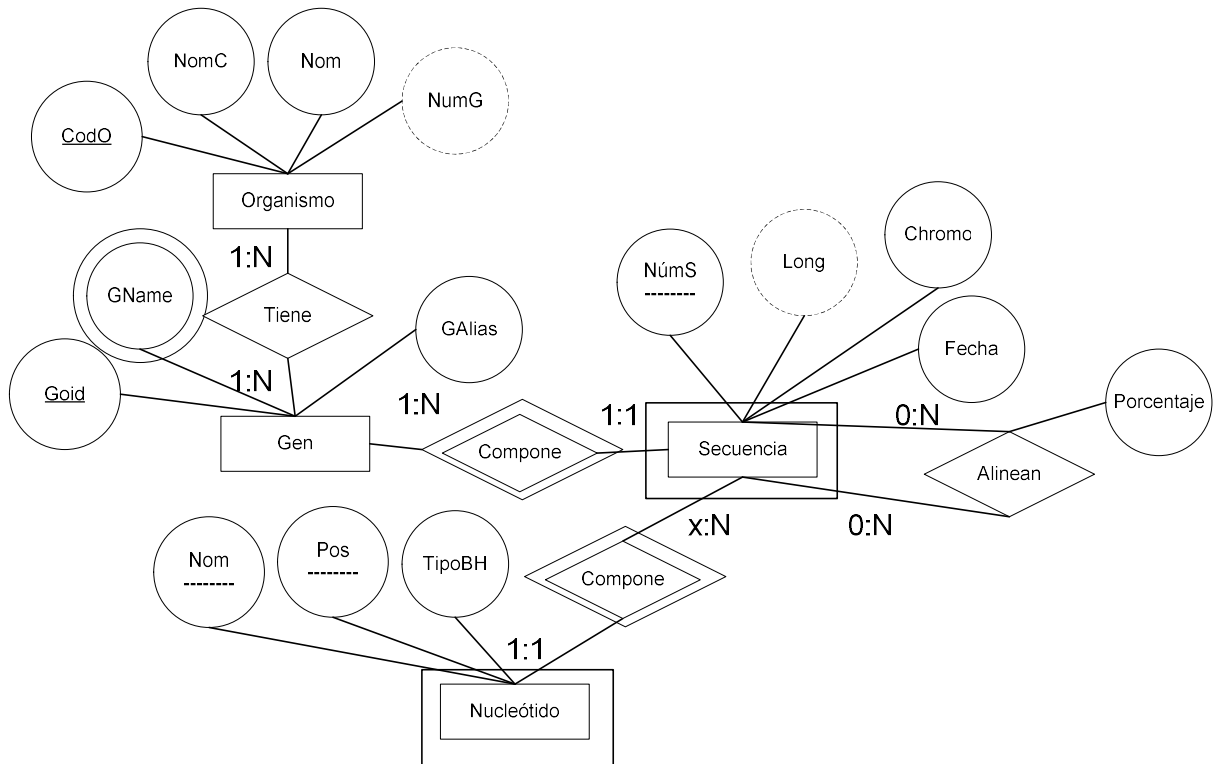
- Parte3:

La base de datos también almacena redes de genes, las cuales han sido descubiertas computacionalmente. De una red necesitamos conocer un identificador y el número de nodos que la forman. La red está compuesta por un conjunto de aristas (1 como mínimo). A su vez, cada arista relaciona a 2 genes entre sí. Además de querer almacenar un número para cada arista, que se corresponde con el orden en que los genes se combinan dentro de la red, las identificaremos mediante un código propio más el código de la red a la que pertenece.

Una red está vinculada, como mínimo, a un proceso bioquímico o pathway. Hay que tener en cuenta que el hecho de que un gen contenga una secuencia codificadora que intervenga en un pathway no quiere decir que tenga que aparecer necesariamente en una red vinculada a ese pathway en concreto.

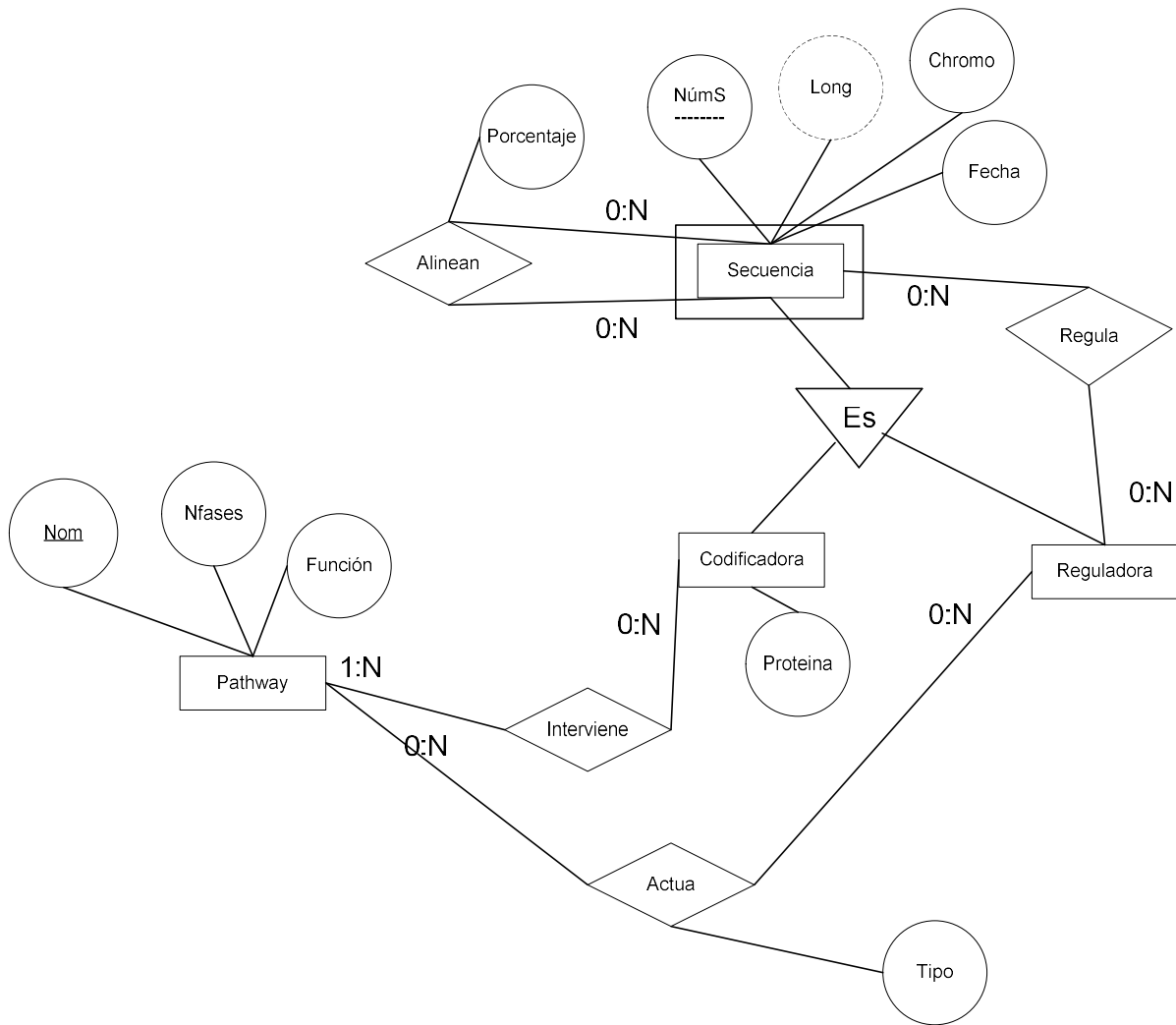


PARTE1



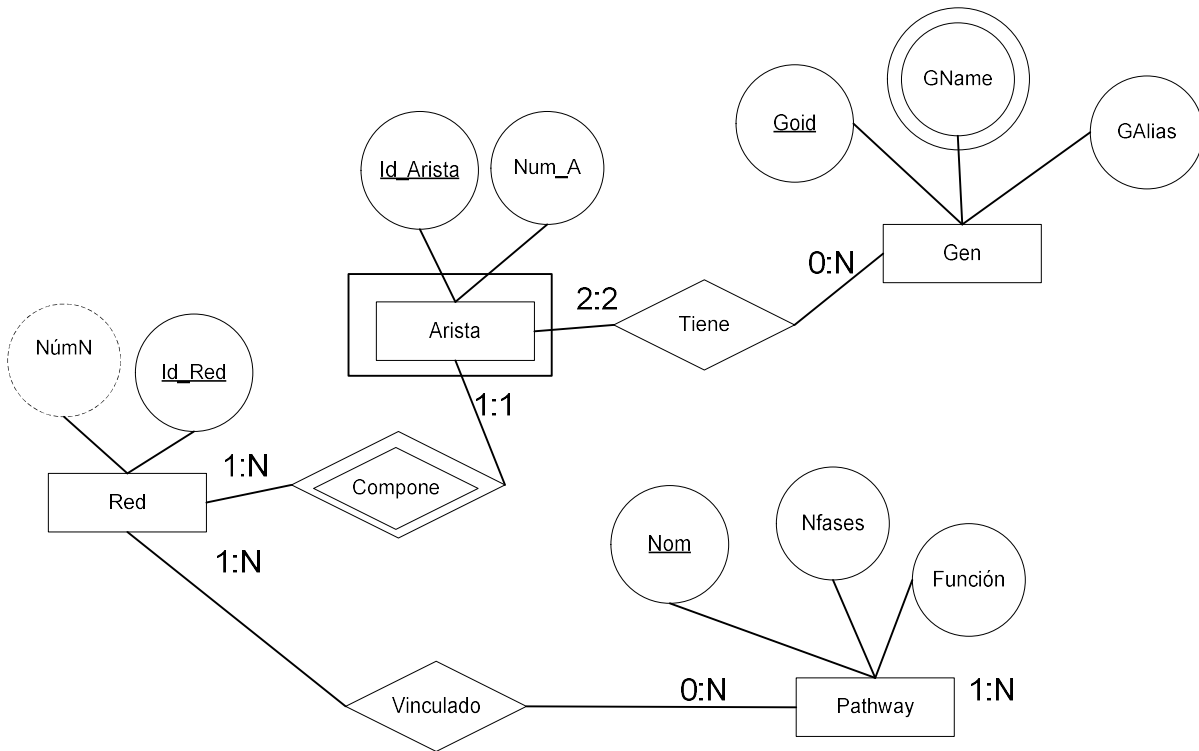


PARTE2





PARTE3





MODELO RELACIONAL

