

INSTITUTO TECNOLÓGICO DE COSTA RICA

ESCUELA DE INGENIERÍA EN COMPUTACIÓN

MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

INTRODUCCIÓN A LA BIOLOGÍA MOLECULAR COMPUTACIONAL

Proyecto #1: Generador de Mapas Cromosómicos Reporte Escrito

Autores:

Olger Calderón Achío
Wilberth Castro Fuentes
Diego Pérez Arroyo

Profesor:

Francisco Torres Rojas, PhD



18 de Abril del 2015

Índice

1. Introducción	2
2. Primera Parte: Acerca de los Mapas Cromosómicos	2
3. Segunda Parte: Aplicación Creada	2
3.1. Detalles técnicos de implementación	2
3.2. Imágenes ilustrativas de la aplicación	3
3.3. Problemas resueltos por nuestra aplicación	3
3.4. Problemas que trascienden el alcance del proyecto	3
4. Conclusiones	3

1. Introducción

El presente documento representa una recopilación de la información más importante concerniente a la implementación del primer proyecto programado del curso. Dicho proyecto se centra en la generación de mapas cromosómicos dado que se poseen algunos datos sobre los genes. Dichos mapas representan gráficamente las posiciones relativas entre genes de un mismo cromosoma, y esto para el biólogo molecular tiene mucha importancia práctica. El proyecto fue implementado en C sobre Linux, más concretamente, Ubuntu.

En la primera sección del documento se analizan conceptos tales como: recombinación de cromosomas, probabilidades de *crossover*, relación entre mapas genéticos y mapas físicos, centimorgans, entre otros. También se adjunta nuestro análisis sobre algunos papers de importancia en el tema. A pesar de que varios investigadores presentaron algoritmos concretos y eficientes para tratar la problemática de generación de mapas, éstos no se ajustaban perfectamente al enunciado del proyecto. Algunos papers presentaban propuestas para encontrar el mejor mapa dado algún criterio previamente escogido, mientras que en nuestro caso, lo que interesa es poder generar todos los mapas posibles. Así que en este caso nos aventuramos a crear un algoritmo original, no sin tomar antes en cuenta información obtenida de la literatura.

En la segunda parte del documento se profundiza un poco más en la aplicación creada. Primero se mencionan los desafíos técnicos y herramientas utilizadas. Luego se exponen varias imágenes ilustrativas, detallando la funcionalidad lograda en la aplicación. Se finaliza con algunos ejemplos de problemas resueltos por nuestra aplicación, así como con una reflexión de aquellos que trascienden este proyecto, pero que por supuesto tienen un interés en la comunidad científica.

2. Primera Parte: Acerca de los Mapas Cromosómicos

2.1. Algunos conceptos importantes

2.2. Unidades de medida

2.3. Mapas genéticos y mapas físicos

2.4. Gregor Mendel

2.5. Literatura consultada

3. Segunda Parte: Aplicación Creada

3.1. Detalles técnicos de implementación

Nos basamos principalmente en las herramientas de desarrollo de GNOME y en la cadena de herramientas de GNU. De estas últimas podemos citar el compilador gcc, la herramienta de construcción make y el GNU Build System (autotools). Por el lado de GNOME, se utilizó el IDE Anjuta, el diseñador de interfaces gráficas Glade, y la colección de componentes gráficos GTK+3.0.

Una pregunta importante de implementación fue: ¿Cómo graficamos los mapas cromosómicos?. Contempla-

mos varias opciones entre ellas: graficar los mapas usando primitivas de Cairo, usando el canvas Clutter, e incluso, OpenGL. Realizar los mapas usando Cairo parecía caro y tal vez no se iba a lograr un resultado sorprendente. Por otra parte Clutter parecía algo complejo y tal vez mucho más de lo necesario. Así que se optó por investigar si existía algún canvas de dibujo específico para bioinformáticos. Para nuestro asombro encontramos Genome Tools (<http://genometools.org/>), que es posiblemente el marco de trabajo más importante para bioinformáticos que desean hacer aplicaciones en C. Dicho proyecto, no es solamente una biblioteca, sino también un conjunto de aplicaciones basadas sobre la misma. Para nuestros propósitos nos interesó usar *Annotation Sketch*, un conjunto de clases y módulos con la capacidad de crear diagramas para representar información genómica. La biblioteca es muy personalizable, y claramente cumplía con nuestros requerimientos. En el fondo *Annotation Sketch* hace llamados a las primitivas de Cairo. Lo único que se debe proveer es un grafo de objetos representando la información de los genes que se desea graficar y la relación entre ellos. La información no está limitada a genes, se puede incluir también: cromosomas, exones, intrones, segmentos de ARN mensajero, entre otros. A pesar de ello, la herramienta estaba más orientada a graficar mapas físicos (en términos de pares de bases), por lo que requirió cierto trabajo el ajustar los gráficos para que pareciesen más mapas de ligamiento de genes. Para una mayor información sobre las ventajas de usar Genome Tools en términos de eficiencia, y para una introducción a sus representaciones se puede consultar [1].

3.2. Imágenes ilustrativas de la aplicación

3.3. Problemas resueltos por nuestra aplicación

3.4. Problemas que trascienden el alcance del proyecto

4. Conclusiones

El proyecto fue sumamente provechoso, desde muchos puntos de vista. Por un lado se recopiló información sobre temas de biología, como es el de los mapas de ligamiento genéticos. Se comprendió su importancia práctica y repercusiones en el campo de la Biología Molecular.

Desde el punto de vista algorítmico tan bien fue muy valioso. Nos dimos cuenta como en efecto, muchos problemas de la naturaleza pueden modelarse en términos de árboles y grafos. A veces resulta sorprendente como conceptos como árboles de expansión minimal y componentes fuertemente conexos representan de forma muy aproximada algún concepto del problema que se trata. Se tuvieron discusiones y disyuntivas sobre cuál era la mejor forma de atacar el problema. Como se mencionó en la introducción del documento, al final se optó por la creatividad del grupo para la implementación del algoritmo, aunque esto representara cambiar ligeramente la forma de modelar y tratar el problema. Toda esta experiencia estimula el analizar los algoritmos más detenidamente.

Por el lado de programación, nos adentramos de nuevo al mundo de programación en C, recordando así el potencial del lenguaje. Al encontrar una biblioteca como Genome Tools, nos dimos cuenta de que existe una comunidad que realiza aplicaciones para bioinformáticos. Ellos siguen estándares y apuestan en C como en el lenguaje apropiado para la implementación. No está de más conocer la existencia de tan útil proyecto. Por otra parte se sacó provecho de las herramientas libres de GNOME, donde GTK+ resalta con un kit muy completo de componentes gráficos.

Referencias

- [1] Gordon Gremme, Sascha Steinbiss, and Stefan Kurtz. Genometools: a comprehensive software library for efficient processing of structured genome annotations. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, 10(3):645–656, 2013.