

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

<b>1. Introducción</b>	<b>2</b>
<b>2. Primera Parte: Acerca de los Mapas Cromosómicos</b>	<b>2</b>
2.1. Algunos conceptos importantes . . . . .	2
2.2. Unidades de medida . . . . .	4
2.3. Mapas genéticos y mapas físicos . . . . .	5
2.4. ¿Por qué Gregor Mendel no reportó el fenómeno de recombinación? . . . . .	5
2.5. Literatura consultada . . . . .	5
<b>3. Segunda Parte: Aplicación Creada</b>	<b>6</b>
3.1. Detalles técnicos de implementación . . . . .	6
3.2. Imágenes ilustrativas de la aplicación . . . . .	7
3.3. Algunos problemas resueltos por nuestra aplicación . . . . .	13
3.3.1. Ejercicio #1 . . . . .	13
3.3.2. Ejercicio #2 . . . . .	14
3.4. Problemas que trascienden el alcance del proyecto . . . . .	16
<b>4. Conclusiones</b>	<b>16</b>

# 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 (o mapas de ligamiento genético) 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

Conviene repasar un poco la importancia de los cromosomas, y como el confirmar sus papeles en la herencia causó cierta revolución sobre lo que se pensaba que era un “gen”. Recopilamos cierta información de [9] y [7], pero lo que se expone a continuación es nuestra interpretación de los conceptos.

Al principio el concepto de “gen” era sumamente abstracto. No se tenía clara su forma, su ubicación, incluso su forma de operar. ¿Había alguna relación entre los distintos genes? ¿O eran totalmente independientes? Con la reivindicación de Gregor Mendel se le dio mucha preponderancia a sus trabajos en la comunidad de biólogos. Mendel afirmaba que los genes actuaban de forma independiente y se mezclaban en las nuevas generaciones de una manera uniforme. Sin embargo en la década de 1910, Thomas Hunt Morgan corroboró la todavía controversial teoría cromosómica Boveri-Sutton. Así quedó claro que los cromosomas controlaban la herencia, y si existía algún lugar “real” donde ubicar los genes, era dentro de los “cuerpos de colores”. Todo esto seguía siendo abstracto, pero abrió la puerta a nuevos marcos del pensamiento.

Con sus trabajos con la mosca de la fruta y los cromosomas sexuales, Morgan también dedujo al observar las proporciones con la que aparecían ciertos fenotipos, que éstos no seguían las leyes Mendelianas al pie de la letra. No parecía que hubiese una independencia real entre algunos grupos de genes. Con esto, Morgan se aventuró a plantear que algunos genes posiblemente estaban ligados. Los genes ligados no seguían los patrones de distribución de las leyes mendelianas, ya que parecían heredarse juntos con más frecuencia. Vienen en paquete se podría decir.

Hoy día sabemos que este fenómeno tiene que ver con la recombinación de cromosomas (o *crossover* en inglés). Durante la meiosis, porciones de material genético de la madre se mezclan con el material genético

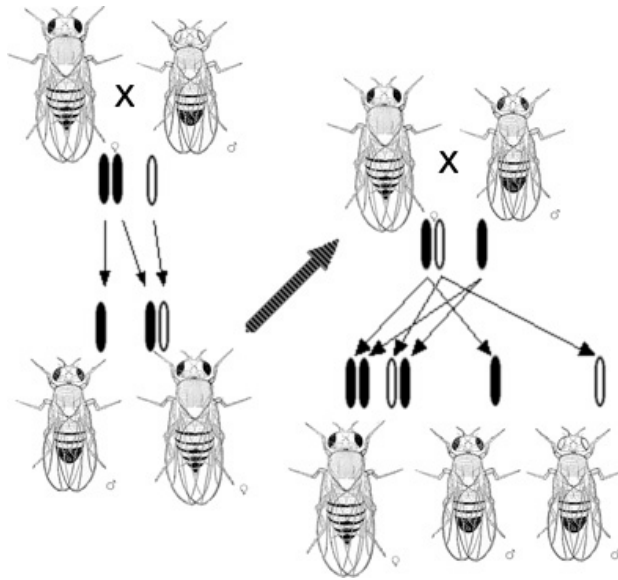


Figura 1: Experimentos de Morgan con la mosca de la fruta. Fuente de imagen: [5]

del padre, formando cromosomas híbridos. Es una forma de intercambiar material genético entre los seres vivos y de “sacudirlo” un poco. Todo esto juega un papel importante en la evolución.

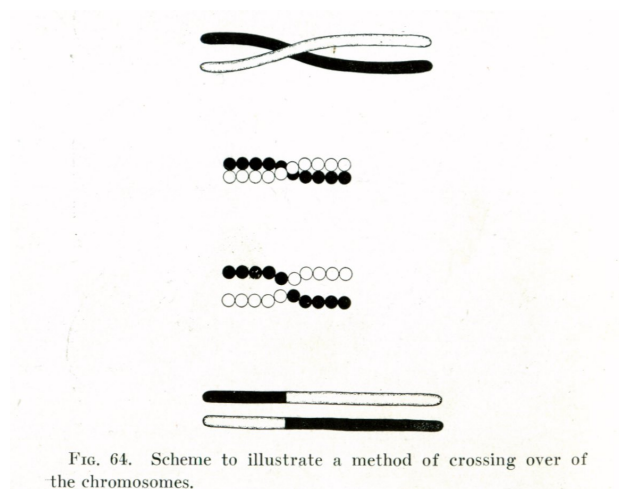


Figura 2: “Crossover simple”. Fuente de imagen: [4]

A muy alto nivel el proceso es como sigue. La meiosis es un tipo de división celular que da paso a la creación de 4 células sexuales o gametos. Cada uno de estos gametos sólo posee una copia de cada cromosoma a diferencia de la mayoría de otras células del cuerpo que poseen dos. En un proceso de meiosis, después de haber ocurrido dos juegos de divisiones, quedan 4 versiones de cada cromosoma que son distribuidas en un gameto cada una. Un cromosoma es idéntico al del padre del individuo, otro es idéntico al de la madre, y los otros dos usualmente se recombinan. En términos muy simplistas, el recombinar dos cromosomas es cortarlos en uno o más puntos, y luego intercambiar segmentos homólogos entre los mismos. Nótese que si todo sale bien, ambos cromosomas recombinados estarán completos después del proceso.

Es un buen momento para definir el concepto de probabilidad de “crossover” que es vital para este proyecto. La probabilidad de recombinación entre dos genes es básicamente, la probabilidad de que al escoger un gameto

Gametes produced by a dihybrid heterozygous individual with linked loci			
Condition	Meiotic tetrad	Gametes	Combination
No crossover	$\begin{array}{cc} a^+ & b^+ \\ a^+ & b^+ \\ a & b \\ a & b \end{array}$	$\begin{array}{cc} a^+ & b^+ \\ a^+ & b^+ \\ a & b \\ a & b \end{array}$	Parental P P P
Single crossover: markers in <i>cis</i> $a^+b^+//ab$	$\begin{array}{cc} a^+ & b^+ \\ a^+ & b^+ \\ a & \updownarrow b \\ a & b \end{array}$	$\begin{array}{cc} a^+ & b^+ \\ a^+ & b \\ a & b^+ \\ a & b \end{array}$	P Recombinant Recombinant P

Figura 3: Resultados de meiosis. Fuente de imagen: [7]

cualquiera, cada gen provenga de un progenitor distinto. En otras palabras y de una forma equivalente, es la probabilidad de que los genes que formaban parte de un mismo cromosoma, sean separados durante la meiosis.

Sea  $P$  una probabilidad de recombinación entre dos genes  $G1$  y  $G2$ , se cumple siempre que  $0 \leq P \leq 0,5$ . Esto es sencillo de explicar. Hay dos situaciones posibles. Que los genes sean parte del mismo grupo de ligamiento (en cuyo caso es muy posible que estén en el mismo cromosoma), o bien, que estén en diferentes grupos de ligamiento (diferentes cromosomas). En el último caso, es cuando se deben cumplir las predicciones de Mendel ( $P = 0,5$ ), ya que realmente hay una independencia en la ubicación física de los genes. En caso de que los genes estén ligados, la probabilidad de que sean separados es menor que  $0,5$ .

Lo interesante de todo lo anterior es que la probabilidad de recombinación es directamente proporcional a la distancia física entre los genes. Si están más cerca, costará más que haya un corte que los separe. Por eso entre menor sea la probabilidad de “crossover” entre dos genes, estos tenderán más a viajar juntos de una generación a otra. A la posición que ocupa un gen en un cromosoma se le llama loci. Si dos genes están tan cerca que cuesta demasiado que se separen se podría decir que están en el mismo loci.

La magia del trabajo de Morgan fue, con base en observaciones de los fenotipos, intuir cuáles eran las probabilidades de “crossover” entre los genes. Sin saber nada acerca de la composición química de los genes, sus predicciones fueron de lo más certeras. Uno de sus aprendices, Alfred Henry Sturtevant estudió el primer mapa de ligamiento del cromosoma X de *Drosophila melanogaster* [8]. Estos mapas son una representación gráfica de la distancia relativa que hay entre los genes, distancias que como se dijo anteriormente son proporcionales a las probabilidades de recombinación entre los genes. La generación de estos mapas son el foco de este proyecto.

Aplicaciones prácticas de crear estos mapas pueden ser: tratar de predecir la posición de un gen responsable de alguna enfermedad, asistir en las mejoras genéticas, descubrir asociaciones entre genes y asistir en la clonación de genes.

## 2.2. Unidades de medida

En honor al trabajo de Morgan, la principal unidad de medida utilizada en estos mapas es el cM (centimorgan) = 1 unidad de mapa [9]. Un centimorgan equivale a aproximadamente una fracción de recombinación del 1 % o  $0,01$ . En el presente proyecto, los diagramas de los mapas hacen uso de esta medida para representar

que tan ligados están los genes. Un 1 cM corresponde a aproximadamente un millón de pares de bases en los seres humanos [2]. Este número puede variar de especie a especie.

### 2.3. Mapas genéticos y mapas físicos

Los cM, y en general, los mapas de ligamiento genéticos no son utilizados para medir distancias físicas. Para eso existen los mapas físicos [6].

Los primeros mapas genéticos fueron generados experimentando con generaciones sucesivas de organismos y analizando los porcentajes de segregación de ciertas características. En un mapa de ligamiento interesa representar el orden y posiciones de los genes en el cromosoma. Sin embargo no son tan útiles para medir las distancias físicas (pares de bases), ya que si dos genes están demasiado cerca, ocuparán el mismo loci en el mapa.

Los mapas físicos reflejan la distancias reales, y operan a un nivel mayor de detalle. Usualmente se usan otras técnicas distintas para construirlos. Finalmente otro recurso que opera a un nivel mayor de detalle sería la secuenciación [6].

### 2.4. ¿Por qué Gregor Mendel no reportó el fenómeno de recombinación?

Cabe preguntarse ¿porqué Gregor Mendel, el padre la genética, no descubrió lo que Morgan y Sturtevant relatan en sus papers muchos años después [3] y [8]? Mendel en sus reportes concluyó que las características heredadas son independientes unas de las otras, sin ninguna relación entre sí. Lo cual sabemos que no siempre es cierto.

Cuando los genes están ligados usualmente se expresan juntos (tienden a actuar todos como un sólo gen) sin embargo, durante la meiosis puede haber recombinación, lo cual resulta en que genes ligados se separen en un individuo particular. Por lo general un grupo de ligamiento se encuentra en un cromosoma [10][p3] y cuando los genes están ubicados en diferentes cromosomas se heredan independientemente uno del otro. La razón por la cual Mendel no reportó el fenómeno de recombinación es porque las características seleccionadas para su análisis estaban en diferentes grupos de ligamiento, por ello siempre se iban a expresar independientemente.

### 2.5. Literatura consultada

El paper de Sturtevant [8] y el de Morgan [3], más el primero que el segundo, nos sientan las bases de los mapas cromosómicos de genes ligados. Nos explica cómo se llegó a la idea de que la probabilidad de crossover estaba relacionada con la distancia de los genes en los cromosomas. Morgan define conceptos que son fundamentales tener claros; estos dos papers son la base que había que tener clara antes de si quiera tratar de pensar en la solución al problema. Sin embargo, una vez que se han leído y entendido, para los efectos del proyecto, no proporcionan mucha ayuda. Reafirma la teoría vista en clase. No trata temas como probabilidades faltantes, mapeo de genes no ligados y otros detalles.

El paper de más ayuda fue el llamado *Efficient and accurate construction of genetic linkage maps from noisy and missing genotyping data* [10]. Una vez leído el título pensamos que iba a resolver todas nuestras dudas, porque parece ser justo el mismo tema del proyecto: construcción eficiente de mapas de ligamiento, manejo de datos faltantes. Empero, en realidad el problema que trata es distinto. El objetivo de los autores es construir el mejor mapa de ligamiento posible de forma eficiente. Parten de una población de individuos, en los cuales hay genes que se expresan de una u otra forma. Estiman la probabilidad de crossover entre

genes al determinar el número de veces que un gen se expresa de la misma forma en todos los individuos.

Una vez determinadas las probabilidades de crossover entre los genes entonces pareciera encontrarse en el punto inicial de nuestro proyecto. El siguiente paso del paper es hacer clustering para determinar los grupos de ligamiento de genes. Es importante destacar en este punto que gracias a lo leído nos dimos cuenta que la probabilidad de crossover entre 2 genes ligados está entre  $[0 - 0.5]$  esto porque si la probabilidad es de 0.5 significa que los genes se expresan independientemente y por lo general estarían ubicados en diferentes cromosomas. Del paper entendimos que básicamente se determinan los grupos de genes ligados al realizar un recorrido en profundidad del grafo (matriz inicial del proyecto) ignorando las aristas con valor a 0.5 o muy cercanas a dicho valor. El resultado son componentes conexos del grafo original, cada uno de ellos es un grupo de ligamiento.

Es en este punto en que el paper nos deja de ser útil. Los siguientes pasos se centran en determinar el orden óptimo de los genes para que el mapa sea de menor tamaño. Varios métodos son presentados e incluso uno propio por parte de los autores. Mas, para nuestros efectos no eran de ayuda pues el problema del proyecto es generar todos los posibles mapas no sólo el óptimo. De igual forma cuando se trataba sobre datos faltantes el paper no nos fue de utilidad porque de nuevo el enfoque de los autores difiere del nuestro.

El paper llamado *Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph* [11] es el mismo que el anterior sólo que en una edición un año después por parte de los mismos autores. No nos aporta nada que no lo haya hecho su predecesor.

## 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?. Contemplamos 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

Al abrir la aplicación, por omisión muestra una tabla con una configuración para tres genes, como se muestra en la figura 4.

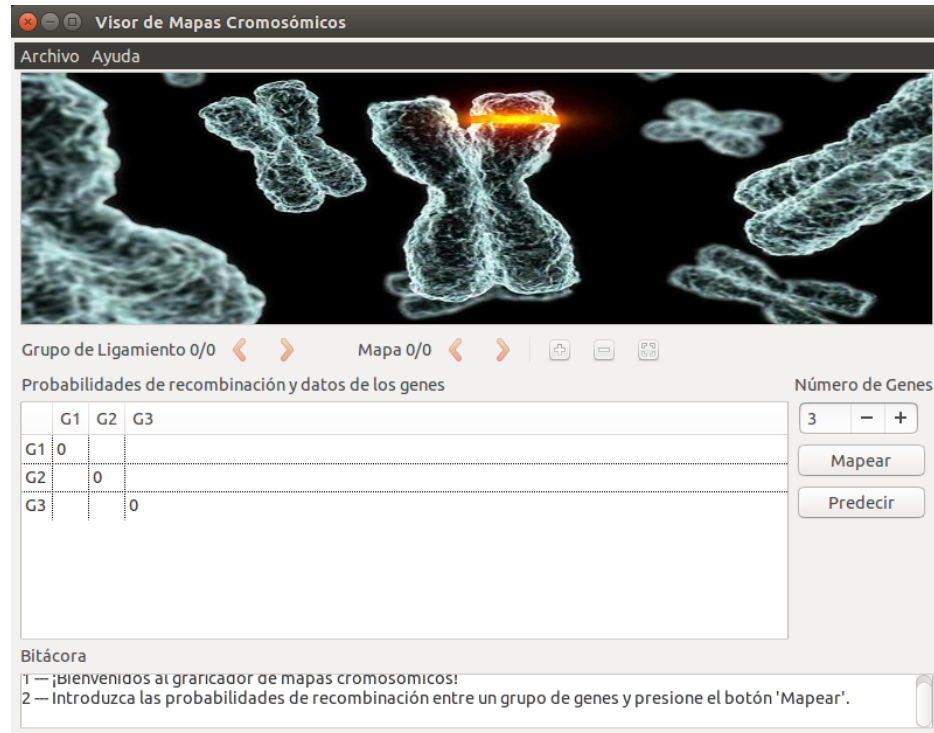


Figura 4: Configuración por omisión

Aunque es posible incluir más genes en el modelo, esto se logra por medio del control que se encuentra sobre el botón "Mapear", es posible incrementar o decrementar en 1 el número por medio del control o digitar directamente en la casilla con el teclado. El rango valido es  $[2, 20]$ .

Para iniciar con el proceso es necesario introducir las probabilidades de recombinación entre al menos un par de genes. Mientras más completa esté la tabla, más específico será el resultado.

Se permiten probabilidades dentro de un rango de  $[0, 0.5]$  y al editar un valor de la tabla, haciendo click sobre él, su homólogo es actualizado automáticamente, por ejemplo, si se edita la primer celda vacía de la segunda fila, el nuevo valor aparecerá también en la primer celda vacía de la primera fila.

También es posible editar el nombre de los genes, pero solamente haciendo click sobre su nombre en la fila que los representa, al hacerlo, su nombre en la columna correspondiente es actualizado automáticamente, como se ve en la figura 8.



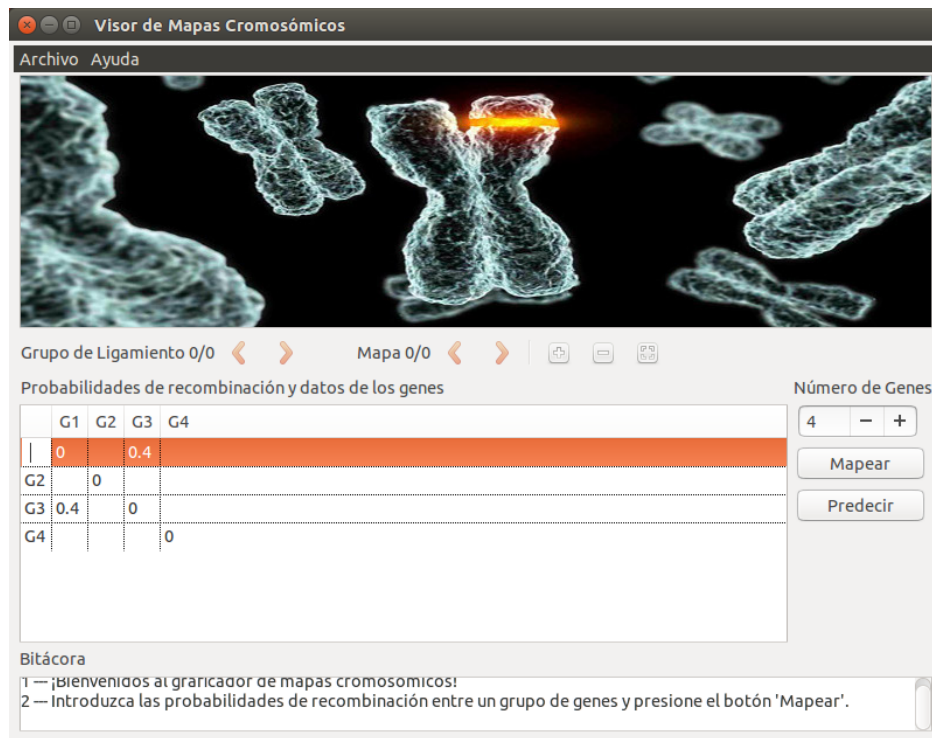


Figura 5: Tabla con 4 genes, un valor y nombre del G1 en edición

A lo largo del procedimiento, la aplicación mostrará notificaciones en el panel de texto inferior.

A continuación se muestra con un ejemplo (los nombres de genes y probabilidades de recombinación se han tomado del ejercicio #2), el uso de la aplicación.

Después de establecer el modelo a partir del cual se desea graficar, se debe presionar el botón "Mapear", para el modelo correspondiente al ejercicio #2 se obtendría una tabla y un primer mapa (visible al mapear) como los que se ven en la figura 6.

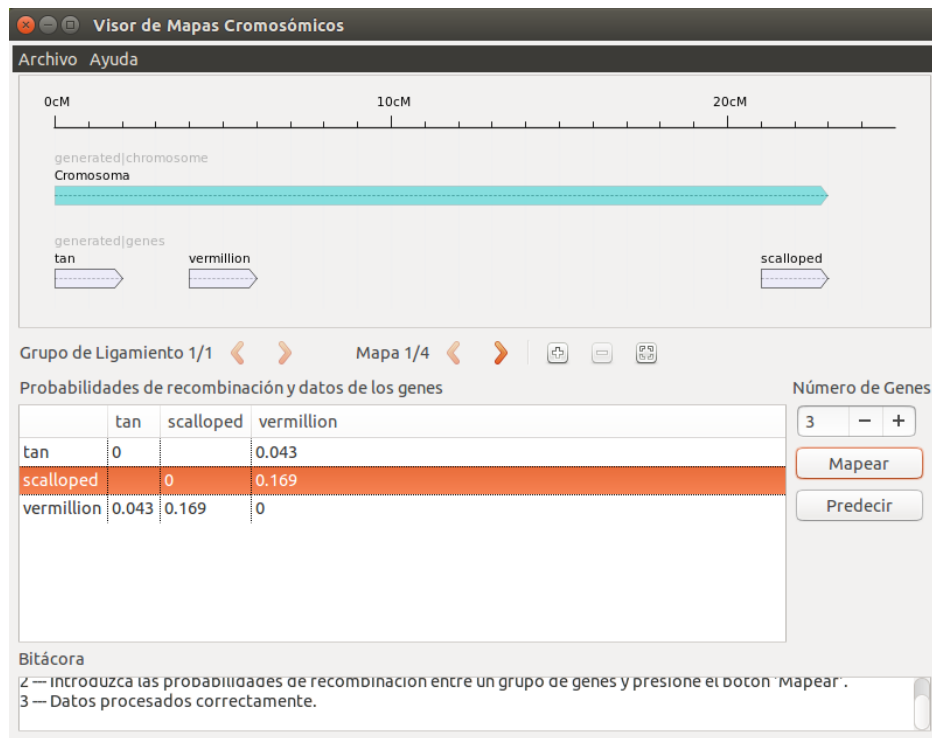


Figura 6: Primer mapa generado a partir de la tabla para el ejercicio #2

Para ese ejercicio solo se obtiene un grupo de ligamiento, aunque en caso de obtener más, sus posibles mapas se visualizarían navegando con las flechas al lado de la etiqueta correspondiente.

Al elegir el grupo de ligamiento de interés, se puede navegar entre sus posibles mapas usando el otro par de flechas. Para el ejercicio #2 tenemos cuatro mapas (dos más sus inversos), el primero es el que vemos en la figura 6 y el segundo en la figura 7.

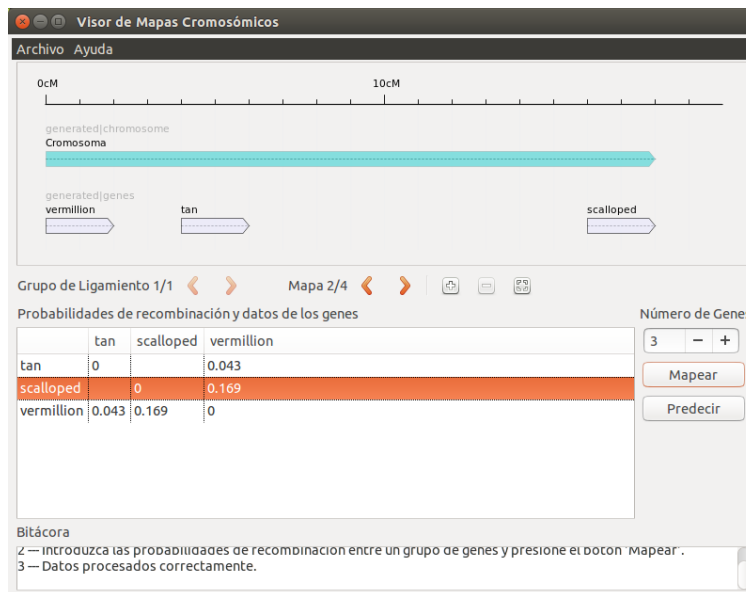
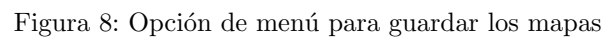


Figura 7: Segundo mapa generado a partir de la tabla para el ejercicio #2

En caso de desear guardar los mapas, esto se puede hacer yendo a la barra de menú superior, en el menú “Archivo” se encuentra la opción de “Guardar”, al elegirla aparece una ventana de diálogo para seleccionar la ubicación de los mapas, así como se muestra en las figuras 8 y 9. El formato del archivo será CSV.



De forma similar es posible abrir los mapas guardados, esto se hace yendo a la barra de menú superior, en el menú “Archivo”, la opción “Abrir”, al elegirla aparece una ventana de diálogo para seleccionar la ubicación de los mapas, así como se muestra en la figura 10.

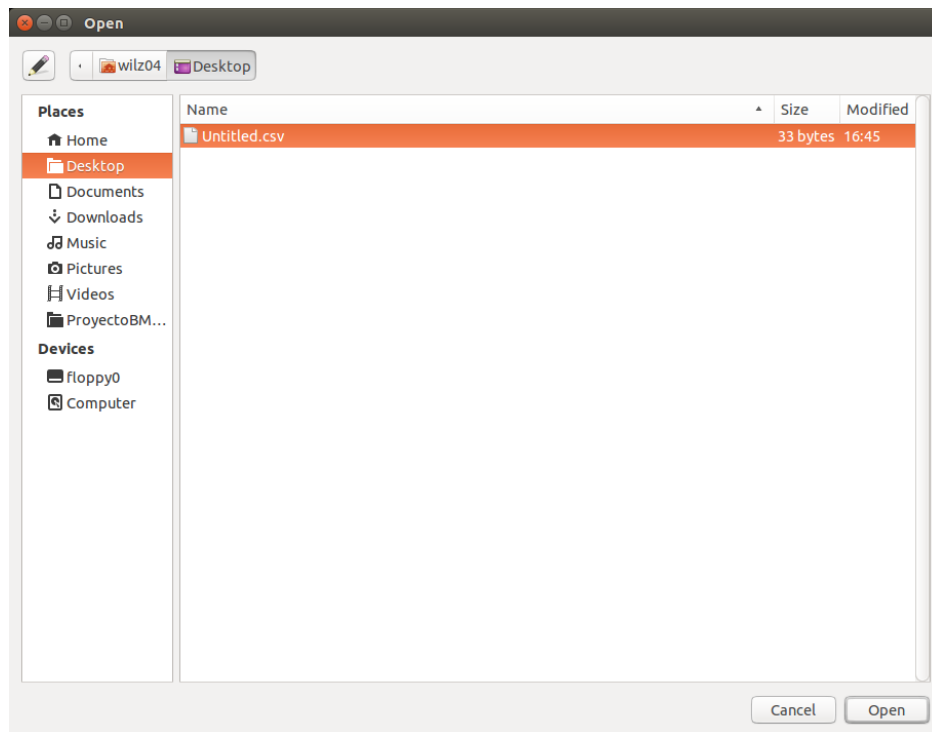


Figura 10: Ventana de diálogo para abrir los mapas

O bien, también es posible reestablecer el modelo a la configuración por defecto utilizando la opción “Nuevo” que se ve en la figura 5.

La aplicación cuenta también con información de autoría, se encuentra en el menú de ayuda, la información es la que se muestra en las figuras 11 y 12.



Figura 11: Ventana de información de la aplicación



Figura 12: Ventana de información de los autores

Al finalizar la sesión de trabajo, es posible salir de la aplicación presionando la opción “Salir” (que se ve en la figura 8) o la -x- en la barra de título.

### 3.3. Algunos problemas resueltos por nuestra aplicación

#### 3.3.1. Ejercicio #1

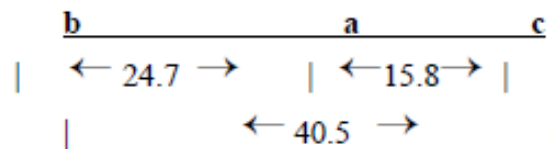
Éste es un problema super sencillo encontrado en [7]. Se tienen 3 genes,  $a$ ,  $b$  y  $c$ .

$$(b \leftrightarrow a) \text{ Map distance} = \frac{112 + 102 + 18 + 15}{1000} \times 100 = 24.7\% = 24.7 \text{ m.u.}$$

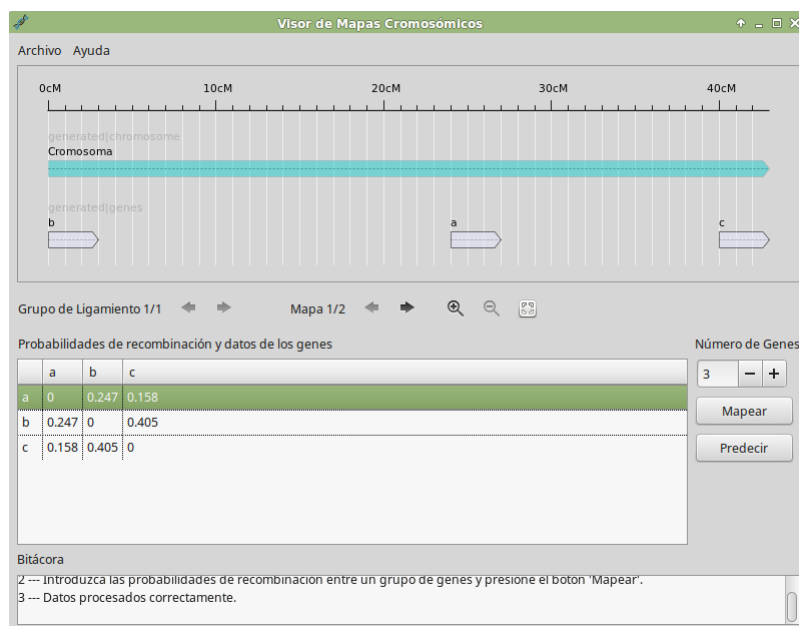
$$(a \leftrightarrow c) \text{ Map distance} = \frac{66 + 59 + 18 + 15}{1000} \times 100 = 15.8\% = 15.8 \text{ m.u.}$$

$$(b \leftrightarrow c) \text{ Map distance} = 24.7 \text{ m.u.} + 15.8 \text{ m.u.} = 40.5 \text{ m.u.}$$

Ésta es la respuesta esperada:



Y éste es el mapa generado por nuestra aplicación:



### 3.3.2. Ejercicio #2

Acá tenemos un ejercicio un poco más complicado del mismo libro. Éste requiere calcular las probabilidades de recombinación a partir de datos observados. Se tiene el siguiente enunciado:

**Tan body, scalloped wing female *Drosophila* were crossed with vermilion eye males.**

vi) The testcross result showed

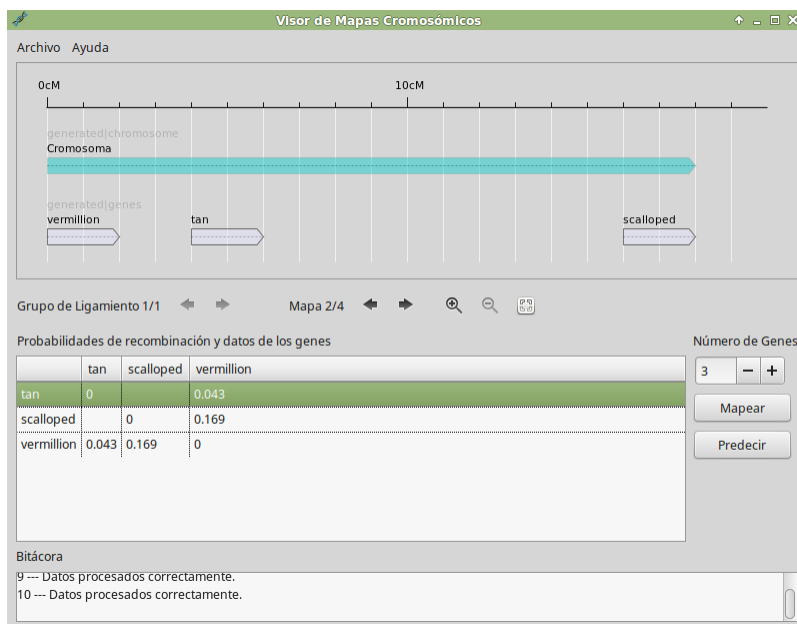
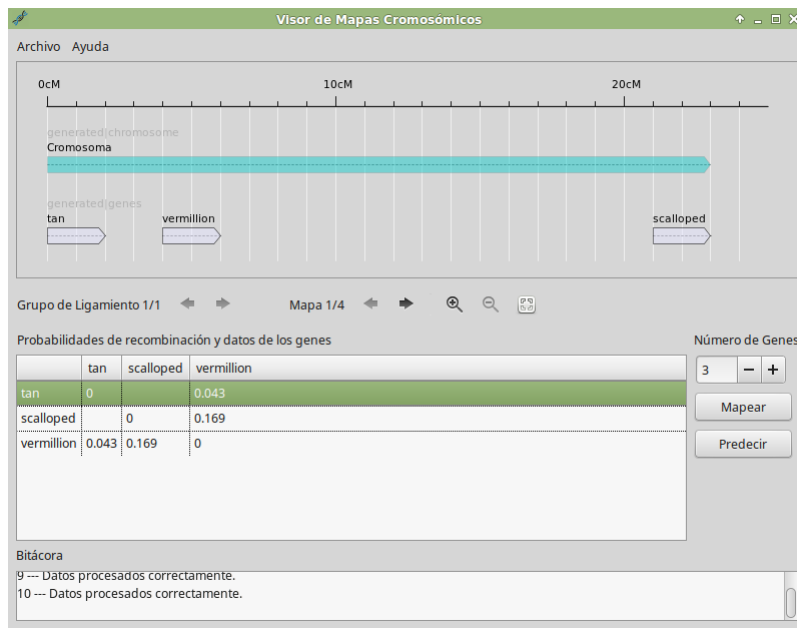
413 tan scalloped	84 tan
380 vermilion	80 vermilion scalloped
18 tan vermilion	2 tan vermilion scalloped
20 scalloped	3 wild-type

**Determine the map distance between each pair of loci.**

Así podemos decir que los machos tienen el siguiente genoma: no-tan no-scalloped vermilion. Y las hembras el siguiente: tan scalloped no-vermilion. Son en total mil datos.

Las dos clases del final (tan scalloped vermilion y wild-type) corresponden a casos de doble “crossover”. Esto se puede deducir porque son las que presentan un menor número de ocurrencias [7]. De acá mismo podemos deducir que el gen vermilion es el que está fuera de lugar si se compara un tipo de doble “crossover” contra un tipo parental. Por ejemplo: si se compara (tan scalloped), tipo parental, contra (tan scalloped vermilion), tipo de doble “crossover”, el gen que no hace “match” es vermilion. Un razonamiento similar aplica para cualquier pareja de clase parental y doble “crossover”. Y ésto nos dice que precisamente el gen vermilion está entre el de tan y el de scalloped. Por lo tanto vamos a calcular la probabilidad de “crossover” de tan-vermilion y vermilion-scalloped.

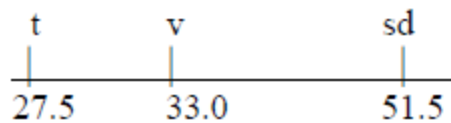
Primero entre tan-vermilion:  $(18 + 20 + 2 + 3)/1000 = 0,043$ . Y ahora entre vermilion-scalloped:  $(84 + 80 + 2 + 3)/1000 = 0,169$ . Introducimos los datos en nuestra aplicación, y nos topamos con el siguiente resultado:



La aplicación nos genera dos pares de mapas equivalentes. Gracias a nuestro conocimiento de que vermillion es el gen del medio podemos decir que el primer mapa es el correcto. Y nótese que si calculamos la probabilidad de “crossover” de tan-scalloped, nos topamos que:  $(18 + 20 + 84 + 80)/1000 = 0,202$ , lo cuál se acerca a  $0,043 + 0,169 = 0,212$ . Ésto soporta la elección de nuestro mapa.

Resultados esperados:





### 3.4. Problemas que trascienden el alcance del proyecto

Nuestro proyecto hace una serie de suposiciones que simplifica un poco su alcance. Por ejemplo no aplicamos ciertas funciones de transformación. Cuando hay más de un “crossover” en el mismo evento meiótico, puede que haya que transformar las probabilidades de recombinación a una cantidad de cM distinta. Funciones que realizan estas transformaciones son las de Haldane y Kosambi. Por ejemplo la función de Haldane sobre una probabilidad de 0,27, retorna 39 cM, mientras que la de Kosambi 30cM [9].

El tema se puede tornar tan complicado como se quiera, y hay aspectos que se vuelven más difíciles de modelar. Entre algunos de estos aspectos se puede mencionar las probabilidades de mutación y desequilibrio de ligamiento. El paquete LINKAGE contempla mucho de estos problemas y más [9].

## 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 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.
- [2] Harvey Lodish. *Molecular cell biology*. Macmillan, 2008.
- [3] Thomas H Morgan. Sex limited inheritance in drosophila. *Science, New Series*, 32(812):120–122, 1910.
- [4] Thomas Hunt Morgan. *A Critique of the Theory of Evolution*. Princeton University Press, 1916.
- [5] Thomas Hunt Morgan. *The physical basis of heredity*. JB Lippincott, 1919.
- [6] Joao Carlos Setubal and Joao Meidanis. *Introduction to computational molecular biology*. PWS Pub., 1997.
- [7] Carr Steven M and Innes David J. *Biology 2250 Principles of Genetics*. Memorial, University of Newfoundland, 2003.
- [8] Alfred Henry Sturtevant. The linear arrangement of six sex-linked factors in drosophila, as shown by their mode of association. *Journal of experimental zoology*, 14(1):43–59, 1913.
- [9] Joseph Douglas Terwilliger. *Handbook of human genetic linkage*. JHU Press, 1994.
- [10] Yonghui Wu, Prasanna Bhat, Timothy J Close, and Stefano Lonardi. Efficient and accurate construction of genetic linkage maps from noisy and missing genotyping data. In *Algorithms in Bioinformatics*, pages 395–406. Springer, 2007.
- [11] Yonghui Wu, Prasanna R Bhat, Timothy J Close, and Stefano Lonardi. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS genetics*, 4(10):e1000212, 2008.