Capítulo 3

SOLUCIÓN DEL MODELO MATEMÁTICO

En este capítulo se describirá el procedimiento que se siguió para resolver el modelo matemático planteado en el capítulo 2. En la primera sección se discutirán varios aspectos de la solución inicial y se justificará el método empleado para obtener una lista inicial de preguntas. Posteriormente, en la sección 3.2 se desarrollará el algoritmo para obtener las preguntas de la primera etapa. A partir de estas preguntas y a través de un programa computacional, construiremos una matriz de bialelos reducida, en la tercera sección del presente capítulo. El código del programa computacional y el diagrama de flujo que lo explica aparecen también en la sección 3.3. Después, en la última sección, presentaremos el método para encontrar la segunda etapa de preguntas.

3.1 SOLUCIÓN INICIAL

Trabajaremos en esta sección con la matriz de bialelos construida en el capítulo anterior durante el desarrollo del modelo matemático. Esta matriz absorbe las características principales de nuestro problema. Toma en cuenta la estructura de los datos como nos fueron presentados originalmente, representa en cada columna la forma en que se realizan las preguntas al código genético y en cada entrada la manera en que se reciben las respuestas utilizando 0’s y 1’s, lo que equivale a *No* y *Sí* respectivamente. Además, sus líneas representan pares de alelos (bialelos) en la manera en que podemos identificarlos mediante preguntas, es decir, a través del análisis molecular utilizando reactivos en el laboratorio. Esto nos permite, en un primer acercamiento, concentrarnos al estudio de esta matriz para obtener una solución al problema que nos concierne.

El problema principal que encontramos al buscar una solución inicial es la pérdida de información. Ésta tiene su origen en la forma simultánea en que se hacen las preguntas a dos alelos concatenados de un ser humano, donde la respuesta es positiva si lo es para alguno de los alelos (sin saber cuál) o si lo es para ambos. Como se verá enseguida, esto nos creará situaciones ambiguas que no podrán ser esclarecidas en algunos casos.

Lo primero que hay que tomar en cuenta es que el problema no necesariamente tiene solución. Es posible, y de hecho es el caso, que aún haciendo todas las preguntas posibles en todas las posiciones de nuestro arreglo de bialelos, las respuestas arrojadas no nos permitan diferenciar ciertos bialelos de otros. En algunos casos afortunados, estas ambigüedades no representan un problema, pues los bialelos que se confunden entre sí pertenecen a la misma biclase. Y es la biclase lo que realmente estamos buscando, no un bialelo particular. Sin embargo, esto no sucede siempre que encontramos un bialelo idéntico a otro. En ocasiones dos bialelos idénticos pertenecen a biclases distintas y será imposible diferenciarlos bajo las condiciones establecidas en esta tesis. Como se dijo, aún haciendo todas las preguntas posibles en todas las posiciones, no será posible distinguir entre algunos bialelos y otros.

Afortunadamente, los pares de bialelos idénticos que pertenecen a biclases distintas son muy escasos en nuestro conjunto total. De hecho, de los 31626 bialelos que pueden aparecer, sólo 568, el 1.79% aproximadamente, no pueden ser ubicados en su biclase. Es decir, si un posible receptor o donador de órgano cae dentro de este 1.79%, no podremos decir si el bialelo que posee pertenece a una de dos (o en pocos casos tres) biclases determinadas, utilizando los métodos que se describieron en el problema biológico del capítulo 2.

Si bien no analizaremos aquí la forma de clasificar este pequeño porcentaje de bialelos ambiguos por rebasar los alcances de esta tesis, dicha clasificación es posible utilizando otros métodos de análisis molecular [4].

Aún no hemos dicho cómo se encontraron los alelos ambiguos, así que lo explicaremos a continuación. Se construyó la matriz de bialelos de 31626 filas y 265 columnas que se describió en el modelo matemático. Dado que cada fila es una sucesión de 265 0's y 1's, ésta puede interpretarse como un número binario. Transformando cada número binario a la notación decimal tenemos un identificador compacto para cada fila. Sin embargo, el número 2^{265} es demasiado grande para que pueda ser expresado en toda su expansión decimal, por lo que las posiciones fueron separadas en 8 grupos de 30 posiciones y un grupo de 25. Así, las primeras treinta posiciones son un número binario que transformamos a notación decimal y llamamos identificador uno (ID1), las segundas treinta posiciones generan el identificador 2 (ID2), etcétera. Finalmente tenemos nueve identificadores para cada fila (o bialelo) y dos filas son iguales si y sólo si sus nueve identificadores coinciden simultáneamente.

Bialelo	1 0 0 1	0 0 1 0	0 1 0 0	0 1 0 0	0 0 1
Potencia	16 8 4 2	16 8 4 2	16 8 4 2	16 8 4 2	8 4 2
Suma	16 + 0 + 0 + 2	0 + 0 + 4 + 0	0 + 8 + 0 + 0	0 + 8 + 0 + 0	0 + 0 + 2
Identificador	ID1 = 18	ID2 = 4	ID3 = 8	ID4 = 8	ID5 = 2

Tabla 6: Construcción de los identificadores.

Para aclarar esta idea consideremos el ejemplo de la Tabla 6. Ahí aparecen cuatro filas, la primera de las cuales representa un bialelo de 19 posiciones divididas en 4 grupos de cuatro posiciones y un grupo de 3 posiciones.. La segunda presenta la potencia de 2 por la cual será multiplicada cada posición del alelo. En la tercera observamos una expresión, cuyos sumandos son el resultado de multiplicar cada posición del alelo por la potencia de 2 que le corresponde. Y en la parte inferior, se encuentran los 5 identificadores así construidos.

Notemos que las potencias de 2 no empiezan con $2^0 = 1$. Esto se hizo para que la potencia mayor coincida con el número de posiciones en cada grupo. O sea que, en realidad, estamos pensando que cada número binario termina con un 0 adicional a la derecha, lo cual no afecta la unicidad de los identificadores. De manera análoga se construyeron los identificadores de la matriz original de bialelos, sólo que con los grupos que se describieron arriba.

En una segunda etapa, se ordenaron los identificadores de manera ascendente para poder compararlos más fácilmente y encontrar las ambigüedades. Posteriormente, se descartaron los pares de bialelos idénticos pertenecientes a una misma biclase y el resto pasaron a formar el grupo del 1.79% que no podrán ser clasificados en esta tesis.

En resumen, dadas las características del análisis al código genético humano que estamos suponiendo en este trabajo, no nos será posible clasificar a la totalidad de los bialelos en sus biclases, debido al 1.79% de bialelos no clasificables. De modo que concentraremos nuestro estudio, en adelante, solamente al subconjunto de bialelos que pueden ser encasillados en una biclase, que forma el 98.21% del conjunto total. Y tomaremos como una solución a nuestro problema a la lista de preguntas que clasifique adecuadamente este subconjunto. Además, debido a que no podremos clasificar todos los bialelos, aún si hacemos todas las preguntas posibles en todas las posiciones, tomaremos como solución inicial a esta lista de preguntas. Dicho de otra manera, nuestra solución inicial consta de las 265 preguntas que aparecen representadas en cada columna de la matriz de bialelos generada a partir de todos los datos.

El método para mejorar la solución inicial será desarrollado a lo largo de las secciones restantes de este capítulo. En primer lugar, se obtendrá una lista de preguntas que denominaremos la *primera etapa*. Y posteriormente encontraremos varias listas de preguntas que formarán la *segunda etapa*. Así, dependiendo de las respuestas obtenidas en la primera etapa, tendremos que decidir si es necesario aplicar alguna o algunas de las listas de preguntas de la segunda etapa.

3.2 OBTENCIÓN DE PREGUNTAS DE LA PRIMERA ETAPA

El objetivo de esta sección es encontrar una lista de preguntas más reducida que la considerada como solución inicial, que clasifique un bialelo en la biclase que le corresponde. Para cumplir con este objetivo, volveremos a trabajar con la matriz original de datos (alelos HLA-I-A), pues la cantidad de información que nos provee es más manejable.

Empezaremos por construir un árbol de decisión que nos arrojará una lista de preguntas, la cual nos permitirá clasificar alelos en su clase correspondiente. De esta manera obtenemos un conjunto de posiciones útiles para la clasificación de alelos en clases. Después, a partir de este conjunto de preguntas, utilizaremos un procedimiento para obtener una lista inicial de preguntas destinadas al conjunto de bialelos, las preguntas de la primera etapa.

3.2.1 Construcción de un árbol de decisión

Elaboraremos un árbol de decisión, escogiendo preguntas que nos dividan los datos en dos partes iguales aproximadamente. Nuestro criterio para elegir las preguntas será que el número de clases se divida más o menos a la mitad con cada pregunta. Luego, cuando alguna rama del árbol agrupe alelos pertenecientes a una sola clase, diremos que hemos alcanzado una *punta* y no haremos más preguntas en esa rama. Una vez que todas las ramas del árbol terminen en una punta, habremos concluido la construcción del árbol.

Notemos que un nodo de nuestro árbol de decisión representa una pregunta que origina dos respuestas (Sí y No), las cuales, a su vez, generan dos nuevas ramas del árbol. Una rama agrupa un conjunto de alelos que responden *Sí* a la pregunta del nodo anterior y la otra agrupa un conjunto que responde *No* a dicha pregunta.

Posición	5	8	17	19	24	25
A*0101	C	C	G	A	A	T
A*0102	C	C	G	A	A	A
A*0103	C	C	G	A	A	C
A*0301	T	A	T	G	T	T
A*0302	C	A	T	G	C	A
A*0303	T	A	C	A	T	T
A*0304	T	T	A	C	T	T
A*290101	C	T	C	T	C	A
A*290102	C	T	G	C	C	A
A*6901	C	C	G	A	T	A
A	0	3	1	5	3	5
G	0	0	6	2	0	0
C	7	4	1	2	3	1
T	3	3	2	1	4	4

Tabla 7: Ejemplo de matriz de alelos HLA.

Consideremos el ejemplo de la Tabla 7 para esclarecer el procedimiento de elaboración de un árbol de decisión. Tenemos un conjunto de 10 alelos clasificados en las clases 01, 03, 29 y 69, cada una de ellas con distinto número de elementos. Ahora debemos analizar los datos ahí presentados para elegir una “buena” pregunta. Como explicábamos al principio de esta subsección, lo primero que debemos tomar en cuenta para escoger una pregunta es que ésta nos divida el número de clases aproximadamente a la mitad. Podríamos hacer la pregunta ¿existe una G en la posición 17? Sin embargo, hay tres clases cuyos alelos arrojan la respuesta Sí y sólo una en donde responden No, lo cual no corresponde con nuestro criterio de separar los datos en partes iguales.

Otra pregunta no conveniente para estos datos sería ¿existe una C en la posición 5?. Los alelos que responden afirmativamente son todos los de las clases 01, 29 y 69, además de un alelo de la clase 03. Lo cual nos deja con un grupo de cuatro clases por un lado y de una (la 03) por el otro.

La pregunta ideal que podemos hacer inicialmente a la información de la Tabla 7 es ¿existe una C en la posición 8? A esta pregunta, los alelos que responden Sí son todos los de las clases 01 y 69, por lo que con ella separamos los datos en dos grupos con dos clases cada uno.

El siguiente paso consiste en separar los datos en dos partes. Por un lado los que responden Sí a la primera pregunta y por otro los que responden No, como podemos observar en la Tabla 8. Tenemos un primer nodo (la primera pregunta) y las dos primeras ramas del árbol de decisión que estamos construyendo en nuestro ejemplo. Ahora debemos elegir preguntas para obtener dos nodos más, uno en cada rama.

¿Existe una C en la posición 8?

Sí						
Posición	5	8	17	19	24	25
A*0101	C	C	G	A	A	T
A*0102	C	C	G	A	A	A
A*0103	C	C	G	A	A	C
A*6901	C	C	G	A	T	A
A	0	0	0	4	3	2
G	0	0	4	0	0	0
C	4	4	0	0	0	1
T	0	0	0	0	1	1

No						
Posición	5	8	17	19	24	25
A*0301	T	A	T	G	T	T
A*0302	C	A	T	G	C	A
A*0303	T	A	C	A	T	T
A*0304	T	T	A	C	T	T
A*290101	C	T	C	T	C	A
A*290102	C	T	G	C	C	A
A	0	3	1	1	0	3
G	0	0	2	2	0	0
C	3	0	1	2	3	0
T	3	3	2	1	3	3

Tabla 8: Datos separados mediante la primera pregunta.

Analicemos la rama que responde Sí a la primera pregunta. Para poder separar las dos clases involucradas podemos hacer la pregunta ¿existe una A en la posición 24? Aunque también podemos elegir la pregunta complementaria ¿existe una T en la posición 24? Supongamos que elegimos la primera. Ésta nos genera un nuevo nodo cuyas ramas son puntas, pues en una sólo hay elementos de la clase 01 y en la otra el único elemento de la clase 69.

Ahora, observemos la rama del primer nodo que responde No. Como podrá notar el lector, no existe una pregunta que separe las dos clases involucradas de manera disjunta. Si hacemos, por ejemplo, la pregunta ¿existe A en la columna 25?, obtendremos en la rama que responde Sí, los dos elementos de la clase 29 y uno de la clase 03, y en la rama que responde No agrupamos tres elementos de la clase 03. Esto nos deja con dos clases en una rama y una clase en la otra. Sin embargo este problema no es eludible por ningún medio, pues no existe una posición en la que podamos hacer una pregunta ideal. Así que estamos obligados a hacer alguna pregunta mixta, como la que consideramos anteriormente. Demos esto por hecho y observemos los resultados obtenidos en la Tabla 9.

¿Existe A en la posición 25?

Sí						
Posición	5	8	17	19	24	25
A*0302	C	A	T	G	C	A
A*290101	C	T	C	T	C	A
A*290102	C	T	G	C	C	A
A	0	1	0	0	0	3
G	0	0	2	1	0	0
C	3	0	0	1	3	0
T	0	2	1	1	0	0

No						
Posición	5	8	17	19	24	25
A*0301	T	A	T	G	T	T
A*0303	T	A	C	A	T	T
A*0304	T	T	A	C	T	T
A	0	2	1	1	0	0
G	0	0	0	1	0	0
C	0	0	1	1	0	0
T	3	1	1	0	3	3

Tabla 9: Tercer nodo del árbol de decisión.

En esta tabla observamos las dos ramas que se obtuvieron al hacer la pregunta que establecimos en el párrafo anterior. Podemos notar que la rama que contesta No ya es una punta. En cambio, la rama que contesta Sí todavía tiene elementos de dos clases, por lo que es necesario agregarle una última pregunta o nodo. Podemos elegir entre varias preguntas para separar las dos clases restantes. Por ejemplo, ¿existe A en la posición 8?, ¿existe T en la posición 17?, entre otras. Elegimos arbitrariamente la segunda y con esto separamos las clases que faltaban.

En la Figura 1 está representado el árbol de decisión que se obtuvo a partir de nuestro pequeño ejemplo. En cada nodo aparece la pregunta que lo generó y las ramas están marcadas con la respuesta que dan a la pregunta del nodo. Las puntas están sombreadas y llevan el número de la clase que agrupan.

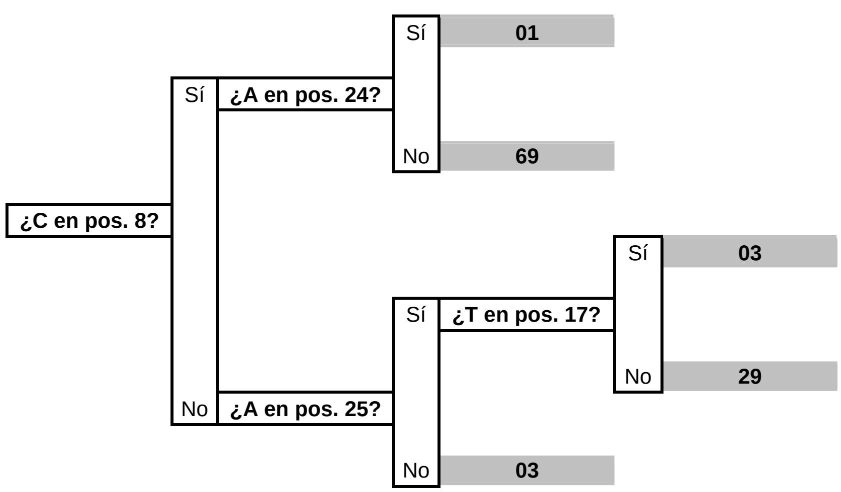


Figura 1: Ejemplo de árbol de decisión.

Cabe mencionar que la situación en la que nos vimos obligados a realizar una pregunta mixta, que no separa clases enteras en una rama y en otra, sucede con frecuencia cuando elaboramos el árbol de decisión con los datos reales. Esto se debe a que todos los alelos son distintos entre sí y dentro de una clase encontramos columnas que poseen varias letras diferentes.

El árbol de decisión obtenido a partir de los datos reales, que puede ser consultado en el apéndice A, consta de 31 nodos y 32 puntas. No obstante, el número de preguntas distintas asciende a 25, ya que algunas de éstas se repiten en nodos de ramas diferentes.

A continuación explicaremos la manera en que obtendremos la lista de preguntas de la primera etapa, a partir del árbol de decisión construido utilizando el proceso descrito en la presente subsección.

3.2.2 Procedimiento para obtener una lista de preguntas

A partir del árbol de decisión con el que contamos en este momento, obtendremos un conjunto inicial de preguntas que debemos verificar si es solución al problema de clasificación que nos ocupa. Dicho árbol sólo servirá de ayuda para encontrar un conjunto de preguntas que serán hechas simultáneamente. Más adelante comprobaremos si dicho conjunto es en verdad una solución, o que al menos clasifique a la mayoría de los bialelos con los que trabajamos.

Para obtener esta lista de preguntas, primero consideramos el conjunto de nodos del árbol de decisión. Como se mencionó en la subsección anterior, éste consta de un total de 31 elementos, de los cuales 6 están repetidos. Es decir, tenemos 25 preguntas distintas que nos clasifican alelos en clases. Estas 25 preguntas involucran 23 posiciones, o sea que en dos casos se hicieron preguntas distintas en una misma posición. Esto puede observarse en la Tabla 10, que enlista el conjunto de 25 preguntas que nos proporcionó el árbol de decisión.

1	¿C en pos. 5?	**14**	¿G en pos. 234?
2	¿G en pos. 8?	**15**	¿A en pos. 268?
3	¿A en pos. 24?	**16**	¿A en pos. 282?
4	¿A en pos. 29?	**17**	¿G en pos. 289?
5	¿A en pos. 71?	**18**	¿C en pos. 312?
6	¿C en pos. 107?	**19**	¿G en pos. 341?
7	¿G en pos.184?	**20**	¿G en pos. 345?
8	¿A en pos. 184?	**21**	¿A en pos. 380?
9	¿A en pos. 197?	**22**	¿A en pos. 416?
10	¿G en pos. 209?	**23**	¿A en pos. 429?
11	¿C en pos. 226?	**24**	¿A en pos. 451?
12	¿A en pos. 226?	**25**	¿G en pos. 498?
13	¿A en pos. 228?		

Tabla 10: Preguntas arrojadas por el árbol de decisión.

Como puede verse en esta tabla, las preguntas 7 y 8 involucran a la posición 184 y las preguntas 11 y 12 involucran a la posición 226. Y éstos son los únicos casos en que se repite la posición. Por lo tanto, tenemos un total de 23 posiciones involucradas en este conjunto de preguntas.

Ahora lo que haremos será eliminar temporalmente todas las demás posiciones del conjunto de datos, para ver si con nuestro pequeño conjunto de 23 posiciones tenemos suficiente información para resolver el problema de clasificación de bialelos. Esto nos genera una matriz de 251 filas (alelos) y 23 columnas (posiciones), a la cual le aplicaremos el primer y segundo pasos de la transformación descrita en el modelo matemático del capítulo 2.

El primer paso consiste en asignar a cada una de estas 23 posiciones, tantas columnas como letras diferentes aparezcan en la posición en todos los datos (ver Tabla 2 en el capítulo 2), lo cual nos deja con una matriz de 60 columnas. Luego, en el segundo paso insertamos una nueva línea que servirá de encabezado, en el cual se escriben las letras que aparecen en cada posición y generamos una matriz de 0's y 1's, como se ejemplifica en la Tabla 3 del capítulo anterior. Cada columna representa ahora una pregunta y en cada fila aparecen las respuestas que arroja el alelo correspondiente a estas preguntas.

El encabezado de la matriz de 60 columnas en cuestión puede verse en la Tabla 11. Recordemos que estas preguntas se leen por posición y letra. Por ejemplo, las primeras dos preguntas son ¿existe C en la posición 5? y ¿existe T en la posición 5? Así, preguntamos a cada posición la existencia de la letra que aparece debajo.

Posición	5	5	8	8	8	24	24	29	29	29	71	71	107	107	107	183	183
Pregunta	**C**	**T**	**A**	**G**	**C**	**A**	**T**	**A**	**C**	**T**	**A**	**C**	**G**	**C**	**T**	**G**	**C**

184	184	184	197	197	197	209	209	226	226	226	228	228	234	234	268	268
A	**G**	**T**	**A**	**C**	**T**	**G**	**C**	**A**	**C**	**T**	**A**	**G**	**G**	**C**	**A**	**C**

282	282	282	289	289	312	312	341	341	341	345	345	345	345	380	380	416
A	**G**	**C**	**G**	**T**	**C**	**T**	**A**	**G**	**C**	**A**	**G**	**C**	**T**	**A**	**C**	**A**

416	429	429	429	451	451	451	498	498
G	**A**	**G**	**C**	**A**	**G**	**T**	**G**	**T**

Tabla 11: Lista de preguntas de la primera etapa.

Ahora contamos con la primera lista de preguntas, es decir la de la primera etapa. Lo siguiente que debemos hacer es ver si estas preguntas son suficientes para clasificar bialelos en biclases. En caso de no clasificarlos todos, necesitamos verificar que clasifique un porcentaje alto y, posteriormente, habrá que encontrar una segunda etapa de preguntas para clasificar los restantes.

En la siguiente sección del presente capítulo, construiremos una matriz de biclases a partir de las preguntas de la primera etapa. Emplearemos el tercer paso de la transformación descrita en el modelo matemático, pero antes habrá que hacer una depuración de los datos que será explicada enseguida.

3.3 CONSTRUCCIÓN DE LA MATRIZ DE BICLASES

La lista de 60 preguntas de la primera etapa nos genera una matriz de 0's y 1's con 60 columnas y 251 líneas (alelos). Sin embargo, por el hecho de haber eliminado columnas de la matriz original de datos, es de esperarse que algunas de estas líneas sean iguales. Sabemos que todos los alelos son distintos entre sí, pero si ignoramos algunas posiciones es posible encontrar grupos de alelos que coincidan en todas las demás.

Ahora bien, debido a que el conjunto de preguntas que estamos considerando fue arrojado por el árbol de decisión construido en la sección precedente, podemos estar seguros de que las líneas que coinciden en nuestra matriz de 60 columnas pertenecen a una misma clase. Dicho árbol de decisión, nos proporcionó una clasificación en 25 preguntas, las cuales bastan para garantizar la ubicación de todos los alelos en su clase. Nosotros añadimos otras 35 preguntas a este conjunto con la esperanza de lidiar con el problema de pérdida de información que se nos presenta.

En consecuencia, el siguiente paso a seguir es depurar la matriz de 60 columnas y 251 líneas, que es la que consideramos momentáneamente como nuestros datos totales. La depuración consiste en eliminar los alelos iguales para dejar sólo un representante de cada grupo. Como sabemos que no pueden coincidir alelos de diferentes clases, el proceso se simplifica si depuramos clase por clase. Esta depuración resulta muy conveniente, pues reduce el número de alelos que vamos a considerar y por supuesto también el número de bialelos.

De hecho, una vez depurada la matriz de 60 columnas y 251 líneas, obtenemos otra del mismo número de columnas pero con 115 líneas, que puede consultarse en el apéndice B. O sea que, en lugar de tener 251(252)/2 = 31626 bialelos, contamos solamente con 115(116)/2 = 6670 bialelos que es una cantidad mucho más manejable.

A continuación se presenta el programa computacional que fue utilizado para obtener la matriz de 6670 bialelos a la que nos referimos en el párrafo anterior, construida a partir de la matriz depurada consistente en 115 alelos y 60 columnas que representan preguntas. La matriz de 6670 bialelos puede verse en el apéndice C de esta tesis.

3.3.1 Programa para obtener la matriz de biclases

El lenguaje de programación JAVA fue empleado para generar la matriz de biclases que estamos buscando. Abajo se muestra el código de dicho programa, el cual será explicado mediante un diagrama de flujo en la siguiente subsección. Sólo cabe aclarar que la definición de la matriz fue eliminada del código, debido a que ocupa una gran cantidad de espacio y sólo consta de la lista de 0's y 1's que puede ser consultada en el apéndice B. El texto del código que aparece después del símbolo "//" sólo representa una nota y no forma parte del programa.

```
import java.awt.*;
import java.applet.Applet;

public class Matriz0y1 extends Applet
{
  Label etiq1, etiq2;
  TextField coef;
  Button boton, boton2;
  int lin, k, c, pos, numclases, numbiclases;
  int[] a = new int[2], b = new int[3];
  double temp, pas, suma;
  double[][][] mat1= new double[2][3][3];
  double[][][] mat2= new double[3][6][3];
  boolean pintar, pintar1, pintar2, pintar3, logica;

 public void init()
  { etiq1 =new Label("Este programa convierte la matriz de clases en la matriz de biclases");
    add(etiq1);
        boton = new Button("Crear matriz de biclases");
        boton.setBounds(150,150,10,2);
        add(boton);
        boton2 = new Button("Imprimir matriz de biclases");
        boton2.setBounds(200,150,10,2);
        add(boton2);
  }

 public boolean action (Event e, Object o)
  {
   if (e.target==boton)
    {// aquí definimos manualmente las entradas de la matriz de clases mat1[][][]:
        numclases=21;  //definimos el número de clases en mat1
        a[0]=3; a[1]=2;  // etc, definimos los tamaños de las clases 1, 2, etc. de mat1
        pos=60;  //definimos el número de columnas en la matriz de clases (mat1)
        Hacer(); //llamamos el método para construir la matriz de biclases
    }

   if (e.target==boton2)
   {pintar=true;}
   repaint();
   return true;
  }
```

```
public void Hacer()
{k=0;
 numbiclases=0; //inicialización del numero de biclases
 for (int i=0; i<numclases; i++)
  {c=i; lin=0;
   for (int d=c; d<numclases; d++)
    {if(d==c)                        //se construye una biclase a partir de una sola clase
      {
       for (int e = 0; e < a[d]; e++) {
         for (int f = e; f < a[d]; f++) {
           for (int p = 0; p < pos; p++) {
             suma = mat1[c][e][p] + mat1[d][f][p];
             if (suma == 0)
             {
               mat2[k][lin][p] = 0;
             }
             else {
               mat2[k][lin][p] = 1;
             }
           }
           lin++;
           b[k]++;
         }
       }
       numbiclases++;
      }

      if (d > c)                  //se construye una biclase a partir de dos clases distintas
       {
       for (int g = 0; g < a[c]; g++) {
         for (int h = 0; h < a[d]; h++) {
             for (int p = 0; p < pos; p++) {
               suma = mat1[c][g][p] + mat1[d][h][p];
               if (suma == 0)
               {
                 mat2[k][lin][p] = 0;
               }
               else {
                 mat2[k][lin][p] = 1;
               }
             }
           lin++;
           b[k]++;
         }
       }
       numbiclases++;
      }
      lin = 0;
      k++;
    }
  }
}



public void paint (Graphics g)
```

```
{ int x=50,y=40;
  int r=0, s=0, t=0, v=20;
  int biclase=0;
  if (pintar)
  {g.drawString("La matriz de biclases que se obtuvo es:", x, y);
    while(biclase<numbiclases)
    {while(r<b[biclase])
     {for(int col=0; col<pos; col++)
      {g.drawString(“”+mat2[biclase][r][col], x+u, y+v);
       u+=10;
      }
       r++;
       v+=20;
       u=0;
     }
     biclase++;
     r=0;
     v+=10;
    }
  }
 }
}
```

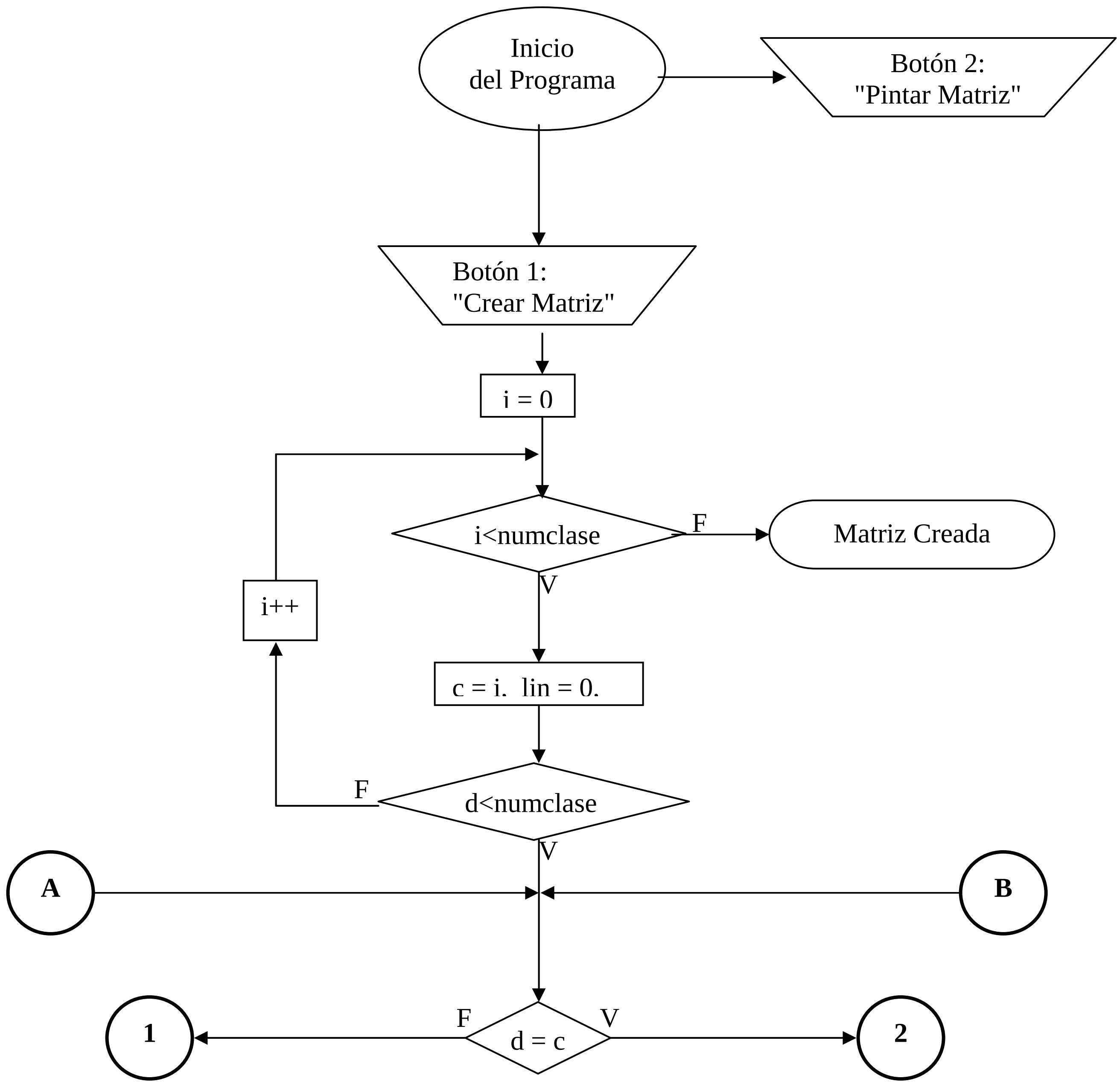
3.3.1 Diagrama de flujo del programa para obtener la matriz de biclases

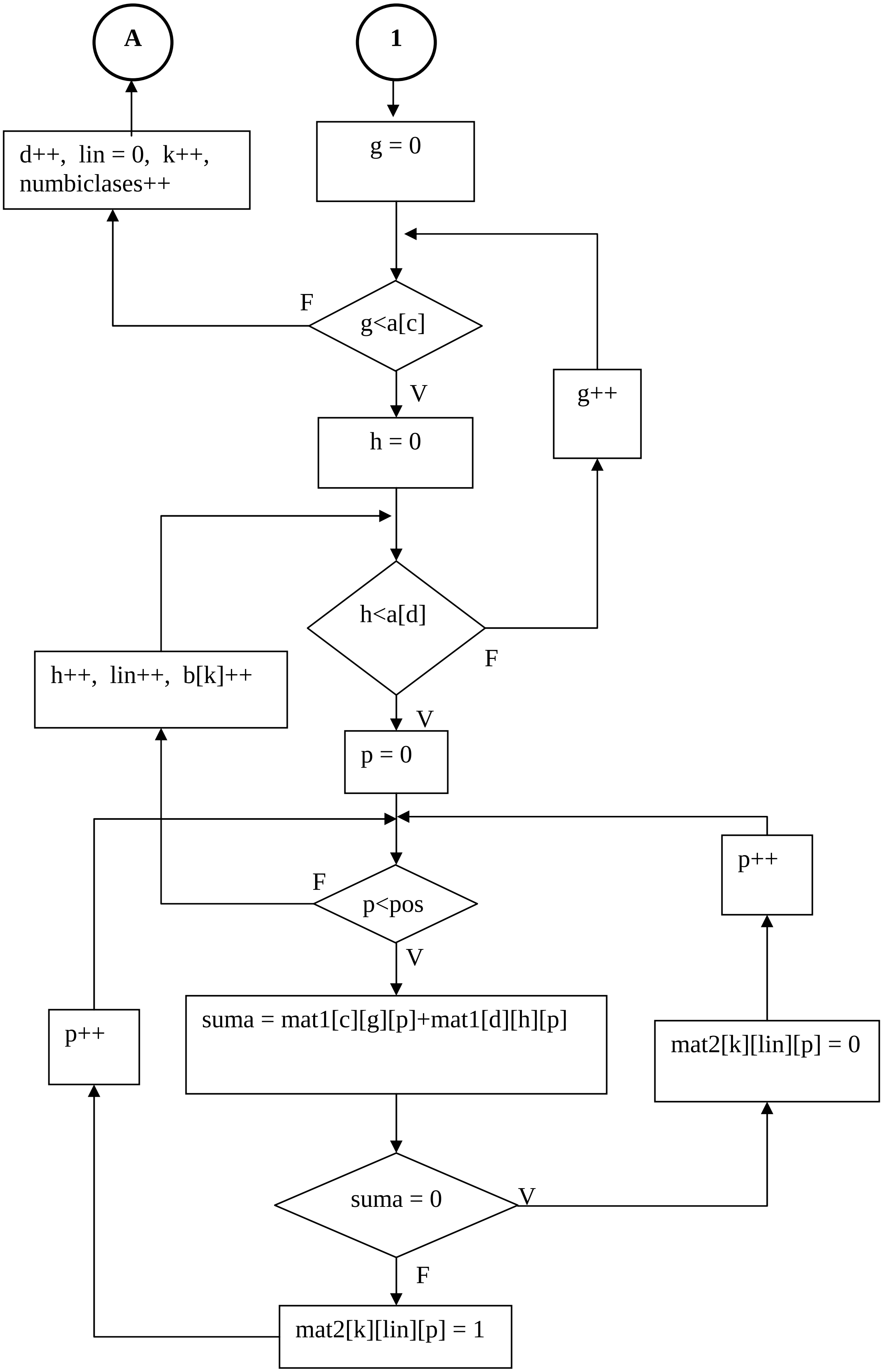
En esta subsección presentaremos un diagrama de flujo que nos explica cómo funciona el programa en JAVA para obtener la matriz de biclases cuyo código se encuentra en la subsección anterior. Para una mejor comprensión del diagrama se dan las siguientes definiciones:

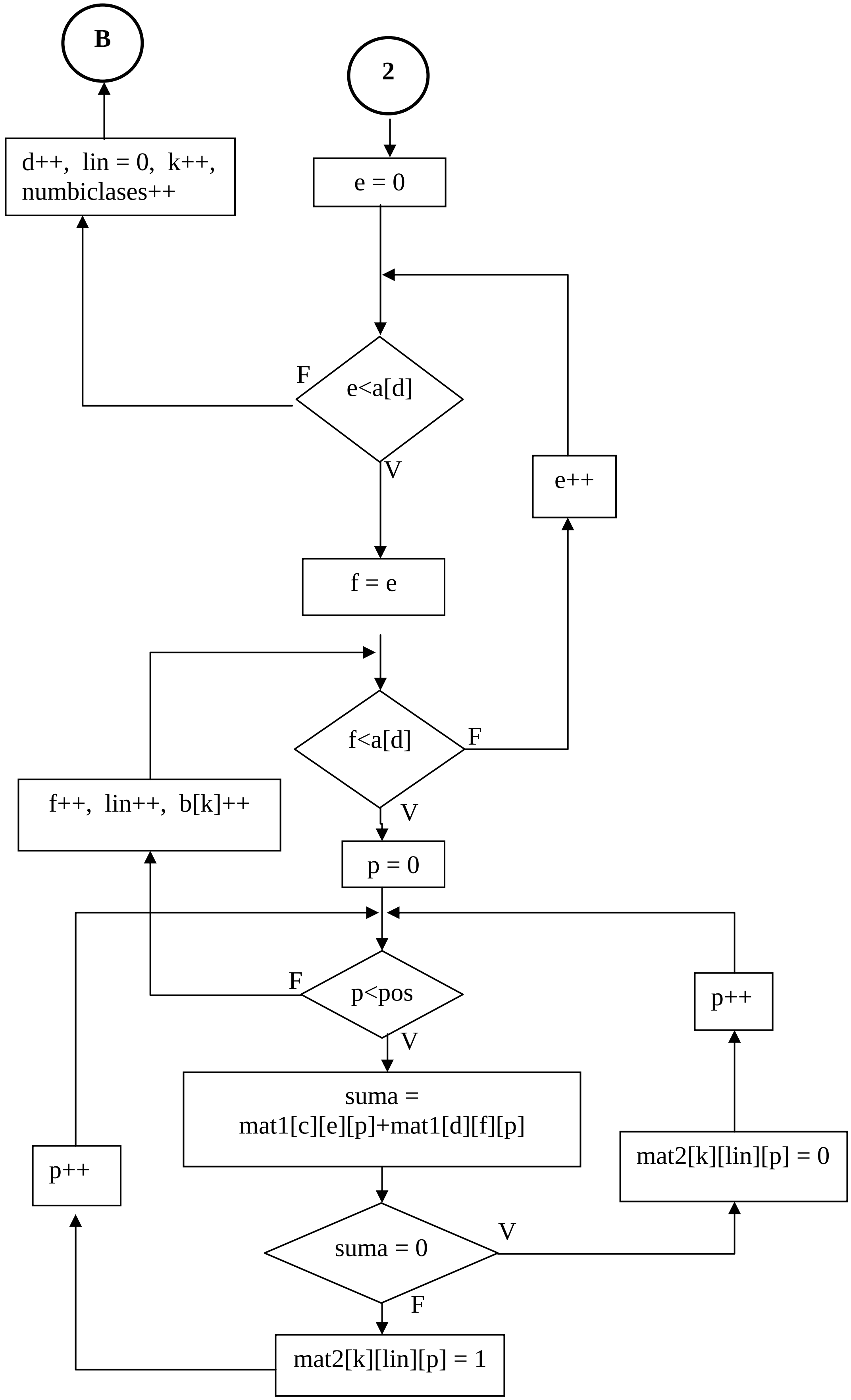
<u>Variables</u>:

a) **mat1[x][y][z]** = entrada de la matriz de clases, correspondiente a la clase x, fila (alelo) y, columna (pregunta) z.
b) **mat2[u][v][w]** = entrada de la matriz de biclases, correspondiente a la biclase u, fila (bialelo) y, columna (pregunta) z.
c) **Numclases** = número de clases en la matriz original de alelos
d) **Numbiclases** = contador de biclases de la matriz generada de bialelos
e) **a[m]** = tamaño de la clase m de la matriz mat1.
f) **b[n]** = tamaño de la biclase n de la matriz mat2.
g) **pos** = el número total de columnas de la matriz (60).

Nota: El comando “++” después de una variable significa sumar uno a la variable (i.e. var++ equivale a var = var + 1).


Inicio
del Programa
Botón 2:
"Pintar Matriz"
Botón 1:
"Crear Matriz"
i = 0
i<numclase
F
Matriz Creada
V
i++
c = i. lin = 0.
F
d<numclase
V
A
B
1
F
d = c
V
2


A
1
d++, lin = 0, k++,
numbiclases++
g = 0
F
g<a[c]
V
g++
h = 0
h<a[d]
F
h++, lin++, b[k]++
V
p = 0
p++
F
p<pos
V
p++
suma = mat1[c][g][p]+mat1[d][h][p]
mat2[k][lin][p] = 0
suma = 0
V
F
mat2[k][lin][p] = 1


B
2
d++, lin = 0, k++,
numbiclases++
e = 0
F
e<a[d]
V
e++
f = e
f<a[d]
F
f++, lin++, b[k]++
V
p = 0
F
p<pos
p++
V
suma =
mat1[c][e][p]+mat1[d][f][p]
p++
mat2[k][lin][p] = 0
suma = 0
V
F
mat2[k][lin][p] = 1

3.4 OBTENCIÓN DE PREGUNTAS DE LA SEGUNDA ETAPA

Una vez obtenida la matriz de biclases en la sección 3.3, debemos verificar si hay ambigüedades entre bialelos. Es decir, tenemos que encontrar bialelos iguales y determinar si pertenecen a la misma biclase o no. Los casos en que los dos bialelos son miembros de la misma biclase los hacemos un lado, pues sabemos que nuestro objetivo es determinar la biclase a la que pertenece el bialelo de un ser humano, no encontrar el bialelo particular que posee. En cambio, cuando dos bialelos idénticos, considerando las 60 preguntas de la primera etapa, pertenecen a biclases distintas, entonces tenemos un problema. Éste debe ser resuelto encontrando una segunda lista de preguntas que nos permita diferenciarlos y ubicarlos en su biclase correspondiente.

3.4.1 Detección y clasificación de ambigüedades

Lo primero que debemos hacer es construir identificadores que nos permitan comparar bialelos. Este proceso se describió en la sección 3.1 y se ejemplificó en la Tabla 6. La matriz a la que queremos asignar identificadores es la construida en la sección precedente, que consta de 60 columnas (preguntas) y 6670 filas (bialelos).

Siguiendo el procedimiento antes referido, dividimos las 60 posiciones en dos grupos de 30 que consideramos como números binarios. Transformamos éstos a su notación decimal y así obtenemos dos identificadores (ID1 y ID2) para cada línea. De esta manera, dos bialelos son iguales si sus pares de identificadores coinciden exactamente.

Note que pudimos haber escogido grupos más pequeños para construir los identificadores como, por ejemplo, seis grupos de diez posiciones o tres grupos de veinte posiciones. Sin embargo, esto nos genera una complicación a la hora de verificar si todos los identificadores coinciden. Tomando en cuenta la enorme cantidad de bialelos que debemos analizar, lo más conveniente es tener pocos identificadores. Por otro lado, tampoco fue posible construir un solo identificador que agrupara las 60 posiciones, ya que el número 2^{60} es demasiado grande para poder ser interpretado por la computadora. En consecuencia, se encontró que dos grupos de treinta posiciones empleados para construir identificadores al mismo tiempo redujo la cantidad de operaciones que debían realizarse para diferenciar los bialelos y generó números lo suficientemente pequeños para poder ser interpretados por la computadora.

Para hallar las ambigüedades o bialelos iguales, seguimos un procedimiento sencillo. En primer lugar, ordenamos los pares de identificadores, primero por ID1 y luego por ID2, obteniendo una lista de bialelos ordenada por sus identificadores. Después, en cada línea de dos columnas adicionales hacemos la resta del identificador de esa línea con el de la línea inmediata inferior. Así, cuando las restas de ambos identificadores sean iguales a cero, habremos encontrado dos bialelos iguales. Finalmente, verificamos la biclase a la que pertenecen estos bialelos y sabremos si tenemos una ambigüedad positiva (misma biclase) o una ambigüedad negativa (biclases distintas).

	Bialelos	ID1	ID2
1	1x1	2	16
2	1x1	8	16
3	1x2	32	24
4	1x2	24	24
5	1x2	2	16
6	2x3	8	24
7	2x3	2	32
8	3x4	24	32
9	3x4	10	16
10	3x4	24	32

	Bialelos	ID1	ID2	Resta ID1	Resta ID2	Ambigüedades
1	1x1	2	16	0	0	negativa
2	1x2	2	16	0	-16	
3	2x3	2	32	-6	16	
4	1x1	8	16	0	-8	
5	2x3	8	24	-2	8	
6	3x4	10	16	-14	-8	
7	1x2	24	24	0	-8	
8	3x4	24	32	0	0	positiva
9	3x4	24	32	-8	8	
10	1x2	32	24			

Tabla 12: Ordenamiento y comparación de identificadores.

En la Tabla 12 podemos ver un ejemplo del ordenamiento y comparación de los identificadores. Al lado izquierdo de la tabla hay una lista de bialelos, representados por sus biclases, y junto a ellos sus dos identificadores. Del lado derecho de la tabla, se encuentran ordenados los bialelos por sus identificadores, primero por ID1 y después por ID2, como se describió arriba. También encontramos la resta del identificador de cada línea con el de debajo en dos columnas adicionales. Y en la última columna se marcan las ambigüedades, que corresponden a las líneas en que ambas restas son iguales a cero. La primera ambigüedad es negativa, pues el alelo de la primera línea pertenece a la biclase 1x1 y el de la segunda a la biclase 1x2. En cambio, la otra ambigüedad es positiva, ya que los dos alelos que intervienen pertenecen a la misma biclase, la 3x4.

De esta forma detectamos y clasificamos las ambigüedades en positivas y negativas, según el criterio que se definió, de la matriz de 6670 bialelos construida en la sección anterior. Encontramos mediante este proceso un total de 435 ambigüedades, de las cuales 117 son negativas y 318 positivas.

En consecuencia, debemos encontrar una segunda etapa de preguntas que nos permita diferenciar estos 117 casos de ambigüedades negativas. Sin embargo, es importante tener en cuenta que dentro de estos 117 casos de ambigüedad se encuentra el conjunto de bialelos que no pueden ser ubicados en su biclase. Es decir, el 1.79% de bialelos (568 de 31626) no clasificables definido en la solución inicial, sección 3.1. Este conjunto no pudo ser eliminado previamente, pues los alelos que forman los bialelos ambiguos también forman otros bialelos que sí son clasificables. Así que nos vemos obligados a "cargar" con ellos hasta que podamos identificarlos y descartarlos al final de nuestro estudio

En la siguiente subsección encontraremos la forma de clasificar la mayoría de los 117 casos que permanecieron ambiguos después de las 60 preguntas de la primera etapa. El resto será descartado por ser imposible de clasificar bajo las hipótesis de esta tesis.

3.4.2 La segunda etapa

El procedimiento para clasificar los bialelos que escaparon a la primera etapa consistirá de varios pasos. Primero descartaremos las 60 preguntas utilizadas en la primera etapa. En la descripción del modelo matemático del capítulo 2 se mostró que el total de preguntas útiles que podemos hacer a nuestro conjunto de datos es de 265, por lo que nos quedaremos con un total de 205 preguntas posibles al eliminar la primera etapa.

En este momento estamos suponiendo que conocemos las respuestas a la lista de la primera etapa para los 117 casos que queremos analizar. Esta información es muy útil, pues podemos asignar a cada respuesta, una nueva lista de preguntas. Es decir, no hace falta encontrar una segunda lista de preguntas que nos clasifique a los 117 casos simultáneamente, excluyendo los imposibles. Una lista de preguntas con estas características es más grande que una individual para cada una de las 117 respuestas.

Lo que haremos será considerar solamente las preguntas que estén lo bastante "cargadas" de información, bajo el siguiente criterio. Como se expuso en el capítulo 2, a lo largo del modelo matemático, los datos originales que nos fueron proporcionados habían sido previamente depurados. En la depuración fueron eliminadas las posiciones que tenían la misma letra en todos los alelos, dejando solamente las que tenían al menos dos letras diferentes en todo el conjunto de alelos. Como resultado, tenemos algunas posiciones o columnas con la misma letra en 250 alelos y una distinta en 1 solo alelo. Otras columnas tienen 249 letras iguales y 2 distintas. Estos dos tipos de columna las consideraremos como poco "cargadas" de información y no serán tomadas en cuenta para clasificar los 117 bialelos que escaparon a la primera etapa.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Position										1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	2	2	2	4	4	5	5	5	5	2	2	3	3	4	4	5	5	6	6	6	6	7	7	8	8	8	8	8	9	9
	5	5	5	8	8	0	0	3	3	7	7	0	0	6	6	5	5	5	5	7	7	0	0	6	6	6	8	8	2	2
Pregunta	**A**	**C**	**T**	**A**	**C**	**C**	**T**	**A**	**G**	**A**	**G**	**A**	**G**	**A**	**G**	**A**	**G**	**A**	**G**	**G**	**T**	**G**	**T**	**A**	**G**	**C**	**G**	**C**	**G**	**C**

	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Position	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3
	9	9	1	1	1	1	1	1	2	2	3	3	4	4	4	4	4	4	4	4	4	9	9	9	9	9	9	1	1	2
	8	8	7	7	7	9	9	9	9	9	8	8	0	0	1	1	4	4	4	6	6	0	0	5	5	5	5	8	8	4
Pregunta	**A**	**G**	**G**	**C**	**T**	**A**	**G**	**C**	**A**	**G**	**C**	**T**	**G**	**C**	**C**	**T**	**A**	**G**	**T**	**G**	**C**	**A**	**G**	**A**	**G**	**C**	**T**	**G**	**T**	**C**

	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
Position	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	2	2	2	3	3	3	3	4	4	5	5	7	7	2	2	3	3	4	4	4	4	5	5	5	5	5	5	5	6	6
	4	9	9	8	8	9	9	0	0	0	0	5	5	4	4	3	3	4	4	8	8	3	3	3	4	4	4	4	5	5
Pregunta	**T**	**G**	**C**	**C**	**T**	**G**	**C**	**A**	**G**	**C**	**T**	**C**	**T**	**C**	**T**	**A**	**G**	**A**	**G**	**C**	**T**	**G**	**C**	**T**	**A**	**G**	**C**	**T**	**C**	**T**

	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109
Position	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5
	6	6	6	7	7	8	8	8	8	8	8	8	8	8	8	9	9	1	1
	6	6	6	2	2	2	2	6	6	6	6	7	7	7	7	7	7	0	0
Pregunta	**A**	**G**	**T**	**C**	**T**	**G**	**T**	**A**	**G**	**C**	**T**	**A**	**G**	**C**	**T**	**G**	**C**	**C**	**T**

Tabla 13: Preguntas de la segunda etapa.

Encontramos 96 posiciones poco cargadas de información que no serán tomadas en cuenta. Así, al ignorar las 60 preguntas de la primera etapa y las 96 con poca información, tenemos un conjunto de 109 preguntas que integrarán la segunda etapa. Esta lista puede observarse en la Tabla 13. Se encuentran agrupadas en cuatro conjuntos, tres de 30 preguntas y uno de 19. Esto nos servirá en el capítulo 4 para construir los identificadores y éstos, a su vez, servirán para decidir cuál o cuáles de estos cuatro conjuntos de preguntas será necesario hacer para clasificar un bialelo particular, según la respuesta obtenida a la lista de la primera etapa.

En el capítulo 4 daremos una interpretación a los resultados obtenidos en este trabajo de tesis. Se explicará cómo deben leerse las respuestas arrojadas a las 60 preguntas de la primera etapa. Asimismo, se dará a conocer el criterio que nos ayudará a decidir si es necesario pasar a la segunda etapa. Y, en caso de pasar a la segunda etapa, explicaremos cómo elegir una o varias de las 4 listas de preguntas que se obtuvieron en la segunda etapa para ubicar un bialelo en su biclase.