



CAPÍTULO 3

PROBLEMA MATEMÁTICO

En este capítulo se presenta de manera matemática el problema así como el desarrollo del manejo de datos con el fin de que el lector comprenda mejor la problemática. También se definirán algunos términos importantes que se utilizarán en el transcurso de la tesis y principalmente en la programación.

3.1 CATÁLOGOS HLA

Dado que existe una gran variedad de información contenida en cada alelo que se descubre, cada 4 meses se actualizan los datos que se encuentran en los catálogos HLA [G2005].

A continuación en la Tabla 3.1.1 se presenta la nomenclatura que se utiliza en los catálogos HLA que esta establecida por la European Bioinformatics Institute [B2005] [TAN2006]:

Nomenclatura	Descripción
HLA	Prefijo utilizado para nombrar el gen y proporciona la información de que se trata de la región HLA
HLA-A	Nombre de un gen en particular



HLA-A*01	Familia a la que pertenece el alelo dentro del gen
HLA-A*0102	Alelo en específico
HLA-A*0104N	Alelo nulo
HLA-A*010102	Alelo con mutación sinónima
HLA-A*03020101	Alelo con mutación fuera de la región de codificación
HLA-A*04020101N	Alelo nulo con mutación fuera de la región de codificación

Tabla 3.1.1 Descripción de la nomenclatura usada en los catálogos HLA

La información contenida en estos catálogos varía ya que, como se mencionó anteriormente, esta se divide en clases que a su vez están formadas por cierto número de subclases. En la Tabla 3.1.2 se muestra el número de subclases y el número de alelos contenidos en cada clase que se obtuvo en Octubre del 2005:

Clase	Número de Alelos	Subclases
HLA-A*	278	21
HLA-A	379	21
HLA-Cw	207	14
HLA-DQB1	67	5
HLA-DRB1	420	13

Tabla 3.1.2 Contenido de las clases del mes de Octubre del 2005 (la clase que tiene el * indica el catálogo utilizado en la tesis de Peña [P2004] utilizada como base para la programación)

Todas las clases contienen sus alelos y donde cada uno de estos tiene una longitud de 545 posiciones (columnas) con la combinación de los componentes de las bases nitrogenadas adenina, timina, guanina y citosina (A, T, G, C).



Existen otro tipo de clases por ejemplo la HLA-B, pero esta tesis se enfocó principalmente en las descritas en la Tabla 3.1.2 por contar con la información completa. Pero es necesario mencionar que el programa en lenguaje Java (Apéndice **A**) es capaz de resolver cualquier catálogo siempre y cuando se actualice la información de cada catálogo (ver manual: Apéndice **B**).

Los catálogos de lectura para el lenguaje Java se localizan en el Apéndice **C** que son los archivos de entrada para resolver la clasificación de alelos HLA con sus respectivos nombres.

3.2 MODELO MATEMÁTICO

Para poder resolver el problema de clasificación de los alelos es necesario transformar los datos a un lenguaje matemático. Dado que son varios catálogos, se presentarán tablas para mostrar los datos que se obtuvieron en el transcurso del proyecto de tesis.

El problema biológico se transforma en el modelo matemático basado en la siguiente estructura [L2003]:

1. Se tienen S_i conjuntos que son las subclases que contiene el catálogo, en donde $i=1, 2, \dots, s$.
2. Cada S_i se compone de a_i elementos que corresponden a los alelos para cada i .
Por lo que $S_{ij} = \{S_{i1}, S_{i2}, \dots, S_{ia_i}\}$ donde $j=1, 2, \dots, a_i$.
3. También cada S_{ij} tiene un número p de posiciones, quedando $S_{ij} = \{S_{i1}^1, S_{i2}^2, \dots, S_{ia_i}^{545}\}$ en donde cada posición está caracterizado por uno de los elementos de $B = \{A, T, G, C\}$.



Cabe mencionar que cada una de las S_{ij} solo pertenece a una subclase S_i .

Para entender el modelo matemático, se toma como ejemplo el catálogo HLA de la clase ficticia “X” y los alelos que contiene son los que se muestran en la Fig. 3.2.1, compuesto por subclases (S_i), elementos de las clases que son los alelos (S_{ij}) y combinaciones de los elementos del conjunto B . La información que se va presentando en las Tablas del ejemplo del catálogo “X” en excel se localizan en el Apéndice D.

Las subclases se obtienen con los primeros dos dígitos del nombre del alelo, por ejemplo la subclase 01 esta formada por 3 alelos: A*010101, A*010102 y A*0102, la subclase 02 tiene 5 alelos: A*020101, A*020102, A*020103, A*020104 y A*020105 y así con cada una de las subclases.

Alelos		1	2	3	4	5	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Subclase 01 S_1	A*010101	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
	A*010102	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
	A*0102	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	C	T	T	G	A
Subclase 02	A*020101	A	T	G	C	T	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
	A*020102	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
	A*020103	A	T	G	C	T	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
	A*020104	A	T	G	C	T	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
	A*020105	A	T	G	C	T	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
Subclase 03	A*03010101	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
	A*03010102N	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
Subclase 04	A*4301	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	A	T	T	G	C

Combinaciones del conjunto B
 S_{ij}

Elementos de la Subclase S_{ij}

Fig. 3.2.1 Composición del catálogo HLA de la clase “X”

Dado que es demasiado manejo de datos, se requiere eliminar aquellas posiciones que no proporcionen información. Esto se refiere a que una posición que contiene todas sus letras iguales no proporciona información que sirva para la clasificación.



Tomando el ejemplo del catálogo de alelos HLA de la clase “X” se observa en la Fig. 3.2.1 las posiciones (columnas) que no dan información, con lo que eliminando estas columnas se reducen el número de posiciones a manejar como se observa en la Fig. 3.2.2.

Alelos	1	2	3	4	5	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
A*010101	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
A*010102	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
A*0102	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	C	T	T	G	A
A*020101	A	T	G	C	T	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
A*020102	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
A*020103	A	T	G	C	T	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
A*020104	A	T	G	C	T	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
A*020105	A	T	G	C	T	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
A*03010101	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
A*03010102N	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	T	T	T	G	A
A*4301	A	T	G	C	C	C	T	A	A	A	C	G	G	T	G	T	A	C	G	C	G	T	A	T	T	G	C

Fig. 3.2.2 Las posiciones que están marcadas con el recuadro rojo son las que proporcionan información y las que se están en color gris son las que se eliminan.

Es importante mencionar que estas columnas respetan su posición original, ya que se debe mantener las posiciones consecutivas. Por ejemplo si solamente quedan tres posiciones con letras diferentes ya sea el 5, el 25 y el 29 en este caso, el 5 no toma la posición del número 1, ni el 25 con el 2 debido a que más adelante se seguirán manejando los datos con sus respectivas posiciones, pues en la naturaleza no llega a suceder esto y los datos que se obtendrían serían falsos.

Con la eliminación de columnas se tiene una matriz depurada D con $a \times c$ elementos en donde a es el número de alelos del catálogo y c es el número de columnas no eliminadas por contener información.

Este procedimiento de depuración permite un mejor manejo de datos y es realizado por el programa en lenguaje Java dejando para cada catálogo la información que se muestra en



la Tabla 3.2.1:

Catálogo	No. de Posiciones para trabajar	Tamaño de la matriz D $a \times c$
HLA-A*	129	278*129
HLA-A	153	379*153
HLA-Cw	105	207*105
HLA-DQB1	64	67*64
HLA-DRB1	115	420*115

Tabla 3.2.1 Matriz D a utilizar para resolver el problema.

3.2.1 CÁLCULO DE PREGUNTAS

Una vez que se tiene la matriz D el siguiente paso es obtener las preguntas que se aplicarán más adelante.

Estas preguntas aplicadas al órgano del receptor o donador permiten dar una clasificación a los alelos a una biclase en específico. En donde la biclase se define como la combinación de los alelos de una clase con otra, más adelante se mostrará la obtención de estas biclases.

Estas preguntas se aplican en la vida real para determinar la biclase del órgano a través de reactivos que permiten definir la letra que se encuentra en cierta posición.

Como ya se mencionó anteriormente, estas preguntas se componen de 20 posiciones consecutivas que corresponden a la combinación de los elementos de la base nitrogenada adenina, timina, guanina y citosina (A, T, G, C) [P2004].

Partiendo de la matriz D se obtienen las preguntas de 20 posiciones, pero dado que se



eliminaron columnas en la matriz D hay que respetar el número de esas posiciones eliminadas.

Aplicando esto al ejemplo del catálogo de alelos HLA de la clase “X” se tiene que la matriz D es la que se muestra en la Tabla 3.2.1.1.

Alelo	5	25	20
A*010101	C	T	A
A*010102	C	T	A
A*0102	C	C	A
A*020101	T	T	A
A*020102	C	T	A
A*020103	T	T	A
A*020104	T	T	A
A*020105	T	T	A
A*03010101	C	T	A
A*03010102N	C	T	A
A*4301	C	A	C

Tabla 3.2.1.1 Elementos de la Matriz D

Las preguntas en este ejemplo se obtienen de la siguiente manera, partiendo de la posición 5 se tienen que tomar las 20 posiciones tomando en cuenta la misma posición 5 con el fin de obtener una pregunta, lo que da como resultado que el contenido de la primera pregunta queda con las posiciones 5, 6, 7, ..., 23, 24, pero dado que las posiciones de la matriz ya no están completas únicamente se colocan las que se encuentran indicadas en la matriz D . Otro punto importante es que muchas de las letras localizadas en las mismas posiciones de dos o más diferentes alelos van a coincidir, lo que implica que muchas preguntas serán iguales en las mismas posiciones por lo que es necesario eliminar aquellas que aparezcan nuevamente si ya han sido registradas.

Para explicar mejor lo anterior se toma nuevamente el ejemplo del catálogo de alelos



HLA de la clase “X”. La primera pregunta que se obtiene parte de la posición 5 hasta la 24, dado que las columnas del 6 a la 24 han sido eliminadas queda únicamente la posición 5 con las letras C y T en sus respectivos alelos (renglones) pero siempre en la misma posición (columna). Esto proporciona únicamente 2 preguntas para esta posición. Obsérvese la Fig. 3.2.1.1.

Preguntas de la Posición 5 a la 24

Alelo	5	
A*010101	C	→ Primera pregunta
A*010102	C	→ Ya existe
A*0102	S	
A*020101	T	→ Segunda pregunta
A*020102	C	→ Ya existe...
A*020103	T	
A*020104	T	
A*020105	T	
A*03010101	C	
A*03010102N	C	
A*4301	C	

PREGUNTA / POSICIÓN	5
0	C
1	T

Fig. 3.2.1.1 A la izquierda se van seleccionando las preguntas que no han salido anteriormente y a la derecha se encuentra la lista de preguntas seleccionadas para el rango de 5-24 posiciones

Realizando el mismo procedimiento para todo el catálogo de “X” se obtienen 7 preguntas dando como resultado la Tabla 3.2.1.2.

PREGUNTA	RANGO DE PREGUNTA	POSICIÓN		
		5	25	29
0	5-24	C	---	---
1	5-24	T	---	---
2	25-44	---	T	A
3	25-44	---	C	A
4	25-44	---	A	C
5	29-48	---	---	A
6	29-48	---	---	C

Tabla 3.2.1.2 Total de Preguntas para el catálogo “X”



Estas preguntas se obtienen a través del programa en Java el cual da como resultado el número de preguntas que se obtuvieron para los catálogos seleccionados con anterioridad y se encuentran localizados en el Apéndice E con el nombre de “archivo_preguntas_formato.txt”. Estas preguntas forman el conjunto P en donde $P = \{P1, P2, P3, \dots, k\}$ donde k es el número total de preguntas. El número de preguntas resultantes se muestran en la Tabla 3.2.1.3.

Catálogo	No. de Preguntas
HLA-A*	1117
HLA-A	1570
HLA-Cw	693
HLA-DQB1	388
HLA-DRB1	1809

Tabla 3.2.1.3 Preguntas obtenidas por el programa Java para cada uno de los catálogos

Ya que se tienen las preguntas que se aplicarán para poder determinar la biclase es necesario realizar un paso más, el cálculo de los bialelos.

3.2.2 CÁLCULO DE BIALELOS

Obtenidas las posiciones que no se eliminan y las preguntas que se aplicaran el siguiente paso es realizar las combinaciones entre alelos para obtener los *bialelos* que, como se describió en el capítulo 1, son la combinación genética de parte de la madre y del padre.



Esta combinación se realiza entre cada alelo de cada una de las subclases de las posiciones que no se eliminaron, tomando en cuenta la combinación de un alelo en específico consigo mismo. Con estas combinaciones entre los alelos de cada subclase se llega a formar lo que se llama una *biclase*.

A continuación se muestra un ejemplo en la Fig. 3.2.3, tomando los dos primeros dígitos del nombre de los alelos para deteminar la biclase.

		Bialelos	5	25	29	
Combinación del alelo A*010101 con el mismo	{	A*010101	C	T	A	} Biclase 01*01
		A*010101	C	T	A	
		A*010101	C	T	A	
		A*010102	C	T	A	
Combinación del alelo de la subclase 01 con el Alelo de la subclase 02	{	A*010101	C	T	A	} Biclase 01*02
		A*020101	T	T	A	
		A*010101	C	T	A	
		A*020102	C	T	A	
		A*010101	C	T	A	
		A*020103	T	T	A	
		A*010101	C	T	A	
		A*020104	T	T	A	
		A*010101	C	T	A	
		A*020105	T	T	A	
}	} Biclase 01*03	A*010101	C	T	A	
		A*03010101	C	T	A	
		A*010101	C	T	A	
}	} Biclase 01*04	A*03010102N	C	T	A	
		A*010101	C	T	A	
		A*4301	C	A	C	
...						



...				
A*020101	T	T	A	} Biclase 02*02
A*020101	T	T	A	
A*020101	T	T	A	
A*020102	C	T	A	
A*020101	T	T	A	
A*020103	T	T	A	
...				
...				
A*03010101	C	T	A	} Biclase 03*03
A*03010101	C	T	A	
A*03010101	C	T	A	
A*03010102N	C	T	A	
A*03010101	C	T	A	} Biclase 03*04
A*4301	C	A	C	
...				
...				
A*4301	C	A	C	} Biclase 04*04
A*4301	C	A	C	

Fig. 3.2.2.1 Combinación de alelos de cada subclase para formar las biclases

Como García [G2003] lo expresa, se tiene la biclase “X×Y” que es la combinación de los bialelos formados por todos los alelos de la clase “X” con todos los alelos de la clase “Y” y dado que la suma entre los alelos es conmutativa, la biclase también es conmutativa por lo que “X×Y” es igual que “Y×X”. Dado que estas biclases son equivalentes, únicamente es necesario colocar una sola combinación tomando como biclase “X×Y” cuando $X \leq Y$. El conjunto de las biclases forman la matriz **B** de tamaño $b \times c$ donde **b** es el número de bialelos y **c** el número de columnas.

Para calcular el número de bialelos contenidos en la matriz **B** se utiliza la fórmula de combinación siguiente [P2004]:

$$No. de Combinaciones = \frac{n(n+1)}{2} \quad (3.2.2.1)$$



donde n es el número de alelos. En la Tabla 3.2.2 se muestra el número de bialelos y el tamaño de la matriz B obtenidos para cada catálogo a través del programa realizado en Java y la información de cada uno de los bialelos está en el E con el nombre de “archivo_bialelos_formato.txt”.

Catálogo	No. de Bialelos	Tamaño de la matriz B $b \times c$
HLA-A*	38781	38781*129
HLA-A	72010	72010*153
HLA-Cw	21518	21518*105
HLA-DQB1	2278	2278*64
HLA-DRB1	88410	88410*115

Tabla 3.2.2.1 Preguntas obtenidas del programa en Java para cada uno de los catálogos

3.2.3 APLICACIÓN DE PREGUNTAS A BIALELOS

Ya obtenida la matriz B así como las preguntas, se realiza el proceso de aplicar estas preguntas a la información genética que se tiene del donador o receptor. Esta parte consiste principalmente en realizar la pregunta del conjunto P en la posición correspondiente a los bialelos del receptor o donador para obtener como respuesta: Sí o No (Binario).

La respuesta va a depender de un Sí si la letra o el conjunto de letras que conforman la pregunta se presentan en cualquiera de los dos alelos en esa posición y se va a representar con un “1” y un No si no hay coincidencia de la pregunta con cualquiera de los dos alelos y su representación es por medio de un “0” [P2004].



Así que el enfoque va dirigido a encontrar la subclase a la que pertenece el donador y la subclase a la que pertenece el receptor a través las preguntas de manera positiva con el fin de que, como lo menciona Arroyo [G2003] “aquel receptor cuyo par de subclases coincida con el del donador es un candidato seguro a recibir el órgano en cuestión”.

Al realizar las preguntas a los bialelos se obtiene la matriz R la cual contiene la información de las respuestas que se obtienen con este procedimiento. Esta matriz es de dimensión $b \times p$ donde b es el número de bialelos (renglones) y p es el número de preguntas (columnas) que se obtuvieron en la sección 3.2.1 y 3.2.2 respectivamente. En la Tabla 3.2.3.1 se presenta el tamaño de cada matriz para los catálogos correspondientes.

Catálogo	Tamaño de la matriz R $b \times p$
HLA-A*	38781*1117
HLA-A	72010*1570
HLA-Cw	21518*693
HLA-DQB1	2278*388
HLA-DRB1	88410*1809

Tabla 3.2.3.1 Tamaño de las matrices de respuesta obtenidos de l programa Java para cada uno de los catálogos

Aplicando este procedimiento al ejemplo del catálogo “X”, la matriz R se obtiene aplicando las preguntas de la Tabla 3.2.1.2 a los bialelos de la Fig. 3.2.2.1 dando como resultado la Tabla 3.2.3.1.



	Bialelos	P0	P1	P2	P3	P4	P5	P6
Biclase 01*01	01*01	1	0	1	0	0	1	0
	01*01	1	0	1	0	0	1	0
	01*01	1	0	1	1	0	1	0
...								
Biclase 01*02	01*02	1	1	1	0	0	1	0
	01*02	1	0	1	0	0	1	0
...								
Biclase 01*03	01*03	1	0	1	0	0	1	0
	01*03	1	0	1	0	0	1	0
...								
Biclase 02*02	02*02	0	1	1	0	0	1	0
	02*02	1	1	1	0	0	1	0
	02*02	0	1	1	0	0	1	0
...								
Biclase 03*03	03*03	1	0	1	0	0	1	0
	03*03	1	0	1	0	0	1	0
...								
Biclase 03*04	03*04	1	0	1	0	1	1	1
	03*04	1	0	1	0	1	1	1
...								
Biclase 04*04	04*04	1	0	0	0	1	0	1

Tabla 3.2.3.1 Matriz de respuestas **R** para el catálogo “X”

Con respecto a los catálogos analizados en este proyecto, también se les aplicó las preguntas a las combinaciones de bialelos obteniendo sus matrices de respuestas y los cuales se encuentran registrados en el Apéndice **E** con el nombre de “matriz_respuestas.txt”.



El siguiente paso es aplicar algún método que permita obtener una mejor solución a la clasificación de los alelos con el fin de determinar la subclase a la que pertenece la información genética del donador o receptor del órgano. En este caso se parte de la Entropía de Shannon y la Ganancia de la Información.