



CAPÍTULO 2

ESTADO DEL ARTE

En este capítulo se introduce con más detalle el desarrollo científico que ha llevado a crear nuevas formas de resolver el problema del rechazo de órganos, dando una breve historia del descubrimiento de la genética y los avances que ha permitido en este ámbito. También se proporciona una breve descripción de estudios anteriores dedicados a clasificar los alelos HLA.

2.1 ASPECTOS HISTÓRICOS

Durante mucho tiempo, el hombre ha buscado la manera de reducir los malestares y molestias que le pueden ocasionar una enfermedad o la carencia del buen funcionamiento de alguno de sus órganos. Por lo cual en el aspecto médico y científico se han realizado, desde ya hace varios siglos una gran cantidad de estudios genéticos enfocados a mejorar la condición de vida o, en el peor de los casos, minimizar estos males.

Como lo describe Oliver [O1977] “la genética es aquella disciplina de la biología que estudia los fenómenos de herencia y las causas de las variaciones entre organismos”, por lo cual es necesario enfocarse a los estudios genéticos.



A mediados del siglo XIX, M.J. Schleider y Teodor Schwann desarrollaron la “Teoría Celular”, la cual abre paso a nuevos descubrimientos sobre la composición del ser humano. En 1869, el alemán Friedrich Miescher aisló por primera vez el ácido desoxirribonucleico (ADN), el cual es un componente de la célula. Para 1886 Gregorio Mendel, considerado el “Padre de la Genética”, desarrolló conceptos que establecieron las leyes de la herencia y como resultado las bases de la genética. A finales del siglo XIX, el alemán Theodor Boveri y el estadounidense W.S. Sutton dieron a conocer que la localización de los genes se encuentra en el “cromosoma” al igual que el ADN, descubierto por el alemán Robert Feulgen en 1914. Y en los años ´20 el bioquímico P.A. Levene realizó estudios al ADN, degradándolo hasta convertirlo en un azúcar de cinco carbonos, un grupo fosfato y cuatro bases nitrogenadas: adenina (A), guanina (G), timina (T) y citosina (C) [O1977] [SC1981].

Todos estos descubrimientos llevaron a los científicos a realizar experimentos médicos, algunos de ellos enfocados al trasplante de órganos. Al querer transplantar un órgano se encontraban con un problema muy grave, el rechazo de éste. Dado que antes del siglo XX no sabían las causas, muchos pacientes murieron debido a la gravedad del rechazo. No fue hasta 1958 que Van Roux descubrió el sistema HLA (*Human Leukocyte Antigen*) el cual, como ya se mencionó en el capítulo anterior, “establece que el organismo reconoce como distintos los tejidos de otro individuo, a pesar de pertenecer a la misma especie y es responsable del fenómeno del rechazo” [CP2003].



Para poder resolver los problemas biológicos que se presentan, no solamente en el proceso del trasplante de órganos, es necesario la manipulación de una gran cantidad de datos que crecen de manera exponencial, por lo que a finales del siglo XX los científicos deciden utilizar herramientas computacionales para poder apoyarse en éstas y llegar a encontrar soluciones a los problemas biológicos a corto o mediano plazo. A esta nueva ciencia se le conoce como *bioinformática*, la cual L. Garreta y Flechas [GF2006] la definen como “la aplicación y desarrollo de la computación, las matemáticas y la estadística al manejo, análisis y entendimiento de datos biológicos”.

Dando una definición exacta de la bioinformática, el Centro Nacional para la Información Biotecnológica "National Center for Biotechnology Information" (NCBI) la define de la siguiente manera:

“Bioinformática es un campo de la ciencia en el cual confluyen varias disciplinas tales como: biología, computación y tecnología de la información. El fin último de este campo es facilitar el descubrimiento de nuevas ideas biológicas así como crear perspectivas globales a partir de las cuales se puedan discernir principios unificadores en biología” [NCBI2006].

2.2 INVESTIGACIONES ANTERIORES

En base a la “Bioinformática” han surgido investigaciones y trabajos dedicados y enfocados a la solución de problemas biológicos apoyados con herramientas computacionales. Pero para poder establecer una comunicación entre la biología y la



computación es necesario el uso de algoritmos con el fin de transmitir la información entre ambas partes.

Uno de los trabajos ya existentes enfocados a clasificar los alelos HLA es la tesis de Luna [L2003], la cual está enfocada en transformar el problema biológico a un lenguaje matemático proporcionando un método a través de una definición y teorema basados en el módulo 2. En este trabajo se logran definir 14 preguntas de una sola letra para la clase HLA-A y 18 para la clase HLA-B. Esta tesis únicamente proporciona algunas ideas de las cuales se pueden partir para buscar una mejor solución a este problema ya que aún no proporciona una solución real.

Otro de los trabajos enfocados a la tipificación de los alelos del sistema HLA es la tesis realizada por García [G2003], el trabajo se divide en tres etapas y se basa principalmente en árboles de decisión para obtener las preguntas necesarias para clasificar el órgano del receptor o donador a una biclase.

La primera etapa se enfoca a depurar datos innecesarios, ya que debido a la complejidad de manejar un gran número de datos realiza una limpieza del catálogo de alelos pero sin perder información valiosa. La segunda etapa es la construcción de un árbol de decisión basándose en una división equitativa para obtener las ramas del árbol y así llegar a la punta de la rama. Esta punta significa que se ha llegado a una biclase definida, ya que contesta las preguntas afirmativamente. Y la tercera etapa se divide en dos pasos, el primer paso parte del árbol de decisión de donde obtiene el catálogo de 60 preguntas y



crea identificadores para distinguir entre cada biclase, en caso de que no se pueda clasificar el par de alelos en esta parte se aplica el segundo paso el cual consiste en obtener otro catálogo de preguntas, en este caso se obtienen 109 preguntas sin tomar en cuenta las preguntas que resultaron del primer paso para evitar no encontrar la clasificación de los alelos. En este paso también se buscan los identificadores que van a permitir distinguir a un alelo de dos o más biclases, llegando así a clasificar el 98.21% de los alelos. Cabe mencionar que este trabajo es uno de los primeros intentos por clasificar los alelos HLA por lo que no se toman en cuenta varios aspectos biológicos que ocurren en la realidad y como resultado únicamente permite dar una visión de general de cómo se puede convertir el problema biológico en un problema matemático a través de árboles de decisión.

Otra investigación es realizada por Santillán et al. [SCGGO2005] el cual se basa principalmente en el uso del algoritmo para resolver el “set covering” basado en el método “branch and bound”. Dado que el problema de set covering crece exponencialmente se pueden aplicar algoritmos exactos (Subset of subset pruning, Unique element subset) y algoritmos de aproximación. El trabajo se basa en implementar y examinar dos algoritmos de aproximación: algoritmo voraz mejorado y la programación dinámica.

Se realizaron las pruebas al algoritmo exacto, concluyendo que la solución que éste ofrece se puede optimizar y con respecto al algoritmo voraz se examinó obteniendo una cota inicial en un tiempo razonablemente rápido.



El algoritmo voraz se comparó con la especificación DLV, el cual consiste en programación lógica que intenta procesar código en lenguaje natural. Esta comparación se realizó con el fin de corregir errores que hayan sido generados durante la implementación. Los resultados que se obtuvieron del algoritmo en la mayoría de las veces fueron más rápidos que el DLV, pero en algunas instancias no se presentó esta respuesta.

Estos estudios se aplicaron a los datos del Polimorfismo en la Longitud de los Fragmentos de Restricción (RFLP por sus siglas en inglés) del problema de tipificación del HPV y RFLP y los datos de hibridación del problema de tipificación del HLA-DQ. Los resultados se obtuvieron de diferentes aplicaciones, a los dos primeros problemas se les aplicaron los algoritmos exacto y voraz con una respuesta en un tiempo aceptable y con el mismo resultado óptimo.

Otro trabajo enfocado en el algoritmo para el “set covering” es la tesis de Bravo [B2005] en la cual realiza una evaluación de los datos originales para ir depurando datos que no proporcionen información útil, ya que al manejar un gran número de datos es complicado llegar a resolver este problema de tipificación. Una vez obtenido el nuevo catálogo depurado realiza la creación de preguntas las cuales también depura para evitar que se repita información. Compara su método usado con el de Peña [P2004] obteniendo un catálogo de 20 preguntas con un porcentaje de clasificar un alelo del 99.9995%, pero es importante mencionar que tanto las metodologías como los catálogos usados son diferentes ya que estos últimos pertenecen al año del 2004 en el caso de Peña [P2004] y



el de Bravo [B2005] es del año del 2005, habiendo en el segundo un incremento en el número de alelos.

El algoritmo que utiliza para resolver el set covering es el algoritmo greedy (voraz). Para poder clasificar los alelos con este algoritmo es necesario determinar qué preguntas serán aplicadas, y las obtiene seleccionando el conjunto con la cardinalidad máxima (el conjunto que contenga mayor número de elementos) en cada una de las iteraciones, con el fin de determinar las preguntas con mayor información combinando las que contestan positivo con negativo. También hace uso del DLV para corregir fallas en la implementación y principalmente para encontrar todos los modelos estables de éste problema.

Así como estos trabajos hay otros que utilizan inteligencia artificial basada en agentes como el de Gómez [GF2005], pero este proyecto de tesis se basará principalmente, como ya se mencionó, en el trabajo realizado por Peña [P2004] que se desarrolla a través de la Entropía de Shannon y la Ganancia de Información aplicado al catálogo de alelos HLA-A.

En los siguientes capítulos se presentará la descripción del problema en forma matemática y las bases teóricas para desarrollar el modelo computacional en lenguaje Java que resolverá el problema de la clasificación de alelos para cualquier catálogo.