

HERRAMIENTA COMPUTACIONAL PARA EL ANÁLISIS DE MUTACIONES DEL VIH ASOCIADAS A SUSCEPTIBILIDAD O RESISTENCIA A DROGAS¹

SANDRA MILENA ACERO BARRAZA

*Ingeniera de Sistemas
Universidad del Norte
smacero@uinorte.edu.co*

GUILLERMO JOSÉ CERVANTES ACOSTA

*Ph.D. en Virología e Inmunología
Universidad del Norte
guicerva@uinorte.edu.co*

EDUARDO ENRIQUE ZUREK VARELA

*Ph.D. en Ingeniería
Universidad del Norte
ezurek@uinorte.edu.co*

Fecha de Recibido: 26/03/2009

Fecha de Aprobación: 15/07/2009

RESUMEN

Este artículo presenta el desarrollo e implementación de un sistema que sirva como herramienta para el estudio y exploración del VIH a partir del análisis computacional de sus mutaciones, teniendo en cuenta la información genética del mismo. El análisis se realiza empleando la información contenida en la secuencia genética del VIH y comparándola con secuencias de referencia. Se presenta una descripción de algoritmos para alineamiento, identificación de mutaciones y análisis de resistencia aplicados a la secuencia genética del virus.

PALABRAS CLAVE: Bioinformática, Resistencia, ADN, Aminoácido, Nucleótido, Proteasa, Transcriptasa Inversa.

ABSTRACT

This paper presents the development and implementation of the system that serves as a tool for the study and exploration of Human Immunodeficiency Virus (HIV) from computational analysis of these mutations, taking genetic information itself. The analysis is performed using the information contained in the HIV genetic sequence and comparing it to reference sequences. Descriptions of the algorithms for alignment, mutation identification and resistance analysis applied to the genetic sequence of the virus are presented

KEY WORDS: Bioinformatics, Resistance, DNA, Amino acid, Nucleotide, Protease, Reverse Transcriptase.

1. Resultado de proyecto Joven Investigador financiado por Uninorte y COLCIENCIAS en el marco del convenio 125-2007. Febrero 25 2008 – Febrero 25 2009.

1. INTRODUCCIÓN

Este artículo presenta el desarrollo e implementación de un sistema que sirva como herramienta para el estudio y exploración del VIH a partir del análisis computacional de sus mutaciones, teniendo en cuenta la información genética del mismo.

La tipificación de virus y bacterias es de importancia estratégica para el desarrollo de investigaciones que lleven a la generación de medicamentos para la inmunización o cura de enfermedades.

El VIH se ha caracterizado por su heterogeneidad, debido a una serie de procesos que incluyen sustitución de nucleótidos, supresiones, inserciones y reordenamientos resultantes de eventos de recombinación. Cada vez que un virus infecta una célula, este replica su información genética dando lugar a un nuevo ente capaz de infectar a otras células. En este proceso de réplica ocurren algunos errores no corregidos causando que la copia de la información genética no sea exactamente igual al genoma del virus parental. A este cambio se le denomina mutación [1]. Este hecho es la principal dificultad que se presenta en la búsqueda de una droga antirretroviral exitosa. Es por esto que resulta de gran importancia el desarrollo de pruebas que faciliten el estudio y análisis de las mutaciones del virus para determinar su reacción a determinadas drogas [2]. Lo anterior muestra la importancia que tiene el uso de las tecnologías de la computación en el desarrollo de herramientas para identificar susceptibilidad viral a drogas. Algunos ejemplos de herramientas desarrolladas son: ADRA desarrollada por Los Alamos National Laboratory [3], Geno2Pheno [4] y DR_SEQAN [5]. El desarrollo de este tipo de herramientas es importante no sólo para la creación de nuevas drogas, sino también como instrumento de soporte a decisiones médicas pertinentes a predicción de resistencia a drogas.

La herramienta descrita en este artículo permite realizar, a partir de una muestra de ADN de VIH, un análisis de la resistencia adquirida a algunos medicamentos antirretrovirales, empleando la información contenida en la secuencia genética del virus. La secuencia de ADN de muestra es comparada con el virus tipo salvaje determinando las mutaciones existentes. Estas mutaciones son comparadas con mutaciones previamente establecidas, asociadas con resistencia a determinada droga [6, 7].

2. METODOLOGÍA

2.1 Conceptos generales

Un virus es un ente biológico capaz de auto replicarse, proceso para el cual requiere de una célula a la cual infectar y usar para reproducirse. Cada virus o virión es potencialmente patógeno. La gran mayoría de virus producen en el organismo infectado una reacción inmune, generando anticuerpos que controlan la infección. Las vacunas previenen futuras infecciones debido a que estimulan en el organismo la producción de anticuerpos antes de que el virus lo infecte, esto produce inmunidad durante un tiempo prolongado.

Un virión está constituido por una molécula de ADN o ARN, una envoltura viral, la cual es opcional, una matriz que es la encargada de darle estabilidad estructural al virus, una cápside de proteínas que es la que protege el material genético del virus [8, 9].

Los virus pueden experimentar cambios genéticos a través de varios mecanismos:

EL genoma del virus puede sufrir mutaciones (cambios de nucleótidos en la secuencia de ARN o ADN) que pueden resultar perjudiciales para el proceso de replicación e infección del virus haciéndolo inofensivo, o pueden no producir ningún cambio importante, o puede resultar en ventajas evolutivas que le brinden al virus características tales como resistencia a los fármacos antirretrovirales.

Recombinación genética: es el proceso por medio del cual el genoma de un virus primero se divide y luego se une a una molécula de ADN diferente y ambas recombinan su información genética creando especies de virus nuevas [10].

El VIH pertenece al género Lentivirus de la familia Retroviridae. Se caracteriza por infectar y destruir de forma selectiva los linfocitos T CD4 del sistema inmune o también conocidos como células CD4. Se conocen dos tipos virales que pueden infectar humanos: VIH-1 y VIH-2.

El VIH debido a su característica de altamente cambiante relacionada a la gran cantidad de mutaciones que se producen en el proceso de replicación, no ha permitido a los investigadores hallar una vacuna o medicamento

que permita eliminar su poder de infección. Sin embargo, el uso de algunas drogas antirretrovirales ha permitido que las personas contagiadas extiendan su esperanza de vida por un tiempo mayor al que tendrían de no someterse a estos tratamientos. El objetivo de este tipo de tratamientos es disminuir considerablemente la carga viral a niveles indetectables durante periodos de tiempo lo más largo posibles[11].

Los antirretrovirales contra el VIH inhiben enzimas esenciales como la Transcriptasa Inversa y la Proteasa con lo que reducen la multiplicación del VIH.

La Transcriptasa Inversa se encarga de transcribir el ARN del virus a ADN para que pueda ser incluido en el ADN humano y así infectar la célula. La Proteasa a su vez es la encargada de catalizar el proceso de proteólisis de las poliproteínas para la obtención de proteínas funcionalmente activas y por ende de virus maduros.

Los tratamientos antiproteasa intentan mantener los niveles de infección bajos evitando el cambio de estado de virus inmaduro a virus maduro y contagioso, mientras que los inhibidores de la Transcriptasa Inversa impiden la síntesis de la nueva cadena de ADN[11].

2.2 Tipos de drogas Antirretrovirales

NRTI: Inhibidores Nucleósidos de la Transcriptasa Reversa.

Fueron el primer tipo de inhibidores desarrollados. No garantizan la eliminación total del virus ya que con el tiempo el virus puede mutar y no ser afectado por este tratamiento.

Como se dijo, un virión posee originalmente ARN. Cuando éste infecta una célula, para poder incluir su material genético a ésta requiere transcribir el ARN a ADN, este proceso lo realiza una enzima denominada reversa transcriptasa.

Los NRTI son una versión modificada que imita la estructura química que el VIH necesita para replicarse. Cuando el VIH utiliza un NRTI en vez de una estructura química normal, se bloquea la actividad de la Transcriptasa Inversa deteniéndose el proceso de replicación e infección del virus.

PI: Inhibidores de la proteasa

Como mencionábamos previamente, mediante el mecanismo de proteólisis por parte de la Proteasa se

logra la maduración de los viriones en formación. Los PI desactivan la Proteasa, causando que los viriones sean defectuosos y por tanto no infecciosos.

NNRTI Análogos No Nucleósidos para la Transcriptasa Inversa

Actúan en la misma etapa que los NRTI. Igualmente bloquean la reversa transcriptasa, pero utilizando un proceso químico diferente, consistente en su ligación y posterior desactivación.

2.3 Resistencia del VIH a drogas Antirretrovirales

Se denomina resistencia a la capacidad que posee el virus de replicarse y continuar infectando nuevas células aun ante la presencia de antirretrovirales. La causa de la resistencia se encuentra principalmente en los cambios que experimenta en su genoma, mutaciones. Debido a esto un medicamento antirretroviral no garantiza el control del virus resistente. Los análisis de resistencia ayudan a tomar mejores decisiones en la prescripción de medicamentos de los pacientes.

El VIH se torna resistente de manera aleatoria. Una persona puede infectarse por un virus que ya es resistente a uno o más medicamentos. Cuanto más se multiplica en VIH más mutaciones aparecen. Cuando estas mutaciones están localizadas en ciertas posiciones dentro de algunos genes han sido vinculadas con la resistencia a los medicamentos[12].

Investigadores de la sociedad Internacional del SIDA (IAS-USA) han logrado una recopilación que incluye mutaciones asociadas con resistencia viral a drogas antirretrovirales proporcionadas a pacientes que conviven con VIH [13]. Esta información ha sido obtenida a través de: pruebas de susceptibilidad realizadas en laboratorios aislados; secuenciamiento de nucleótidos de muestras de pacientes en los que la droga no ha resultado exitosa; estudios de correlación en pacientes expuestos a la droga.

2.4 Descripción de la herramienta de software

Teniendo en cuenta lo anterior se diseñó una interfaz de software stand-alone que realiza análisis de resistencia sobre una secuencias de ADN.

El programa fue desarrollado en Python versión 2.6 bajo el sistema operativo Microsoft XP. El desarrollo se realizó a través de programación orientada a objetos

y se empleó el patrón de arquitectura Model-View-Controller (MVC).

Esta aplicación permite que el usuario obtenga como resultado una tabla con las drogas hacia las cuales la muestra analizada ha desarrollado resistencia. Esta información puede servir de apoyo para la determinación de tratamientos a pacientes con virus similares al de la muestra.

El software consta de tres módulos. La distribución de los módulos corresponde a la implementación del modelo MVC (Model-View-Controller):

Un módulo GUI.py que genera la interfaz gráfica del programa.

Un módulo que realiza las operaciones necesarias para analizar la resistencia.

Un módulo que maneja la conexión a la base de datos para hallar las mutaciones.

2.5 Algoritmo para alineación: Needleman-Wunsch

La alineación de secuencias en bioinformática es una forma de comparar y representar la similitud de dos o más secuencias de ADN, ARN o proteína.

En general, existen dos tipos de métodos para realizar alineamiento: alineamiento global y alineamiento local. El alineamiento global es una forma de optimización global que obliga a la secuencia a alinear a ocupar la longitud total de toda la secuencia de referencia, mientras que el alineamiento local identifica regiones similares dentro de las secuencias. El alineamiento local es más útil en el caso en que las secuencias a alinear sean de longitudes similares, y, el local, cuando las secuencias son muy divergentes[14]

Los métodos de alineamiento también se pueden clasificar en: Alineamiento de pares (pairwise alignment) y Alineamiento múltiple (Multiple alignment), dependiendo de la cantidad de secuencias a alinear. Los métodos de alineamiento de pares se usan para alinear dos secuencias, mientras que los de alineamiento múltiple se usa para alinear múltiples secuencias simultáneamente[15].

El método para alineamiento global de secuencias implementado en la herramienta es el propuesto por Needleman y Wunsch en 1970[16]. Es un método de

alineamiento de pares para alineamiento global basado en programación dinámica.

El algoritmo consta de tres pasos:

Paso uno del algoritmo para alineación: Inicialización

Para realizar el alineamiento la herramienta requiere de dos secuencias: una secuencia de referencia R elegida por el usuario entre: HXBR, Consensus B, SF2 y RF4, y la secuencia de entrada E, cargada de un archivo en formato Fasta.

En el proceso de inicialización se crea una matriz de tamaño $n+1 \times m+1$ donde n corresponde a la longitud de la secuencia E y m a la longitud de R. Primero se inicializan la primera fila y columna de la matriz con:

$$D(i, j) = \begin{cases} 2 * i, & \text{si } j = 0 \\ 2 * j, & \text{si } i = 0 \end{cases} \quad (1)$$

A continuación se inicializa con 1 (uno) aquellas posiciones que corresponden a letras iguales en ambas secuencias y cero en las demás posiciones. Ejemplo: R = ATCAGAGTC y E= TTCAGTC

longitud de R = 9 y longitud de E = 7.

Tabla 1. Paso 1 del algoritmo de inicialización

	A	T	C	A	G	A	G	T	C
0	-1	-2	-3	-4	-5	-6	-7	-8	-9
T	-1	0	1	0	0	0	0	1	0
T	-2	0	1	0	0	0	0	1	0
C	-3	0	0	1	0	0	0	0	1
A	-4	1	0	0	1	0	1	0	0
G	-5	0	0	0	0	1	0	1	0
T	-6	0	1	0	0	0	0	1	0
C	-7	0	0	1	0	0	0	0	1

Paso dos del algoritmo para alineación: Llenado de la matriz

Para llenar la matriz de similitudes se tienen en cuenta lo siguiente:

$$D(i, j) = \max \begin{cases} D(i-1, j-1) + S(x_i, y_j) \\ D(i-1, j) + g \\ D(i, j-1) + g \end{cases}, \quad (2)$$

$i, j \geq 1$ (2)

Donde i, j es la posición i, j de la matriz de similitudes, es el valor de substitución de dado por:

$$s(x_i, y_j) = \begin{cases} 1, & \text{si } x_i = y_j \\ 0, & \text{de otra forma} \end{cases} \quad (3)$$

Y g corresponde a un valor de penalización (gap penalty), $g = 1$.

El cálculo de un valor para una posición i,j dada depende de los valores inmediatamente adyacentes: arriba, izquierda y diagonal superior izquierdo.

Tabla 2. Cálculo de un valor i,j en la matriz

C(i-1,j-1)	C(i-1,j)
C(i,j-1)	C(i,j)

Para la matriz de ejemplo dada:

Tabla 3. Paso 2 del algoritmo de inicialización

	A	T	C	A	G	A	G	T	C	
T	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
T	-1	0	0	-1	-2	-3	-4	-5	-6	-7
T	-2	-1	1	0	-1	-2	-3	-4	-4	-5
C	-3	-2	0	2	1	0	-1	-2	-3	-3
A	-4	-2	-1	1	3	2	1	0	-1	-2
G	-5	-3	-2	0	2	4	3	2	1	0
T	-6	-4	-2	-1	1	3	4	3	3	2
C	-7	-5	-3	-1	0	2	3	4	3	4

Paso tres del algoritmo para alineación: Búsqueda hacia atrás - traceback

Una vez se han llenado todas las posiciones de la matriz de similitudes. Se calcula el traceback, es decir, la deducción de la mejor alineación de la matriz. Siempre comienza por la última celda que se llenó. A través del cálculo de los valores óptimos se determina un camino hasta llegar a la celda superior izquierda. Para esto, existen tres formas posibles de moverse: hacia arriba, hacia la izquierda o en diagonal hacia arriba-izquierda.

	A	T	C	A	G	A	G	T	C	
T	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
T	-1	0	0	-1	-2	-3	-4	-5	-6	-7
T	-2	-1	0	-1	-2	-3	-4	-4	-5	-5
C	-3	-2	0	2	1	0	-1	-2	-3	-3
A	-4	-2	-1	1	3	2	1	0	-1	-2
G	-5	-3	-2	0	2	4	3	2	1	0
T	-6	-4	-2	-1	1	3	4	3	3	2
C	-7	-5	-3	-1	0	2	3	4	3	4

Ilustración 1. Traceback

Cada avance horizontal indica un hueco y cada avance diagonal indica no cambio. Para la secuencia de ejemplo el alineamiento resultante es:

A	T	C	A	G	A	G	T	C
T	T	-	C	-	A	G	T	C

2.6 Método para determinar resistencia.

El método implementado para determinar la resistencia consta de los siguientes pasos:

Paso uno del método para determinar resistencia: Lectura de la secuencia a analizar.

Para identificar la resistencia adquirida a ciertos medicamentos de una muestra de ADN de VIH, la herramienta requiere: una secuencia de referencia R elegida por el usuario entre: HXBR, Consensus B, SF2 y RF4, y la sección correspondiente a la proteína pol de la secuencia muestra M, cargada de un archivo en formato Fasta.

Paso dos del método para determinar resistencia: Traducción de secuencia de ADN a secuencia de aminoácidos.

La secuencia de ADN M leída en el paso anterior es traducida de nucleótidos a aminoácidos de acuerdo con el código genético[17]

Paso tres del método para determinar resistencia: Búsqueda de mutaciones en la secuencia.

La secuencia M es comparada contra una cepa de referencia correspondiente a la secuencia de referencia R elegida por el usuario entre: HXBR, Consensus B, SF2 y RF4. A partir de esta comparación se identifican las mutaciones, es decir, las posiciones y aminoácidos que cambiaron en la secuencia M a partir de la secuencia de referencia R.

Paso cuatro del método para determinar resistencia: Carga de mutaciones de la Base de Datos.

La herramienta utiliza una base de datos en la cual se encuentra la información correspondiente a las mutaciones y las drogas para las cuales causan resistencia. En este paso se carga esta información.

Paso cinco del método para determinar resistencia: Construcción de la tabla de resultados.

A partir de las mutaciones determinadas en el paso tres, se identifican cuales son determinan resistencia a antirretrovirales. Con esta información se crea una tabla con la droga y la correspondiente mutación hallada en la muestra analizada.

2.7 Método para determinar mutaciones

Las mutaciones no son más que los cambios de los nucleótidos en la secuencia de ARN o ADN del virus. Para hallar las mutaciones presentes en una muestra es necesario compararla con una cepa de referencia y hallar las diferencias entre ambas secuencias.

3. IMPLEMENTACIÓN Y RESULTADOS

Se desarrollaron módulos en Python para realizar análisis de resistencia a drogas de un segmento de determinada secuencia de ADN de VIH. Para lo cual se realizó: primero, la traducción de la secuencia de ADN (nucleótidos) a aminoácidos teniendo en cuenta el código genético. Segundo, la búsqueda de las mutaciones presentes en la secuencia de entrada con respecto a una secuencia de referencia, que puede ser elegida entre las siguientes: HXBR, CONSENSUS B, SF2 y RF4 [18, 19]. Tercero, conexión a una base de datos local que contiene una tabla que relaciona las mutaciones y la correspondiente droga para la cual hace al virus resistente. Cuarto, Visualización de la lista de las drogas de hacia las cuales la muestra analizada ha generado resistencia.

Además de lo anterior, el software cuenta con un módulo que realiza alineamiento global de una secuencia con respecto a una secuencia de referencia que puede ser elegida entre las siguientes: HXBR, CONSENSUS B, SF2 y RF4. Para esto, se implementó el algoritmo de Needleman-Wunsch[20] para alineamiento global. Adicionalmente, se cuenta con módulos para realizar análisis filogenético de varias secuencias de ADN cargadas a partir de un archivo en formato Fasta y dando como resultado, un árbol de análisis filogenético. Esto se logró a través de la implementación de los algoritmos: UPGMA, Parsimony y Neighbor-Joining para análisis filogenético[21]. Algo que vale la pena comentar es la

importancia de alinear. El análisis de resistencia y en general cualquier estudio realizado sobre secuencias genéticas, implica que las secuencias se encuentren en una posición dada, es decir, alineadas. La alineación implica encontrar la posición óptima para conseguir un mejor resultado en el análisis de la secuencia.

Los resultados desde el punto de vista computacional son: se desarrollaron clases para el manejo básico de secuencias genéticas; se desarrollaron clases para conexión a base de datos MySQL; se desarrollaron clases muy flexibles y escalables; se incorporaron paquetes para Python disponibles bajo licencia GPL; cada uno de los paquetes desarrollados está debidamente documentado.

Los resultados desde el punto de vista de aplicaciones para el usuario son: se toman segmentos de interés en secuencias genéticas de nucleótidos y se transforman a secuencias de aminoácidos[22]; se tiene la opción de alinear secuencias de nucleótidos o de aminoácidos con el método Needleman-Wunsch, contra una secuencia de referencia que pueden ser elegida de entre las siguientes: HXBR, CONSENSUS B, SF2 y RF4; las secuencias de aminoácidos se comparan contra una secuencia de referencia que pueden ser elegida de entre las siguientes: HXBR, CONSENSUS B, SF2 y RF4; con base en la comparación se identifican las mutaciones; con las mutaciones identificadas se busca en una base de datos y se correlaciona la mutación y la droga hacia la cual hay resistencia.

La dinámica de las pruebas realizadas al software fue la siguiente, primero, se realizaron pruebas sobre cada uno de los componentes de manera individual y se verificó el funcionamiento de cada uno de los paquetes. Estas pruebas se realizaron durante el desarrollo y una vez finalizado el proceso de programación se repitieron.

Se ejecutaron pruebas en consola sobre cada una de las clases y finalmente sobre el paquete en general. En esta etapa se generaron errores de falta de memoria al procesarse datos de secuencias de 1000 o más nucleótidos, pero esto logró superarse. Posteriormente se requirió el rediseño de algunas fórmulas y procedimientos para superar las limitaciones.

A continuación se presenta una comparación con otras herramientas similares:

Tabla 4. Comparación con otras herramientas.

Característica	Herramienta propuesta	ADRA	Geno2Pheno	DR_SEQAN
Plataformas Soportadas				
Windows XP, Vista	✓	✓	✓	✓
Linux	✓	✓	✓	
Especificaciones mínimas				
Memoria Ram	512 MB	512 MB	512 MB	512 MB
Espacio requerido	1GB	1GB	1GB	1GB
Procesador	1.8 GHz x86	1.8 GHz x86	1.8 GHz x86	1.8 GHz x86
Conexión a Internet		✓	✓	
Otras características				
Adecuación previa de la secuencia	✓	✓	✓	✓
Formato requerido de la secuencia	FASTA o txt	FASTA o digitación directa	FASTA	FASTA o txt
Persistencia de los resultados	✓	24 Hrs.		
Confidencialidad de resultados	Cómo los datos se manejan localmente, la confidencialidad de los datos depende del manejo dado por el usuario de la herramienta	Los datos viajan por una conexión web no segura. Es responsabilidad del usuario no enviar información confidencial.	Los datos viajan por una conexión web no segura. Es responsabilidad del usuario no enviar información confidencial.	Los datos viajan por una conexión web no segura. Es responsabilidad del usuario no enviar información confidencial.

4. CONCLUSIONES

En este proyecto se diseñó e implementó una interfaz de software para análisis e interpretación de mutaciones del VIH para determinar susceptibilidad a drogas. Como resultado de este proyecto, se presenta un sistema de información diseñado para identificar posible resistencia a ciertas drogas antirretrovirales asociadas a cambios en la proteasa y la reversa transcriptasa. El sistema permite procesar secuencias de ADN de VIH de hasta 5000 nucleótidos de longitud. Sin embargo, para realizar análisis de mutaciones se requiere como dato principal una secuencia de muchísima menor longitud que la que el software es capaz de procesar, debido a que dicho análisis se realiza a partir de la sección correspondiente a la proteasa o la reversa transcriptasa del virus.

5. REFERENCIAS

[1] C. Dykes, K. Fox, A. Lloyd, M. Chiulli, E. Morse, and L. M. Demeter, "Impact of clinical reverse transcriptase sequences on the replication capacity of HIV-1 drug-resistant mutants," *Virology*, vol. 285, pp. 193-203, 2001.

[2] L. A. N. Laboratory, "ADRA: Antiviral Drug Resistance Analysis," Los Alamos National Laboratory, 2007.

[3] S. J. Little and S. Holte, "Antiretroviral-Drug Resistance among Patients Recently Infected with HIV," *The New England Journal of Medicine*, vol. 347, pp. 385-394, Agosto 8 2002 2002.

[4] E. K. Wagner and M. J. Hewlett., Basic Virology, 2, Illustrated ed.: Blackwell Publishing, 2004.

[5] Mount and W. David, "Bioinformatics: Sequence and Genome Analysis," New York: Cold Spring Harbor Laboratory Press., 2001.

[6] M. H. Schierupa and J. Hein, "Consequences of Recombination on Traditional Phylogenetic Analysis," *Genetics*, vol. 156, pp. 879-891, Octubre 2000 2000.

[7] M. Schena, "DNA Microarrays: A practical Approach," O. U. Press, Ed., 1999, p. 232.

[8] G. C. Menéndez-Arias, "DR_SEQAN: a PC/Windows-based software to evaluate drug," *BMC Infectious Diseases*, vol. 6, 2006.

[9] H. Tao, "Drug Resistance Mutations Superimposed on the Structures of HIV-1 Protease and Reverse Transcriptase," in *HIV Molecular Immunology Database New Mexico: Theoretical Biology and Biophysics Group*, 1997.

[10] D. L. Hartl, *Essential Genetics: A Genomics Perspective*, 3 ed.: Jones and Bartlett Publishers, 1995, 1995.

[11] B. N. Fields, D. M. Knipe, P. M. Howley, and D. E. Griffin, *Fields' virology*, 5 ed.: Lippincott Williams & Wilkins, 2006.

[12] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *J. Mol. Biol.*, vol. 48, pp. 443-453, 1970.

[13] N. Beerenwinkel, T. Lengauer, J. Selbig, B. Schmidt, H. Walte, K. Korn, R. Kaiser, and D. Hoffmann., "Geno2pheno: interpreting genotypic HIV drug resistance tests," *Intelligent Systems, IEEE*, vol. 16, pp. 35-41, nov - dic 2001 2001.

[14] C. Pasquier, N. Millot, R. Njouom, K. Sandres, M. Cazabat, J. Puel, and J. Izopet, "HIV-1 subtyping using phylogenetic analysis of pol gene sequences," *Journal of Virological Methods*, vol. 92, pp. 45-54, Abril 30 2001 2001.

[15] M. S. Hirsch, "HIV Drug Resistance - A Chink in the Armor," *The New England Journal of Medicine*, vol. 347, pp. 438-439, Agosto 8 2002 2002.

[16] C. Yabar, P. Chavez, Z. Varas, and R. Rodriguez, "Identificación molecular de mutaciones puntuales relacionadas con resistencia a drogas en VIH-1 de pacientes peruanos," *Revista Peruana de Medicina Experimental y Salud Publica*, vol. 23, pp. 149-157, Julio - Septiembre 2006 2006.

- [17]N. J. Dimmock, A. Easton, and K. Leppard, *Introduction to Modern Virology*, 5, Illustrated ed.: Blackwell Publishing, 2001.
- [18]R. G. Cotton, E. Edkins, and S. Forrest, "Mutation Detection: A practical Approach," Oxford University Press, 1999, p. 264.
- [19]S. Clark, C. Calef, and J. Mellors, "Mutations in Retroviral Genes Associated with Drug Resistance " in *HIV Sequence Compendium 2006/200*, F. B. Thomas Leitner T, Hahn B, Marx P, McCutchan F, Mellors J, Wolinsky S, Korber B., Ed.: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM. LA-UR 07-4826, 2007, pp. 58-158.
- [20]W. M. Fitch and T. F. Smith, "Optimal Sequence Alignments," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 80, pp. 1382-1386, 1983.
- [21]M. M. Miyamoto and J. Cracraft, "Phylogenetic Analysis of DNA Sequences," New York: Oxford University Press, Incorporated, 1991, p. 369.
- [22]D. Costagliola, L. Assoumou, and L. Morand-Joubert, "Prevalence of HIV-1 drug resistance in treated patients - A French nationwide study," *JAIDS-Journal Of Acquired Immune Deficiency Syndromes*, vol. 46, pp. 12-18, septiembre 1 2007 2007.
- [23]J. M. Coffin, S. H. Hughes, and H. E. Varmus, *Retroviruses*, illustrated ed.: CSHL Press, 1997.
- [24]Y. Xing, "Sequence Alignment Algorithm," in *School of electronics Engineering and Computer Science Beijing: Peking University*, 2005.
- [25]A. Abecasis, A. Vandamme, and P. Lemey, "Sequence Alignment in HIV Computational Analysis " in *HIV Sequence Compendium 2006/2007: Theoretical Biology and Biophysics Group*, Los Alamos National Laboratory, Los Alamos, NM. LA-UR 07-4826, 2007.
- [26]S. Yerly, L. Kaiser, E. Race, J. Bru, F. Clavel, and LPerrin, "Transmission of antirretroviral drug-resistant HIV-1 variants," *Lancet*, vol. 354, p. 729, 1999.
- [27]V. A. Johnson and F. Brun-Vézinet, "Update of the drug Resistance Mutations in HIV-1," *Topics in HIV Medicine*, vol. 16, pp. 62-68, 2008.
- [28]W. E. Highsmith, Q. Jin, A. J. Nataraj, J. M. O'Connor, V. D. Burland, W. R. Baubonis, F. P. Curtis, N. Kusukawa, and M. M. Garner, "Use of a DNA toolbox for the characterization of mutation scanning methods. I: Construction of the toolbox and evaluation of heteroduplex analysis," *Electrophoresis*, vol. 20, pp. 1186-1194, Mayo 26 1999 1999.
- [29]V. G. S. Merino, "Variacion de los genes accesorios y su implicacion en la biología del vih-1," in *Departamento de microbiología II. vol. Doctor Madrid: Universidad Complutense de Madrid*, 2001, p. 700.

6. AGRADECIMIENTOS

Este artículo es resultado del proyecto Joven Investigador financiado por Uninorte y COLCIENCIAS en el marco del convenio 125-2007. Febrero 25 2008 – Febrero 25 2009.

7. CURRÍCULUM

Sandra Milena Acero Barraza, Ingeniera de Sistemas, Universidad del Norte. Joven investigadora. Docente catedrático universidad del Norte. smacero@uninorte.edu.co

Guillermo José Cervantes Acosta, Químico farmacéutico, Universidad del Atlántico, Ph.D. en Virología e Inmunología, Universidad de Montreal (Canadá). Docente tiempo completo Universidad del Norte. guicerva@uninorte.edu.co

Eduardo Enrique Zurek Varela, Ingeniero de Sistemas, Universidad del Norte, Ph.D. Dpto. de Ingeniería Eléctrica, University of South of Florida (Tampa, Florida). Docente tiempo completo Universidad del Norte. ezurek@uninorte.edu.co