

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

FACULTAD 6



**TÍTULO: “SISTEMA DE ANÁLISIS DE SECUENCIAS
MRNA PARA EL DISEÑO DE siRNA. V 1.1”**

Autores:

Naivys Tirado Ruíz

Kader Antonio Román Stable

Tutores:

Dr. Ricardo Bringas Perez

Ing. Vladir Antonio Parrado Cruz

Lic. Milena Ossorio Lamí

Ciudad de la Habana, Cuba, Junio de 2008

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Nayvis Tirado Ruiz

Kader Antonio Román Stable

Ricardo Bringas Perez

Vladir Antonio Parrado Cruz

Milena Ossorio Lamí

DATOS DEL CONTACTO

Tutor:

Dr. Ricardo Bringas Perez

Centro de Ingeniería Genética y Biotecnología, Ciudad de La Habana, Cuba.

Email: ricardo.bringas@cigb.edu.cu

Tutor:

Ing. Vladir Antonio Parrado Cruz

Universidad de las Ciencias Informáticas, Ciudad de La Habana, Cuba.

Email: vparrado@uci.cu

Tutor:

Lic. Milena Ossorio Lamí

Universidad de las Ciencias Informáticas, Ciudad de La Habana, Cuba.

Email: milena@uci.cu

Agradecimientos

A nuestro padre del cielo por permitir este sueño. Por ser la luz del camino.

A nuestros padres y familiares por su apoyo constante y su confianza en nosotros.

A la Revolución por darnos la oportunidad de realizar nuestros sueños.

A los profesores de la Universidad de las Ciencias Informáticas por su dedicación y entrega.

A aquellos que ya no están y a los que nos han acompañado a lo largo de los cinco años de carrera. Especialmente a los profes Vladir y Milena por su ayuda incondicional y preocupación.

A Vilmaivis, Alieski y Liudmila por tanta ayuda en esta recta final.

A nuestros amigos y compañeros de la vida por el apoyo brindado aún en situaciones difíciles.

A los especialistas del CIGB Alexander y Ricardo Bringas por sus consejos y su ayuda.

A todos los que de una forma u otra contribuyeron a nuestra formación profesional.

Dedicatoria

A mis abuelos Pilar, Tatín, Blanquita y Carlos. Ustedes han sido fuente de mi inspiración.

A mi mamá, por estar siempre y ser tan preocupada conmigo. Por apoyarme tanto. Por siempre haber sido mamá.

A mi papá, por ser el personaje principal de esta carrera, de estos cinco años. Por nunca haberme dado la espalda en nada. Eres el mejor padre del mundo.

A mi hermano, por su cariño y atención conmigo. Por siempre haber estado a mi lado en cualquier circunstancia.

A mi cuñada, por sus actos y preocupaciones conmigo.

A toda mi familia, por su generosidad a lo largo de estos cinco años.

A mis amigos del alma, Ariannis (mi mejor amiga), Magdiel (mi mejor amigo), Juan Antonio, Antúnez, Luna y Noemi. Por las palabras de aliento que nunca faltaron; por la mano amiga que siempre estuvo y por el cariño con que me han tratado.

A mis hermanos de la fe en la universidad, son especiales y siempre los serán. Tienen un lugar exclusivo en mi corazón. Especialmente a mi grupo pequeño de amigos llamado clan. A estas personas tan especiales, Daimara, Ariannis, Juan Antonio y Mariannis, los amo mucho y nunca los olvidaré, han calado mi corazón.

A mis hermanos de la fe que no están en la universidad. A Antúnez (hijo), a Noemi, Dayana, Gikleryz, Antúnez (padre), Marisol, Nena, Odalis, Samuel, Dorca, Henry, Alexander, Alexei, Alejandro, Julito, Sonia y Adrian. Ustedes han marcado una diferencia en mi vida.

A mis compañeros y amigos de estudio, nunca los olvidaré, especialmente a Yaniel, Yuneimy, Cantillo, Yaisel, Persy y Adrián (el congri), los quiero mucho.

Kader Antonio Román Stable

A mi mamita del alma Vivian, a mis abuelitos Anadelia y Reynaldo y a mi tía Ninfa por estar ahí siempre, por hacer el papel de padre y madre. Ustedes son el motivo de mis mayores esfuerzos.

A mi papá por ser mi ejemplo, anhelo y guía en este batallar por llegar a la cima.

A mis hermanitos Luisi y Angel Luis y a mi prima Dayi. Los quiero con la vida.

A toda mi familia en general por su apoyo y preocupación a los largo de estos cinco años.

A mis hermanitas de la escuela Yadi, Adita, Yali, Yane y Yeni por estar ahí cuando las lágrimas no ayudaban, por compartir tanto. ¡Las quiero mucho!

Naiuys Tirado Ruiz

Resumen

El uso de los RNA pequeños de interferencia (siRNA) para el silenciamiento de genes se perfila como una poderosa herramienta terapéutica para combatir enfermedades no solo en humanos, sino también en otros organismos. Un componente fundamental de esta tecnología lo constituyen las herramientas informáticas para el diseño de los siRNA. A pesar de la existencia en Internet de aplicaciones que permiten realizar este diseño, el Centro de Ingeniería Genética y Biotecnología (CIGB) y la Universidad de las Ciencias Informáticas (UCI) se han dado a la tarea de crear una herramienta capaz de dar solución a los problemas que traen consigo el uso de estas herramientas, con una mayor seguridad de los resultados que se obtengan. Para ello se hizo una primera versión del sistema que cumplía con las funcionalidades requeridas por los especialistas en una primera etapa de definición del proyecto. Luego surgió la necesidad de extenderlo a una segunda versión, tomando como base la anterior, con el objetivo de hacerle mejoras al proceso de diseño e incluir nuevas funcionalidades.

Índice

Introducción	1
Capítulo 1 Fundamentación teórica	4
Introducción.....	4
1.1 La informática y su relación con las ciencias biológicas.....	4
1.1.1 miRNA y siRNA [11]. Semejanzas y diferencias	5
1.1.2 mRNA	5
1.1.3 SNP	6
1.2 Sistemas existentes para el diseño de siRNA.....	6
1.2.1 Target Finder siRNA design [3].....	6
1.2.2 SiRNA Design Software [4]	6
1.2.3 Dharmacon siDESIGN Center [5].....	7
1.2.5 DEQOR [6].....	7
1.2.6 IDT's RNAi Design [7].....	7
1.2.7 GenScript siRNA Target Finder [8]	7
1.2.4 siDirect [9].....	8
1.3 Limitaciones de los sistemas existentes	8
1.4 Tecnologías, metodologías y herramientas	9
1.4.1 Tecnología J2EE	9
1.4.2 Lenguaje JAVA.....	10
1.4.3 Herramientas de desarrollo utilizadas	11
1.4.4 Gestor de Base de Datos	12
1.4.5 Servidor de Aplicaciones	12
1.4.6 Lenguaje de Modelado (UML)	13
1.4.7 Proceso de desarrollo de software (OpenUp).....	13
1.5 Resumen de la tecnología a utilizar	13
1.6 Conclusiones	14
Capítulo 2 Características del sistema	15
Introducción.....	15
2.1 Objeto de estudio	15
2.1.1 Objetivos estratégicos de la organización.	15
2.2 Procesos objeto de automatización.	16
2.2.1 Procesos que serán objeto de automatización.	16

2.3 Propuesta de sistema.....	18
2.3.1 Descripción general de la propuesta de sistema	18
2.3.2 Análisis comparativo de otras soluciones existentes con la propuesta	18
2.4 Especificación de los requisitos de software	19
2.4.1 Requerimientos Funcionales.....	19
2.4.2 Requerimientos no funcionales	20
2.5 Definición de actores y casos de uso del sistema.....	22
2.5.1 Actores.....	22
2.6.2 Casos de Uso definidos	23
2.5.3 Diagrama de casos de uso del sistema	24
2.5.4 Descripción detallada de casos de usos.	25
2.7 Conclusiones	32
Capítulo 3 Análisis y diseño del sistema.....	33
Introducción.....	33
3.1 Arquitectura de Software del sistema.....	33
3.2 Objetivos del diseño	37
3.3 Diagramas de Clases del Diseño.....	37
3.4 Comparación de diseño.....	43
3.5 Diagramas de Interacción (Secuencia, Colaboración)	45
3.6 Diseño de la Base de Datos	45
3.6.1 Modelo lógico de datos	46
3.6.2 Modelo físico de datos	47
3.7 Descripción de las Tablas de la Base de Datos.	48
Capítulo 4 Implementación y prueba	57
4.1 Diagrama de Componentes.....	57
4.1.1 Diagramas de Componentes de la aplicación	57
4.2 Optimización del costo de ejecución del algoritmo de búsqueda de sitios blancos	61
4.3 Obtención del algoritmo para la búsqueda de sitios blancos en la secuencia de mRNA teniendo en cuenta los SNP.....	63
4.4 Modelo de Prueba	63
4.4.1 Casos de prueba de caja negra	63
4.5 Conclusiones.....	74
Conclusiones Generales	75
Recomendaciones	76

Referencias bibliográficas	77
Bibliografía	80
Anexos	82
Glosario de Términos	97

Introducción

El silenciamiento de genes mediante los RNA pequeños de interferencia (siRNA) se perfila como una poderosa herramienta terapéutica en el futuro. Hace aproximadamente nueve años, en estudios de otra índole, se descubrió la existencia de genes que permitirían el silenciamiento de otros genes; es decir, que serían capaces de inhibir su acción en el organismo. Ahora se sabe que está conservado en casi todos los organismos, por lo que puede ser utilizado como una forma de terapia génica: si cierta enfermedad es el efecto del aumento en la expresión de un gen específico y se conoce su identidad, se podría “apagar” su expresión por medio de inyecciones de RNA pequeños de doble cadena y curar así la enfermedad. En estudios más profundos se dedujo, que la creación de forma artificial de un RNA como el siRNA sería una de las principales herramientas experimentales para combatir y buscar soluciones a problemas de enfermedades en animales, plantas y el genoma humano. [1]

Actualmente, existe un gran número de científicos y especialistas que realizan estudios sobre el diseño de siRNA y su aplicación en el silenciamiento de genes sobre la base de herramientas online publicadas en Internet, de las cuales podemos mencionar a: Target Finder SiRNA design tools [2], SiRNA Design Software [3], Dharmacon siDESIGN Center [4], DEQOR [5], RNAi Design [6], y GenScript siRNA Target Finder [7], siDirect [8]. Uno de los problemas de la utilización de tales herramientas es que no se conocen los algoritmos que se han utilizado para la predicción de los siRNA y no son configurables para el uso de algoritmos propios. No dan la posibilidad de hacer un análisis con el por ciento de exactitud que se estime conveniente de acuerdo al estudio que se esté realizando y no existe una unidad en las ideas expuestas en estos sitios, razón que provoca que haya diversidad sobre el mismo tema.

Lo anteriormente planteado se deriva de la carencia de un software que reúna todas las condiciones y cumpla con los requisitos de diseño que desean los especialistas, brindándoles cualquier servicio que necesiten sobre la base de sus investigaciones.

Hoy, los especialistas cubanos no cuentan con una herramienta propia capaz de dar solución al tema tratado sobre el diseño de siRNA. Para satisfacer esta necesidad el centro de Ingeniería Genética y Biotecnología (CIGB) y la Universidad de las Ciencias Informáticas (UCI) se han dado a la tarea de crear una aplicación capaz de dar solución a este problema. Para ello se hizo una primera versión de esta con las funcionalidades requeridas por los especialistas para aquella etapa, pero no con la capacidad de dar solución a todas las problemáticas planteadas en este tema de diseño de siRNA. [9]

Para una versión superior del sistema, donde se plantea un nuevo diseño de implementación, perfilar

las funcionalidades logradas e implementar nuevas funcionalidades, el trabajo se orienta a **¿Cómo automatizar el proceso de diseño de siRNA?**

Para la obtención de un siRNA específico, a través de las investigaciones de los especialistas, de manera eficiente y en el tiempo requerido, se hace necesario estudiar el **proceso de análisis de secuencia mRNA para el diseño de siRNA**. Para esto, se incide en las aplicaciones informáticas Cliente-Servidor para el diseño de un siRNA. Para dar solución al problema señalado se traza como **objetivo general** desarrollar la versión 1.1 del sistema de análisis de secuencia mRNA para el diseño de siRNA.

Para dar cumplimiento al objetivo general del trabajo se trazaron los siguientes objetivos específicos:

- Optimizar el costo de ejecución del algoritmo de búsqueda de sitios blancos implementado en la versión anterior.
- Rediseñar el sistema.
- Obtener el algoritmo para búsqueda de sitios blancos en la secuencia de mRNA teniendo en cuenta los polimorfismos de nucleótido simple (SNP).
- Implementar el sistema.

Para cumplir con dichos objetivos, se trazaron las siguientes **tareas**:

- Análisis del algoritmo de búsqueda de sitios blancos en secuencias mRNA, implementado en la versión anterior, en busca de mejoras.
- Implementación de un algoritmo más eficiente para la búsqueda de sitios blancos en secuencias mRNA.
- Confección del diagrama de casos de uso del sistema
- Confección de los diagramas de clases del diseño para cada caso de uso y los diagramas de interacción por cada escenario de los mismos.
- Diseño de una nueva interfaz de usuario que incorpore nuevas funcionalidades.

- Introducción en la base de datos de la información de los SNP.
- Diseño del algoritmo de búsqueda de sitios blancos en secuencias mRNA teniendo en cuenta los SNP.
- Introducción en la base de datos de la información de los genes capaces de regular la expresión de otros genes de manera natural (miRNA).
- Confección de diagramas de componentes.
- Implementación de la aplicación haciendo uso del patrón de arquitectura Controlador Frontal.
- Realización de pruebas al sistema.

Se pretende terminar una versión estable del Sistema de análisis de secuencia mRNA para el diseño de siRNA, incluyendo nuevas funcionalidades y perfeccionando lo que se ha hecho hasta el momento para conseguir una mayor velocidad y eficiencia. Se enfoca el trabajo principalmente hacia la implementación del algoritmo que dadas la reglas a aplicar y la secuencia para analizar, devuelva los sitios blancos candidatos para el diseño de siRNA. Este trabajo incluye las búsquedas en la base de datos y en los ficheros que contienen la información de los cromosomas del genoma humano, para darle a los especialistas la posibilidad de diseñar, a partir de los resultados obtenidos, un siRNA específico para un gen analizado.

De forma general, la herramienta debe:

- Tratar todo lo referente a la información de los genes y transcritos.
- Ensamblar la secuencia de mRNA a partir de la información almacenada.
- Analizar una secuencia mRNA dada.
- Mostrar, luego del análisis, los sitios posibles a silenciar (sitios blancos) y una representación gráfica de la localización de los mismos en dicha secuencia.
- Mostrar los sitios blancos teniendo en cuenta los SNP.
- Mostrar información de los miRNA.

Capítulo 1

Fundamentación teórica

Introducción

En este capítulo se hace un análisis de la relación de la informática con las ciencias biológicas. Se presentan los fundamentos teóricos y prácticos que cimientan las soluciones escogidas. Se aborda la importancia del desarrollo de la aplicación. Además se mencionan y analizan una serie de aplicaciones ya desarrolladas vinculadas con el tema de diseño de siRNA. Se fundamentan las tendencias y tecnologías actuales utilizadas para el logro de una herramienta o producto con calidad.

1.1 La informática y su relación con las ciencias biológicas.

Durante la última década del siglo XX, los avances de la ingeniería genética y las nuevas tecnologías de la información, condicionaron el surgimiento de una disciplina que creó vínculos indisolubles entre la Informática y las ciencias biológicas: la Bioinformática.

Ella es la combinación de las ciencias de la vida con las ciencias de la información. Es un campo científico que se propone la investigación y el desarrollo de sistemas que faciliten la comprensión del flujo de información desde los genes a las estructuras moleculares, su función bioquímica, su conducta fisiológica y finalmente su influencia en las enfermedades y la salud.

Entre los principales factores que han favorecido el desarrollo de esta disciplina, se encuentra el impresionante volumen de datos sobre secuencias generadas por los distintos proyectos del genoma (tanto el humano como el de otros organismos); los nuevos enfoques experimentales, que permiten obtener datos genéticos a gran velocidad, bien de genomas individuales (mutaciones, polimorfismos), o de enfoques celulares (expresión génica); y el desarrollo de Internet, que permiten el acceso mundial a las bases de datos de información biológica. [10]

En la actualidad, con todo este avance, se han creado herramientas principalmente publicadas en sitios online para el diseño de siRNA. Ellas han sido fuentes utilizables en el presente trabajo como guía base para tomar experiencia sobre las técnicas de diseño de representación de la información que estos emplean.

1.1.1 miRNA y siRNA [11]. Semejanzas y diferencias

Un RNA de interferencia (iRNA) es un RNA de cadena simple que tiene la capacidad de inhibir o atenuar la expresión de genes. Los iRNA son transcritos a partir de ADN, pero no son traducidos a proteínas, estos pasan por varias fases de refinamiento (mecanismos de regulación en la célula) hasta formarse y estar listo para ejercer su función. De esta forma un iRNA es complementario a una parte de uno o más ARN mensajeros. El apareamiento del iRNA con el mRNA inhibe su traducción a proteína, y facilita la rotura del mRNA y de esta forma “apaga” la expresión de un gen [11].

El siRNA y el miRNA son ambos iRNA y su función por tanto es la misma.

El siRNA es creado de forma artificial por los especialistas en sus laboratorios una vez identificado el gen que se desea silenciar y los sitios blancos de su acción. La cadena de nucleótidos de un siRNA por tanto se diseña de acuerdo al objetivo específico del especialista teniendo en cuenta su investigación. Esto constituye una ventaja puesto que una vez creado el siRNA aumenta la probabilidad de que se silencie el gen deseado o se ejerza la acción de la forma que se quiere.

A diferencia de los siRNA, **el miRNA es creado naturalmente por el organismo**. El apareamiento de los miRNA no es siempre exacto, por tal causa, en ocasiones, la expresión del gen no es inhibida, sino solo atenuada.

1.1.2 mRNA

El RNA mensajero es el ácido ribonucleico que contiene la información genética procedente del ADN para utilizarse en la síntesis de proteínas, es decir, determina el orden en que se unirán los aminoácidos. Sale del núcleo y se asocia a ribosomas, donde se construye la proteína. Cada tres nucleótidos (codón) corresponde un aminoácido distinto. Así, la secuencia de aminoácidos de la proteína está configurada a partir de la secuencia de los nucleótidos del mRNA [12].

1.1.3 SNP

Un polimorfismo de nucleótido simple es una variación en la secuencia de ADN que ocurre cuando un nucleótido simple (A, T, C, G) en el genoma (u otra secuencia compartida) difiere entre miembros de una especie (o entre pares de cromosomas en un individuo). Por ejemplo, dos fragmentos de ADN secuenciados de diferentes individuos AAGC'C'TA y AAGC'T'TA, contienen una diferencia en un nucleótido simple. En este caso se dice que son dos alelos: C y T. Para que una variación sea considerada SNP, debe ocurrir al menos en el 1% de la población. Es importante notar que hay variaciones entre la población, de forma que un alelo de un SNP que es común en un grupo geográfico o étnico puede ser mucho más raro en otro.

Las variaciones en las secuencias de ADN de humanos pueden afectar la forma en que los humanos desarrollan las enfermedades y responden a agentes patógenos, químicos, medicamentos y vacunas. Para el presente trabajo, tener en cuenta esta información es imprescindible, pues los especialistas harán uso del sistema para reducir la búsqueda a los siRNA que no contienen SNP. Esto posibilita su aplicación en el diseño de futuros fármacos, efectivos para la mayor cantidad de personas que lo requieran, disminuyendo el riesgo de obtener resultados no esperados para determinados individuos. [13]

1.2 Sistemas existentes para el diseño de siRNA

Hoy en día existen cierto número de herramientas publicadas en Internet dedicadas específicamente al diseño de siRNA. Las más conocidas a nivel internacional son:

1.2.1 Target Finder siRNA design [3]

Es una herramienta creada para encontrar los sitios blancos en una secuencia de mRNA de interés. Esta herramienta no da la posibilidad de consultar una base de datos para obtener la cadena mRNA, sino que solo da la opción de entrar la cadena ya conformada por el usuario y para el análisis de esta secuencia cuenta con solo cuatro reglas.

1.2.2 SiRNA Design Software [4]

SDS (software para el diseño de siRNA) toma una secuencia del mRNA y hace uso de las herramientas de diseño ya existentes para dar como resultado una lista de sitios blancos. No es

conocido el algoritmo que implementa para el análisis de las secuencias de mRNA y cuenta con pocas reglas para arrojar un resultado de calidad. Además se tarda considerablemente al hacer el análisis de la secuencia solicitada.

1.2.3 Dharmacon siDESIGN Center [5]

Es una herramienta de diseño con notables ventajas dentro de las que se destaca una interfaz de usuario amigable y significativas mejoras en el proceso de identificación de siRNA. No obstante a pesar de estas ventajas, posee solo cuatro reglas para realizar la búsqueda de sitios blancos y no da la posibilidad de analizar una secuencia entrada por el usuario, en caso de que se necesite realizar la búsqueda en una cadena que no esté en la base de datos.

1.2.5 DEQOR [6]

Es una herramienta que opera mediante un uso excesivo de secciones que muestran las distintas soluciones según las reglas seleccionadas por el usuario. Evalúa la potencia inhibidora de las secuencias potenciales de siRNA e identifica las regiones del gen que tienen un alto poder de silenciamiento.

Posee una interfaz poco amigable y estructurada. No se tiene conocimiento del algoritmo que utiliza para el análisis de las secuencias. Brinda demasiadas reglas en cuanto a la configuración de vistas que complican el trabajo del usuario. El análisis de las secuencias se torna tedioso debido al tiempo que emplea en obtener y mostrar los resultados. Las vistas que ofrece no cuentan con la calidad que necesitan los especialistas puesto que muestra los resultados de forma poco intuitiva.

1.2.6 IDT's RNAi Design [7]

Esta aplicación a pesar de las ventajas que posee hace consultas a una base de datos externa (Centro Nacional para la Información de Biotecnología NCBI) que resulta confiable debido al prestigio del centro, pero no siempre los datos de esta pueden estar disponibles. No se conoce el algoritmo de diseño que emplea.

1.2.7 GenScript siRNA Target Finder [8]

Esta herramienta brinda la opción de configurar las reglas del análisis. Permite seleccionar la cadena de una base de datos según el identificador de un gen o entrar ya la cadena conformada y posibilita

realizar estudios sobre varios organismos. Tiene como una importante limitación que usa un algoritmo propietario. Además es necesario estar registrado y autenticado para un uso eficiente de la aplicación, lo cual provoca pérdida de tiempo y desmotivación por parte del usuario.

1.2.4 siDirect [9]

Es una aplicación web online para diseñar los siRNA que acepta una secuencia como entrada y devuelve como resultado una lista de sitios blancos. Posee grandes ventajas en cuanto al diseño pero como ya se ha señalado en aplicaciones anteriores, no se conoce al algoritmo que emplea para realizar el análisis de las secuencias. Hace consultas a la base de datos del NCBI lo que constituye una limitante como ya se ha señalado anteriormente. Cuando se presentan los resultados del análisis de una secuencia, las vistas que muestra no poseen la calidad requerida y tampoco brinda con claridad la información representada.

1.3 Limitaciones de los sistemas existentes

Las herramientas anteriormente presentadas, tienen su uso básico sobre Internet. Estas usualmente suelen sufrir cambios repentinos e ignorados por los usuarios. Se puede afirmar que los resultados obtenidos de los análisis, haciendo uso de estos sistemas, carecen de total confidencialidad. No se tienen en cuenta criterios de búsqueda como en el caso donde hay que considerar la existencia de los SNP. Tampoco se trata ningún tipo de información en relación a los genes (miRNA) que silencian de forma natural al gen que está siendo analizado.

La mayoría de estas herramientas solo permiten hacer el análisis mediante una cadena mRNA entrada por el usuario. Se desconocen los algoritmos implementados para el análisis de las secuencias de mRNA y las reglas que estos algoritmos aplican para la predicción de los sitios blancos de siRNA.

El año pasado se desarrolló la primera versión del Sistema de Análisis de Secuencias mRNA para el Diseño de siRNA. Esta, aunque cumplía con todos los requerimientos de los especialistas, definidos en aquel período, no resolvía todas las problemáticas planteadas en el tema de diseño de siRNA para el silenciamiento genético. Presentaba todas las reglas que se necesitaban pero no con el dinamismo requerido. El algoritmo implementado para la predicción de sitios blancos no era el más eficiente. La arquitectura de este software no permitía su extensión a nuevas funcionalidades fácilmente, es decir, con un mínimo de cambios y esfuerzos en la implementación. Las vistas que mostraba al usuario, resultado de las solicitudes realizadas, no tenían la calidad suficiente y eran poco funcionales. Además,

la herramienta estaba orientada fundamentalmente a la presentación de la información y no a las funcionalidades que realizaba.

Por lo anteriormente planteado es que se origina la necesidad de desarrollar una nueva versión de este sistema. Versión que incluye una arquitectura más extensible, un algoritmo más eficiente y dinámico que el anterior, en cuanto a las reglas presentadas para su ejecución. En esta versión, para el análisis de los genes, se tiene en cuenta la presencia de los SNP y de los miRNA para una mayor efectividad y eficiencia en los resultados.

1.4 Tecnologías, metodologías y herramientas

Las metodologías de desarrollo del software describen los pasos para desarrollar un proyecto de software. Definen detalladamente qué es lo que se debe hacer, cómo hacerse y quién debe hacerlo. Precisan cuáles son las tareas a desarrollarse durante el ciclo de vida del proyecto, los artefactos que han de ser construidos, cómo deben hacerse y el orden en que serán realizados, además de designar responsables por cada tarea. Especifican la estructura que ha de tener el proyecto para un mejor desenvolvimiento de todos los procesos llevados a cabo durante su desarrollo. De esta forma se obtiene como producto final, una solución eficaz, factible y de calidad al problema que le dio origen.

Para el desarrollo de este sistema se ha hecho un estudio de las principales herramientas y entornos de desarrollo acorde a las especificaciones de los clientes. Luego de un análisis riguroso, entre los desarrolladores y los clientes, de las diferentes tecnologías se tomaron los acuerdos de cuales elegir para la aplicación. A continuación se presentan los principales aspectos considerados que fundamentan nuestra elección.

1.4.1 Tecnología J2EE

La tecnología Java 2 Enterprise Edition (J2EE) proporciona una completa y potente plataforma orientada al desarrollo de aplicaciones corporativas distribuidas y a los servicios web.

Las aplicaciones desarrolladas sobre esta plataforma suelen tener una arquitectura en capas, que es lo que se pretende realizar en esta versión del sistema. Esto permite una mejor organización, reutilización y extensión del código en nuevas versiones.

Entre las características de este tipo de aplicaciones se encuentran las siguientes:

- Necesidad de alta productividad en el desarrollo de la aplicación
- Integración con los sistemas existentes
- Libertad de elección de plataformas de desarrollo y producción
- Escalabilidad
- Modelos flexibles de seguridad. [14]

Se utiliza J2EE como plataforma de desarrollo, además de las características que se vieron anteriormente, debido a la seguridad que aporta, cuenta con librerías y frameworks, la mayoría gratuitas y de código abierto (*open source*).

Existe la opción de escoger otra plataforma, como por ejemplo la plataforma .NET. Esta brinda todas y cada una de las ventajas anteriormente mencionadas, pero no es ni libre ni gratuita. Además, tampoco es multiplataforma (no puede utilizarse en distintos sistemas operativos).

1.4.2 Lenguaje JAVA

Java ofrece todas las ventajas de un lenguaje potente y robusto, pues fue diseñado para crear software altamente fiable. Es un lenguaje de programación basado en clases y orientado a objetos. Sus características de memoria liberan a los programadores de responsabilidades y errores.

Java es compilado en un código intermedio más abstracto que el código de máquina que es ejecutado por la máquina virtual de Java. Hereda su sintaxis de los lenguajes C y C++, pero cuenta con un modelo de objetos mucho más sencillo y elimina elementos de trabajo a bajo nivel como los punteros. Una de las características más significativas de Java es que posee una arquitectura neutral, es decir, el compilador Java compila su código a un fichero objeto de formato independiente de la arquitectura de la máquina en que se ejecutará. La independencia de la arquitectura representa solo una parte de su portabilidad. Java especifica los tamaños de sus tipos de datos básicos y el comportamiento de sus operadores aritméticos, de manera que los programas son iguales en todas las plataformas.

Para el desarrollo del software, ya que es una aplicación web, se podría escoger como lenguaje de desarrollo PHP. Es un lenguaje fácil de aprender y de usar. Consume menos recursos y menos tiempo de desarrollo que Java, pero no brinda la seguridad de este último. PHP es un lenguaje interpretado, y no compilado como Java, lo que lo hace más lento. No maneja la persistencia de objetos tanto como lo

hace Java, pues este es más robusto en ese aspecto.

De igual forma se cuenta con el conocido lenguaje C#, pero este tiene las mismas limitantes que se abordaron de la plataforma .NET.

Por las razones anteriormente expuestas y por solicitud del cliente, fue que se escogió Java como lenguaje para el desarrollo de esta aplicación.

1.4.3 Herramientas de desarrollo utilizadas

Existen varias herramientas para el desarrollo de un software, de ellas hacemos uso de las CASE y las herramientas IDE.

Se puede definir a las Herramientas CASE como un conjunto de programas y ayudas que dan asistencia a los analistas, ingenieros de software y desarrolladores, durante todos los pasos y procesos del Ciclo de Vida de desarrollo de un proyecto. [15]

Visual Paradigm como Herramienta CASE

Es una herramienta CASE que soporta el UML como lenguaje de modelado. Además provee el Modelado del Proceso de Negocio, un generador de mapeo objeto-relación para Java, .NET y PHP. Sirve de soporte también para 13 tipos de diagramas de UML. Una característica clave es su habilidad de generar no solo código del modelo de clases, sino también de la estructura de la base de datos relacional adecuados para sostener persistentemente la información contenida en las clases llamadas "entidad" [16].

Como otra herramienta CASE muy similar al Visual Paradigm tenemos el Rational Rose, pero por no ser una herramienta multiplataforma y tener menos facilidades que el Visual Paradigm, se escogió esta última.

NetBeans 6.0 como entorno integrado de desarrollo (IDE)

La plataforma NetBeans permite que las aplicaciones sean desarrolladas a partir de un conjunto de componentes de software llamados módulos. Estos a su vez pueden ser desarrollados independientemente, razón por la que las aplicaciones basadas en NetBeans pueden ser extendidas fácilmente por otros desarrolladores de software. Este IDE está desarrollado para la programación en Java, pero puede servir para cualquier otro lenguaje. Existe además un número importante de módulos para extender el IDE NetBeans. [17]

Por todas las ventajas que ofrece NetBeans es que se decide a nivel de polo (Bioinformática) hacer uso de esta herramienta para el desarrollo de la versión de la aplicación.

1.4.4 Gestor de Base de Datos

PostgreSQL

Es un servidor de base de datos relacional orientada a objetos de software libre. Está considerado como la base de datos de código abierto más avanzada del mundo. [18] PostgreSQL aproxima los datos a un modelo objeto-relacional, y es capaz de manejar complejas rutinas y reglas. Ejemplos de su avanzada funcionalidad son las consultas SQL declarativas, el control de concurrencia multi-versión, el soporte de multi-usuario, transacciones, optimización de consultas, herencia, y arreglos. En este proyecto se usa PostgreSQL además de las ventajas que posee, porque la estructura de la base de la tecnología del cliente presenta dicho gestor de base de datos y se debe de cumplir con sus peticiones. Además de que es uno de los gestores que estableció nuestro polo de Bioinformática. [19]

1.4.5 Servidor de Aplicaciones

Tomcat 6.0

Tomcat es un servidor web con soporte de Servlets y JSPs. Incluye el compilador Jasper, que compila páginas JSP convirtiéndolas en Servlets. El motor de servlets de Tomcat a menudo se presenta en combinación con el servidor web Apache.

Tomcat puede funcionar como servidor web por sí mismo. En sus inicios existió la percepción de que el uso de Tomcat de forma autónoma era sólo recomendable para entornos de desarrollo con requisitos mínimos de velocidad y gestión de transacciones. Hoy en día ya no existe esa percepción y Tomcat es usado como servidor web autónomo en entornos con alto nivel de tráfico y alta disponibilidad. [20].

Es necesario señalar que el cliente de este proyecto tiene JBoss como servidor de aplicaciones y este utiliza internamente un servidor Tomcat. En este trabajo se decide usar este último como servidor de aplicaciones para gastar menos recursos de cómputo debido a que no se cuenta con toda la infraestructura necesaria para el correcto funcionamiento de JBoss. Además de que la versión de Tomcat que se utiliza es totalmente compatible con la versión de JBoss que utiliza el cliente y cumple

con las políticas del polo al que pertenece este proyecto.

1.4.6 Lenguaje de Modelado (UML)

UML es el lenguaje que utiliza RUP para visualizar, especificar, construir y documentar cada una de las partes que constituye un sistema de forma general. Se hace uso de este, dadas sus facilidades para el modelado de estos artefactos, está compuesto por diversos elementos gráficos.

EL UML (Lenguaje Unificado de Modelado) es una de las herramientas más gustadas en el mundo para desarrollo de software, esta particularidad de su uso se debe a que permite a los desarrolladores de sistemas, generar diseños conforme a sus ideas de forma fácil y entendible.

1.4.7 Proceso de desarrollo de software (OpenUp)

Para el desarrollo correcto de la aplicación se utiliza la metodología OpenUP. Esta preserva las características fundamentales de RUP, que incluye un desarrollo iterativo, casos de uso y escenarios, administración de riesgos, entre otros elementos. Excluye la mayoría de las partes opcionales de RUP e incluye varios elementos nuevos. Es un proceso ágil y ligero, que promueve al desarrollo del software sobre las buenas prácticas, haciéndolo un proceso pequeño y extensible si es necesario (híbrido, incluye partes de otros modelos). Debido a estas ventajas que presenta OpenUp el polo de Bioinformática la adoptó para aplicarla en todos sus proyectos, y por consiguiente en el presente trabajo.

1.5 Resumen de la tecnología a utilizar

Como resultado del estudio de las tecnologías y herramientas de desarrollo a utilizar se ha llegado a la conclusión de que el sistema se desarrollará sobre la plataforma **J2EE** usando como lenguaje de programación **Java**, y como gestor de bases de datos **PostgreSQL 8.2**. Se determinó además hacer uso del **Lenguaje Unificado de Modelado (UML)** por su estrecha integración con RUP (*Rational Unified Process*) que constituye la base de **OpenUP**, metodología adoptada. Como herramienta de modelado se utiliza **Visual Paradigm**. Como entorno de desarrollo integrado se decidió que debe ser **NetBeans 6.0** y como servidor de aplicaciones se adoptó **Tomcat 6.0**.

1.6 Conclusiones

En este capítulo se adentró en los principales conceptos vinculados al tema para una mejor comprensión del lector. Se ha hecho un análisis de los sistemas actualmente utilizados para dar solución al problema del diseño de siRNA. Se detallaron las limitantes de estos sistemas de manera general para ver cuáles eran las posibles funcionalidades de interés en el presente trabajo. Se hace además una fundamentación de las tecnologías a usar por los desarrolladores, programadores y analistas para el proceso de desarrollo de la herramienta en cuestión.

Capítulo 2

Características del sistema

Introducción

En el presente capítulo se hace una descripción del objeto de estudio, y se hace alusión a los problemas existentes que dieron pie al desarrollo de la herramienta. Se definen los requerimientos, tanto funcionales como no funcionales y se brinda una propuesta de solución para la construcción de la misma haciendo uso de los artefactos y procesos que ofrece OpenUp, obteniendo como resultados el diagrama de casos de uso del sistema y las descripciones de los correspondientes casos de uso.

2.1 Objeto de estudio

Para desarrollar el presente trabajo se hace necesario estudiar las aplicaciones cliente-servidor con el objetivo de detallar su funcionamiento y ventajas. Esto proporcionará un mejor entendimiento del funcionamiento de la aplicación y su desarrollo.

2.1.1 Objetivos estratégicos de la organización.

En la actualidad, existen gran cantidad de centros que enfocan sus investigaciones al campo de la genética. Particularmente en nuestro país uno de ellos es el CIGB. Este tiene un papel integrador en la esfera de la biotecnología cubana, asumiendo además la responsabilidad de contribuir directamente al desarrollo económico y social de nuestro país. Su impacto está destinado a la salud humana, las producciones agropecuarias, acuícolas, y al medio ambiente [21]. Debido a la gran importancia de las investigaciones que allí toman lugar ha surgido la idea de crear un sistema capaz de predecir los sitios blancos para el mecanismo de silenciamiento de genético, pues es sabida su gran utilidad en la curación de enfermedades tanto humanas como en animales y plantas.

2.2 Procesos objeto de automatización.

2.2.1 Procesos que serán objeto de automatización.

En el proceso de estudio de secuencias mRNA para el diseño de siRNA, se pretende automatizar todo lo referente a la información genética que se maneja actualmente en el CIGB, específicamente para el diseño de siRNA. Quedan enmarcadas en este proceso las reglas necesarias para el diseño establecidas por los especialistas.

Las secuencias de mRNA se pueden analizar de varias formas. La segunda versión del sistema se centrará principalmente en la búsqueda de sitios blancos de la acción del siRNA tomando en cuenta los SNP. Incluirá el algoritmo desarrollado para esta búsqueda implementado en la versión anterior, con mejoras en referencia al tiempo de ejecución; y añadirá una nueva funcionalidad relacionada con los miRNA que permitirá buscar y mostrar para un gen determinado los genes que lo silencian.

Análisis simple:

Primeramente se necesita una secuencia de mRNA que será objeto del análisis. Puede ser entrada por el usuario o resultar de consultas realizadas por el mismo a la base de datos. En este último caso, se necesitaría la información correspondiente al criterio de búsqueda seleccionado (nombre para el caso de un transcrito y símbolo o id de un gen). Cuando el especialista solicita la búsqueda de un gen, luego de obtener los datos correspondientes, se procede a buscar los transcritos asociados a este. Posteriormente se muestra de forma gráfica los datos de dichos transcritos para luego dar la opción al especialista de seleccionar uno de estos y realizar el análisis. Luego de haber mostrado los datos del transcrito solicitado se da la posibilidad de realizar el análisis de la secuencia de nucleótidos del mismo.

Después de esta última acción, el especialista debe especificar las reglas necesarias para llevar a cabo el proceso de análisis para obtener los sitios blancos. Estas reglas son:

- La secuencia comienza con AA.
- Tiene T en la posición 10.
- Tiene una A en la posición 3.
- Tiene una A en la posición 19
- Contenido de CG de 30-50% (se da la posibilidad de escoger a conveniencia los

porcentajes).

- Ausencia de GC en la posición 19.
- Presenta G en la posición 13.
- Tiene 3 o más A/T entre las bases 15 y 19.
- De las reglas seleccionadas se desea que se cumplan al menos un número específico seleccionado.
- Excluir las regiones SNP del diseño (no se incluye para el caso de introducir una secuencia).

Como resultado del análisis se obtiene una representación gráfica de los sitios blancos que corresponden a la cadena de nucleótidos, incluyendo aquellos que contienen SNP. También se da la posibilidad de ver de forma individual la información de cada sitio blanco.

Es notable resaltar que la aplicación da la opción de realizar el análisis de la secuencia completa de un gen.

Análisis teniendo en cuenta los SNP:

Al hacer el análisis anteriormente descrito se tienen en cuenta los SNP. Estos, localizados en los exones, son la primera causa de variabilidad genética; producen cambios en el código que altera la cadena de aminoácidos [12]. Debido a que son porciones de código que implican una diferencia genética, no se pueden tener en cuenta para el diseño de los siRNA, pues estas variaciones en la secuencia del ADN pueden afectar la respuesta de los individuos a enfermedades, bacterias, virus, productos químicos, fármacos. Como resultado del análisis de la secuencia se muestran solo los sitios blancos que no contienen SNP.

Búsqueda de miRNA:

Se buscan y muestran gráficamente los miRNA que actúan en determinado gen señalando la posición donde actúan. Se visualizan además datos significativos del miRNA. Además se brinda la opción de mostrar solamente los de un valor de significancia mínimo seleccionado en adelante.

Se pretende automatizar las actividades de análisis anteriormente señaladas, para así obtener como

resultado un producto funcional y extensible a cambios estructurales, haciendo posible que las investigaciones se realicen de manera independiente. El objetivo final es obtener un sistema que reúna todos los procesos requeridos por los especialistas con una interfaz de configuración con facilidades para su utilización.

2.3 Propuesta de sistema

2.3.1 Descripción general de la propuesta de sistema

Persiguiendo mejorar la calidad de las investigaciones llevadas a cabo en el CIGB, referentes al diseño de siRNA, se pretende desarrollar un sistema centrado en cómo llevar a cabo de manera rápida y eficaz el estudio de los genes, específicamente para lograr un diseño de siRNA más estable y seguro. Para alcanzar este nivel se desarrollará un sistema capaz de almacenar y manipular la información necesaria para realizar los estudios de silenciamiento genético.

El sistema permitirá las siguientes funcionalidades:

Analizar secuencias mRNA: Permitirá analizar una secuencia mRNA, introducida por el especialista o formada a partir de la información de genes y transcritos almacenada en la base de datos. Brindará la posibilidad de elegir las reglas deseadas para el análisis y el tipo de análisis que se quiere realizar (excluyendo los SNP o no). Luego de esta operación se mostrará la localización de los sitios blancos, así como sus datos (representación gráfica, secuencia, posiciones inicial y final, y el número de reglas que se cumplen).

Mostrar Genes: Se brindará la posibilidad de visualizar la información referente a los genes, con sus transcritos asociados y los datos más importantes de estos y conocer además de forma gráfica los datos de los miRNA que actúan sobre los mismos, y en que región lo hace.

Mostrar Transcritos: Permitirá a los especialistas conocer la información referente a los transcritos. Mostrando además los datos de los exones asociados y entre estos, las regiones codificantes y las no traducidas.

2.3.2 Análisis comparativo de otras soluciones existentes con la propuesta

Considerando otras soluciones existentes, las que son accesibles a través de Internet y el sistema

anteriormente desarrollado, se llega a la conclusión que la propuesta que se brinda con el presente trabajo es la más acertada como solución al problema planteado por los especialistas.

- Se brindan más funcionalidades que las que implementan las demás herramientas (análisis con SNP y visualizar miRNA), visualizando cada resultado de forma gráfica para un mejor entendimiento y valoración del mismo.
- Se implementan algoritmos propios bien conocidos por los usuarios finales.
- Mantiene la independencia de internet lograda con la aplicación anterior lo cual favorece el avance en las investigaciones.
- Cuenta con una conexión a una base de datos segura, solo accedida por el administrador del sistema, lo cual posibilita que se mantenga la confiabilidad de la información y se mantenga disponible siempre que se necesite.

En comparación con la solución anterior el sistema que se propone posee ventajas. Se pretende obtener un algoritmo para la búsqueda de sitios blancos tomando en cuenta los SNP, esto no se tuvo en cuenta en esta primera versión del producto. Se propone una arquitectura de implementación eficiente y ventajosa, no utilizada anteriormente, que permitirá que el sistema sea extensible a cualquier otra funcionalidad que se desee agregar. Para el análisis de un gen determinado esta versión incluye la información de los miRNA, que aumenta la efectividad de este proceso. El sistema quedará concebido no como un portal web, que es una de las deficiencias del sistema anterior, sino como una herramienta potente capaz de brindar servicios de calidad y eficiencia a los investigadores.

2.4 Especificación de los requisitos de software

2.4.1 Requerimientos Funcionales

Los requerimientos funcionales son capacidades o condiciones que el sistema debe cumplir. Permiten expresar una especificación más detallada de las responsabilidades del sistema en cuestión. Los requerimientos funcionales del software propuesto son los siguientes:

CU: Buscar datos de un gen.

- **R 1.1:** Buscar datos de un gen.

- **R 1.2:** Visualizar datos de un gen.
- **R 1.3:** Buscar datos de transcritos asociados al gen.
- **R 1.4:** Visualizar datos de transcritos asociados al gen.
- **R 1.5:** Buscar datos de los miRNA asociados al gen.
- **R 1.6:** Visualizar datos de los miRNA asociados al gen.

CU: Buscar datos de un transcrito.

- **R 2.1:** Buscar los datos de un transcrito indicado.
- **R 2.2:** Visualizar gráficamente los datos de un transcrito indicado.
- **R 2.3:** Buscar los SNP asociados a un transcrito.
- **R 2.4:** Visualizar gráficamente los SNP asociados a un transcrito

CU: Realizar análisis de las secuencias mRNA

- **R 3.1:** Buscar sitios blancos en la secuencia mRNA.
- **R 3.2:** Mostrar gráfica de la secuencia de mRNA resaltando sitios blancos ordenados por reglas cumplidas.
- **R 3.3:** Mostrar SNP asociados a la secuencia de mRNA.
- **R 3.4:** Mostrar datos de un sitio blancos indicado.

2.4.2 Requerimientos no funcionales

Los requerimientos no funcionales son propiedades o cualidades que el producto debe tener. Representan las características del mismo.

Apariencia o interfaz externa:

La herramienta deberá tener un diseño de interfaz amigable y agradable, de forma tal que el usuario haga uso de la misma sin dificultad alguna, ajustándose a los estándares establecidos para el

desarrollo de un buen diseño.

Usabilidad:

El sistema brindará la posibilidad de ser usado por cualquier persona cuyos conocimientos en cuanto al trabajo con los ordenadores sean básicos, solo es necesario contar con conocimientos especializados en biología para poder entender los resultados brindados por la aplicación.

Rendimiento:

El sistema esta concebido para un ambiente cliente/servidor, que garantiza que la eficiencia del mismo en cuanto a la rapidez de respuesta ante las solicitudes de los usuarios sea la mejor, al igual que la velocidad de procesamiento de la información. Para ello se realiza la validación de los datos de entrada y la manipulación de eventos en el cliente y en el servidor, las que por cuestiones de seguridad o de acceso a los datos lo requieran. De esta forma se logra una mayor velocidad de procesamiento y un mayor aprovechamiento de los recursos.

Soporte:

El sistema debe dar la posibilidad de mejoramiento y su ampliación a otras especificaciones que se le incorporen en un futuro. Además debe ser de fácil instalación, adaptable a numerosas plataformas y de fácil mantenimiento.

Portabilidad:

El sistema esta concebido para que sea multiplataforma.

Confiabilidad:

El sistema debe brindar el máximo de confiabilidad y precisión de la información al usuario, para evitar errores.

Software:

Se debe disponer de Windows 95 (o versiones superiores) y Linux para la instalación de la aplicación. La aplicación se realizará en un ambiente Web, la base de datos es independiente de la aplicación. Se debe disponer de un navegador web cualquiera.

Hardware:

Para el desarrollo y puesta en práctica del proyecto se requieren máquinas servidoras con los siguientes requisitos:

Para el servidor de aplicaciones:

- Procesador Pentium 3 o superior.
- 512 MB de RAM.
- Un mínimo de 4 GB de espacio disponible en disco duro.

Para el servidor de base de datos.

- Procesador Pentium 3 o superior.
- 512 MB de RAM.
- Un mínimo de 2 GB de espacio disponible en disco duro

Para las PC clientes se necesita:

- Un procesador Pentium 3 o superior.

2.5 Definición de actores y casos de uso del sistema

2.5.1 Actores

Un actor es una entidad externa al sistema que interactúa con él. Siempre se beneficia de la realización de un caso de uso. Estimula el sistema con eventos de entrada y salida. Es un rol de un usuario que puede intercambiar información o puede ser un recipiente pasivo de información. Representa un ser humano, una máquina o un software que interactúa con el sistema.

Nombre del actor	Descripción
Especialista	Representa el usuario del sistema que tiene la capacidad profesional de entender e interactuar con todas las funcionalidades del mismo.

Tabla #1 Actores del sistema.

2.6.2 Casos de Uso definidos

Los casos de uso son “fragmentos” de funcionalidad que ofrece el sistema para aportar un resultado de valor para sus actores. Son una secuencia de acciones a llevar a cabo por el sistema como respuesta a los requerimientos funcionales. Se utilizan para obtener información de cómo debe trabajar el sistema. Describen el comportamiento del sistema mediante el intercambio acción y reacción.

Los casos de uso definidos son los siguientes:

1. Realizar análisis de las secuencias mRNA.
2. Buscar datos de un gen.
3. Buscar datos de un transcrito.

CU 1	Realizar análisis de la secuencia mRNA
Actor	Especialista
Descripción	Obtener un diseño de siRNA a partir de un adecuado análisis de la secuencia mRNA que se haya escogido.
Referencia	1. R3. 2. CU Mostrar transcrito.

Tabla #2 CU Realizar análisis de la secuencia mRNA

CU 2	Buscar datos de un gen.
Actor	Especialista

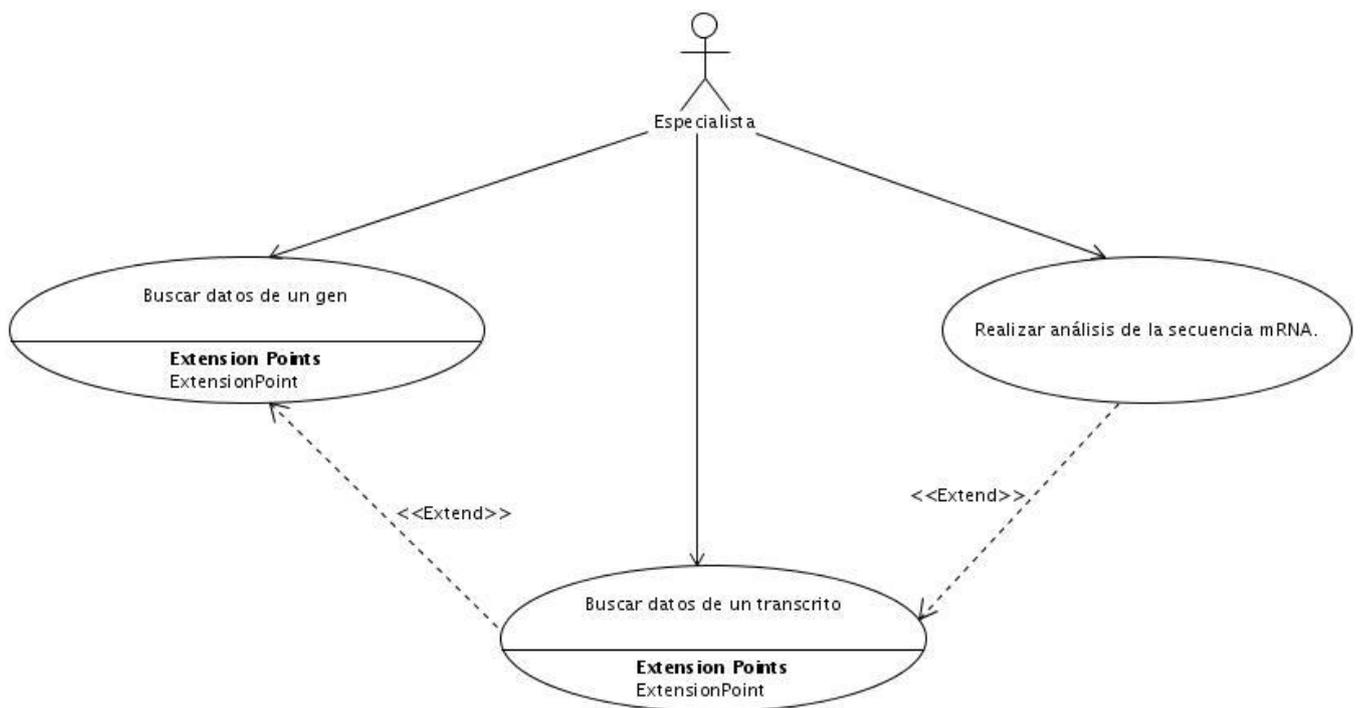
Descripción	Tiene como propósito mostrar información de un determinado gen.
Referencia	R1

Tabla #3 CU Buscar datos de un gen.

CU 3	Buscar datos de un transcrito.
Actor	Especialista
Descripción	Su objetivo es mostrar al especialista información relevante de un determinado transcrito.
Referencia	R2.

Tabla #4 CU Buscar datos de un transcrito

2.5.3 Diagrama de casos de uso del sistema



2.5.4 Descripción detallada de casos de usos.

Caso de Uso:	Realizar análisis de la secuencia mRNA
Actores:	Especialista.
Propósito:	Obtener al predicción de sitios blancos en la secuencia mRNA a partir de un adecuado análisis indicado.
Resumen:	El caso de uso se inicia cuando luego de haber buscado y mostrado determinado gen o transcrito y el especialista decide analizar la secuencia encontrada; o accede a la opción “Analyze mRNA sequence” y procede a analizarla según las reglas definidas obteniéndose la predicción de los sitios blancos para la acción del siRNA.
Referencia:	R 3.1; R 3.2; R 3.3; R 3.4; R 3.5; R 3.6
Precondiciones:	
Poscondiciones:	
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
1) El especialista selecciona la opción “Search and Analyze Gene” para analizar la secuencia de un gen.	2) El sistema visualiza la interfaz para buscar los datos de un gen (cromosoma, id y símbolo) y luego analizar su secuencia.
3) El especialista selecciona un criterio de búsqueda (id o símbolo del gen) y ordena la búsqueda.	El sistema : 4) Valida que se entren datos. 5) Busca los datos del gen especificado (cromosoma, id y símbolo, cadena de nucleótidos). 6) Muestra los datos del gen dando la posibilidad de realizar un análisis visualizando las reglas a seleccionar para el mismo.

<p>7) El especialista selecciona las reglas para analizar la secuencia mRNA y ordena el análisis.</p>	<p>8) El sistema verifica que al menos se seleccione una regla y que la secuencia sea correcta.</p>
	<p>9) El sistema verifica si el especialista ha seleccionado la regla de excluir los SNP de los resultados (“Exclude SNP regions from the design (No For Pasted Sequence)”).</p> <p>10) El sistema busca sitios blancos en la secuencia mRNA y los SNP asociados a la misma.</p> <p>11) El sistema muestra gráficamente los resultados de la búsqueda de los sitios blancos en la cadena mRNA y señala los sitios blancos con SNP ordenados por la cantidad de reglas cumplidas de las seleccionadas.</p> <p>12) El sistema muestra una gráfica con los SNP de la secuencia.</p> <p>13) El sistema brinda la posibilidad de visualizar en detalle la información de un sitio blanco.</p>
Flujos Alternos	
Acción del Actor	Respuesta del Sistema
<p>13.1) El especialista selecciona un sitio blanco para ver su información.</p>	<p>13.2) El sistema visualiza los datos del sitio blanco seleccionado (secuencia de nucleótidos, posición inicial y posición final de la misma y regiones SNP).</p>
<p>1.1) El especialista accede a la opción “Analyze mRNA sequence”.</p>	<p>1.2) El sistema muestra la interfaz para realizar el análisis de una secuencia mRNA dada.</p>
<p>1.3) El especialista introduce la secuencia mRNA, selecciona las reglas y ordena el análisis.</p>	<p>1.4) El sistema verifica que se entró la secuencia y que esta es correcta.</p>

	<p>1.5) El sistema busca sitios blancos en la secuencia mRNA.</p> <p>1.6) El sistema muestra gráficamente los resultados de la búsqueda de los sitios blancos en la cadena mRNA ordenados por la cantidad de reglas cumplidas.</p> <p>1.7) Se va a la acción 13 del flujo normal de eventos.</p>
	<p>9.1) El sistema busca los sitios blancos en la cadena excluyendo los que contienen SNP.</p> <p>Se va a la acción 1.6 de los flujos alternos.</p>
	<p>4.1) En caso de que el especialista no introduzca los datos el sistema muestra un mensaje de error.</p>
Prioridad: Prototipo Interfaz	Critico.

Tabla #5 Descripción de CU Realizar análisis de la secuencia mRNA

Caso de Uso:	Buscar datos de un gen.
Actores:	Especialista.
Propósito:	Mostrarle información relevante a los especialistas.
Resumen:	El caso de uso se inicia cuando el especialista escoge una de las opciones de búsqueda de secuencias (id _ gen o símbolo del gen) y desea conocer más información sobre los genes, sin necesidad de realizar un análisis previo. Entonces se especifica un criterio de búsqueda, y a partir del mismo se busca en la base de datos la información necesaria para que sea consultada por los especialistas.
Referencia:	R 1.1; R 1.2; R 1.3; R 1.4
Precondiciones:	

Poscondiciones:	
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
	1) El sistema brinda la opción de escoger un criterio de búsqueda para buscar y mostrar los datos de un gen (“By gene id”, “By gene symbol”).
2) El especialista: Selecciona un criterio de búsqueda. Introduce los datos requeridos(id o símbolo del gen)	El sistema: 3) Valida que se entre la información de un criterio de búsqueda. 4) Busca los datos del gen (número del cromosoma, id y símbolo). 5) Busca los datos de los transcritos asociados al gen (posición inicial y final, cantidad de exones, nombre y regiones codificantes y no traducidas de los exones asociados a los transcritos). 6) Busca los datos de los miRNA asociados al gen (nombre, pvalue, SCORE, posición inicial y final). 7) Muestra los datos del gen. 8) Muestra gráficamente los datos de los transcritos asociados. 9) Muestra gráficamente los datos de los miRNA asociados al gen.
	10) El sistema brinda la posibilidad de seleccionar un pvalue (nivel de significancia) para visualizar los miRNA asociados al gen con un nivel de significancia que como máximo sea el especificado.

	13) El sistema da la posibilidad mostrar en detalle un transcrito y analizar su secuencia de nucleótidos.
14) Se realiza el caso de uso "Buscar datos de un transcrito" y se ejecuta el caso de uso "Realizar análisis de la secuencia mRNA" y así termina la ejecución del caso de uso.	
Flujos Alternos	
Acción del Actor	Respuesta del Sistema
2.2) El especialista introduce la información del criterio de búsqueda de un gen que no tiene miRNA.	2.2.1) El sistema no muestra los miRNA.
2.3) El especialista introduce la información del criterio de búsqueda de un gen que no tiene transcritos.	2.3.1) El sistema no muestra los transcritos.
13.1) Se culmina la ejecución del caso de uso.	
10.1) El especialista selecciona un pvalue y ordena que se muestren los miRNA.	10.2) El sistema muestra los miRNA que tengan un valor de significancia menor que el especificado.
	3.1) Si el especialista no introduce los datos requeridos el sistema muestra un mensaje de error.
Prioridad: Prototipo Interfaz	Critico.

Tabla #6 Descripción de CU Buscar datos de un gen

Caso de Uso:	Buscar datos de un transcrito.
---------------------	--------------------------------

Actores:	Especialista.
Propósito:	Mostrarle información relevante a los especialistas.
Resumen:	El caso de uso se inicia cuando luego de haber generado a partir la especificación de un criterio de búsqueda la secuencia del gen requerido se muestran los transcritos asociados al mismo, el especialista escoge un transcrito y del mismo se muestra su información o el especialista ordena la búsqueda por el criterio de búsqueda nombre del transcrito.
Referencia:	R 2.1; R 2.2; R 2.3; R 2.4
Precondiciones:	
Poscondiciones:	
Flujo Normal de Eventos	
Acción del Actor	Respuesta del Sistema
	1) El sistema brinda la opción de escoger un criterio de búsqueda para buscar y mostrar los datos de un transcrito.
2) El especialista selecciona el criterio de búsqueda "By transcript name", introduce la información requerida (nombre del transcrito) y ordena la búsqueda.	3) El sistema valida que se entre la información referente a al criterio de búsqueda. 4) El sistema busca los datos del transcrito (posición inicial y final, cantidad de exones, nombre y regiones codificantes y no traducidas de los exones asociados al transcrito y regiones SNP). 5) El sistema busca la secuencia de nucleótidos del transcrito. 6) El sistema muestra gráficamente los datos del transcrito. 7) El sistema muestra la secuencia de nucleótidos.
	8) El sistema da la posibilidad de analizar la secuencia de nucleótidos del transcrito visualizando las reglas para dicho análisis.

9) Se realiza el caso de uso " Realizar análisis de la secuencia mRNA" a partir de la acción 4 del flujo normal de eventos.	
Flujos Alternos	
Acción del Actor	Respuesta del Sistema
	1.1) El sistema visualiza una interfaz donde muestra los transcritos asociados a un gen y sus datos y da la posibilidad de seleccionar uno de ellos para visualizar en detalle sus datos.
1.2) El especialista selecciona el nombre del transcrito y ordena la búsqueda.	1.3) El sistema busca los datos del transcrito (secuencia de nucleótidos y regiones SNP). 1.4) El sistema muestra gráficamente los datos del transcrito (posición inicial y final, cantidad de exones, nombre y regiones codificantes y no traducidas de los exones asociados al transcrito y regiones SNP). 1.5) Se realizan las acciones de la 6 a la 9 del flujo normal de eventos.
	3.1) El sistema muestra un mensaje de error en caso de que el especialista no haya introducido la información solicitada.
9.1) Se sale del caso de uso.	
Prioridad:	Critico.
Prototipo Interfaz	

Tabla #7 Descripción de CU Buscar datos de un transcrito. Buscar datos de un transcrito.

2.7 Conclusiones

En este capítulo se describió el objeto de estudio del presente trabajo y se plantearon los problemas existentes que dieron pie al desarrollo de la herramienta. Se ha brindado una definición de los requisitos que deberá cumplir el sistema. A través de estos se han detallado las especificaciones que han hecho los especialistas como parte del flujo de trabajo de levantamiento de requisitos. Se redefinió el diagrama de casos de uso del sistema planteado en la tesis anterior, retirando casos de uso que se consideraron innecesarios, lo cual permitió que se representaran detalladamente las funcionalidades requeridas, lográndose una modelación más clara de la interacción del sistema y el especialista.

De esta forma se concretó el funcionamiento del sistema que se pretende desarrollar y quedaron sentadas las bases para las siguientes fases del proceso de diseño e implementación de la herramienta.

Capítulo 3

Análisis y diseño del sistema

Introducción

En el presente capítulo se ofrece una descripción de la arquitectura del sistema y de los principales objetivos del diseño. Se muestran los diagramas de clases web del diseño y se hace una breve comparación del diseño de la versión anterior de la herramienta con el utilizado en la actual tomando como base un caso de uso específico. Se muestran los diagramas de interacción generados en el flujo de trabajo actual y además se define el diseño de la base de datos a través del modelo lógico y el modelo físico. Luego se da una breve descripción de las tablas más relevantes de la base de datos creada. Se muestra el diagrama de despliegue de la aplicación, y se hace énfasis en la seguridad de la misma y el tratamiento de errores implementado.

3.1 Arquitectura de Software del sistema

La Arquitectura del Software es la organización fundamental de un sistema formado por sus componentes, las relaciones entre ellos y el contexto en el que se implantarán, y los principios que orientan su diseño y evolución. Para que un sistema sea extensible y reutilizable, como se pretende que sea el presente proyecto, su arquitectura tiene que estar diseñada a en base a ello. Para tal propósito y buen uso de la misma se utilizan los patrones de arquitectónicos.

Patrón de arquitectura: Los patrones arquitectónicos expresan el esquema de organización estructural fundamental para sistemas de software. Provee un conjunto de subsistemas predefinidos, especifica sus responsabilidades e incluye reglas y pautas para la organización de las relaciones entre ellos.

Dentro de los varios patrones arquitectónicos que existen, en este proyecto se escogió el Controlador Frontal, que implementa el patrón Modelo Vista Controlador (MVC) con páginas Java.

Vista lógica: La vista lógica describe el diseño más importante de las clases y su organización en paquetes y subsistemas, y la organización de éstos en capas. Muestra cómo la funcionalidad es

diseñada en el interior del sistema, en términos de la estructura estática y comportamiento dinámico del sistema.

Para el desarrollo del presente trabajo se escogió un estilo arquitectónico en Capas, dado que se adapta muy bien a las necesidades del negocio y es aprobado internacionalmente para el desarrollo de aplicaciones web. A continuación se presenta un diagrama con las capas diseñadas y sus relaciones, estas son: capa **Vista**, capa **Controlador**, y capa **Modelo**.

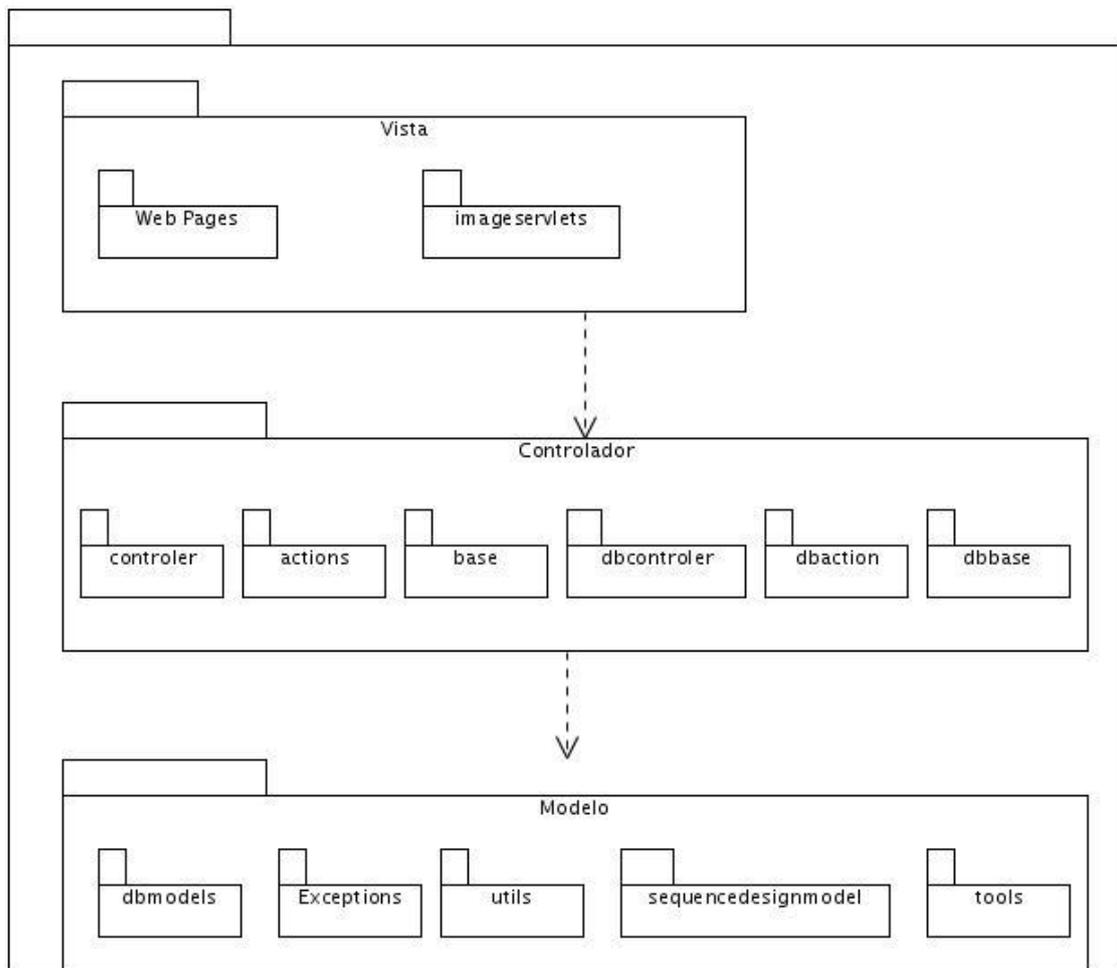


Fig. #1. Arquitectura del sistema.

Vista: Capa que contiene las clases dedicadas puramente a la creación de objetos de presentación. Para trabajar con las vistas hemos empleado el uso de plantillas.

La capa Vista (o presentación), fue separada en 2 paquetes. El primero *Web Pages*, contiene a las páginas servidoras (<<server page>> según extensión a UML para aplicaciones web) encargadas de producir las páginas cliente (<<client page>>) que se presentan en el cliente web (navegador web). El segundo paquete *imageservlets* contiene los servlets encargados de generar las distintas imágenes que se visualizan de forma dinámicas al cliente a través del navegador web. No todas las vistas (paginas jsp) tienen asociada un servlet generador de imágenes, ejemplo de esto es la página “geneSequence.jsp”. Sin embargo con la página “findgeneID.jsp” no sucede esto, esta tiene asociado el servlet “imageFindGeneServlet.java” encargado de generar la imagen correspondiente.

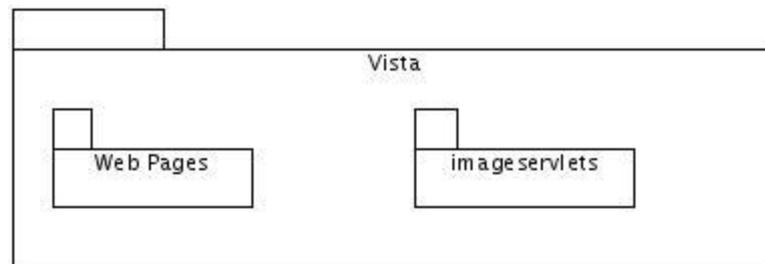


Fig. #2 Paquetes de la Capa Vista

Entre cada una de las páginas de estos paquetes se establecen relaciones de tipo <<link>> y <<submit>> que no fueron mostradas en el diagrama para no cargarlo de información irrelevante para este nivel de abstracción.

Controlador: Capa que contiene todas las clases encargadas de garantizar la lógica de la aplicación. Según el patrón Controlador Frontal, se ha trabajado con estructuras denominadas acciones que no son más que clases encargadas de llevar el control del flujo en el sistema.

La capa Controlador (o lógica de aplicación), fue separada en 6 paquetes. Cada uno de estos paquetes contiene clases que son las encargadas de llevar a cabo el flujo de la aplicación y preparar la información que va a ser enviada a la capa Vista para ser mostrada.

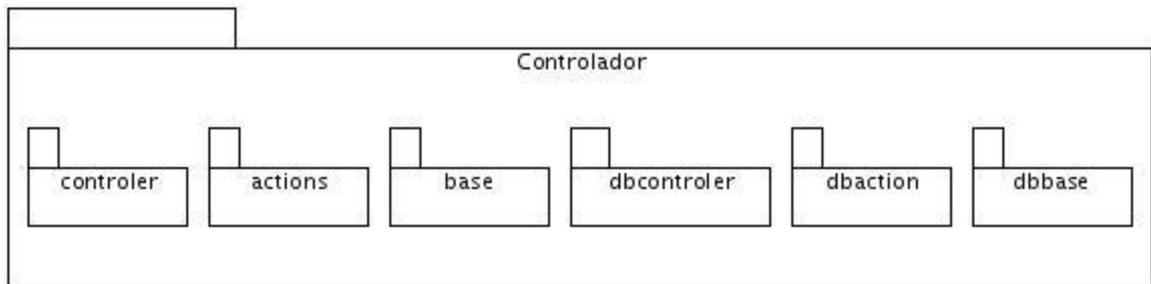


Fig. #3 Paquetes de la Capa Controlador

En el sistema existen dos tipos acciones:

- Las que se llevan a cabo sin conexión a base de datos.
- Y las que se llevan a cabo con conexión a base de datos.

Las acciones que se llevan a cabo sin conexión a base de datos son manipuladas por las clases situadas en los paquetes controller, action y base. El paquete controller contiene la clase “controllerServlet” que controla a todas las acciones de este tipo. El paquete action brinda la Interfaz “Action” de donde tienen que heredar todas las acciones para poder implementarse y acoplarse al sistema correctamente. Y el paquete base, brinda un conjunto de clases que son las que dan la estructura general a la armazón y un correcto funcionamiento a todas las acciones que son realizadas sin conexión a la base de datos.

Las acciones que se llevan a cabo con conexión a la base de datos son manipuladas por las clases situadas en los paquetes dbcontroller, dbaction y dbbase, que realizan la misma función que los paquetes controller, action y base respectivamente, pero con las acciones que se llevan a cabo con conexión a la base de datos.

Modelo: Capa que contiene todas las clases que toman las decisiones importantes desde el punto de vista empresarial y del negocio a informatizar.

La capa Modelo (dominio o lógica de negocio), está dividida en 5 paquetes, y agrupan las clases responsables de la toma de decisiones del negocio en el contexto de cada acción.

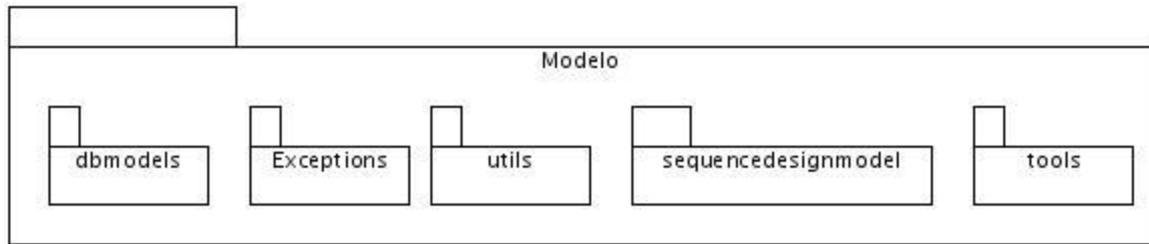


Fig. #4 Subsistemas de la capa: Modelo

El paquete *dbmodels* contiene todos los modelos de las acciones que se conectan a la base de datos. El paquete *exceptions* contiene las clases que definen los distintos tipos de excepciones que puede dar la aplicación.

El paquete *utils* contiene las clases persistentes de la base de datos, con las que se trabajan en cada consulta realizada a esta base de datos.

El paquete *sequencedesignmodel* contiene las clases responsables del negocio para las acciones que no se conectan a la base de datos.

El paquete *tools* contiene las clases responsables del acceso a algún recurso externo que sea necesitado por algún modelo. En este caso, a través la clase contenida en este paquete accedemos a los ficheros que contienen información de los cromosomas del genoma humano.

3.2 Objetivos del diseño

El diseño tiene como principales objetivos comprender detalladamente los requisitos funcionales y no funcionales, sistemas operativos, tecnologías de distribución, restricciones relacionadas con el lenguaje de programación, tecnologías de interfaz de usuario. Además, el diseño crea un punto de partida para las actividades de implementación que siguen.

La realización física de los casos de uso del sistema se ve descrita mediante el modelo de diseño. Este se centra en como influyen los requisitos funcionales y no funcionales en el sistema en cuestión.

3.3 Diagramas de Clases del Diseño

A continuación se muestran los diagramas de clases web del diseño por cada caso de uso. Estos permiten que se describa detalladamente y de forma gráfica las especificaciones de las clases de

software. Estos diagramas contienen: Clases, asociaciones y atributos, métodos, información sobre los tipos de atributos, navegabilidad, dependencias.

El lenguaje de modelado UML cuenta con una extensión para el modelado de aplicaciones Web. La misma es usada para el diseño de las clases. Los estereotipos que usa esta extensión son:

<<Server Page>> Representa la página Web con código que se ejecuta en el servidor y además interactúa con recursos en el mismo.

<<Client Page>> Es una página Web, con formato HTML; mezcla de datos, presentación y lógica de trabajo.

<<Form>> Grupo de elementos de entrada que forman parte de una página cliente que constituyen sus atributos.

Existen otros componentes como las clases JavaScript para hacer las validaciones en la entrada de datos.

CU Buscar datos de un gen

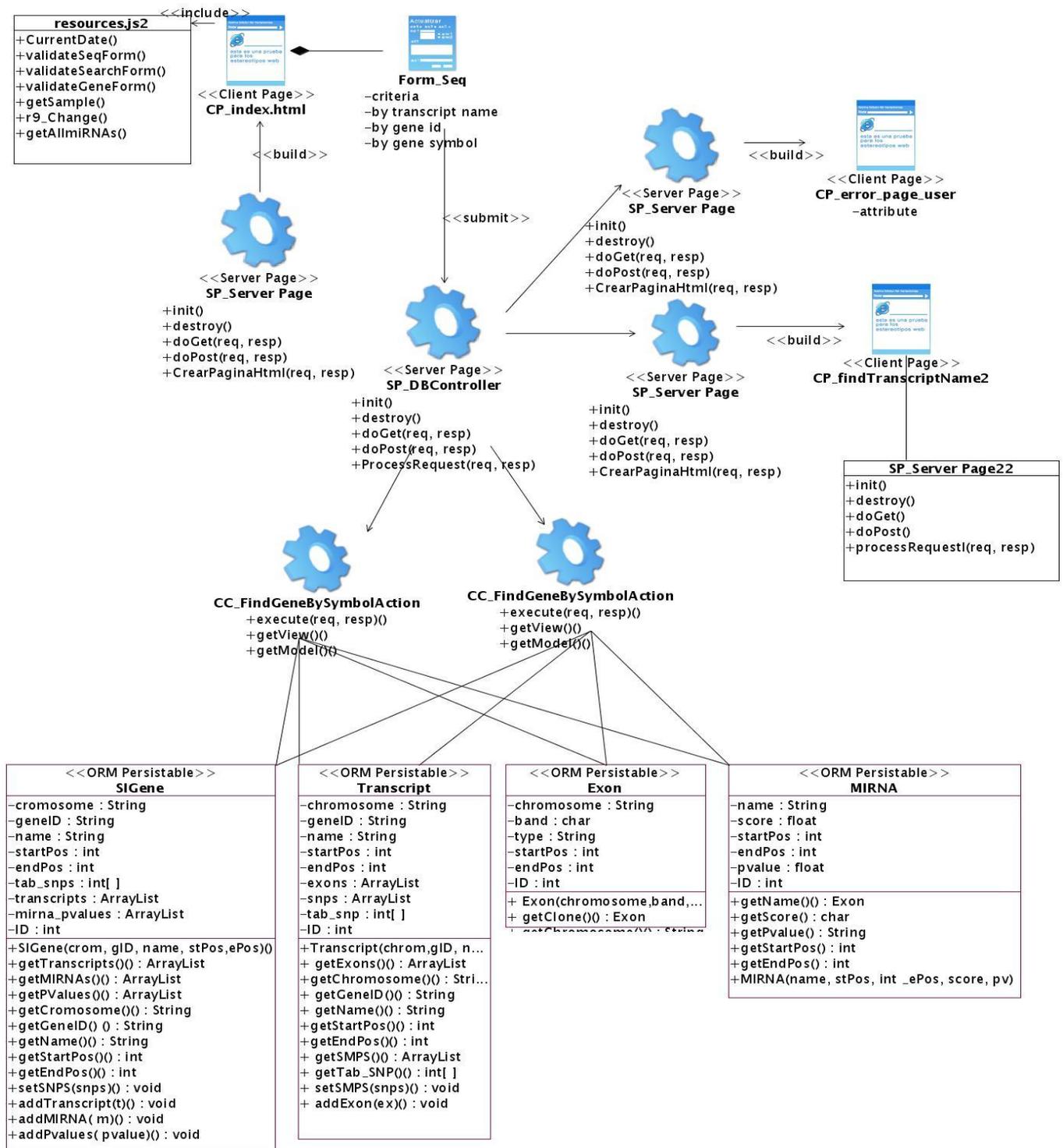


Fig. #5 Diagrama de clases del diseño de CU Buscar datos de un gen. V 1.1

CU Buscar datos de un transcrito

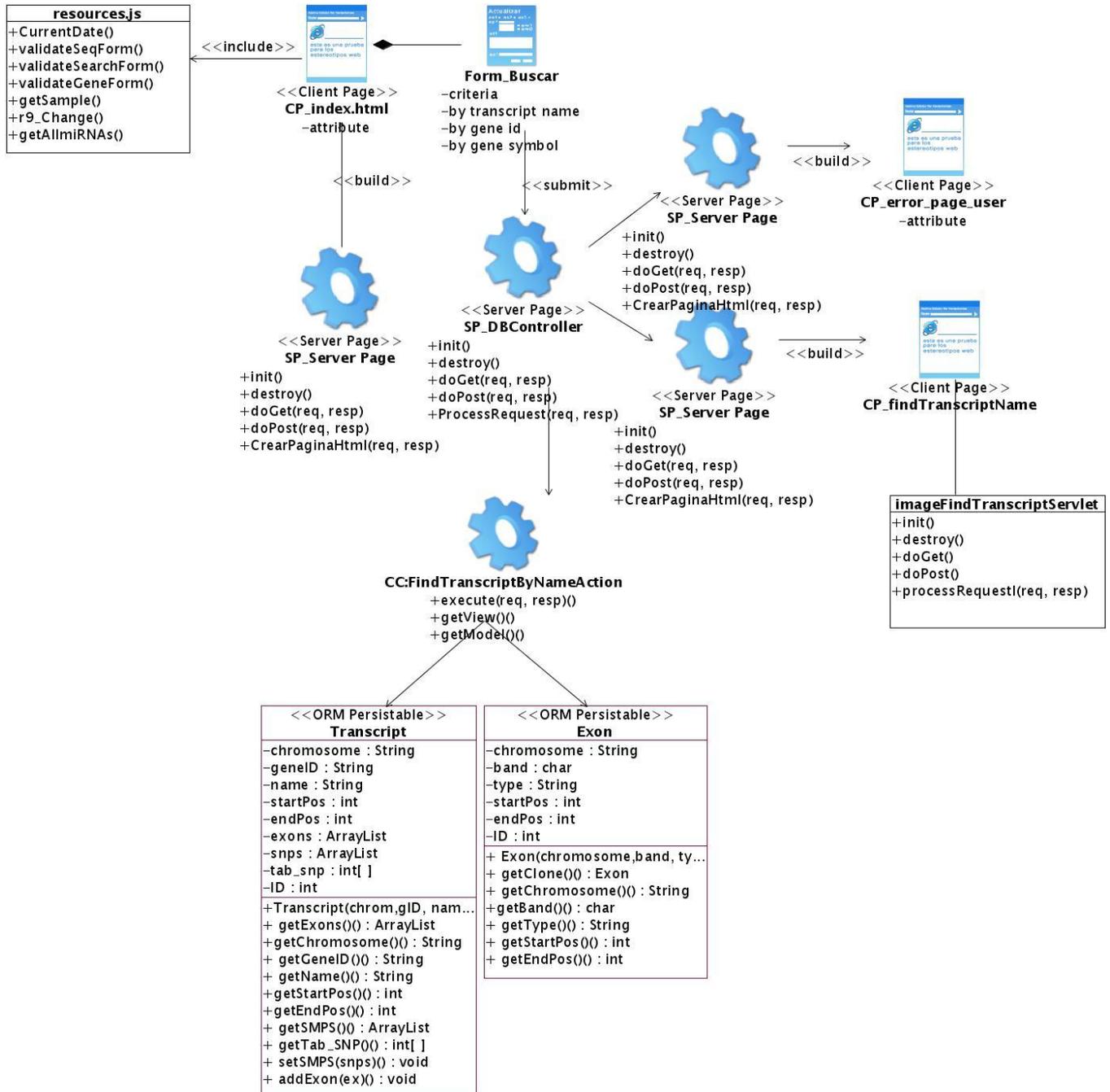


Fig. #6 Diagrama de clases del diseño de CU Buscar datos de un transcrito. V 1.1

CU Realizar análisis de secuencia mRNA

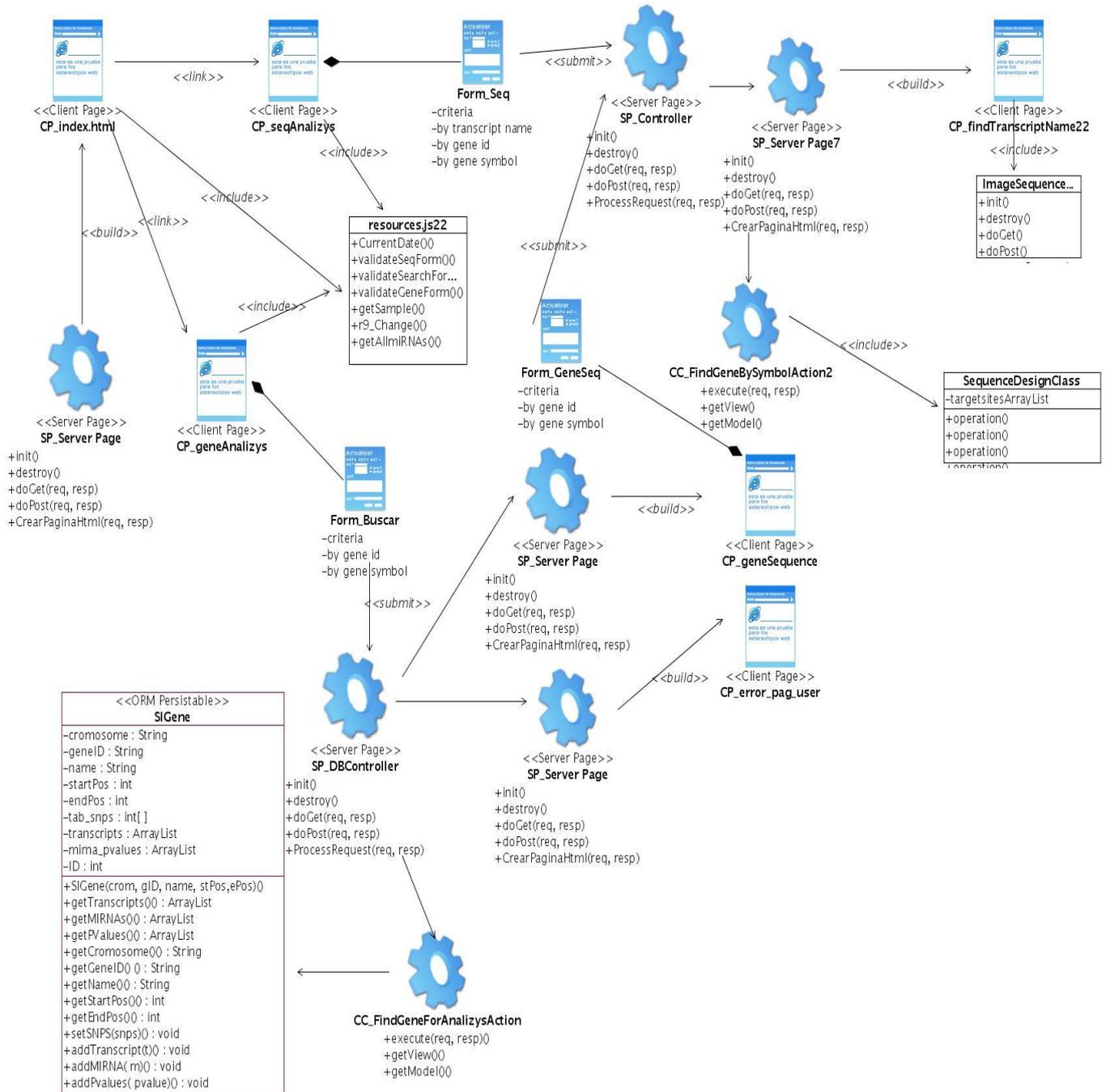


Fig. #7 Diagrama de clases del diseño de CU Realizar análisis de secuencia mRNA. V 1.1

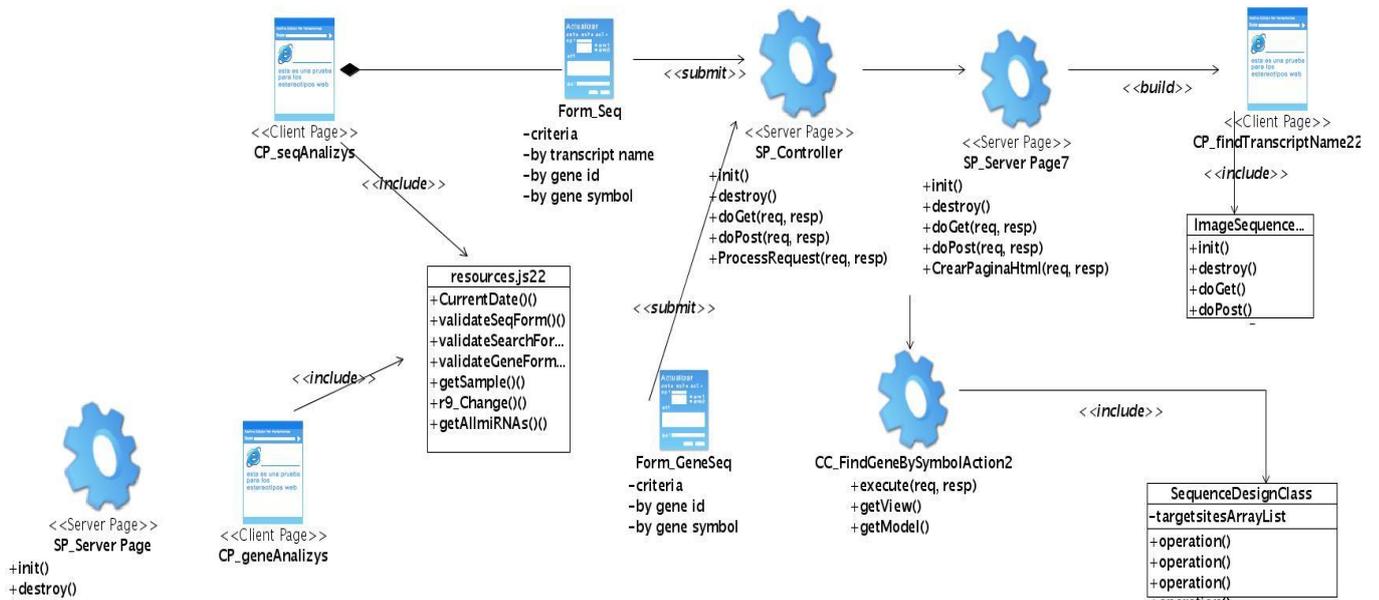


Fig. #8 Diagrama de clases del diseño de CU Realizar análisis de secuencia mRNA. V 1.1

(Parte 1)

Se ha dividido el diagrama de clases del diseño de CU Realizar análisis de secuencia mRNA. V 1.1 en dos figuras para una mejor visualización del mismo. Quedan evidenciadas en el diagrama original (Fig. # 7) las relaciones existentes entre las clases del diseño que la componen.

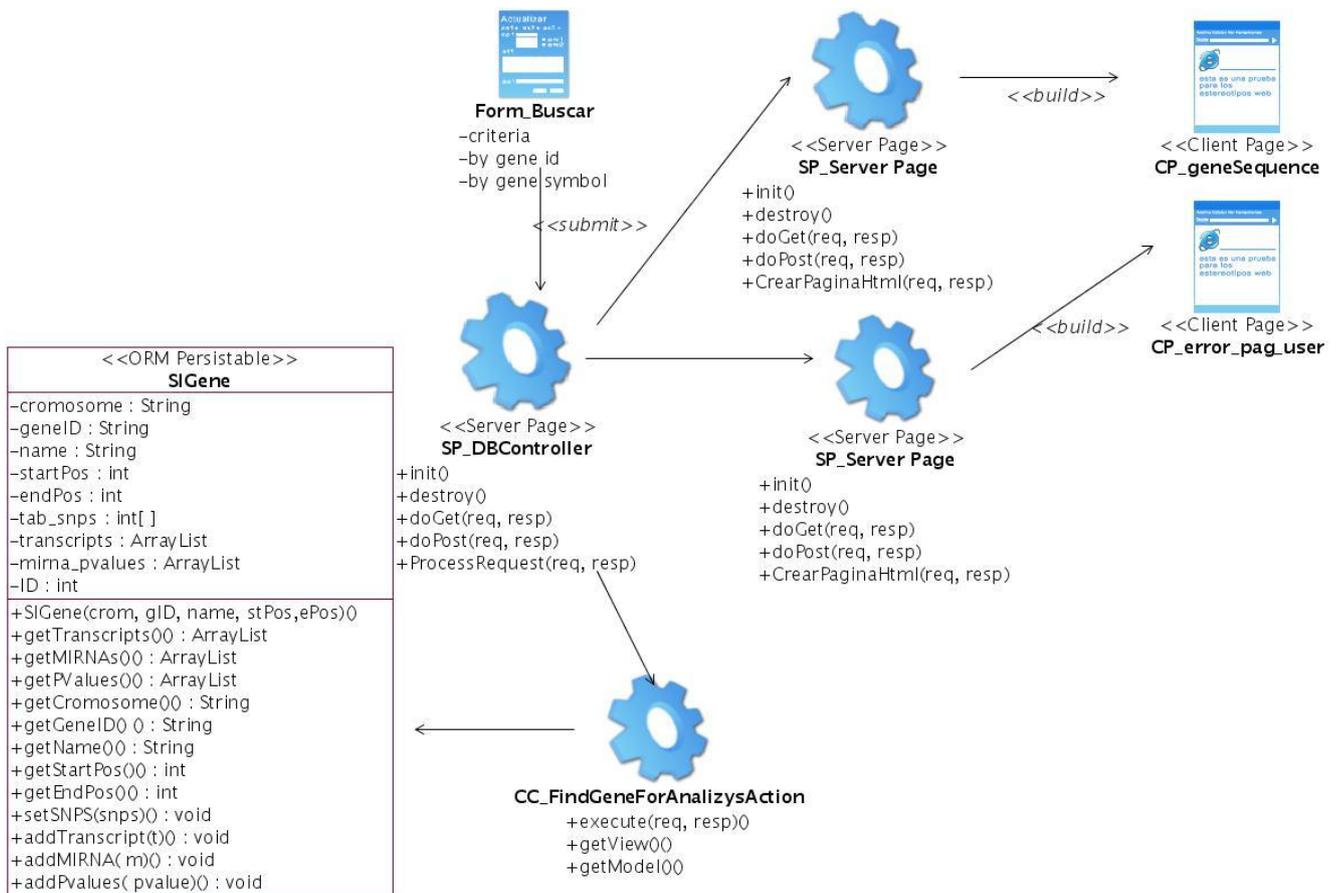


Fig. #9 Diagrama de clases del diseño de CU Realizar análisis de secuencia mRNA. V 1.1

(Parte 2)

3.4 Comparación de diseño

En los objetivos específicos definidos en el presente trabajo se hace alusión al rediseño del sistema, para ello primeramente se hizo necesario, siguiendo la metodología de desarrollo OpenUp, generar los artefactos de los flujos previos al análisis y diseño.

Para rediseñar del sistema se siguieron los principios del patrón de arquitectura Modelo-Vista-Controlador específicamente implementado por el patrón Controlador Frontal. Como resultado de ello se hizo necesario crear nuevas clases para el diseño y suprimir otras utilizadas en la versión anterior

de forma que la modelación del sistema quedara más clara y estructurada.

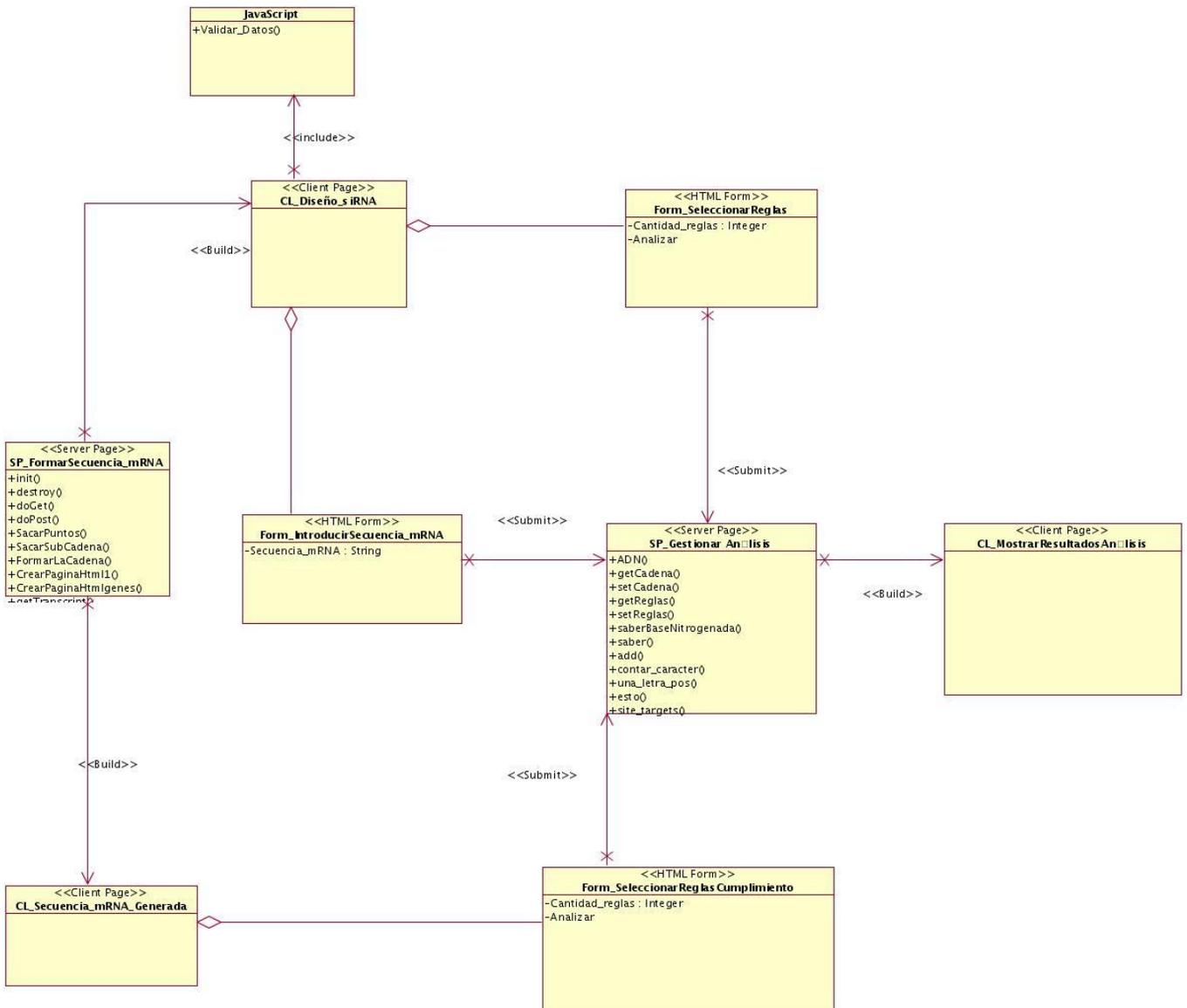


Fig. #10 Diagrama de clases del diseño de CU Realizar análisis de secuencia mRNA. V1.0 [9]

Como se puede observar en el diagrama de diseño de la Fig. #10, correspondiente a la versión anterior del sistema propuesto, hay clases del diseño que no tienen los atributos y responsabilidades que la describen. Esto hace que este diagrama sea poco comprensible por parte de los programadores. Si vemos el diagrama de la versión actual no ocurre de igual manera. La implementación de la arquitectura a través del diagrama de diseño de esta segunda versión es clara y precisa, ya que se

puede notar como el sistema maneja el flujo de sus acciones y de los posibles errores. La arquitectura de la versión anterior a pesar sus desventajas con la actual, tampoco es legible a través del diagrama de clases del diseño de este [9]. Se puede notar también que hay clases en esta primera versión que son innecesarias en el sistema actual. La razón está dada a que este último implementa el patrón MVC, aspecto del cual se carece en esta primera versión.

3.5 Diagramas de Interacción (Secuencia, Colaboración)

Los diagramas de interacción son utilizados para la modelación de los aspectos dinámicos de un sistema. Esto conlleva modelar instancias concretas, componentes y nodos junto con los mensajes enviados entre ellos que representan su interacción. Un diagrama de casos de uso presenta una visión externa del sistema y la funcionalidad de estos se recoge como un flujo de eventos y para ello, se hace uso de interacciones entre sociedades de objetos.

Los diagramas de interacción tienen dos formas de manifestarse:

- Diagramas de secuencia.
- Diagramas de colaboración.

Un diagrama de secuencia destaca la ordenación temporal de los mensajes; el diagrama de colaboración a su vez destaca la organización estructural de los objetos que envían y reciben mensajes.

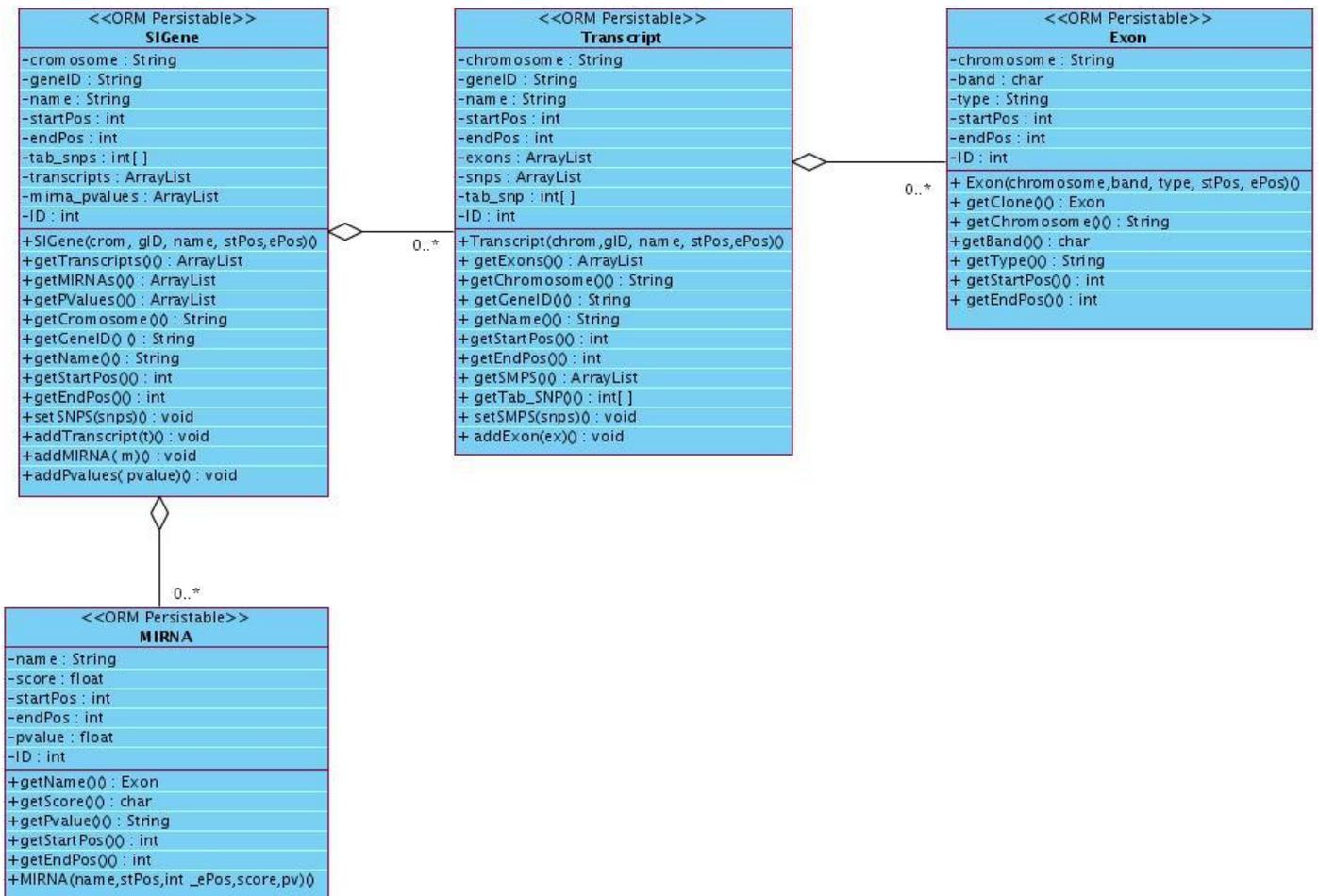
Diagramas de Secuencia. Ver [Anexo 1](#)

3.6 Diseño de la Base de Datos

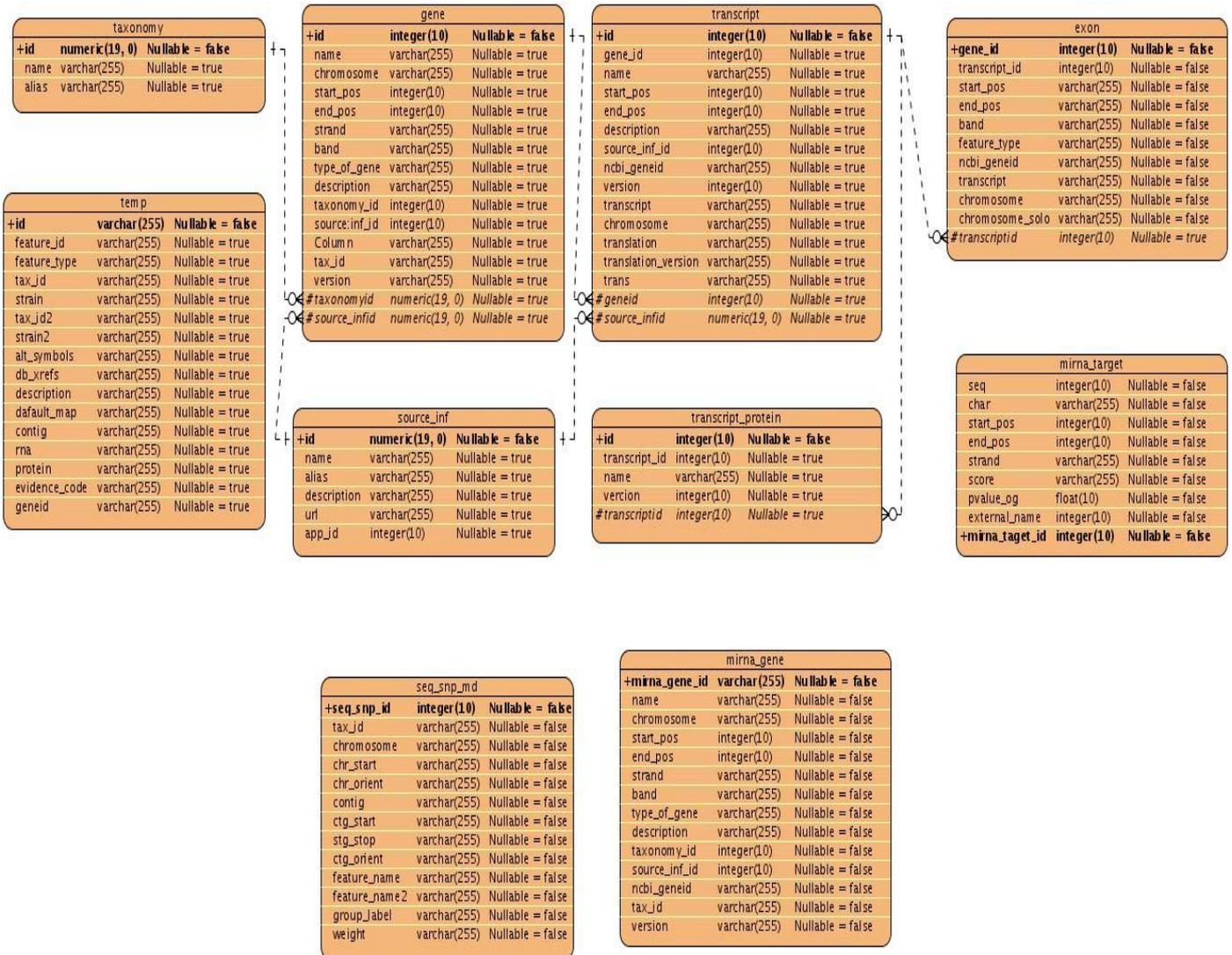
Para realizar el diseño de la base de datos de la aplicación se confeccionaron el diagrama de clases persistentes y el modelo de datos. Las clases persistentes son clases entidades que pueden mantener su valor en el espacio y el tiempo. En contrapartida las clases temporales son aquellas de las cuales el sistema se encarga de manejar y almacenar en tiempo de ejecución, razón por la que dejan de existir cuando termina la ejecución del programa.

A continuación se presenta el diagrama de clases persistentes que está compuesto por dichas clases y las correspondientes relaciones entre ellas y el modelo físico de datos que es una representación de las tablas de la base de datos y sus relaciones.

3.6.1 Modelo lógico de datos



3.6.2 Modelo físico de datos



3.7 Descripción de las Tablas de la Base de Datos.

Nombre: gene		
Descripción: En esta tabla se almacenan las informaciones referentes a los genes.		
Atributo	Tipo	Descripción
id	integer	Identificador autogenerado del gen.
name	varchar	Nombre del gen de de acuerdo al NCBI
chromosome	varchar	Número del cromosoma donde se localiza el gen
start_pos	int	Posición inicial en el cromosoma.
end_pos	int	Posición final en el cromosoma.
strand	char	Orientación del gen en el cromosoma, cadena (+) o cadena (-)
band	varchar	La banda del gen en la que se encuentra.
type_of_gene	varchar	Tipo del gen (codificador, RNA, etc.)
description	text	Descripción de la función del gen.
taxonomy_id	int	Id del organismo a que pertenece el gen de acuerdo a la base Taxonomy del NCBI
source_inf_id	int	Identificador de la fuente de información
ncbi_geneid	varchar	Identificador del gen de acuerdo al NCBI
tax_id		Identificador de la taxonomía a la que pertenece
version	varchar	Versión del gen.

Nombre: transcript		
Descripción: En esta tabla se almacenan las informaciones de los transcritos.		
Atributo	Tipo	Descripción
id	varchar	Identificador autogenerado del transcrito
gene_id	int	Gene _ id del gene en la tabla Gene
name	varchar	Nombre del transcrito de de acuerdo al NCBI
start_pos	int	Posición inicial en el cromosoma
end_pos	int	Posición final en el cromosoma
description	text	Descripción del transcrito
source_inf_id	int	Identificador de la fuente de recursos de donde provino.
ncbi_geneid	varchar	Identificador del gen que produce el transcrito
version	int	Versión del transcrito
transcript	varchar	Identificador del transcrito de acuerdo al NCBI
chromosome	varchar	Cromosoma del gen que produce el transcrito
translation	varchar	Identificador de la secuencia de aminoácidos de acuerdo al NCBI.
translation_version	varchar	Versión de la secuencia aminoácidos.

Nombre: exón		
Descripción: En esta tabla se almacenan las informaciones de los exones.		
Atributo	Tipo	Descripción
id	varchar	Identificador autogenerado del exón
gene_id	int	Gene _ id del gene en la tabla Gene
transcript_id	int	transcript_ id del transcrito gene en la tabla Transcript
start_pos	varchar	Posición inicial en el cromosoma
end_pos	varchar	Posición final en el cromosoma
band	varchar	La banda del cromosoma donde se encuentra
feature_type	varchar	Tipo de exón (UTR o CDS)
ncbi_geneid	varchar	Identificador del gen a que pertenece
transcript	varchar	Identificador del transcrito a que pertenece
chromosome	varchar	Cromosoma al que pertenece
chromosome_solo	varchar	Cromosoma al que pertenece

Nombre: chromosome		
Descripción: En esta tabla se almacenan las informaciones referentes a los cromosomas.		
Atributo	Tipo	Descripción
id	varchar	Identificador autogenerado del cromosoma
name	varchar	Nombre del cromosoma (1,2,3 ... X, Y)
known_genes	varchar	Genes conocidos.
unknown_genes	varchar	Genes desconocidos.
SNP	int	cantidad de SNP
length	varchar	Longitud del cromosoma
taxonomy_id	varchar	Identificador del organismo.

Nombre: source_Info		
Descripción: En esta tabla se almacenan toda la información de la fuente.		
Atributo	Tipo	Descripción
id	varchar	Identificador autogenerado de la fuente de información.
name	varchar	Nombre de la fuente.
alias	varchar	Alias de la fuente.
descripcion	varchar	Descripción de la fuente.
url	varchar	URL de la fuente.
app_id	varchar	Identificador de la aplicación que generó la información

Nombre: taxonomy		
Descripción: En esta tabla se almacenan todo lo referente a la información de los organismos.		
Atributo	Tipo	Descripción
id	varchar	Identificador del organismo.
name	varchar	Nombre del organismo.
alias	varchar	Alias del organismo.

Nombre: mirna_gene		
Descripción: En esta tabla se almacenan las informaciones referentes a los genes.		
Atributo	Tipo	Descripción
mirna_gene_id	integer	Identificador autogenerado del gen.
name	varchar	Nombre del gen de de acuerdo al NCBI
chromosome	varchar	Número del cromosoma donde se localiza el gen
start_pos	int	Posición inicial en el cromosoma.
end_pos	int	Posición final en el cromosoma.
strand	char	Orientación del gen en el cromosoma, cadena (+) o cadena (-)
band	varchar	La banda del cromosoma donde se encuentra

type_of_gene	varchar	Tipo del gen (codificador, RNA, etc.)
description	text	Descripción de la función del gen.
taxonomy_id	int	Id del organismo a que pertenece el gen de acuerdo a la base Taxonomy del NCBI
source_inf_id	int	Identificador de la fuente de información
ncbi_geneid	varchar	Identificador del gen de acuerdo al NCBI
tax_id		Identificador de la taxonomía a la que pertenece
Versión	varchar	Versión del gen.

Nombre: mirna_target		
Descripción: En esta tabla se almacenan las informaciones referentes a los genes.		
Atributo	Tipo	Descripción
mirna_target_id	integer	Identificador autogenerado del gen.
seq	varchar	Nombre del gen de de acuerdo al NCBI
chr	varchar	Número del cromosoma donde se localiza el gen
start_pos	int	Posición inicial en el cromosoma.
end_pos	int	Posición final en el cromosoma.
strand	char	Orientación del gen en el cromosoma, cadena + o cadena
band	varchar	La banda del cromosoma donde se encuentra
score	varchar	Valores propios de los SNP almacenados
pvalue_og	text	Valor de significancia del SNP
external_name	varchar	Gen al que silencia.

Se han incluido nuevas tablas a la base de datos ya existente en la versión anterior del sistema permitiendo que se pueda mostrar más información de los genes y hacer consultas más detalladas. Con esto se logra obtener resultados experimentales para los especialistas de mayor eficiencia y calidad

3.8 Diagrama de despliegue

El modelo de despliegue es una descripción de cómo va a estar distribuido físicamente el sistema. Se muestra a través de nodos de cómputo y las relaciones que existen entre los mismos.

Se puede observar lo siguiente sobre el modelo de despliegue:

- Los nodos están relacionados de modo que dichas relaciones representan la forma en que se comunican en términos de protocolos, dígame HTTP, TC/IP.
- Cada nodo representa un equipo de cómputo, normalmente un procesador o dispositivo. Pueden estar incluidas impresoras.

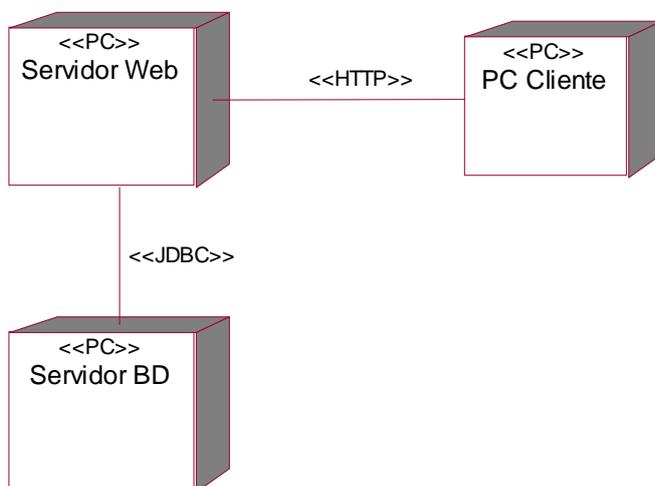


Fig. #11 Diagrama despliegue

3.9 Tratamiento de errores

Con el fin de prevenir los errores que pueda cometer el usuario al interactuar con el sistema, siempre se verifican los datos de entrada antes de ser utilizados. Esto evita incidentes y búsquedas innecesarias en la base de datos.

La aplicación hace uso de mensajes para señalarle al especialista que debe rectificar los datos introducidos antes de llevar a cabo las operaciones solicitadas por el mismo. Es de gran importancia verificar la integridad de la información introducida para de esta forma obtener un resultado de calidad. Un ejemplo de las validaciones que se realizan es cuando se selecciona un criterio de búsqueda pero no se introduce un valor. (*Fig. 7*).

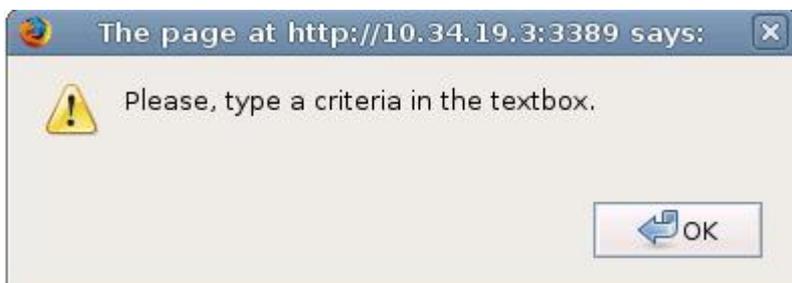


Fig. #12 Utilización de mensajes para rectificar datos nulos.

En la aplicación también se tratan los errores a nivel de servidor en cuanto a las búsquedas que se realizan en la base de datos. Cuando un usuario busca una información que no está almacenada, el sistema a través de excepciones propias trata este tipo de error. El problema a resolver con esto es que cada vez que se realiza una búsqueda, esta desencadena dos o tres búsquedas más, si la primera existe. Una vez que se comprueba que no existe la primera, el sistema controla que las demás no se ejecuten a través de las excepciones que el implementa. Con esto no solo se evita búsquedas innecesarias sino también demora de tiempo en los procedimientos. (*Fig. 8*)

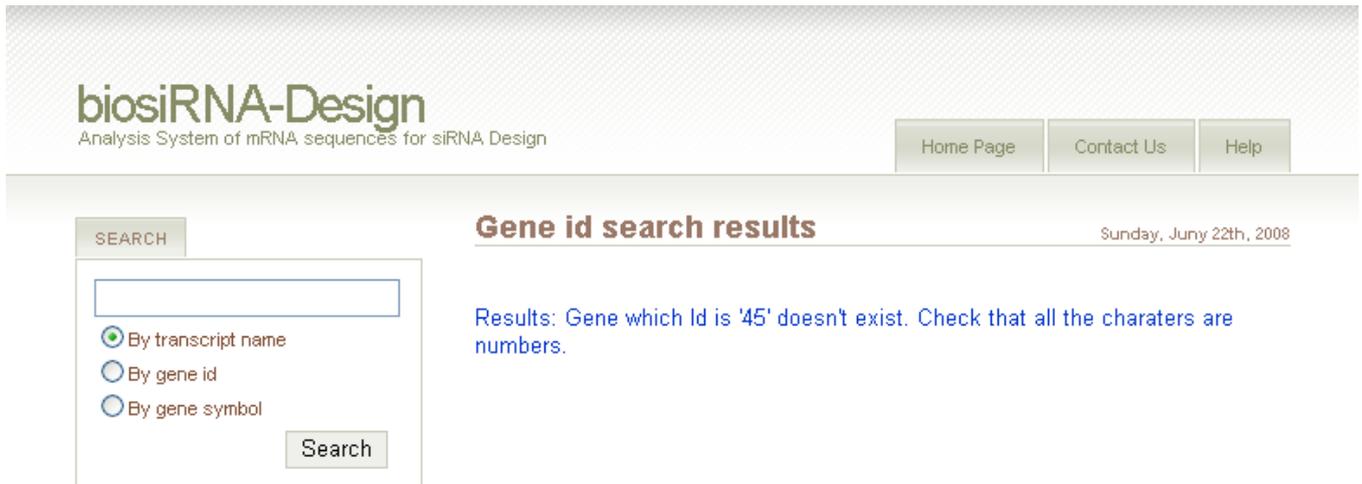


Fig. #13 Utilización de mensajes para evitar búsquedas innecesarias en la base de datos.

Cuando en la aplicación ocurren errores inesperados, esta los trata de manera que no se pierde su integridad. Al ocurrir estos sucesos, el sistema los captura y redirecciona el usuario a una página de error mostrándole el mensaje del error y la indicación de que lo reporte al administrador del sistema.

3.10 Seguridad

La seguridad de la herramienta que se presenta se evidencia a través del servidor de base de datos PostgreSQL. A la información almacenada en este solo tendrá acceso el administrador del sistema, evitando así cambios no deseados en los datos.

En la aplicación, se evidencia también la seguridad a través de la dirección web que se muestra en el navegador donde se está ejecutando el sistema. Todas las páginas se muestran visualizan bajo casi una misma dirección web, y las direcciones no coinciden con el nombre de la página mostrada. Con esto se evita que se acceda a las distintas páginas de la aplicación sin las precondiciones correspondientes, ya que se afectaría la integridad de toda la herramienta.

3.11 Conclusiones

En este capítulo se ha determinado el diseño como una redefinición del ya establecido en la versión anterior de la aplicación. Para ello se hizo primeramente una descripción de la arquitectura que tendrá la herramienta y luego se presentaron los diagramas de clases web del diseño y de interacción, sentando así las bases para una implementación más detallada y organizada de los procesos que deberá llevar a cabo el sistema. Además se definió el diseño de la base de datos a través del modelo lógico y el modelo físico. Como paso posterior se creó la base de datos y se describieron las principales tablas contenidas en la misma, incluyendo nuevas tablas como parte de las mejoras hechas al proceso de diseño de siRNA. Se obtuvo el modelo de despliegue con el cual quedó representado la distribución física del sistema y se mostró la forma en que será implementada la seguridad del mismo y el tratamiento de errores. Con todos estos artefactos generados mediante la puesta en práctica de la metodología de desarrollo OpenUp se dejaron claras las bases para la implementación de la herramienta.

Capítulo 4

Implementación y prueba

Introducción

En este capítulo se hará el modelo de implementación como artefacto de más importancia en el flujo de trabajo actual, teniendo en cuenta las bases creadas por el flujo de trabajo de diseño anteriormente desarrollado. Se optimizará el costo de ejecución del algoritmo logrado en la versión anterior de la aplicación, para la búsqueda de sitios blancos. Se obtendrá el algoritmo para la búsqueda de sitios blancos en una secuencia de mRNA teniendo en cuenta los SNP detallando su funcionamiento. Se van a describir en detalle las pruebas que se le han realizado al sistema, y como evidencia de ello se tendrán en cuenta esencialmente los casos de prueba de caja negra.

4.1 Diagrama de Componentes.

Un diagrama de componentes es la representación de la forma en que los componentes físicos de un sistema serán separados (pueden ser: archivos, cabeceras, módulos, paquetes). Muestra las dependencias entre estos componentes.

Son utilizados para modelar la vista estática de un sistema. Muestra la organización y las dependencias que existen entre un conjunto de componentes.

4.1.1 Diagramas de Componentes de la aplicación

Se confeccionaron dos diagramas de componentes. Cada uno fue realizado de acuerdo al tipo de acción con la que se relaciona.

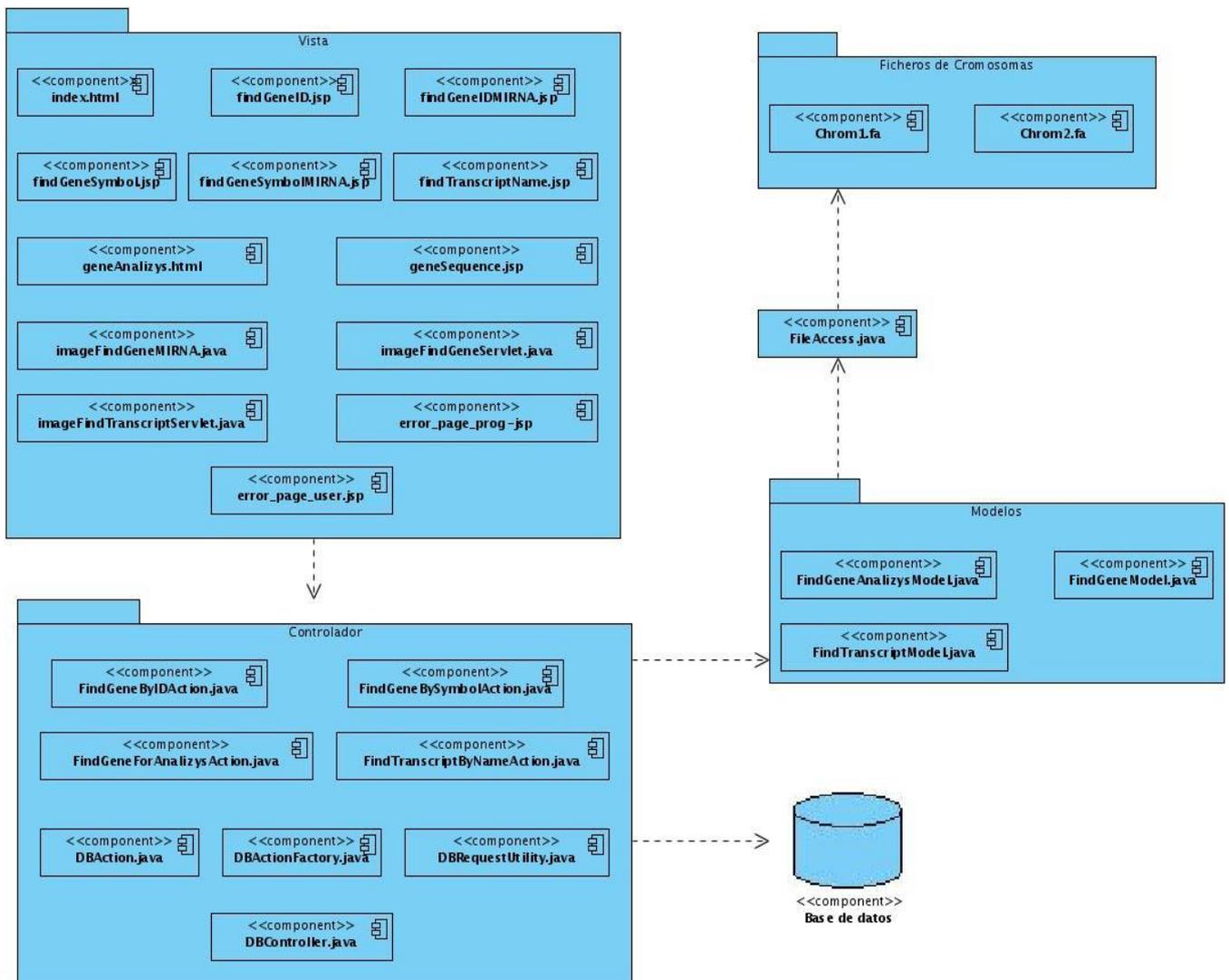


Fig. #14. Diagrama de componentes que agrupa las acciones con conexión a la base de datos.

Este primer diagrama (Fig. #14) agrupa todos los componentes que se relacionan cuando se llevan a cabo las acciones que tienen conexión con la base de datos. A continuación se describe cada uno de los paquetes y componentes relacionados.

- Vista: Este paquete contiene todos los ficheros que se encargan de visualizar la información resultante de las operaciones solicitadas.

- Controlador: En este paquete están agrupados los ficheros que se encargan de capturar todos los eventos que suceden como producto de la interacción del cliente con el sistema.
- Base de Datos: Este componente representa el contenido almacenado en la base de datos del sistema.
- Modelos: Este paquete agrupa los ficheros encargados de almacenar con una lógica definida la información recuperada de la base de datos resultado de las solicitudes del usuario.
- FileAcces.java: Este fichero es el responsable de acceder y extraer los datos solicitados por el usuario, de los ficheros que contienen la información de los cromosomas del genoma humano.
- Ficheros de Cromosomas: Este paquete contiene todos los ficheros que contienen la información de los cromosomas.

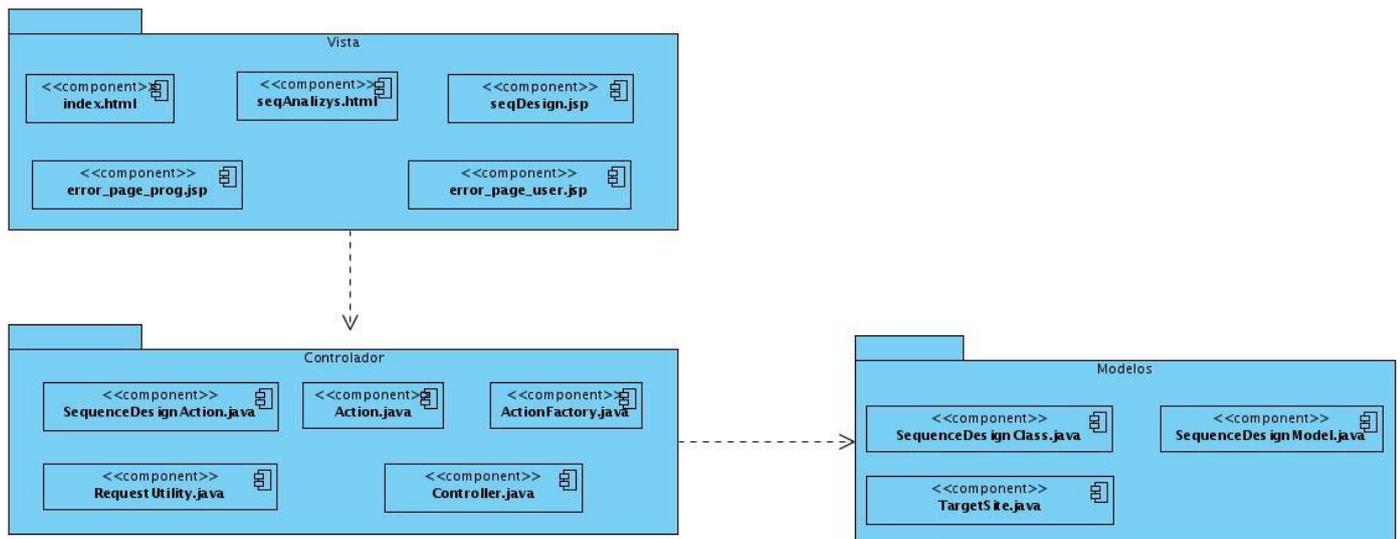


Fig. #15. Diagrama de componentes que agrupa las acciones sin conexión a la base de datos.

Con este segundo diagrama (Fig. #15) se muestran las relaciones que tienen los paquetes de ficheros responsables de llevar a cabo las acciones que no se realizan con conexión a la base de datos. A continuación se describen los paquetes relacionados.

- Vista: Este paquete contiene todos los ficheros que se encargan de visualizar la información

resultante de las operaciones solicitadas.

- **Controlador:** En este paquete están agrupados los ficheros que se encargan de capturar todos los eventos que suceden como producto de la interacción del cliente con el sistema.
- **Modelos:** Este paquete agrupa los ficheros encargados de almacenar con una lógica definida la información manejada por la aplicación de acuerdo al tipo de procedimiento que se esté realizando.

Con el diagrama de componentes se detalla el funcionamiento interno de la herramienta propuesta haciendo uso del patrón arquitectónico Controlador Frontal. Este patrón separa los datos con que trabaja la aplicación, la interfaz de usuario y la lógica de control en tres componentes distintos permitiendo que se estructuren las operaciones que se realizan de una mejor manera.

Un flujo de eventos ejemplo puede ser el siguiente:

1. El usuario interactúa con la interfaz de alguna forma.
2. El controlador recibe (por parte de los objetos de la interfaz-vista) la notificación de la acción solicitada por el usuario. El controlador gestiona el evento que llega.
3. El controlador accede al modelo, realizando la operación de acuerdo a la acción solicitada por el usuario.
4. El controlador delega a los objetos de la vista la tarea de desplegar la interfaz de usuario. La vista obtiene sus datos del modelo para generar la interfaz apropiada para el usuario donde se refleja los cambios en el modelo.
5. La interfaz de usuario espera nuevas interacciones del usuario, comenzando el ciclo nuevamente.

En la implementación de la versión anterior no se utilizó ningún patrón de arquitectura [9], lo cual no permitió que se implementara estructurada y organizadamente a diferencia de la versión actual 1.1. Esto limitó que se le hicieran cambios con el objetivo de incorporar nuevas funcionalidades sin alterar la estructura del sistema, a diferencia de esta nueva versión. Como hace uso del MVC se convierte en una herramienta fácilmente extensible de forma que las diferentes “piezas” que lo constituyen se pueden construir por separado y luego unirlos en tiempo de ejecución. Si uno de sus componentes posteriormente se quiere retirar o agregar otros, puede realizar el cambio sin que los otros componentes se vean afectados.

4.2 Optimización del costo de ejecución del algoritmo de búsqueda de sitios blancos

Para mejorar el proceso de diseño que se plantea con el siguiente trabajo, no solo en cuanto a estructura interna del producto como la arquitectura, sino también en cuanto a la rapidez con que se procesan y responden las solicitudes del usuario, surge la necesidad de optimizar el costo de ejecución de los algoritmos logrados en la versión anterior de la aplicación, específicamente el de la búsqueda de los sitios blancos. Para ello se ha creado un algoritmo más óptimo que el implementado anteriormente y como prueba de ello se calcula a continuación el costo de ejecución de los mismos.

Descripción del algoritmo de búsqueda de sitios blancos en la versión 1.0 [9]

- Se inicializan las variables locales para contar el número de reglas marcadas y el número de reglas cumplidas en la subsecuencia que se está analizando.
- Se recorre con una estructura iterativa la secuencia completa analizando cada subsecuencia de 21 nucleótidos.
- Se comprueba en la subsecuencia si comienza con AA.
- Se cuenta, con una estructura iterativa, la cantidad de C y de G que existe para comprobar si cumple con el por ciento especificado por el usuario.
- Se comprueba, con una estructura iterativa de orden constante, cuantas A y T existen entre las posiciones 15 y 19.
- Se verifica si hay una A en la posición 19, una T en la posición 10 y una A en la posición 3.
- Se comprueba si no hay ni C ni G en la posición 19, y si tampoco existe G en la posición 13.
- Al final de todas estas comprobaciones, se verifica si el número de reglas marcadas es igual al número de reglas cumplidas, si coinciden se añade el sitio blanco y si no, no se añade.
- Después se pasa a comprobar los siguientes 21 nucleótidos, a partir del primero de la subsecuencia anterior.

Orden del algoritmo anteriormente descrito: $O(kn)$, k es una constante.

Descripción del algoritmo de búsqueda de sitios blancos de la versión 1.1

- Se inicializan las variables locales para contar el número de reglas marcadas, el número de reglas cumplidas en la subsecuencia que se está analizando y la cantidad de C y G que contiene la primera subsecuencia. Este último paso se realiza a través de una estructura iterativa, pero una sola vez, la primera. Después esta variable se modifica a medida de que se avanza en la secuencia completa.
- Se recorre con una estructura iterativa la secuencia completa analizando cada subsecuencia de 21 nucleótidos.
- Se comprueba en la subsecuencia si comienza con AA, si hay una T en la posición 10, si hay una A en la posición 3 y en 19.
- Se comprueba después la cantidad de C y de G que se tiene hasta el momento para comprobar si cumple con el por ciento especificado por el usuario.
- Se comprueba, con una estructura iterativa de orden constante, cuantas A y T existen entre las posiciones 15 y 19.
- Se comprueba si no hay ni C ni G en la posición 19, y si tampoco hay una G en la posición 13.
- Al final de todas estas comprobaciones, se verifica si el número de reglas marcadas es igual al número de reglas cumplidas, o si el número de reglas marcadas es mayor o igual a un número previamente definido por el usuario. Si una de estas condiciones se cumplen, se añade el sitio blanco y si no, no se añade.
- Después se pasa a comprobar los siguientes 21 nucleótidos, a partir del primero de la subsecuencia anterior.

Orden del algoritmo anteriormente descrito: $O(n)$

4.3 Obtención del algoritmo para la búsqueda de sitios blancos en la secuencia de mRNA teniendo en cuenta los SNP

Dentro de los requerimientos funcionales del sistema está realizar el análisis de las secuencias mRNA. Existen dos formas de analizar una secuencia como ya se ha visto: teniendo en cuenta los SNP y de forma simple; o sea, mostrar los sitios blancos en la cadena, si importan si contienen SNP o no. Para el primer caso se hizo necesario estudiar las reglas que brindan los especialistas para la obtención de un algoritmo que cubra sus necesidades teniendo en cuenta sus especificaciones. Luego de esto se ha obtenido el algoritmo que se aprecia a continuación:

Algoritmo para reportar sitios blancos que contienen SNP:

- Una vez que se ha obtenido el sitio blanco a través del algoritmo para la obtención de los mismos, se verifica si existe la información de los SNP.
- Posteriormente se escoge en la información de los SNP la subsecuencia correspondiente a la subsecuencia que estamos analizando.
- Se comprueba si existe regiones SNP y se identifican. En el caso en que el usuario solicite la exclusión de los sitios blancos con SNP, este sitio blanco no se añade a la colección y se pasa a la siguiente subsecuencia de nucleótidos. En caso de que el usuario no especifique la opción anteriormente comentada, se prosigue como a continuación se describe.
- Se añade al sitio blanco los datos de la existencia de SNP y se añade a la colección de sitios obtenidos.
- Posteriormente se sigue a la próxima subsecuencia de nucleótidos.

4.4 Modelo de Prueba

El objetivo del flujo de trabajo de prueba es medir la calidad del producto de software creado. En todas las fases del desarrollo del proyecto se deben ir realizando las pruebas aunque es principalmente en esta donde se desarrolla con mayor rigor.

4.4.1 Casos de prueba de caja negra

Las pruebas de caja negra se refieren a las pruebas que se llevan a cabo sobre la interfaz del software, por lo que los casos de prueba tienen como objetivo demostrar que las funcionalidades del software son operativas, que los datos de entrada se aceptan de forma adecuada y que se produce una salida correcta garantizando la integridad de la información que se almacena y procesa. Se examinan

fundamentalmente algunos aspectos del modelo del sistema sin profundizar mucho en la estructura interna del software.

Caso de Prueba 1: Analizar una secuencia mRNA

Flujo Central:

- 1) El especialista selecciona la opción “Search and Analyze Gene” para analizar la secuencia de un gen.
- 2) El sistema visualiza la interfaz para buscar los datos de un gen (cromosoma, id y símbolo) y luego analizar su secuencia.
- 3) El especialista selecciona un criterio de búsqueda (id o símbolo del gen) y ordena la búsqueda.
 - El sistema:
- 4) Se valida que se entren datos.
- 5) Se busca los datos del gen especificado (cromosoma, id y símbolo, cadena de nucleótidos).
- 6) Se muestra los datos del gen dando la posibilidad de realizar un análisis visualizando las reglas a seleccionar para el mismo.
- 7) El especialista selecciona las reglas para analizar la secuencia mRNA y ordena el análisis.
- 8) El sistema verifica que al menos se seleccione una regla y que la secuencia sea correcta.
- 9) El sistema verifica si el especialista ha seleccionado la regla de excluir los SNP de los resultados (“Exclude SNP region from the design (No For Pasted Sequence)”).
- 10) El sistema busca sitios blancos en la secuencia mRNA y los SNP asociados a la misma.
- 11) El sistema muestra gráficamente los resultados de la búsqueda de los sitios blancos en la cadena mRNA y señala los sitios blancos con SNP ordenados por la cantidad de reglas cumplidas de las seleccionadas.
- 12) El sistema muestra una gráfica con los SNP de la secuencia.
- 13) El sistema brinda la posibilidad de visualizar en detalle la información de un sitio blanco.

Flujo Alternativo:

- 13.1) El especialista selecciona un sitio blanco para ver su información.
 - 13.2) El sistema visualiza los datos del sitio blanco seleccionado (secuencia de nucleótidos, posiciones inicial y final de la misma y regiones SNP).
 - 1.1) El especialista accede a la opción "Analyze mRNA sequence".
 - 1.2) El sistema muestra la interfaz para realizar el análisis de una secuencia mRNA dada.
 - 1.3) El especialista introduce la secuencia mRNA, selecciona las reglas y ordena el análisis.
 - 1.4) El sistema verifica que se entró la secuencia y que esta es correcta.
 - 1.5) El sistema busca sitios blancos en la secuencia mRNA.
 - 1.6) El sistema muestra gráficamente los resultados de la búsqueda de los sitios blancos en la cadena mRNA ordenados por la cantidad de reglas cumplidas.
 - 1.7) Se va a la acción 13 del flujo normal de eventos.
 - 9.1) El sistema busca los sitios blancos en la cadena excluyendo los que contienen SNP.
- Se va a la acción 1.6 de los flujos alternos.

Iteraciones:

Clases válidas	Clases inválidas	Resultado esperado	Resultado de la prueba	Observaciones	
<Se selecciona en Search la opción By gene id el siguiente dato: 2>		<El sistema verifica que el dato entrado sea el id de un gen, luego muestra la información del transcrito seleccionado y las reglas para el análisis, luego busca sitios blancos en la secuencia mRNA, los SNP pertenecientes a la secuencia, luego muestra la secuencia	<Satisfactorio>		

		<p>mRNA resaltando sitios blancos y sus datos, muestra una grafica de la secuencia</p> <p>mRNA resaltando sitios blancos, muestra una gráfica con los SNP asociados al sitio blanco correspondiente</p> <p>y brinda la opción de seleccionar de forma individual el segmento que contiene un sitio blanco específico luego muestra una nueva ventana con la información de los sitios blancos (Segmento de la cadena correspondiente, número de reglas que cumple el transcrito al que pertenece, posición inicial y final, número y SNP en caso de tener).</p> <p>></p>			
<Se selecciona en Search la opción By gene id y se entra el siguiente dato: MNP >		<El sistema verifica que este no es el id de un gen y muestra el siguiente mensaje "Gene which Id is 'mnp' doesn't exist. Check that all the characters are numbers">	<Satisfactorio>		
<Se selecciona la		<El sistema reconoce	<Satisfactorio>		

<p>opción Analyze mRNA sequence y se entra el siguiente dato: 30 ></p>		<p>la secuencia mRNA, busca sitios blancos en la secuencia mRNA, busca los SNP pertenecientes a la secuencia, luego muestra la secuencia mRNA resaltando sitios blancos y sus datos y muestra una grafica de la secuencia mRNA resaltando sitios blancos, si la secuencia solicitada no presenta SNP el sistema no muestra la gráfica con SNP.</p>	<p>orio></p>		
---	--	--	-----------------	--	--

Caso de Prueba 2: Buscar datos de un gen.

Flujo Central.

1) El sistema brinda la opción de escoger un criterio de búsqueda para buscar y mostrar los datos de un gen (“By gene id”, “By gene symbol”).

2) El especialista:

Selecciona un criterio de búsqueda.

Introduce los datos requeridos (id o símbolo del gen)

Ordena la búsqueda.

El sistema:

- 3) Valida que se entre la información de un criterio de búsqueda.
- 4) Busca los datos del gen (número del cromosoma, id y símbolo).
- 5) Busca los datos de los transcritos asociados al gen (posición inicial y final, cantidad de exones, nombre y regiones codificantes y no traducidas de los exones asociados a los transcritos).
- 6) Busca los datos de los miRNA asociados al gen (nombre, PValue, SCORE, posición inicial y final).
- 7) Muestra los datos del gen.
- 8) Muestra gráficamente los datos de los transcritos asociados.
- 9) Muestra gráficamente los datos de los miRNA asociados al gen.
- 10) El sistema brinda la posibilidad de seleccionar un PValue (nivel de significancia) para visualizar los miRNA asociados al gen con un nivel de significancia que como máximo sea el especificado.
- 12) El sistema da la posibilidad mostrar en detalle un transcrito y analizar su secuencia de nucleótidos.
- 13) Se realiza el caso de uso “Buscar datos de un transcrito” y se ejecuta el caso de uso “Realizar análisis de la secuencia mRNA” y así termina la ejecución del caso de uso.

Flujos Alternos:

2.2) El especialista introduce la información del criterio de búsqueda de un gen que no tiene miRNA.

2.2.1) El sistema no muestra los miRNA.

2.3) El especialista introduce la información del criterio de búsqueda de un gen que no tiene transcritos.

2.3.1) El sistema no muestra los transcritos.

13.1) Se culmina la ejecución del caso de uso.

10.1) El especialista selecciona un PValue y ordena que se muestren los miRNA.

10.2) El sistema muestra los miRNA que tengan un valor de significancia menor que el especificado.

3.1) Si El especialista no introduce los datos requeridos el sistema muestra un mensaje de error.

Iteraciones.

Clases válidas	Clases inválidas	Resultado esperado	Resultado de la prueba	Observaciones
<Se selecciona en		<El sistema	<Satisfactorio	

<p>Search la opción By gene id y se introduce el siguiente dato: 400></p>		<p>verifica que el dato entrado es correcto brinda la opción de introducir la información a partir del criterio de búsqueda seleccionado, muestra datos de gen(id y símbolo, cromosoma correspondiente y cantidad de transcritos asociados), busca transcritos asociados al gen, muestra datos de los transcritos asociados al gen (posición inicial y final en la cadena, nombre, exones, regiones codificantes y regiones no traducidas) y muestra los miRNA correspondientes al gen y sus datos ></p>	<p>></p>	
--	--	---	-------------	--

<p><Se selecciona en Search la opción By gene symbol y se entra el siguiente dato: ARL1></p>		<p><El sistema verifica que el dato entrado es válido, brinda la opción de introducir la información a partir del criterio de búsqueda seleccionado, muestra datos de gen(id y símbolo, cromosoma correspondiente y cantidad de transcritos asociados), busca transcritos asociados al gen, muestra datos de los transcritos asociados al gen (posición inicial y posición final en la cadena, nombre, exones, regiones codificantes y regiones no traducidas) y muestra los miRNA correspondientes al gen y sus datos></p>	<p><Satisfactorio ></p>	
<p><Se selecciona en Search a opción By gene symbol y se entra el</p>		<p><El sistema verifica que el dato entrado es incorrecto y muestra el mensaje</p>	<p><Satisfactorio ></p>	

siguiente dato: GSR>		de error: “Gene which symbol is ‘gsr’ doesn’t exist. Check that the letters are capitalized”>		
<Se selecciona la opción By gene id y se entra el siguiente dato: ADR2>		<El sistema verifica que el dato entrado es incorrecto y muestra el siguiente mensaje de error: “Gene which Id is ‘adr2’ doesn’t exist. Check that all the characters are numbers”>	<Satisfactorio >	

Caso de Prueba 3: Mostrar un transcrito

Descripción General

Mediante este caso de uso el sistema da la posibilidad de mostrarle información relevante a los especialistas.

Flujo Central:

El sistema brinda la opción de escoger un criterio de búsqueda para buscar y mostrar los datos de un transcrito.

- 2) El especialista selecciona el criterio de búsqueda “By transcript name”, introduce la información requerida (nombre del transcrito) y ordena la búsqueda.
- 3) El sistema valida que se entre la información referente a al criterio de búsqueda.
- 4) El sistema busca los datos del transcrito (posición inicial y final, cantidad de exones, nombre y regiones codificantes y no traducidas de los exones asociados al transcrito y regiones SNP).
- 5) El sistema busca la secuencia de nucleótidos del transcrito.
- 6) El sistema muestra gráficamente los datos del transcrito.
- 7) El sistema muestra la secuencia de nucleótidos.

8) El sistema da la posibilidad de analizar la secuencia de nucleótidos del transcrito visualizando las reglas para dicho análisis.

9) Se realiza el caso de uso " Realizar análisis de la secuencia mRNA" a partir de la acción 4 del flujo normal de eventos.

Flujos Alternos:

1.1) El sistema visualiza una interfaz donde muestra los transcritos asociados a un gen y sus datos y da la posibilidad de seleccionar uno de ellos para visualizar en detalle sus datos.

1.2) El especialista selecciona el nombre del transcrito y ordena la búsqueda.

1.3) El sistema busca los datos del transcrito (secuencia de nucleótidos y regiones SNP).

1.4) El sistema muestra gráficamente los datos del transcrito (posición inicial y final, cantidad de exones, nombre y regiones codificantes y no traducidas de los exones asociados al transcrito y regiones SNP).

1.5) Se realizan las acciones de la 6 a la 9 del flujo normal de eventos.

3.1) El sistema muestra un mensaje de error en caso de que el especialista no haya introducido la información solicitada.

9.1) Se sale del caso de uso.

Iteraciones:

Clases válidas	Clases inválidas	Resultado esperado	Resultado de la prueba	Observaciones
<Se selecciona en la sección "Search" la opción By transcript name y se entra el siguiente dato: NM_001025091>		<El sistema verifica que el dato entrado es correcto, busca exones asociados al transcrito, muestra id y nombre del transcrito y cantidad de exones, muestra una gráfica con datos de los exones asociados	<Satisfactorio >	

		al transcrito y muestra una gráfica de los SNP presentes en el transcrito y su cadena mRNA.>		
<Se selecciona la opción By transcript name y se entra un transcrito que no tiene SNP>		<El sistema verifica que la entrada es válida, y no muestra SNP>	<Satisfactorio >	
<Se selecciona la opción By transcript name y se entra el siguiente dato: NMOR>		<El sistema verifica que el dato entrado en incorrecto y muestra el mensaje de error: "The transcripts which name is 'nmor' doesn't exist. Check that the letters are capitalized.">	<Satisfactorio >	

Con las pruebas realizadas se demuestra que el sistema es seguro no solo en cuanto a la información que se maneja, sino también en referencia a los resultados que se generan como resultado de las solicitudes del usuario. Se controla que no hayan entradas incorrectas para brindar soluciones de calidad, así como también se controla y valida cada solicitud realizada a través de la dirección web.

4.5 Conclusiones

Se hizo una descripción detallada de los diagramas de componentes creados, correspondientes al modelo de implementación. Se analizó además el algoritmo para la búsqueda de sitios blancos obtenido en la versión anterior y se optimizó el tiempo de ejecución del mismo, con lo que se logró un mejoramiento en cuanto al tiempo requerido por el sistema para ejecutarlo y mostrar los resultados. Se obtuvo un nuevo algoritmo con este mismo fin, pero que incluye nuevos criterios (SNP). Se implementó el sistema a partir de los artefactos generados en los flujos de trabajo anteriores. Se realizaron las pruebas de caja negra mediante las que se comprobó la calidad y el nivel de eficiencia de las funcionalidades logradas. Durante la etapa de prueba se encontraron algunas inconformidades en cuanto a la concepción del sistema. Las mismas fueron rectificadas a medida que se fueron desarrollando las pruebas posteriores.

Conclusiones Generales

Con el presente trabajo se brinda una propuesta de solución al problema planteado por el especialista referente al diseño de los siRNA y el silenciamiento génico, capaz de brindar servicios de eficiencia no solo en materia de funcionalidad sino también de confiabilidad y seguridad de la información que se maneja. La herramienta se crea con el propósito de que sea utilizada principalmente por del Centro de Ingeniería Genética y Biotecnología en Cuba y demás instituciones de carácter tanto nacional como internacional que necesiten hacer uso de la misma. Como resultado del trabajo desarrollado se ha obtenido una aplicación capaz de realizar estudios especializados de carácter básico enfocados al diseño de los siRNA, brindando resultados de valor para los especialistas en los que se muestra suficiente información de los sitios blancos que se utilizarán para los estudios que se llevarán a cabo posteriormente por parte de los especialistas. Una adecuada explotación de esta herramienta informática, puede llegar a convertirla en una poderosa fuente de avance en los estudios científicos en el campo de la genética.

El sistema se desarrolló siguiendo la metodología OpenUp, y se utilizaron representaciones de todos los artefactos generados por la metodología para la modelación de todas las fases del proyecto. Como resultado del desarrollo el sistema cuenta con ambiente fácil de entender y usar por el personal calificado para el uso del mismo. Se optimizó el costo de ejecución del algoritmo para la búsqueda de sitios blancos implementado en la versión 1.0 del sistema y se obtuvo un algoritmo para realizar el análisis de las secuencias tomando en consideración los SNP. El sistema adaptó una nueva arquitectura haciendo uso del patrón MVC utilizando específicamente el patrón Controlador Frontal tomando una estructura flexible y adaptable a próximos cambios.

Por todo lo anteriormente expuesto se concluye que los objetivos trazados para el desarrollo del presente trabajo han sido cumplidos satisfactoriamente. Se incluyen en el mismo una serie de consideraciones a tener en cuenta para el seguimiento de este proyecto.

Recomendaciones

Con el desarrollo del trabajo se han cumplido el objetivo general del mismo y los objetivos de carácter específico, a pesar de ello todavía existen nuevas expectativas e ideas que han ido surgiendo, lo cual da la posibilidad de que se llegue a crear una aplicación más potente que brinde toda clase de servicios sobre la base de las investigaciones en el campo de la microbiología y específicamente en los estudios genéticos, por tanto recomendamos:

- Continuar el desarrollo de la herramienta con el objetivo de extender el propósito de la misma con otras funcionalidades de interés, que se desean para versiones superiores, como el efecto “off-target”.
- Aplicar el sistema a todas las instituciones nacionales e internacionales donde pueda ser usado, con el propósito de aumentar el desarrollo en la esfera de los estudios genéticos.
- Hacer uso de un framework en versiones superiores del sistema, con el objetivo de hacer más flexible y reutilizable la arquitectura lograda.

Referencias bibliográficas

[1] Caenorhabdtis elegans COMO MODELO DE ESTUDIO DEL DESARROLLO. [Consultado el: 27 de septiembre de 2007]. Disponible en:
<http://bq.unam.mx/mensajebioquimico>

[2]- siRNA Design Tools. [Consultado el: 6 de Diciembre de 2007]. Disponible en:
http://www.Ambion.com/techlib/misc/siRNA_tools.html

[3]- siRNA Design Software. [Consultado el: 6 de Diciembre de 2007]. Disponible en:
<http://i.cs.hku.hk/~sirna/software/sirna.php>

[4]- Custom siRNA Design Tool. [Consultado el: 6 de febrero de 2007]. Disponible en:
<http://www.dharmacon.com/sidesign/default.aspx>

[5]-DEQOR. [Consultado el 17 de junio de 2008]. Disponible en:
<http://cluster-1.mpi-cbg.de/Deqor/deqor.html>

[6]IDT's RNAi Design.[Consultado el: 6 de Diciembre de 2007]. Disponible en:
<http://www.idtdna.com/scitools/applications/rnai/rnai.aspx>

[7]GenScript siRNA Target Finder [Consultado el: 6 de Diciembre de 2007]. Disponible en:
<https://www.genscript.com/ssl-bin/app/rnai>.

[8]siDirect. [Consultado el 10 de febrero de 2008]. Disponible en:
<http://genomics.jp/sidirect/index.php?type=fc>

[9] Correa Y.,Díaz Y.C. Sistema de análisis de secuencias mRNA para el diseño de siRNA, Julio, 2007.

[10]-Impacto de la Bioinformática en las ciencias biomédicas. [Consultado el: 9 de Diciembre de 2007]
Disponible en:

http://bvs.sld.cu/revistas/aci/vol11_4_03/aci07403.htm

[11]-SiRNA [Consultado el: 11 de Diciembre de 2007].Disponible en:

<http://es.wikipedia.org/wiki/SiRNA>

[12]-ARN mensajero [Consultado el : 15 de Diciembre de 2007]. Disponible en:

http://es.wikipedia.org/wiki/ARN_mensajero

[13]Single nucleotide polymorphism.[Consultado el 17 de Junio de 2008

[\]http://en.wikipedia.org/wiki/Single_nucleotide_polymorphism](http://en.wikipedia.org/wiki/Single_nucleotide_polymorphism)

[14]- La tecnología J2EE [Consultado el 8 de Febrero de 2008]. Disponible en:

<http://www.jtech.ua.es/j2ee/2003-2004/present.htm>

[15]- Herramientas CASE [Consultado el 8 de Febrero de 2008]. Disponible en:

<http://www.cyta.com.ar/biblioteca/bddoc/bdlibros/proyectoinformatico/libro/c5/c5.htm>

[16]- Visual Paradigm. [Consultado el: 8 de Febrero de 2008]. Disponible en:

http://alarcos.inf-cr.uclm.es/per/fgarcia/isofware/doc/LabTr1_VP.pdf.

[17]-NetBeans. [Consultado el: 8 de Febrero de 2008]. Disponible en:

http://es.wikipedia.org/wiki/NetBeans#NetBeans_IDE

[18]- PostgreSQL [Consultado el: 8 de Febrero de 2008]. Disponible en:

<http://es.wikipedia.org/wiki/PostgreSQL>

[19]- Características de PostgreSQL [Consultado el: 8 de Febrero de 2008]. Disponible en:

<http://www.sobl.org/traduccion/practical-postgres/node19.html>

[20]- Tomcat [Consultado el : 8 de Febrero de 2008]. Disponible en:

http://es.wikipedia.org/wiki/Apache_Tomcat#Tomcat_5.5.x

[21]- Misión del CIGB. [Consultado el 9 de Febrero de 2008]. Disponible en:
<http://www.cigb.edu.cu/pages/>

Bibliografía

- ARN mensajero [Consultado el: 15 de Diciembre de 2007]. Disponible en:
http://es.wikipedia.org/wiki/ARN_mensajero

- Caenorhabditis elegans COMO MODELO DE ESTUDIO DEL DESARROLLO. [Consultado el: 27 de septiembre de 2007]. Disponible en:
<http://bq.unam.mx/mensajebioquimico>

- Características de PostgreSQL [Consultado el: 8 de febrero de 2008]. Disponible en:
<http://www.sobl.org/traduccion/practical-postgres/node19.html>.

- Correa Y., Díaz Y.C. “Sistema de análisis de secuencias mRNA para el diseño de siRNA”, Julio, 2007.

- Custom siRNA Design Tool. [Consultado el: 6 de febrero de 2007]. Disponible en:
<http://www.dharmacon.com/sidesign/default.aspx>

- DEQOR. [Consultado el 17 de junio de 2008]. Disponible en:
<http://cluster-1.mpi-cbg.de/Deqor/deqor.html>

- siRNA Design Tools. [Consultado el: 6 de diciembre de 2007]. Disponible en:
http://www.Ambion.com/techlib/misc/siRNA_tools.html

- siRNA Design Software. [Consultado el: 6 de Diciembre de 2007]. Disponible en:
<http://i.cs.hku.hk/~sirna/software/sirna.php>

- GenScript siRNA Target Finder [Consultado el: 6 de diciembre de 2007]. Disponible en:
<https://www.genscript.com/ssl-bin/app/rnai>

- Herramientas CASE [Consultado el 8 de febrero de 2008]. Disponible en:
<http://www.cyta.com.ar/biblioteca/bddoc/bdlibros/proyectoinformatico/libro/c5/c5.htm>

- IDT's RNAi Design. [Consultado el: 6 de diciembre de 2007]. Disponible en:
<http://www.idtdna.com/scitools/applications/rnai/rnai.aspx>

- Impacto de la Bioinformática en las ciencias biomédicas.[Consultado el: 9 de diciembre de 2007]
Disponible en
http://bvs.sld.cu/revistas/aci/vol11_4_03/aci07403.htm

- La tecnología J2EE [Consultado el 8 de febrero de 2008]. Disponible en:
<http://www.jtech.ua.es/j2ee/2003-2004/present.htm>

- Misión del CIGB. [Consultado el 9 de febrero de 2008]. Disponible en:
<http://www.cigb.edu.cu/pages/>

- NetBeans. [Consultado el: 8 de febrero de 2008]. Disponible en:
http://es.wikipedia.org/wiki/NetBeans#NetBeans_IDE

- PostgreSQL [Consultado el: 8 de febrero de 2008]. Disponible en:
<http://es.wikipedia.org/wiki/PostgreSQL>

- siDirect. [Consultado el 10 de febrero de 2008]. Disponible en:
<http://genomics.jp/sidirect/index.php?type=fc>

- Single nucleotide polymorphism.[Consultado el 17 de junio de 2008]
http://en.wikipedia.org/wiki/Single_nucleotide_polymorphism

- SiRNA [Consultado el: 11 de diciembre de 2007].Disponible en:
<http://es.wikipedia.org/wiki/SiRNA>

- Tomcat [Consultado el: 8 de febrero de 2008]. Disponible en:
http://es.wikipedia.org/wiki/Apache_Tomcat#Tomcat_5.5.x

- Visual Paradigm. [Consultado el: 8 de febrero de 2008]. Disponible en:
http://alarcos.inf-cr.uclm.es/per/fgarcia/isoftware/doc/LabTr1_VP.pdf.

Anexos

Anexo 1 “Diagramas de secuencia”

Diagrama de secuencia CU Mostrar datos de un gen. (Escenario 1)

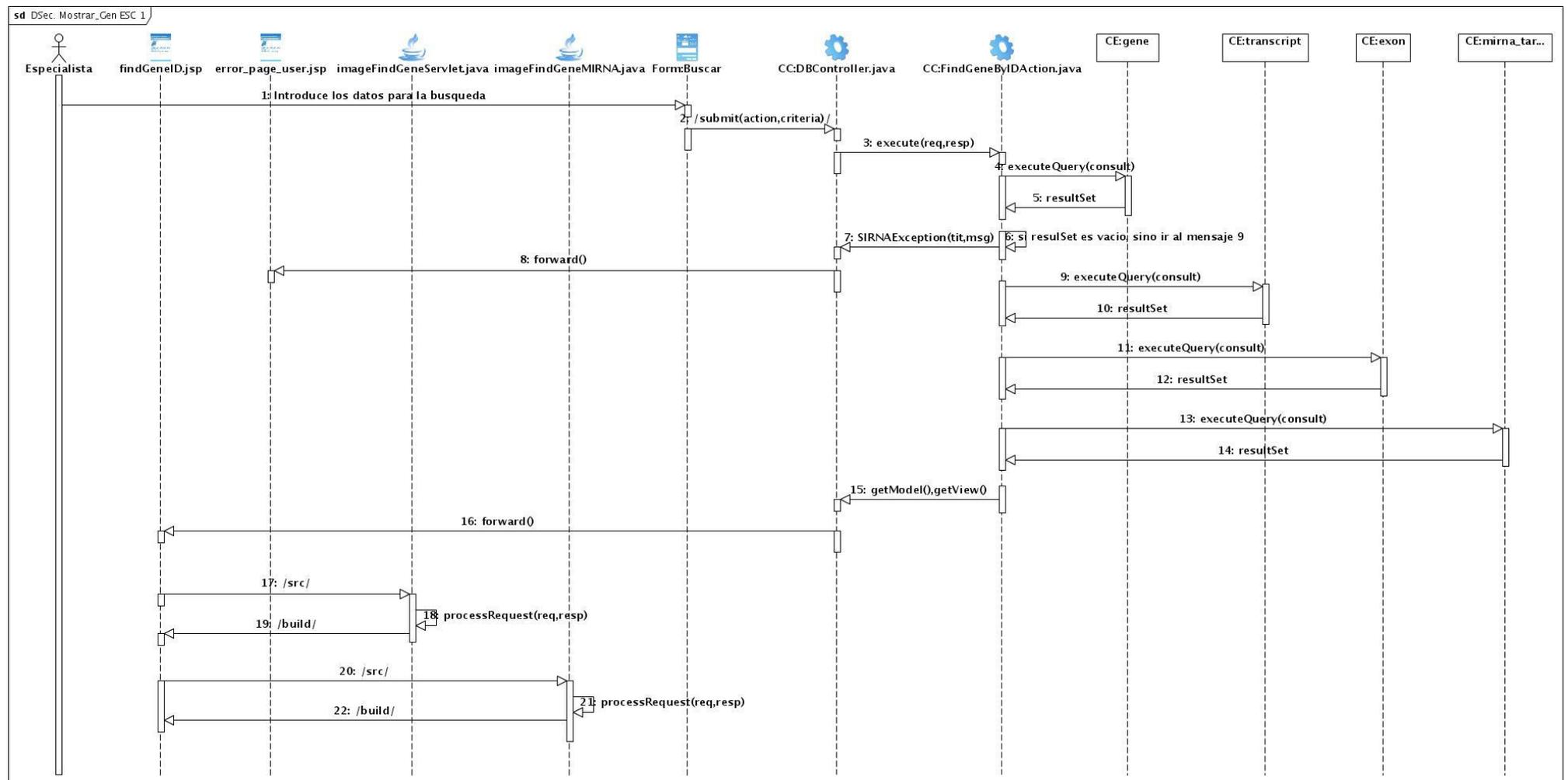


Diagrama de secuencia CU Mostrar Genes. (Escenario 2)

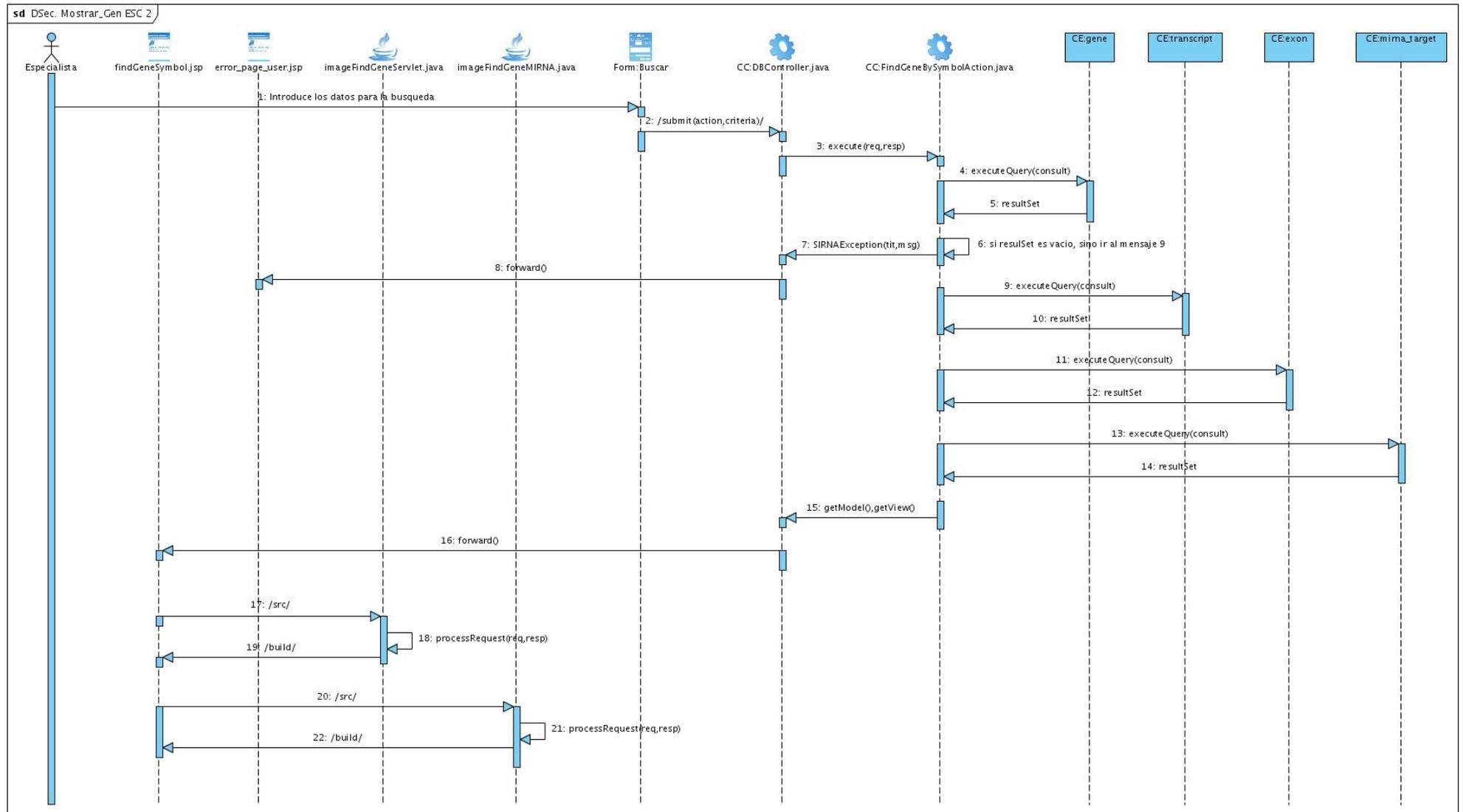


Diagrama secuencia CU Buscar datos de un transcrito

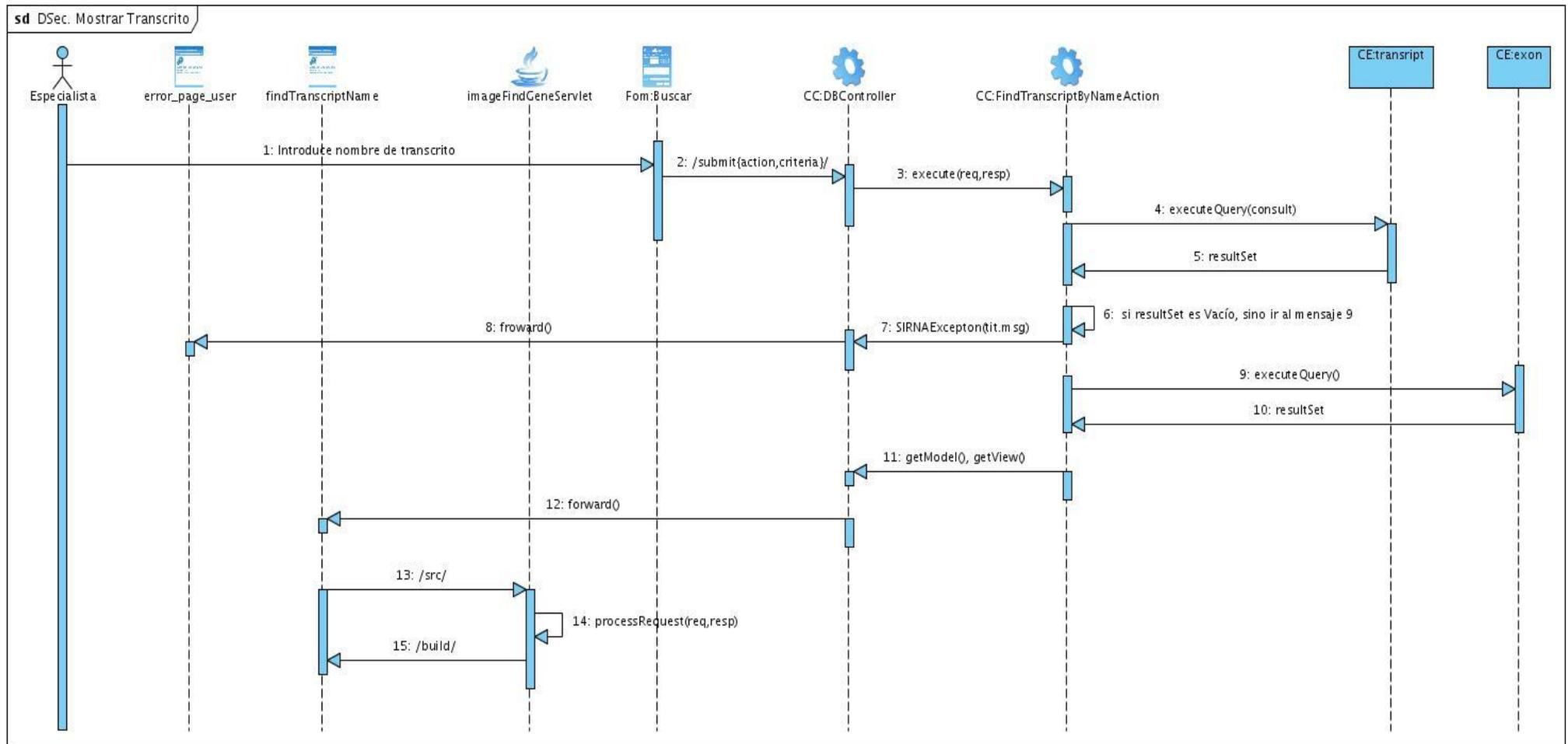


Diagrama de secuencia CU Realizar análisis de la secuencia mRNA (Escenario 1)

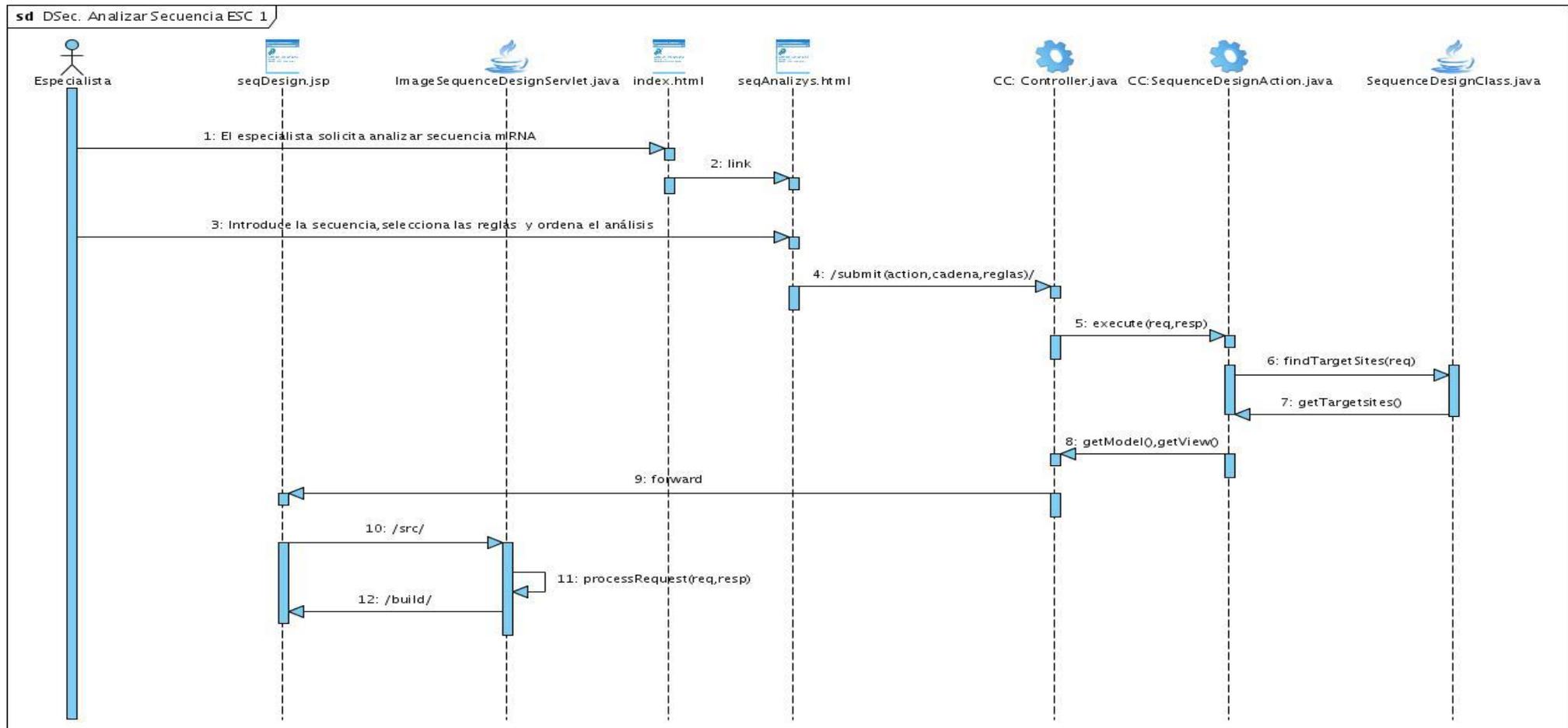
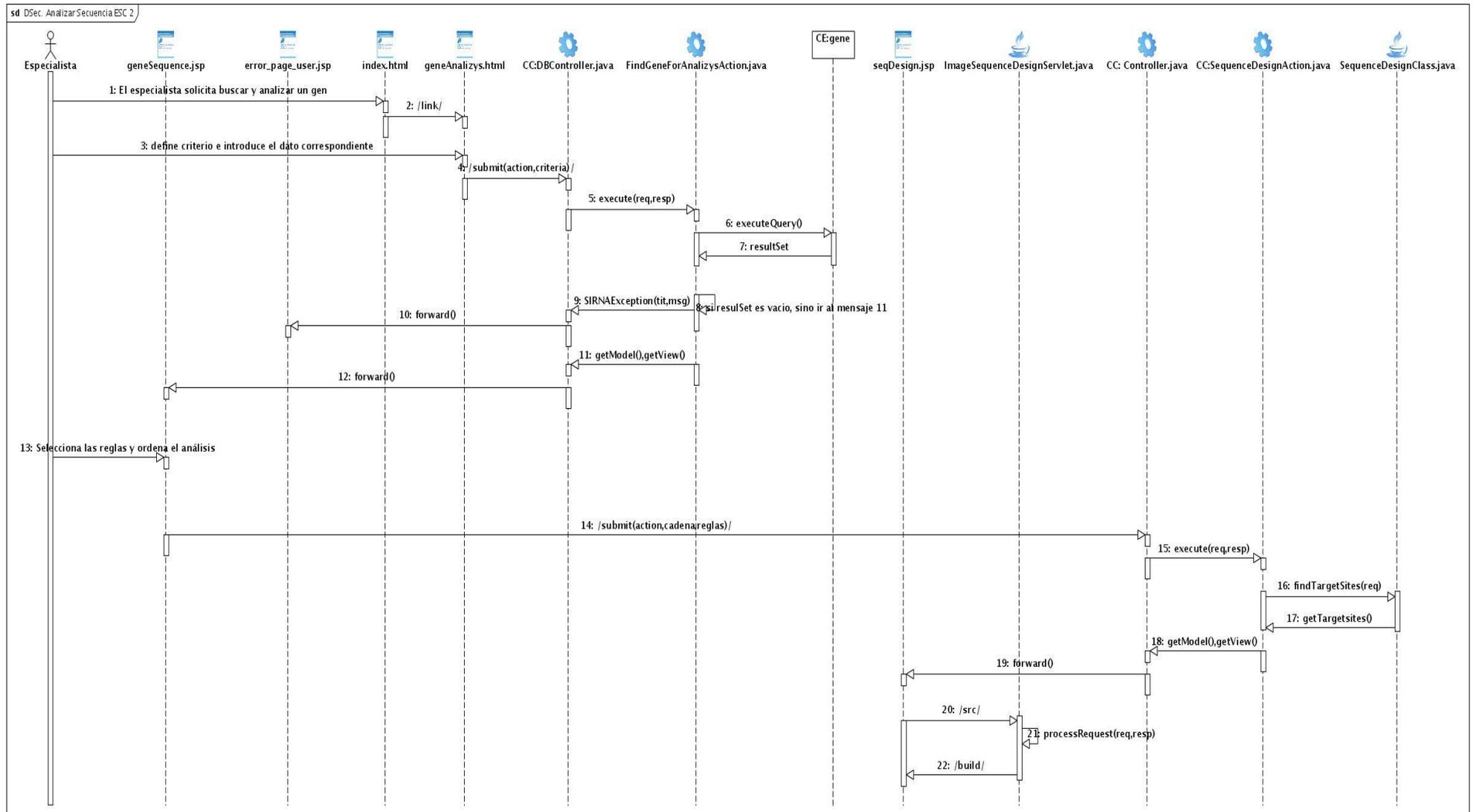


Diagrama de secuencia CU Realizar análisis de la secuencia mRNA. (Escenario 2)



Anexo 2 “Pantallas del sistema”

Inicio

En esta primera página aparece una descripción del propósito de la herramienta. Contiene un formulario para realizar la búsqueda de un gen o un transcrito determinado, según un criterio de búsqueda que puede ser nombre (para un transcrito) e id o símbolo (para un gen). Contiene dos vínculos a formularios: uno para analizar la cadena de nucleótidos completa de un gen y otro para analizar una cadena de mRNA brindada por el especialista. Además contiene una sección con vínculos a sitios de interés.

biosiRNA-Design
Analysis System of mRNA sequences for siRNA Design

Home Page Contact Us Help

SEARCH

By transcript name
 By gene id
 By gene symbol

Search

OPTIONS

Analyze mRNA sequence
Search and analyze gene

OTHER SITES

CIGB
NCBI
EBI
Heber Biotec S.A
Bussines Group
OMS
GAVI
EMBnet

Home Page Saturday, Juny 21th, 2008

What does biosiRNA-Design do?

biosiRNA-Design tool has been created to find target sites in a mRNA sequence taking into consideration SNPs. You can choose whether the analysis is going to include SNPs or not. You can input the sequence that is going to be analyzed or select it from our database, and then you can select a set of rules for the design.

This tool shows up a graphic result of the analysis with the target sites, exons, coding and untranslated regions in the sequence.

Búsqueda de un gen

En esta pantalla se muestra cuando se selecciona el criterio de búsqueda "By gene id", se introduce el id del gen (número) y se ordena la búsqueda. Como consecuencia se muestran todos los transcritos asociados a dicho gen y los miRNA que lo silencian. Se brinda la posibilidad de seleccionar un transcrito para analizar su cadena de nucleótidos.

biosiRNA-Design

Analysis System of mRNA sequences for siRNA Design

[Home Page](#)
[Contact Us](#)
[Help](#)

SEARCH

By transcript name
 By gene id
 By gene symbol

Gene id search results

Saturday, June 21th, 2008

Chromosome: 6 Gene id: 23 Symbol: ABCF1 [+]

5' 3'

30667288 30647149

transcript: NM_001025091 [exons:27]
transcript: NM_001090 [exons:26]

- coding region
 - untranslated region

Choose one transcript to analyze it

miRNAs reported for gene 'ABCF1'

miRNAs Total: 42

5' 3'

30667288 30647149

		start: 30666830
mir-0029	PValue: 1.52254E-4 SCORE: 16.8174	end: 30666852
		start: 30666830
mir-0029	PValue: 1.52254E-4 SCORE: 16.5116	end: 30666852

En la pantalla siguiente se muestran los resultados de la búsqueda de un gen específico según el símbolo.

biosiRNA-Design

Analysis System of mRNA sequences for siRNA Design

Home Page
Contact Us
Help

SEARCH

By transcript name
 By gene id
 By gene symbol

OPTIONS

Analyze mRNA sequence
Search and analyze gene

OTHER SITES

CIGB
NCBI
EBI
Heber Biotec S.A
Bussines Group
OMS
GAVI
EMBNet

Gene symbol search results Saturday, June 21th, 2008

Chromosome: 2 Gene id: 98 Symbol: ACYP2 [+]

5' 3'

54385785 54195914

[exons:6]
transcript: NM_138448

■ - coding region ■ - untranscribed region

Choose one transcript to analyze it

▼

miRNAs reported for gene 'ACYP2'

Graphic of miRNAs

miRNAs Total: 30

5' 3'

54385785 54195914

mir-00522	PValue: 1.40547E-4 SCORE: 18.4224	start: 54385757 end: 54385777
mir-00651	PValue: 1.4163E-4 SCORE: 16.7375	start: 54385628 end: 54385647
mir-0030	PValue: 3.56269E-4 SCORE: 18.3101	start: 54385576 end: 54385599

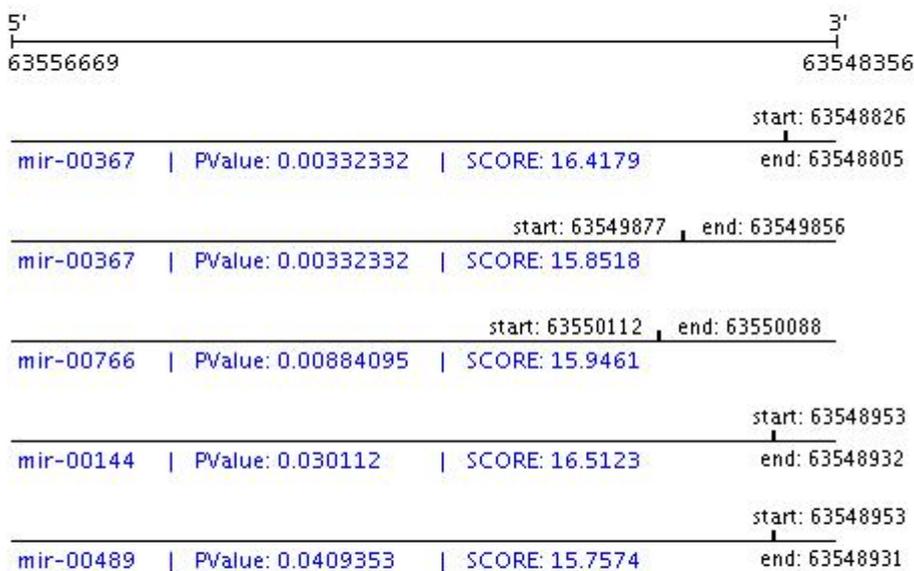
89

Una vez mostrados los datos del gen, los cuales incluyen los miRNA que actúan sobre el mismo, el sistema brinda la posibilidad de especificar un PValue para mostrar solo los miRNA que poseen un valor menor que el especificado.

miRNAs reported for gene 'A1BG'

Graphic of miRNAs

miRNAs Total: 5



Choose one pvalue to be the minimum value

- 0.00332332
- 0.00884095
- 0.030112
- 0.0409353

Show

Búsqueda y análisis de un transcrito

Se selecciona el criterio de búsqueda ("By transcript name") y se ordena la operación. Luego se muestran gráficamente los datos del transcrito especificado como son los exones. De estos últimos se diferencia la región codificante de la no traducida. Se muestran otros datos del transcrito como son el nombre y el id del gen al que pertenece, la posición inicia y la posición final. Se visualizan además los SNP contenidos en el transcrito y la cadena de nucleótidos del mismo. Se da la posibilidad de analizar la secuencia de dicho transcrito mostrando las reglas de restricción que debe seleccionar el especialista antes de ordenar el análisis.

SEARCH

By transcript name
 By gene id
 By gene symbol

OPTIONS

Analyze mRNA sequence
 Search and analyze gene

OTHER SITES

[CIGB](#)
[NCBI](#)
[EBI](#)
[Heber Biotec S.A](#)
[Bussines Group](#)
[OMS](#)
[GAVI](#)
[EMBnet](#)

Transcript name search results

Saturday, Juny 21th, 2008

Transcript Name: NM_001025091 Gene ID: 23

5'

30647149

3'

30667288

3'

30667288

- coding region
 - untranslated region
 - snp region

Analyze transcript sequence

GCGCCAGCTTGGAGAGCCAGCCCCATCGGGGTCCCCGCCGCCG

The algorithm will apply these rules to find target sites:

- It begins with AA
- It contains T in 10th position
- It contains A in 3rd position
- It contains A in 19th position
- There is CG in - %
- There isn't GC in 19th position
- The mRNA sequence contains G in 13th position
- There is 3 or more A/T between the bases 15 and 19
- From the selected rules, to be fulfilled at least:
- Exclude SNPs region from the design (No For Pasted Sequence)

Analizar una secuencia entrada por el usuario

Cuando el especialista accede a la opción "Analyze mRNA sequence" se muestra la pantalla que se ve a continuación donde se da la posibilidad de entrar una cadena de nucleótidos cualquiera para buscar los sitios blancos. Se excluye la regla de incluir los SNP puesto que el sistema no tiene conocimiento de la localización de los mismos en la secuencia que el especialista introdujo.

The screenshot shows the 'Analyze mRNA Sequence' page of the biosiRNA-Design application. The page has a header with the logo 'biosiRNA-Design' and the tagline 'Analysis System of mRNA sequences for siRNA Design'. Navigation links for 'Home Page', 'Contact Us', and 'Help' are in the top right. The main content area is titled 'Analyze mRNA Sequence' and includes a date stamp 'Sunday, Juny 22th, 2008'.

On the left side, there are three sections:

- SEARCH:** A search input field and three radio buttons: 'By transcript name' (selected), 'By gene id', and 'By gene symbol'. A 'Search' button is at the bottom.
- OPTIONS:** Two links: 'Analyze mRNA sequence' and 'Search and analyze gene'.
- OTHER SITES:** A list of external links: CIGB, NCBI, EBI, Heber Biotec S.A, Bussines Group, OMS, GAVI, and EMBnet.

The main content area contains:

- A text prompt: 'Paste your mRNA sequence (5' to 3')'.
- A large text input field with a 'Sample' button to its right. A note above the field says 'Ensure that sequence contains no numbers'.
- A section titled 'The algorithm will apply these rules to find target sites:' followed by a list of rules with checkboxes:
 - It begins with AA
 - It contains T in 10th position
 - It contains A in 3rd position
 - It contains A in 19th position
 - There is CG in 30 - 50 %
 - There isn't GC in 19th position
 - The mRNA sequence contains G in 13th position
 - There is 3 or more A/T between the bases 15 and 19
 - From the selected rules, to be fulfilled al least: 8
- 'Analyze' and 'Reset' buttons at the bottom right.

Luego de realizar el análisis de una secuencia el sistema muestra los resultados obtenidos y da la posibilidad de ver los datos de un sitio blanco específico dando click sobre él.

SEARCH

By transcript name
 By gene id
 By gene symbol

Search

OPTIONS

Analyze mRNA sequence
Search and analyze gene

OTHER SITES

[CIGB](#)
[NCBI](#)
[EBI](#)
[Heber Biotec S.A](#)
[Bussines Group](#)
[OMS](#)
[GAVI](#)
[EMBNet](#)

Sunday, Juny 22th, 2008

Sequence Design

Transcript Name: NM_001085 Gene id: 12 Target Sites Total: 4

5' _____ 3'

mRNA length: 1590

- target site
 - snp region
 - target site with snp

No. Rules

5

5

5

5

mRNA sequence(5' to 3')

```

ATTCATGAAAATCCACTACTCCAGACAGACGGCTTTGGAATCCA
CCAGCTACATCCAGCTCCCTGAGGCAGAGTTGAGAATGGAGAG
AATGTTACCTCTCCTGGCTCTGGGGCTCTTGGCGGCTGGGTTCT
GCCCTGCTGCTCCTTGCCACCCTAACAGCCCACTTGACGAGGA
GAATCTGACCCAGGAGAACCAAGACCGAGGGACACACGTGGA
          
```

The algorithm will apply these rules to find target sites:

It begins with AA
 It contains T in 10th position
 It contains A in 3rd position
 It contains A in 19th position
 There is CG in - %
 There isn't GC in 19th position
 The mRNA sequence contains G in 13th position
 There is 3 or more A/T between the bases 15 and 19
 From the selected rules, to be fulfilled at least:

Exclude SNPs region from the design (No For Pasted Sequence)

Cuando se le da click sobre el sitio blanco se muestra una ventana flotante con los datos correspondientes. En este caso contiene un SNP que se diferencia de los demás nucleótidos en su color para visualizar de forma mas clara su localización.

Target Site Information

Target Site Fields

Target Site 2:

Number of rules: 5 | Start In: 477 | End In: 498

Segment: GCAGCTGAGTATGGGAAATGC

SNP area: GCAGCTG'A'GTATGGGAAATGC

Close

biosiRNA-Design
Analysis System of mRNA sequences for siRNA Design

Sequence Design

Transcript Name: NM_0

5'

mRNA length: 1590

■ - target site ■ - snp region ■ - target site with snp

mRNA sequence(5' to 3')

```

ATTCATGAAAATCCACTACTCCAGACAGACGGCTTTGGAATCCA
CCAGCTACATCCAGCTCCCTGAGGCAGAGTTGAGAATGGAGAG
AATGTTACCTCTCTGGCTCTGGGGCTCTGGCGGCTGGGTTCT
GCCCTGCTGTCCTCTGCCACCCTAACAGCCCACTTGACGAGGA
GAATCTGACCCAGGAGAACCAAGACCAGGGACACACGTGGA

```

The algorithm will apply these rules to find target sites:

- It begins with AA
- It contains T in 10th position
- It contains A in 3rd position
- It contains A in 19th position

SEARCH

By transcript name
By gene id
By gene symbol

Search

OPTIONS

Analyze mRNA sequence
Search and analyze gene

OTHER SITES

CIGB
NCBI
EBI
Heber Biotec S.A
Bussines Group
OMS
GAVI
EMBNet

Analizar la secuencia de un gen

El especialista accede a la opción “Search and analyze gene” del menú “Options” y procede a buscar la cadena de nucleótidos de un gen, según un criterio de búsqueda seleccionado (id o símbolo del gen). Luego de obtener los resultados de la búsqueda el especialista tiene la posibilidad de buscar los sitios blancos para la acción del siRNA en toda la cadena de nucleótidos del gen.

biosiRNA-Design

Analysis System of mRNA sequences for siRNA Design

Home Page
Contact Us
Help

SEARCH

By transcript name
 By gene id
 By gene symbol

OPTIONS

[Analyze mRNA sequence](#)
[Search and analyze gene](#)

OTHER SITES

[CIGB](#)
[NCBI](#)
[EBI](#)
[Heber Biotec S.A](#)
[Bussines Group](#)
[OMS](#)
[GAVI](#)
[EMBNet](#)

Sunday, Juny 22th, 2008

Analyze gene

Search gene to analyze it

56

By gene id
 By gene symbol

Analyze gene sequence

Chromosome: 11 | Gene id: 56 | Symbol: ACRV1

TGCTTTAACCATGTGAAATTTATTGAAAAAAAAATCAGGAATATT
 GAGAGAAAGAGTTGGAGCAGGGAAGACAGGACAGAGAAGAATG
 GATAAAGCATGAATAGAAGAAGAAAGATAGGATGTTTGAGCAT
 CCTAATGCTCAGGCAGAGGCAGATGTGGTCAGTTGTTGACTGG
 GGAAGGAACTAAATATAAATGAGCCAAATAGAGTGATGGAGGC

The algorithm will apply these rules to find target sites:

- It begins with AA
- It contains T in 10th position
- It contains A in 3rd position
- It contains A in 19th position
- There is CG in - %
- There isn't GC in 19th position
- The mRNA sequence contains G in 13th position
- There is 3 or more A/T between the bases 15 and 19
- From the selected rules, to be fulfilled al least:
- Exclude SNPs region from the design (No For Pasted Sequence)

Glosario de Términos

Transcrito: Molécula de RNA transcrita a partir de una hebra complementaria de ADN.

Gen: Es la unidad de información hereditaria de los seres vivos y una secuencia lineal de nucleótidos en la molécula de ADN o RNA, que contiene la información necesaria para la síntesis de una macromolécula con función celular específica. El gen es considerado como la unidad de almacenamiento de información y unidad de herencia al transmitir esa información a la descendencia.

Cromosomas: Son las estructuras físicas de la célula eucariota que portan los genes. Estos cromosomas solo son visibles durante la división celular. Desde el punto de vista de su composición los cromosomas están formados de ADN y proteínas.

Exones: Son cada una de las regiones de un gen que contienen la información para producir la proteína codificada en el gen. Cada exón codifica una porción específica de la proteína completa.

Inhibición: Es la condición de interrupción de algunas de las funciones de un determinado gen para resolver alguna acción importante de la función de los genes en la expresión genética del organismo.