



**BIOINFORMÁTICA: INTERFAZ GRÁFICA PARA COMPARACIÓN DE GENOMAS VÍA
WEB**

Memoria del Proyecto Fin de Carrera
De Ingeniería en Informática
realizado por
Oscar Castro García
dirigido por
Jordi Gonzàlez Sabaté
y codirigido por
Mario Huerta
Bellaterra 10 de Junio de 2009



El sotasignat, Jordi Gonzàlez Sabaté

Professor/a de l'Escola Tècnica Superior d'Enginyeria de la UAB,

CERTIFICA:

Que el treball a què correspon aquesta memòria ha estat realitzat sota la seva direcció per en Oscar Castro García

I per tal que consti firma la present.

Signat:

Bellaterra,de.....de 200.....



El sotasignat, Mario Huerta

de l'empresa, Institut de Biotecnologia i de Biomedicina de la UAB

CERTIFICA:

Que el treball a què correspon aquesta memòria ha estat realitzat en l'empresa sota la seva supervisió mitjançant conveni amb la Universitat Autònoma de Barcelona.

Així mateix, l'empresa en té coneixement i dóna el vist-i-plau al contingut que es detalla en aquesta memòria.

Signat:

Bellaterra,de.....de 200.....

ÍNDICE

	Página
1.- Introducción	2
1.1 Motivación del proyecto	2
1.2.- Estado del arte	2
1.3.- Objetivos	6
1.4.- Organización	6
2.- Fundamentos teóricos	8
2.1.- Introducción biológica	8
2.2- Bioinformática y genómica comparativa	11
2.3.- Maximal Unique Matchings	12
3.- Fases del proyecto	15
3.1.- Formación	15
3.2.- Pre-proceso	15
3.3.- Análisis de requisitos	18
3.3.1.- Requisitos funcionales	19
3.3.2.- Requisitos no funcionales	20
3.4.- Modelo de desarrollo	21
3.5.- Implementación	22
3.5.1.- ActionScript 2	22
3.5.2.- Diseño y estructura básica	23
3.5.3.- Estructura de las capas	39
3.5.4.- Estructura de datos	40
3.5.5.- Entorno gráfico	44
4.- Problemas encontrados y soluciones propuestas	49
5.- Conclusiones	55
6.- Referencias	57
7.- Resumen	59
Anexos	
Anexo A	60
Anexo B	61

1.- INTRODUCCIÓN

1.1.- Motivación del proyecto

Este proyecto forma parte de un proyecto mayor de investigación del IBB (Institut de Biotecnologia i de Biomedicina). Dicho proyecto trata de la creación de una herramienta de análisis de genomas vía web que recibe el nombre de FILOMUMS para proporcionar información sobre los procesos evolutivos que han llevado a la aparición de los diferentes organismos, así como otros avances en el ámbito de la biología molecular.

La colaboración en un proyecto de investigación sobre genómica y adquisición de conocimientos necesarios tanto en el ámbito de genética como en el de programación para web, suponen un atractivo reto desde el punto de vista formativo además de la realización personal.

1.2.- Estado del arte

Hace más de 30 años Theodosius Dobzhansky afirmaba: "Nada en biología tiene sentido excepto a la luz de la evolución". En la actualidad, una gran parte de los análisis bioinformáticos por simples que sean, utilizan la comparación de secuencias entre especies a fin de poder determinar la función biológica de una nueva secuencia.

Como contexto histórico de los avances vinculados a la bioinformática podemos puntualizar hitos que han marcado el desarrollo de la genómica [1]:

- 1977: secuenciación BACTERIOFAGO fX174
- 1981: genoma mitocondrial humano
- 1982: GeneBank Version 3
- 1985: FAST (Pearson-Lipman)
- 1986: Proyecto Genoma humano
- 1988: Se funda en Europa el NCBI
- 1990: BLAST (Altschul-Lipman)
- 1995: Secuenciación de la 1ª bacteria (*Haemophilus influenzae*)
- 1996: Secuenciación del 1ª archeobacteria (*Methanococcus jannaschii*)

- 1998: Secuenciación del 1º organismo multicelular (*Caenorabdistis elegans*)
- 2001: Secuenciación del genoma humano

Gracias a los recientes avances en la secuenciación de genomas coordinados con los avances tecnológicos, el coste económico de la decodificación del ADN de un organismo se ha visto reducido drásticamente, dando como resultado un incremento importante de proyectos de secuenciación de genomas.

La información que proporcionan y proporcionarán, permitirá estudiar la evolución de especies emparentadas o cercanas desde un punto de vista genético, gracias a la comparación de genomas.

Para dicha comparación existen diversas herramientas que podemos ver en la Figura 1 donde se muestra la filogenia aproximada de aplicaciones de los últimos 30 años, entre las que cabe destacar las siguientes en referencia a nuestro proyecto:

- MUMmer (1999). Aplicación para la alineación de genomas completos. Se basa en la búsqueda de MUMs, sub-secuencias únicas idénticas de longitud máxima comunes en los dos genomas comparados. Solo puede realizar la comparación de dos genomas simultáneamente. Para ello se construye un único *suffix-tree* o árbol de sufijos que contiene los sufijos de ambos genomas, donde las ramas más largas son los MUMs o sub-secuencias coincidentes entre genomas, concepto en el cual profundizaremos más adelante. [2]
- MUMs On-Line (2002). Evolución del cálculo de los MUMs a partir del algoritmo *suffix-trees*. Introduce una modificación para reducir el espacio de construcción de éstos, que consiste en crear un único *suffix-tree* que contendrá sólo los sufijos de un genoma y que una vez creado podrá usarse para comparar dicho genoma con el resto de genomas. Esto implica que el espacio a ser utilizado no es lineal respecto a la longitud de todos los genomas como venía siendo, sino a la del genoma más pequeño. El algoritmo MUMs On-line es posible gracias a una variación en el algoritmo de construcción de los *suffix-trees* y en el *suffix-tree* generado. Esta variación recibe el nombre de *slide suffix-trees*. [3]

- MALGEN (2003). Es el acrónimo de *Multiple ALignment of GENomes* y es una herramienta para la exploración de relaciones entre secuencias de ADN. Está basada en el cálculo de MUMs y pueden ser calculadas y representadas por más de dos secuencias simultáneamente haciendo uso del algoritmo de MUMs On-Line para encontrar los MUMs. Esta herramienta es accesible vía web. [4]

- M-GCAT (2006). Es el acrónimo de *Multiple Genome Comparison and Alignment Tool* y es una herramienta interactiva para la comparación de genomas. Puede realizar comparaciones de varios genomas simultáneamente. Se basa en *slide suffix-trees* añadiendo un post proceso de *anchoring*, *recursive anchoring*, *filtering* y *clustering*.
 - *Anchoring*: Para poder alinear eficientemente todos los genomas es necesario limitar el espacio de programación dinámica a través de la búsqueda heurística. Este anclaje es un método heurístico que puede ser usado para establecer un marco de secuencia conservada entre todas las secuencias que se comparan.
 - *Recursive anchoring*: Este paso implica la búsqueda de varios *anchors* comunes en todos los genomas. El objetivo es recorrer los genomas con la mayor concordancia posible por medio de búsquedas en las regiones que se encuentran entre las anclas para la creación de nuevas regiones lo suficientemente pequeñas como para ser alineados de manera eficiente.
 - *Filtering*: El objetivo de este paso es la eliminación de ruido.
 - *Clustering*: Organización y agrupamiento de las regiones conservadas en los pasos previos.

Esta aplicación no está orientada a web, sino que es accesible para descarga y ejecución local. Con ella se obtienen los cálculos de MUMs con una mayor eficiencia temporal. [5]

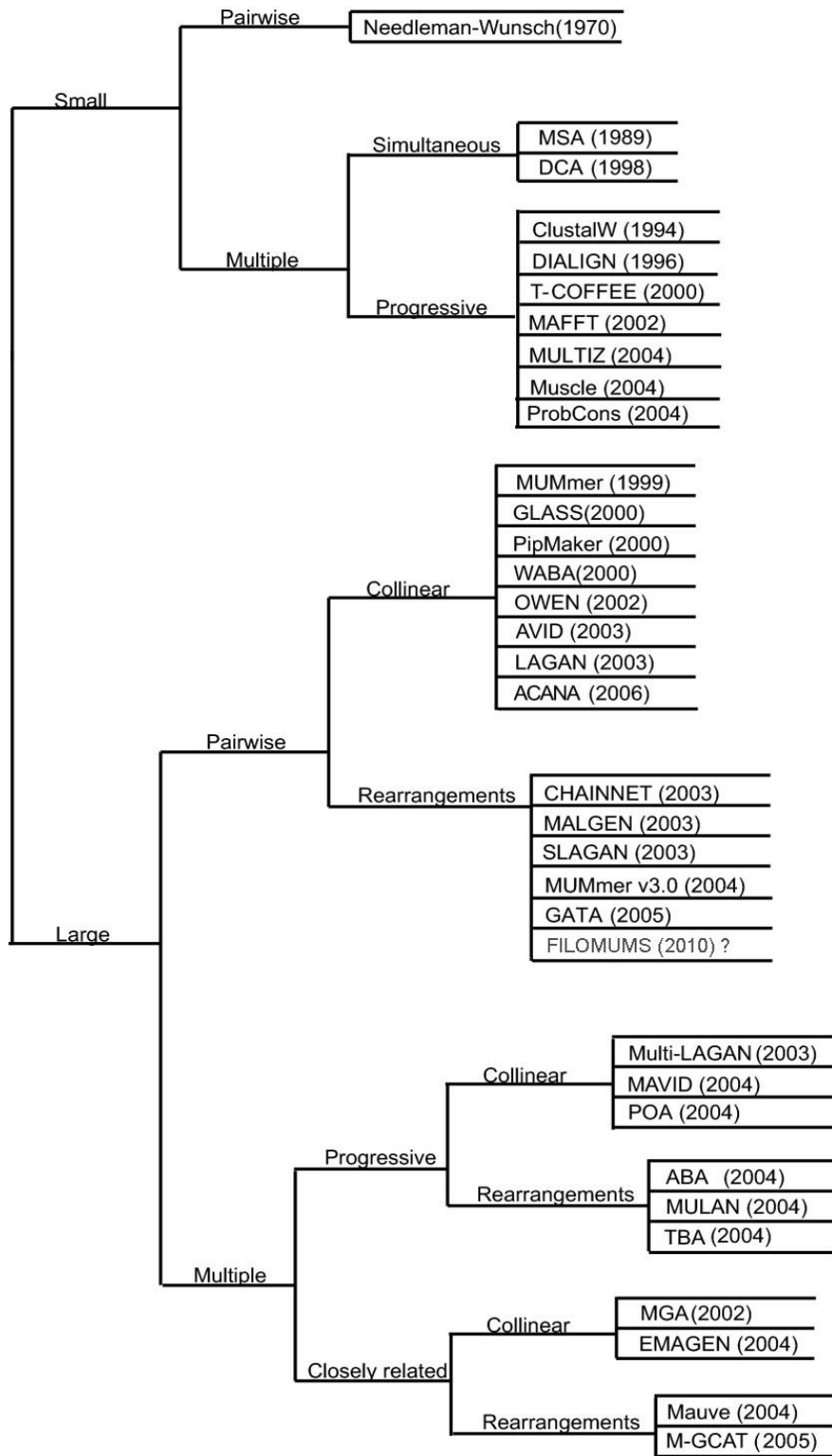


Figura 1. Árbol filogenético de aplicaciones en los últimos 30 años

1.3.- Objetivos

Creación de una interface gráfica para mostrar similitudes entre genomas mediante secuencias únicas que aparecen en éstos y aportan información sobre las coincidencias existentes entre diferentes especies en su ADN.

La aplicación debe recibir por parámetro hasta 8 genomas diferentes, permitiendo representar en primer lugar, una comparación entre dos de ellos mostrando sub-secuencias coincidentes. La herramienta debe realizar Zooms, desplazamientos laterales y desplazamientos individuales de una sola de las cadenas para la comparación explícita entre dos regiones o sub-secuencias de interés.

En segundo lugar podremos representar una comparación encadenada que mostrará las secuencias coincidentes entre cada genoma con su contiguo, con un máximo de hasta 8 comparaciones simultaneas, pudiendo modificar su orden, seleccionar el número de muestras simultaneas a visualizar, hacer Zooms y desplazamientos laterales.

Finalmente, esta aplicación deberá integrarse en la aplicación general (FILOMUMS) del proyecto del IBB en el servidor para poder hacer uso de ésta, de manera interactiva. La aplicación será realizada en lenguaje de programación Flash ActionScript 2¹ y deberá ser lanzada desde una aplicación principal en lenguaje de programación Java.

1.4.- Organización

Para la organización del proyecto se han establecido hitos u objetivos parciales sujetos a adaptaciones o modificaciones durante su elaboración con el fin de mejorar su funcionalidad, previa consulta o sugerencia del codirector de proyecto Mario Huerta.

Cada paso va seguido de un periodo de testeo para verificar que la funcionalidad obtenida es la deseada. Se ha estipulado una duración para cada hito intentando respetar, en la medida de lo posible, los plazos inicialmente establecidos.

¹ Lenguaje de programación orientado a objetos para Macromedia Flash. Está basado en el estándar ECMA-262 propuesto por los creadores de Javascript.

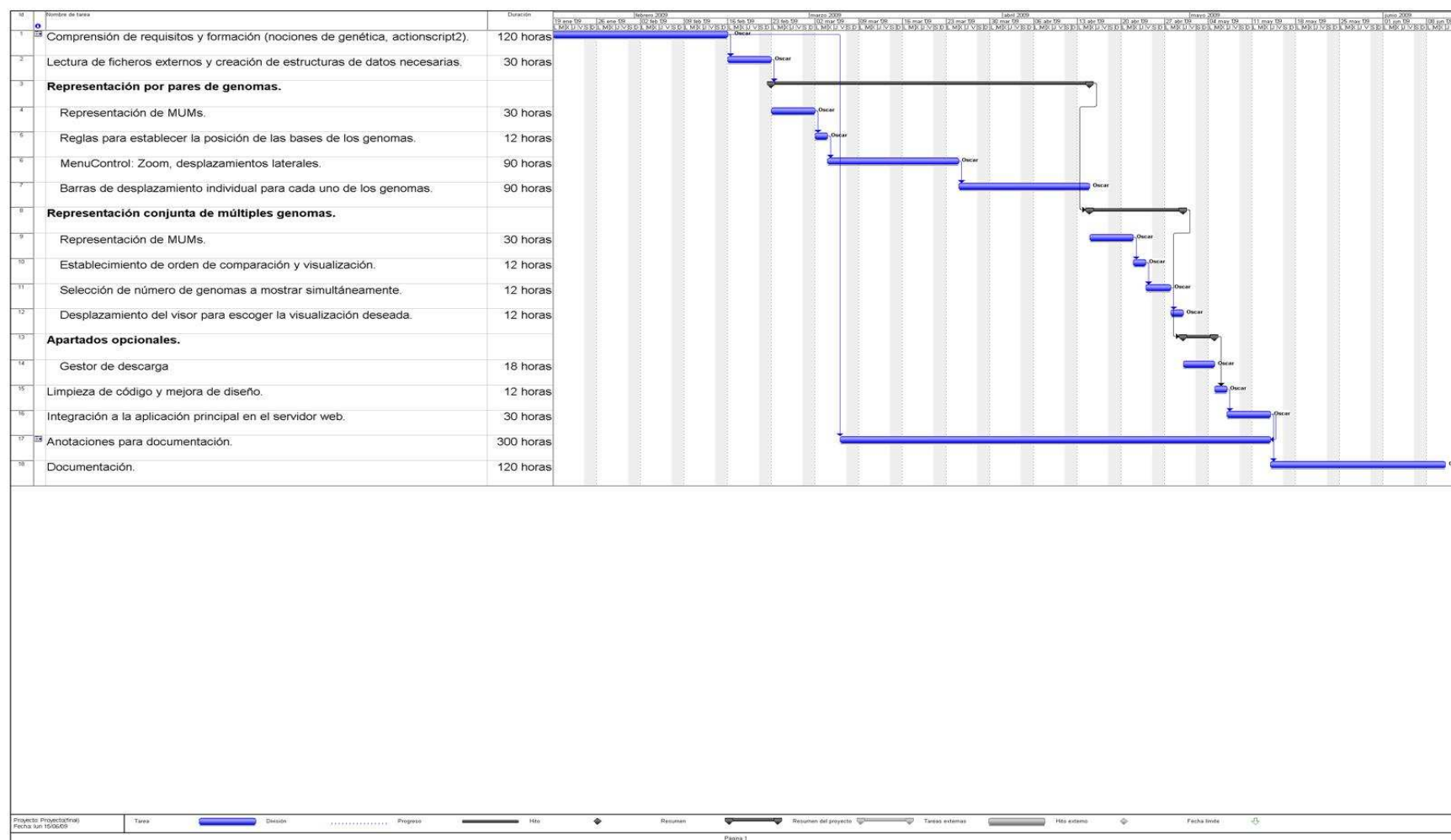


Figura 2. Diagrama final de Gantt del proyecto

2.- FUNDAMENTOS TEÓRICOS

2.1.- Introducción biológica

La Genética es el estudio de la naturaleza, organización, función, expresión, transmisión y evolución de la información genética codificada de los organismos.

Se puede dividir en las siguientes áreas:

- Genética clásica (transmisión y localización de los genes en los cromosomas²)
- Genética molecular (estructura y el control de la expresión del material genético)
- Genética evolutiva (referente a los procesos evolutivos de poblaciones)
- Genómica (correspondiente al análisis e interpretación de los genomas)

Este proyecto está orientado a la **genómica comparativa** que estudia las relaciones entre genomas de diferentes especies o razas. El objeto que se busca es el de beneficiarse de la información proporcionada por las firmas de la selección natural para entender la función y los procesos evolutivos que actúan sobre los genomas.

Las secuencias de ADN que constituyen la unidad básica y funcional de la herencia se denominan genes³, los cuales contienen la información genética de un genoma, y al conjunto de toda la información correspondiente a un organismo lo denominamos su genotipo.

Cada célula de un organismo contiene el genoma completo, la diferencia que encontramos entre dos células distintas es que algunos genes están activos y otros no.

Podemos distinguir 3 dominios para clasificar las células según un modelo evolutivo de clasificación, propuesto por Carl Woese, basado en las diferencias encontradas en las secuencias de nucleótidos en los ribosomas⁴ y RNAs⁵ de transferencia de la célula, la estructura de los lípidos de la membrana, y la sensibilidad a los antibióticos.

² Filamento condensado de ácido desoxirribonucleico, visible en el núcleo de las células durante la mitosis.

³ Gen, proviene de la palabra griega γένος y significa "raza, generación".

⁴ Orgánulo en el que tienen lugar las últimas etapas de la síntesis de proteínas.

⁵ Ácido ribonucleico.

Tipo	Reino	Descripción
Bacteria (Eubacteria)	mycoplasmas cyanobacteria bacterias Gram-positivas bacterias Gram-negativas	<ul style="list-style-type: none"> Son procariotas, contiene el ADN en el citoplasma. Son organismos microscópicos y en su mayor parte unicelulares. Son sensibles a los antibióticos antibacterianos tradicionales.
Archaea (Archaeobacteria)	metanógeno halófilos extremos termoacidófilos	<ul style="list-style-type: none"> Son procariotas, contiene el ADN en el citoplasma. Son organismos unicelulares. No son sensibles a algunos antibióticos que afectan a las Bacterias. Viven a menudo en ambientes extremos.
Eucariota (Eukaryota)	Hongos Protistas Plantas Animales	<ul style="list-style-type: none"> Tienen un núcleo celular diferenciable que contiene el ADN. No son sensibles a los antibióticos antibacterianos tradicionales.

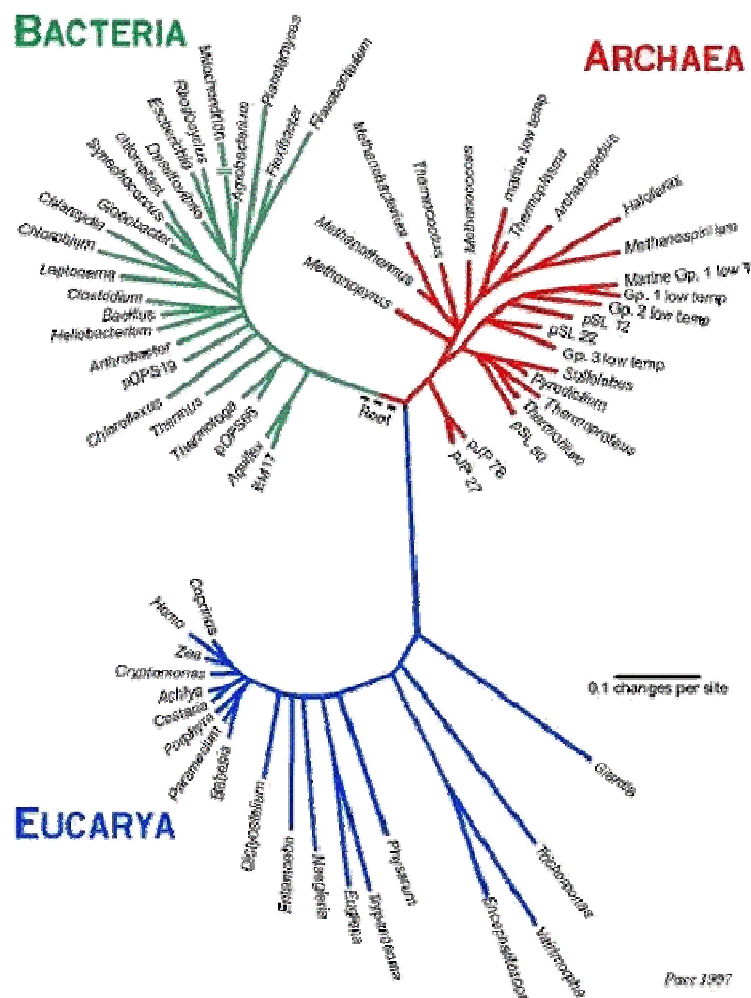


Figura 3: Árbol de los 3 Dominios planteados por Carl Woese. Esquema de Norman Pace (1977)

Podemos resumir que el dominio eucariota lo conforman plantas, hongos y animales, donde situamos a los mamíferos y por lo tanto al ser humano. Mientras que los dominios de bacteria y archaea los conforman respectivamente bacterias y arcobacterias.

Desde el punto de vista químico, el ADN es un polímero⁶ de nucleótidos y cada nucleótido, a su vez, está formado por un azúcar (la desoxirribosa), una base nitrogenada (que puede ser adenina→*A*, timina→*T*, citosina→*C* o guanina→*G*) y un grupo fosfato que actúa como enganche de cada uno con el siguiente.

Lo que distingue nucleótido de otro es, entonces, la base nitrogenada, y por ello la secuencia del ADN se especifica nombrando sólo la secuencia de sus bases. La disposición secuencial de estas cuatro bases a lo largo de la cadena es la que codifica la información genética: por ejemplo, una secuencia de ADN puede ser *ATGCTAGATCGC...*

En los organismos vivos, el ADN se presenta como una doble cadena de nucleótidos, en la que las dos hebras están unidas entre sí por unas conexiones denominadas puentes de hidrógeno.

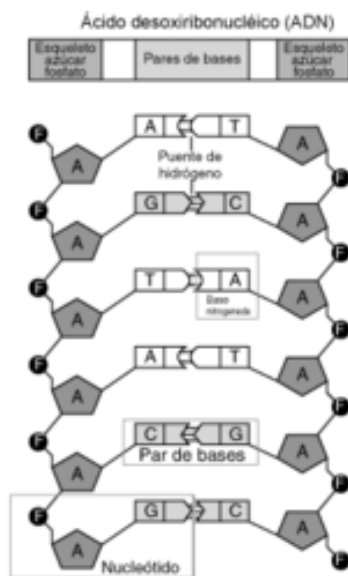


Figura 4: Nucleótido

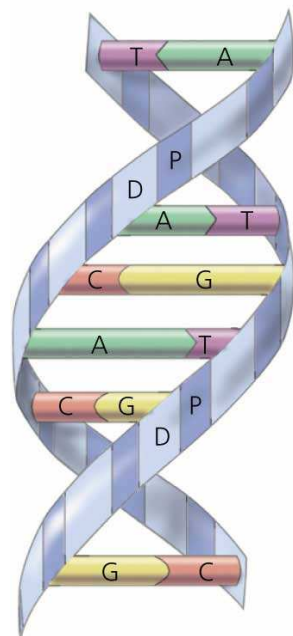


Figura 5: ADN

⁶ Un polímero es un compuesto formado por muchas unidades simples conectadas entre sí.

Simplificando podemos decir que el ADN es un almacén cuyo contenido es la información de un organismo, este conjunto de información se denomina genoma, el cual se compone de una cadena de genes, y que un gen es una sub-secuencia lineal de nucleótidos de la secuencia completa que compone un genoma.

2.2.- Bioinformática y genómica comparativa

La bioinformática consiste en la gestión y el análisis de datos biológicos haciendo uso de la computación y tecnologías a nuestra disposición, abarcando los ámbitos de la matemática aplicada, la estadística, la inteligencia artificial, la química, la bioquímica, la informática o las ciencias de la computación. Podemos decir que es la ciencia dedicada al estudio de los fenómenos biológicos de la microbiología molecular desde un ámbito computacional, cuyo objeto es conseguir métodos robustos que faciliten la comprensión, simulación y predicción de comportamientos biológicos observados en los seres vivos, mediante la utilización de recursos computacionales para buscar soluciones, para el análisis de datos o la simulación de sistemas y mecanismos de índole biológica (habitualmente a nivel molecular).

La bioinformática, unida al avance de las técnicas de manipulación⁹ del ADN, ha permitido determinar la secuencia completa del genoma de un organismo. De hecho parte de la importancia que tiene actualmente se debe al nacimiento del Proyecto Genoma Humano (HGP).

La secuencia que obtenemos de cada genoma aporta gran cantidad de información biológica de interés ya que permite predecir y catalogar el número total de genes y su estructura, así como definir la organización básica del organismo y las diferentes clases funcionales de proteínas (energía, comunicación, información, etc..).

Un ejemplo de la utilidad de los ámbitos en los que juega un papel fundamental, es el descubrimiento de nuevos fármacos. Además, combinada con las nuevas técnicas de biotecnología, la bioinformática permite identificar genes y proteínas causantes de enfermedades o de mecanismos de resistencia a antibióticos, que de otra forma serían

complicados de detectar, permitiendo intervenir en los procesos con fármacos más específicos o, incluso, introduciendo nuevos paradigmas terapéuticos, como la terapia génica.

En relación con este proyecto, la comparación de secuencias entre diferentes especies resulta de gran importancia para comprender la evolución de éstas así como para entender la organización y funcionalidad de ciertos genes comunes entre dos especies distintas.

El estudio de dicha comparación al que nos referimos, se denomina genómica comparativa, que busca solución a ciertas cuestiones como “¿cuál es el número mínimo de genes que son indispensables para crear un organismo?” o “¿qué familias de proteínas son universales y cuales son específicas de un organismo?”

2.3.- Maximal Unique Matchings

Un método para realizar la tarea de comparación de dos secuencias genómicas distintas es buscar la sub-secuencia única más larga coincidente en ambos genomas, que debe aparecer una sola vez a lo largo la secuencia completa que conforma cada uno de los genomas. Estas sub-secuencias reciben el nombre de *Maximal Unique Matchings* (MUM) y el conjunto de MUMs obtenidos en una comparación determina el *skeleton*⁷ sobre el cual se puede hacer una comparación global.

La búsqueda clásica de MUMs se realiza mediante la construcción de *suffix-tress* (árboles de sufijos que podemos ver en la Figura 6) donde se representan todos los sufijos de las secuencias en un árbol en el que comparten sus prefijos comunes. El proceso de construcción del árbol tiene un alto coste en tiempo de proceso, por ello, la variante *slide suffix-trees* permite buscar los MUMs de varias secuencias, construyendo el árbol para una sola secuencia y recorriéndolo después para comparar con el resto.

En este proyecto, en particular, se ha creado un entorno grafico que permite interpretar de forma visual las coincidencias entre dos o más genomas mostrando los MUMs para todos los genomas secuenciados hasta el momento. Éstos son calculados para cada par de genomas secuenciados para bacteria, archea y eukariota, para luego ser analizados desde una interface

⁷ Estructura básica de comparación de genomas que tiene significado biológico, que nos permite ver los cambios evolutivos de una especie a otra.

Java vía web a partir de un árbol que muestra la relación de dependencia entre los genomas en base a estos MUMs.

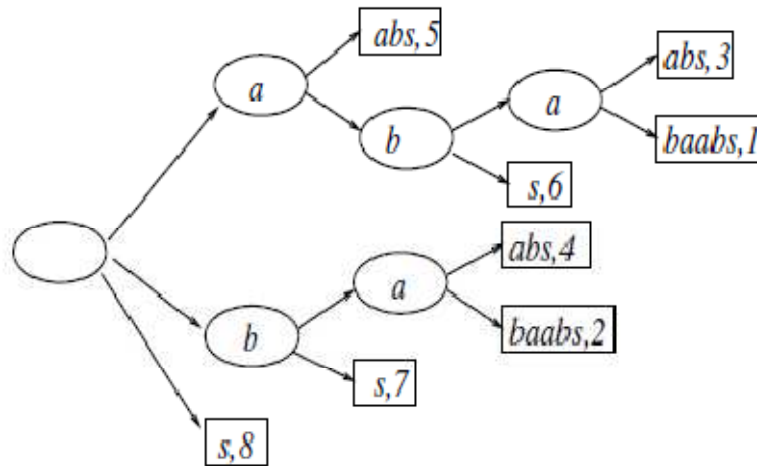


Figura 6. Ejemplo de árbol Suffix-tree

Desde este árbol pueden seleccionarse los genomas cuyo *skeleton* en base a los MUMs desea estudiarse en detalle utilizando la interfaz gráfica creada en el proyecto.

En el proceso evolutivo de las especies, en ocasiones la cadena de un gen muta invirtiéndose completamente *ATGCTAGA*. \Rightarrow *AGATCGTA* y aquí aparece el concepto de MUMs inversos, que significa que existe una coincidencia con otra cadena, entendiendo una de las secuencias de ha invertido. Como aclaración, debo añadir que los MUMs inversos serán representados en rojo en la aplicación mientras que los directos, lo serán en verde tal y como vemos en la Figura 7.

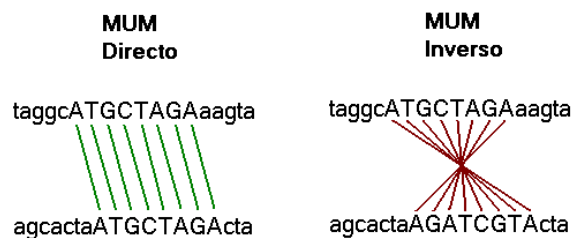


Figura 7. MUMs directos e inversos

A continuación podemos ver un ejemplo de un MUM directo para nuestra aplicación:

Genoma1
 ...agctcgat**GGGCTTTAGACTCTCGATA**ggcgagagc...

Genoma2
 ...agacctaaagtctatcc**GGGCTTTAGACTCTCGATA**aagtt...

En el genoma1, el MUM comienza en la novena base y en el segundo, en la decimoctava y la longitud de éste es de 19 bases.

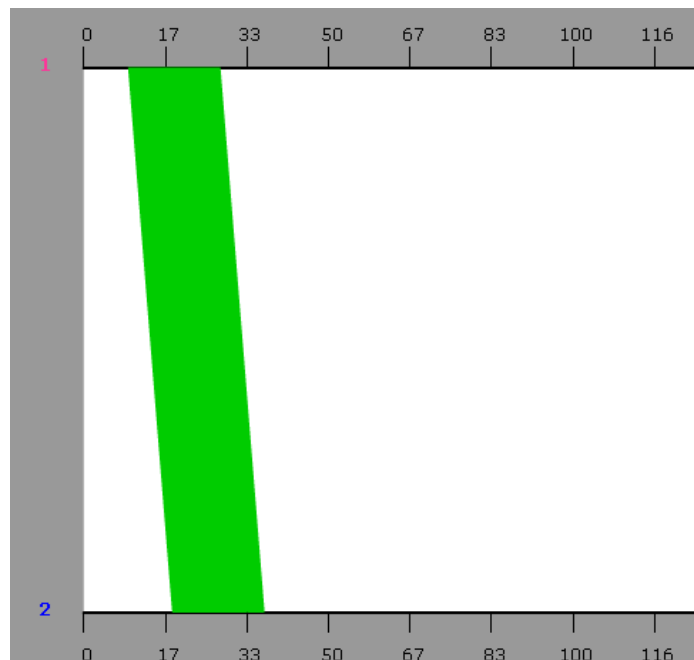


Figura 8: Representación del ejemplo en la aplicación

Habría que especificar que este ejemplo es poco real ya que por lo general los genomas con los que trabajamos contienen secuencias del orden de millones de bases y la longitud mínima que se ha establecido para la longitud es de 20 bases.

3.- FASES DEL PROYECTO

3.1.- Formación

La principal base formativa para el desarrollo de este proyecto es el aprendizaje del lenguaje de programación ActionScript2 propio de Flash, ya que a pesar de que se trata de un entorno pensado para la creación de gráficos, animaciones y multimedia sin amplios conocimientos de programación, no es este el caso y se hace un mayor uso del lenguaje que de su interface gráfica.

Se requiere la lectura de archivos, la generación de estructuras de datos y la manipulación de éstos mediante un cálculo constante bajo la demanda del usuario, respetando con la mayor exactitud y precisión posibles las posiciones, las dimensiones y la manipulación de la representación.

También ha sido necesaria una base de formación, esencial en el área a la que está dirigido el proyecto (la genómica), para poder entender lo que se esperaba obtener y la motivación perseguida.

En último lugar, también, ha sido necesario recuperar conocimientos adquiridos a lo largo de la carrera como la programación en Java o PHP para objetivos auxiliares como son la integración de la herramienta a la aplicación FILOMUMS.

3.2.- Pre-proceso

El método de alineamiento seguido es el de MUMs, el cual consiste en la creación de un árbol de sufijos sobre un genoma A e insertar los sufijos del genoma B para detectar las secuencias repetidas. De las secuencias encontradas se descartan aquellas que no alcanzan una longitud mínima así como aquellas cuyo sufijo pertenezca a otra secuencia. Esto corresponde al algoritmo *slide Suffix-trees*.

Para la realización del cálculo de los MUMs se utiliza el programa Mumix, que consiste en una colaboración de 4 subprogramas: malloc.cc (para la definición de funciones de

memoria), st4lm.cc (algoritmo por el que obtendremos los MUMs), st5.cc (para la construcción del árbol de sufijos) y mumol.cc (encuentra todos los MUMs entre 2 aminoácidos). Mumix nos da los MUMs entre 2 genomas, pero para obtener el sumatorio del tamaño de todas las coincidencias entre cada par de genomas, necesitamos un segundo programa que llamara al primero tantas veces como sea necesario (lo cual dependerá del número de genomas a ser comparados). Esta tarea la lleva a cabo Lanza_mums.cc, que recibe por parámetro el directorio donde se encuentran los genomas y genera un fichero de salida llamado factors.txt. Este fichero guarda la suma de todas las longitudes de los MUMs entre cada par de genomas, a partir del cual queremos generar un grafo para su representación gráfica.

La representación gráfica que queremos, es la de un grafo que nos muestre la relación que guardan los genomas de un mismo dominio entre sí y, para ello, solo deseamos unir los genomas que guarden una mayor correlación, sin que ninguno de ellos quede fuera de la estructura.

La solución para nuestras necesidades es un árbol conexo máximo o *maximum spanning tree*, que es un grafo conexo sin ciclos cuyos vértices serán los genomas y sus aristas serán la relación que guardan entre sí. Para su construcción usamos el algoritmo de PRIM mediante Prim.c.

El Grafo interactivo: tiene por objetivo mostrar los genomas de forma visual y ofrecer diferentes herramientas para estudiarlos. Está programado en Java y, debido a la necesidad de trabajar con funciones para grafos, se utiliza la librería JUNG (*Java Universal Network / Graph Framework*). El entorno de programación es Eclipse⁸.

La interface gráfica consta de 3 *applets*, uno para cada organismo (Archaea, Bacteria, Eucariota). En la Figura 9 podemos ver el *applet* para bacteria y a su derecha se sitúa la interface de usuario, que es un contenedor *JPanel*, el cual contiene las siguientes opciones:

⁸ Plataforma de desarrollo open source i multiplataforma basado en Java.

- Zoom-In:
 - Mspath. Selecciona y muestra el camino entre 2 genomas seleccionados, marcando los genomas intermedios y poder observar rasgos evolutivos.
 - MUMs analysis. Lanza la aplicación gráfica para la comparación de los genomas seleccionados. (Aquí se debe lanzar la aplicación de este proyecto)
- Search: Es un *JTextField* que permite al usuario buscar un genoma concreto.
- Genoma Info:
 - Correlations
 - Genoma. Opción de enlace al NCBI, pasando por referencia el genoma seleccionado.
- Custom Ops:
 - Zoom.
 - Mode. Checkbox Transforming / Picking que le da la opción al usuario de interactuar con el gráfico, ya sea para trasladarlo o para seleccionar genomas.
- Filter Correlations: Filtro del número de aristas que veremos en el grafo. De esta manera no se visualizarán las aristas de peso menor al indicado.
- Show Ops: esta sección será para un conjunto de opciones que nos permitirán cambiar los nodos y aristas que visualizamos y como las visualizamos.
 - Correlations. Visualiza o no las aristas entre cada par de genomas del fichero factors.
 - Gen. Names. Muestra por pantalla o no el nombre de los genomas.
 - Show-mstree. Muestra las aristas que pertenecen al *mstree*.
 - Paint. Representación en colores de las aristas según su peso y de los clusters a los que pertenecen los genomas.

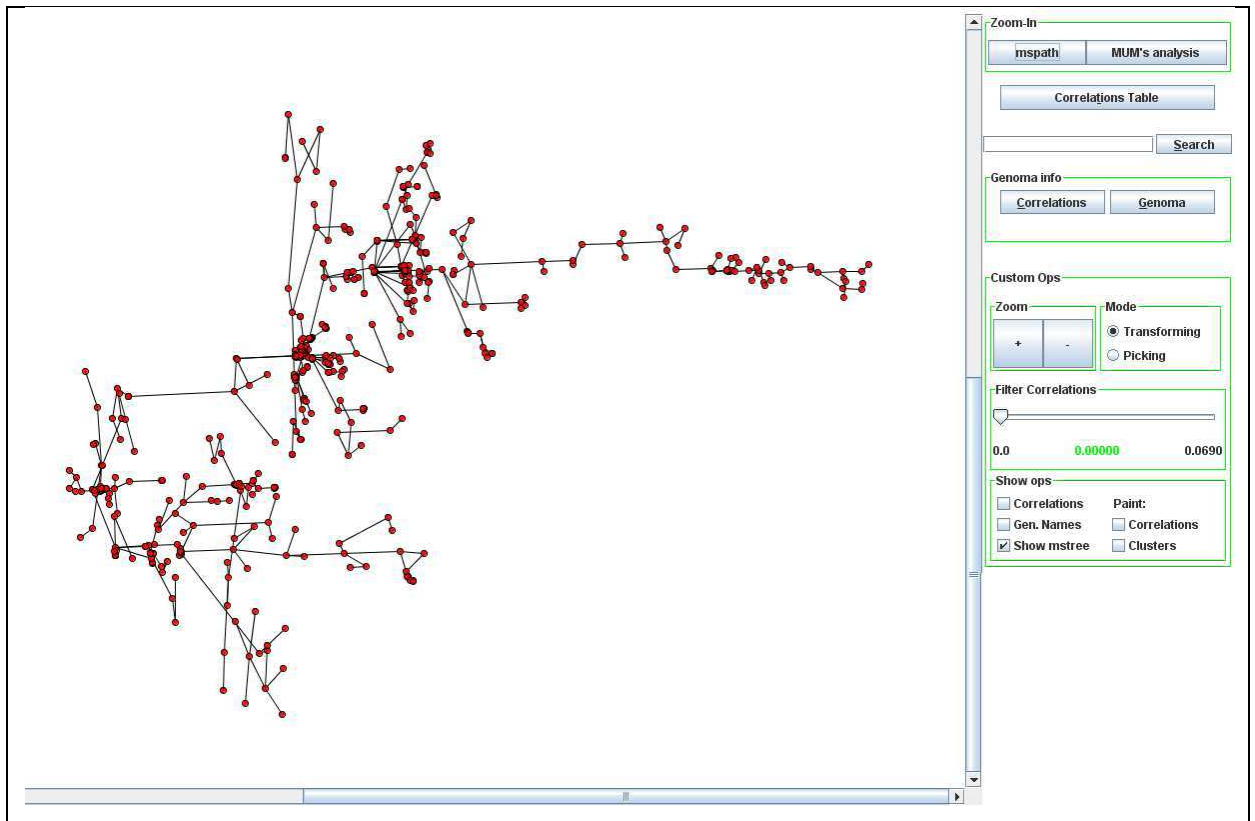


Figura 9. Applet Bacteria desde el que se lanza la aplicación

De todo este pre-proceso he tenido que modificar la aplicación `Lanza_mums.cc` para que además de generar el `factors.txt` también generase los ficheros de MUMs con el formato y nombre necesarios para ser cargados en mi aplicación. Para ello he usado el compilador GNU gcc o g++. Finalmente he adaptado la interface del grafo para añadir la visualización en detalle de las comparaciones, es decir, crear la llamada a la aplicación de comparación de genomas, que es la que se ha realizado en este proyecto (botón *MUMs analysis* de la figura 9) en el *applet* Java.

3.3.- Análisis de requisitos

Los requisitos necesarios para la creación de una herramienta gráfica vía web de las características concretas estipuladas previamente, condicionaran constantemente la toma de decisiones así como su desarrollo y su resultado.

3.3.1.- Requisitos funcionales

En primer lugar especificaremos que la herramienta constaría de tres opciones contempladas en un menú integrado:

- ***Pair wise comparison*** o comparación simple (Figura 22), aquí debían ser comparados dos genomas, siendo posible seleccionar que par debían ser visionados del total de genomas cargados, que pueden ser hasta 8, desde la aplicación de la figura 9.

Dentro de esta opción habría un control para poder desplazar lateralmente cada genoma de manera individual, con objeto de poder focalizar dos regiones concretas de interés. Este control se realizaría mediante dos barras de desplazamiento (Figura 23), una por cada genoma donde el grosor y la posición de la barra deslizante determinarían el área que se estaba mostrando del genoma.

- ***Multiple comparison*** o comparación múltiple (Figuras 25 y 26), aquí se mostraría una comparación de hasta ocho genomas de cada uno con su siguiente, con la opción de poder modificar el orden en que estos serían mostrados y comparados en el visor de la aplicación.

Dentro de esta opción habría dos controles: un selector a modo de *combobox* donde se podría seleccionar el número de genomas simultáneos a ser mostrados en el visor de la aplicación y dos botones para permitir subir o bajar el dicho visor, en función de las muestras de genomas a comparar. Esta última opción solo es válida en el caso de que el número de genomas simultáneos fuera menor que el número total de genomas cargados en la aplicación, lo que significa, que si hay por ejemplo, 5 genomas cargados en la aplicación y tenemos seleccionados 5 en el *combobox* de número de genomas simultáneos, ya los estaríamos visualizando todos, por lo que tendría sentido desplazar el visor.

- **Opción de descarga** (Figura 28), esta opción debía permitir descargar los archivos de datos de los MUMs con los que trabaja la aplicación renombrándolos en la descarga con sus nombres reales en vez de con su identificador, que es como residen en el

servidor. Esto serviría para que el usuario que estaba realizando una comparación, si deseaba hacer un análisis más profundo, podría descargarse éstos. Este requisito era opcional para este proyecto.

Tanto para la comparación de pares como para la múltiple debían existir dos menús de control (Figura 27):

- Un control que permitiera seleccionar los MUMs que debían ser visionados, directos, inversos o ambos simultáneamente. Esto se ha realizado mediante dos botones a modo de *checkbox* en los que se selecciona cuales de estos se desean visionar.
- Un menú de controles para la manipulación de la visualización de la gráfica que constaría de lo siguiente:
 - Zoom (ampliación), pudiendo hacer un zoom por defecto en un punto o seleccionando el área sobre la que se desease realizar el zoom.
 - Zoom (reducción), reduciendo a la mitad la escala y establecer como centro de la nueva visualización el punto de la gráfica seleccionado.
 - Escala original, retornando a las dimensiones originales 1:1.
 - Desplazamiento lateral de la gráfica.

Los zooms a realizar serán horizontales ya que por los datos que se desean mostrar, la distancia vertical no aporta ninguna información.

Dentro de este menú de control también se ha agregado un botón que abriría una ventana de ayuda (ver Anexo B).

3.3.2.- Requisitos no funcionales

Para la solución del problema se optó por Flash, por los siguientes motivos:

- La solución debía integrarse a un navegador y para ello el archivo generado por Macromedia Flash es una película reproducible en cualquier navegador mediante un

plugin muy extendido en todos los navegadores como es “Macromedia Flash Player”, el cual resulta fácilmente instalable, si no está ya integrado en estos.

- La herramienta debía ser un instrumento rápido e intuitivo para la consulta de similitudes entre genomas, esto implica rapidez, lo que podemos traducir como eficiencia temporal (optimización de los cálculos) y espacial (reducido tamaño de la película que debe descargarse completamente antes de ser cargada por el navegador). Otra cuestión es la carga de los ficheros externos a la aplicación que deberán ser cargados desde otra ubicación del servidor web y variará su volumen de datos en función de los genomas que vayan a consultar.
- Por último es necesario aclarar que ya existe una aplicación gráfica destinada a mostrar resultados de experimentos sobre genes y se pretendía respetar al menos visualmente su diseño, además, de reutilizar funciones ya creadas para ActionScript2 que aunque no eran funcionales en este proyecto, si podían establecerse como punto de partida así como, también, su estructura.

Además de ser funcional, la aplicación debe ser ampliable (Reusabilidad) ya que en un futuro inmediato se le adaptaran nuevas funcionalidades por lo que se ha optado por una estructura de fácil comprensión para poder ser retomado con el menor esfuerzo posible. Sin embargo Flash y concretamente ActionScript2 no favorece precisamente la reutilización de código ya que está concebido para crear contenidos multimedia como animaciones, juegos o reproducciones de audio y video específicos.

3.4.- Modelo de desarrollo

Aunque se conocían los requisitos inicialmente, éstos estaban sujetos a cambios durante el proceso de implementación con el fin de mejorar su funcionalidad en cada una de las etapas del proceso. Por ello se descartó el modelo de desarrollo en cascada, que exige un diseño completo inicial, ya que cualquier cambio o error de diseño detectado en la etapa de prueba conduce necesariamente al rediseño y nueva programación del código afectado, aumentando los costes del desarrollo.

Finalmente, el modelo con el que se ha realizado este proyecto es el de desarrollo por etapas, que es similar al modelo de prototipos, en el que el cliente (en nuestro caso codirector

de proyecto) ve y comprueba la nueva funcionalidad desarrollada en cada etapa siendo ésta susceptible a cambios o mejoras. Pero a diferencia del modelo de prototipos, en éste, las especificaciones no son conocidas al detalle al inicio del proyecto, por lo cual se desarrollan simultáneamente con las diferentes versiones del código.

Para cada etapa del diseño se ha pasado por el siguiente ciclo:

- Especificación conceptual.
- Análisis de requisitos. En caso de haber algún conflicto se debe aclarar con el cliente.
- Diseño inicial.
- Test. Resolución de problemas si los hay y propuesta de soluciones.
- Diseño y codificación detallados y liberación

Al mismo tiempo, durante la realización de una etapa, se hallaban errores que no habían sido apreciados o aspectos susceptibles de mejora de etapas anteriores.

3.5.- Implementación

3.5.1.- ActionScript 2

En este apartado haremos un breve análisis de ActionScript 2, que como veremos no sigue un modelo de programación estructurada clásica ni tampoco es puramente orientada a objetos. El objeto de este análisis es el de ayudar a comprender como se ha estructurado la aplicación posteriormente.

Hay tres tipos de objetos en flash:

- *MovieClips* (películas)
- *Buttons* (botones)
- *Graphics* (imágenes)

El motivo por el que cualquier objeto debe pertenecer a uno de estos tipos es por la orientación de este lenguaje hacia los contenidos multimedia o aplicaciones con alto contenido gráfico y bajas en complejidad lógica, para su publicación en internet.

Un archivo de Flash tiene un marco principal llamado escenario y un objeto principal de tipo *MovieClip* de nivel 0 de donde deben partir el resto de objetos que conformarán la película final e interactuarán con ésta. Cada objeto deriva del principal, situándose en un nivel determinado de la jerarquía de objetos.

Además de los niveles, Flash trabaja con capas, que establecen el orden de visualización, superponiéndose unas a otras, en caso de que existan varias capas con contenido visual (de manera similar a como trabajan aplicaciones de diseño gráfico como podrían ser Photoshop, Corel Draw, etc.).

El código de ActionScript debe situarse también en estas capas y los objetos de tipo *MovieClip* pueden disponer a su vez de sus propias capas, así como los de tipo *Button* tienen un área propia para la programación de eventos (como pueden ser pulsar, soltar, pasar el puntero por encima).

En la creación de la aplicación, no se usa una de las principales herramientas de flash que es la línea de tiempo, útil cuando se pretenden hacer animaciones o establecer un orden de aparición de los clips de película, exceptuando un clip de espera, que se carga únicamente durante la lectura de los ficheros que contienen la información que deseamos analizar y éstos pueden ser de muy diverso volumen.

Cabría especificar, que se ha optado por la carga de todos los ficheros en el inicio de la aplicación, ya que ésta carga por defecto la opción “*Pair Wise Comparison*” o Comparación Simple (con los genomas 1 y 2 seleccionados) y para ello debe haber generado previamente la estructura de datos con la que trabajará. Flash gestiona el contenido de los ficheros como una variable de tipo *string* y esto implica un retraso temporal, por lo que no se carga la gráfica hasta que no se ha generado dicha estructura.

3.5.2.- Diseño y estructura básica

En este apartado vamos a definir como se estructuran todos los objetos que intervienen en la aplicación, especificando: como están instanciadas, de que tipo son, además de una descripción de cada uno de ellos, comenzando por el inicio de la aplicación y siguiendo por

las tres opciones de menú que ésta ofrece. En los diagramas representados, cada objeto se encasilla en una elipse y los estados u opciones en rectángulos.

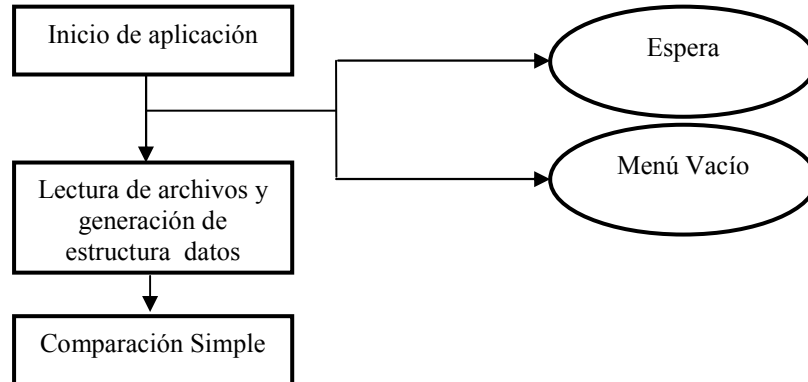


Figura 10: Diagrama de inicio de la aplicación

Espera

Nombre Objeto	Wait
Tipo Objeto	MovieClip
Nombre Instancia/s	Waiting
Accesibilidad	Inicialización de aplicación
Descripción	Animación simple con mensaje de espera.

Menú Vacío

Nombre Objeto	MenuVacio
Tipo Objeto	MovieClip
Nombre Instancia/s	MenuVacio
Accesibilidad	Inicialización de aplicación
Descripción	Panel vacio en la posición en que se carga el Panel de Gráficas Simple y múltiple, para evitar su carga sin inicialización de datos.

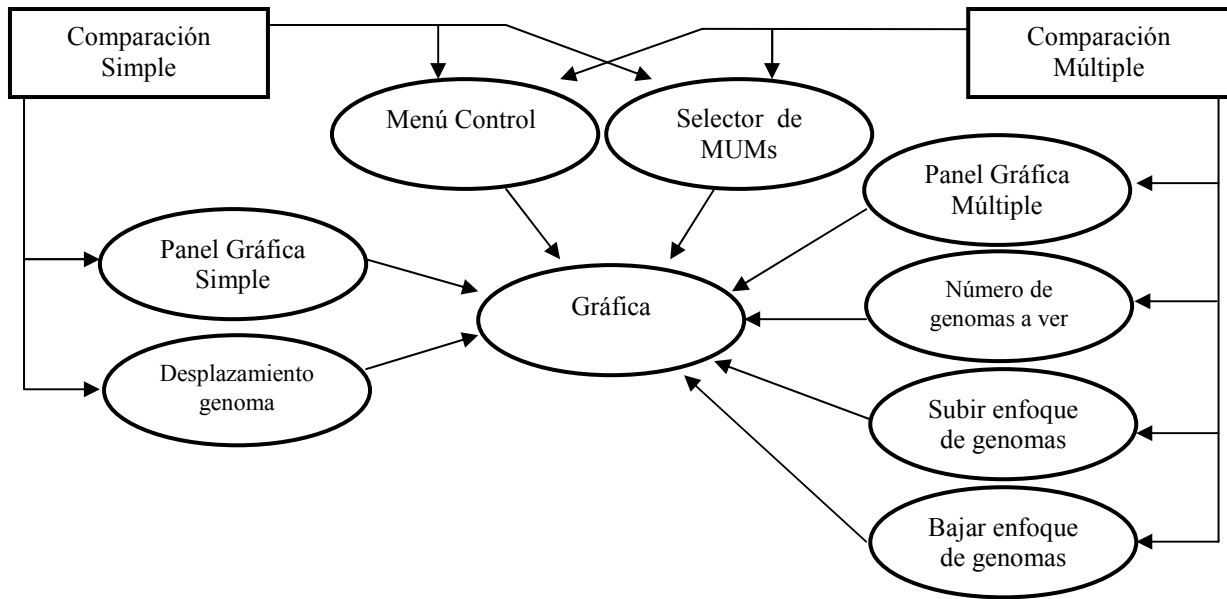


Figura 11. Diagrama completo A. Opciones de comparación

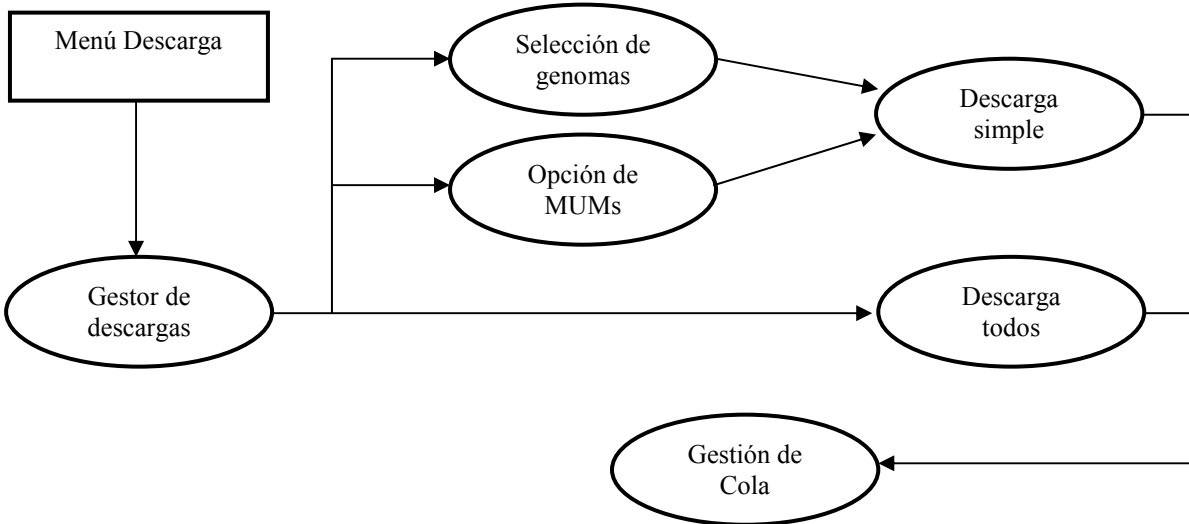


Figura 12. Diagrama completo B. Opción descarga

Selector de MUMs

Nombre Objeto	CheckBoxes
Tipo Objeto	MovieClip
Nombre Instancia/s	CheckBoxes
Accesibilidad	Comparación Simple / Comparación Múltiple
Descripción	Permite seleccionar los MUMs a ser mostrados en la gráfica, ya sean directos, inversos o ambos.

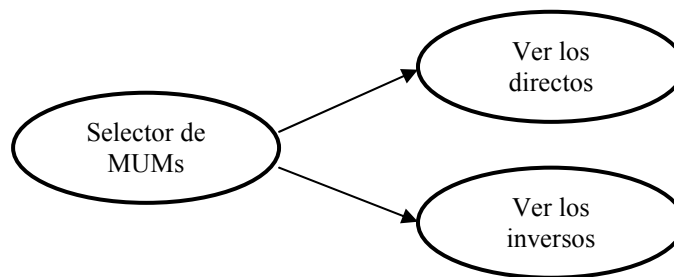


Figura 13. Diagrama del panel de selección de MUMs

Ver los directos

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	BotonDir
Accesibilidad	Comparación Simple / Comparación Múltiple
Descripción	Botón que funciona a modo de interruptor, seleccionando o deseleccionando la visualización de los MUMs directos y redibuja la gráfica.

Ver los inversos

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	BotonInv
Accesibilidad	Comparación Simple / Comparación Múltiple
Descripción	Botón que funciona a modo de interruptor, seleccionando o deseleccionando la visualización de los MUMs inversos y redibuja la gráfica.

Menú Control

Nombre Objeto	MenuControl
Tipo Objeto	MovieClip
Nombre Instancia/s	MenuControl
Accesibilidad	Comparación Simple / Comparación Múltiple
Descripción	Permite manipular la gráfica, desde ampliaciones o reducciones hasta desplazamientos laterales. También contiene la opción de ayuda.

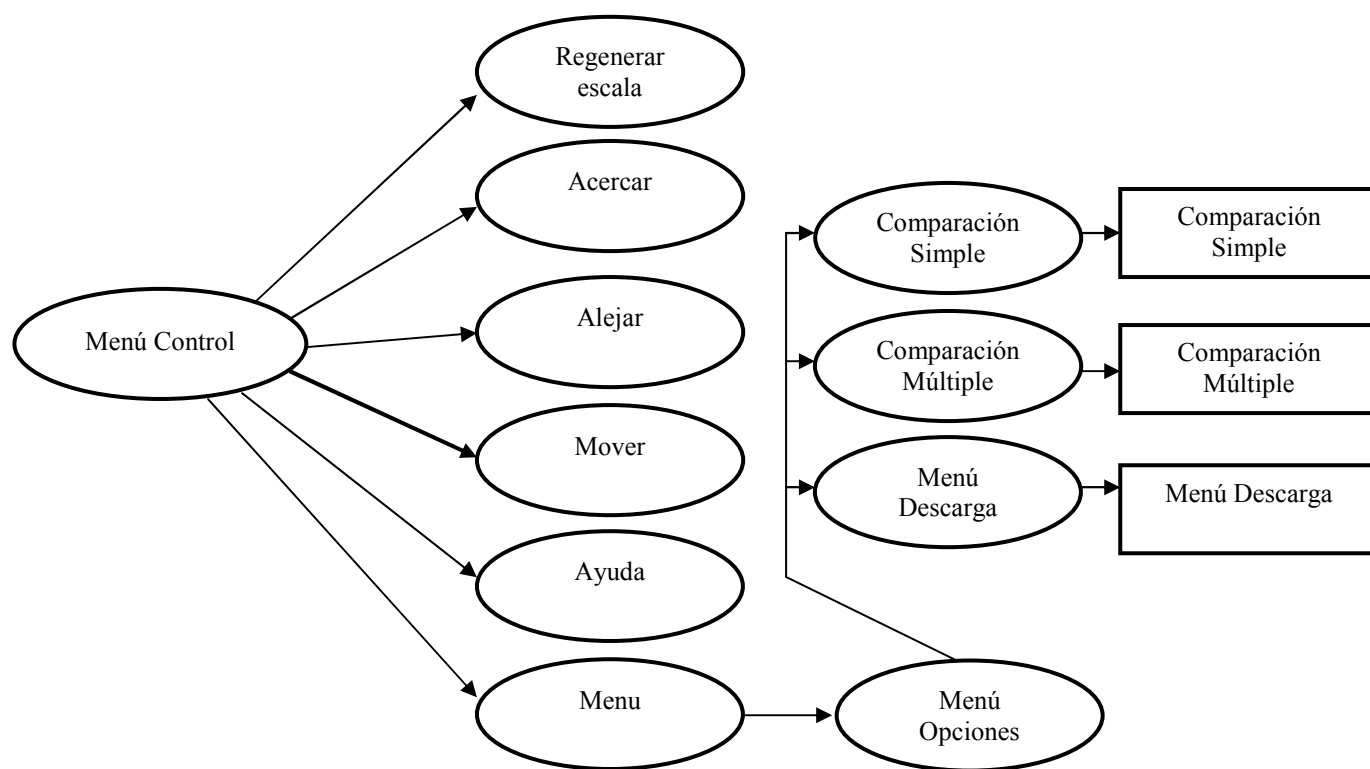


Figura 14. Diagrama del panel Menú Control

Regenerar escala

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	RegenerarEscala
Accesibilidad	Comparación Simple / Comparación Múltiple
Descripción	Redimensiona la gráfica que estamos viendo a su escala original

Acercar

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	Acercar
Accesibilidad	Comparación Simple / Comparación Múltiple
Descripción	Activa el control de Zoom(in) permitiendo situarnos sobre la gráfica y hacer una ampliación de la escala x5 o seleccionar el área sobre la que queremos hacer el zoom.

Alejar

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	Alejar
Accesibilidad	Comparación Simple / Comparación Múltiple
Descripción	Activa el control de Zoom(out) permitiendo situarnos sobre la gráfica y hacer una reducción de la escala x0.5.

Mover

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	Mover
Accesibilidad	Comparación Simple / Comparación Múltiple
Descripción	Activa el control de <i>Drag</i> permitiendo situarnos sobre la gráfica y arrastrarla a izquierda y derecha en busca del área en la que estemos interesados.

Ayuda

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	Ayuda
Accesibilidad	Siempre
Descripción	Llama a una nueva ventana de navegador donde se cargará una guía de usuario, sobre el manejo de la herramienta. (Véase Anexo B)

Menú

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	Menu
Accesibilidad	Siempre
Descripción	Carga el menú de opciones principales de la aplicación

Menú Opciones

Nombre Objeto	MenuOpciones
Tipo Objeto	MovieClip
Nombre Instancia/s	MenuOpciones
Accesibilidad	Botón MenuControl.Menu
Descripción	Menú de selección de las 3 opciones principales de la aplicación

Comparación Simple

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	SelecciónSimples
Accesibilidad	MenuOpciones
Descripción	Carga la opción de comparación simple

Comparación Múltiple

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	SelecciónMultiples
Accesibilidad	MenuOpciones
Descripción	Carga la opción de comparación múltiple

Menú Descarga

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	SelecciónDownload
Accesibilidad	MenuOpciones
Descripción	Carga la opción de Menú Descarga

Panel Gráfica Simple

Nombre Objeto	GraficasSimple
Tipo Objeto	MovieClip
Nombre Instancia/s	GraficaSimple
Accesibilidad	Comparación Simple
Descripción	Panel para seleccionar que par de genomas se desea comparar. Contiene tantas selecciones como pares de genomas haya: $\text{NumeroSelecciones}_n = \text{NumeroSelecciones}_{n-1} + (n-1)$, siendo n el número de genomas cargados en la aplicación y $\text{NumeroSelecciones}_2=1$.

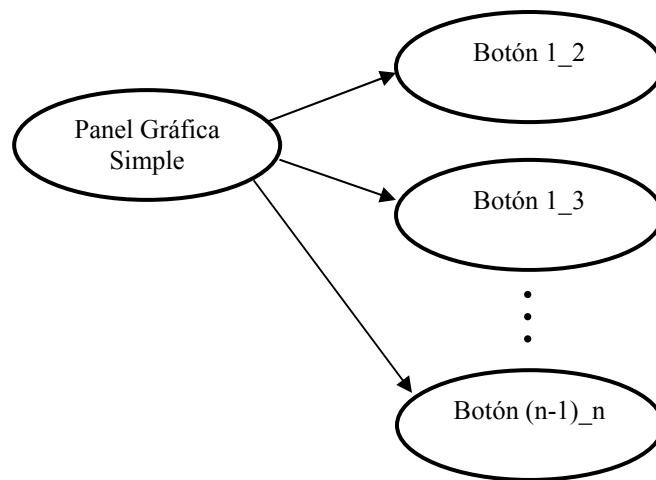


Figura 15. Diagrama del panel de Gráfica Simple

Botón x_y

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	Bxy (habrá $\text{NumeroSelecciones}_n$ instancias, para $x = 1..(n-1)$ y $y = (x+1)..n$)
Accesibilidad	Panel Gráfica Simple
Descripción	Selecciona los genomas a ser comparados y redibuja la gráfica.

Desplazamiento Genoma

Nombre Objeto	ScrollBar
Tipo Objeto	MovieClip
Nombre Instancia/s	ScrollBarEjeX1, ScrollBarEjeX2
Accesibilidad	Comparación Simple
Descripción	<p>Son dos barras de desplazamiento que permiten situar la posición deseada del genoma en el visor, para poder comparar dos regiones de interés, distintas en los genomas seleccionados.</p> <p>Estas barras están sincronizadas con la gráfica representada, así cualquier cambio de escala de la gráfica o desplazamiento lateral del Menú Control se verá reflejado en éstas.</p>

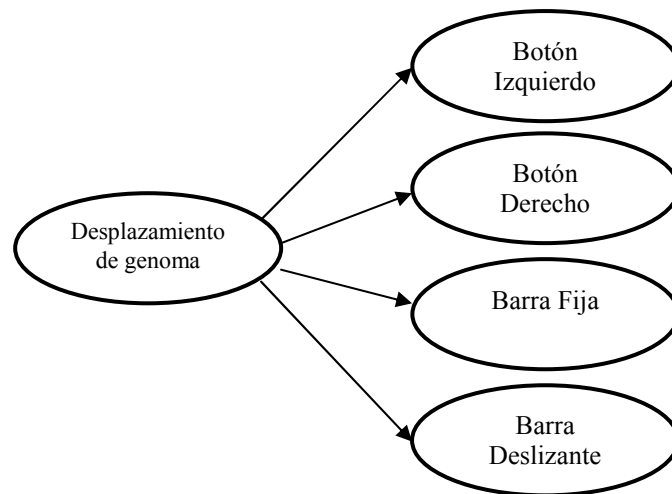


Figura 16. Diagrama del control Barra de desplazamiento de genoma

Botón Izquierdo

Nombre Objeto	BIzquierdo
Tipo Objeto	MovieClip
Nombre Instancia/s	BIzquierdo
Accesibilidad	Desplazamiento de genoma
Descripción	Desplaza la posición del genoma a la derecha, afectando al área contenida en el visor.

Botón Derecho

Nombre Objeto	BDerecho
Tipo Objeto	MovieClip
Nombre Instancia/s	BDerecho
Accesibilidad	Desplazamiento de genoma
Descripción	Desplaza la posición del genoma a la izquierda, afectando al área contenida en el visor.

Barra Fija

Nombre Objeto	BarraFija
Tipo Objeto	MovieClip
Nombre Instancia/s	BarraFija
Accesibilidad	Desplazamiento de genoma
Descripción	Barra trasera que representa la totalidad del tamaño del genoma al que está vinculado. Cuando se presiona en un punto, se produce un desplazamiento hasta esa zona de la barra deslizable.

Barra Deslizable

Nombre Objeto	BarraDeslizable
Tipo Objeto	MovieClip
Nombre Instancia/s	BarraDeslizable
Accesibilidad	Desplazamiento de genoma
Descripción	Barra frontal que se desliza por encima de la fija. Su tamaño y posición muestran la parte de genoma y la región que se está mostrando en el visor.

Panel Gráfica Múltiple

Nombre Objeto	GraficasMultiple
Tipo Objeto	MovieClip
Nombre Instancia/s	GráficasMultiple
Accesibilidad	Comparación Múltiple
Descripción	Panel en el que se muestran los nombres y el orden de muestra de todos los genomas cargados en la aplicación. También nos permite modificar dicho orden, mediante $2*n$ botones, siendo n el número total de genomas cargados en la aplicación.

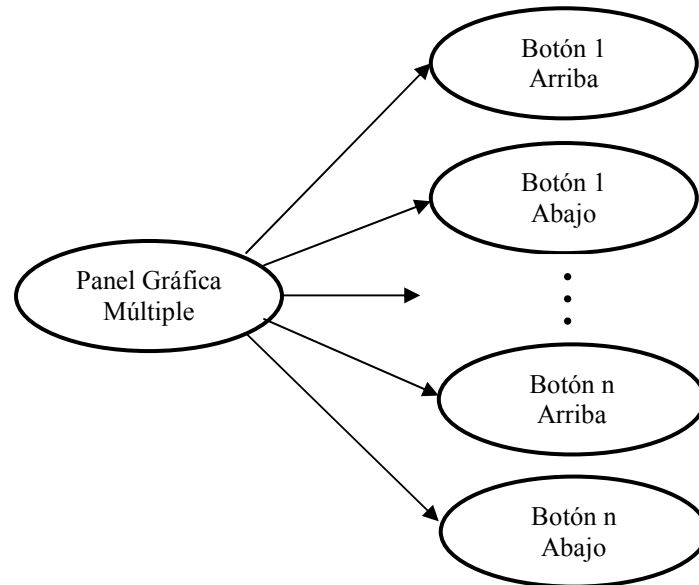


Figura 17. Diagrama del panel de Gráfica Múltiple

Botón x Arriba

Nombre Objeto	BotonArribaP
Tipo Objeto	Button
Nombre Instancia/s	Orden_Upx (n instancias para x=1..n)
Accesibilidad	Panel Gráfica Múltiple
Descripción	Botón que modifica el orden de muestra de la lista de genomas, subiendo una posición el genoma al que está vinculado y bajando el inmediatamente superior. Estos botones son transparentes si están deshabilitados o pueden estar iluminados para mostrar que están dentro del conjunto que se está mostrando en el visor.

Botón x Abajo

Nombre Objeto	BotonAbajoP
Tipo Objeto	Button
Nombre Instancia/s	Orden_Downx (n instancias para x=1..n)
Accesibilidad	Panel Gráfica Múltiple
Descripción	Botón que modifica el orden de muestra de la lista de genomas, bajando una posición el genoma al que está vinculado y subiendo el inmediatamente inferior. Estos botones son transparentes si están deshabilitados o pueden estar iluminados para mostrar que están dentro del conjunto que se está mostrando en el visor.

Número de genomas a ver

Nombre Objeto	ComboBox
Tipo Objeto	ClipCompilado (Herencia: MovieClip > Clase UIObject > Clase UIComponent > ComboBase > ComboBox)
Nombre Instancia/s	NumGens
Accesibilidad	Comparación Múltiple
Descripción	Es una lista desplegable que nos permite seleccionar el número de genomas que deseamos mostrar simultáneamente en el visor para ser comparados cada uno con su siguiente. El número puede ser de un mínimo de 2 y un máximo del número de genomas que hayan cargados en la aplicación. Cada vez que se cambia dicha selección se redibuja la gráfica.

Subir enfoque de genomas

Nombre Objeto	BotonArriba_G
Tipo Objeto	Button
Nombre Instancia/s	ScrollArriba
Accesibilidad	Comparación Múltiple
Descripción	Botón que sirve para subir el enfoque del visor respecto a la lista de muestras ordenada, solo es aplicable cuando el número de genomas seleccionados para comparar en el visor es menor que el número total de genomas cargado en la aplicación y el visor no está mostrando el primero de la lista. Cuando no es aplicable se deshabilita y se puede ver por su transparencia.

Bajar enfoque de genomas

Nombre Objeto	BotonAbajo_G
Tipo Objeto	Button
Nombre Instancia/s	ScrollAbajo
Accesibilidad	Comparación Múltiple
Descripción	Botón que sirve para bajar el enfoque del visor respecto a la lista de muestras ordenada, solo es aplicable cuando el número de genomas seleccionados para comparar en el visor es menor que el número total de genomas cargado en la aplicación y el visor no está mostrando el último de la lista. Cuando no es aplicable se deshabilita y se puede ver por su transparencia.

Gráfica

Nombre Objeto	Grafica
Tipo Objeto	MovieClip
Nombre Instancia/s	Grafica
Accesibilidad	Comparación Simple / Comparación Múltiple
Descripción	Éste es el objeto donde se visualizarán los resultados.

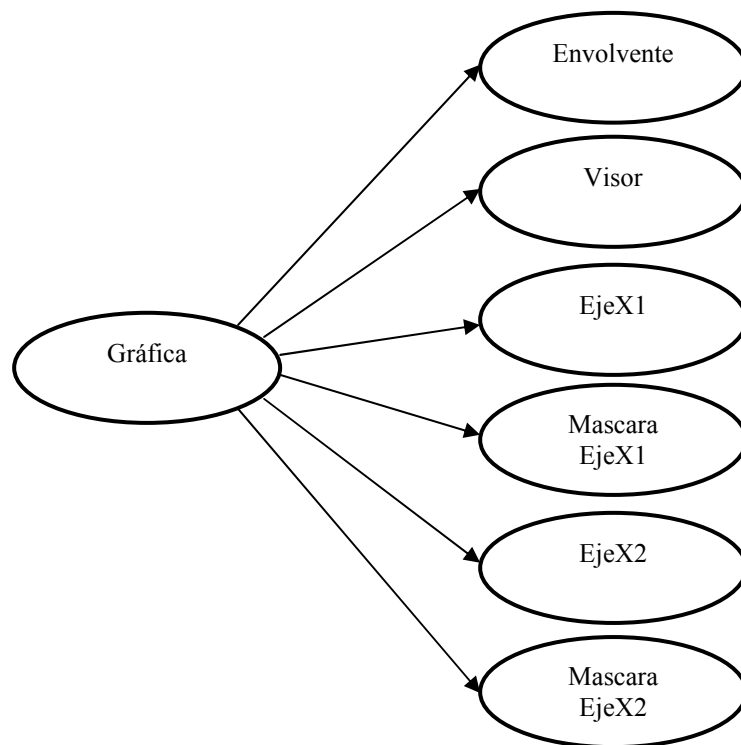


Figura 18. Diagrama de objetos de Gráfica, que contiene los resultados

Envolvente

Nombre Objeto	EmptyMovieClip
Tipo Objeto	MovieClip
Nombre Instancia/s	Envolvente
Accesibilidad	Gráfica
Descripción	Contenido de la gráfica, es el lienzo en el que se trazan las líneas de la comparación de los dos genomas.

Visor

Nombre Objeto	EmptyMovieClip
Tipo Objeto	MovieClip
Nombre Instancia/s	Visor
Accesibilidad	Gráfica
Descripción	Mascara de la Envolvente, es la ventana de los resultados que nosotros vemos de la totalidad de la Envolvente.

EjeX1

Nombre Objeto	EmptyMovieClip
Tipo Objeto	MovieClip
Nombre Instancia/s	EjeX1
Accesibilidad	Gráfica
Descripción	Contiene la regla del eje superior.

Mascara EjeX1

Nombre Objeto	EmptyMovieClip
Tipo Objeto	MovieClip
Nombre Instancia/s	MascaraEjeX1
Accesibilidad	Gráfica
Descripción	Mascara del EjeX1. Área visible del EjeX1, su tamaño se corresponde al tamaño del visor.

EjeX2

Nombre Objeto	EmptyMovieClip
Tipo Objeto	MovieClip
Nombre Instancia/s	EjeX2
Accesibilidad	Gráfica
Descripción	Contiene la regla del eje inferior.

Mascara EjeX2

Nombre Objeto	EmptyMovieClip
Tipo Objeto	MovieClip
Nombre Instancia/s	MascaraEjeX2
Accesibilidad	Gráfica
Descripción	Mascara del EjeX2. Área visible del EjeX2, su tamaño se corresponde al tamaño del visor.

Gestor de descargas

Nombre Objeto	GestorDescarga
Tipo Objeto	MovieClip
Nombre Instancia/s	GestorDescarga
Accesibilidad	Menú Descargas
Descripción	Objeto que cumple la función de realizar descargas de los archivos de MUMs con los que trabajamos, renombrándolos. Nos da la opción de realizar una descarga simple o la de descargar todos los archivos. Ver Figura 12. (página 27)

Selección de genomas

Nombre Objeto	ComboBox
Tipo Objeto	ClipCompilado (Herencia: MovieClip > Clase UIObject > Clase UIComponent > ComboBase > ComboBox)
Nombre Instancia/s	Genoma1, Genoma2
Accesibilidad	Gestor de descargas
Descripción	Selección de los dos genomas cuya comparación de MUMs deseamos descargar en el modo descarga simple.

Opción de MUMs

Nombre Objeto	ComboBox
Tipo Objeto	ClipCompilado (Herencia: MovieClip > Clase UIObject > Clase UIComponent > ComboBase > ComboBox)
Nombre Instancia/s	Direccion
Accesibilidad	Gestor de descargas
Descripción	Selección de archivo/s que deseamos, solo MUMs directos, solo inversos o ambos.

Descarga simple

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	DescargaSimple
Accesibilidad	Gestor de descargas
Descripción	Botón que inicia la descarga de uno o dos archivos en función de las selecciones de Opción de MUMs y Selección de genomas

Descarga todos

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	DescargaMultiple
Accesibilidad	Gestor de descargas
Descripción	Botón que inicia la descarga de todos los archivos tanto directos como inversos de los MUMs.

Gestión de Cola

Nombre Objeto	Cola
Tipo Objeto	MovieClip
Nombre Instancia/s	Cola
Accesible desde	Gestor de descargas
Descripción	Se encarga de gestionar el proceso de descarga cuando son varios los archivos a descargar.

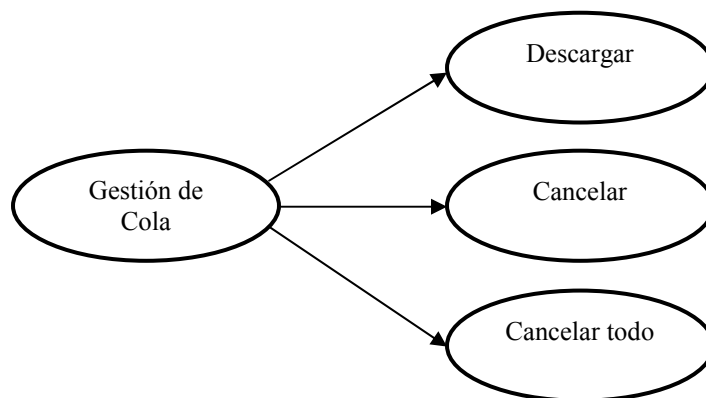


Figura 19. Diagrama de la Gestión de Cola

Descargar

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	SiguienteDescarga
Accesibilidad	Gestión de Cola
Descripción	Botón que inicia la siguiente descarga de la cola.

Cancelar

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	CancelThis
Accesibilidad	Gestión de Cola
Descripción	Botón que cancela la siguiente descarga de la cola.

Cancelar todo

Nombre Objeto	Button
Tipo Objeto	Button
Nombre Instancia/s	CancelAll
Accesibilidad	Gestión de Cola
Descripción	Botón que cancela todas las descargas que hay en cola.

3.5.3.- Estructura de las capas

La aplicación está estructurada en 4 capas, las cuales contienen código y una de ellas también el entorno gráfico:

- **Dibujo**, es una capa de código que contiene las funciones básicas con las que trazamos los MUMs en la envolvente de la gráfica.
- **Botones**, es la capa de código que contiene todas las funciones que corresponden a la interacción con el usuario, como botones, controles y funciones auxiliares que completan las funcionalidades de la aplicación.
- **Graficas**, en esta capa además del código asociado a la generación de las gráficas y las reglas, generamos las máscaras para el visor y para éstas últimas, también contiene

el entorno gráfico, es aquí donde se ubican todos los objetos gráficos que completan la aplicación.

- **Datos**, esta es la capa de código en la que se efectuará la lectura de ficheros, la generación de las estructuras de datos con las que trabajaremos y desde donde se inicializa la aplicación.

Además de estas capas que forman parte de la película principal, también existe otra capa de código contenida en el objeto Desplazamiento genoma o *ScrollBar*. Ésta es la capa “acciones” y contiene, como su nombre indica, todas las acciones con las que se establece la funcionalidad de las barras de desplazamiento de genomas en la Comparación Simple.

3.5.4.- Estructura de los datos

La aplicación recibe por parámetros los códigos que identifican los genomas a ser analizados y sus respectivos nombres. A partir de aquí la aplicación se encarga de encontrar los ficheros ubicados en el servidor para ser leídos y parseados. Estos ficheros han sido generados por un proceso previo (visto en el apartado de Post-proceso), tras analizar y aplicar un algoritmo de *slide Suffix-trees* sobre los datos descargados del *GenBank* del NCBI (National Center for Biotechnology Information) localizado en Maryland(EUA).

Hay dos tipos de ficheros, el que contiene los MUMs directos y el que contiene los inversos. Cada fichero consiste en un texto plano con tres columnas de datos; cada fila es un MUM, la primera columna contiene la base del primer genoma en la que comienza el MUM, la segunda columna es la base del segundo genoma y la tercera es la longitud de este MUM.

Generamos un objeto global que será accesible desde cualquier parte de la aplicación llamado “Datos” (ver Figura 20). Éste contendrá la información recogida de la siguiente forma:

- Datos.VarId[1..8], identificador de cada uno de los genomas cargados en la aplicación.
- Datos.VarIdInfo[1..8].Nombre, el nombre de cada genoma y .Tipo su tipo (Eucariota, Bacteria o Archaea)

- Datos.NumVariables, el número de genomas que se han cargado en la aplicación.
- Las variables Datos.(CE / CE_ok / i / j) se utilizan para la etapa del parsing ya que en determinados momentos no se pueden usar variables locales y necesitamos modificar sus valores desde otras funciones.
- Datos.Colores.Variable[1..8] contiene un valor hexadecimal que corresponderá a cada uno de los genomas.
- Datos.fichero[1..8].(Indice / Nombre / Invertido) es una estructura que utilizamos para la apertura de ficheros y su posterior parseo.
- FicheroMums_dir y FicheroMums_inv son las estructuras que más uso recibirán durante la ejecución de la aplicación. Se podrían definir como un *array* tridimensional, un ejemplo de uso podría ser:

Datos.FicheroMums_dir[id_fichero][num_columna][num_fila], lo que nos retornará un valor que pertenece a un fichero concreto y está en la fila num_fila y la columna num_columna.

Los valores de id_fichero se corresponden a la posición de los genomas en la lista. Un ejemplo podría ser 37, si por ejemplo deseamos comparar el tercer genoma con el séptimo de la lista.

Además también se calculan los valores máximos y mínimos de cada columna almacenándose en Datos.FicheroMums_dir[id_fichero][num_columna].Max / Min, esto se hace en la etapa de *parsing* para evitar tener que calcularlos cada vez que se tenga que redibujar la gráfica.

- Por último Datos.\$cargaEntrada[1..8] son objetos del tipo LoadVars con las que realizaremos el volcado del contenido de los ficheros.

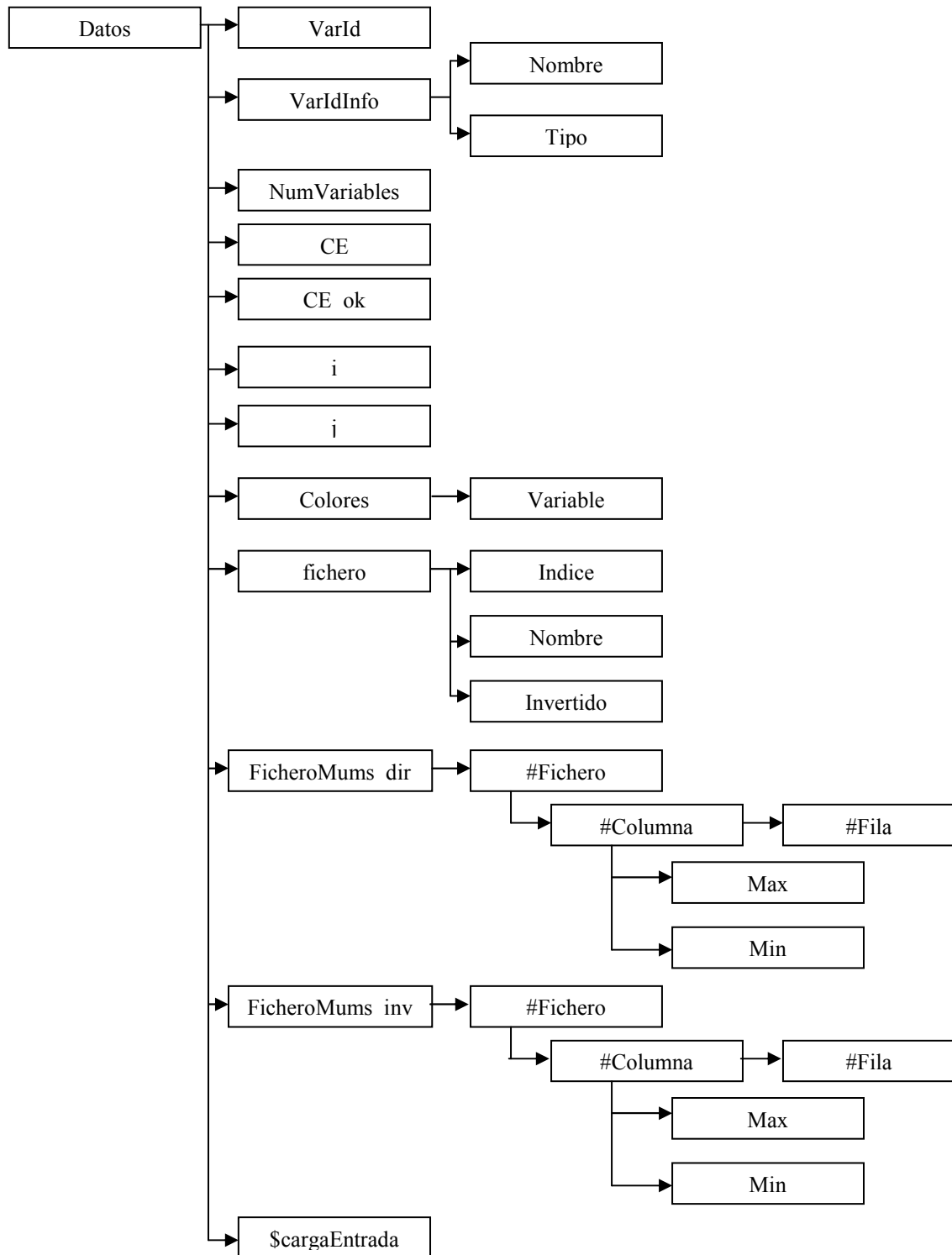


Figura 20. Estructura de la variable global Datos

Durante la ejecución los parámetros de una gráfica se almacenan en una variable que es un array de estructuras denominada Graficas en la que podemos almacenar valores con los que definiremos tanto las gráficas como el comportamiento de la aplicación.

La variable está en la raíz por lo tanto siempre accedemos especificando la gráfica sobre la que deseamos modificar u obtener alguna propiedad (“grafica_actual”) de la siguiente manera:

`_root.Graficas[grafica_actual].Propiedad`

Gráficas

Propiedad	Tipo de gráfica	Descripción
Nombre	Simple y Múltiple	<i>String</i> : Nombre que se le asigna a la gráfica
Tipo	Simple y Múltiple	<i>String</i> : Puede ser “simple” o “multiple”
Dir	Simple y Múltiple	<i>Bool</i> : Valor booleano que determina si la visualización de MUMs directos está activa o no
Inv	Simple y Múltiple	<i>Bool</i> : Valor booleano que determina si la visualización de MUMs inversos está activa o no
Escala	Simple y Múltiple	<i>Number</i> : Determina la escala en porcentaje horizontal a la que se dibuja la gráfica, el valor inicial es 1 y equivale a 100%
EscalaX	Simple y Múltiple	<i>Number</i> : Determina la escala horizontal a la que se dibuja la gráfica sobre la envolvente, por ejemplo 3,5 equivale a 1 : 3,5
EscalaY	Múltiple	<i>Number</i> : No funciona como una escala real, sino que marca la distancia vertical que existe entre dos genomas de la comparación múltiple
PosicionX	Simple y Múltiple	<i>Number</i> : coordenada horizontal en la que se sitúa el inicio de la Envolvente que contiene la gráfica
PosicionY	Simple y Múltiple	<i>Number</i> : coordenada vertical en la que se sitúa el inicio de la Envolvente que contiene la gráfica
Max	Múltiple	<i>Number</i> : Valor máximo de las bases que deben ser representadas en los genomas de la gráfica múltiple
MaxX1	Simple	<i>Number</i> : Valor máximo de las bases que deben ser representadas del genoma vinculado al eje superior de la gráfica simple
MaxX2	Simple	<i>Number</i> : Valor máximo de las bases que deben ser representadas del genoma vinculado al eje inferior de la gráfica simple

Desplazamiento_X1	Simple	<i>Number</i> : Valor de desplazamiento sobre la PosicionX en el genoma del eje superior que afecta a la representación de los datos en éste
Desplazamiento_X2	Simple	<i>Number</i> : Valor de desplazamiento sobre la PosicionX en el genoma del eje inferior que afecta a la representación de los datos en éste
ArrayOrden[1..8]	Múltiple	<i>Array::Number</i> : Lista de que establece el orden de los genomas que se comparan en la gráfica múltiple

3.5.5.- Entorno gráfico

Las bases para el diseño estaban establecidas respecto a otra aplicación gráfica que tenía por objeto el *Pattern-Analysis* que es un análisis de la relación entre genes. Por lo que se pretendía respetar el formato de la aplicación.

A partir de aquí vamos a mostrar el entorno gráfico referenciando los objetos y funciones con los que se trabaja en la aplicación:

- Inicialización de la aplicación. Durante la carga de ésta puede existir una demora según el volumen de carga de ficheros.

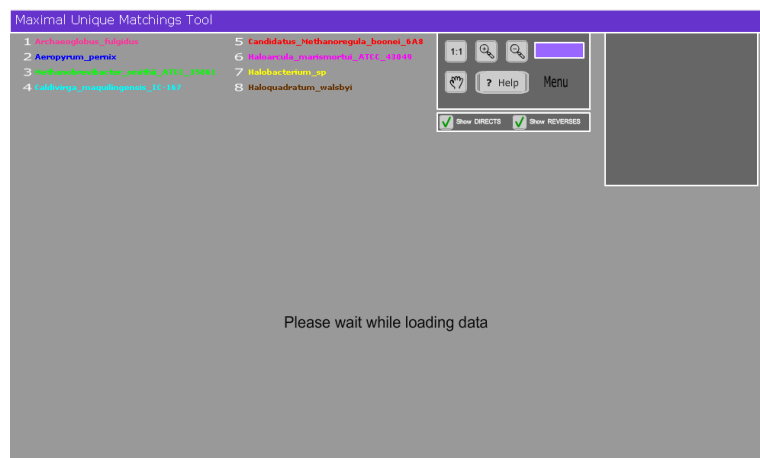


Figura 21. Captura de inicialización de la aplicación

- *Pair Wise Comparison* o comparación simple.

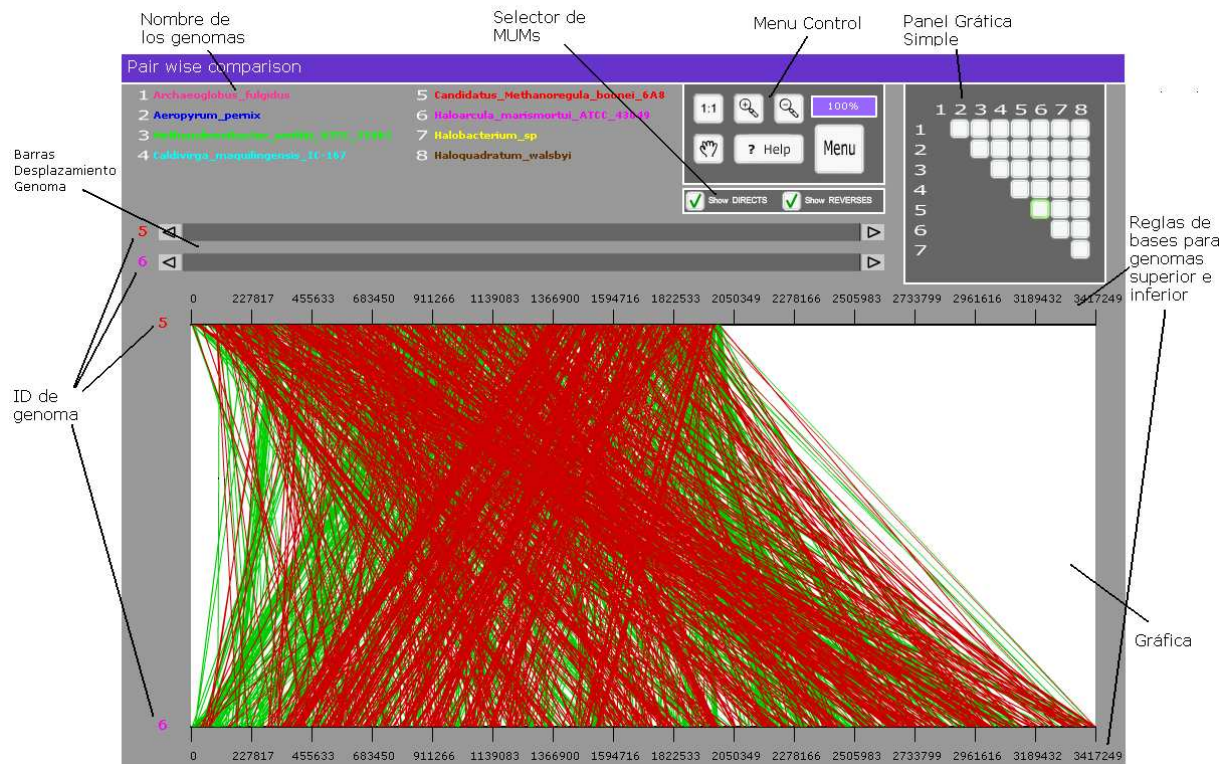


Figura 22. Formato de la comparación simple

Como podemos ver en la Figura 22, la gráfica nos muestra los MUMs tanto directos como inversos (verdes y rojos respectivamente) existentes entre el genoma 1 y el 2; en el Panel gráfica simple podemos seleccionar que dos genomas deseamos comparar y las dos barras de “Desplazamiento genoma” nos permitirán desplazar de manera exclusiva el área que deseamos visualizar.

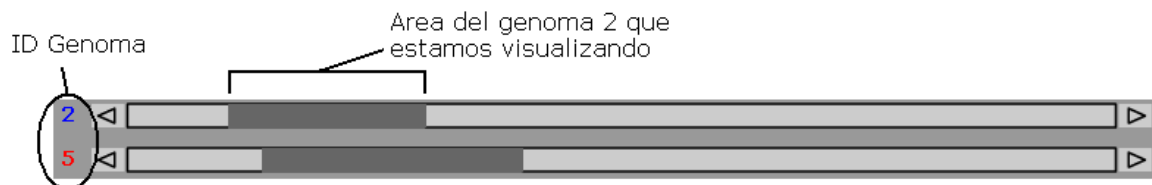


Figura 23. Barras de desplazamiento de genoma

En el siguiente ejemplo (Figura 24), podemos apreciar el desplazamiento de la barra del genoma superior hacia la derecha y la correspondiente actualización de la visualización de la gráfica.

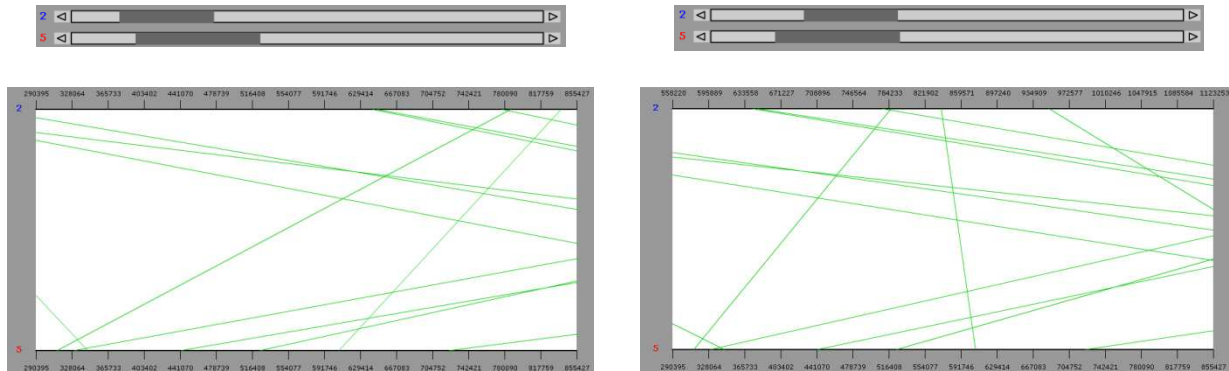


Figura 24. Comportamiento de las barras de desplazamiento de genomas

- Comparación Múltiple.

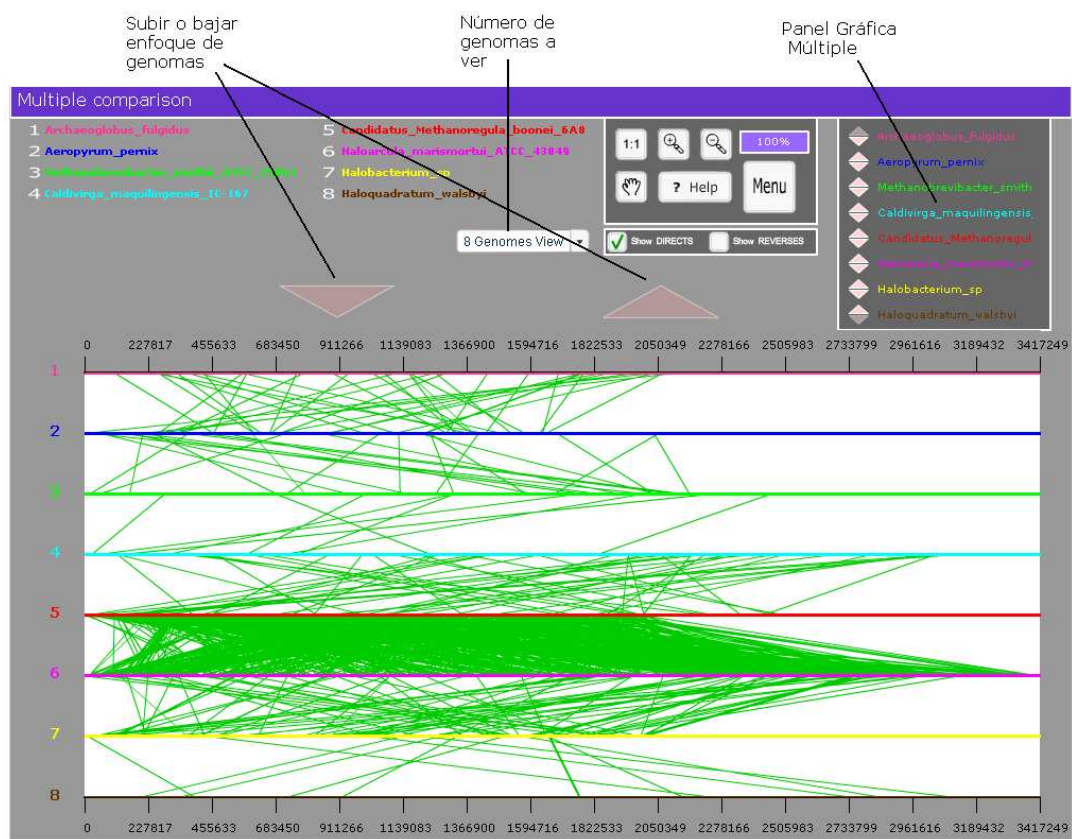


Figura 25. Formato de la comparación múltiple

Ahora simularemos una comparación con un número menor de genomas, por ejemplo 4 mediante el selector de número de genomas, bajaremos el enfoque suponiendo el contexto en el que el primer genoma no nos interesa y además deseamos ver la relación que guarda el octavo con el segundo y el tercero respectivamente, por lo que reordenaremos las posiciones mediante el Panel Gráfica Múltiple.

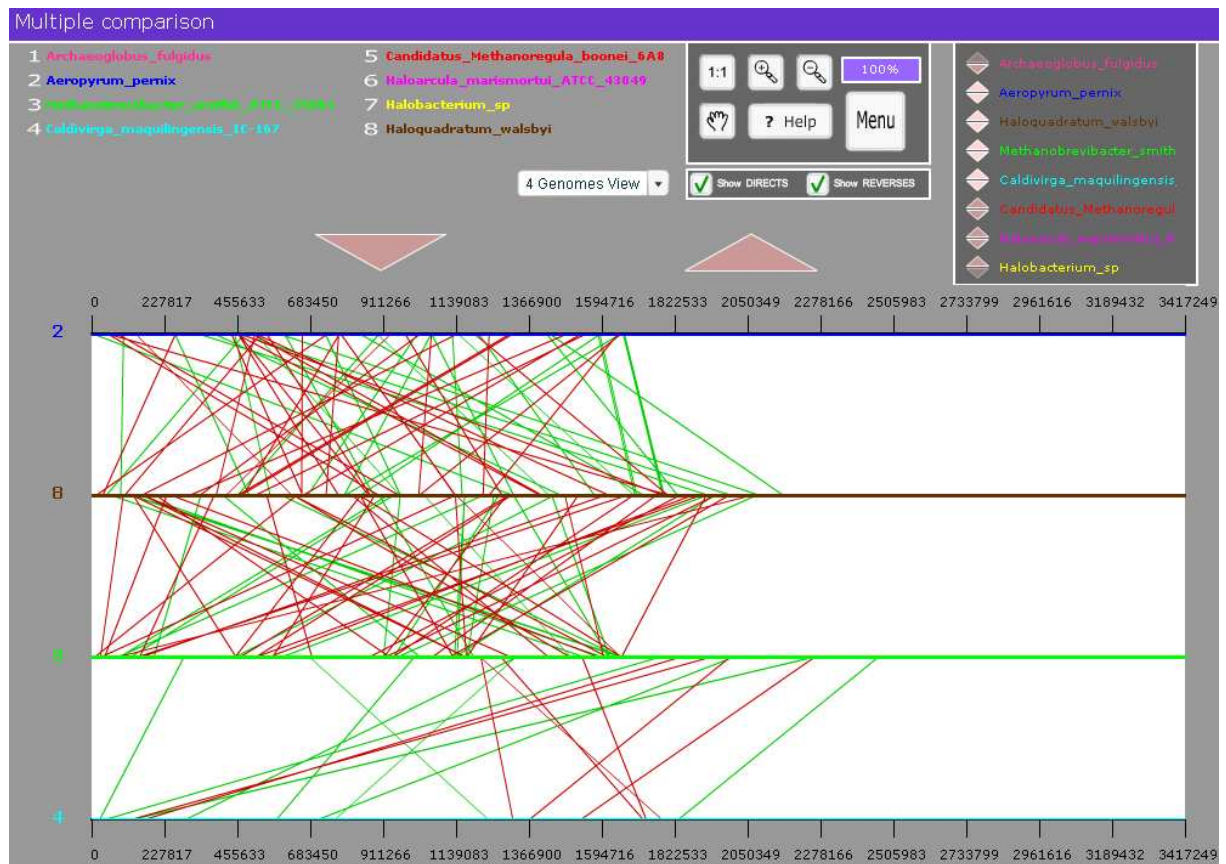


Figura 26. Comparación múltiple sujeta a la interacción del usuario

Menú Control y Selector de MUMs.

Estos paneles contienen controles comunes a la comparación simple y múltiple.

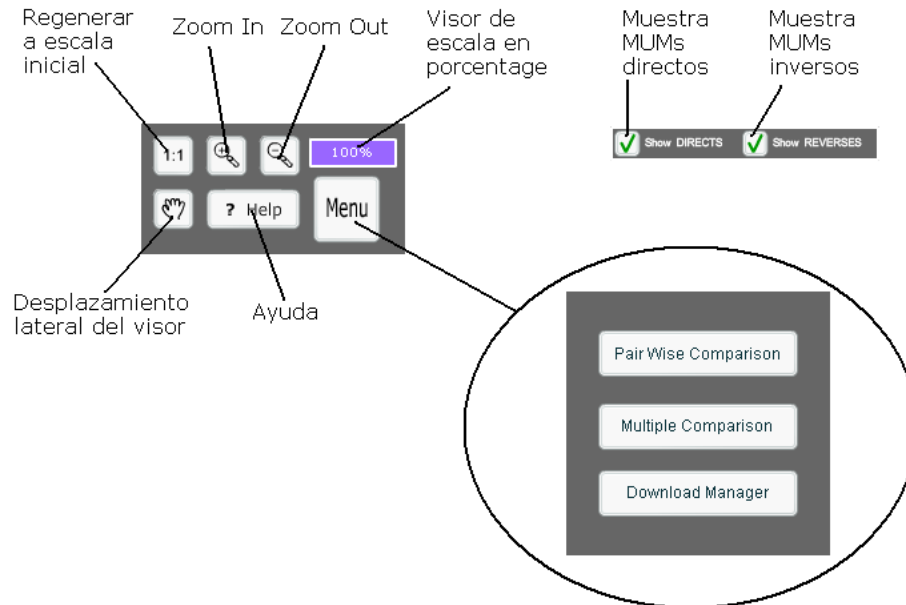


Figura 27. Menú control y Selector de MUMs

Menú Descarga

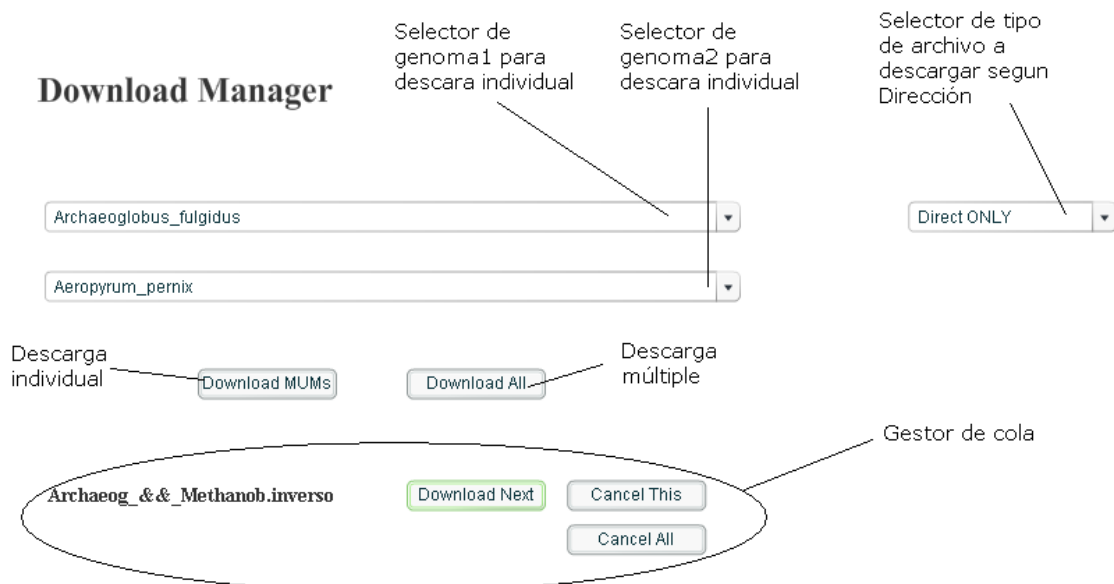


Figura 28. Opción de Menú Descarga

4.- PROBLEMAS ENCONTRADOS Y SOLUCIONES PROPUESTAS

A lo largo del desarrollo de la aplicación aparecieron múltiples problemas que provocaban parones en el proceso, de éstos vamos a mencionar los más relevantes o los que implicaban decisiones a continuación:

Problema 1:	<p>Carga de datos inconsistentes.</p> <p>En la carga de datos, durante la lectura y <i>parsing</i> de los ficheros, algunos de estos no se abrían a pesar de contabilizarlos, dando lugar a una estructura de datos incompleta que provocaba errores en la aplicación.</p> <p>El problema surgía a raíz del bucle en el que se gestionaba la carga de archivos mediante la función <i>LoadVars.onData()</i>.</p>
Posibles soluciones planteadas:	El problema se producía por que el bucle se completaba sin haber efectuado la carga de ficheros, ya que la función de volcado era más lenta que el resto de la ejecución y necesitaba variables de control que fueran accesibles tanto desde la función de parsing como de la de de volcado.
Solución escogida:	Aprovechando la estructura de Datos se crearon dos variables en la raíz de ésta que llevarían el control de la carga Datos.CE y Datos.CE_ok. Una vez finalizada la carga de todos los datos éstas son eliminadas de la estructura Datos.

Problema 2:	<p>Demora en las ampliaciones de la gráfica con el Zoom.</p> <p>Una de las claves, ya planteada en los objetivos, era los problemas que podrían surgir a raíz del análisis de genomas grandes con número de bases del orden de Gigas.</p> <p>El principal problema surgía cuando realizaba un Zoom, a partir de cierta escala (alrededor del 20.000% del tamaño original 1:1), la gráfica tardaba mucho en redibujarse, llegando a demorarse varios minutos cuanto mayor era el Zoom. Lo peor es que Flash Player acaparaba la CPU durante estos periodos, ralentizando todo el PC.</p>
--------------------	--

Posibles soluciones planteadas:	<p>En un principio interpreté erróneamente que el problema se debía al tamaño del lienzo (la envolvente sobre la que dibuja la gráfica), ya que flash no estaría concebido para contener <i>MovieClips</i> de tamaños exageradamente grandes, por lo que propuse la siguiente solución:</p> <ul style="list-style-type: none"> ▪ Limitar el Zoom a 35.000% cuya demora era asumible y la separación entre las bases podía ser suficiente para su análisis. <p>Sin embargo más tarde indagando en ello, descubrí que el problema no era el tamaño de la gráfica sino de las reglas y más concretamente el bucle, en el que se dibujaban las marcas y se calculaban los valores (números de bases) asociadas a éstas sobre los ejes de las reglas, por lo que propuse por la siguiente solución:</p> <ul style="list-style-type: none"> ▪ Reducir dicho bucle, dibujando solo las marcas y calculando los valores que aparecieran a la altura de las mascarar de las reglas, además de las inmediatamente anteriores y posteriores. ▪
Solución escogida:	<p>Obviamente opté por la segunda solución, sin embargo se planteaba un problema; cuando se realizaba una desplazamiento de la gráfica a través del Menú de Control, que solucioné con una función (<i>AmpliaEjes()</i>) que ampliaba las marcas y sus valores, en la dirección en la que se realizase el desplazamiento.</p>

Problema 3:	<p>Tamaño de texto para las bases grandes.</p> <p>En base a los archivos de ejemplo con los que trabajaba durante la primera etapa de la realización del proyecto (del orden de 40.000 bases) estipulé un tamaño para los valores de la regla. Sin embargo se me advirtió más adelante que trabajaría con MUMs del orden de millones de bases por lo que debía hacer más pequeño el tamaño del texto para que no quedaran superpuestas las bases.</p>
Posibles soluciones planteadas:	<ul style="list-style-type: none"> ▪ Establecer un tamaño menor para el texto de las bases ajustándose a la máxima longitud que podían alcanzar éstas. ▪ Ajustar dinámicamente el tamaño del texto en función de la gráfica que tuviera que ser representada un su valor máximo de base.

Solución escogida:	La segunda opción fue la escogida ya que tan solo era necesario hacer un pequeño cálculo de longitud dependiendo de cual fuera la gráfica activa.
--------------------	---

Problema 4:	<p>Error de grosor de línea para longitudes altas.</p> <p>Al trazar una línea en Flash de longitud mayor a 2^{15} pixels el grosor de ésta se altera, por ejemplo, a pesar de que se le asigne un grosor de línea de 1 flash puede darle un valor aleatorio, eliminando la aproximación de la gráfica con los datos reales. Esto se debe a un <i>bug</i> de Flash Player que podemos encontrar en la página de <i>bugs</i> de Adobe (anexo A).</p> <p>Este problema nos afecta cuando hacemos zooms grandes, ya que el tamaño de la gráfica en la Envolvente es muy superior a este límite y pueden existir multitud de líneas afectadas por este problema.</p>
Posibles soluciones planteadas:	<p>Se han propuesto 2 soluciones a este problema:</p> <ul style="list-style-type: none"> ▪ No representar en la gráfica las líneas que superan esta longitud, ya que éstas a efectos prácticos no aportan ninguna información válida para la comparación sobre el área del visor. ▪ Subdividir las líneas de mayor longitud dibujándolas por partes. Esta solución afecta a la rapidez de la presentación de la gráfica, ya que a determinados zooms estamos multiplicando el número de líneas que se deben dibujar en la envolvente incluyendo sus cálculos.
Solución escogida:	Se ha escogido la segunda solución para respetar el realismo de los datos aun considerando la demora temporal que produce; al fin y al cabo ésta se produce cuando llegamos a zooms algo desorbitados.

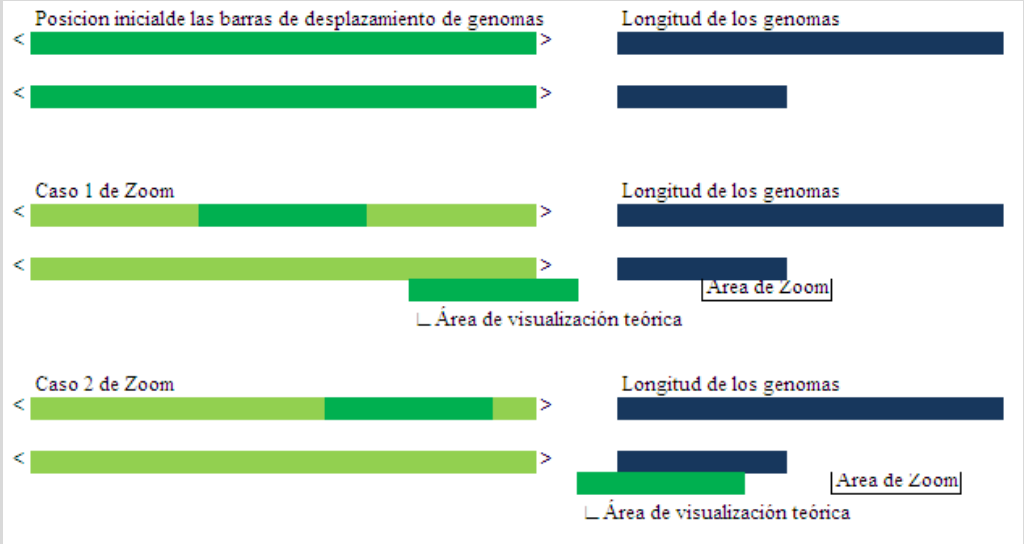

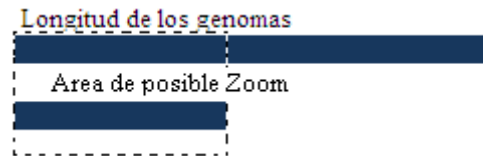
Problema 5:	<p>Incoherencia de visualización en función de la región del Zoom.</p> <p>Cuando nos encontramos en la escala original 1:1 y pretendemos hacer un Zoom, nos surge un problema si al realizar éste lo hacemos fuera del área de uno de los genomas. Esto se debe a que normalmente los genomas tienen diferente longitud (siempre teniendo en cuenta, que no representamos los genomas sino sus relaciones con otros mediante los MUMs).</p> <p>En la siguiente figura veremos las barras de desplazamiento donde se aprecia el tamaño total del genoma y el área que estamos visualizando a la izquierda, mientras que en la derecha vemos la región de los genomas sobre la que hacemos el Zoom:</p>  <p>Figura 29. Estudio de comportamientos Zoom</p> <p>Como podemos apreciar si en el caso 1, una parte del genoma quedaría fuera del visor, mientras que en el segundo caso uno de los genomas estaría completamente fuera de éste.</p>
Posibles soluciones planteadas:	<p>Teniendo en cuenta que nunca debía existir la posibilidad de que en el visor no estuviera representado alguno de los genomas o parte de ellos, el límite de visualización debía ser como máximo el final de genoma como se puede ver en la siguiente figura.</p> 

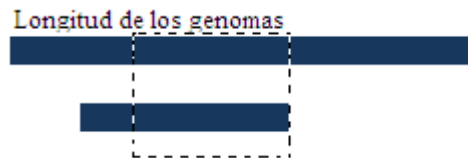
Figura 30. Limite del visor

Por lo que se optó por dos posibles soluciones:

- Que solo se pudiera realizar Zoom sobre un área que contuviera información de los dos genomas en cuestión.

**Figura 31. Área ampliable**

- La segunda opción contemplaba la posibilidad de desplazar el genoma hasta el marco del visor. Además también se le imponía un límite de área ampliable correspondiente a la longitud total del tamaño del genoma más pequeño, de esta manera nunca quedaría el visor desocupado.

**Figura 32. Desplazamiento del genoma más pequeño**

Solución
escogida:

La segunda opción fue la seleccionada por motivos funcionales. El objeto era poder comparar el genoma pequeño con el área que nos pudiera interesar del genoma más grande.

Problema 6:	Descarga de fichero único. <p>En la tarea opcional de descargar los ficheros si era requerido, se me pedía que existiera la opción de descargase a través de un botón todos los ficheros con los que estaba trabajando la aplicación. Para ello cree un archivo externo php junto a una librería de compresión en zip para este lenguaje. Comprobé su funcionalidad en un servidor web propio pero no funcionaba en el servidor del IBB por que este tenía deshabilitados los permisos de php para crear archivos.</p>
Posibles soluciones planteadas:	<ul style="list-style-type: none"> ▪ Habilitar los permisos en el servidor de IBB ▪ Crear un gestor de descargas dentro del propio programa para descargar los archivos uno a uno.
Solución escogida:	<p>La segunda opción fue la escogida y ya que se había optado por ésta, amplié la funcionalidad para que se pudiera escoger realizar una descarga individual escogiendo los genomas involucrados en el fichero.</p>

5.- CONCLUSIONES

Se ha presentado una interfaz gráfica interactiva eficiente que cumple con todos los requisitos funcionales y no funcionales especificados. El objeto: visualizar la comparación de genomas basada en sus MUMs. Para ello se ha cuidado la precisión de las representaciones gráficas, superando problemas importantes debidos, por ejemplo, al tamaño de los genomas o a la complejidad de sincronizar parámetros de escala, desplazamientos y sobretodo el desplazamiento individual de genomas utilizado para situar a éstos en la posición deseada del visor.

Esta aplicación puede resultar muy útil a la comunidad científica para observar las similitudes entre genomas, las diferentes cadenas de genes que tienen en común, las evoluciones que han ido apareciendo entre diferentes organismos o los genes que comparten unas especies con otras, incluida la humana, ya que nos aportará mucha información incluso sobre nosotros mismos según se vayan descubriendo las funciones que cumplen éstos.

El ámbito de la genómica abarca una gran cantidad de áreas de estudio, cuya elevada complejidad obliga a la especialización. Sin embargo, aunque estamos aún lejos de desvelar el significado de cada una de las secuencias que componen un genoma, los avances de los diferentes estudios son exponenciales y la investigación genómica puede contribuir ampliamente a mejorar la condición y calidad de vida humana, por ejemplo al descubrir que genes nos provocan ciertos males y como inhibirlos

La aplicación FILOMUMS desde la que arranca la herramienta creada en este proyecto, aporta respecto a los programas existentes mencionados en el apartado del estado del arte, una interfaz para el usuario desde la que se pueden seleccionar los genomas desde un árbol, en el que ya se puede apreciar la similitud existente entre ellos, mientras que hasta ahora se había tenido que seleccionar los genomas por su nombre para ser analizados.

Con motivo de la ampliación de funcionalidades de la aplicación de este proyecto en un futuro, se ha respetado una estructura clara, con todo el código comentado. Como ejemplo de dichas ampliaciones funcionales podemos apuntar a la posibilidad de poder localizar los genes dentro del genoma y así entender el significado biológico de éstos, gracias a los MUMs, entre

un genoma y otro; también, se podría acceder a las Bases de datos del NCBI para consultar estos genes y la información biomédica conocida sobre estos; poder seguir rastros entre las similitudes por MUMs, es decir, si una región de un genoma se corresponde por MUMs con la región de otro genoma y ver si esta otra región se corresponde con otra región concreta de un tercer genoma; o como poder localizar los genes mediante un buscador y visualizar la región del genoma correspondiente a estos genes.

Como apunte personal quiero decir que me siento realizado por el desarrollo de este proyecto y no descarto la posibilidad de orientar mi carrera profesional al ámbito de la investigación.

6.- REFERENCIAS

1. Isabel A. Nepomuceno-Chamorro, Seminario Bioinformática
2. **Alignment of Whole Genomes.** A.L. Delcher, S. Kasif, R.D. Fleischmann, J. Peterson, O. White, and S.L. Salzberg, *Nucleic Acids Research*, 27:11 (1999), 2369-2376.
3. Huerta,M. and Messeguer,X. (2002) **Efficient space and time multicomparison of genomes.** Research Report LSI-02-64-R, Dep. Llenguatge i Sistemes Informàtics, Universitat Politècnica de Catalunya.
4. Ferre D, Roset R, Huerta M, Adsuara JE, Rosello L, Alba MM, Messeguer X: **Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN.** *Nucleic Acids Res* 2003, **31**(13):3651-3653.
5. T. Treangen and X. Messeguer. **M-GCAT: Interactively and efficiently constructing large-scale multiple genome comparison frameworks in closely related species.** *BMC Bioinformatics* 2006, 7:433.

FIGURAS

Figura 1: An approximate phylogeny of genome comparison tools over the past 30 years.

Imagen modificada, <http://www.biomedcentral.com/1471-2105/7/433/figure/F1>

Figura 3: Árbol de los 3 Dominios planteados por Carl Woese. Esquema de Norman Pace (1977)

Figura 4: Nucleótido. <http://es.wikipedia.org/wiki/Nucle%C3%B3tido>

Figura 5: ADN. <http://haplogroup-i.com/img/A4dna.jpg>

Figura 6: Suffix tree. Revolution research: Didactic Material

BIBLIOGRAFÍA ON-LINE

Genómica y Bioinformática:

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

<http://bioinformatica.uab.es/diaposcurso/>

<http://revolutionresearch.uab.es/main.html>

http://danival.org/100%20biolomar/4000notasbio/clas/clas_3_dominios.html

http://bioinfo.cipf.es/courses/UBA_CIPF_3/?q=leer_mas

<http://www.lsi.us.es/~janepo/janepo/Seminario-LSI-BioInf.pdf>

ActionScript 2:

<http://www.adobe.com/support/documentation/es/flash/>

<http://www.manualespdf.es/manual-lenguaje-actionscript-2-0/>

http://livedocs.adobe.com/flash/9.0/main/flash_as2_learning.pdf

<http://www.gotoandlearn.com/>

7.- RESUMEN

Castellano

El proyecto consiste en un entorno gráfico cuyo fin es el de visualizar, estudiar e interpretar la conservación de código genético existente entre los diferentes genomas.

Una interface que permite cargar hasta ocho genomas para ser comparados en detalle, por pares o entre todos ellos a la vez.

El gráfico que se muestra en la interfaz, representa los *Maximal Unique Matchings* entre cada par de genomas, lo que significa coincidencias de la mayor longitud posible no repetidas, en las secuencias de ADN de las especies comparadas.

La finalidad es el estudio de las evoluciones que han ido apareciendo entre diferentes organismos o los genes que comparten unas especies con otras.

Català

El projecte consisteix en un entorn gràfic la fi del qual és el de visualitzar, estudiar i interpretar la conservació del codi genètics entre els diferents genomes.

Una interfície que permet carregar fins a vuit genomes per a ser comparats en detall, per parells o entre tots ells alhora.

El gràfic que es mostra a la interfície, representa els *Maximal Unique Matchings* entre cada parell de genomes, el que significa coincidències de la major longitud possible no repetides, en les seqüències d'ADN de les espècies comparades.

La finalitat és l'estudi de les evolucions que han anat apareixent entre diferents organismes o els gens que comparteixen unes espècies amb unes altres.

English

The project consists of a graphical environment whose aim is to visualize, explore and interpret the conservation of the genetic code between different genomes.

An interface that allows you to upload up to eight genomes to be compared in detail, in pairs or all at once.

The graph shown in the interface, represents Maximal Unique Matchings between each pair of genomes, which means the longest possible non-repeated matches in DNA sequences of the compared species.

The purpose is to study the evolution that have emerged between different organisms or genes shared with other species.

ANEXO A

[FP-495] Corrupted line thickness for line segments longer than 2^15 pixels - Adobe Bug System

Page 1 of 1

Issue Details

(XML)

Flash Player

Corrupted line thickness for line segments longer than 2^15 pixels

Created: 02/10/08 12:35 PM Updated: 09/05/08 04:29 PM

Component/s:

Rendering Shapes

Security Level:

Public (All JIRA Users)

File Attachments:

Manage Attachments

1. LineStyleBug.mxml (1.0 kb)

2. LineStyleWorkaround.mxml (1 kb)

Severity:

Incorrectly Functioning with Workaround

Reproducibility:

Every Time

Discoverability:

Medium

Found in Version:

SDK Moxie RC4

Milestone:

Flash Player 9

Affected OS(s):

Windows - All Windows

Injection:

No

Steps to Reproduce:

Steps to reproduce:

1. Run attached LineStyleBug.mxml

2. Drag the slider to set the length of the drawn line to a value larger than 2^15 pixels

Actual Results:

Line thickness becomes inconsistent, varies with each redraw.

Expected Results:

Line thickness should stay at the set value of 1 pixel.

Workaround (if any):

Draw long lines using segments shorter than 2^15 pixels, as shown in LineStyleWorkaround.mxml.

This bug was observed with SDK version 3.0.0.338.

Language Found:

English

Bugbase Id:

none

Triaged:

No

QA Owner:

Joan Lafferty

Resolved by:

Charles Liss

Participants:

Joan Lafferty, Julian Frumar, Simon Ratner and Tinic Uro

All

Comments

Sort Order:

Julian Frumar - [07/21/08 06:26 PM]

Happens in OS X Leopard too.

[Permalink | « Hide]

Joan Lafferty - [07/29/08 05:30 PM]

I think that this is a player limitation.

[Permalink | « Hide]

Tinic Uro - [08/07/08 04:07 AM]

This is an elected limitation. We can not change this without a large performance impact, we would have to switch to a larger number type which is very expensive. The rasterizer in the Flash Player is using 16.16 fixed point integers. The way we use them still results in better performance than pure 32bit floating point which has it's own problems. (jittering and such. You would hate it more than the support for very large strokes, believe me).

[Permalink | « Hide]

ANEXO B

Help :: MUMs Interface

Functionality

The first direct application is to study the common DNA subsequences between two different genomes. For this comparison we use the Maximal Unique Matchings, which means the longest coincidences between two genome sequences that must be uniques and can't be repeated along each genome.

In this tool, can be loaded from two to eight genomes which names can be viewed at the top-left, for making comparisons in couples and user can choose three options in the Menu:

The first option "pair wise comparison" is a simple comparison of two genomes that can be selected in the top-right panel. User can manipulate the graphic with the Control Menu, the panel up centered and selecting the direct, reverses or both MUMs (Maximal Unique Matchings) to be displayed in the MUMs panel under the Control Menu. User can shift the genomes to make coincide two areas of the genomes where user is interested in too, by means of the Shift Genome Control Bars displayed on the left over the graphic.







The second option "multiple comparison" is a multiple comparison of a few genomes that means the comparison between each one with the next in the list. This list can be changed in the top-left panel and the Genome View Controls, the number of genomes to be viewed in the graphic can be selected by means of a list which it's default selection is the total number of genomes loaded in the application, for example "8 Genomes View" placed in the center. If user have chosen a lower number of genomes to be displayed than the total number loaded, user can shift the graphic vertically for viewing the other genomes of the list. User can manipulate the graphic with the Control Menu, the panel up centered and selecting the direct, reverses or both MUMs (Maximal Unique Matchings) to be displayed in the MUMs panel under the Control Menu.

The third option "download manager" contains a little form for downloading the MUM files with the data showed in the graphics, there user can download a simple MUM's file or all the files loaded in the application.

Control Menu

The Control Menu offers different tools that allow the user to perform basic manipulations on the graphic plots. Here is the list of the options provided.



Control Menu Options

Control	Name	Description	Function
	Regenerate Scale	Original Scale	Clicking will regenerate the plot to a 100% scale plot with 1:1 scale ratios between the graphic horizontal axis, allowing the user to see the all data displayed in the same scale along the plot area.
	Zoom-in	Zoom-in Tool	Clicking will allow the user to draw a zoom window on any area of the plot and the Flash Interface will perform a zoom-in operation in it.
	Zoom-out	Zoom-out Tool	Clicking will allow the user to click on any point of the plot and the Flash Interface will perform a zoom-out operation reducing the scale to a 50% and centering the resulting plot on the clicked point.
	Drag	Drag Tool	Clicking will allow the user to click on any point of the plot and move it in horizontal direction. The Flash Interface will move the plot accordingly with the user movements.
	Help	User help	Clicking will allow the user to open a simple user guide to understand the controls of the application.
	Menu	Menu Button	Clicking will make the Flash Interface display the different options available in the secondary menu area: Pair wise comparison, multiple comparison and download manager. See the below the information offered for this options.

MUM's Panel

The MUM's Panel offers the option of checking or unchecking the viewing of the direct and reverse MUMs between each couple of genomes on the graphic plots. Here the user can decide what kind of MUMs wants to be displayed. Here is the list of the options provided.

MUM's Panel Options

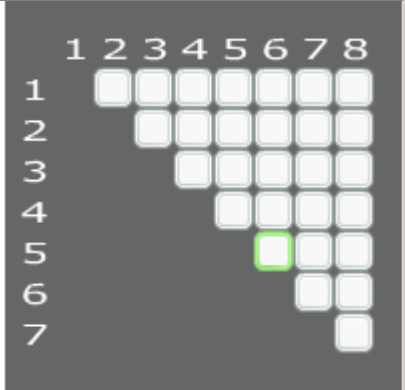
Control	Name	Description	Function
	Show DIRECTS	Show Direct MUMs	Clicking will set the option of enabling / disabling the showing of the direct MUMs of the comparison.
	Show REVERSES	Show Reverse MUMs	Clicking will set the option of enabling / disabling the showing of the reverse MUMs of the comparison.

PAIR WISE COMPARISON

The Pair Wise Comparison let the user compare graphically two genomes by its MUMs.


The Pair Wise Comparison Panel

The Pair Wise Comparison Panel selection dialog allows the user to select the different genome couple combinations to been plot.

Panel	Name	Description	Function
	Pair Wise Comparison Selection	Pair Wise Comparison Selection dialog	The user can select any button on the grid to view the associated comparison of genomes plot. The grid is structured with the rows representing the loaded genomes from one to seven and the columns representing the loaded genomes from one to eight. This means that user can choose the comparison between any genome with another loaded. For example, in the default state, the selection is the comparison between the first and the second genome.

Shift Genome Control Bar

Each Shift Genome Control Bar is bound to a genome and allows the user to shift the genome horizontally to set the position wanted in the window viewer and let compare two different areas of the genomes.











Control	Name	Description	Function
	Shift Genome Control Bar	Shift Genome Control Bars	This control works like a ScrollBar and there is a meaning for the length and the position of the slide bar, it indicates the client the area of the genome which is been showed at the window viewer, while the length of the scroll bar represents the entire genome. When the user shift the bar, he's shifting the genome individually over the window viewer, the other genome isn't shifted and neither the graphic.

MULTIPLE COMPARISON

The Pair Wise Comparison let the user compare graphically few genomes, each one with the next by its MUMs.




The Multiple Comparison Panel

The Multiple Comparison Panel shows the list of the genomes loaded and the order they must be shown for the comparison. This list may be manipulated by the assigned buttons of each genome in the list.

Control	Name	Description	Function
 Archaeoglobus_fulgidus  Aeropyrum_pernix  Haloquadratum_walsbyi  Methanobrevibacter_smith  Caldivirga_maquilingensis  Candidatus_Methanoregul  Haloarcula_marismortui_A  Halobacterium_sp	Pair Wise Comparison Panel	The Pair Wise Comparison Panel	List of genomes that establish the order to be showed and compared.
	Move Up	Move Up Button	Change the order of the list by moving up the concrete genome doing an exchange with the prior genome
	Move Down	Move Down Button	Change the order of the list by moving down the concrete genome doing an exchange with the following genome




Genome View Controls

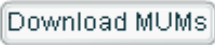

The Genome View control are controls that allow the user to select the number of genomes to be shown at the same time in the Multiple comparison and move up and down the window viewer to see those user wants.

Control	Name	Description	Function
	Num Genome View	Num Genome View Selector	ComboBox where user selects the number of genomes to be shown at the window viewer.
	Move Up	Move Up Button	Clicking will shift the viewer up to see the upper genomes that are not showed at the window viewer
	Move Down	Move Down Button	Clicking will shift the viewer down to see the lower genomes that are not showed at the window viewer

DOWNLOAD MANAGER


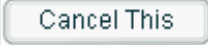

The Download Manager allows the user to select a concrete comparison for downloading the MUM's data file, or download all the MUM's data files bound to the comparison of all genomes loaded.

Control	Name	Description	Function
	Genome1	First genome selection	Used for a simple file download. Selects the first genome of the comparison that the user wants to download.
	Genome2	Second genome selection	Used for a simple file download. Selects the second genome of the comparison that the user wants to download.
	Direction	Direct/Reverse file	Used for a simple file download. Selects what kind of MUMs, the user wants to download, there are three options "Only Directs", "Only Reverses" and "Both", with the last one will download two files.

	Download MUMs	Simple file download	Clicking will start the download with the selected parameters, genomes and direction.
	Download All	Multiple file download	Clicking will start the download of all the MUM's data files.

Queue SubControls

The Queue SubControl is only shown when there are multiple files to be downloaded and lets the user control the file downloading one by one.

Control	Name	Description	Function
	Download Next	Start next download	Download the next file in the queue.
	Cancel This	Cancel this download	Cancel the actual file downloading and find next file in the queue.
	Cancel All	Cancel all downloads in queue	Clicking will cancel all downloads in course.