

Método de secuenciación Sanger

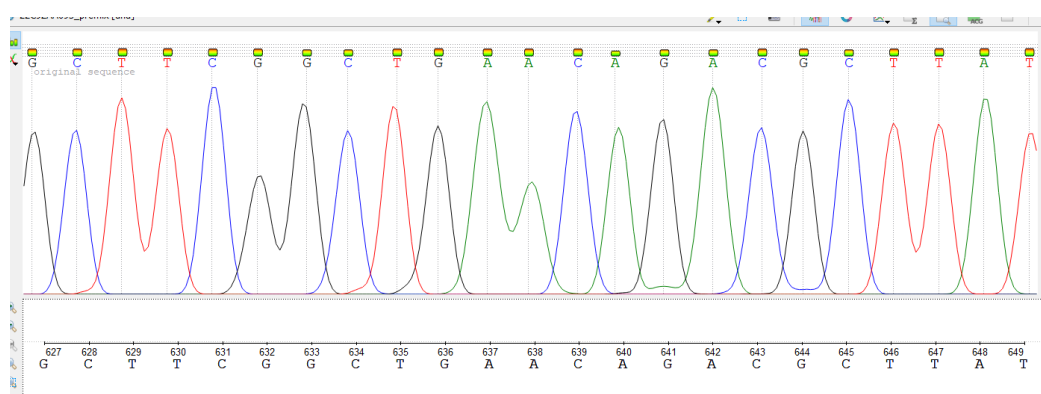
La secuenciación genética es una tecnología que permite conocer y descifrar el código genético que tienen todos los seres vivos. En 1977, el inglés Frederick Sanger desarrolló el método de secuenciación de ADN conocido como método de Sanger basado en la polimerización del ADN y el uso de dideoxinucleótidos como terminadores de la reacción. Para entender el método de secuenciación desarrollado por Sanger es importante recordar que el genoma es el conjunto de genes de un organismo o especie. Estos se encuentran codificados en el ácido desoxirribonucleico (ADN), que está formado por cuatro tipos de subunidades, llamadas bases nitrogenadas. Las bases nitrogenadas del ADN son la adenina, timina, citosina y guanina. Cada secuencia de ADN que codifica una proteína se denomina gen. El ADN es una doble cadena en forma de hélice, a cada lado de la cadena hay dos bases nitrogenadas complementarias: la adenina es complementaria a la timina y la citosina es complementaria a la guanina. Por lo tanto, si sabemos que en un punto de una de las dos cadenas hay una citosina, sabemos que en ese mismo punto de la cadena complementaria hay una guanina. En resumen, la secuenciación de ADN es el proceso que determina la secuencia de bases de los nucleótidos (As, Ts, Cs y Gs) de un fragmento de ADN

El método de secuenciación de Sanger parte de una única hebra de ADN, que ha sido separada previamente de su hebra complementaria. El proceso de síntesis de ADN se realiza por la enzima polimerasa, que es añadida a la hebra de ADN que se quiera secuenciar. La polimerasa, en presencia de una hebra de ADN, una secuencia a partir de la cual inicial la síntesis (llamada “cebador”) y bases nitrogenadas, es capaz de sintetizar la hebra complementaria de ADN. En la actualidad, este proceso se ha refinado utilizando una variante de la PCR en la que se incorporan dideoxinucleótidos (ddNTPs) marcados con fluoróforos, y posteriormente se resuelven los fragmentos resultantes mediante electroforesis capilar de alta precisión.

El mercado de la secuenciación automática basada en el método de Sanger está controlado principalmente por Applied Biosystems. La información obtenida en los secuenciadores Sanger se guarda en ficheros binarios (**.ab1**; **.phd.1**; **.scf**; **.seq** o **.fasta**). Estos ficheros pueden incluir, además del cromatograma procesado, el dato crudo leído por el secuenciador automático, la secuencia de nucleótidos y las calidades.

Cromatogramas:

El cromatograma es el resultado gráfico de la electroforesis capilar. Se observan diferentes picos de colores, donde cada color corresponde a un nucleótido. El conjunto de estos picos ordenados corresponde a la secuencia de un fragmento de ADN. Los secuenciadores de Applied Biosystems generan los datos en formato **.ab1**. Estos ficheros solamente se pueden visualizar con programas especializados.



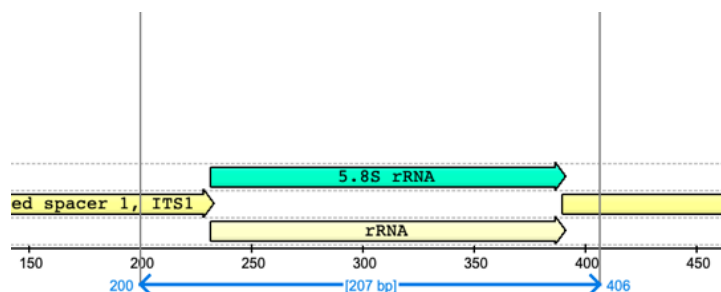
Otro formato en el que se guardan los cromatogramas es el **scf**. es un formato libre no controlado por una empresa y sólo incluye el cromatograma procesado y la secuencia.

Secuencias de nucleótidos

Los secuenciadores también generan las secuencias de cada reacción en formato **.fasta**. Este archivo es un archivo bruto, que no ha sido procesado, por lo que generalmente incluye nucleótidos con mala calidad. Por ello recomendamos que este primer archivo NO SE UTILICE para ningún tipo de análisis.

Nombre	Fecha de modificación	Tamaño	Clase
ELI569_29985693_29985693.ab1	17/1/23, 16:24	330 KB	Applied Biosystems ABIF file
ELI569_29985693_29985693.fasta	17/1/23, 16:24	16 KB	Documento TextEdit.app
ELI569_29985693_29985693.scf	17/1/23, 16:24	145 KB	Standard Chromatogram Format file

Existen otros archivos muy útiles como son los de formato GenBank (.gb). Este formato de archivo se utiliza para almacenar secuencias de ADN y ARN, así como información de anotación asociada, como genes, exones, CDS (secuencias codificantes), regiones regulatorias y otros elementos funcionales.



Ejercicio 1

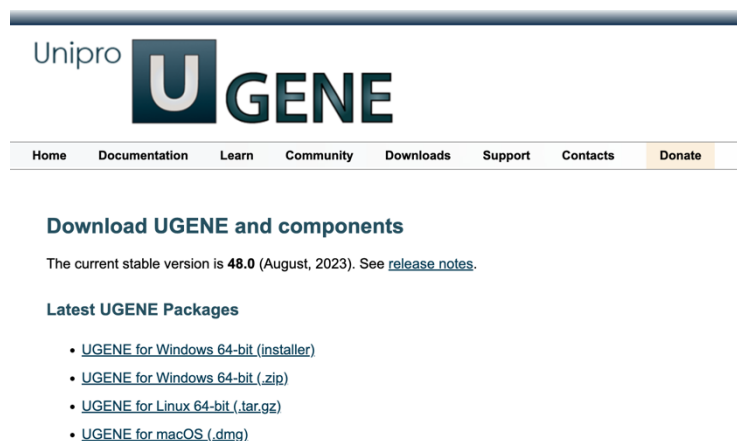
E.1.1 Descarga del software UGENE

Los software bioinformáticos son herramientas computacionales esenciales para analizar y comprender datos biológicos, especialmente de secuenciación de ADN, ARN y proteínas. Estos programas permiten analizar secuencias, ensamblar genomas, predecir estructuras de proteínas y ARN, analizar expresión génica, identificar variantes genéticas, construir árboles filogenéticos y visualizar datos biológicos, entre muchas otras cosas. Algunos ejemplos de softwares o programas conocidos son: MEGA, Vector NTS, SnapGene, Geneious, **UGENE** y muchos más.

Durante el presente curso vamos a trabajar con el *software* bioinformático UGENE, el cual es un programa de acceso libre con una *interface* que nos permitirá analizar de manera sencilla datos de secuencias. UGENE (Unipro UGENE) es un software bioinformático ampliamente utilizado que proporciona una amplia gama de herramientas para el análisis de datos genómicos y biológicos. Es desarrollado por el Centro de Biotecnología de Instrucción e Investigación "Unipro" en Rusia. UGENE se destaca por su enfoque integral, lo que significa que integra varias herramientas y funciones para diferentes tipos de análisis en una sola plataforma.

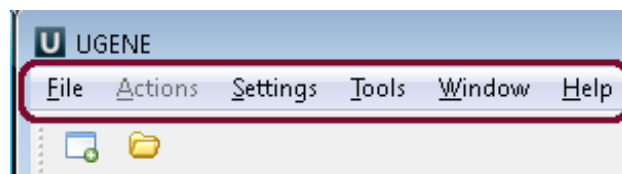
Aquí hay algunas características y capacidades clave de UGENE:

Disponible para sistemas operativos Windows, Linux y MacOS
Interfaz gráfica de usuario amigable
Análisis de secuencias de ADN y proteínas
Ensamblaje de genomas
Análisis de expresión génica
Visualización de datos
Predicción de motivos y estructuras de proteínas y ARN
Automatización y Flujo de Trabajo
Integración de Bases de Datos



Link de acceso para la descarga del software: <http://ugene.net/download-all.html>

¿Qué puedes hacer en UGENE?



Usando estos menús puede ejecutar muchos algoritmos, configurar diferentes configuraciones y obtener ayuda

Tabla 1. Descripción de las opciones del menú del software UGENE

Menú	Descripción
Files	Un conjunto de operaciones a nivel de proyecto/archivo. Ejemplo: crear, abrir, etc. un proyecto; abrir un documento; acceder a la base de datos remota para descargar un archivo.
Actions	Varias acciones asociadas a la ventana activa. Ejemplo: exportar, eliminar, editar, analizar una secuencia usando diferentes complementos (para la vista de secuencia); editar, alinear, cambiar el modo de consenso (para el Editor de alineamiento).
Settings	Configuración de aplicaciones, complementos y herramientas.
Tools	Varias herramientas, independientes de una ventana activa. Este menú se amplía con diferentes complementos. Ejemplo: herramientas HMMER2 / HMMER3, SITECON, <i>Workflow Designer</i> .
Window	Una lista de ventanas activas y manipulaciones básicas con las ventanas. Ejemplo: cerrar todas las ventanas, mosaicos de ventanas, seleccionar la siguiente ventana.
Help	Ayuda de la aplicación y buscar actualizaciones.

Estos menús se pueden completar dinámicamente con nuevas acciones agregadas por *plugins*, que se pueden descargar aquí: <http://ugene.net/download-all.html>

1. Ver, editar y anotar secuencias.

La *vista de secuencia* es una de las principales vistas de objetos en UGENE destinada a visualizar y editar secuencias y anotaciones, cromatogramas, modelos 3D, datos estadísticos, etc. Para cada archivo, UGENE analiza el contenido del archivo y automáticamente abre la vista más adecuada. Para activar la *vista de secuencia*, abra cualquier archivo con al menos una secuencia.

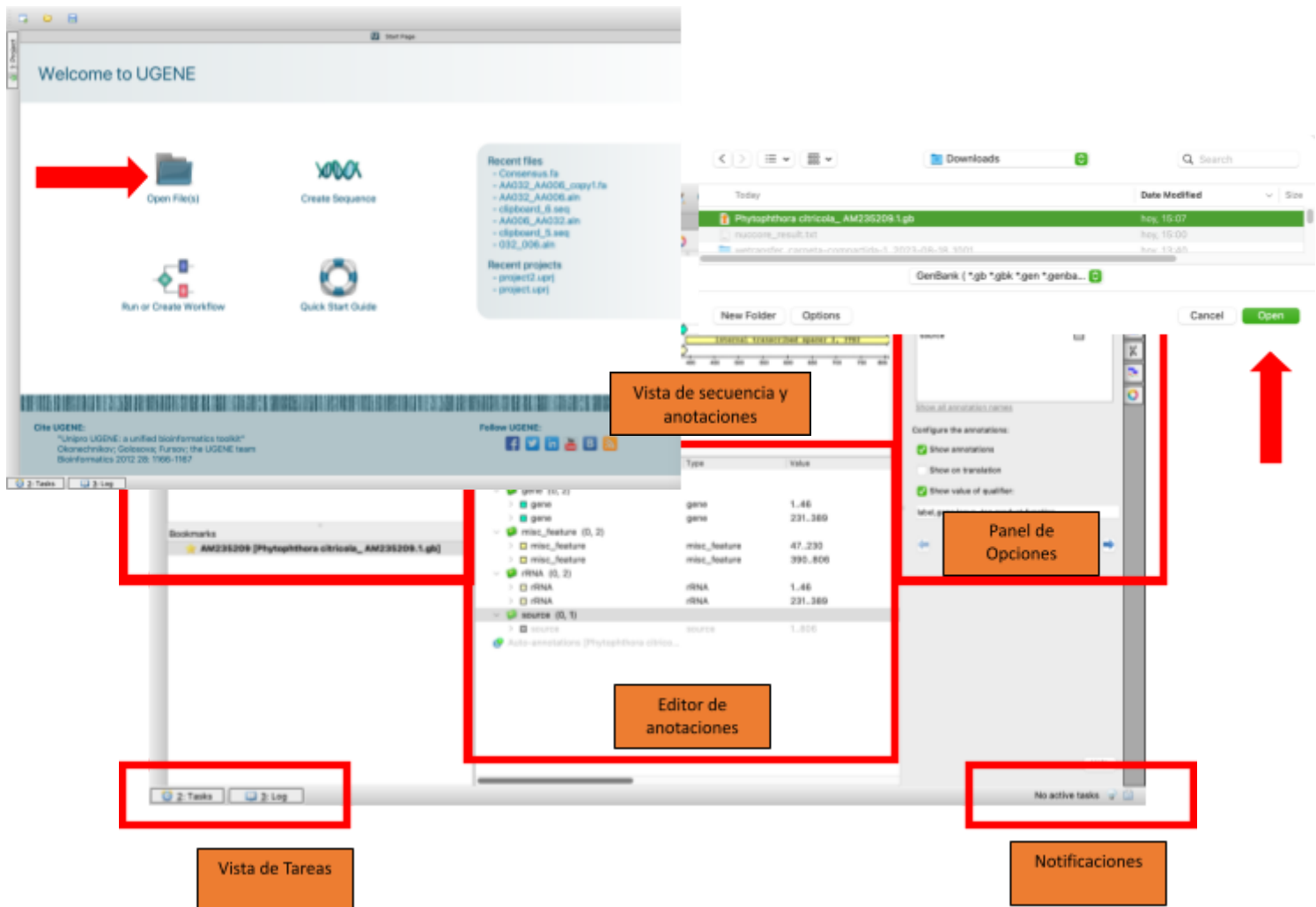
Veamos un ejemplo, puede utilizar el archivo.

Visualización de una secuencia depositada en la base de datos:

1. Descargar secuencia de base de datos: Link de acceso: <https://www.ncbi.nlm.nih.gov/nuccore/AM235209.1>



2. Abrir la secuencia descargada en el software UGENE:
Open File? Seleccionar la secuencia descargada “Phytophthora citricola_AM235209.1.gb”

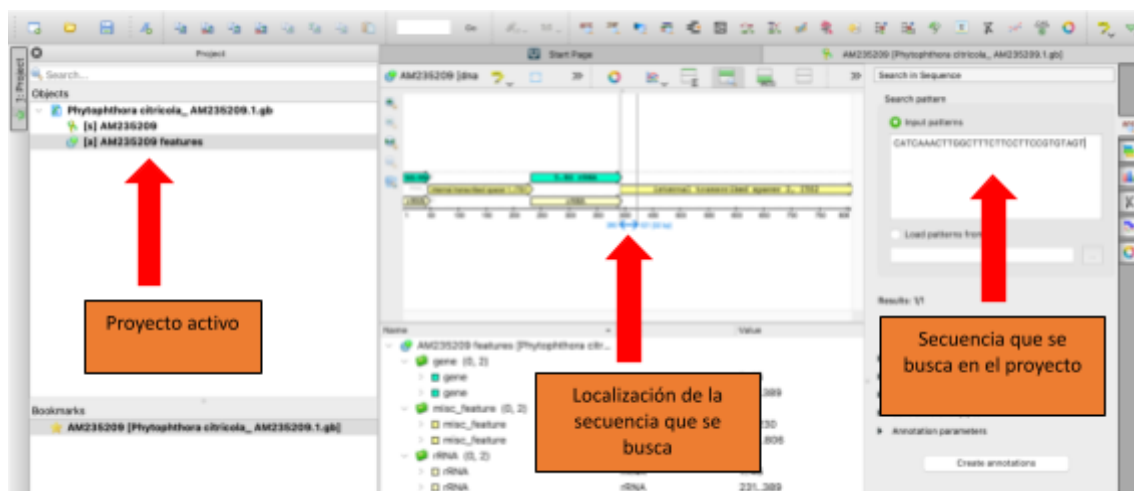


- **Vista del proyecto** : un componente visual que se utiliza para administrar el proyecto activo y los marcadores.
- **Vista de secuencia y anotaciones** : una vista de objeto destinada a visualizar secuencias de ADN, ARN o proteínas junto con sus propiedades, como anotaciones, cromatogramas, modelos 3D, datos estadísticos, etc.
- **Editor de anotaciones** : contiene herramientas para manipular las anotaciones de una secuencia.
- **Panel de opciones**: es el panel con diferentes pestañas de información y pestañas con configuraciones para Vista de secuencia .

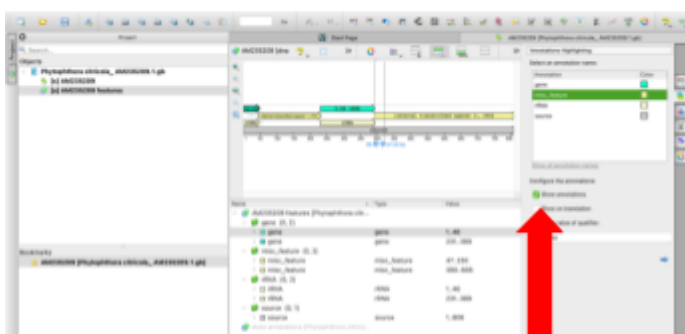
- **Vista de tareas** : muestra tareas activas, por ejemplo, cálculos de algoritmos.
- **Vista de registro** : muestra la información de registro del programa.
- **Notificaciones** : muestra notificaciones para informes de tareas.

3. Uso del panel de opciones

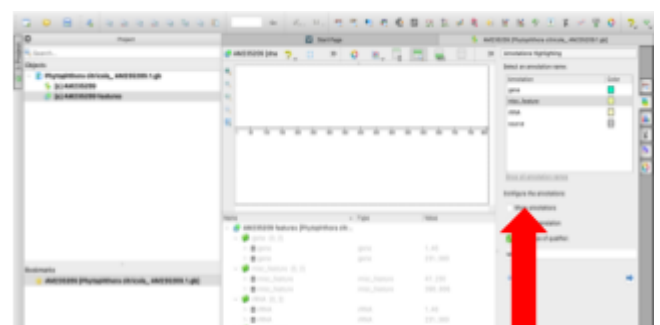
En la pestaña  del panel de opciones tenemos la opción de buscar una secuencia determinada. Por ejemplo: si copiamos la siguiente secuencia: “CATCAAACTTGGCTTTCTCCTTCCGTGTAG” y la pegamos en el recuadro “Input patterns”, el software es capaz de buscarnos esa secuencia en el objeto incluido en el proyecto activo



En la pestaña  del panel de opciones tenemos las opciones referentes a las anotaciones de las secuencias, en el podemos ver de qué tipo de anotaciones de trata (genes, ORFs, Fuente de la secuencia, rRNA, etc.) y también activar o desactivar las éstas anotaciones. De igual forma, en el panel de anotaciones se pueden ver éstas mismas anotaciones.

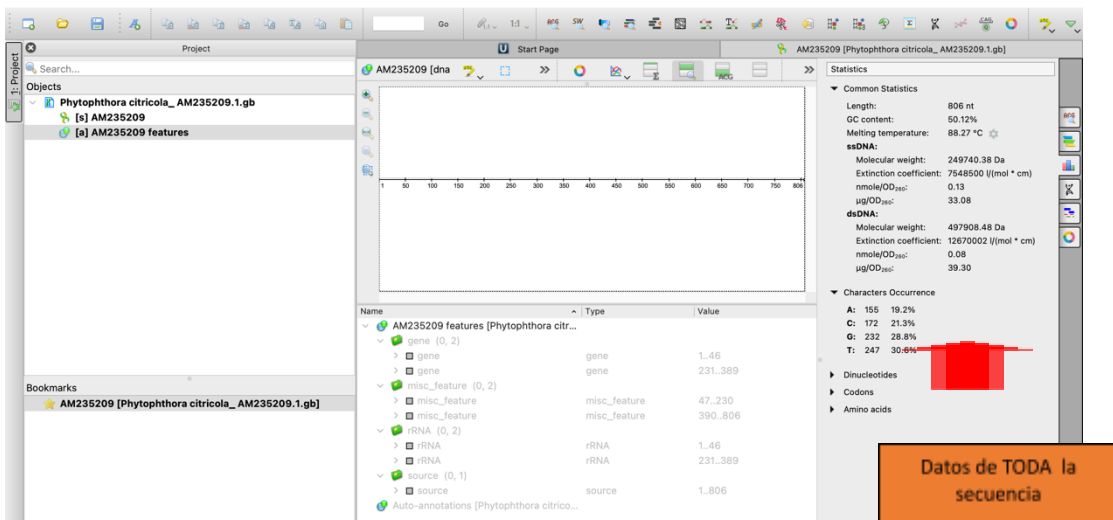


Mostrando todas las anotaciones

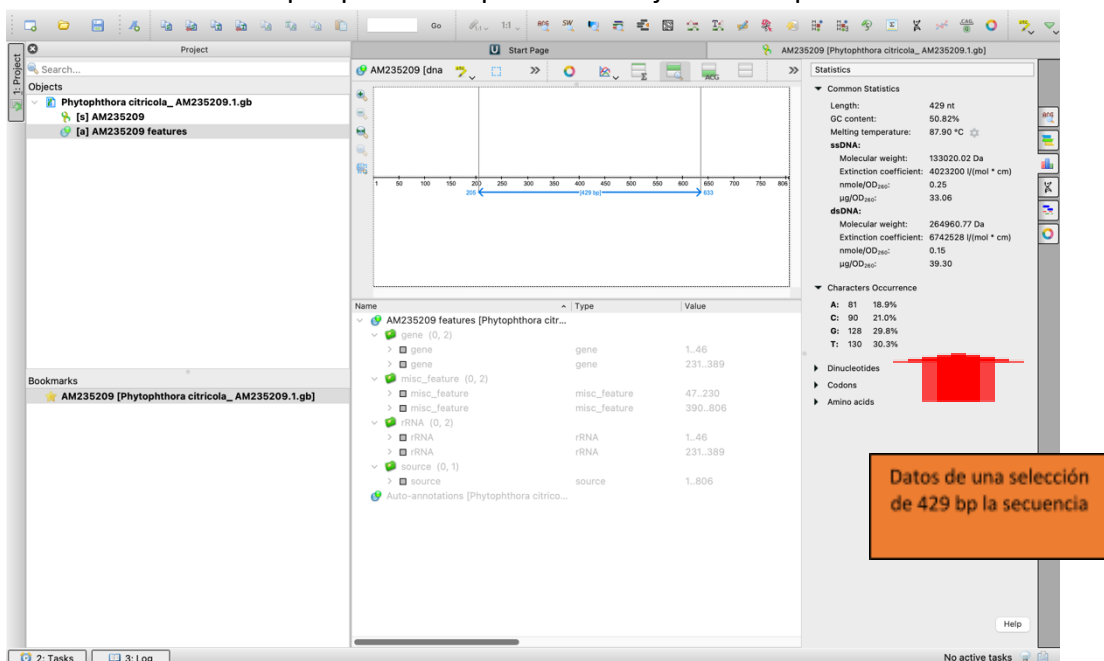


Ocultando todas las anotaciones

En la pestaña  se pueden apreciar los datos de la secuencia como son el tamaño de la secuencia, el contenido de GC, la temperatura de *melting*, y otros datos de interés como puede ser el peso molecular etc etc. Es importante destacar que los datos que aparecen en este apartado hacen referencia siempre a la secuencia que se tenga seleccionada en la vista de secuencia, veamos un ejemplo:



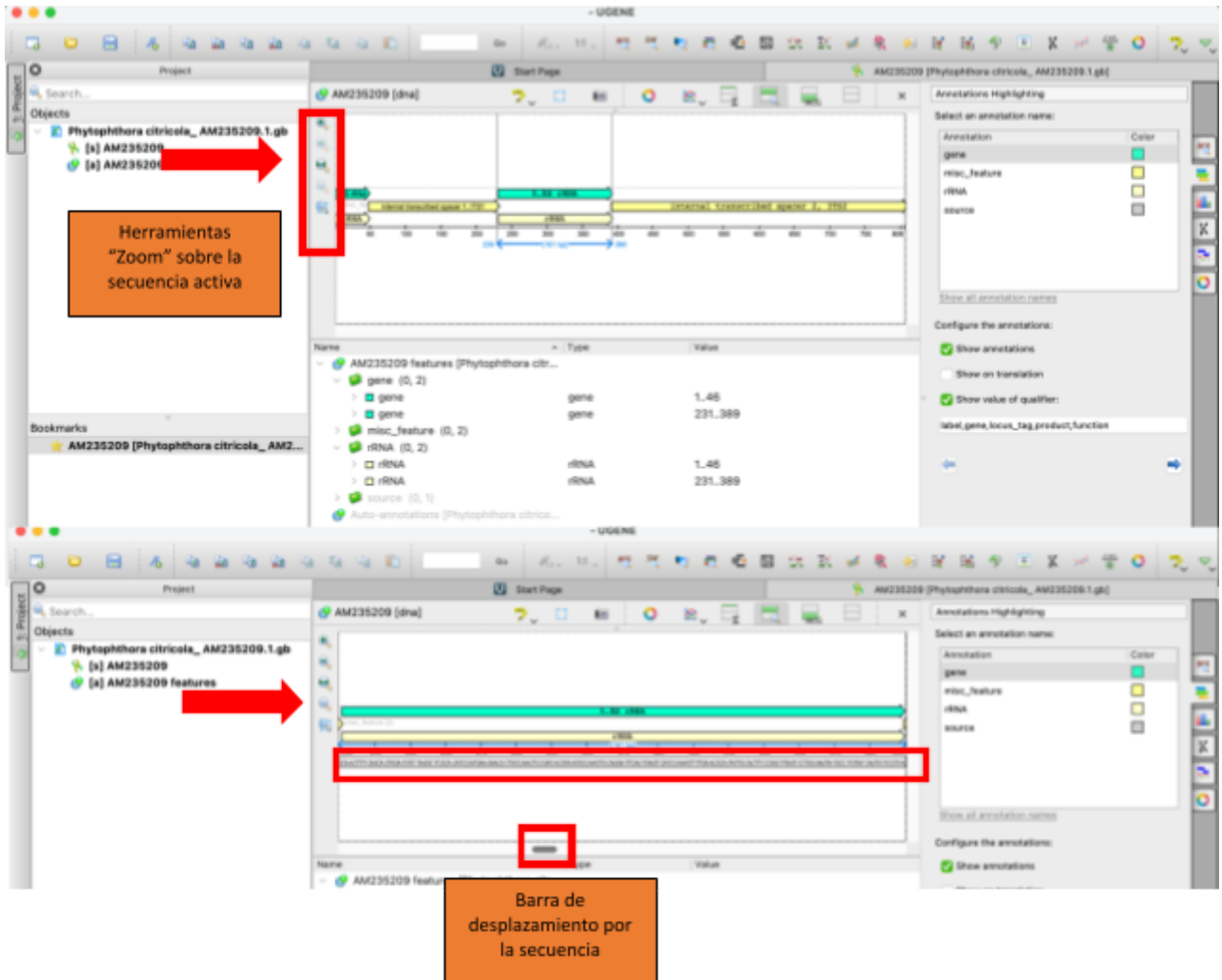
Existen otras pestañas en el panel de opciones que sirven para realizar PCRs *in silico* sobre una secuencia determinada, circularizar la secuencia, entre otros, que no vamos a ver durante este curso porque se escapan de los objetivos del presente curso.



4. Uso del panel de visualización



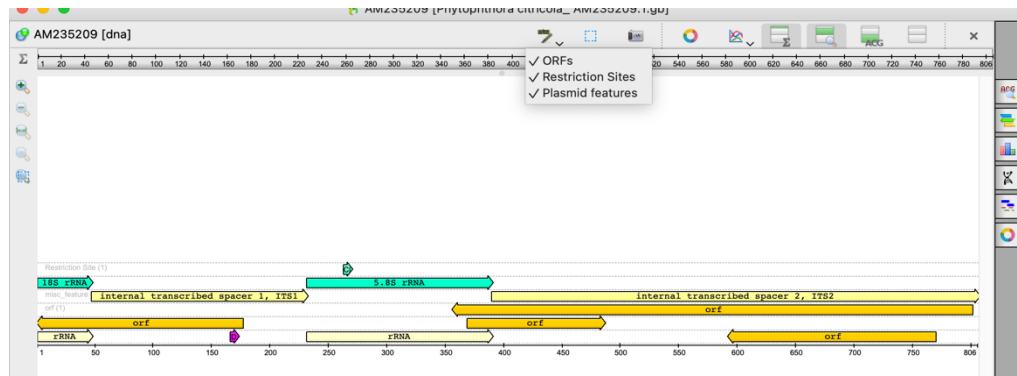
ZOOM: Estas pestañas nos permiten desplazarnos por la secuencia desde un nivel de nucleótidos hasta un nivel más gráfico que incluye toda la secuencia. De tal forma que podemos ver los nucleótidos que componen la secuencia.



ANOTACIÓN AUTOMÁTICA

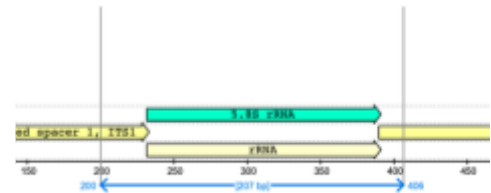
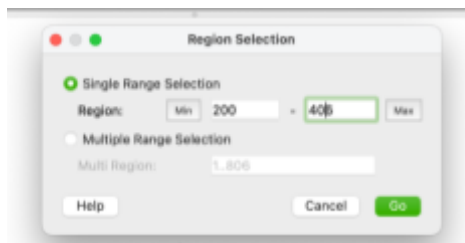
Se pueden anotar open reading frames (ORFs), Sitios de restricción o ciertas características de plásmidos. Esta opción no se suele usar mucho porque se basa en la base de datos del software, la cual de momento es bastante deficiente.

También es importante saber cuándo utilizarla, porque si activamos esta opción en un gen ribosomal, como es el que estamos usando, el software nos anotaría proteínas donde no las hay. Ver ejemplo:



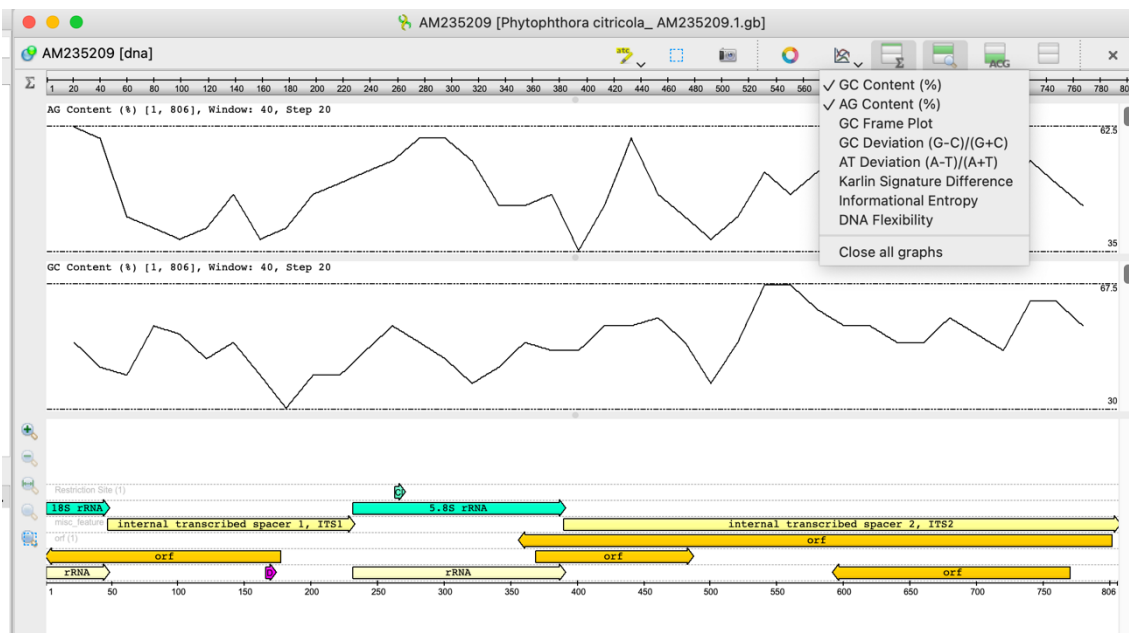
- SELECCIÓN DE UNA REGIÓN CONCRETA

Con esta opción podemos seleccionar un o varios rangos concretos de la secuencia, tan solo debemos introducir las posiciones que lo componen. Ver ejemplo: Selección de un rango comprendido entre la posición 200 y la posición 406



- GRÁFICOS

Con esta opción podemos activar algunos gráficos de interés sobre la secuencia, en relación al % de contenido de nucleótidos (AG; GC), desviaciones (AG; GC), flexibilidad del ADN entre otras.





DETALLES DE LA SECUENCIA

Estas funciones sirven para modificar la visualización de las secuencias o cromatogramas que están abiertos en el proyecto. Por ejemplo la opción  sirve para activar o desactivar la numeración referente a las posiciones de los nucleótidos (figura X.a)

A



CROMATOGRAMA (para poder visualizar el cromatograma se necesita el fichero **.ab1** o **.scf**). Al abrir este fichero se podrá ver la calidad de la secuencia, que será buena si se observan picos diferenciados (pico verde para adenina, pico rojo para timina, pico negro para guanina y pico azul para citosina). Un sistema de color (tipo semáforo) te indica la confianza en la lectura del pico. En la parte inferior se puede observar la posición numérica de nucleótido y la secuencia.

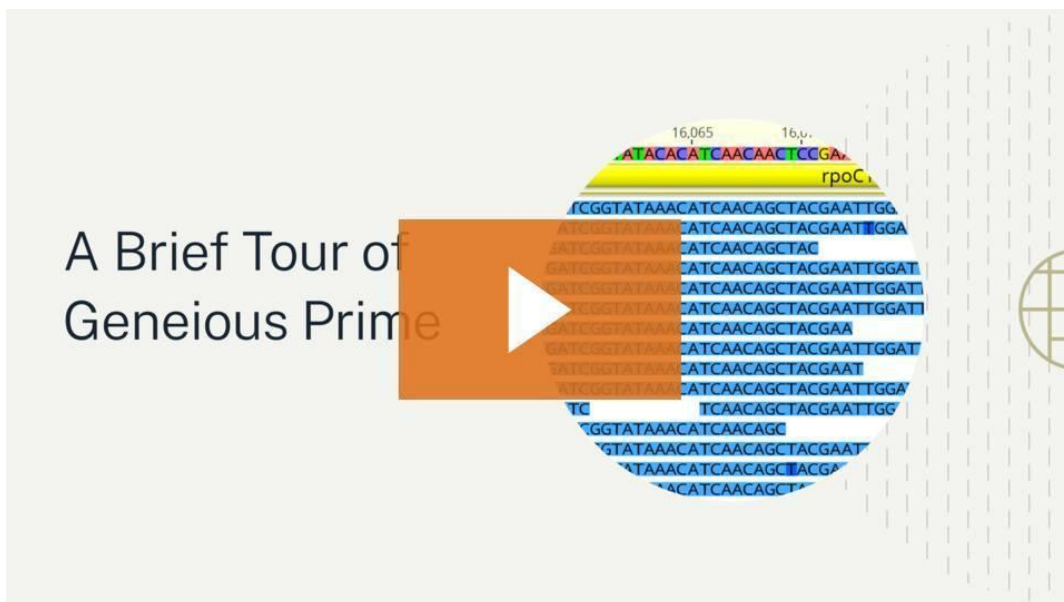


E.1.2 Descarga del software Geneious prime

Como ya hemos mencionado, existen numerosos softwares bioinformáticos que nos permiten trabajar con datos de secuenciación. Otro software que nos permite realizar análisis bioinformático de una manera amigable o “friendly” es Geneious.

Geneious ofrece una variedad de herramientas para tareas como el ensamblaje de secuencias, el análisis de genomas, la anotación de genes, la alineación de secuencias y la visualización de datos. También es compatible con una amplia gama de formatos de archivos, lo que facilita la integración de datos de diferentes fuentes.

Una característica notable de Geneious es su interfaz de usuario intuitiva, que facilita la navegación y el uso de sus diversas funciones. Además, el software se actualiza regularmente para incorporar nuevas herramientas y mantenerse al día con los avances en la investigación genómica.



[Features | Geneious Prime](#)

Link de acceso para la descarga del software 
<https://www.geneious.com/download/previous-versions/#geneious-prime-2022>

Durante el presente curso, **utilizaremos la versión Geneious Prime 2022 de prueba de 14-30 días**, por lo que **rogamos a los participantes que se lo descarguen 1-2 días antes de la sesión 2**

¿Qué puedes hacer en Geneious?

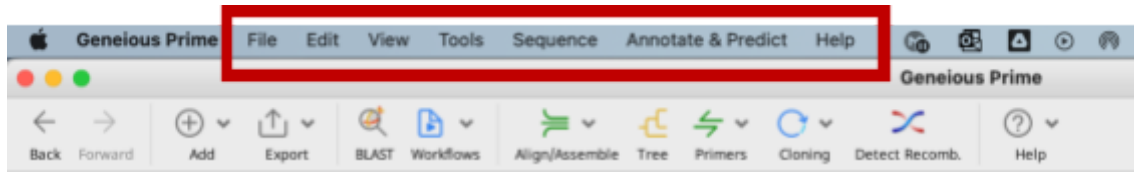
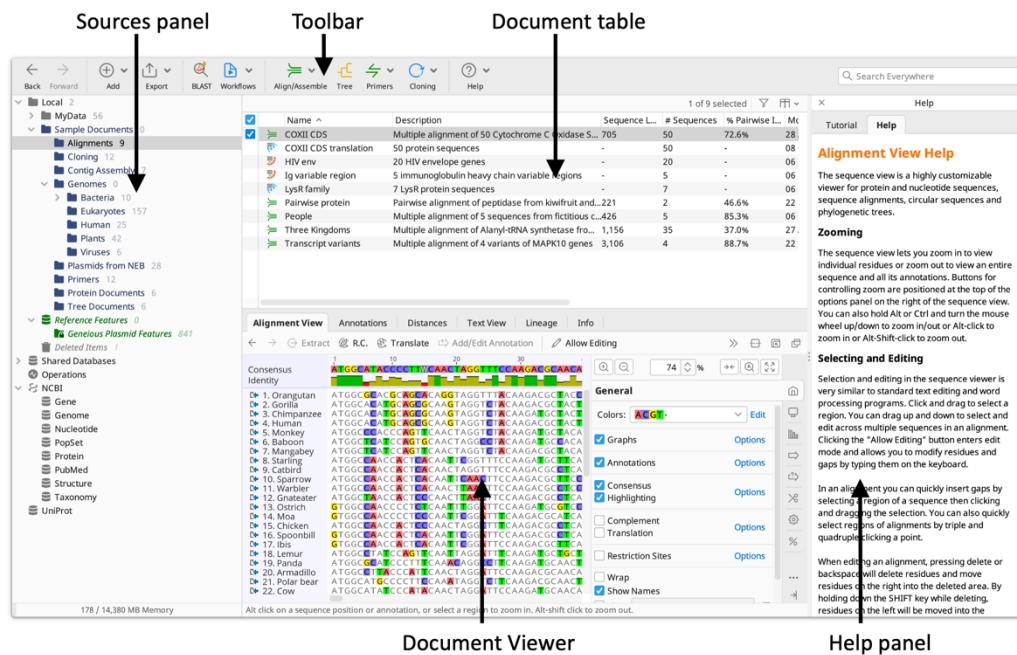


Tabla 1. Descripción de las opciones del menú del software Geneious Prime

Menú	Descripción
Files	Un conjunto de operaciones a nivel de carpetas. Ejemplo: crear una nueva carpeta, secuencia, grupo de enzimas, o un simple texto. También para eliminar, mover, guardar, exportar o importar archivos. Éstas ultima acciones requieren de la selección previa del archivo en cuestión.
Edit	Varias acciones asociadas a la ventana activa. Ejemplo:
View	Configuración de aplicaciones, complementos y herramientas.
Tools	Herramientas para el análisis de los archivos importados, ejemplo: alineamientos, árboles filogenéticos, diseño de cebadores, etc.
Sequence	Análisis para la manipulación y creación de secuencias nucleotídicas, por ejemplo: introducir manualmente una secuencia, invertir la dirección de la secuencia, convertir DNA en RNA...
Annotate & Predict	Nos permite la introducción de metadatos a secuencias o archivos con los que se está trabajando: anotar regiones de mala calidad, anotar pautas de lectura abierta, búsqueda de repeticiones en la secuencia, búsqueda de factores de transcripción, etc.
Help	Acceso al manual de usuario del software, acceso al soporte web e información relacionada con la licencia del software.

La ventana principal de Geneious tiene cinco áreas importantes o 'paneles'.

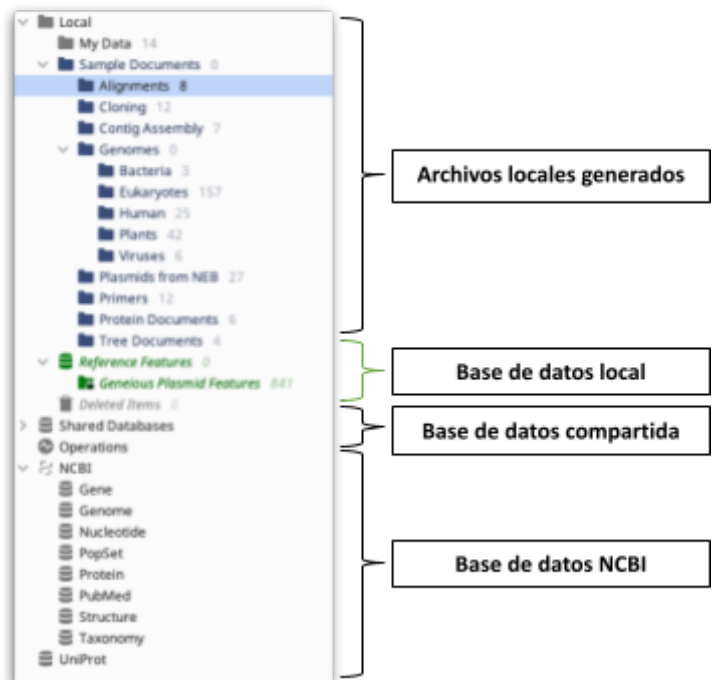


1. Panel de fuentes

El Panel de Fuentes muestra tus documentos almacenados y contiene los servicios que Geneious Prime ofrece para almacenar y recuperar datos. El símbolo de la flecha (>) indica que una carpeta contiene subcarpetas. Haz clic en este símbolo para expandir o contraer carpetas.

El Panel de Fuentes te permite acceder a:

- Tus documentos locales, incluidos los elementos eliminados.
- La base de datos de características de plásmidos de Geneious.
- Las bases de datos del NCBI: Gene, Genome, Nucleotide, PopSet, Protein, Pubmed, SNP, Structure y Taxonomy.
- La base de datos Uniprot (EMBL).
- Bases de datos compartidas, si están configuradas.



2. Tabla de documentos

La Tabla de Documentos muestra resúmenes de cada documento en una carpeta o carpetas seleccionadas, presentados en forma de tabla. Una carpeta local puede contener una mezcla de documentos, como secuencias de ADN, secuencias de proteínas, artículos de revistas, alineaciones de secuencias, árboles filogenéticos, etc. Cada tipo de documento contenido en la tabla se muestra con diferentes iconos. Los documentos no leídos, como los generados recientemente, estarán en negrita y tendrán un "•" en una columna adicional (U).

Documento no
leído

Secuencia .fasta

Cromatograma .ab1

	Name ^	Description	Sequence L...	# Sequences
•	COI Assembly	2 reads from COI assembled using Geneious	758	2
•	COI Assembly 2	2 reads from COI assembled using Geneious	726	-
	COI Assembly Report	-	-	-
	COI fwd read	Forward Sanger read of COI gene	724	-
	COI rev read	Reverse Sanger read of COI gene	721	-
	COXII CDS	Multiple alignment of 50 Cytochrome C Oxidase Subunit II genes	705	50
	Gel Image	An image	-	-
	HIV	Human immunodeficiency virus 1, complete genome	9,181	-
	Ig variable region	5 immunoglobulin heavy chain variable regions	-	5
	M13-F (-20)	M13 forward sequencing primer, 20bp upstream	17	-
	M13-R (-26)	M13 reverse sequencing primer, 26bp upstream	17	-
	Moa translation	Dinornis giganteus mitochondrion, complete genome	228	-
	Notes	-	-	-
	pUC19	Cloning vector pUC19, complete sequence	2,686	-
	Three Kingdoms consensus tree	Neighbor-joining consensus tree Three Kingdoms	1,156	35

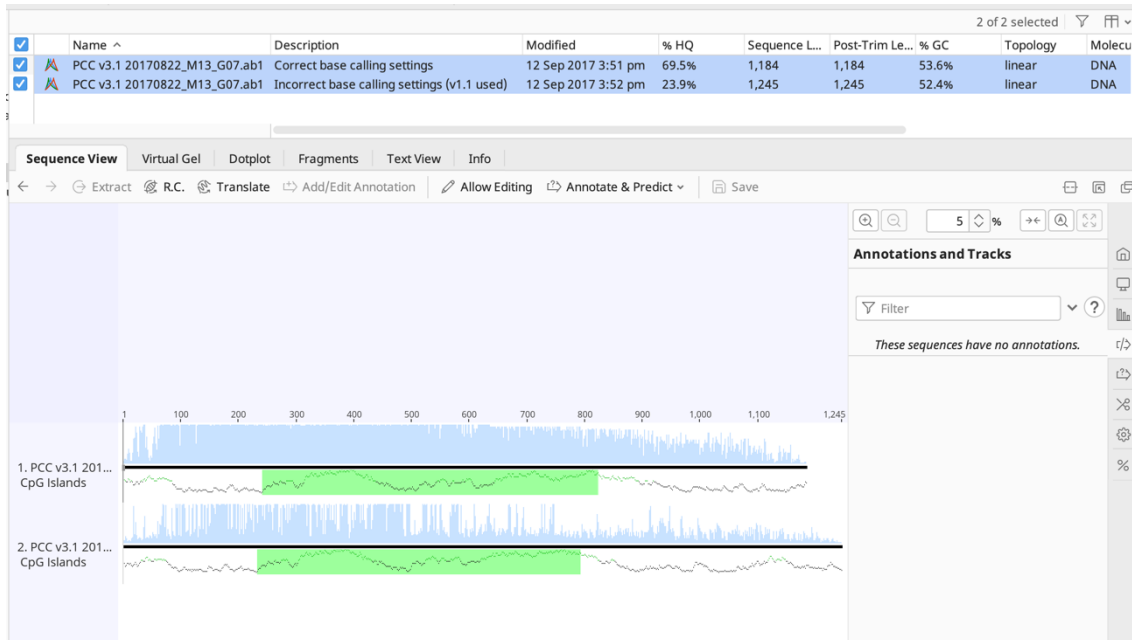
Iconos

- Nucleotide sequence
- Nucleotide sequence list
- Chromatogram (Sanger)
- Circular sequence (nucleotide)
- Oligonucleotide
- Amino acid sequence
- Protein list sequence
- BLAST nucleotide hit
- BLAST protein hit

- Contig (nucleotide assembly)
- Alignment
- Tree
- 3D structure
- Restriction enzyme collection
- Publication
- Image
- PDF
- Text Document

Seleccionar un documento en la Tabla de Documentos mostrará sus detalles en el Panel de Vista de Documentos. Seleccionar varios documentos mostrará una vista de todos los documentos seleccionados si son de tipos similares, por ejemplo, seleccionar dos secuencias mostrará ambas en la vista de secuencias.

La forma más sencilla de seleccionar varios documentos es haciendo clic en las casillas de verificación en el lado izquierdo de la tabla. También se pueden utilizar los controles estándar del teclado (Mayús y Ctrl/clic en Mac OS X).



Características de la Tabla de Documentos

Edición

Se pueden introducir valores en las columnas de la tabla. Esta es una forma útil de editar la información en un documento. Para editar un valor en particular, primero haz clic en el documento y luego haz clic en la columna que deseas editar. Ingresa la nueva información correspondiente y presiona enter. Sin embargo, no se pueden editar ciertas columnas, como el número de acceso de NCBI.

Copia

Los valores de las columnas se pueden copiar. Este es un método rápido para extraer información que se puede buscar, como un número de acceso. Para copiar un valor, haz clic derecho (Ctrl+clic en Mac OS X) sobre él y elige la opción "Copy name", donde "name" es el nombre de la columna.

Ordenar

Todas las columnas se pueden ordenar alfabética, numérica o cronológicamente, según el tipo de datos. Para ordenar por una columna dada, haz clic en su encabezado. Si tienes diferentes tipos de documentos en la misma carpeta, haz clic en la columna "Icono" para ordenarlos según su tipo.

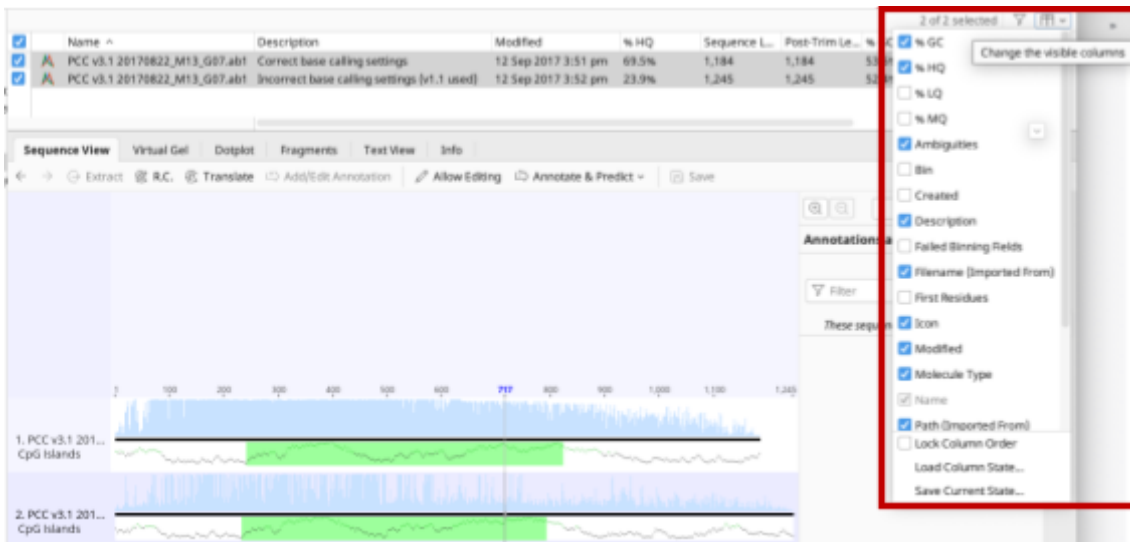
Exportar

El contenido de la tabla de documentos se puede exportar en formato .csv o .tsv, lo que te permite exportar secuencias y sus estadísticas para un conjunto de documentos. Para obtener más información, consulta Exportar a archivos CSV o TSV.

Gestión de columnas

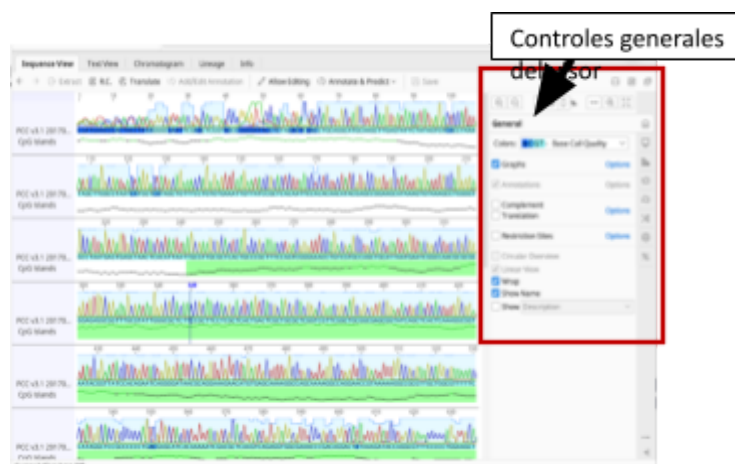
Puedes reorganizar las columnas a tu gusto. Haz clic en el encabezado de la columna y arrástrala a la posición horizontal deseada.

También puedes elegir qué columnas quieres que sean visibles haciendo clic derecho (Ctrl+clic en Mac OS X) en cualquier encabezado de columna o haciendo clic en el pequeño botón de encabezado sobre la esquina superior derecha de la tabla. Esto mostrará un menú emergente con una lista de todas las columnas disponibles.



3. Panel de Visualización

El Panel de Visualización de Documentos muestra el contenido de cualquier documento al que se haya hecho clic en la Tabla de Documentos, permitiéndote ver secuencias, alineaciones, árboles, estructuras 3D, resúmenes de artículos de revistas y otros tipos de documentos en una vista gráfica o de texto plano.



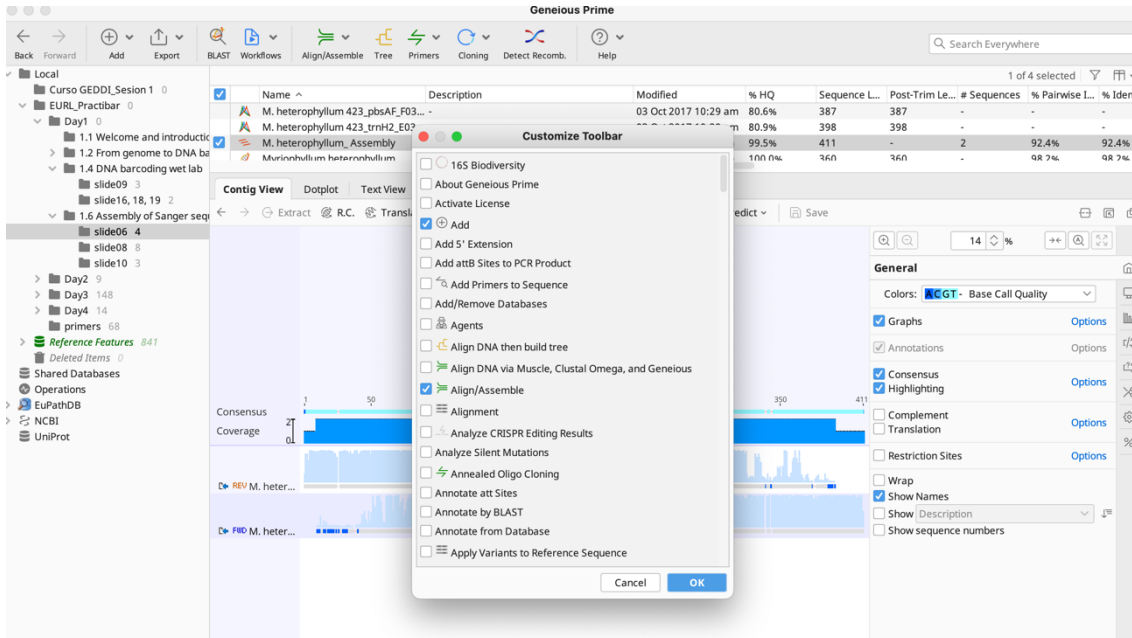
Visualización de un cromatograma

4. La Barra de Herramientas

La barra de herramientas contiene botones que ofrecen accesos directos a funciones comunes en Geneious Prime, incluyendo BLAST, Flujos de Trabajo, Alinear,

Construcción de **Árboles**, **Diseño de Cebadores** y Clonación. La barra de herramientas también proporciona accesos directos para Importar y Exportar archivos, así como para acceder a las opciones del menú de Ayuda.

Las opciones Atrás y Adelante te ayudan a desplazarte entre vistas anteriores y son análogas a los botones de retroceso y avance en un navegador web. Para retroceder y avanzar entre carpetas a las que has accedido recientemente, utiliza Shift+clíc.



La barra de herramientas se puede personalizar haciendo clic derecho en ella (Ctrl+clíc en Windows/Linux, comando + clic en Mac OS X). Esto mostrará un menú emergente con las siguientes opciones:

Mostrar Etiquetas: Activa o desactiva las etiquetas de texto.

Iconos Grandes: Cambia entre iconos grandes y pequeños.

Personalizar: Muestra todas las funciones que se pueden agregar a la barra de herramientas.

Seleccionar/deseleccionar botones mostrará/ocultará los botones en la barra de herramientas.

5. El Panel de Ayuda

El Panel de Ayuda está desactivado por defecto y se puede abrir haciendo clic en el icono de Ayuda en la barra de herramientas y seleccionando Ayuda Rápida. El Panel de Ayuda tiene una pestaña de Ayuda y una pestaña de Tutorial.