

sonificADN

*Recurso informático para la generación automática de secuencias
sonoras a partir de secuencias genéticas*



Javier Forment

First system of musical notation. The top staff is a treble clef. The bottom staff is a bass clef. The notes are: Treble: G, A, T, A, A, A, T, C, T, G, G, T, C, T, T, A, T, T, T, C, C. Bass: G, G, A, T, A, A, A, C, G, C, T, A, A, C, T, T, A, A, A, C.

Second system of musical notation. The top staff is a treble clef. The bottom staff is a bass clef. The notes are: Treble: C, T, G, G, T, C, T, T, A, T, T, T, C, C, G, A, T, A, A, A, T. Bass: A, A, A, C, G, C, T, A, A, C, T, T, A, A, -, A, C.

Third system of musical notation. The top staff is a treble clef. The bottom staff is a bass clef. The notes are: Treble: G, C, T, A, A, C, G, C, T, A, C, C, G, G, T, C, T, T, A, A. Bass: G, A, T, A, A, A, T, C, T, A, A, A, C, A, T, T, T, A, C.

Fourth system of musical notation. The top staff is a treble clef. The bottom staff is a bass clef. The notes are: Treble: G, G, A, T, A, A, A, C, G, C, T, A, -, T, T, A, A, -. Bass: C, T, G, G, T, C, T, T, A, T, T, T, C, C, G, A, T, A, A, A, T.

Guía didáctica

sonificADN

*Recurso informático para la generación automática de secuencias
sonoras a partir de secuencias genéticas*



Proyecto Daniel Carasso Fellowship “El reto de la digitalización de semillas:
sostenibilidad, big data y el movimiento social por sistemas de semillas de código abierto”
dirigido por Raquel Ajates (UNED)

Comisariado 1:

Marco Ranieri. Banco Móvil de Semillas

Comisariado 2:

Aina Socies, Lluís Vidaña, Nívola Uyà y Enric Socias. La Danza de las Semillas

Comisariado 3:

Javier Forment. Semillas Sonoras

Con la colaboración de PLANEA Red de arte y escuela



Los contenidos de esta publicación se pueden distribuir, copiar y remezclar
citando la fuente, sin usos comerciales y manteniendo esta licencia



Índice

*sonificADN. Recurso informático para la generación automática de
secuencias sonoras a partir de secuencias genéticas*



1

Resumen

2

Introducción

3

Antecedentes

4

Programa informático

5

Guía de instalación

6

Guía de uso

1 Resumen

¿A qué suena un gen? ¿En qué se parece un gen a una partitura? ¿Puede un gen crear una canción? El presente recurso ofrece a alumnado y docentes la posibilidad de asentar e interiorizar tanto el concepto de secuencia genética como el concepto de pieza sonora a partir del establecimiento de un paralelismo entre una secuencia de nucleótidos (un gen) y una secuencia de sonidos (una canción) mediante un programa informático. Dada la multidisciplinariedad del recurso, que combina conceptos de genética, música e informática, se propone su uso mediante la colaboración entre docentes de las áreas de Biología, Arte y Tecnología.



Tiempo: 2 horas

Lugar: Aula con ordenadores con Linux

sonificADN es un recurso educativo que forma parte de una obra artística encuadrada en el marco del proyecto de investigación Daniel Carasso Fellowship “El reto de la digitalización de semillas: sostenibilidad, big data y el movimiento social por sistemas de semillas de código abierto” dirigido por Raquel Ajates (UNED). El objetivo principal del proyecto era analizar la viabilidad de modelos alternativos de gobernanza de semillas, explorando cómo el concepto de bien común y los movimientos de código abierto digital pueden ayudar a proteger las semillas de la creciente pérdida de variedades y su privatización. Es esencial reconocer el rol clave de las semillas para mantener los derechos de las personas productoras, la sostenibilidad del sistema alimentario, y la diversidad cultivada en nuestros campos y platos. El proyecto incluyó tres comisariados artísticos que dieron lugar a la creación de tres obras que exponen con inspiración, creatividad y gran pericia, las problemáticas sociales y medioambientales a las que se enfrentan las semillas y la diversidad cultivada, invitándonos a reflexionar sobre su belleza e importancia para la salud humana y planetaria, y llamando a la acción para conservarlas. Se trata de las obras «Banco móvil de semillas» de Marco Ranieri, «La danza de las semillas. La germinación del conocimiento» de la Associació Varietats Locals, y «Semillas Sonoras» de Javier Forment. Cada obra ha generado un recurso educativo interdisciplinar con material multimedia para que su impacto vaya más allá de las exposiciones y llegue a más audiencias.”

Introducción

Los genes constituyen el material hereditario de un organismo, y desempeñan un papel clave en la determinación de su aspecto y la forma en que funciona. El ADN de cada gen contiene las instrucciones específicas (como si se tratara de una receta de cocina) para fabricar una proteína distinta, que son el componente fundamental de la maquinaria celular de todo tipo de seres vivos. En el cuerpo humano, cada célula contiene aproximadamente 20,000 genes. La edad y la exposición a factores ambientales, como la alimentación, ejercicio, medicamentos y sustancias químicas, pueden activar o desactivar genes sin cambiar la secuencia del ADN. El estudio de estos cambios que no alteran el ADN se llama epigenética, una ciencia que nos ayuda a crear un puente de entendimiento entre los genes, el ambiente en que vivimos y nuestros hábitos.

Un gen es una larga cadena formada por numerosas moléculas unidas una a continuación de otra. También puede pensarse en un gen como un largo collar formado por la unión de numerosas perlas. Al igual que un collar podría tener perlas de, por ejemplo, cuatro colores distintos, formando un collar multicolor, el gen está formado por la unión de 4 moléculas distintas, conocidas como nucleótidos o bases: adenina (A), timina (T), citosina (C) y guanina (G). De este modo, los genes se diferencian unos de otros, no por sus componentes, que son siempre esas cuatro bases, sino por el orden en el que se encadenan, del mismo modo que dos collares formados por exactamente las mismas perlas de 4 colores se diferencian no por las perlas que contienen, sino por el orden en que éstas se insertan en el collar.

Por ejemplo, un gen podría tener la secuencia AATCGAGTCATC... y otro gen podría tener la secuencia ATAATATCGCGC... Ambos están formados por adeninas (A), timinas (T), guaninas (G) y citocinas (C), pero en un orden distinto. Serían genes distintos que tendrían funciones distintas en el organismo.

Del mismo modo, una canción está compuesta por una secuencia de sonidos representados por notas. Existen tan sólo 12 notas distintas en la escala cromática, la escala musical occidental a la que estamos acostumbrados: Do, Do#, Re, Re#, Mi, Fa, Fa#, Sol, Sol#, La, La# y Si (o, en la representación alternativa de origen anglosajón, C, C#, D, D#, E, F, F#, G, G#,

A, A# y B). Y sin embargo, a pesar de ser tan sólo 12 notas, se pueden producir prácticamente infinitas melodías, que producen en el oyente distintas emociones, gracias a encadenar dichas notas en secuencias diferentes.

Por ejemplo, una melodía podría ser Re-Sol-La-La-Re-Re (3 Res, 2 Las y 1 Sol) y otra podría ser Sol-La-Re-La-Re-Re (también 3 Res, 2 Las y 1 Sol, pero en un orden distinto). Son melodías distintas que pueden causarnos distintas impresiones al escucharlas.

Para poner de manifiesto dicho paralelismo entre las secuencias genéticas y las secuencias sonoras, el presente recurso consiste en un programa informático que crea secuencias sonoras a partir de secuencias genéticas. En concreto, el programa toma una secuencia genética proporcionada por el usuario, y cuatro o más sonidos también proporcionados por el usuario y genera una secuencia sonora formada por los sonidos proporcionados dispuestos en el mismo orden en el que están dispuestos los nucleótidos en la secuencia genética proporcionada.

A modo de ejemplo, si al programa le proporcionamos la secuencia genética AATCGG y cuatro sonidos de piano con las notas Do, Re, Sol y La, el programa asignará, por ejemplo, la nota Do a la base C, la nota Sol a la base G, la nota La a la base A y la nota Re a la base T, y generará la melodía de piano formada por la secuencia de las notas La-La-Re-Do-Sol-Sol. En la guía de uso del recurso se indica el modo de asignar las distintas notas a cada una de las cuatro bases A, T, C y G.

O, si al programa le proporcionamos la secuencia genética TATCTG y cuatro sonidos de percusión (por ejemplo bombo, caja, platillo y timbal), el programa asignará, por ejemplo, la caja a la base C, el bombo a la base T, el platillo a la base G y el timbal a la base A y generará el ritmo de percusión formada por la secuencia bombo-timbal-bombo-caja-bombo-platillo. En la guía de uso del recurso se indica el modo de asignar los distintos instrumentos a cada una de las cuatro bases A, T, C y G.

En definitiva, el programa hace que la secuencia genética proporcionada se convierta en una partitura que será ejecutada por el instrumento formado por los sonidos que le proporcionemos. Ello establece otro paralelismo entre la genética y la música, ya que, del mismo modo que una partitura guía a los músicos para producir una determinada obra musical y no otra, los genes guían a las células para producir un determinado organismo y no otro.

En concreto, el recurso consiste en:

1

Un programa informático de código abierto para la generación automática de secuencias sonoras a partir de secuencias genéticas.

El programa consiste en un script Python (pequeño programa formado por un único fichero de texto con órdenes en el lenguaje de programación Python) que toma una secuencia genética y una serie de sonidos y crea una secuencia sonora consistente en una sucesión de dichos sonidos determinada por la sucesión de nucleótidos (A, T, C, G) en la secuencia genética.

2

Una guía de instalación de dicho programa en un ordenador con sistema operativo de código abierto (cualquier versión de Linux).

La guía de instalación detalla el proceso de instalación del script Python en un ordenador con Linux, así como todos los requerimientos necesarios para su correcto funcionamiento.

3

Una guía de uso del programa informático, incluyendo el modo de obtener secuencias genéticas y el modo de obtener sonidos para que el programa los combine.

La guía de uso detalla (1) el modo de obtener secuencias genéticas de distintos organismos (por ejemplo, semillas y animales), (2) el modo de obtener sonidos, (3) dónde deben estar situados los ficheros con las secuencias genéticas y los sonidos que se deseen usar, (4) el modo de asociar los distintos sonidos a los distintos nucleótidos (A, T, C, G) de la secuencia genética, y (5) el modo de correr el programa para obtener un fichero .mp3 con la correspondiente secuencia sonora.

Antecedentes

Las piezas sonoras resultantes se enmarcan en la corriente de creación musical conocida con los nombres de “música generativa”, “música procedural”, “música algorítmica”, “música aleatoria”, “serialismo”, o similares y cuenta con representantes como John Cage, Arvo Pärt, Karlheinz Stockhausen, Edgard Varèse, Iannis Xenakis, Pierre Boulez, Charles Ives, Henry Cowell, Earle Brown, Cornelius Cardew, Elliot Carter, Alvin Lucier, Terry Riley, Steve Reich, Joseph Schillinger o Lejaren Hiller. En este tipo de música, la pieza se genera automáticamente a partir de una serie de elementos musicales iniciales y una serie de reglas y procesos formales pre-establecidos. En esta forma de creación, la persona compositora está más interesada en definir dichas condiciones y dejar que la pieza se genere a partir de ellas sin su intervención, que en crear una pieza musical de principio a fin de acuerdo con unos criterios estéticos. El objetivo es, pues, el proceso más que el resultado, con lo que las piezas resultantes pueden apartarse en mayor o menor grado de la estética musical habitual.

Ejemplos:

- **Music of Changes, de John Cage**, basado en permutaciones obtenidas a partir del “Libro de las Mutaciones” (I Ching):
<https://sulponticello.com/iii-epoca/john-cage-y-el-i-ching>
https://www.youtube.com/watch?v=B_8-B2rNw7s
- **Le Marteau Sans Maître, de Pierre Boulez**, destacado representante del estilo musical conocido como serialismo:
<https://www.youtube.com/watch?v=4uxwxOdb8PO>
- **Studie I y Studie II, de Karlheinz Stockhausen**, ambos con un planteamiento compositivo basado en procesos de serialización de parámetros físicos y en proporciones derivadas de operaciones matemáticas, pero con resultados sonoros diferentes:
<https://www.youtube.com/watch?v=H4QaMwpVXVM>
<https://www.youtube.com/watch?v=bwj6ZptPnDo>
- **Metastaseis, de Iannis Xenakis**, compuesta a partir de las progresiones geométricas arquitectónicas de Le Corbusier o la progresión matemática de Fibonacci:
<https://es.laphil.com/musicdb/pieces/2339/metastaseis>
<https://www.youtube.com/watch?v=SZazYFchLRI>
- **Tabula Rasa, de Arvo Pärt**, destacado exponente de la conocida como música generativa:
<https://www.youtube.com/watch?v=8HON4AswPVk>

Programa informático

El programa se puede descargar en un fichero .zip desde la URL [_](#). Dicho .zip contiene tanto el script Python en que consiste el programa (el fichero 'sonificADN.py'), como su fichero de configuración (el fichero 'config.txt'), una secuencia genética de ejemplo (el fichero 'seq.fasta' que hay en el directorio 'seqs') y una serie de sonidos de ejemplo para usar en la sonificación (los ficheros que hay en los directorios '0.perc', '1.inst' y '2.voz' que hay dentro del directorio 'sonidos'). Con dichos ficheros de ejemplo de secuencia genética y sonidos de percusión, un instrumento y unas voces se puede usar directamente el programa para sonificar dicha secuencia genética utilizando dichos sonidos. Abajo se describe cómo obtener otras secuencias genéticas y otros sonidos para experimentar con otras sonificaciones.

PLANEA

CENTRO DE RECURSOS

¿Qué es el centro de recursos? Requisitos Colecciones Participa Revistas Comité

BUSCAR RECURSOS →

Inicio / Recursos / SonificADN

SonificADN



Recurso informático para la generación automática de secuencias sonoras a partir de secuencias genéticas

Recurso en torno a la guitarra flamenca preparada, un concepto vinculado al piano de John Cage y a otros artistas y prácticas experimentales del siglo XX y XXI. A lo largo del recurso se explica cómo prestar atención a las modificaciones que sufre el sonido de las cuerdas utilizando algo tan simple como la masilla que nos sirve para pegar papeles, el Blu Tack, o un puente de bambú que cambia el timbre de las cuerdas, palillos que nos permiten percudir el instrumento o un ventilador de mano modificado para conseguir trémolos infinitos. No hay necesidad de que el alumnado esté previamente vinculado con la guitarra y puede usarse por pequeños grupos. A través de esta experiencia vamos a descubrir y a preguntarnos cuestiones como ¿Qué es la música? ¿Cómo "se hace" la música? ¿Qué idea tenemos sobre qué es "la melodía"? ¿Puede entenderse la vibración de un golpe como música? ¿Qué necesito para hacer música? ¿Qué otras prátesis musicales conocemos? El acercamiento experimental al instrumento conecta con una tradición y un patrimonio que nos hace recordar que en cuestiones de aprendizaje no solo hay un camino posible.

Este recurso se puede distribuir, copiar y remezclar citando la fuente y manteniendo esta licencia.

COLECCIÓN

Destripando la tecnología para crear

PLANEA

¿Qué necesitas?
—
Un ordenador y conexión a internet

Nivel formativo
—
Bachillerato, ESO,

Materia
—
Educación musical, Lengua y literatura, Matemáticas, Tecnología, Transversal

Formato
—
Documento

Idiomas
—
Español

Fecha
—

Guía de instalación

5.1.

Se requiere su instalación en un ordenador con el sistema operativo de código abierto Linux, preferiblemente con la distribución de Linux Ubuntu. En su defecto, podría usarse un Windows o un Mac con una máquina virtual Linux en él, como por ejemplo, VirtualBox (<https://ubuntu.com/tutorials/how-to-run-ubuntu-desktop-on-a-virtual-machine-using-virtualbox>). En principio el programa podría también instalarse y usarse en un ordenador Windows o Mac, pero este recurso se ha desarrollado en Linux y no hay garantías de que funcione adecuadamente en Windows o Mac. Las instrucciones de esta guía están por tanto encaminadas a su instalación y uso en Linux.

5.2.

Descargar y descomprimir el .zip descargado en el punto 5.1. Se generará así un directorio llamado 'sonificADN' con el programa Python y su fichero de configuración, así como dos subdirectorios ('seqs' y 'sonidos') con ejemplos de una secuencia genética y de distintos sonidos de percusión, guitarra y voz, respectivamente, para utilizar directamente sin necesidad de descargar nuevas secuencias y diferentes sonidos.

5.3.

Instalar la librería Python 'musicpy' (<https://musicpy.readthedocs.io/en/latest/>) siguiendo los pasos indicados en el apartado 'Installation' de <https://musicpy.readthedocs.io/en/latest/Introduction/>.

5.4.

Instalar el módulo 'daw' de 'musicpy' siguiendo los pasos indicados en el apartado 'Preparation before importing' de <https://musicpy.readthedocs.io/en/latest/Musicpy%20daw%20module/>

Guía de uso

6.1. Obtención de secuencias genéticas de distintos organismos

Existen numerosas bases de datos públicas en internet desde donde pueden descargarse libremente multitud de secuencias genéticas de distintos organismos. La principal de ellas es el repositorio internacional de secuencias genéticas GenBank. Para descargar secuencias genéticas desde GenBank y usarlas con sonificADN se pueden seguir los siguientes pasos:



Abrir la página web <https://www.ncbi.nlm.nih.gov/genbank>

The screenshot shows the GenBank website interface. At the top, there is a navigation bar with the NIH logo and the text 'National Library of Medicine National Center for Biotechnology Information'. Below this is a search bar with 'Nucleotide' selected and a 'Search' button. A navigation menu includes links for 'GenBank', 'Submit', 'Genomes', 'WGS', 'Metagenomes', 'TPA', 'TSA', 'INSDC', 'Documentation', and 'Other'. The main content area is titled 'GenBank Overview' and includes a 'What is GenBank?' section with text about the database and a 'GenBank Resources' sidebar with links like 'GenBank Home', 'Submission Types', 'Submission Tools', 'Search GenBank', and 'Update GenBank Records'.

- a) En el cuadro de búsqueda ('Search') que aparece arriba, seleccionar 'Nucleotide' (viene seleccionado por defecto) e introducir un término de búsqueda (por ejemplo, 'Homo sapiens' para buscar secuencias de humano).

- b) Tras darle al botón “Search” aparecerá una lista de items de la base de datos que contienen el término usado en la búsqueda.

The screenshot shows a search results page with the following elements:

- Source databases:** INSDC (GenBank) (49,316,973), RefSeq (13,326,088), Customize ...
- Sequence Type:** Nucleotide (52,303,894), EST (8,864,030), GSS (1,795,947)
- Genetic compartments:** Chloroplast (5), Mitochondrion (216,268), Plasmid (60,213), Plastid (2,047)
- Sequence length:** Custom range...
- Release date:** Custom range...
- Revision date:** Custom range...
- Clear all** and **Show additional filters** links.
- Items: 1 to 20 of 62963871** (Page 1 of 3148194)
- Item 1:** [Lynx pardinus genome assembly, contig: lp23s36493, whole genome shotgun sequence](#)
4,643,072 bp linear DNA
Accession: CAAGRJ010006848.1 GI: 1603550370
[BioProject](#) [BioSample](#) [Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- Item 2:** [Homo sapiens Kidd blood group protein \(SLC14A1\) gene, exons 4, 5 and partial cds](#)
441 bp linear DNA
Accession: KJ946236.1 GI: 683576128
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- Item 3:** [Homo sapiens syntaxin 16 isoform C-like protein \(STX16\) mRNA, 3' UTR](#)
608 bp linear mRNA
Accession: DQ268529.1 GI: 82780403
[Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- Item 4:** [Homo sapiens sideroflexin 4 \(SFXN4\), RefSeqGene on chromosome 10; nuclear gene for mitochondrial product](#)
31,948 bp linear DNA
Accession: NG_033895.2 GI: 2302520031
[Protein](#) [PubMed](#) [Taxonomy](#)
- Search details:** "Homo sapiens"[Organism] OR Homo sapiens[All Fields]
- Recent activity:** Turn Off Clear
Homo sapiens (62963871) Nucleotide

- c) Muchos de los resultados contienen secuencias muy largas que le llevarían al programa mucho tiempo para sonificar, por lo que se recomienda filtrar los resultados por tamaño de la secuencia usando la herramienta de filtrado que aparece a la izquierda de la página: “Sequence length - Custom range...”, para seleccionar secuencias de un rango de tamaños más manejable. Una secuencia de 200 nucleótidos (como la que se proporciona como ejemplo junto con el programa) le cuesta al programa menos de un minuto para sonificarla en un portátil estándar y produce una secuencia sonora de 53 segundos utilizando un ritmo de 150bpm (como el que se indica en el fichero de configuración proporcionado junto con el programa). Se recomienda hacer pruebas con distintos tamaños de secuencia y distintos bpm.

The screenshot shows the same search results page as above, but with a **Custom range** dialog box open over the **Sequence length** filter. The dialog box contains:

- Input fields for **100** and **1000**.
- Apply** and **Clear** buttons.

The background page shows the same search results, with the **Sequence length** filter now set to the custom range of 100 to 1000.

- d) Tras aplicar el filtrado por tamaño de secuencia, aparecerá una lista de items dentro de dicho rango de tamaños.

The screenshot shows the NCBI search results for 'Homo sapiens'. The search criteria are set to 'Nucleotide' and 'Homo sapiens'. The results are filtered by sequence length from 100 to 1000. The first two results are:

- Homo sapiens Kidd blood group protein (SLC14A1) gene, exons 4, 5 and partial cds**
1. 441 bp linear DNA
Accession: KJ946236.1 GI: 683576128
Taxonomy: [Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)
- Homo sapiens syntaxin 16 isoform C-like protein (STX16) mRNA, 3' UTR**
2. 608 bp linear mRNA
Accession: DQ268529.1 GI: 82780403
Taxonomy: [Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

The interface includes a left sidebar with filters for Species, Molecule types, Source databases, Sequence Type, Genetic compartments, and Sequence length. The main content area shows a list of items with their accession numbers and GI numbers. The right sidebar contains sections for 'Filter your results', 'Results by taxon', 'Find related data', 'Search details', and 'Recent activity'.

- e) Para descargar la secuencia de uno de ellos hay que usar el link "FASTA" que aparece debajo de él. Tras pinchar en "FASTA", aparecerá la secuencia genética seleccionada.

The screenshot shows the FASTA page for the selected sequence: 'Homo sapiens syntaxin 16 isoform C-like protein (STX16) mRNA, 3' UTR'. The accession number is DQ268529.1. The FASTA sequence is displayed as follows:

```
>DQ268529.1 Homo sapiens syntaxin 16 isoform C-like protein (STX16) mRNA, 3' UTR
TTATATACTAACTACATAATGACCTGTTCAAACAGACTCTATTTAATGAAGTGTGAATTTACACAGAGG
CCATTTTAAATGGGTACCCCATTTAGGATTAGTGGATCTCAAATTTAAACAAACATCACTCCATTTT
AAAGTAAAATATTACGCCACAGCGATTGGATTTATTTGGCTCTTCCATTGCCACTGAGCAATGCCAGGA
AGCAGGCACATTGCCAAGGACTGGGGCATCTTGACTAGGAAGCTCCTCGTTTGTGTGAGCGTGGGTGAG
ACCGCCAAAGTAGGACTTCATCGTTTACCTACCTATTATCAATATGGTCTTGAATATATTCCTATAACT
GTAGAACAGTGTGAAAGTGATGGTAACTTTGAAGTTTTCATGTTTATTGGCTTTGTTTAAATTGACA
CTAGTAGAGTTTATGGTCTTTGTGTAATTTTATGGCTTTTCATTGTTTGTGCTCTCTAAAAGTTTAGAG
ATATGACTCATTAAATGTGCTGTTATTGGACAGTCACACAAATGTCTCTAGAGTGACTTAAAGTGTACA
GAATTAACATCAGAGATGAGTGTCAACAGCTCATGGCTAGTGTCAA
```

The page includes a 'FASTA' dropdown menu, a 'Send to' dropdown, and a 'Change region shown' dropdown. The right sidebar contains sections for 'Customize view', 'Analyze this sequence', 'Related information', and 'LinkOut to external resources'.

- f) Para descargarla en nuestro ordenador, hay que usar el link “Send to” que aparece arriba a la derecha, seleccionando “Complete Record” y “Choose Destination → File”. Tras ello, seleccionar “Format → FASTA” y darle a “Create File”.

The screenshot shows the NIH website interface. At the top, there's a navigation bar with the NIH logo and 'National Library of Medicine' text. Below that, a search bar is visible. The main content area displays the FASTA format for the 'Homo sapiens syntaxin 16 isoform C-like protein (STX16) mRNA'. A 'Send to' dropdown menu is open, showing options for 'Complete Record' and 'Gene Features'. Under 'Choose Destination', 'File' is selected. The 'Format' dropdown is set to 'FASTA', and the 'Create File' button is visible. The FASTA sequence is displayed in the background.

- g) Dependiendo de cómo tengamos configurado el navegador, se nos pedirá que seleccionemos un nombre para el fichero con la secuencia, así como un directorio de nuestro ordenador donde descargarla, o el navegador descargará directamente la secuencia en el directorio que tenga asignado para las descargas.

Por otro lado, hay que decir que sonificADN puede sonificar cualquier secuencia genética que se tenga en el ordenador, sea ésta real (descargada desde una base de datos pública como se indica arriba) o inventada. Una opción, pues, es crear una secuencia genética mediante la escritura en un fichero de texto de una secuencia de letras A, T, G y C inventada. El único requisito es emplear un editor de texto simple para ello (en Ubuntu se puede usar 'gedit', o 'nano') para guardar la secuencia en formato de texto simple y no en otro formato que incluya información más allá que simplemente caracteres de texto (como color, tamaño o tipo de letra, o tamaño de los márgenes, cursiva, negrita y otras informaciones de maquetación) y no incluir caracteres distintos de A, T, C y G.

6.2. Obtención de sonidos para usar en la sonificación

En teoría cualquier conjunto de ficheros .wav o .mp3 podría usarse para realizar la sonificación de una secuencia sonora. Lo más adecuado, no obstante, es utilizar ficheros .wav o .mp3 que contengan cada uno de ellos un sonido simple y corto, para evitar el solapamiento excesivo que se produciría entre sonidos largos al encadenarlos en la secuencia sonora cuando la duración del sonido es mayor que el intervalo entre ellos.

Existen distintas bases de datos de sonidos de donde descargar ficheros con sonidos que cumplan estas características. Uno de ellos, remarcable por tratarse de una iniciativa de subida y descarga de ficheros de audio con licencia Creative Commons, es Freesound (<https://freesound.org>). Para descargar ficheros de sonidos desde Freesound y usarlos con sonificADN se pueden seguir los siguientes pasos:

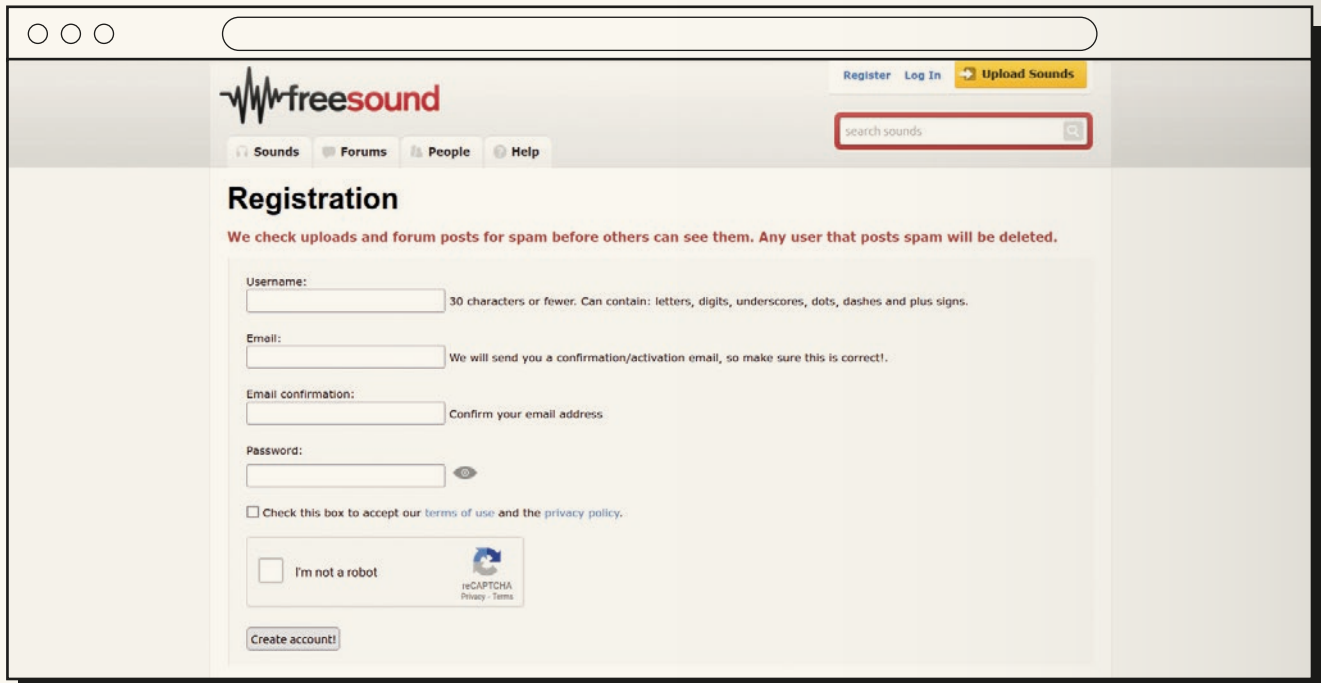


Abrir la página web <https://freesound.org>

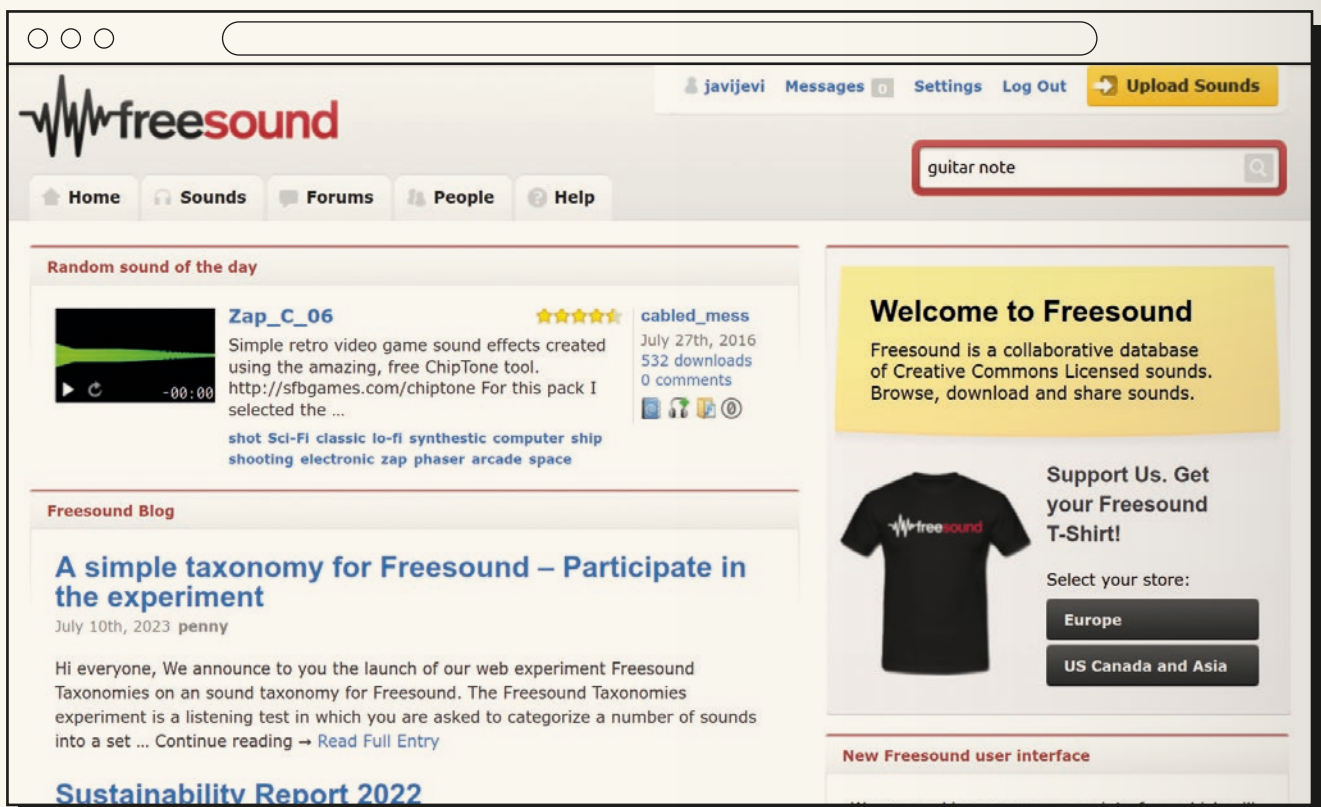
The screenshot shows the Freesound website homepage. At the top, there is a navigation bar with the Freesound logo, a search bar, and buttons for 'Register', 'Log In', and 'Upload Sounds'. Below the navigation bar, there are tabs for 'Sounds', 'Forums', 'People', and 'Help'. The main content area is divided into several sections:

- Random sound of the day:** Features a sound titled 'Zap_C_06' by 'cabled_mess', described as 'Simple retro video game sound effects created using the amazing, free ChipTone tool.' It includes a play button, a progress bar, and a star rating.
- Freesound Blog:** Contains two blog entries: 'A simple taxonomy for Freesound – Participate in the experiment' and 'Sustainability Report 2022'.
- Welcome to Freesound:** A yellow banner with the text: 'Freesound is a collaborative database of Creative Commons Licensed sounds. Browse, download and share sounds.'
- Support Us. Get your Freesound T-Shirt!:** A section promoting a t-shirt, with buttons to select a store: 'Europe' and 'US Canada and Asia'.
- New Freesound user interface:** A section announcing a new user interface, with a 'Try the new UI' button.

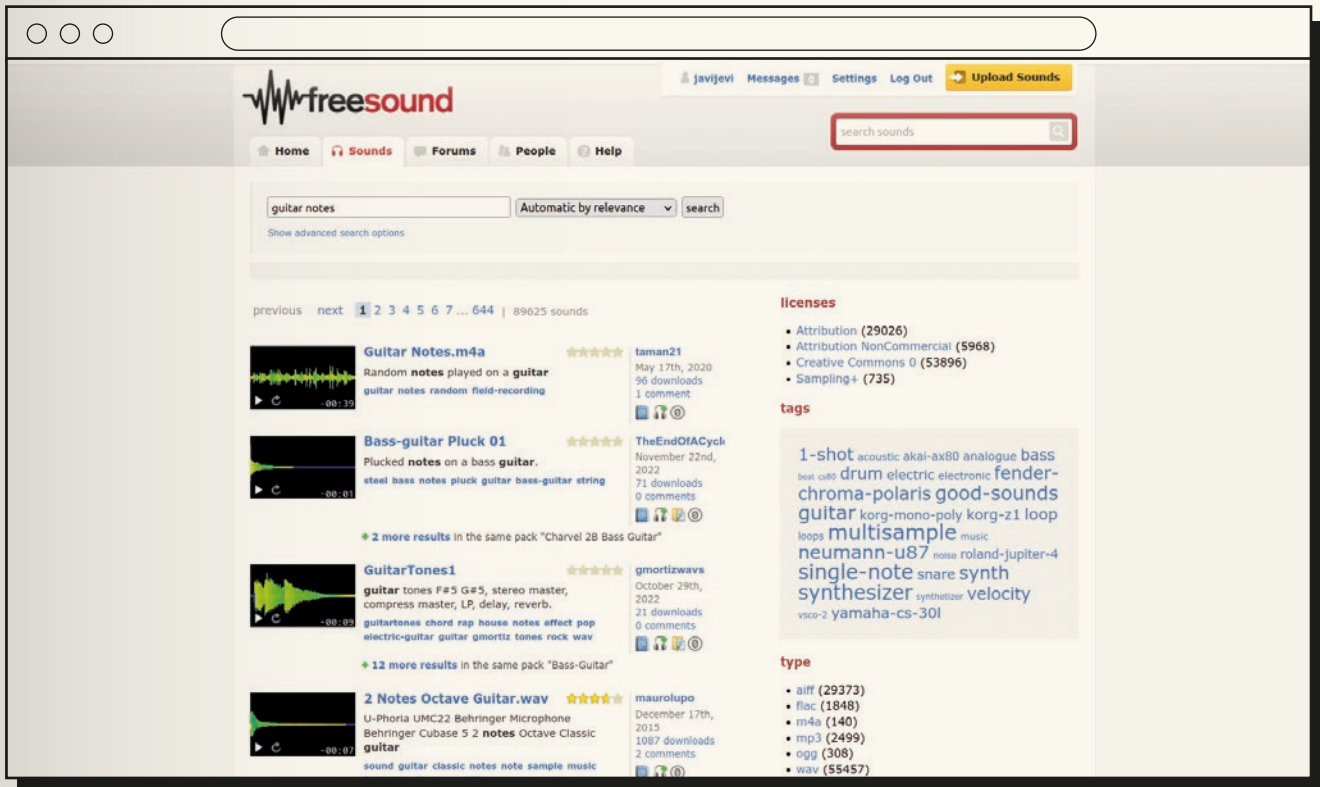
- a) Para poder descargar ficheros de audio desde Freesound es necesario estar registrado y haber hecho login. Si no estamos registrados en Freesound hay que hacerlo usando el enlace “Register” de arriba a la derecha. El registro es gratuito y sólo requiere elegir un nombre de usuario, proporcionar un email válido para recibir un enlace de confirmación del registro y elegir una contraseña para hacer login.



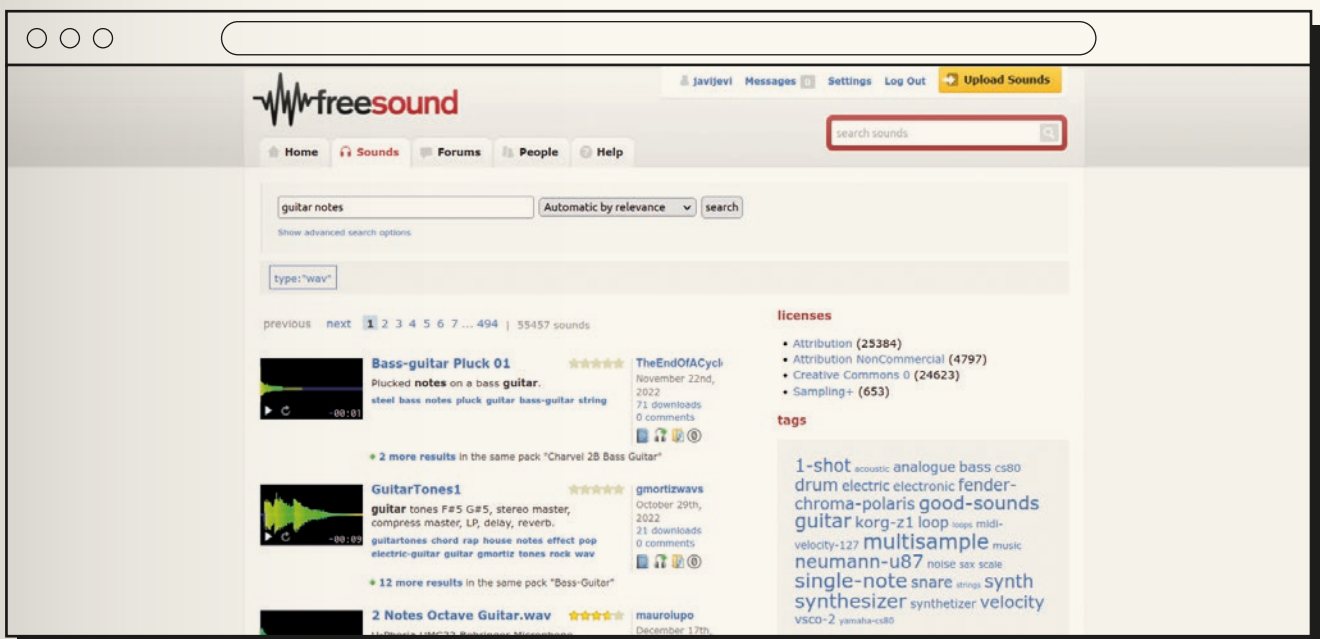
- b) Tras hacer login, para hacer una búsqueda de sonidos hay que introducir un término de búsqueda (por ejemplo, 'guitar notes' para buscar notas sueltas de guitarra) en el cuadro de búsqueda ('search sounds') que aparece arriba a la derecha.



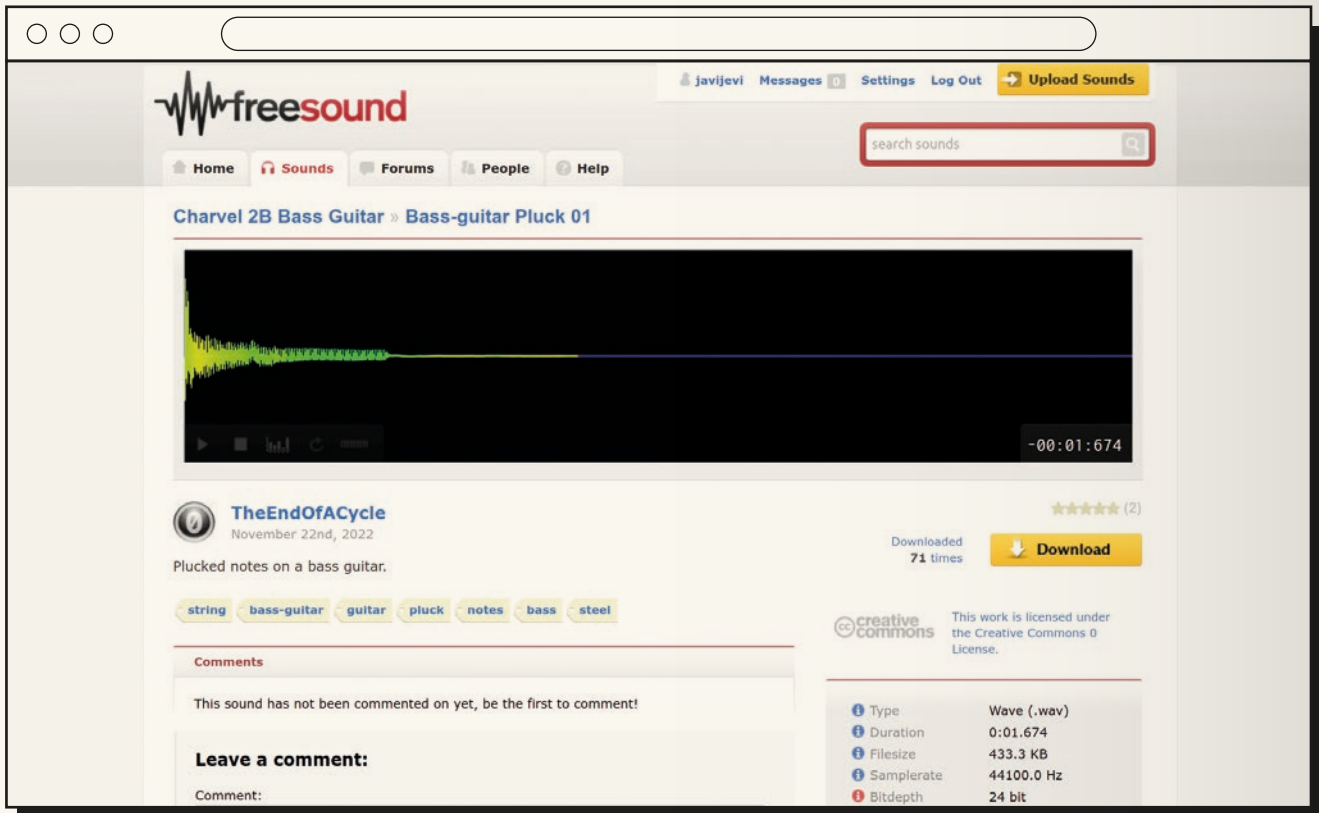
- c) Tras darle al icono de la lupa para iniciar la búsqueda, aparecerá una lista de items en la base de datos que contienen el término usado en la misma.



- d) Dichos items contienen sonidos en distintos formatos (.wav, .mp3, .flac, .ogg, etc.), por lo que se recomienda usar la herramienta de filtrado que aparece a la derecha para seleccionar sólo sonidos en formato .wav (aunque los sonidos en formato .mp3 y .ogg deberían poder utilizarse también).
- e) Tras aplicar el filtrado por formato de audio, aparecerá una lista de items en dicho formato. En la imagen que aparece junto al título de cada uno de ellos aparece un icono de "Play" que se puede usar para escuchar el sonido del item y poder decidir el que se quiera descargar.



- f) Tras hacer login, para hacer una búsqueda de sonidos hay que introducir un término de búsqueda (por ejemplo, 'guitar notes' para buscar notas sueltas de guitarra) en el cuadro de búsqueda ('search sounds') que aparece arriba a la derecha.



- g) Dependiendo de cómo tengamos configurado el navegador, se nos pedirá que seleccionemos un nombre para el fichero con la secuencia, así como un directorio de nuestro ordenador donde descargarla, o el navegador descargará directamente la secuencia en el directorio que tenga asignado para las descargas.

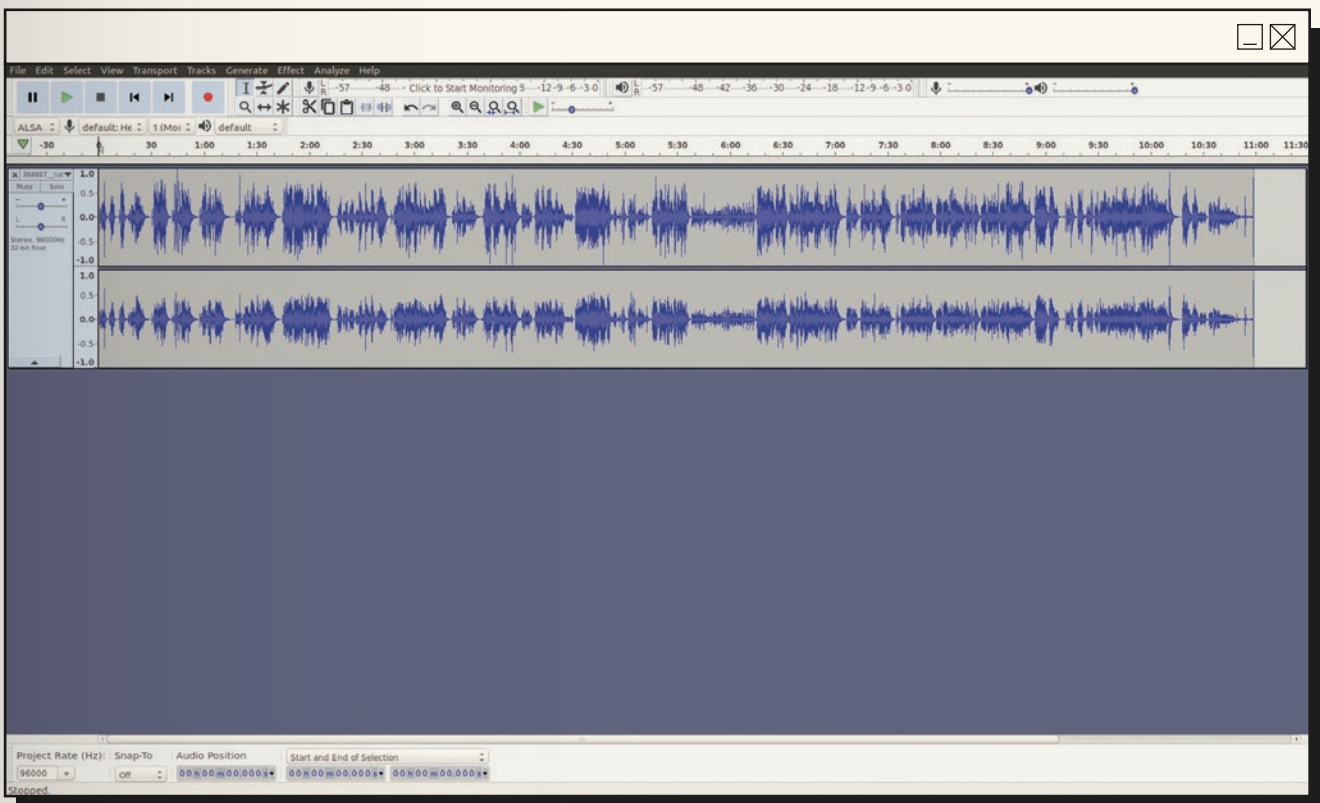
Otro modo de conseguir ficheros .wav o .mp3 con sonidos simples y cortos es recortarlos y extraerlos a partir de ficheros de audio largos (por ejemplo de canciones) usando cualquier software de edición de audio. Se recomienda la herramienta Audacity (<https://audacity.es>), por ser software gratuito, de código abierto y multiplataforma, siguiendo los siguientes pasos:



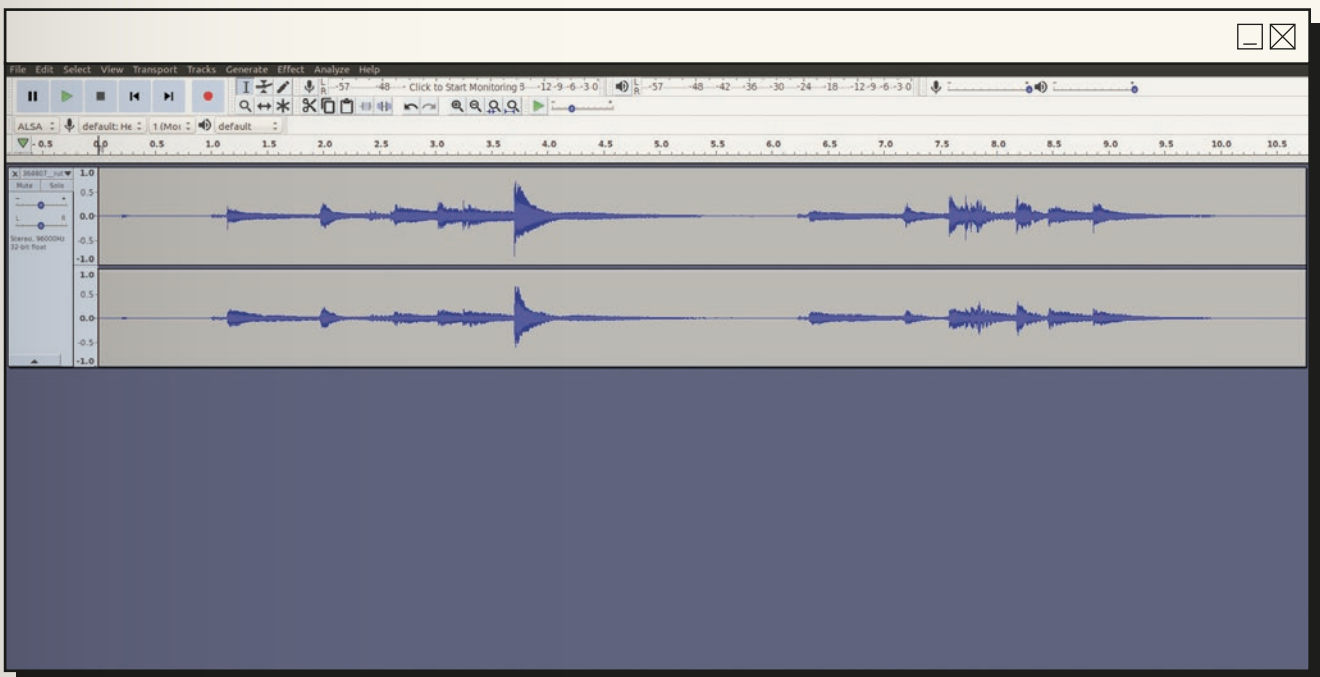
Instalar Audacity en nuestro ordenador

(se puede instalar y usar tanto en Linux como en Windows y Mac).

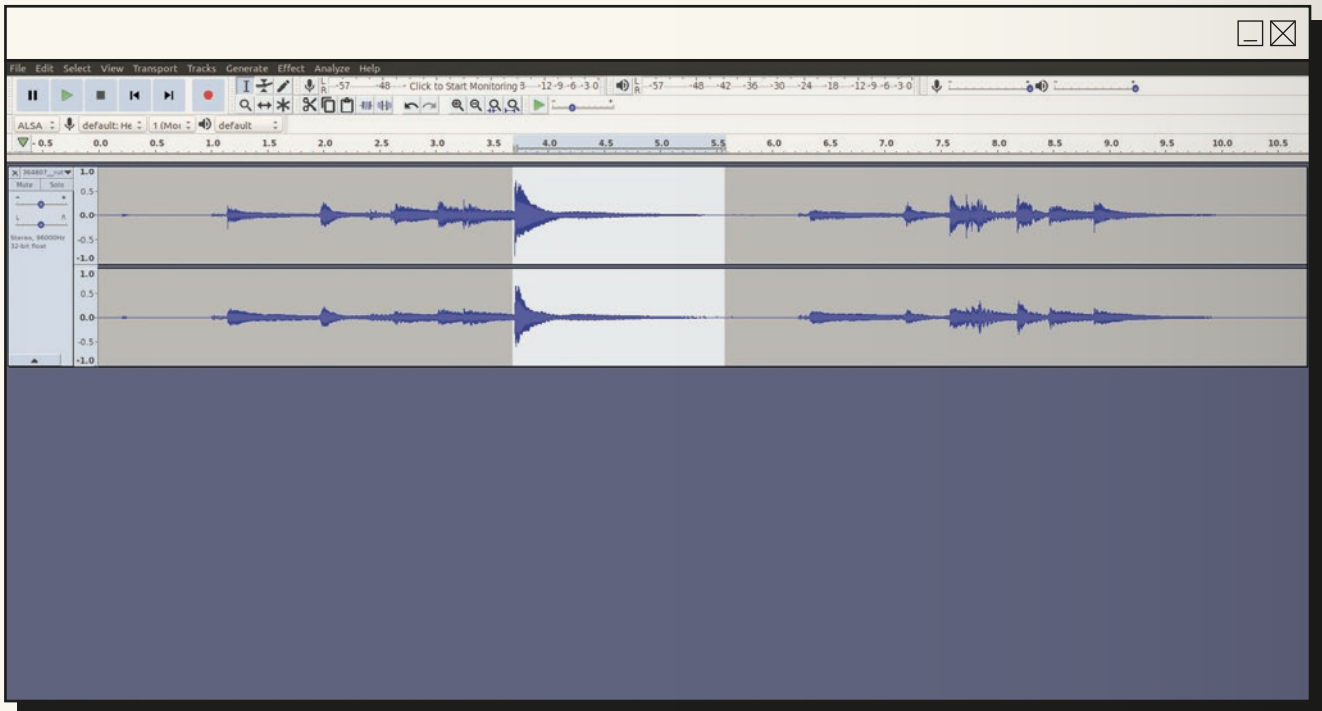
a) Abrir Audacity y abrir el fichero de audio del que queremos recortar un sonido simple y corto.



b) Darle al “Play” para escuchar el audio y así identificar en el audiograma el fragmento que queremos recortar y extraer. Haciendo previamente zoom-in con Ctrl+1 varias veces se consigue una mayor precisión a la hora de hacer dicha selección. Para hacer zoom-out, usar Ctrl+3.



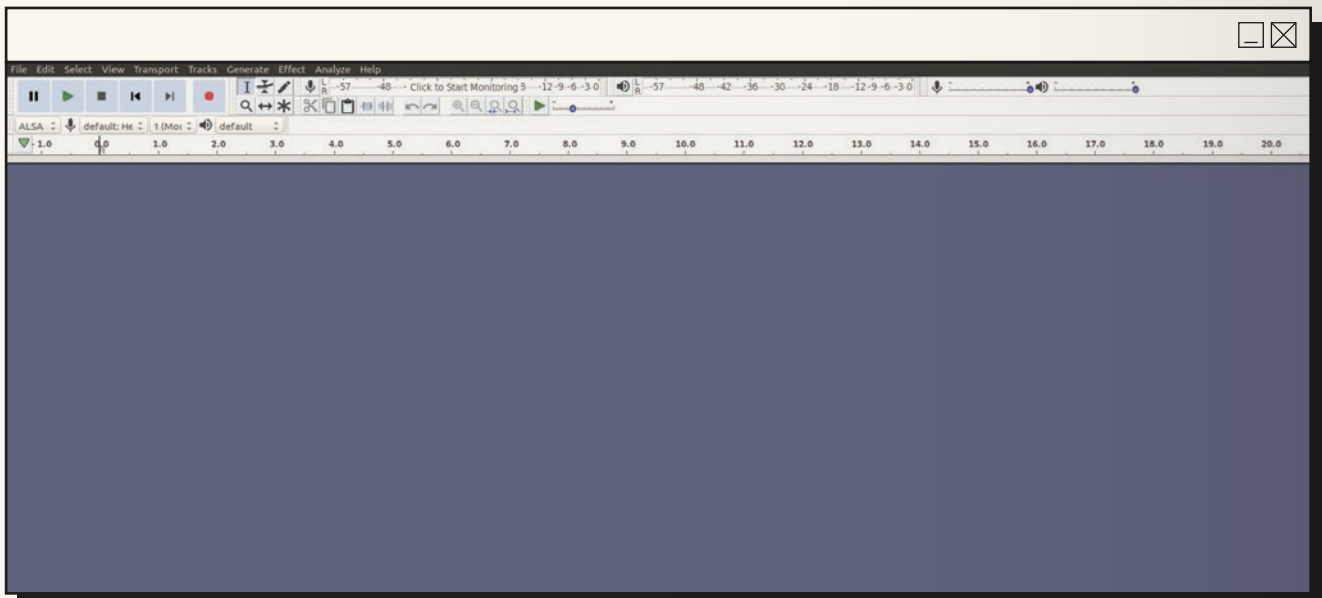
c) Una vez identificado el fragmento que se quiere recortar y extraer, seleccionarlo con el ratón.



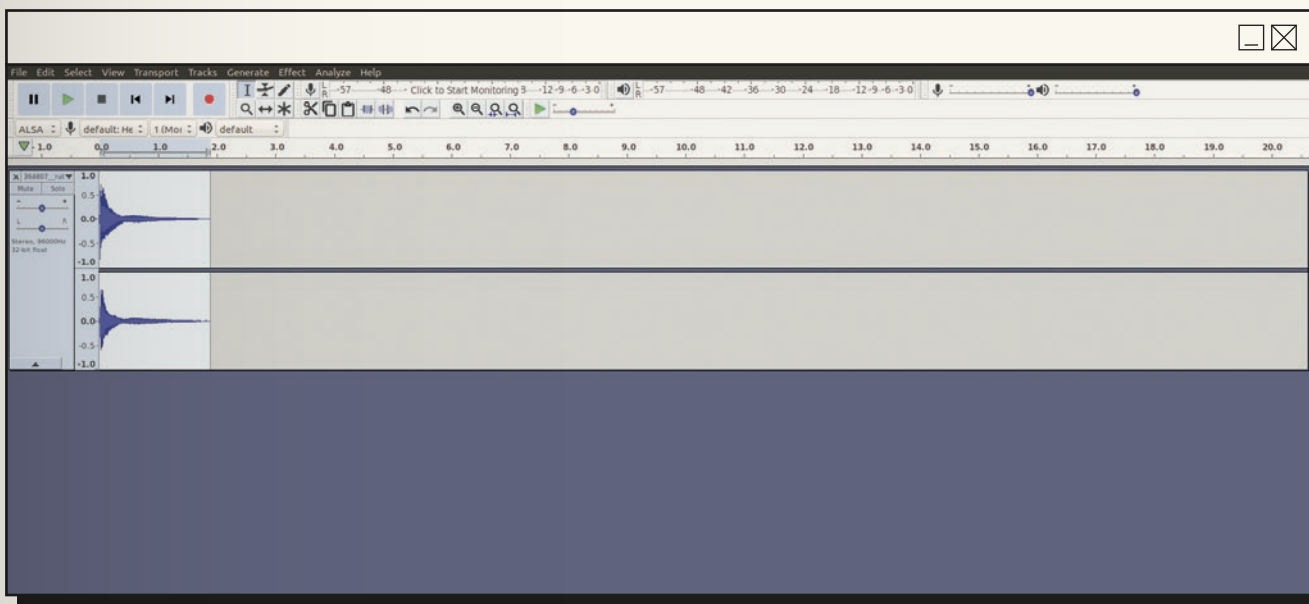
d) Una vez seleccionado con el ratón, se puede usar de nuevo el “Play” para escuchar únicamente dicho fragmento y ajustar más la selección tantas veces sea necesario para conseguir que suene únicamente el fragmento que se desea.

e) Copiar el fragmento (“Editar → Copiar”).

f) Abrir un audio nuevo (“Fichero → Nuevo”).



g) Pegar el fragmento copiado en el audio nuevo (“Editar → Pegar”)

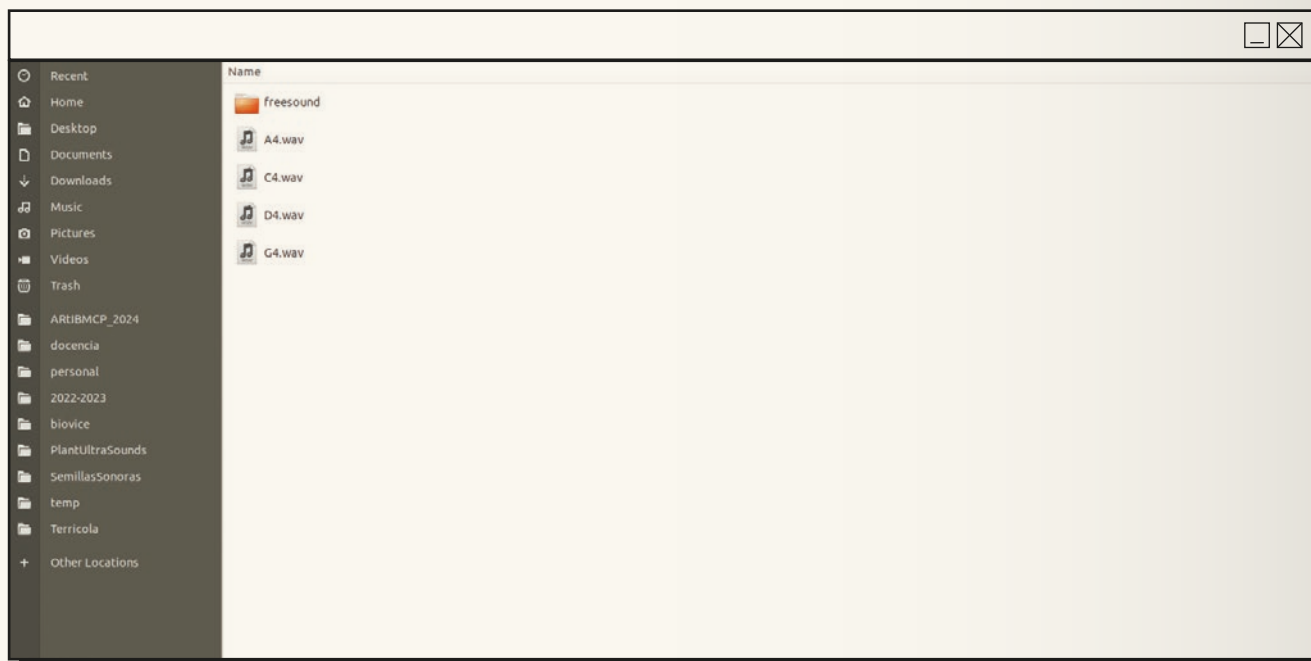


- h)** En este punto, se puede usar distintas herramientas de Audacity para mejorar el sonido obtenido (recortar extremos, hacer fade-in al principio, hacer fade-out al final, amplificar o reducir el volumen, cambiar el tono, distorsionar, etc.).
- i)** Exportar como .wav o .mp3 (“Fichero → Exportar → Exportar como WAV” o “Fichero → Exportar → Exportar como MP3”, respectivamente).
- j)** Se abrirá un cuadro de diálogo donde se puede seleccionar el nombre del fichero .wav o .mp3 que queremos crear con el sonido recortado y el directorio de nuestro ordenador donde lo queremos guardar.
- k)** Al darle a “Guardar” se nos abrirá un cuadro de diálogo donde podemos elegir metadatos para el fichero (nombre del artista, título de la pista, año, estilo, etc.).
- l)** Finalmente, dando a “OK” se creará el fichero y se guardará con el nombre seleccionado previamente y en el directorio seleccionado previamente.

6.3. Asociación entre sonidos y nucleótidos

La base de la sonificación de una secuencia genética consiste en la asociación de cada uno de los cuatro nucleótidos de una secuencia genética (A, T, C y G) a uno de los cuatro sonidos seleccionados de uno o más de distintos instrumentos, percusiones y/o voces. Esta asociación es la que hace que se pueda generar una secuencia sonora en la que el orden de los sonidos está determinado por el orden de los nucleótidos en la secuencia genética.

El modo que sonificADN realiza dicha asociación es la colocación de los cuatro ficheros con los cuatro sonidos de un instrumento en un directorio determinado, los de los cuatro sonidos de una percusión en otro directorio y los cuatro sonidos de una voz en otro, y el renombrado de dichos ficheros a A4.wav, C4.wav, D4.wav y G4.wav (o A4.mp3, C4.mp3, D4.mp3 y G4.mp3). El motivo de tener que utilizar dichos nombres es que la librería 'musicpy' de Python requiere que dichos ficheros tengan nombres de notas en la notación anglosajona (A, B, C, D, E, F y G en lugar de Do, Re, Mi, Fa, Sol, La, Si y Do, respectivamente) e incluyan el valor de la octava de la nota (aunque el sonido que contienen dichos ficheros no corresponda ni a esa nota ni a esa octava).



El programa asocia:

- el nucleótido adenina (A) con los sonidos de los ficheros con nombre A4.wav (o A4.mp3)
- el nucleótido citosina (C) con los sonidos de los ficheros con nombre C4.wav (o C4.mp3)
- el nucleótido guanina (G) con los sonidos de los ficheros con nombre G4.wav (o G4.mp3)
- y el nucleótido timina (T) con los sonidos de los ficheros con nombre D4.wav (o D4.mp3).

Aunque en tres de las cuatro asociaciones la letra que designa al nucleótido coincide con la letra que designa a la supuesta (que no real) nota del fichero correspondiente, la elección de dichas asociaciones es totalmente arbitraria.

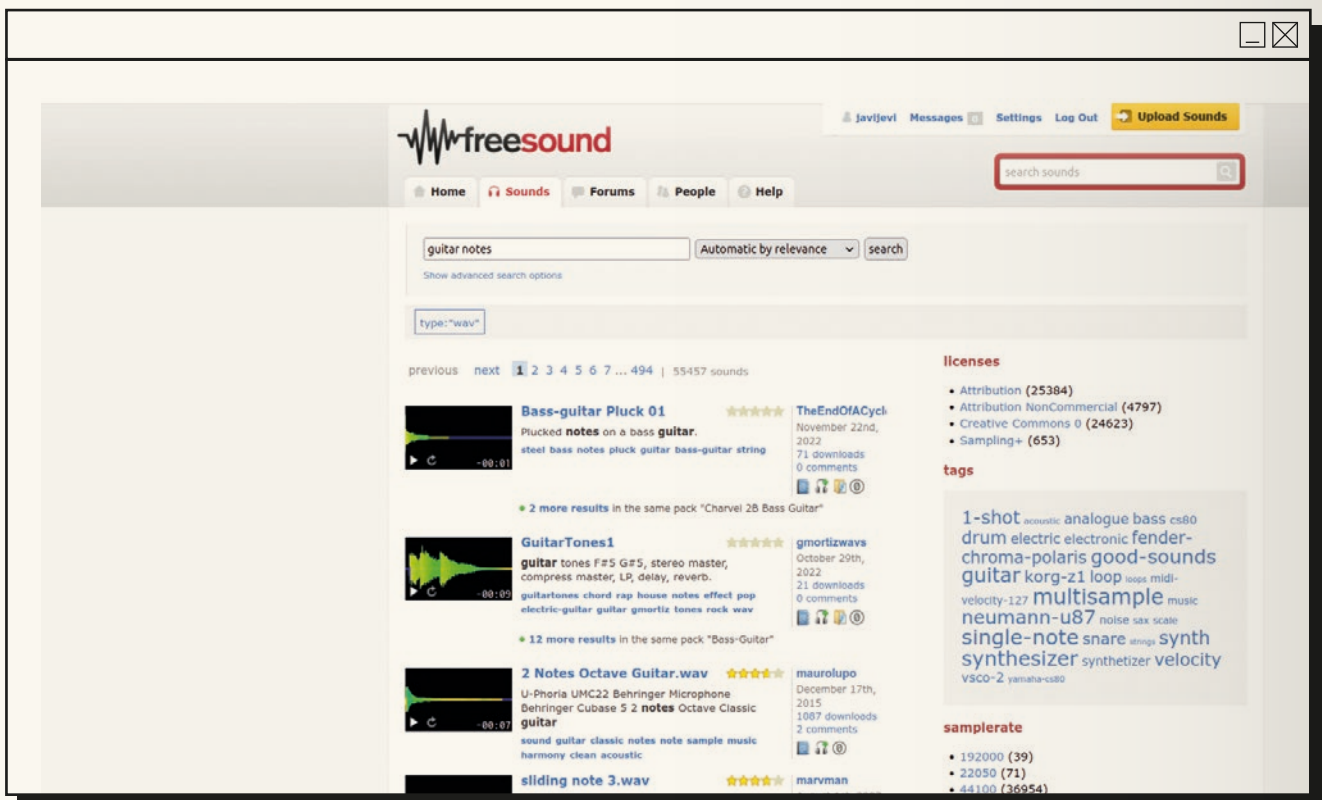
De este modo, los ficheros con sonidos de percusión eventualmente descargados en el apartado “3.2. Obtención de sonidos para usar en la sonificación” deberán renombrarse a A4.wav, C4.wav, D4.wav y G4.wav (o A4.mp3, C4.mp3, D4.mp3 y G4.mp3), independientemente de la nota y la octava que contengan, y situarlos en un mismo directorio del ordenador. Obviamente, dicho directorio debe ser distinto del utilizado para los sonidos de un instrumento y del utilizado para los sonidos de voces.

De igual modo, los ficheros con sonidos de un instrumento eventualmente descargados en el apartado “3.2. Obtención de sonidos para usar en la sonificación” deberán renombrarse a A4.wav, C4.wav, D4.wav y G4.wav (o A4.mp3, C4.mp3, D4.mp3 y G4.mp3), independientemente de la nota y la octava que contengan, y situarlos en un mismo directorio del ordenador. Obviamente, dicho directorio debe ser distinto del utilizado para los sonidos de percusión y del utilizado para los sonidos de voces.

De igual modo, los ficheros con sonidos de voces eventualmente descargados en el apartado “3.2. Obtención de sonidos para usar en la sonificación” deberán renombrarse a A4.wav, C4.wav, D4.wav y G4.wav (o A4.mp3, C4.mp3, D4.mp3 y G4.mp3), independientemente de la nota y la octava que contengan, y situarlos en un mismo directorio del ordenador. Obviamente, dicho directorio debe ser distinto del utilizado para los sonidos de percusión y del utilizado para los sonidos de un instrumento.

Las rutas de estos directorios con los sonidos que se desea usar en la sonificación deben ser indicadas en el fichero de configuración 'config.txt', donde se especifican otros parámetros del programa, tales como el nombre del fichero .mp3 que se generará, la velocidad del ritmo de la secuencia sonora que se generará, la ruta del fichero con la secuencia genética que se sonificará, el porcentaje de sonidos que se desea silenciar para evitar una secuencia sonora excesivamente monótona y el volumen que se desea para la percusión, para el instrumento y para las voces (ver apartado "5.1. Fichero de configuración").

En el fichero de configuración que viene con el .zip descargado en el apartado "1. Programa informático" se indican las rutas de los directorios con los sonidos que se proporcionan como ejemplo.

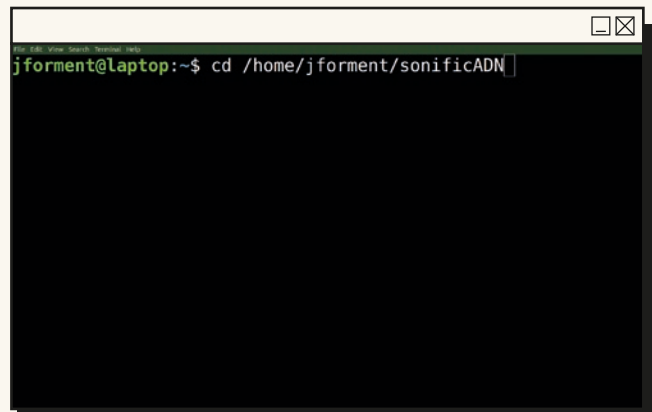
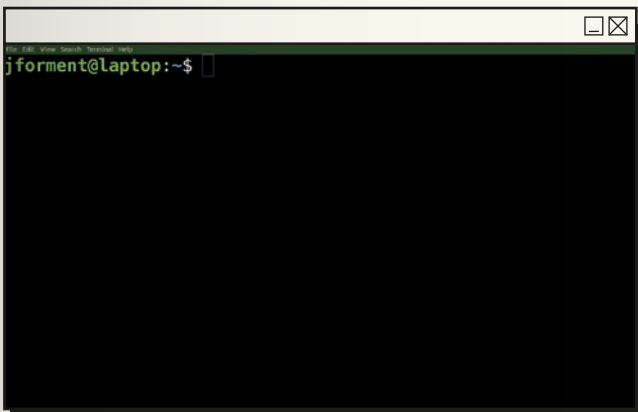


6.4. Modo de correr el programa

Una vez seguidos los pasos anteriores, el modo de correr el programa 'sonificADN' es utilizando la terminal de Linux (también conocida como "línea de comandos", "shell" o "bash"). Aunque el uso de la terminal Linux puede parecer complejo al principio, en realidad no lo es tanto. Es tan sólo un entorno al que puede no estarse acostumbrado. Lo importante es seguir los pasos indicados con exactitud, escribiendo exactamente los comandos indicados, recordando que mayúsculas y minúsculas no son lo mismo, que todos los espacios en blanco y las barras (/) indicados son estrictamente necesarios, sin omitir ni añadir nada.

Para realizar la sonificación de la secuencia genética de ejemplo usando los sonidos de ejemplo proporcionados en el .zip que nos hemos descargado, los pasos son los siguientes:

- a) Abrir una terminal Linux (también llamada "línea de comandos"). En Ubuntu existen diversos modos de hacerlo (la mayoría de ellos también funcionan en otras distribuciones Linux), como por ejemplo pulsar ALT+F2, escribir 'gnome-terminal' en el cuadro que se abre y darle a 'Intro'.



- b) Usar el comando 'cd' para moverse en la terminal al directorio 'sonificADN' que se ha creado al descomprimir el .zip con el programa. Por ejemplo, si dicho directorio está en la ruta /home/usuario/sonificADN (donde 'usuario' es el usuario con el que hemos iniciado sesión en Linux), el comando sería escribir lo siguiente en la terminal y dándole a 'Intro':

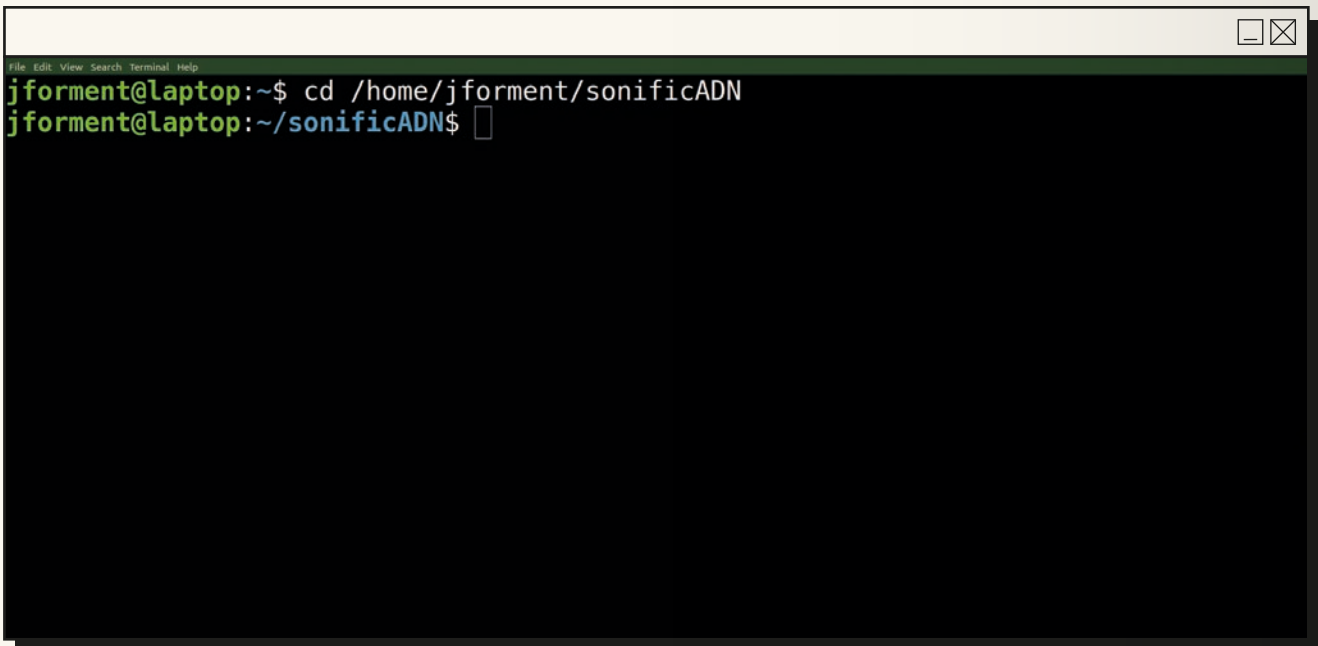
cd /home/jforment/sonificADN

Si no recordamos dónde hemos descomprimido el .zip, podemos localizar el directorio 'sonificADN' con el comando 'locate' escribiendo lo siguiente en la terminal y dándole a 'Intro':

locate sonificADN

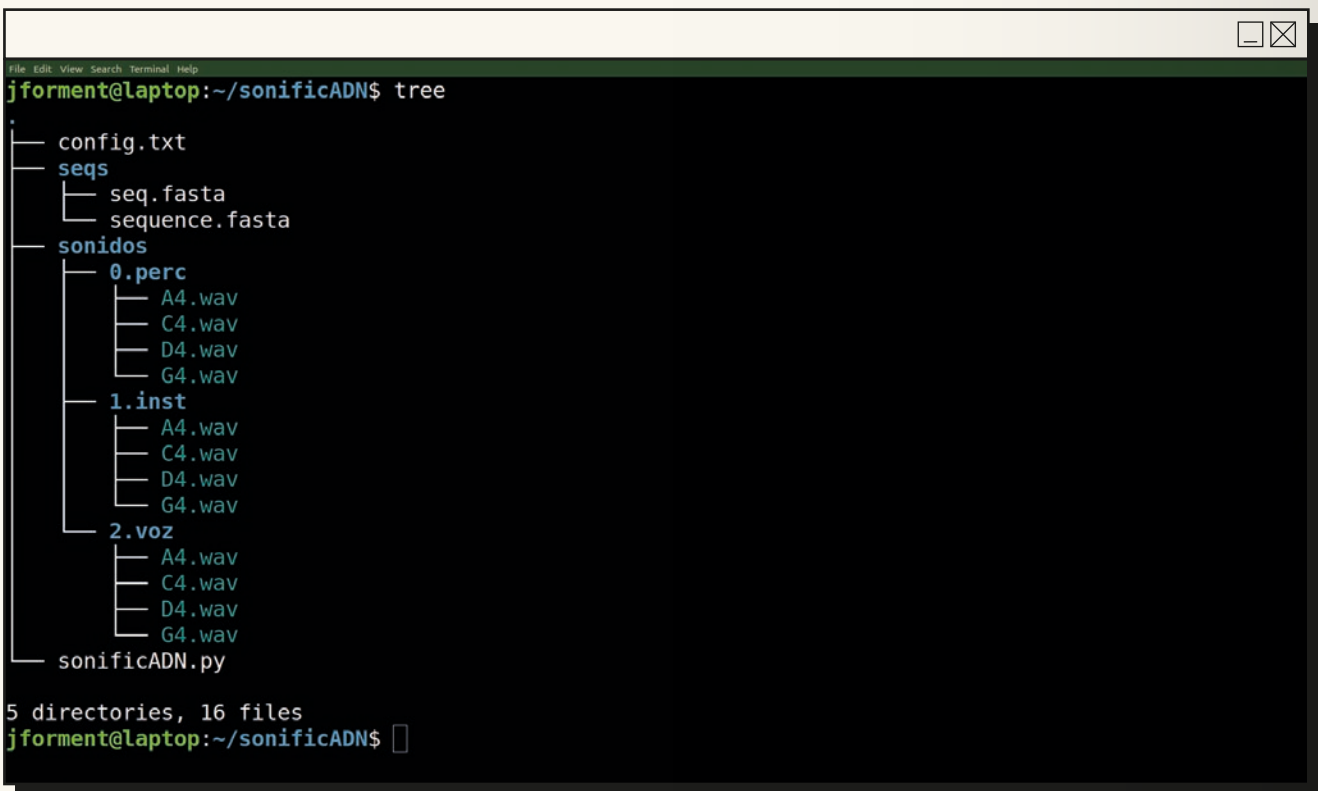
La primera ruta que aparezca es la ruta que deberemos usar con el comando 'cd' (en lugar de '/home/usuario/clases/sonificADN' en el ejemplo de arriba).

- c) Una vez situados correctamente en el directorio 'sonificADN' (nos aparecerá parcialmente en la terminal la ruta terminada en 'sonificADN', justo a la izquierda del cursor).



```
File Edit View Search Terminal Help
jforment@laptop:~$ cd /home/jforment/sonificADN
jforment@laptop:~/sonificADN$
```

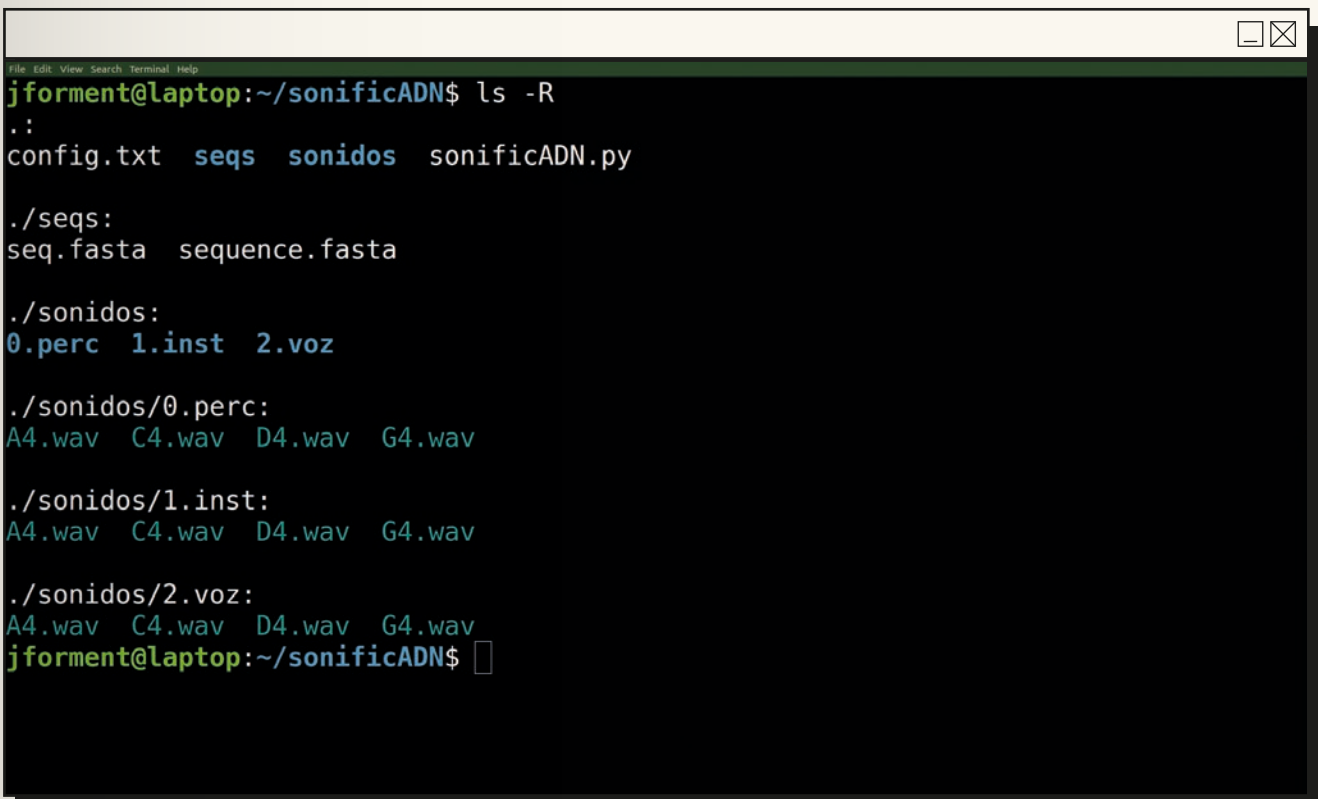
- d) Podemos comprobar que tenemos allí todos los ficheros necesarios para correr el programa con el comando 'tree', que nos muestra la estructura de ficheros y subdirectorios



```
File Edit View Search Terminal Help
jforment@laptop:~/sonificADN$ tree
.
├── config.txt
├── seqs
│   ├── seq.fasta
│   └── sequence.fasta
├── sonidos
│   ├── 0.perc
│   │   ├── A4.wav
│   │   ├── C4.wav
│   │   ├── D4.wav
│   │   └── G4.wav
│   ├── 1.inst
│   │   ├── A4.wav
│   │   ├── C4.wav
│   │   ├── D4.wav
│   │   └── G4.wav
│   └── 2.voz
│       ├── A4.wav
│       ├── C4.wav
│       ├── D4.wav
│       └── G4.wav
└── sonificADN.py

5 directories, 16 files
jforment@laptop:~/sonificADN$
```

- e) Es posible que el comando 'tree' no esté instalado en nuestro Linux, en cuyo caso podemos comprobar la estructura de ficheros y subdirectorios con 'ls -R' (aunque nos lo muestra de un modo menos gráfico e intuitivo).



```
File Edit View Search Terminal Help
jforment@laptop:~/sonificADN$ ls -R
.:
config.txt  seqs  sonidos  sonificADN.py

./seqs:
seq.fasta  sequence.fasta

./sonidos:
0.perc  1.inst  2.voz

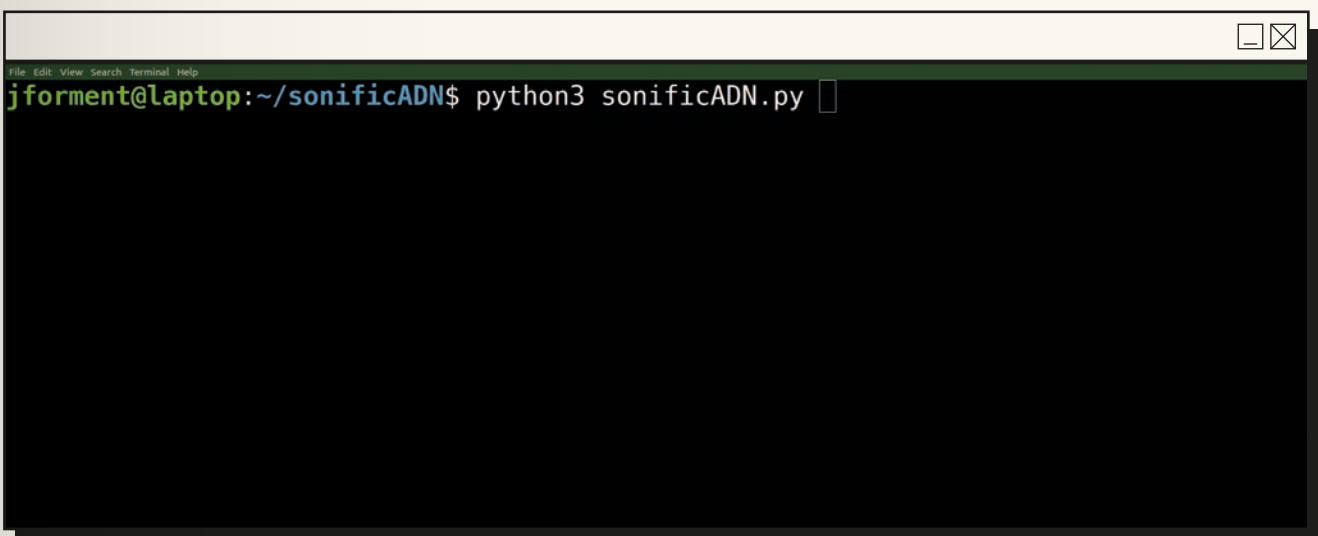
./sonidos/0.perc:
A4.wav  C4.wav  D4.wav  G4.wav

./sonidos/1.inst:
A4.wav  C4.wav  D4.wav  G4.wav

./sonidos/2.voz:
A4.wav  C4.wav  D4.wav  G4.wav
jforment@laptop:~/sonificADN$
```

- f) Una vez hemos comprobado que estamos en el directorio adecuado, se puede correr el programa escribiendo el siguiente comando y dándole a 'Intro':

python3 sonificADN.py



```
File Edit View Search Terminal Help
jforment@laptop:~/sonificADN$ python3 sonificADN.py
```

- g) En la terminal irán apareciendo sucesivamente los siguientes mensajes (con mayor o menor rapidez, dependiendo de la longitud de la secuencia utilizada y la potencia del ordenador utilizado):

rendering track 1/3 channel 1

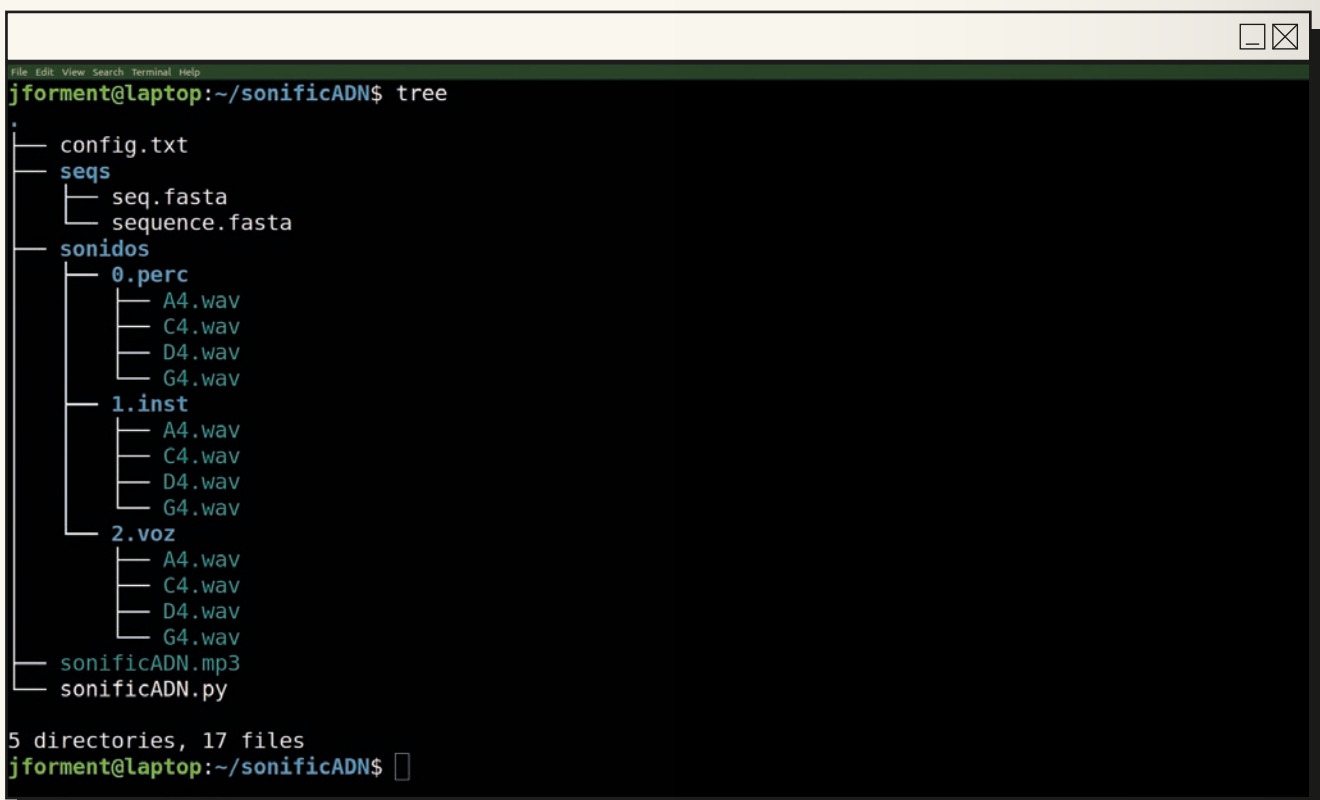
rendering track 2/3 channel 2

rendering track 3/3 channel 3

rendering finished

export finished

- h) Tras ello, volverá a aparecer el cursor (con la ruta parcial terminada en 'sonificADN' justo a su izquierda), lo cual indica que el programa ha terminado. En el directorio en el que estamos ('sonificADN') habrá aparecido un fichero .mp3 que contiene la sonificación. Dicho fichero tendrá el nombre 'sonificADN.mp3' (o el que hayamos indicado en el fichero de configuración 'config.txt' si lo hemos cambiado). Podemos comprobarlo de nuevo con el comando 'tree' o 'ls -R', que nos indican los nombres de los ficheros y directorios que contiene el directorio en el que nos encontramos:



```
File Edit View Search Terminal Help
jforment@laptop:~/sonificADN$ tree
.
├── config.txt
├── seqs
│   ├── seq.fasta
│   └── sequence.fasta
├── sonidos
│   ├── 0.perc
│   │   ├── A4.wav
│   │   ├── C4.wav
│   │   ├── D4.wav
│   │   └── G4.wav
│   ├── 1.inst
│   │   ├── A4.wav
│   │   ├── C4.wav
│   │   ├── D4.wav
│   │   └── G4.wav
│   └── 2.voz
│       ├── A4.wav
│       ├── C4.wav
│       ├── D4.wav
│       └── G4.wav
├── sonificADN.mp3
└── sonificADN.py

5 directories, 17 files
jforment@laptop:~/sonificADN$
```



```
File Edit View Search Terminal Help
jforment@laptop:~/sonificADN$ ls -R
.:
config.txt seqs sonidos sonificADN.mp3 sonificADN.py

./seqs:
seq.fasta sequence.fasta

./sonidos:
0.perc 1.inst 2.voz

./sonidos/0.perc:
A4.wav C4.wav D4.wav G4.wav

./sonidos/1.inst:
A4.wav C4.wav D4.wav G4.wav

./sonidos/2.voz:
A4.wav C4.wav D4.wav G4.wav
jforment@laptop:~/sonificADN$
```

- i) Podemos escuchar dicho fichero .mp3 con la sonificación de la secuencia genética en cualquier dispositivo que reproduzca ficheros .mp3.

6.5. Fichero de configuración

Si queremos utilizar otra secuencia genética distinta de la proporcionada como ejemplo, y que nos habremos descargado previamente (según lo indicado en apartados anteriores), y/o si queremos utilizar otros sonidos distintos de los proporcionados como ejemplo, y que nos habremos descargado previamente (según lo indicado en apartados anteriores), hay que editar el fichero de configuración 'config.txt' para indicarle al programa dónde se encuentran dicha secuencia y/o dichos sonidos. Para ello, seguiremos los siguientes pasos:

- Abrir el fichero de configuración 'config.txt' con un editor de texto. En Ubuntu se puede usar el editor de texto gráfico 'gedit'.
- Cambiar las líneas que comienzan por 'seqfasta' y/o 'percdir', 'instdir' y/o 'vozdir' para que en lugar de 'seqs/sequence.fasta', 'sonidos/0.perc', 'sonidos/1.inst' y/o 'sonidos/2.voz' contengan la ruta del fichero con la secuencia genética que queremos usar, y/o la ruta de los directorios donde se encuentran los ficheros con los sonidos que queremos usar, respectivamente.

En dicho fichero de configuración también podemos cambiar otras características de la sonificación que se desea realizar:

- el nombre del fichero .mp3 resultante: en la línea que comienza con ‘ficheromp3’
- la velocidad del ritmo (en bpm’s, “beats per minute”) de la pieza sonora resultante: en la línea que comienza con ‘bpm’
- el porcentaje de notas, golpes de percusión y voces que queremos silenciar en la pieza sonora final para que ésta resulte menos monótona y más estética musicalmente: en la línea que comienza con ‘silencios’
- el volumen de las notas del instrumento, los golpes de percusión y las voces: en las líneas que comienzan con ‘volumenperc’, ‘volumeninst’ y ‘volumenvoz’, respectivamente.

Se recomienda experimentar con dichos cambios. Como por ejemplo:

- ver qué resulta al silenciar totalmente el instrumento, o la voz, o la percusión (o dos de ellos), poniendo su volumen a 0
- ver qué resulta cambiando el ritmo de la canción (parámetro ‘bpm’)
- ver qué resulta cambiando el porcentaje de silencios
- ver qué resulta utilizando secuencias genéticas compuestas por la repetición de un sólo nucleótido (o dos de ellos)
- ver qué resulta utilizando en el parámetro ‘percdir’ un directorio que no contiene sonidos de percusión sino otro instrumento distinto al existente en ‘instdir’, u otra voz distinta a la existente en ‘vozdir’; o utilizando en ‘instdir’ y/o ‘vozdir’ directorios con sonidos de percusión distintos a los existentes en ‘percdir’.



sonificADN

Recurso informático para la generación automática de secuencias sonoras a partir de secuencias genéticas



¿A qué suena un gen? ¿En qué se parece un gen a una partitura? ¿Puede un gen crear una canción? El presente recurso ofrece a alumnado y docentes la posibilidad de asentar e interiorizar tanto el concepto de secuencia genética como el concepto de pieza sonora a partir del establecimiento de un paralelismo entre una secuencia de nucleótidos (un gen) y una secuencia de sonidos (una canción) mediante un programa informático. Dada la multidisciplinariedad del recurso, que combina conceptos de genética, música e informática, se propone su uso mediante la colaboración entre docentes de las áreas de Biología, Arte y Tecnología.