
Diseñador de sondas de hibridación

Simón Vergara

16 de abril de 2024

Contents:

1. Probe-Designer	1
1.1. descarga module	1
1.2. diseno module	2
1.3. multiplex module	6
1.4. probeDesignerUI module	7
2. Índices y tablas	11
Índice de Módulos Python	13
Índice	15

1.1 descarga module

Para obtener el ADN que se utilizará como referencia para el diseño de las sondas se va a descargar directamente el archivo genético desde la base de datos ‘Nucleotide’ de NCBI (National Center for Biotechnology Information) del NIH (National Institute of Health, EEUU) (<https://www.ncbi.nlm.nih.gov/nuccore>). Esta base de datos contiene millones de registros de secuencias de nucleótidos con mucha información adicional, como la especie, cromosoma, nombre del gen, exones, la fuente de los datos, etc.

Para llevar a cabo la descarga de los archivos genéticos, se implementó un módulo que permite realizar descargas desde esta base de datos a través de la librería Entrez de BioPython y almacenar los archivos en una carpeta. De esta manera se puede automatizar la obtención de las secuencias de referencia.

Por otra parte se implementó la función que retorna una secuencia a partir de un archivo almacenado localmente, además de crear una copia del archivo y almacenarla. Esta función sólo podrá recibir archivos de secuencia anotados como GenBank y GFF3.

El producto de estas funciones implementadas es la secuencia en formato SeqRecord, una estructura de datos propia de la librería BioPython que permite almacenar grandes secuencias con su respectiva notación, como por ejemplo las posiciones de cada uno de los exones, las transcripciones, secciones no codificantes, entre otros. Además contiene información extra de la secuencia tal como la especie, el cromosoma, la fuente de donde salieron los datos, etc.

`descarga.accnun_to_seqrecord(accession_number)`

Función que recibe un accession number de la base de datos “nucleotide” de NCBI <https://www.ncbi.nlm.nih.gov/nuccore>. Genera el archivo .gbk (GenBank) en la carpeta “files” y retorna la secuencia en formato SeqRecord.

Parámetros

accession_number – El accession number del archivo en GenBank, por ejemplo NG_008617.1

Devuelve

La secuencia en formato SeqRecord con todas sus anotaciones.

`descarga.main()`

Ejecutar directamente el archivo “descarga” requiere entregar el parámetro `–accessionnumber` para descargar una secuencia mediante la función `accnun_to_seqrecord`.

Parámetros

--accessionnumber – El accession number del archivo en GenBank, por ejemplo NG_008617.1

`descarga.parse_file_to_seqrecord(filepath)`

Función que recibe la ruta de un archivo genético anotado y lo retorna como SeqRecord. Genera una copia del archivo en la carpeta files. Solo se aceptan archivos en formato «GenBank», o sea “.gbk” y “.gb”. Estos archivos pueden estar comprimidos en formato “.gz”.

Parámetros

filepath – La ruta donde se encuentra el archivo que se quiere utilizar

Devuelve

La secuencia en formato SeqRecord con todas sus anotaciones.

1.2 disen module

El módulo diseño contiene las funciones que permiten ejecutar el algoritmo diseñador y multiplexador de sondas. El diseño de las sondas de hibridación requiere de cierto procesamiento de datos para poder obtener el resultado adecuado, desde obtener las regiones de interés hasta verificar la validez de las sondas.

Este módulo está desarrollado para ser importado e utilizado por una interfaz de usuario.

`diseno.check_probe(seq, iscentral, minlen, maxlen, tmin, tmax, gmin, gmax, dgmin_homodim, dgmin_hairpin, maxhomopol_simple, maxhomopol_double, maxhomopol_triple)`

Función que recibe una sonda (secuencia de nucleótidos) y las restricciones para verificar que la sonda cumpla los parámetros establecidos. Retorna un valor booleano que determina si cumple las restricciones o no.

Parámetros

- **seq** – La secuencia que se quiere verificar.
- **iscentral** – Booleano que indica si la sonda es central (una mitad en un exón y la otra en el siguiente)
- **minlen** – El largo mínimo de la sonda.
- **maxlen** – El largo máximo de la sonda.
- **tmin** – Temperatura de melting mínima.
- **tmax** – Temperatura de melting máxima.
- **gmin** – Porcentaje de GC mínimo.
- **gmax** – Porcentaje de GC máximo.
- **dgmin_homodim** – Delta G mínimo permitido para validación de homodimerización.
- **dgmin_hairpin** – Delta G mínimo permitido para validación de hairpin u horquilla.
- **maxhomopol_simple** – Cantidad máxima permitida de homopolímeros simples (AAAAA)
- **maxhomopol_double** – Cantidad máxima permitida de homopolímeros dobles (AGAGA-GAG)
- **maxhomopol_triple** – Cantidad máxima permitida de homopolímeros triples (AGCAG-CAGC)

Devuelve

Booleano que determina la validez de la secuencia

```
diseno.generate_xlsx(df, name, genes, minlen=60, maxlen=120, tmmin=65, tmmax=80, gcmín=30, gcmax=70,
mindist=0, maxdist=50, minoverlap=25, maxoverlap=50, dgmin_homodim=-10000,
dgmin_hairpin=-10000, maxhomopol_simple=6, maxhomopol_double=5,
maxhomopol_triple=4, multiplex=True, mindg=-13627, maxdt=5, tiempo_diseno=0)
```

Función que genera un reporte excel que contiene todas las sondas generadas y sus características. Además muestra las restricciones iniciales y los grupos de multiplexación. Almacena el reporte en la carpeta “sondas”

Parámetros

- **df** – DataFrame de pandas que contiene las sondas y sus parámetros.
- **name** – Nombre que se le quiere dar al reporte.
- **minlen** – El largo mínimo de la sonda.
- **maxlen** – El largo máximo de la sonda.
- **tmmin** – Temperatura de melting mínima.
- **tmmax** – Temperatura de melting máxima.
- **gcmín** – Porcentaje de GC mínimo.
- **gcmax** – Porcentaje de GC máximo.
- **mindist** – Distancia mínima al borde del exón.
- **maxdist** – Distancia máxima al borde del exón.
- **minoverlap** – Sobrelape mínimo entre sondas.
- **maxoverlap** – Sobrelape máximo entre sondas.
- **dgmin_homodim** – Delta G mínimo permitido para validación de homodimerización.
- **dgmin_hairpin** – Delta G mínimo permitido para validación de hairpin u horquilla.
- **maxhomopol_simple** – Cantidad máxima permitida de homopolímeros simples (AAAAA)
- **maxhomopol_double** – Cantidad máxima permitida de homopolímeros dobles (AGAGA-GAG)
- **maxhomopol_triple** – Cantidad máxima permitida de homopolímeros triples (AGCAG-CAGC)
- **multiplex** – Booleano que indica si el diseño considera multiplexación
- **mindg** – Delta G mínimo para heterodimerización (multiplex)
- **maxdt** – Diferencia máxima de Tm (multiplex)
- **tiempo_diseno** – Tiempo que demoró la ejecución en segundos

Devuelve

Nombre del archivo XLSX generado

```
diseno.get_all_genes(seqrecord)
```

Obtener todos los genes anotados

Parámetros

seqrecord – Subregiones que representan la secuencia codificante o secuencia de exones

Devuelve

Lista con pares de posiciones en una tupla: el fin de un exón y el inicio del siguiente.

`diseño.get_all_transcripts(seqrecord)`

Función que retorna todos los transcritos anotados dentro de una secuencia.

Parámetros

seqrecord – Subregiones que representan la secuencia codificante o secuencia de exones

Devuelve

Lista con pares de posiciones en una tupla: el fin de un exón y el inicio del siguiente.

`diseño.get_probes_from_pos(empalme, record, site, minlen, maxlen, tmmin, tmmax, gcmin, gcmx, mindist, maxdist, minoverlap, maxoverlap, dgmin_homodim, dgmin_hairpin, maxhomopol_simple, maxhomopol_double, maxhomopol_triple)`

Esta función genera un par de sondas a partir de la secuencia de referencia y dos posiciones que representan el punto de empalme entre dos exones. El par de sondas generadas pueden ubicarse en el exón donador, acceptor o en ambos (central).

Parámetros

- **empalme** – Tupla con posiciones que representan al punto de empalme o splicing.
- **record** – Secuencia de referencia.
- **site** – Entero que representa si las sondas son donador, acceptor o central.
- **minlen** – El largo mínimo de la sonda.
- **maxlen** – El largo máximo de la sonda.
- **tmmin** – Temperatura de melting mínima.
- **tmmax** – Temperatura de melting máxima.
- **gcmin** – Porcentaje de GC mínimo.
- **gcmx** – Porcentaje de GC máximo.
- **mindist** – Distancia mínima al borde del exón.
- **maxdist** – Distancia máxima al borde del exón.
- **minoverlap** – Sobrelape mínimo entre sondas.
- **maxoverlap** – Sobrelape máximo entre sondas.
- **dgmin_homodim** – Delta G mínimo permitido para validación de homodimerización.
- **dgmin_hairpin** – Delta G mínimo permitido para validación de hairpin u horquilla.
- **maxhomopol_simple** – Cantidad máxima permitida de homopolímeros simples (AAAAA)
- **maxhomopol_double** – Cantidad máxima permitida de homopolímeros dobles (AGAGA-GAG)
- **maxhomopol_triple** – Cantidad máxima permitida de homopolímeros triples (AGCAG-CAGC)

Devuelve

Lista con dos sondas, cada sonda es una tupla con la secuencia, la posición de inicio, la posición final y la distancia al borde del exón si aplica.

`diseño.get_splicing_pairs(locations)`

Función que obtiene las zonas de empalme a partir de una serie de regiones que representan una transcripción. Las zonas de empalme se representan por dos posiciones: el fin de un exón y el inicio del siguiente.

Parámetros

locations – Subregiones que representan la secuencia codificante o secuencia de exones

Devuelve

Lista con pares de posiciones en una tupla: el fin de un exón y el inicio del siguiente.

`diseno.get_splicings(seqrecord, transcripts)`

Función que retorna los empalmes sin repetición de una secuencia anotada en ciertos transcritos.

Parámetros

- **seqrecord** – Secuencia anotada de referencia
- **transcripts** – Lista de transcritos

Devuelve

Lista con pares de posiciones en una tupla: el fin de un exón y el inicio del siguiente.

`diseno.main()`

`diseno.probe_designer(record, transcripts, progreso_queue=None, minlen=60, maxlen=120, tmmin=65, tmmax=80, gcm=30, gcmax=70, mindist=0, maxdist=200, minoverlap=25, maxoverlap=50, dgmin_homodim=-10000, dgmin_hairpin=-10000, maxhomopol_simple=6, maxhomopol_double=5, maxhomopol_triple=4, ismultiplex=True, mindg=-13627, maxdt=5)`

Función principal diseñadora de sondas. De la secuencia anotada se obtiene una serie de sondas como subsecuencias de ésta, que cumplen varias restricciones ajustadas por los parámetros.

Parámetros

- **record** – Secuencia anotada de referencia para la creación de las sondas.
- **transcripts** – Transcripciones en las cuales se quiere diseñar.
- **minlen** – El largo mínimo de la sonda.
- **maxlen** – El largo máximo de la sonda.
- **tmmin** – Temperatura de melting mínima.
- **tmmax** – Temperatura de melting máxima.
- **gcm** – Porcentaje de GC mínimo.
- **gcmax** – Porcentaje de GC máximo.
- **mindist** – Distancia mínima al borde del exón.
- **maxdist** – Distancia máxima al borde del exón.
- **minoverlap** – Sobrelape mínimo entre sondas.
- **maxoverlap** – Sobrelape máximo entre sondas.
- **dgmin_homodim** – Delta G mínimo permitido para validación de homodimerización.
- **dgmin_hairpin** – Delta G mínimo permitido para validación de hairpin u horquilla.
- **maxhomopol_simple** – Cantidad máxima permitida de homopolímeros simples (AAAAA)
- **maxhomopol_double** – Cantidad máxima permitida de homopolímeros dobles (AGAGAGAG)
- **maxhomopol_triple** – Cantidad máxima permitida de homopolímeros triples (AGCAGCAGC)
- **ismultiplex** – Booleano que indica si el diseño considera multiplexación
- **mindg** – Delta G mínimo para heterodimerización (multiplex)

- **maxdt** – Diferencia máxima de Tm (multiplex)

Devuelve

Dataframe con las sondas y sus características

`diseño.segundos_a_hms(segundos)`

`diseño.verify_specificity(seq, iscentral)`

Función que verifica la especificidad o unicidad de una secuencia usando la herramienta de alineamiento BLAST de manera local. Es necesario tener instalado Blast+ y las bases de datos en la carpeta “refseq”.

Parámetros

- **seq** – Secuencia de nucleótidos que se quiere validar.
- **iscentral** – Booleano que indica si se trata de una sonda central. Se utilizan diferentes criterios para chequear la especificidad de una sonda central.

Devuelve

Booleano que indica si la secuencia se considera específica.

1.3 multiplex module

El módulo multiplex contiene las funciones que utiliza el módulo diseñador para poder llevar a cabo la multiplexación de las sondas diseñadas.

`multiplex.are_sequences_compatible(seq1, seq2, mindg, maxdt)`

Esta función determina si dos secuencias de ADN son compatibles. Debes personalizar las condiciones de compatibilidad según tus necesidades.

Parámetros

- **seq1** – Primera secuencia de ADN.
- **seq2** – Segunda secuencia de ADN.

Devuelve

True si son compatibles, False en caso contrario.

`multiplex.multiplex_sequences(sequences, progress_queue, mindg=-13627, maxdt=5)`

Función principal de multiplexación de sondas. Recibe las secuencias, los parámetros de multiplexación y la cola para informar progreso.

Parámetros

- **sequences** – Lista con todas las diferentes secuencias.
- **progress_queue** – Cola para informar el progreso a la interfaz.
- **mindg** – Mínimo delta G heterodimerización para que dos secuencias sean compatibles.
- **maxdt** – Máxima diferencia de Tm para que dos secuencias sean compatibles.

Devuelve

True si son compatibles, False en caso contrario.

1.4 probeDesignerUI module

El módulo probeDesignerUI despliega la interfaz de usuario para utilizar el diseñador de sondas.

class probeDesignerUI.AutoScrollbar(*master=None, **kw*)

Bases: Scrollbar

A scrollbar that hides itself if it's not needed. Works only if you use the grid geometry manager

pack(***kw*)

Pack a widget in the parent widget. Use as options: *after=widget* - pack it after you have packed widget
anchor=NSEW (or subset) - position widget according to

given direction

before=widget - pack it before you will pack widget *expand=bool* - expand widget if parent size grows
fill=NONE or X or Y or BOTH - fill widget if widget grows *in=master* - use master to contain this widget
in_=master - see "in" option description *ipadx=amount* - add internal padding in x direction *ipady=amount*
- add internal padding in y direction *padx=amount* - add padding in x direction *pady=amount* - add padding
in y direction *side=TOP* or *BOTTOM* or *LEFT* or *RIGHT* - where to add this widget.

place(***kw*)

Place a widget in the parent widget. Use as options: *in=master* - master relative to which the widget is
placed *in_=master* - see "in" option description *x=amount* - locate anchor of this widget at position x of
master *y=amount* - locate anchor of this widget at position y of master *relx=amount* - locate anchor of this
widget between 0.0 and 1.0

relative to width of master (1.0 is right edge)

rely=amount - locate anchor of this widget between 0.0 and 1.0

relative to height of master (1.0 is bottom edge)

anchor=NSEW (or subset) - position anchor according to given direction *width=amount* - width of this
widget in pixel *height=amount* - height of this widget in pixel *relwidth=amount* - width of this widget
between 0.0 and 1.0

relative to width of master (1.0 is the same width as the master)

relheight=amount - height of this widget between 0.0 and 1.0

relative to height of master (1.0 is the same height as the master)

bordermode=»inside» or «outside» - whether to take border width of

master widget into account

set(*lo, hi*)

Set the fractional values of the slider position (upper and lower ends as value between 0 and 1).

class probeDesignerUI.ControladorApp(*root*)

Bases: object

mostrar_pantalla(*nombre, seqrecord=None, transcripciones=None, dict_params=None, df=None,*
filepath=None, folderpath=None, df1=None, df2=None, df3=None, genes=None,
accnum=None)

class probeDesignerUI.DataTable(*master*)

Bases: Frame

```
    create_table()

    mostrar_gen(gen)

    on_button_click_1(row)

    on_button_click_2(row)

class probeDesignerUI.IntegerSelector(master, default_value=0, min_value=None, max_value=None,
                                      increment=1, **kwargs)

    Bases: CTKFrame

    aumentar()

    disminuir()

    validar(nuevo_valor)

class probeDesignerUI.PantallaCarga(root, seqrecord, transcripciones, dict_params, filepath)

    Bases: PantallaInicial

    actualizar_progreso(progreso_queue, tarea_thread, multiplex)

    crear_interfaz()

    ejecutar_diseno(progreso_queue)

class probeDesignerUI.PantallaFinal(root, seqrecord, transcripciones, df, dict_params, filepath,
                                    folderpath, df1, df2, df3)

    Bases: PantallaInicial

    crear_interfaz()

    volver_a_inicio()

class probeDesignerUI.PantallaHistorial(root)

    Bases: PantallaInicial

    crear_interfaz()

    volver_a_seleccion()

class probeDesignerUI.PantallaInicial(root)

    Bases: object

    crear_interfaz()

    seleccionar_archivo()

    ver_historial()

class probeDesignerUI.PantallaParametros(root, seqrecord, transcripciones, filepath)

    Bases: PantallaInicial

    crear_interfaz()

    volver_a_transcripciones()

class probeDesignerUI.PantallaRegistro(root, folderpath, genes, accnum)

    Bases: PantallaInicial
```

```

    crear_interfaz()

    volver_a_inicio()

class probeDesignerUI.PantallaTranscripciones(root, seqrecord, filepath)
    Bases: PantallaInicial

    crear_interfaz()

    eliminar()

    eliminar_todo()

    seleccionar()

    seleccionar_todo()

    volver_a_seleccion()

class probeDesignerUI.ScrollableFrame(parent, *args, **kwargs)
    Bases: Frame

    resize_canvas(event)

class probeDesignerUI.ToolTip(widget, text, image_path=None)
    Bases: object

    mostrar_tooltip(event)

    ocultar_tooltip(event)

class probeDesignerUI.Zoom_Advanced(mainframe, path)
    Bases: Frame

    Advanced zoom of the image

    move_from(event)
        Remember previous coordinates for scrolling with the mouse

    move_to(event)
        Drag (move) canvas to the new position

    scroll_x(*args, **kwargs)
        Scroll canvas horizontally and redraw the image

    scroll_y(*args, **kwargs)
        Scroll canvas vertically and redraw the image

    show_image(event=None)
        Show image on the Canvas

    wheel(event)
        Zoom with mouse wheel

```


CAPÍTULO 2

Índices y tablas

- genindex
- modindex
- search

d

descarga, [1](#)

diseno, [2](#)

m

multiplex, [6](#)

p

probeDesignerUI, [7](#)

A

accnum_to_seqrecord() (en el módulo descarga), 1
 actualizar_progreso() (método de probeDesignerUI.PantallaCarga), 8
 are_sequences_compatible() (en el módulo multiplex), 6
 aumentar() (método de probeDesignerUI.IntegerSelector), 8
 AutoScrollbar (clase en probeDesignerUI), 7

C

check_probe() (en el módulo diseno), 2
 ControladorApp (clase en probeDesignerUI), 7
 crear_interfaz() (método de probeDesignerUI.PantallaCarga), 8
 crear_interfaz() (método de probeDesignerUI.PantallaFinal), 8
 crear_interfaz() (método de probeDesignerUI.PantallaHistorial), 8
 crear_interfaz() (método de probeDesignerUI.PantallaInicial), 8
 crear_interfaz() (método de probeDesignerUI.PantallaParametros), 8
 crear_interfaz() (método de probeDesignerUI.PantallaRegistro), 8
 crear_interfaz() (método de probeDesignerUI.PantallaTranscripciones), 9
 create_table() (método de probeDesignerUI.DataTable), 7

D

DataTable (clase en probeDesignerUI), 7
 descarga
 module, 1
 diseno
 module, 2
 disminuir() (método de probeDesignerUI.IntegerSelector), 8

E

ejecutar_diseno() (método de probeDesignerUI.PantallaCarga), 8
 eliminar() (método de probeDesignerUI.PantallaTranscripciones), 9
 eliminar_todo() (método de probeDesignerUI.PantallaTranscripciones), 9

G

generate_xlsx() (en el módulo diseno), 2
 get_all_genes() (en el módulo diseno), 3
 get_all_transcripts() (en el módulo diseno), 3
 get_probes_from_pos() (en el módulo diseno), 4
 get_splicing_pairs() (en el módulo diseno), 4
 get_splicings() (en el módulo diseno), 5

I

IntegerSelector (clase en probeDesignerUI), 8

M

main() (en el módulo descarga), 1
 main() (en el módulo diseno), 5
 module
 descarga, 1
 diseno, 2
 multiplex, 6
 probeDesignerUI, 7
 mostrar_gen() (método de probeDesignerUI.DataTable), 8
 mostrar_pantalla() (método de probeDesignerUI.ControladorApp), 7
 mostrar_tooltip() (método de probeDesignerUI.ToolTip), 9
 move_from() (método de probeDesignerUI.Zoom_Advanced), 9
 move_to() (método de probeDesignerUI.Zoom_Advanced), 9
 multiplex
 module, 6

`multiplex_sequences()` (en el módulo `multiplex`), 6

O

`ocultar_tooltip()` (método de `probeDesignerUI.ToolTip`), 9

`on_button_click_1()` (método de `probeDesignerUI.DataTable`), 8

`on_button_click_2()` (método de `probeDesignerUI.DataTable`), 8

P

`pack()` (método de `probeDesignerUI.AutoScrollbar`), 7

`PantallaCarga` (clase en `probeDesignerUI`), 8

`PantallaFinal` (clase en `probeDesignerUI`), 8

`PantallaHistorial` (clase en `probeDesignerUI`), 8

`PantallaInicial` (clase en `probeDesignerUI`), 8

`PantallaParametros` (clase en `probeDesignerUI`), 8

`PantallaRegistro` (clase en `probeDesignerUI`), 8

`PantallaTranscripciones` (clase en `probeDesignerUI`), 9

`parse_file_to_seqrecord()` (en el módulo `descarga`), 2

`place()` (método de `probeDesignerUI.AutoScrollbar`), 7

`probe_designer()` (en el módulo `diseño`), 5

`probeDesignerUI`
module, 7

R

`resize_canvas()` (método de `probeDesignerUI.ScrollableFrame`), 9

S

`scroll_x()` (método de `probeDesignerUI.Zoom_Advanced`), 9

`scroll_y()` (método de `probeDesignerUI.Zoom_Advanced`), 9

`ScrollableFrame` (clase en `probeDesignerUI`), 9

`segundos_a_hms()` (en el módulo `diseño`), 6

`seleccionar()` (método de `probeDesignerUI.PantallaTranscripciones`), 9

`seleccionar_archivo()` (método de `probeDesignerUI.PantallaInicial`), 8

`seleccionar_todo()` (método de `probeDesignerUI.PantallaTranscripciones`), 9

`set()` (método de `probeDesignerUI.AutoScrollbar`), 7

`show_image()` (método de `probeDesignerUI.Zoom_Advanced`), 9

T

`ToolTip` (clase en `probeDesignerUI`), 9

V

`validar()` (método de `probeDesignerUI.IntegerSelector`), 8

`ver_historial()` (método de `probeDesignerUI.PantallaInicial`), 8

`verify_specificity()` (en el módulo `diseño`), 6

`volver_a_inicio()` (método de `probeDesignerUI.PantallaFinal`), 8

`volver_a_inicio()` (método de `probeDesignerUI.PantallaRegistro`), 9

`volver_a_seleccion()` (método de `probeDesignerUI.PantallaHistorial`), 8

`volver_a_seleccion()` (método de `probeDesignerUI.PantallaTranscripciones`), 9

`volver_a_transcripciones()` (método de `probeDesignerUI.PantallaParametros`), 8

W

`wheel()` (método de `probeDesignerUI.Zoom_Advanced`), 9

Z

`Zoom_Advanced` (clase en `probeDesignerUI`), 9