Documentación de Software web FungiRegEx



This guide can be useful if you want to use the software.

Nombre del software: FungiRegEx

Versión: 1.0.0

Fecha: 🖰 Aug 10, 2022

TABLE OF CONTENTS

- Pasos previos
- · Resumen del software
- Requisitos para ejecutar el software(Recomendados)
- Configuración inicial
 - Servidor local/Computadora local/Servidor externo
 - Computadora propia
- Configuración del cluster
- Sobre Expresiones Regulares =
- Guía de uso
 - Paso 1 Abriendo consolas y ejecutando comandos para iniciar la aplicación ...
 - Paso 3 Abre la aplicación web
 - Paso 4 Usa FungiRegEx
 - Paso 5 Escribe la expresión regular
 - Paso 6 Búsqueda
- Tutorials
- FAQs



- Question
- Question
- Question
- Additional Resources
- Support

Pasos previos

Antes de ejecutar el software asegurate de cumplir los prerequisitos y de ejecutar los siguientes comandos:

- 1. Recibes la carpeta con el codigo completo, por tanto tu deberás instalar lo que requiere el software para funcionar.
- 2. Primero es necesario instalar las dependencias del proyecto, para ello ejecuta el siguiente comando:
 - 1 npm i
- 3. Posteriormente instala nodemon:
 - 1 npm install -g nodemon
- 4. Instala chromium, revisa la documentación de chromium para verificar pasos detallados acorde al sistema operativo:

```
choco install googlechrome --version=79.0.3945.117
```

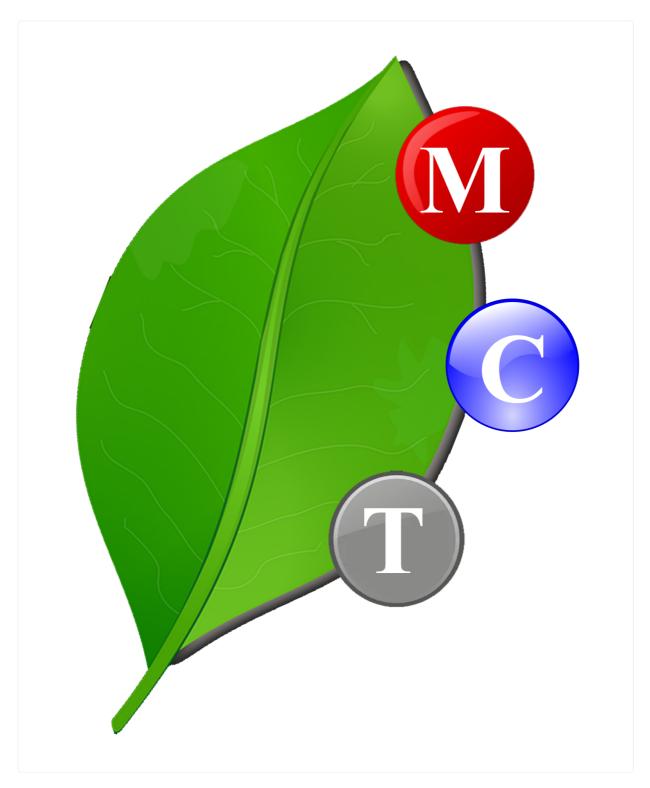
Despues de ejecutar los comandos estas listo para ejecutar el proyecto.

Resumen del software

Esta herramienta es una herramienta web de busqueda de expresiones regulares en proteomas, toda la información se obtiene de JGI (Joinmt Genome Institute) en su base de datos, que por medio de un scraper para todas las especies disponibles se encarga de obtener la cadena proteomica, cabe resaltar que JGI solo tiene información de organizmos de la faminia fungi.

En esta versión se utiliza React JS para front-end y NodeJS + Expresspara el back-end.





Es importante aclarar que el numero de instancias de chromium que la computadora puede abrir es configurable y debe estar acorde a los recursos de la computadora donde se ejecute la aplicación.

Requisitos para ejecutar el software(Recomendados) 🤞

El software ha sido desarrollado y probado en una computadora con las siguientes características:

- 4 GB RAM
- Cent OS 7 y Windows 10 como sistemas operativos
- Procesador Core i7 5th

Sobre las características de software:

- Node JS 16.17.0v
- Chromium 79.0.3945.117v
- React JS 17.0.2v

El proyecto contiene todas las librerias necesarias para desplegar en servidor si se requiere, solo es necesario ejecutar los siguientes comandos en la consola de tu sistema operativo dentro de la carpeta del proyecto.

- npm run start:frontend
- npm run start:backend

Configuración inicial 🍥

Puedes desplegar esta aplicación tanto local como en un servidor si es necesario, esto esta a consideración del usuario de la aplicación y sus necesidades dependiendo de lso recursos computacionales disponibles.

Tome en consideración que si multiples usuarios se conectan al mismo tiempo, se interrumpiran las tareas uno a otro, por tanto la aplicacion es para uso de un usuario a la vez., si el otro usuario encola tareas a la aplicaicon mientras otro usuario esta ejecutando otras se dara prioridad a las del ultimo usuario y se eliminaran las del anterior usuario.

Servidor local/Computadora local/Servidor externo

En caso de que despliegues la aplicación usando un servidor local (que puede ser otra computadora)Sigue las siguientes instrucciones:

- 1. Obten la dirección IP de la computadora.
 - a. Para Windows sigue las siguientes instrucciones:
 - i. https://support.microsoft.com/en-us/windows/find-your-ip-address-inwindows-f21a9bbc-c582-55cd-35e0-73431160a1b9
 - b. Para Linux:



- i. Ubuntu: https://help.ubuntu.com/stable/ubuntu-help/net-findip.html.en
- ii. Cent OS / Red Hat: https://access.redhat.com/documentation/es-es/red_hat_enterprise_linux/7/html/migration_planning_guide/ch04s04s07
- iii. Para otras distribuciones busca información en el sitio oficial
- 2. Navega por el sistema de archivos hasta que encuentres la carpeta raíz que conteiene el proyecto.
- 3. Como la aplicación se ejecuta en una computadora externa se requiere cambiar parametros de la aplicacion relacionados con la IP de la maquina donde se ejecutará. Para ello necesitamos cambiar los siguientes archivos:

```
/../interfacebio/src/components/main/Main.jsx
/../interfacebio/src/components/table/Table.jsx
```

Para el archivo Main.jsx tienes que cambiar la dirección IP en las líneas: 14,117 y 15, por ejemplo, suponga que la computadora externa o servidor externo tiene la IP 555.811.22.33 el código que puedes encontrar en la aplicación tenga lo siguiente:

```
const [products, setProducts] = useState([])
const [progressVal, setProgressVal] = useState(0)
const [blocked, setBlocked] = useState(false);
const endpoint = "http://192.168.1.102:8000/file" //line 14
const progress = "http://192.168.1.102:8000/progress"//line
```

Tienes que cambair los valores a:

```
const [products, setProducts] = useState([])
const [progressVal, setProgressVal] = useState(0)

const [blocked, setBlocked] = useState(false);

const endpoint = "http://555.811.22.33:8000/file"//line 14

const progress = "http://555.811.22.33:8000/progress"//line

15
```

En la línea 117 encontrarás:

```
1 try {
2     await
    axios.post("http://192.168.1.102:8000/resultsScrap", {
3          stringtoBackend
4     })
5
```

Written in **slite**

```
6 }
```

Tienes que cambiar los valores a :

Para el archivo Table.jsx tienes que modificar la dirección IP en la línea 10, por ejemplo, suponga que la computadora externa tiene la IP 555.811.22.33, el código que puedes encontrar en la aplicación tiene lo siguiente:

```
const endpoint = "http://192.168.1.102:3000/listaURLnew.json"
```

Tienes que cambiar el valor de la dirección IP a:

```
const endpoint = "http://555.811.22.33:3000/listaURLnew.json"
```

Computadora propia

Tienes que obtener la dirección IP de tu computadora, y seguir los pasos de las secciones anteriores en la modificación de los archivos, en su defecto puedes sustituir la IP por un locahost.

Posteriormente solo es cambiar los valores de la IP a la adecuada.

La configuración del firewall depende de cada usuario y para esta aplicación si se ejeecutara localmente no es necesario la apertura de puertos a usuarios externos.

Configuración del cluster

El principio de funcionamiento de la aplicación se basa en abrir varias pestañas del navegador chromium, obtener la lista de URL y realizar solicitudes al servidor JGI. La cantidad de pestañas que abres en la computadora donde se ejecuta depende de este parámetro y determinará la cantidad de recursos a usar.



Teniendo en cuenta los requisitos anteriores, no utilizamos más de 50 instancias al realizar las pruebas con el software. Puede configurar este parámetro abriendo el archivo:

```
directorio_donde_descargaste_la_aplicación/interfacebio/backend.js
```

En este archivo verás el siguiente código:

maxConcurrency: 30. Este es el parámetro a modificar, recomendamos no revase las 50 instancias.

Cuanto mas grande sea la cantidad colocada en el parametro mayor cantidad de recursos tendrá la aplicación pero a su vez realizará más rapido la obtención de los datos.

Sobre Expresiones Regulares

Para hacer coincidir una sola "a" seguida de cero o más "b" seguidas de "c", usaría el patrón /ab*c/: el * después de "b" significa "O o más ocurrencias del elemento anterior ." En la cadena "cbbabbbbcdebc", este patrón coincidirá con la subcadena "abbbbc".

Si necesita usar cualquiera de los caracteres especiales literalmente (en realidad, buscando un "*", por ejemplo), debe escaparlo colocando una barra invertida delante de él. Por ejemplo, para buscar "a" seguido de "*" seguido de "b", usaría /a*b/ — la barra invertida "escapa" del "*", haciéndolo literal en lugar de especial.

Para obtener más información acerca de las expresiones regulares, consulte: https://developer.mozilla.org/en-US/docs/Web/JavaScript/Guide/Regular_Expressions

Solo recuerda que no tienes que escribir / al principio y al final, solo coloca los aminoácidos que deseas buscar.

```
1 Characters
2 x The character x
3 \\ The backslash character
```

```
The character with octal value 0n (0 <= n <= 7)
4
    \0n
    \0nn
             The character with octal value 0nn (0 <= n <= 7)
 5
             The character with octal value 0mnn (0 <= m <= 3, 0 <= n
 6
    \0mnn
    <= 7)
    \xhh
             The character with hexadecimal value 0xhh
7
8
    \uhhhh
            The character with hexadecimal value <code>0xhhhh</code>
9
    \t
             The tab character ('\u0009')
             The newline (line feed) character ('\u000A')
    \n
10
             The carriage-return character ('\u000D')
11
    \r
    \f
             The form-feed character ('\u000C')
12
             The alert (bell) character ('\u0007')
13
    ∖a
             The escape character ('\u001B')
14
    \e
             The control character corresponding to \boldsymbol{x}
15
    \cx
16
    Character classes
17
     [abc] a, b, or c (simple class)
18
     [^abc] Any character except a, b, or c (negation)
19
20
     [a-zA-Z]
                     a through z or A through Z, inclusive (range)
                     a through d, or m through p: [a-dm-p] (union)
     [a-d[m-p]]
21
    [a-z&&[def]]
                    d, e, or f (intersection)
22
    [a-z&&[^bc]]
                     a through z, except for b and c: [ad-z]
23
     (subtraction)
24
    [a-z\&\&[^m-p]] a through z, and not m through p: [a-lq-z]
     (subtraction)
25
    Predefined character classes
26
             Any character (may or may not match line terminators)
27
             A digit: [0-9]
    \d
28
            A non-digit: [^0-9]
    \ D
29
            A whitespace character: [ \t\n\x0B\f\r]
    \s
31
    \S
            A non-whitespace character: [^\s]
             A word character: [a-zA-Z_0-9]
32
    \w
33
    \ W
             A non-word character: [^\w]
34
    Boundary matchers
             The beginning of a line
             The end of a line
    $
             A word boundary
37
    \b
             A non-word boundary
    \B
             The beginning of the input
39
    \ A
             The end of the previous match
    \ G
40
```

Written in **slite**

```
\backslash Z
             The end of the input but for the final terminator, if any
41
42
     \backslash z
             The end of the input
43
     Greedy quantifiers
44
             X, once or not at all
45
46
     X *
             X, zero or more times
             X, one or more times
47
     Χ+
     X\{n\} X, exactly n times
48
     X\{n,\} X, at least n times
49
     X\{n,m\} X, at least n but not more than m times
51
     Reluctant quantifiers
52
             X, once or not at all
53
     X??
             X, zero or more times
54
     X*?
55
     X+?
             X, one or more times
     X{n}? X, exactly n times
56
     X\{n,\}? X, at least n times
57
58
     X{n,m}? X, at least n but not more than m times
59
60
     Possessive quantifiers
     X?+
            X, once or not at all
61
             X, zero or more times
62
     X * +
63
     X++ X, one or more times
     X{n}+ X, exactly n times
64
     X\{n,\}+ X, at least n times
65
     X\{n,m\}+X, at least n but not more than m times
67
     Logical operators
68
69
     XY
             X followed by Y
             Either X or Y
     X \mid Y
70
```

Guía de uso 🐣

Antes de ejecutar la aplicación asegurate de tener instalado en la computadora lo siguiente y que sea la version que indica, no se asegura el funcionamiento de la aplicación en otras versiones:

- Node JS 16.17.0v
- Navegador Chromium



Sino tienes instalado NodeJS puedes obtenerlo de: https://nodejs.org/en/ Sino tienes chromium instaldo puedes obtenerlo de: https://www.chromium.org/getting-involved/download-chromium/

O bien ejecutar los codigos desde la consola del sistema operativo indicados al inicio de la presente documentación.

Recuerda considerar el sistema operativo.

Paso 1 - Abriendo consolas y ejecutando comandos para iniciar la aplicación web

Ejecute los siguientes comandos para iniciar la aplicación, es necesario abrir consolas paralelas, una para el frontend y otra para el backend:

```
1 npm run start:frontend
```

Puedes ver la salida del primer comando en la Figura 1.

En la segunda consola, ejecuta el siguiente comando:

```
1 npm run start:backend
```

Puedes ver la salida del comando en la figura 2.

Para detener la aplicación presiona Ctrl+C en las consolas que hayas abierto.

Abre la consola(2 instancias):



Figura 1. Ejecutando el front-end

Figura 2. Ejecutando el back-end

Es importante ver la salida por que esta indicara si hay algun problema en la ejecución de la aplicación. Nota: El back-end no utiliza una versión optimizada de la aplicación react, si quieres desplegar una versión optimizada tienes que ejecutar el siguiente comando en la consola dentro de la carpeta donde está el proyecto.:

```
1 npm run build
```

Este comando generara el front-end optimizado.

Paso 3 - Abre la aplicación web

Después de los pasos 1 y 2 puedes ir a tu navegador, y en la barra de direcciones abrir localhost:3000 o la direccion IP de tu computadora.

En la Figura 1 se puede apreciar que la dirección donde se ejecuta el front-end es http://localhost:3000 pero ten en cuenta que puede ser http://localhost:3000 pero ten en cuenta que puede ser http://localhost:3000 o que puede cambiar a la dirección de cada computadora, solamente asegurese de que tienes la dirección correcta en los parámetros, si todo está bien al entrar a la dirección verás lo siguiente.



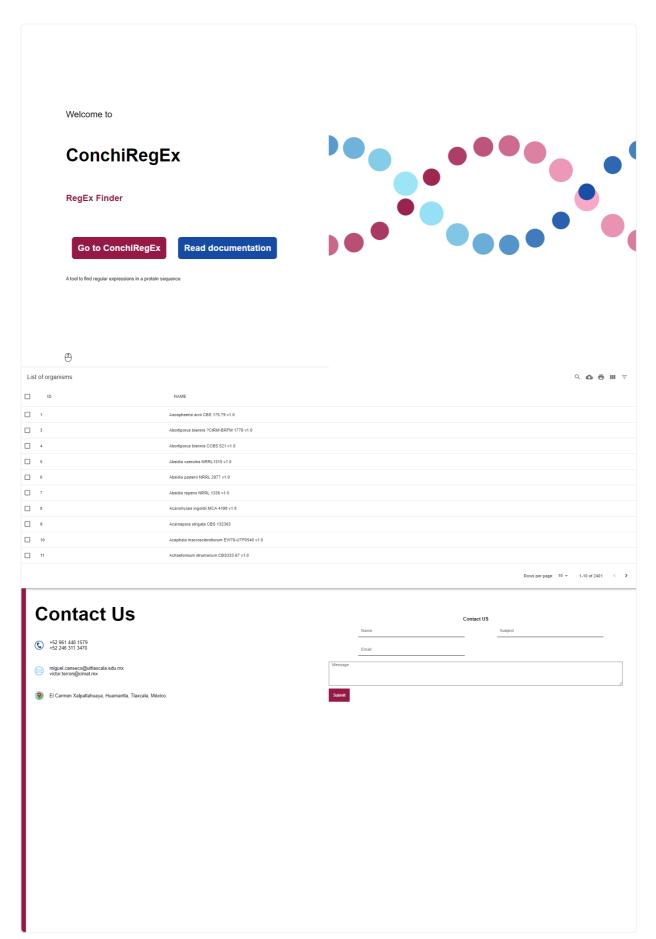


Figura 3. Interface de FungiRegEx

Podemos ver en la Figura 3 la interface que consta de 3 componentes principales:

- Presentación o landing page
- Tabla con las especies incluidas dentro de la version que al 10 de agosto de 2022 estaban disponibles en el portal del Joint Genome Institute
- Un formulario para que puedas contactarnos en caso de requerir soporte o algun comentario o sugerencia

Puedes buscar por el nombre cientifico de la especie, descargar la tabla completa o filtrar por especie.

Paso 4 - Usa FungiRegEx

Despues de los pasos 1, 2 y 3 puedes dar click sobre el boton y ejecutar la aplicación, verás lo siguiente.

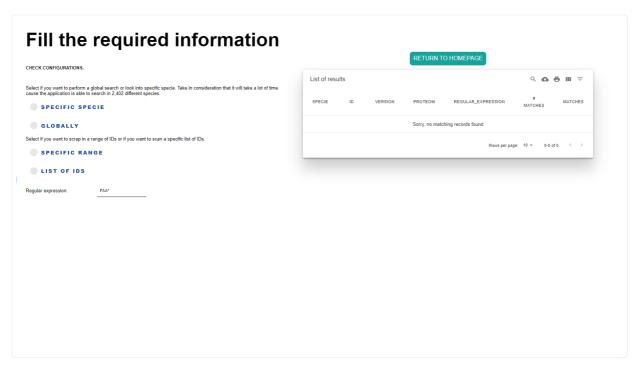


Figura 4. Interface principal

La interfaz principal consta de:

- Tipo de busqueda:
 - Globalmente: Si quieres hacer la busqueda de cierta cadena de aminoacidos en los proteomas de todas las especies

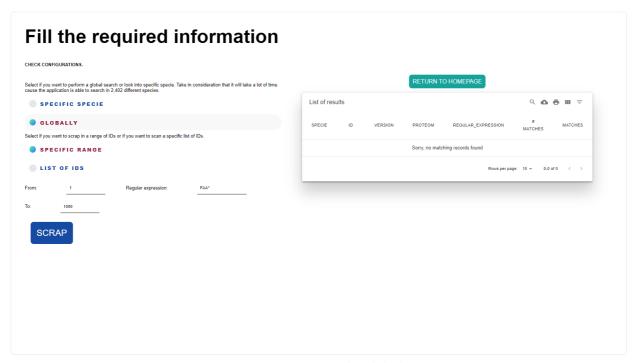


Figura 5. Búsqueda Global

 Especies específicas: Si quieres hacer una busqueda de la expresion regular colocada solo en cierta especie

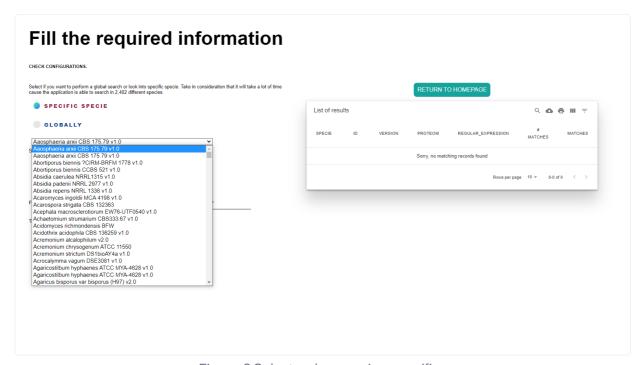


Figura 6. Selector de especie específica

- Rango o lista de Identificadores de proteinas donde realizar la busqueda:
 - Puedes colocar un rango de busqueda o bien colocar una lista específica de ID, identificadores en los que hacer busquedas.

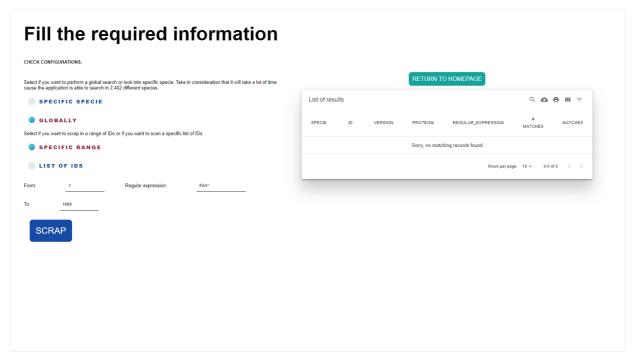


Figura 7. Rango de busqueda

Ene sta versión se incluyeron 2,402 especies unicamente de la familia fungi, mismas que estan disponibles en el selector.

Paso 5 - Escribe la expresión regular

Esta herrameitna se caracteriza por la facilidad de escribir expresiones regulares ya que pueden ser escritas con la secuencia del usuario quiera y tan complejas como las expresiones regulares lo permitan. Para aprender mas sobre las expresiones regulares puedes consultarlo en la sección de Expresiones regulares de la presente documentación, considera que la documentación esta disponible en inglés.

Por ejemplo, para buscar la expresión "FA" seguidas de cero o más "A" s, tienes que usar el patrón FAA*: el asterisco * después de "A" significa "O o mas coincidencias del caracter precedente".



Figure 8. Regular Expression

Paso 6 - Búsqueda

Después de los pasos anteriores puedes dar click sobre el botón de búsqueda, sino realizas los pasos anteriores probablemente la aplicación no funcione correctamente. Considere que al inicio de la ejecución la aplicación puede ser lento, tan pronto como la aplicación realice scrapping sobre la base de datos los datos se mostrarán en la tabla. La velocidad de la búsqueda depende de:

- Carga de los servidores de JGI
- Velocidad del internet del usuario
- Recursos de la computadora
- Configuración del cluster al abrir multiples instancias, este parámetro estpa disponible en el archivo de backend.js, para modificar este parametro consulta la sección "configuración del cluster"

Para este ejemplo, realicé la búsqueda de 1 a 1000 en los identificadores de la especie Trichoderma aethiopicum CBS130628 v1.0 usando la expresión regular FAA*. Los resultados tienen el siguiente orden:

- Especie
- ID
- Scaffold / Versión
- Proteoma
- # de coincidencias
- Coincidencias

Si la consulta al servidor de JGI con las especies y el identificador están vacios significa que el servidor de JGI no tiene información del proteoma especificado con el identificador o no devolvió una respuesta válida y en la tabla verás, NO_SHORTNAME, NO DATA.

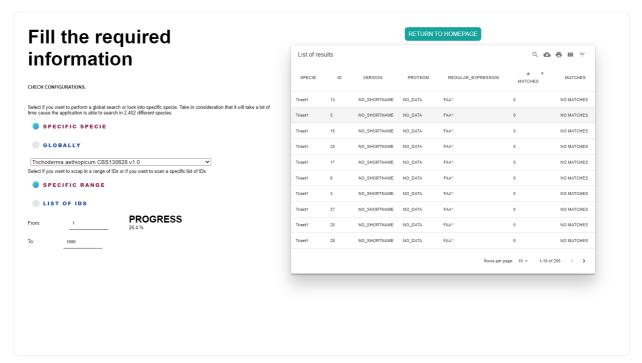


Figura 9. Resultados



Figura 10. Resultados

Durante el proceso de búsqueda, los resultados se mostrarán en la tabla, puedes ordenar por el numero de coincidencias tanto ascendente como descendentemente, igualmente la tabla contendrá el proteoma obtenido.

Ten en consideración que las peticiones al servidor de JGI en ocasiones no regresa información, así que si el resultado está vacio, la aplicación no mostrará esta información en la tabla. Considerando la anterior situación, en la anterior búsqueda de 1000 identificadores que se pasaron como parámetros de busqueda, solo se obtuvieron 995 resultados. Se aclara que esto no sucede confrecuencia y que este fallo no está relacionado con el funcionamiento o ptrogramación de esta herramienta sino mas bien con el servidor de JGI o falta de información en la base de datos de JGI.

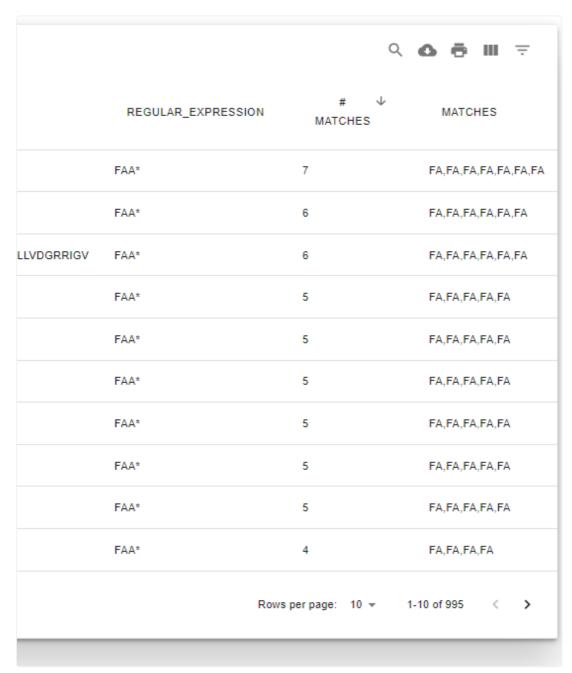
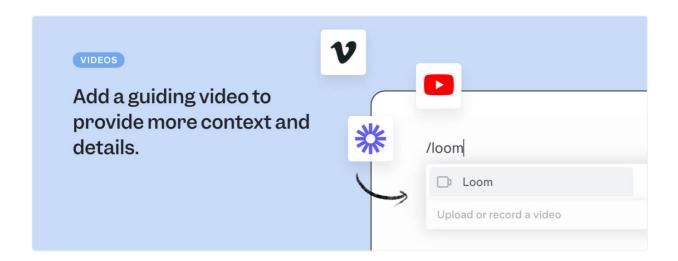


Figura 11. Resultados obtenidos 995 de 1000

Tutorials

Have any videos or knowledge bases to link if new users need additional guidance? Include them here.





Answer and document frequently asked questions below.

Question

Answer

Question

Answer

Question

Answer

Additional Resources 🖑

Include any additional information, forum or documentation that might be relevant to users here for easy access when questions arise.





Support 💬

This application was developed by Technotronic Engineer Victor Miguel Terron M. with advice from Ph.D. Miguel Angel Canseco Pérez. You can contact us:

- ☑ victor.terron@cimat.mx
- 🗹 miguel.canseco@uttlaxcala.edu.mx



Make it your own

Once edited to your liking, save this template to your team's templates list by clicking on the three dots on the right of the screen.

