**Manuscript title**

**RegulonDB v12: Redefiniendo conceptos y reconstrucción**

Heladia Salgado1,\*, Socorro Gama-Castro1, Citlalli Mejia-Almonte1, Paloma Lara-Figueroa1, Gabriel Alarcón-Carranza1, Andrés G. López-Almanzo1, Felipe Betancourt-Figueroa1, Pablo Peña-Loredo1, Shirley Alquicira-Hernández1, Daniela Ledezma-Tejeida1, Lizeth Arizmendi1, Paco2, Karen2, Elizabeth2, Jair S. García-Sotelo3, Víctor M. del Moral-Chávez1, Alfredo J. Hernández-Alvarez, Luis J. Muñiz-Rascado1, César Bonavides-Martínez1, Salvador Capella, Josep Lluis Gelpi-Buchaca, Fanny1, Alberto Santos-Zavaleta1, and Julio Collado-Vides1,4.

1 Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México

2 Instituto Tecnológico de Zacatepec, Zacatepec, Morelos, ####, México

3 Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Querétaro 76230, Querétaro, Mexico.

4 Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Universitat Pompeu Fabra(UPF), Barcelona, Spain.

\* To whom correspondence should be addressed. Tel:+52 777 3132063; Fax: [+52 777 3175581]; Email: [collado@ccg.unam.mx](mailto:collado@ccg.unam.mx)

Present Address: [Author name], Department, Institution, Town, State, Postcode, Country

**ABSTRACT**

The.

**INTRODUCTION**

**MATERIAL AND METHODS**

Backend

datamarts

graphQL lista de genes desde la caja de busqueda .. aceptan operadores lógicos

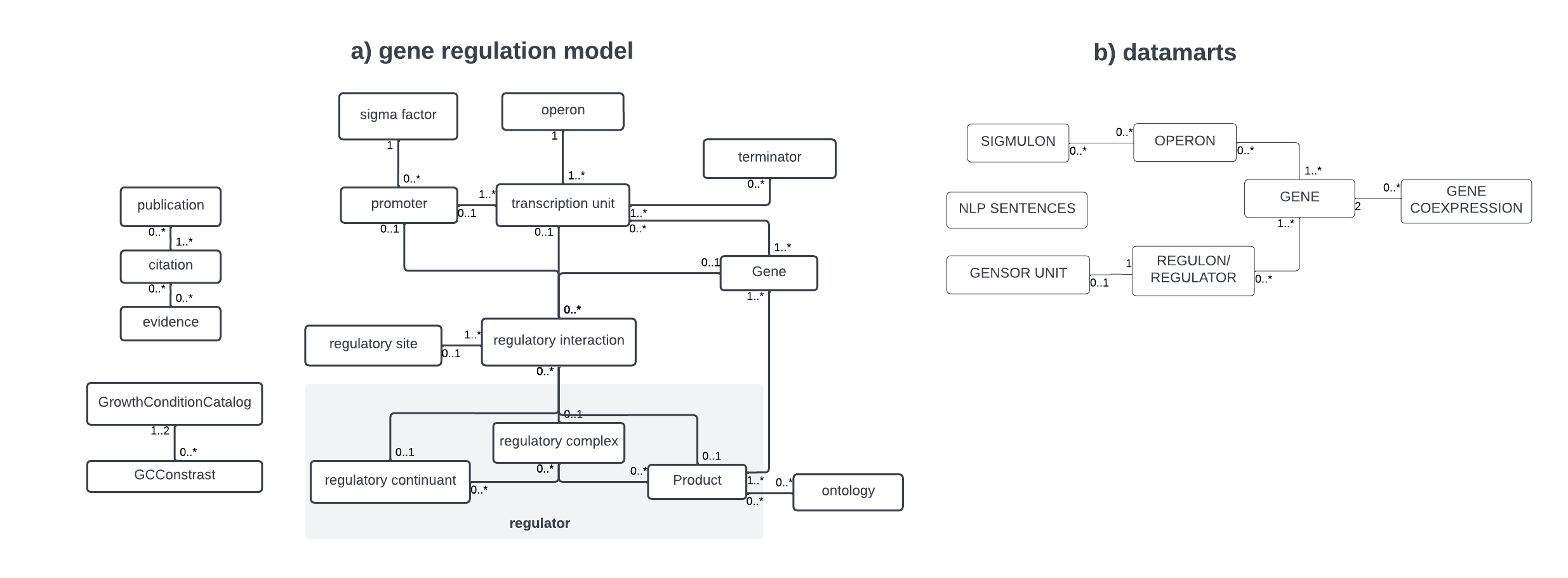
react

**RESULTS**

La regulación genética es un proceso complejo que juega un papel fundamental en la coordinación de la expresión de los genes en los organismos vivos. La comprensión de los mecanismos regulatorios y cómo se interrelacionan entre sí es esencial para desentrañar los circuitos genéticos y su impacto en la función celular. En este contexto, RegulonDB ha sido una herramienta invaluable para la comunidad científica en el estudio de la regulación génica en *Escherichia coli*, uno de los organismos modelo más utilizados en investigación. Desde su primera versión, RegulonDB ha evolucionado de manera significativa, incorporando nuevos avances en datos y tecnologías para proporcionar una visión más completa y actualizada de la regulación genética en esta bacteria.

En esta versión presentamos la reconstrucción y revisión completa de RegulonDB, abordando el modelo de datos subyacente y la interfaz web para brindar una experiencia más intuitiva y accesible a los usuarios. Para lograr esto, hemos llevado a cabo una exhaustiva revisión de los conceptos fundamentales que han evolucionado a lo largo de la historia de RegulonDB, revisando cómo se representan en el modelo actual y en la interfaz web.

En la versión 1.0 de RegulonDB, originalmente definimos esta base de datos como un repositorio de interacciones entre las proteínas reguladoras y sus sitios de unión asociados, así como la organización de las características reguladoras en operones y regulones. Si bien el objetivo principal de RegulonDB no ha cambiado, hemos realizado ajustes significativos para adaptarnos a los avances en datos y tecnologías en el campo de la regulación génica.

En esta nueva versión, RegulonDB ha sido enriquecido con datos de experimentos de secuenciación de próxima generación, permitiéndonos obtener una visión más detallada de los elementos reguladores, sus interacciones y sus efectos en la expresión génica. Además, hemos mejorado la visualización de los resultados, proporcionando herramientas interactivas que facilitan la exploración y el análisis de la regulación genética en *E. coli.* Figura 1: Diagrama de clases de la regulación genética y los datamarts de RegulonDB. La figura 1a presenta el modelo de clases de la regulación genética en RegulonDB. En este diagrama, cada caja representa una entidad o concepto biológico relevante para la regulación genética, y las líneas entre ellas indican las relaciones de cardinalidad. La notación "1..\*" representa una relación de uno o más elementos, mientras que "0..\*" indica una relación de cero o más elementos.Por otro lado, la figura 1b muestra el diagrama de clases de los datamarts de RegulonDB. Estos datamarts se utilizan para integrar una o más entidades biológicas del modelo de clases presentado en la figura 1a.

**Gene y productos**

Comenzaremos nuestra revisión de conceptos con los genes, que son la entidad biológica fundamental en nuestro modelo. En nuestra primera publicación, no definimos explícitamente el concepto, pero se entiende como una región del ADN que codifica para un producto. A lo largo del tiempo, hemos mantenido las propiedades básicas de gene como la posición en el genoma, el strand y la secuencia, así como los sinónimos asociados. Gracias a nuestra estrecha colaboración con Ecocyc, hemos logrado incorporar anotaciones y nuevas propiedades de los genes. Algunas correcciones realizadas en los datos incluyen la actualización de las coordenadas de los genes a través de las nuevas liberaciones del genoma de *E. coli*, la anotación de genes fantasma o pseudogenes, y la anotación de genes fragmentados.

En la primera versión de RegulonDB, se incluyó una tabla llamada "polypeptides" que contenía únicamente los productos que codificaban para proteínas reguladoras. En la versión 3.0, motivados por proporcionar más información sobre los genes regulados, se agregaron todos los productos identificados en el genoma, clasificándolos en tres categorías: polipéptidos reguladores, polipéptidos y RNAs. A partir de la versión 4.0 en adelante, todos los productos de genes estaban contenidos en una entidad llamada "products" con una propiedad que indicaba su tipo o categoría. Al igual que los genes, mantenemos los datos de los productos sincronizados con Ecocyc.

La relación entre un gen y su producto fue originalmente de 1 a 1. Sin embargo, se modificó esta relación para permitir que un gen pueda tener más de un producto, como en el caso de ihfB y otros genes que generan isoformas de proteínas debido a sitios de traducción internos o codones de inicio alternativos. En la actualidad, nuestra base de datos contiene un total de 4736 genes, como se muestra en la tabla 1, de los cuales 10 de ellos tienen más de un producto. Es importante destacar que el número de productos es menor debido a que los genes phantom y algunos pseudogenes no tienen productos asociados.

| **Biological Entity** | **Total** |
| --- | --- |
| **Gene** | 4736 |
| Pseudo gene  Phantom gene | 146  62 |
| **Product** | 4652 |
| polypeptides  tRNAs  tmRNA  rRNA  small RNA | 4432  86  1  22  111 |

Tabla 1. Total de documentos o registros en la colección de gene y productos.

Los datos de los genes y sus productos forman parte de la colección de Genes en los datamarts de RegulonDB. Se puede acceder a la información de estos objetos a través de la caja de búsqueda utilizando el identificador, nombre o sinónimos del gen o producto. Además, en el menú principal se encuentra la opción "Search", y al acceder a las opciones de “Browsing” se mostrará una lista de todos los genes con diversas opciones de filtrado, como se muestra en la figura 2.

La página de genes ha sido completamente rediseñada, con funciones mejoradas para una experiencia de navegación más fluida. Ahora, los usuarios pueden navegar por la página utilizando un menú de secciones intuitivo. La visualización gráfica se realiza a través de una herramienta llamada "drawing traces tool", que ha sido reconstruida en React para proporcionar una experiencia dinámica y ajustable. Con el menú de ajuste, los usuarios pueden controlar el nivel de detalle en la visualización, así como desplazarse por la región genómica de interés. Además, las secuencias de genes se presentan en diferentes formatos y ofrecen opciones de visualización personalizables. En el menú lateral derecho, los usuarios también tienen la posibilidad de descargar toda la información en formato PDF o JSON, según sus necesidades.

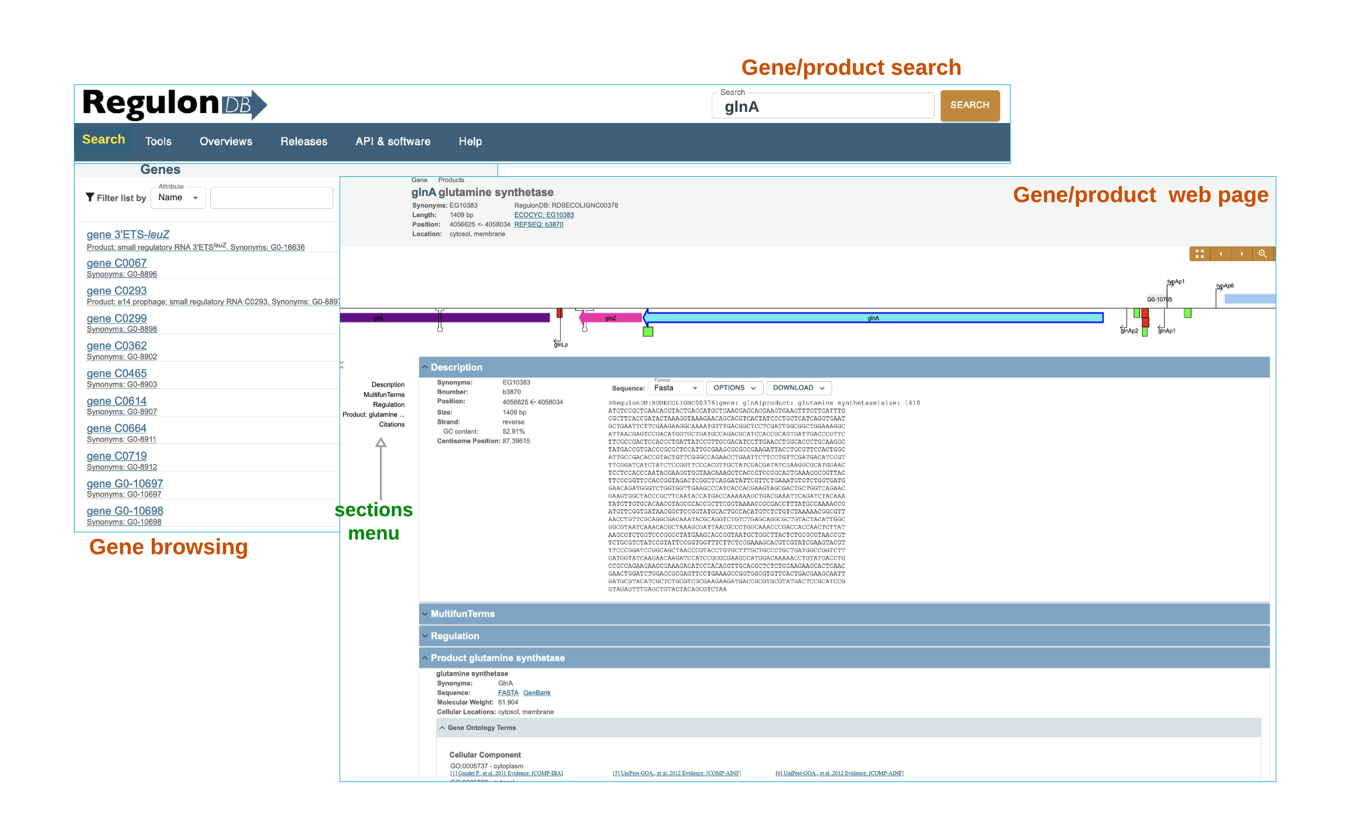


Fig2. Opciones de búsqueda y despliegue de gene y su producto.

**Promotores**

Los promotores son elementos esenciales para la transcripción, ya que determinan el inicio de la transcripción de un gen. En nuestro primer modelo de la base de datos, se incluyeron los promotores como elementos genéticos principales con propiedades específicas para su identificación, como el nombre del promotor, los sinónimos asociados, la hebra de ADN en la cual se encuentra ubicado, el tipo de factor sigma que se une a él, la posición de inicio de la transcripción en el genoma y el operón al que pertenece. Además de estas propiedades, se incluyeron características adicionales como la tasa de transcripción del promotor en ausencia de cualquier regulación, la constante de equilibrio para la unión de la ARN polimerasa al promotor y la constante cinética de la transición del complejo cerrado al complejo abierto.

Hemos avanzado en nuestro conocimiento sobre los promotores desde la versión inicial de nuestra base de datos. Ahora sabemos que no todas las instancias de motivos de secuencia son obligatorias para un promotor, que la unión de Eσ (factor sigma específico) no es suficiente ni necesaria, y que los promotores son específicos del factor sigma. Con estas consideraciones, la definición actual de promotor en RegulonDB es un segmento de ADN esencial para la iniciación de la transcripción en una ubicación definida en una molécula de ADN, aunque esta ubicación puede no ser un solo nucleótido. La identificación del promotor se da por medio del reconocimiento de un factor sigma específico, y este reconocimiento no es necesariamente autónomo (PMID:32665585, citllali).

En nuestra versión actual del modelo, los promotores conservan las propiedades esenciales de la versión 1.0; sin embargo, hemos eliminado las características adicionales debido a que la información al respecto es limitada. Los ajustes realizados se centran en el inicio de la transcripción. En versiones anteriores a la 10.5, cada promotor tenía un solo punto de inicio de transcripción, pero ahora hemos incorporado la posibilidad de tener un inicio de transcripción principal, que es el más frecuente, así como otros inicios de transcripción alternativos. Además, los motivos de unión del factor sigma, conocidos como cajas, se han incluido como un conjunto de elementos que pueden o no estar presentes en el promotor.

Several data have been modified in relation to the classic curation of low throughput data. In relation to this, many promoters were adapted to align with the description in the Mejia-Almonte et al. study.  If a promoter was recognized by two different sigma factors and initiated transcription at the same TSS, it was separated into two promoters with only one sigma factor and with the same TSS. As a result, we created many new promoters and thereby new transcription units.

En las distintas versiones de RegulonDB, se han producido cambios importantes en la conectividad entre los promotores y otros elementos genéticos. En la versión actual del modelo, las interacciones reguladoras (RIs) se asocian a los promotores en una relación de cero o uno, lo que implica que una RI puede estar vinculada o no a un promotor (por el tipo de experimento o falta de información). Asimismo, los promotores se conectan a las unidades de transcripción (TU) como unidades reguladas, estableciendo una relación de uno a muchos, donde una TU puede estar asociada exclusivamente a un promotor.

Para esta versión de RegulonDB, se ha registrado un total de 4050 promotores. La mayoría de estos promotores se clasifican como Sigma70, como se puede apreciar en la figura 3a. Sin embargo, aproximadamente el 17% de los promotores no cuenta con un factor sigma asignado, éstos promotores en su mayoría fueron identificados a través de experimentos de determinación del sitio de inicio de la transcripción sin un mapeo específico del factor sigma, como lo indican las evidencias exp-ida-transcription-init-mapping y exp-ida-hpt-transcr-init-m-race-map.

En los últimos años, hemos dedicado gran esfuerzo en RegulonDB para ofrecer evidencias que respalden todos los promotores registrados en la base de datos. Todos los promotores cuentan con al menos una evidencia y una referencia asociada. De acuerdo con su nivel de confianza, hemos clasificado los promotores en tres categorías: confirmado (79 promotores), strong (1456 promotores) y weak (2515 promotores). La figura 3c muestra el total de evidencias por promotor, destacando que el 60% de ellos solo posee una evidencia. Adicionalmente, en la figura 3d podemos observar que la mayoría de los promotores con una evidencia tienen un nivel de confianza baja (weak), y mientras las evidencias aumentan también lo hace el nivel de confianza.

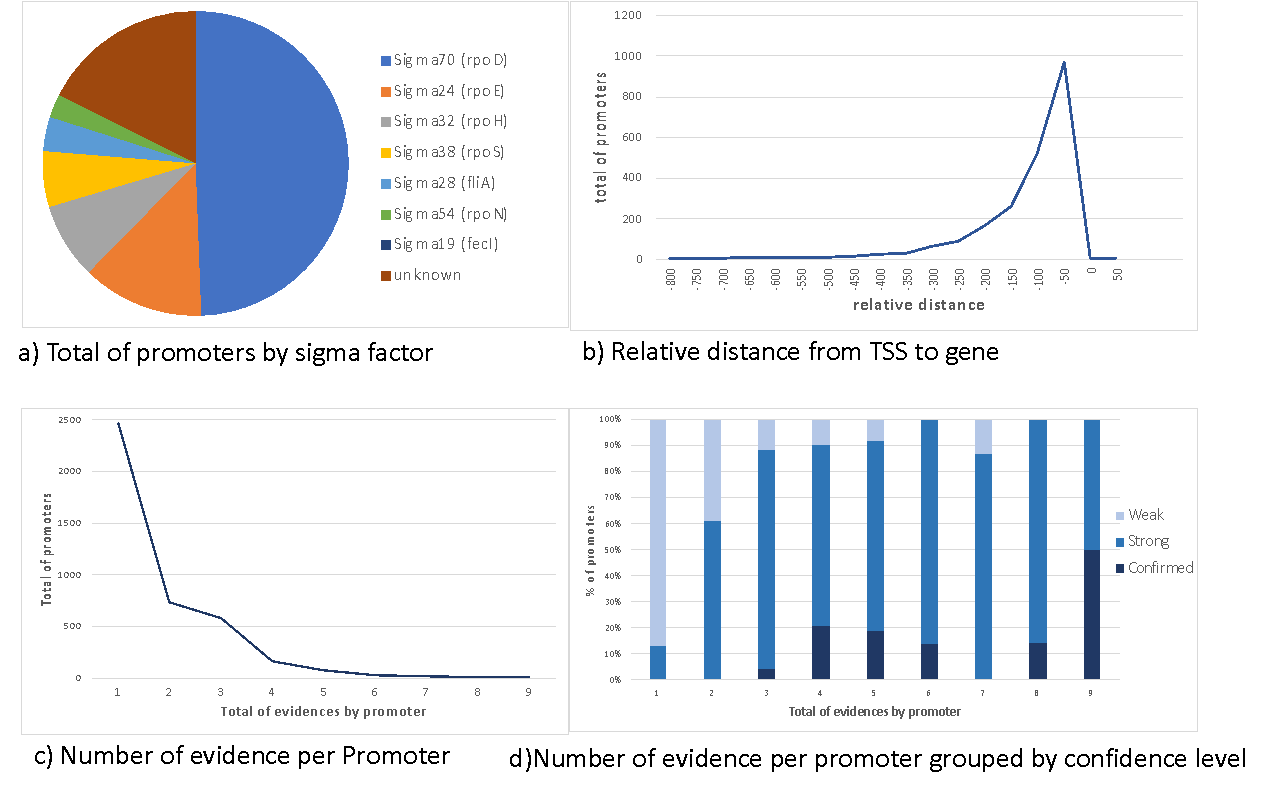
****

Figura 3. Estadísticas de algunas propiedades de los promotores. a) Distribución de promotores según el factor sigma. b) Distribución de promotores según la distancia relativa entre el sitio de inicio de la transcripción (TSS) y el primer gen. c) Distribución de evidencias por promotor. d) Total de evidencias por promotor agrupadas según su nivel de confianza.

**Transcription Units and Operons**

Otro de los conceptos que ha experimentado cambios significativos es el de los operones. Si bien en la definición inicial del modelo se consideraba al operón como una unidad transcrita policistrónica con sus sitios reguladores asociados, este concepto resultó ser limitado. En la tercera publicación de RegulonDB, se introdujo el concepto de unidades de transcripción con el fin de incorporar la regulación por múltiples promotores y la finalización de la transcripción en diferentes puntos dentro de un operón.

La complejidad interna de un operón ha sido discutida en Mejía-Almonte, y ha quedado claro cómo el concepto de unidades de transcripción y operón ha generado confusión a lo largo de los años, incluso en relación a si operones con un solo gen deberían ser considerados como tal. En RegulonDB, definimos una unidad de transcripción como segmentos de ADN que inician en un sitio de inicio de la transcripción (TSS) y finalizan en un sitio de término de la transcripción (TTS) [158,159]. Por otro lado, un operón se refiere a un conjunto de genes adyacentes cuya transcripción está coordinada por una o varias unidades de transcripción que se superponen entre sí, se transcriben en la misma dirección y comparten al menos un gen. Un operón simple es aquel en el que la transcripción de los genes está coordinada mediante una única unidad de transcripción, mientras que un operón complejo es aquel en el que la transcripción se coordina a través de varias unidades de transcripción que se superponen mutuamente, se transcriben en la misma dirección y comparten al menos un gen.

Las relaciones entre promotores, unidades de transcripción (TU) y operones se ilustran en la figura 1, donde cada unidad de transcripción está relacionada exclusivamente con un promotor. Sin embargo, es importante destacar que en algunos casos se han identificado TUs sin promotores, lo cual se ha observado en experimentos de análisis de expresión genética donde se ha estudiado la transcripción de la TU con y sin factores de transcripción, o cuando se han encontrado características de unión de proteínas alrededor del inicio de la TU.

En la versión actual de RegulonDB, se han identificado un total de 2595 operones y 3705 TUs. Como se muestra en la gráfica 4a, aproximadamente el 80% de los operones consta de una única TU, lo que significa que solo el 20% corresponde a operones complejos con múltiples promotores al inicio o dentro del operón. La existencia de promotores o terminadores internos es la razón por la que el número de genes asociados con las TUs y los operones difieren, tal como se ilustra en la figura 4b, donde se pueden observar TUs que contienen solo 1 o 2.

Como mencionamos anteriormente, hemos realizado un esfuerzo para asignar las evidencias correspondientes a cada TU, con el objetivo de determinar el nivel de confianza de la información. En la figura 4c se muestra que aproximadamente el 80% de las TUs tienen asociada una única evidencia. De esta cifra, el 48% corresponde a la categoría de "Inferido computacionalmente sin supervisión humana", mientras que las restantes evidencias (17 en total) respaldan el resto de las TUs. Estas incluyen categorías como "Inferencia automatizada de que una dirección de un solo gen es una unidad de transcripción", "Límites de transcripción identificados experimentalmente" y "Longitud del transcrito determinada experimentalmente". Es importante mencionar que la mayoría de estas evidencias se clasifican como débiles (ver figura 4d). Sin embargo, a medida que las TUs cuentan con más evidencias, su nivel de confianza aumenta, como se muestra en los datos de la figura 4d.

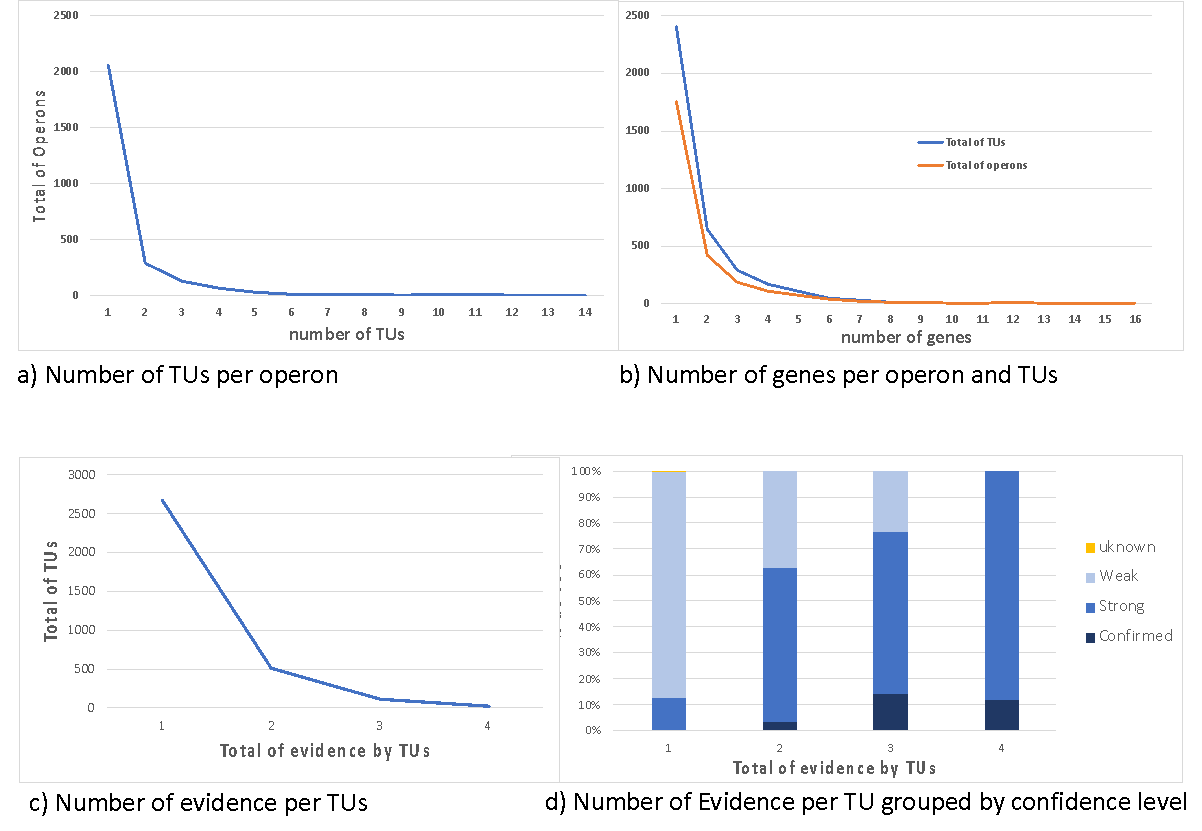


Figura 4.

Los datos relacionados con promotores, TUs y operones forman parte de la colección Operon en los datamarts de RegulonDB. Cada operón tiene un identificador único. Para acceder a los operones, existen dos opciones. La primera es a través de la caja de búsqueda, utilizando el nombre o identificador del promotor, la TU o el operón. La segunda opción es accediendo al menú principal y seleccionando la opción "Search" dentro de la sección "Browsing Operon"

La página de operones ha sido completamente modificada con el objetivo de mejorar la navegación. Es importante recordar que un operón puede tener una o varias TUs, y cada TU contiene información sobre el promotor, los genes transcritos y el terminador. En el diseño anterior de la navegación, se presentaban dificultades a la hora de desplazarse a las diferentes TUs de un operón. Sin embargo, en este nuevo diseño, los usuarios pueden navegar a través del menú de secciones, donde se lista las TU del operon, tal como se ilustra en la figura 5. La representación gráfica de los elementos genéticos relacionados con el operón se despliega al inicio, y dicho elemento gráfico cuenta con una barra de opciones que permite manipular su visualización. Las TUs se muestran en secciones colapsables, y dentro de cada TU hay subsecciones para promotores, la regulación y los terminadores, también colapsables.

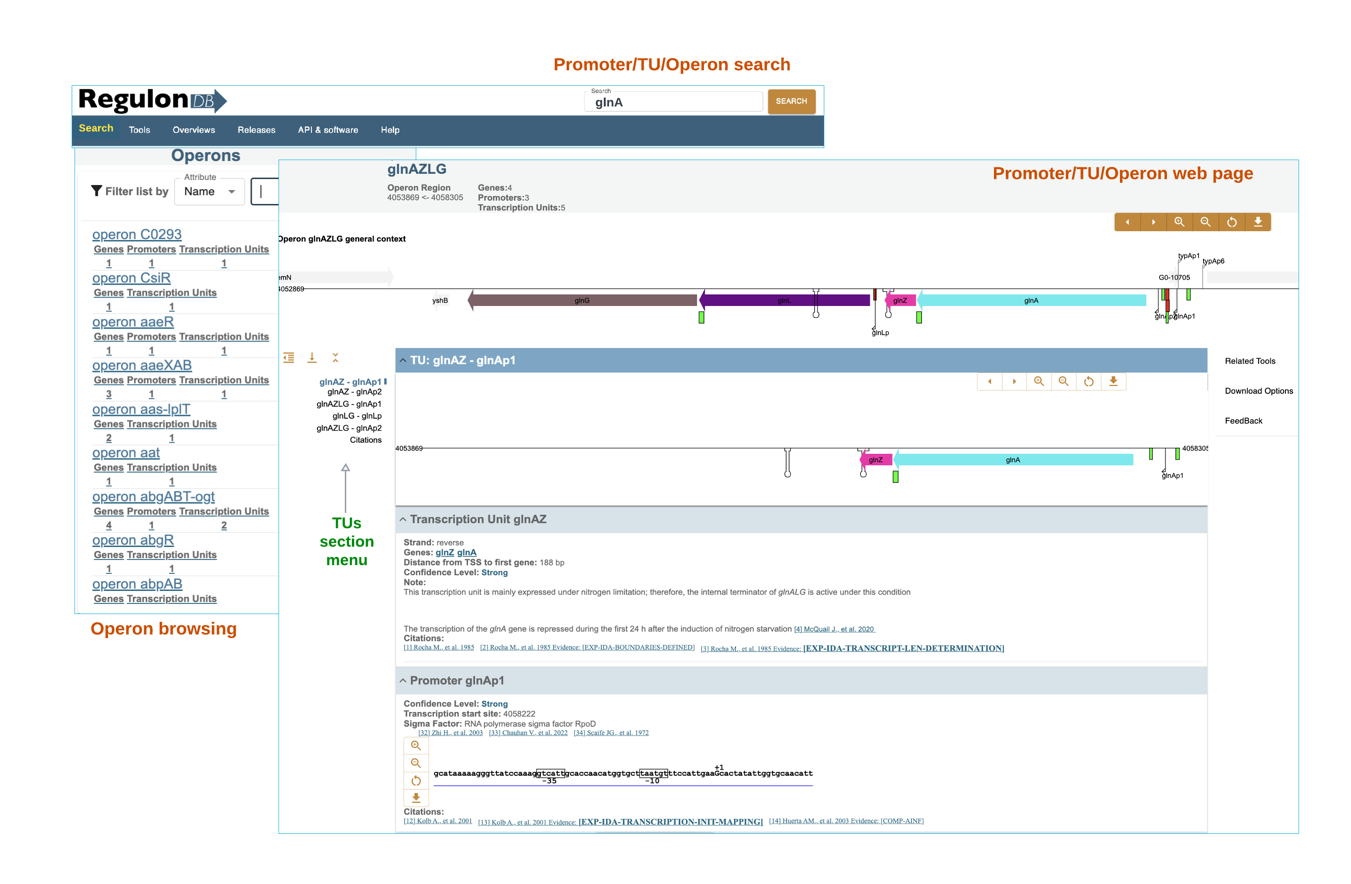


Figure 5. Opciones de búsqueda y despliegue de Operon.

**Transcription factors / Reguladores**

Las proteínas reguladoras descritas en nuestra versión 1.0 de RegulonDB se representaban en el modelo como "protein complex", con propiedades como un identificador, un nombre, sus sinónimos, familia evolutiva, su peso molecular, su secuencia y la simetría de reconocimiento en el ADN (direct repeat, inverted repeat or asymmetric). Además, se incluía información sobre su organización multimérica, indicando si la proteína era un monómero, dímero, tetrámero, etc.

En la versión 3.2, se realizó un ajuste importante para incorporar las conformaciones de las proteínas reguladoras que dependen de sus modificaciones covalentes y alostéricas. Esto permitió tener en cuenta la posibilidad de que estas proteínas se unieran o no a un metabolito efector. Además, se comenzó a utilizar el término "transcription factor" para referirse a estas proteínas en el modelo.

En la versión 6.0, se amplió aún más el alcance del modelo al incorporar otras formas y niveles de regulación. Esto incluyó la regulación por "small RNA", atenuadores y riboswitches. Estos elementos de regulación adicionales permitían una mayor precisión y detalle en la descripción de los procesos de control génico. En la versión 8.0, se añadió la regulación por factores que se unen alostéricamente a la RNA polimerasa directamente, específicamente por ppGpp y DksA. Esta adición fue significativa, ya que tenía un impacto directo en la transcripción de los genes.

Sin embargo, a medida que se incorporaban más y más elementos de regulación al modelo, este se volvía cada vez más complejo y requería de una revisión exhaustiva. Esto condujo a la necesidad de una actualización más reciente del modelo, con el objetivo de simplificar su estructura.

Como se menciona en la revisión de Mejía-Almonte [80-85], la definición clásica de "factor de transcripción" (TF) se refiere a una proteína que se une al ADN para regular la iniciación de la transcripción. Sin embargo, se ha demostrado que existen otros tipos de moléculas que también participan en la regulación de la transcripción, como los RNA reguladores [75,76] y pequeños ligandos, como el ppGpp [77]. Por lo tanto, la definición basada únicamente en la unión al ADN resulta imprecisa.

Para abordar esta limitación, se propone una revisión cuidadosa del concepto de "factor de transcripción" y se propone el término "producto génico regulador". Se define como cualquier producto génico que aumenta o disminuye la expresión de un conjunto específico de genes. Los producto génicos reguladores pueden clasificarse en dos categorías principales: ARN reguladores y proteínas reguladoras. Además, estas categorías pueden subdividirse aún más según el nivel en el que actúan para regular la expresión génica o el mecanismo específico de regulación.

Específicamente, se propone la clase de "proteína reguladora de unión al ADN para la iniciación de la transcripción" como una subclasificación de las proteínas reguladoras que se unen a secuencias específicas de ADN para regular la iniciación de la transcripción. Esta clase incluye a los TF, que son proteínas reguladoras que se unen cerca de un promotor y afectan la iniciación de la transcripción en dicho promotor. También se incluyen los factores sigma, que son proteínas reguladoras de la unión al ADN de la iniciación de la transcripción y que forman parte de la holoenzima de la ARN polimerasa, siendo esenciales para la iniciación específica de la transcripción.

Además, se propone otra subclase de proteínas reguladoras, denominada "proteínas reguladoras centradas en la ARN polimerasa de iniciación de la transcripción". Estas proteínas regulan la iniciación de la transcripción mediante la interacción con la enzima ARN polimerasa y su complejo de iniciación, conocido como Eσ. A diferencia de las proteínas reguladoras de unión al ADN, la especificidad de estas proteínas no se determina directamente por el reconocimiento de secuencias específicas de ADN.

En resumen, la revisión propone una ampliación y clarificación del concepto de "factor de transcripción" mediante la introducción del término "producto génico regulador" y la definición de diferentes clases y subclases de reguladores.

En el modelo no se cuenta con una entidad que englobe todos los tipos de productos génicos reguladores, sino que estos están representados por las diferentes formas en las que se encuentran en su conformación activa. Estas formas pueden ser: a) Producto de un gen que no requiere de ningún otro elemento para regular, como por ejemplo Lrp en su forma de monómero o los small RNA; b) Complejo regulador formado por uno o más productos, ya sean del mismo tipo o diferentes. Algunos ejemplos de estos complejos son el dímero AraC, el complejo IHF compuesto por IhfA y IhfB, CRP que forma un complejo con cAMP, o CsrA que forma un complejo con CsrB RNA; c) Continuante regulatorio, como por ejemplo ppGpp (ver figura 6). Estos continuantes regulatorios también forman parte de la regulación transcripcional.

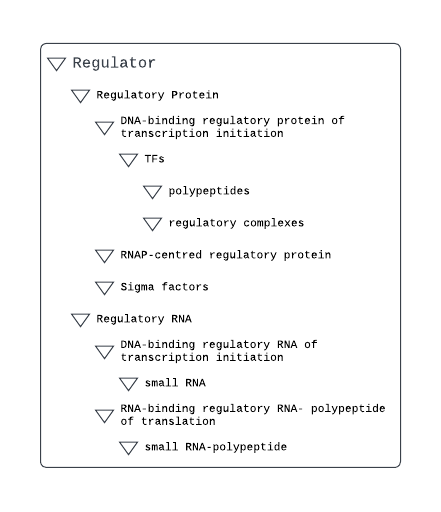


Figure 6. Tipos de reguladores.

**Regulatory Interactions**

En la versión 1.0 de RegulonDB, se definió una interacción reguladora (RI) como una interacción cuádruple que involucra el promotor regulado (P), la proteína reguladora (RP), el sitio de unión del regulador (S) y la función o efecto regulatorio sobre el promotor (F). En esta definición, el sitio al que se hace referencia es un sitio regulador de un factor de transcripción (TFRS), que ejerce un efecto sobre el promotor, a diferencia de los sitios de unión de factores de transcripción (TFBS) que pueden o no tener un efecto. Incluyendo a todos los productos génicos reguladores, definimos un regulator binding site (R-BS) como un sitio de ADN específico donde un regulador se une, mientras que el término sitios reguladores de un regulador (R-RS) se define como el subconjunto de R-BS que participa en la regulación genética.

Otro cambio significativo para éste concepto, es que derivado de experimentos de binding y de análisis de la expresión genética de la tu, donde no se determina el TSS, se tienen interacciones donde el objeto biológico regulado además del promotor puede ser la TU o el gene. Para lidear con esto, hemos sustituido la propiedad de promotor por entidad regulada.

Por lo que la definición de RI sería interacción cuádruple (RE, R, R-RS, F) donde RE es la entidad regulada, R, el regulador, R-RS, el sitio donde se une el regulador y participa en la regulación genética, y F, la función o efecto regulatorio sobre la entidad regulada.

Un análisis detallado de las RIs, las evidencias y su nivel de confianza se muestra en Lara-P.

**Regulon**

En la primera definición de regulon, lo describíamos como un grupo de operones controlados por un regulador. Sin embargo, debido a los ajustes realizados en relación a los reguladores y las interacciones reguladoras, sugerimos modificar esta definición.

Ahora proponemos definir un regulon como un conjunto de unidades de expresión genética que son directamente reguladas por uno o más productos génicos reguladores comunes. Estos productos génicos reguladores pueden ser proteínas o cualquier otro factor de regulación que tenga un efecto directo sobre la expresión de las unidades genéticas del regulon.

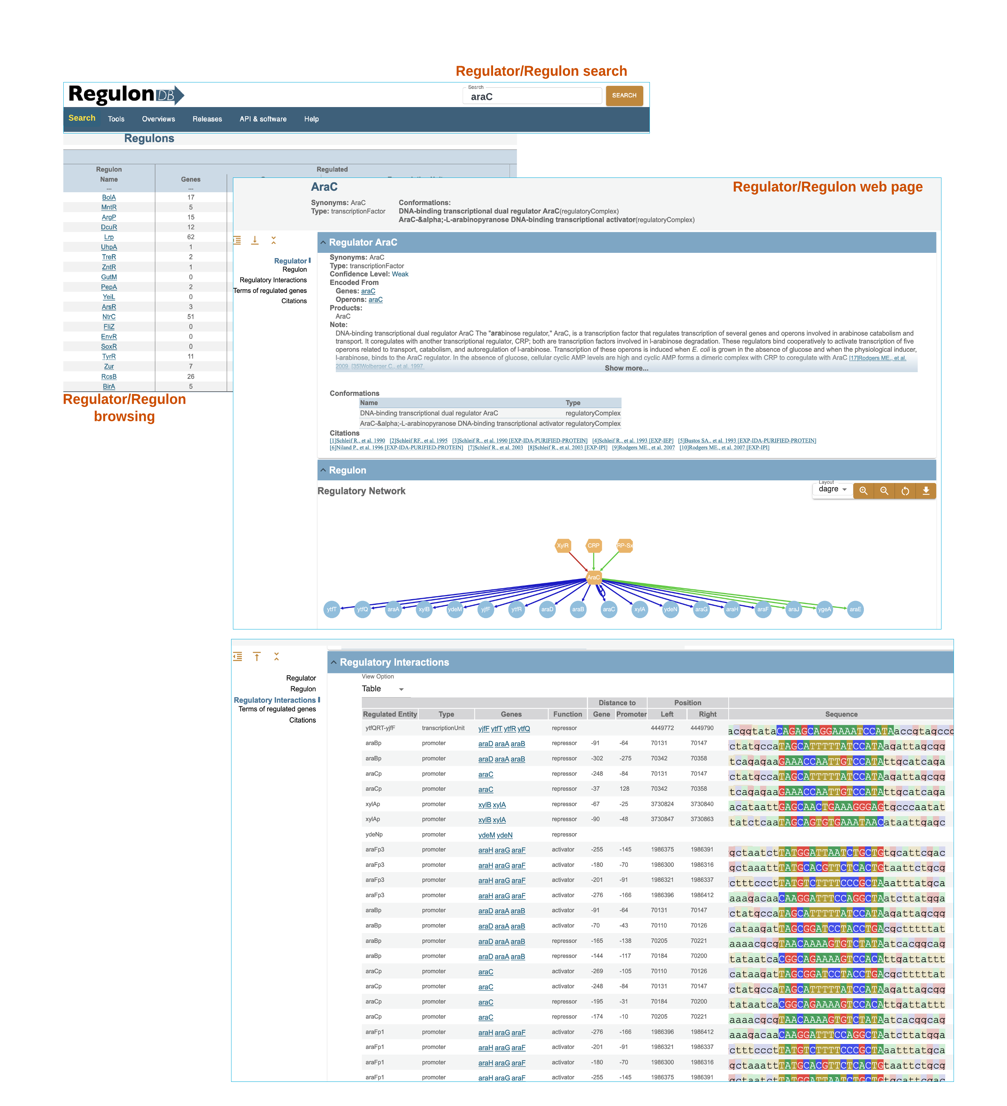


Figure 7. Opciones de búsqueda y despliegue de Regulator/Regulon

Los reguladores, las RIs y el regulon forman parte del datamart Regulon y es a la vez una de las páginas integradoras de RegulonDB. La figura 7 muestra la forma de acceder a un regulon desde la interfaz web. Una opción es a través de la caja de búsqueda y la otra es desde el menú principal en la opción de Search y después en regulon. Al igual que el resto de la pagina integradoras, se tiene el menu de secciones que permite navegar en toda la página. En ésta nueva interfaz, se ha integrado un despliegue del regulon usando el plugin de cytoscape, que tiene una barra de opciones para manipular el gráfico (ver figura 7).

En la sección Regulatory Interactions del regulon, se muestran todos los R-RS del regulador y se indica si la RI es de tipo regulator-promoter, regulator-tu y regulator-gene.

### Gensor Untis

Genetic Sensory Response Units (GENSOR units) are multilevel networks that reflect the functional impact of each TF on the cell via its regulated genes. They show the cascade of information processing that begins with the appearance of a signal, its transformation into the metabolite that binds the TF, the genetic switch that happens upon activation of the TF, and the response promoted by the activated/repressed genes. GENSOR Units have been present in Regulon since version X.X, released in 20?? (las de Soco). The original version comprised only XX manually generated GENSOR units with information retrieved from the literature. In version 9.0, we introduced an automatic assembly pipeline for the data extraction, which allowed us to expand the number to XX, but updates were scarce given that the generation of their graphic representation was still manual. In this version, we have introduced a completely automatized pipeline that allows for timely updates whenever new information is available. The number of objects included in the new release of GENSOR units is summarized in Table XX

The new graphic representation allows for the user to interact with the network [aquí explicar las herramientas con las que el usuario puede interactuar con la GU grafica]

**Curation and tools**

*Relevant classic curation*

Since the latest version of RegulonDB published in NAR, several data have been modified in relation to the classic curation of low throughput data.

It is not so common to add new genes in the database due to the fact more than 4500 genes are in the database. During this period, genes encoding a pair of new small proteins XtpA and SpfP and several small RNAs, such as GlnZ, StfZ, AmeF among others have been added to the database and subjected to curation, including the regulatory interactions of the new small RNAs.

CRP is a global transcriptional factor that regulates a large number of genes and it has been extensively analyzed. Recently, a CRP complex with the polypeptide Sxy was identified. The database now includes the CRP-Sxy complex, responsible for regulating certain genes associated with DNA uptake (natural competence). As described by Søndberg et al. this CRP-Sxy complex recognizes a DNA-binding site with an asymmetric arrangement: it comprises two separate half-sites, with one resembling the typical CRP-cAMP DNA-binding site, while the other half-site shows lower conservation. Other new conformations for other transcription factors have been also added to the database, such as dimer formations for FrlR (Graf von Armansperg et al.),  DgoR (Arya et al.), NanR, ChbR, SlyA and RcdA along with newly identified inactive configurations like FrlR-fructoselysine-6-phosphate (Graf von Armansperg et al.), MprA-salicylate (Arshad et al.), Cbl-thiosulfate, MurR-N-acetyl-D-glucosamine-6-phosphate and AraC-D-fucose. Besides, the acetylated conformation of RcsB was annotated.

In addition, we added the homodimer form as well as the inactive ligand-bound forms of the transcription factor RcdA, RcdA-Tris and RcdA–trimethylamine N-oxide. Four new inactive conformations of the transcription factor [AcrR](https://ecocyc.org/gene?orgid=ECOLI&id=EG12116) were included. AcrR-3,6-diaminoacridine (proflavin), AcrR-rhodamine 6G (RDG).  On the other hand, Galego et al. demonstrated that the functional activity of the transcription factor BolA relies on its phosphorylated state. All of the previous regulatory interactions of BolA are now associated exclusively with the active BolA-phosphate form.

Some proteins have recently been found to have the function of a transcriptional regulator, such as the sigma factor RpoN that was described by Shimada et al. as a transcriptional repressor of Sigma70 promoters in the absence of enhancers. [Turnbull and Gerdes](https://www.ncbi.nlm.nih.gov/pubmed/28266056) identified the [HicB](https://ecocyc.org/gene?orgid=ECOLI&id=G6749) protein, known as the antitoxin of the HicA-HicB toxin-antitoxin system, as a transcription factor that autoregulates the transcription of the hicAB operon. On the other hand, proteins that were predicted as Transcriptional factors were confirmed as proteins with that function upon the discovery of their first regulatory interactions by high throughput analysis; a few instances of these novel TFs include YfeC, YidZ, YciT, YgbI, AaeR, PunR, YieP, YeiE,  YafC, YiaU, YgfI, and FimZ. On the other hand, the DNA-binding consensus sequences of 125 transcription factors and five sigma factors were annotated in the database.

Duarte-Velázquez et al. 2022 described 58 potential transcription factors, encompassing details like their protein family classification, DNA binding domains, potential regulated processes, and the mechanisms governing their regulation  (Duarte-Velázquez et al. 2022), we included these descriptions in the summary of each putative TF.

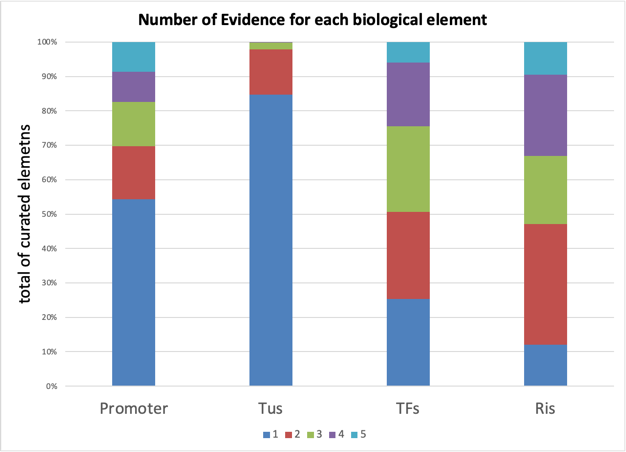
All the objects in RegulonDB contain evidence supporting their presence in E. coli. During this period, some modifications were made to the evidence codes and evidence ontology, leading to the creation of 14 new more specific evidence codes and the reclassification of six existing ones. These evidence codes pertain to distinct experimental approaches, encompassing both traditional and high-throughput methods, and are currently employed in the curation process.

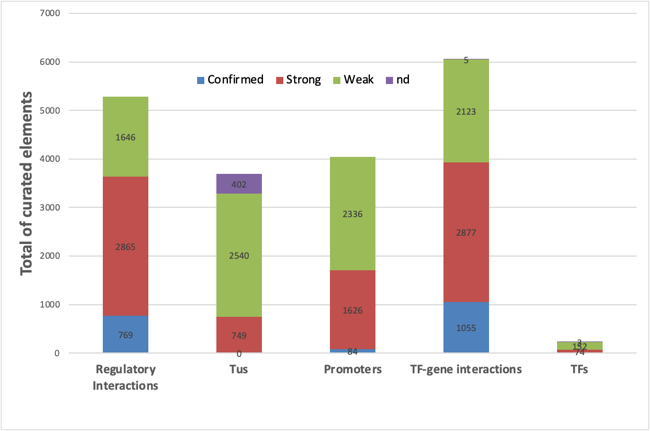
*Lisen&Curate: A platform to facilitate gathering textual evidence for curation y la anotación de frases que se despliegan en RegulonDB*

Lisen&Curate is a web-based platform meticulously designed for the curation of information concerning the regulation of transcription initiation in bacteria. This platform contains interfaces designed for capturing a diverse type of information, like genes, transcription units (TUs), promoters, regulatory interactions (RIs), proteins, compounds, protein features, and terminators. Currently, L&C is used to feed to RegulonDB.

Something very relevant that this platform contains are fields to link the sentences of a paper that describe the properties of an object, like evidence,object  name, regulatory effect, summaries or notes, among others.  The sentences or part of a sentence are named Sources. These sources are offered within RegulonDB. At the moment there are XX number of sources for promoters, X number for regulatory interaction, X number for TFs.

Another feature of this platform is an interface to annotate growth conditions and link them to the different objects of the database. The interface has integrated the MCO ontology to use only the terms annotated in the MCO to bind the terms corresponding to the different items of a specific framework  that form a growth condition. Currently Xnumber of RIs,  have been linked to at least one growth conditions. (o esto no se puede ver aún en RegulonDB?)





**DISCUSSION**

**AVAILABILITY**

**ACCESSION NUMBERS**

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR online.

**ACKNOWLEDGEMENT**

The .

**FUNDING**

This work was supported by ...

**CONFLICT OF INTEREST**

**REFERENCES**

1. Schmitt,E., Panvert,M., Blanquet,S. and Mechulam,Y. (1995) Transition state stabilisation by the 'high' motif of class I aminoacyl-tRNA synthetases: the case of Escherichia coli methionyl-tRNA synthetase. Nucleic Acids Res., 23, 4793-4798.

**TABLE AND FIGURES LEGENDS**