

Análisis comparativo de los sistemas de regulación transcripcional de la absorción de hierro

Sergio Ortí Carrió

Máster en Bioinformática y Bioestadística

Área 2

Ivan Erill Sagales, Consultor

Marc Maceria Duch, Profesor responsable de la asignatura

23 de junio 2021



Esta obra está sujeta a una licencia de Reconocimiento-
NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Análisis comparativo de los sistemas de regulación transcripcional de la absorción de hierro</i>
Nombre del autor:	<i>Sergi Ortí Carrió</i>
Nombre del consultor/a:	<i>Ivan Erill Sagales</i>
Nombre del PRA:	<i>Marc Maceria Duch</i>
Fecha de entrega (mm/aaaa):	06/2021
Titulación:	<i>Máster en Bioinformática y Bioestadística</i>
Área del Trabajo Final:	2
Idioma del trabajo:	Español
Número de créditos:	15
Palabras clave	<i>Comparative genomics, ferric uptake regulator, motif discovery.</i>

Resumen del Trabajo (máximo 250 palabras):

El hierro es un elemento esencial para la supervivencia de la mayoría de las bacterias. Este interviene en su crecimiento y tiene un papel fundamental para la patogénesis de estas. De la misma manera que es esencial, en concentraciones elevadas puede ser tóxico, por lo que precisa de mecanismos para mantener unos niveles constantes. Estos mecanismos conforman la homeostasis del hierro, la cual está regulada en gran parte por la proteína Fur. Esta proteína regula de forma negativa, y algunas veces de forma positiva, la transcripción de varios genes relacionados con la captación de hierro y la patogénesis, por lo que tiene una gran importancia para entender los mecanismos de patogenia y para elaborar posibles tratamientos terapéuticos. En este trabajo se ha estudiado la secuencia de unión al ADN de Fur mediante herramientas de descubrimiento de motivos. A continuación, se ha empleado este motivo encontrado para realizar un análisis de genómica comparativa con CGB. Como resultado de los análisis, se ha obtenido un motivo que sugiere un sitio de unión al ADN de Fur, el cual sigue un modelo propuesto previamente en otro estudio. Tras el análisis de genómica comparativa, se han obtenido elementos regulados por Fur donde se han encontrado diferencias en la regulación para bacterias patógenas y no patógenas, que sugieren un papel de Fur en la patogenia regulando estos genes. Además, se han comparado estos con elementos descritos en la literatura.

Abstract (in English, 250 words or less):

Iron is an essential element for the survival of most bacteria. Iron intervenes in growth and has a fundamental role in pathogenesis. In the same way that it is essential, it can be toxic in high concentrations, so it requires mechanisms to maintain constant levels. These mechanisms are called iron homeostasis, which is regulated by the Fur protein. This protein regulates negatively, and sometimes positively, the transcription of several genes related to iron uptake and pathogenesis, so it is of great importance to understand the mechanisms of pathogenesis and for developing possible therapeutic treatments. In this work, the Fur DNA binding sequence has been studied using motif discovery tools. We found a motif that was then used to perform a comparative genomics analysis with the tool CGB. As a result of the analyzes, a motif has been obtained that suggests a Fur-binding site, similar to a motif previously proposed in another study. After comparative genomics analysis, elements regulated by Fur have been obtained, where differences in regulation have been found for pathogenic and non-pathogenic bacteria, which suggest a role of Fur in the pathogenesis by regulating these genes. These genes have been compared with elements described in different studies.

Índice

1. Resumen.....	7
2. Introducción	9
2.1. Contexto y justificación del estudio	9
2.2. Objetivos	9
2.3. Enfoque y método seguido	9
2.4. Planificación del trabajo.....	10
2.4.1. Tareas	10
2.4.2. Calendario.....	11
2.4.3. Resumen de contribuciones y productos obtenidos	12
2.4.4. Descripción de otros capítulos de la memoria	12
3. Estado del arte.....	13
3.1. Regulación del hierro en bacterias.....	13
3.2. Papel del hierro en la patogénesis	13
3.3. La proteína FUR.....	15
3.4. Fur box.....	16
3.5. Regulación del gen <i>fur</i>	18
3.6. Fur y otras proteínas relacionadas.....	21
3.7. CollecTF	23
3.8. Genómica comparativa	24
3.9. CGB.....	26
4. Materiales y métodos.....	29
4.1. Obtención modelo de unión de proteína Fur	29
4.1.1. Obtención y manipulación de los datos	29
4.1.2. Búsqueda del modelo del sitio de unión de Fur.....	30
4.2. Análisis de genómica comparativa sobre los datos	32
4.2.1. Miniconda.....	33
4.2.2. Programas externos.....	34
4.2.3. Análisis comparativo. CGB	35
5. Resultados y discusión.....	38
5.1. MEME	38
5.2. FIMO.....	40
5.3. CGB.....	41
6. Conclusiones.....	46
7. Bibliografía.....	48
8. Anexos	55

Tabla de ilustraciones

Ilustración 1. Diagrama de Gantt sobre la planificación del trabajo.....	11
<i>Ilustración 2. Secuencia consenso para el sitio de unión al ADN de Fur (Andrews et al., 2003).</i>	16
Ilustración 3. Secuencias de los diferentes modelos para explicar el motivo de unión al ADN de la proteína Fur (14).....	18
Ilustración 4. Genes regulados por Fur en E.coli (Andrews et al., 2003).....	20
Ilustración 5. Resultado de MEME obtenido para Vibrio cholerae donde se muestra en rojo el motivo encontrado que coincide con el modelo C1.	39
Ilustración 6. Motivos encontrados para el Modelo C2 en bacilli, donde se observa cómo los motivos que coinciden con el heptámero del modelo C2 (en rojo) se agrupan para formar la estructura del modelo C1.	40
Ilustración 7. Árbol filogenético obtenido tras el análisis con CGB empleando la proteína RecA.	41

1. Resumen

El hierro es un elemento esencial para la patogénesis y la supervivencia de la mayoría de las bacterias. Interviene en el crecimiento del mayor porcentaje de bacterias, por lo que estas poseen diferentes mecanismos para adquirir el hierro del ambiente. Estos mecanismos se emplean esencialmente en condiciones donde el hierro no es abundante en el ambiente o cuando se encuentran en un huésped, donde este último emplea sus propios mecanismos para secuestrar el hierro y así evitar que la patogenicidad de la bacteria (1). Esta dependencia del hierro se debe a que la atmósfera prebiótica era anaeróbica. Los primeros organismos identificados que no dependían del hierro fueron los lactobacilos, los cuales tienen solo uno o dos átomos de hierro en la célula, niveles demasiado bajos de hierro como para realizar una función biológica (2). También se encuentran bacterias que parece que no requieran hierro para iniciar sus mecanismos de patogenicidad, como *Treponema pallidum*, causante de la sífilis. Esto podría deberse a que es un parásito intracelular obligado, por lo que puede ser que aunque no requieran hierro, empleen los procesos metabólicos dependientes de hierro del huésped para obtener energía y sintetizar elementos funcionales, dependiendo indirectamente de este (3).

Para captar el hierro, las bacterias secretan sideróforos, que son quelantes de gran afinidad del ion férrico. A pesar de ser esencial, el hierro puede ser tóxico cuando las concentraciones son altas en el citoplasma. Esto se debe a la capacidad que tienen de catalizar la reacción de Fenton, la cual provoca la formación de radicales hidroxilos. Debido a esto, es necesario para las bacterias controlar la homeostasis del hierro. Este control se lleva a cabo con la proteína Fur (Ferric uptake regulator) y sus ortólogos. Fur es un represor de la transcripción en presencia de hierro. Se une al ADN y reprime la transcripción de la mayoría de los genes involucrados en el transporte de hierro y la síntesis de sideróforos (1).

El hierro y Fur tienen un papel importante en bacterias patógenas, ya que usan la concentración baja de hierro en un huésped como señal para aumentar la expresión de varias toxinas bacterianas y otros determinantes de virulencia. También existen sistemas de quimiotaxis en bacterias, como por ejemplo en *V.cholerae*, los cuales responden a Fur durante la patogénesis (1).

A pesar de que la proteína Fur ha sido muy estudiada debido a su importancia en la patogenicidad y el comportamiento de las bacterias, no se conoce con exactitud la secuencia de ADN a la cual se une. La secuencia a la cual parece que se une sí que se conoce, la cual se denomina "Fur box" o "Iron box", aunque el motivo de unión no. Se

han presentado varios modelos en varios estudios que sugieren distintos motivos, aunque no se ha concretado.

En este estudio se ha tratado de estudiar Fur en diversas clases de bacterias empleando técnicas de genómica comparativa. Estas técnicas permiten analizar las relaciones evolutivas de diferentes clados del reino bacteria, lo que permite analizar diferentes especies bacterianas patogénicas y no patogénicas y así comprender su papel en la virulencia. Por esto, se buscó extraer un patrón común o un regulón mantenido en diferentes especies gracias a la gran cantidad de datos existentes de secuencias de unión a la proteína Fur, estudiando así posibles diferencias en Fur y los operones que regula esta en bacterias patógenas y no patógenas debido a la importancia que juega en la patogénesis, para así intentar comprender su papel en la patogenicidad y poder usarlo como una diana para antibióticos. Por último, también se ha intentado dilucidar la secuencia de unión al ADN de Fur, ya que hoy en día no se ha aclarado que estructura sigue.

Como resultados del estudio, se ha encontrado una secuencia que es similar en varias especies bacterianas y que podría representar la estructura seguida por el lugar de unión al ADN de Fur. Además, esta estructura se ha empleado para el análisis comparativo de diferentes clados bacterianos, donde se han encontrado diversos elementos regulados por la proteína Fur. Además, se han encontrado genes regulados por Fur que presentaban regulaciones distintas entre las bacterias patógenas y no patógenas, sugiriendo un papel de Fur en la patogenicidad. Por último, estos elementos encontrados se han comparado con literatura existente que reportase esos elementos en las especies analizadas o en otras especies.

2. Introducción

2.1. Contexto y justificación del estudio

Este trabajo de fin de máster consistirá en identificar redes reguladoras centrales en de clados bacterianos, de forma que se pueda observar si hay elementos conservados entre clados o para comprender la función principal del regulón, e identificar elementos de regulón diferenciales entre miembros patógenos y no patógenos de clados, para ver si hay una evolución convergente en los patógenos hacia la cooptación del regulón Fur para la virulencia.

La genómica comparativa juega un papel importante en la patogénesis. Con esta herramienta, es posible analizar las relaciones evolutivas de diferentes clados de bacterias, lo que lleva a comprender su virulencia. Fur es un regulón esencial en las bacterias para evitar daños por ROS, obtener hierro y activar la patogenicidad. La idea de este estudio es intentar encontrar un patrón común o un regulón mantenido en diferentes especies, aprovechando la gran cantidad de datos, para comprender su papel en la patogenicidad y tal vez usarlo como objetivo terapéutico.

2.2. Objetivos

- Como objetivos generales del trabajo, se tienen:
 - Identificar que estructura tiene el motivo de ADN de unión a Fur a lo largo de distintos grupos bacterianos.
 - Identificar elementos diferenciales de regulones entre miembros patogénicos y no patogénicos de clados.
- En cuanto a los objetivos específicos, se busca:
 - Obtención de sitios de unión de Fur al ADN conocidos de la base de datos CollecTF y expandirlos.
 - Realizar descubrimiento de motivos con diferentes restricciones para observar que modelo es el mejor soportado.
 - Realizar análisis de genómica comparativa con CGB de una selección de clados bacterianos.

2.3. Enfoque y método seguido

Para llevar a cabo los objetivos, se ha realizado una minuciosa labor de investigación sobre la proteína Fur. Es muy importante saber que actúa de forma distinta

en bacterias patógenas y no patógenas. Además, es necesario conocer el motivo de unión de Fur al ADN, de forma que se pueda estudiar el transcrito para estudios de genómica comparativa. También se necesita recabar información sobre las secuencias de unión de Fur de una base de datos.

Una vez se obtiene esta información, se puede empezar con el análisis. Es posible seguir dos enfoques, realizando al mismo tiempo la parte teórica y práctica del estudio, o bien separando estas. Debido a la base en biología del autor, se ha decidido empezar primero con una parte de la parte teórica para tener una base en el tema y conocer algunos aspectos importantes del tema para luego aplicarlos en la parte práctica. Tras esto, se continuará con la parte práctica, dado que tendrá una mayor carga de trabajo y será más fácil analizar los resultados con retrospectiva. Cabe destacar que durante la parte práctica habrá ciertos momentos donde sea imprescindible estudiar alguna parte teórica, como el funcionamiento de una herramienta empleada. De esta forma, antes de emplear una herramienta bioinformática se obtendrá información sobre el funcionamiento y el resultado que proporciona antes de emplearla, optando así por trabajar con una herramienta tras tener información suficiente de su funcionamiento para evitar, en lo posible, problemas.

2.4. Planificación del trabajo

El proyecto se ha dividido en tres fases principales, la documentación y obtención de los datos, el descubrimiento de motivos, y la realización de análisis comparativos de regulones de Fur entre los distintos clados.

2.4.1. Tareas

- Documentación y obtención de los datos.
 - Examinar la literatura en busca de información sobre la proteína Fur y su regulación.
 - Compilar datos de la base de datos CollecTF sobre el motivo de unión al ADN de Fur.
 - Examinar la literatura para buscar información adicional no cubierta en CollecTF.
- Descubrimiento de motivos.
 - Adaptar scripts para descargar automáticamente regiones promotoras para genes característicos.

- Realizar descubrimiento de motivos por clado.
- Analizar los resultados y definir un motivo de unión de Fur.
- Inferir motivos de unión a Fur con herramientas de genómica comparativa en clados sin evidencia experimental.
- Realizar análisis comparativos de regulones de Fur entre los clados.
 - Instalar la plataforma CGB y entender su uso.
 - Definir clados para el análisis.
 - Llevar a cabo análisis comparativos de regulones de Fur entre diferentes clados.
 - Analizar los resultados.

2.4.2. Calendario

A continuación, se muestra el calendario del Trabajo de Fin de Máster. En este, se muestran las tareas programadas anteriormente desde el día 17 de marzo de 2021. La primera consiste en obtener información sobre Fur y entender su funcionamiento, además de compilar la información sobre los motivos de las secuencias para empezar el análisis de la parte práctica. El apartado práctico se divide en tres partes, correspondientes a las tareas. La primera, de color verde, abarcará casi la segunda quincena de marzo, la segunda fase, de color amarillo, abarca casi la totalidad del mes de abril, y, por último, la tercera fase, de color rojo, abarca desde el final del mes de abril hasta el final del mes de mayo. En cuanto al apartado teórico, se ha dividido la planificación en cada punto del TFM, destacando la investigación y recopilación de información y los métodos, pues son a los que más tiempo se les ha dedicado.

En un principio, los tiempos de las primeras tareas eran menores, pero debido a complicaciones en la parte práctica, se alargaron estos tiempos.

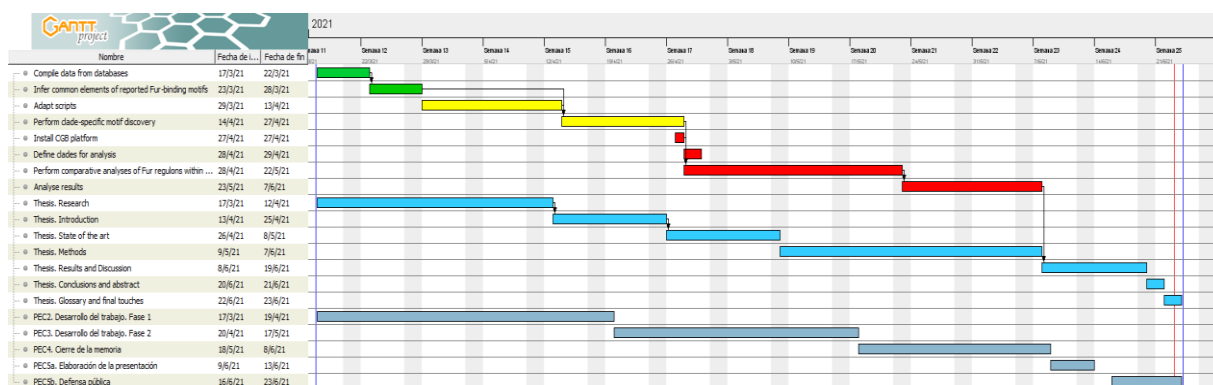


Ilustración 1. Diagrama de Gantt sobre la planificación del trabajo.

En este, se muestran en azul las diferentes secciones de la realización de las partes teóricas del TFM. Para las partes experimentales, se han usado los colores verde, amarillo y rojo, donde cada color corresponde a un objetivo diferente de los que se propusieron. Finalmente, se han marcado en gris los diferentes hitos, los cuales se corresponden con las diferentes PECs que se han ido realizando. Cabe destacar que debido a problemas personales no se ha podido entregar en la fecha estimada el TFM, sino que se ha retrasado su entrega unos días.

2.4.3. Resumen de contribuciones y productos obtenidos

Como resultado del trabajo se han obtenido varios análisis de secuencias de unión a la proteína Fur donde se ha obtenido un motivo de unión para la proteína Fur que se corresponde con un modelo propuesto con anterioridad. Además, se han obtenido varios análisis de genómica comparativa para diferentes clados bacterianos donde se han observado diferencias en la regulación de Fur en bacterias patógenas y no patógenas.

2.4.4. Descripción de otros capítulos de la memoria

Estado del arte: En este apartado se definen varios aspectos teóricos necesarios para tener una base en el conocimiento de la problemática del estudio.

Métodos y materiales: Aquí se han descrito todas las herramientas empleadas para llevar a cabo los análisis pertinentes, así como la descripción de la metodología realizada con esos materiales para la obtención de los resultados.

Resultados y discusión: En esta sección se han descrito los resultados obtenidos para la obtención del motivo de unión al ADN de la proteína Fur y para el análisis comparativo de las secuencias obtenidas. Además, también se han interpretado los resultados e inferido posibles relaciones con los resultados de otros estudios.

Conclusión: Se han descrito los diferentes resultados obtenidos por objetivos, donde se han resumido estos y se ha respondido individualmente a las problemáticas que se buscaban dilucidar.

3. Estado del arte

3.1. Regulación del hierro en bacterias

El hierro es un elemento que forma parte del grupo de los metales de transición. Existe en dos posibles formas en los seres vivos, como ion ferroso o como ion férrico. Se trata de un metal abundante que se requiere para la vida de en la mayoría de los casos, pues se emplea como cofactor para un gran número de enzimas. Participa en numerosos procesos biológicos como la fijación de nitrógeno o la fotosíntesis. En condiciones aeróbicas, el ion Ferroso se convierte en ion férrico debido a la exposición al oxígeno, formando hidróxidos insolubles a un pH neutro que provocan que el hierro disponible para las bacterias escasee. En condiciones anaeróbicas, el Fe^{2+} es soluble a un pH fisiológico, por lo que estos microorganismos pueden obtenerlo del medio. Prácticamente todas las bacterias producen y secretan sideróforos, componentes de bajo peso molecular que emplean para obtener el hierro del medio. El hierro puede ser tóxico bajo condiciones aeróbicas, por lo que los organismos deben equilibrar su captación de hierro para que los niveles de hierro se mantengan en un cierto nivel, pero teniendo en cuenta la formación de residuos de hierro libre que producen toxicidad. Además, el exceso de hierro puede catalizar reacciones de Fenton y formar especies activas de oxígeno. Debido a todo esto, se deben llevar a cabo varios procesos para mantener las concentraciones dentro de unos límites deseados. Estos procesos conforman lo que se denomina como homeostasis del hierro (4).

Las bacterias se encargan de llevar a cabo la regulación del hierro a través de cinco mecanismos principales. Estos mecanismos son la adquisición de hierro del ambiente, el almacenamiento de hierro intracelular para tener depósitos a los que acceder en ambientes con poca disponibilidad de hierro, el uso de sistemas de resistencia al estrés redox, el control del consumo de hierro a través de la regulación de la expresión de proteínas que contienen hierro y de un sistema regulador que reacciona a la presencia de hierro y coordina la regulación de la homeostasis del hierro. De forma normal, las bacterias regulan su metabolismo del hierro en respuesta a la disponibilidad de este. Esta regulación se media a través de la proteína Fur (Ferric-uptake regulator) (3).

3.2. Papel del hierro en la patogénesis

El hierro tiene un papel fundamental en varios procesos biológicos. Por ejemplo, sirve como cofactor en el transporte de electrones y en el metabolismo. Esto influye

tanto en los organismos patógenos como en los huéspedes, por lo que el hierro tiene una gran importancia en la patogénesis. A causa de esta dependencia mutua por el hierro, se llevan a cabo diferentes mecanismos por parte del patógeno y del huésped para conseguir o mantener el hierro que necesitan.

En el caso de los humanos y mamíferos, existen varios mecanismos para secuestrar el hierro libre, como parte de defensas innatas ante microorganismos invasores. Un ejemplo de estos mecanismos es el uso de estructuras que almacenen el hierro libre, como la hemoglobina o la ferritina, las cuales suelen estar situadas en localizaciones donde suele darse la infección. Estas estructuras reducen los niveles de hierro extracelular libre a unas concentraciones de 10^{-18} , las cuales imposibilitan el crecimiento bacteriano. Además, se comienza una producción de proteínas que se unan a la hemoglobina y los grupos hemo, como la hemopexina, para evitar que las bacterias patogénicas puedan extraer hierro de esa fuente (3). También se produce una disminución en la absorción de hierro a través del intestino y un aumento de la producción de las estructuras que almacenan hierro ya mencionadas. Este secuestro de hierro libre tiene una doble función, pues además de retirarlo del ambiente para que organismos patógenos puedan emplearlo y causar infección, también previene el daño oxidativo que se produce cuando el hierro se une al oxígeno (4). A pesar de todos estos mecanismos para secuestrar el hierro, diversos patógenos han desarrollado estrategias para aumentar la afinidad por el hierro, y así competir con el huésped por el hierro libre, o para acceder a los almacenes de hierro del huésped y poder iniciar la patogénesis (5).

En las bacterias, la ausencia de hierro en el ambiente actúa como señal que provoca la activación de la expresión de mecanismos de virulencia, como la toxina Shiga-like I de *E. coli enterohemorrágica* (6). Como mecanismo para lidiar con la ausencia de hierro en el ambiente provocada por el huésped, las bacterias emplean sideróforos para captar el hierro unido a las proteínas de unión al hierro, como la ferritina mencionada previamente, del huésped. Es común que estos mecanismos para captar el hierro del huésped se lleven a cabo mediante sistemas de transporte mediados por el receptor específicos para los complejos de hierro del huésped (3). Además, se cree que Fur está relacionado indirectamente con la expresión de *toxA*, un gen que codifica para la producción de toxinas en *Pseudomonas aeruginosa*. Otros ejemplos son la regulación negativa de la producción de hemolisina en *Vibrio cholerae* (7) y la interacción de Fur con promotores de genes que codifican para proteínas de membrana que se emplean para la invasión y la adherencia al huésped en *Neisseria gonorrhoeae* (8). En *Helicobacter pylori*, Fur regula indirectamente la citotoxina vacuolativa (*vacA*) (5,9).

Además, Fur está relacionado con la colonización del huésped. Un estudio realizado por Bury-Moné et al., en 2004 demostraba como mutantes de *Helicobacter pylori* eran menos eficientes para colonizar un huésped que las cepas salvajes (10). Además, mutantes de *fur* en varios patógenos, como *Listeria monocytogenes* (11), o *Campylobacter jejuni* (12), exhiben una disminución de la virulencia en huéspedes animales.

3.3. La proteína FUR

La proteína Fur (Ferric-uptake regulator) fue descrita por primera vez en 1981 por Hantke, donde describía un mutante de *E.coli* que expresaba funciones inhibidas por el hierro como si este no estuviera presente. Este mutante fue denominado “ferric uptake regulator” debido a la clara relación con la represión de la mayoría de los genes que respondían a concentraciones de Fe. Fur regula más de 90 genes que dependen del hierro para su expresión en *E.coli*. Se trata de un homodímero compuesto por 2 subunidades dando un total de 17 -kDa, donde se encuentra una subunidad de unión al ADN (C-terminal) y otro dominio rico en residuos de Histidina el cual está implicado en la dimerización (N-terminal). Fur se une al hierro bajo condiciones de alta cantidad de hierro, adquiriendo una estructura capaz de unirse a secuencias de ADN, conocidas como Fur boxes. Actúa como un represor positivo, de forma que reprime la transcripción en presencia de su correpresor, el ion ferroso. Esta represión se produce por el aumento de la afinidad de Fur por su sitio de unión al ADN al unirse el ion ferroso a las subunidades correspondientes. En condiciones de escasez de hierro, el equilibrio provoca que se liberen iones ferrosos, por lo que el microorganismo emplea RNA polimerasas para acceder a genes como los encargados de la biosíntesis de sideróforos y genes relacionados con otras funciones del hierro para expresarlos. Normalmente se une un ion ferroso por cada subunidad de Fur, aunque también se pueden unir metales relacionados como el Cobalto o el Manganeso, a pesar de que in vivo sea improbable la unión de estos metales pues se encuentran en muy poca concentración en los seres vivos (4).

El complejo ion ferroso- Fur se suele unir en las posiciones -35 y -10 de los promotores de los genes reprimidos por Fur. Estudiando los sitios de unión de Fur se extrajo una secuencia palindrómica de 19 pares de bases conocida como “iron box” o “Fur box” (Ilustración 2). Los sitios de unión de Fur pueden coincidir en 11/19 pares de bases con la secuencia consenso. Además, estudios de *footprinting* con DNase I muestran que Fur abarca una región mayor que la de “Fur box”, en concreto una región de unos 31 pares de bases, lo que indica que Fur interactúa con el DNA fuera de la región de “Fur box”. Diversos estudios sugieren que varios dímeros de Fur envuelven la

doble hélice del DNA como un sacacorchos, provocando así que Fur se una a regiones fuera de la “Fur box” (13). Hay diferentes modelos que intentan explicar cuál es la secuencia consenso de la caja Fur. Según su orden de aparición, se encuentra el modelo original, un modelo alternativo formado por hexámeros y otro modelo alternativo formado por heptámetros (14).

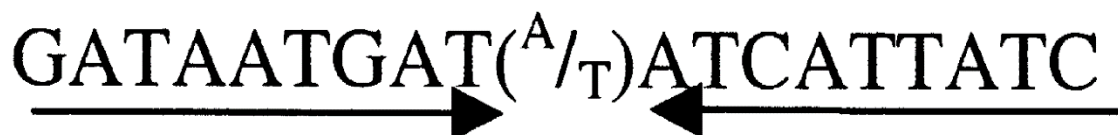


Ilustración 2. Secuencia consenso para el sitio de unión al ADN de Fur (Andrews et al., 2003).

Fur es una proteína muy abundante en bacterias como *E.coli*, alcanzando 10000 copias tras someterla a un estrés oxidativo (15), o 2500 copias en *V.cholerae* (16), aunque no se conoce cuál es el motivo de que existan tantas copias en estas bacterias cuando suelen haber pocas copias en los reguladores en general. Se sospecha que el alto número de copias de Fur puede deberse al gran número de genes controlados por Fur o porque tiene una función adicional como buffer del ion ferroso, uniéndose a este en la célula (3). Esta cantidad de genes puede deberse a que esta proteína participa también en funciones que no están directamente relacionadas con el metabolismo del hierro. Algunos de estos procesos son la respuesta al shock ácido, la defensa ante radicales de oxígeno, la bioluminiscencia y la producción de toxinas y otros factores de virulencia.

3.4. Fur box

Como se ha indicado en el punto anterior, existen diferentes modelos que tratan de explicar la estructura que sigue el lugar de unión de la proteína Fur al DNA. El sitio de unión se denomina Fur box o Iron box. Estos modelos arrastran una controversia, pues se han identificado diferentes modelos a lo largo del tiempo. Estudios realizados en 1987 por de Lorenzo, V. et al. (17) en *E.coli* sugerían como sitio de unión una secuencia de 19 pb palindrómica (GATAATGATAATCATTATC). Este modelo sería el original y el primer presentado. Tras esto, se propuso otro modelo, basado en la misma secuencia, pero dividida en hexámeros (13). Este modelo propone que Fur se une a 3 motivos del hexámero GATAAT. Finalmente, en 2002, Baicho et al., propone un modelo donde la secuencia a la que se une Fur se trata de dos palíndromos con una estructura 7-1-7 de 15 pares de bases que se superponen, formando una secuencia de 21 pares de bases (14).

El problema del modelo original es que es difícil de explicar cómo una proteína dimérica tan pequeña puede interactuar con una región operadora extendida, donde se ha visto que las proteínas que emplean un motivo de unión al DNA de hélice-giro-hélice (Helix-turn-helix), como es el caso de Fur, suelen tener una longitud de 12 pb los motivos más que de 19 pb (18). Además, estudios realizados mediante DNase I footprinting indican que la extensión que abarca Fur al interactuar con el DNA es de 30 pb, además de que Fur y sus homólogos suelen mostrar regios de unión al DNA extendidas en concentraciones altas de la proteína (19).

El segundo modelo surgió para dar respuesta a los problemas que hacían que el modelo original no cuadrara. Por ello, Escolar et al., tras realizar un estudio con oligonucleótidos sintéticos, propusieron un modelo donde la secuencia de 19 pares de bases podía ser estructurada como 3 motivos de seis pares de bases con la estructura directo-a-directo-a-inverso ($\rightarrow\rightarrow\leftarrow$) de la misma secuencia, GATAAT. Además, consideraron al hexámero GATAAT como la unidad mínima de reconocimiento para Fur. De esta forma, el núcleo del hexámero podría interactuar potencialmente con una proteína dimérica como Fur y que un mínimo de tres dímeros de unión se unirían de forma cooperativa, aunque el estudio relatado como ese modo de unión al DNA no se conoce entre reguladores de procariontes, aunque se parece a algunos tipos de unión de proteínas que contienen Zn en eucariotas (20).

En cuanto al tercer modelo a discutir, el modelo sostiene que el motivo de unión al DNA de Fur son dos motivos palindrómicos solapados de heptámeros. El cual es derivado de un estudio donde se realizó el alineamiento de sitios de unión al DNA para Fur en *B.subtilis*. Este modelo sugiere que cada proteína dimérica con la estructura hélice-giro-hélice reconoce secuencias en las ranuras principales adyacentes. Los dos dímeros reconocen secuencias compuestas por 6pb y se unen al DNA por las caras opuestas. En la investigación, descubrieron que los sitios formados por heptámeros conservaban la base central y que en el estudio de DNase I footprinting eran protegidos con una concentración de solo 10mM, lo que sugería que eran sitios de unión de gran afinidad. En este estudio, realizaron un estudio EMSA con una estructura similar pero formada por hexámeros (6-1-6), pero obtuvieron que Fur se unía con una afinidad significativamente reducida a este motivo de hexámeros, por lo que se concluyó que el motivo heptamérico era la unidad mínima de reconocimiento con gran afinidad para la unión de Fur (14).

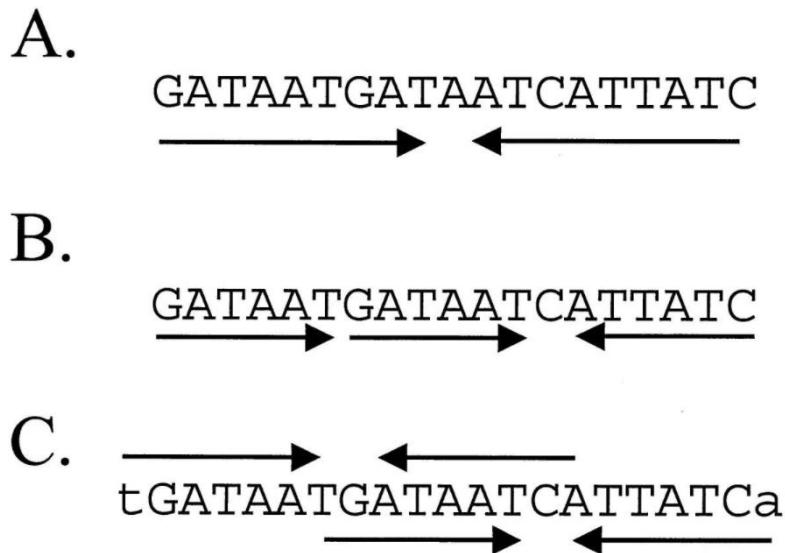


Ilustración 3. Secuencias de los diferentes modelos para explicar el motivo de unión al ADN de la proteína Fur (14).

Como se puede apreciar, no hay discrepancias en la secuencia a la cual se une Fur, sino en la interpretación del patrón funcional de esa secuencia. Existen estudios que discrepan de esto, donde creen que la secuencia consenso puede no representar completamente la secuencia diana para proteínas Fur bacterianas como un todo. Un ejemplo es el estudio realizado por Yali E. Friedman y Mark R. O'Brian en 2003, donde daban a conocer una secuencia de unión de Fur diferente de la expuesta en la Fur box (21). Esto fue realizado en *Bradyrhizobium japonicum* en un estudio de represión del gen *irr*, el cual está involucrado en el metabolismo del hierro. Además, la afinidad de unión con esta nueva secuencia era la misma que la de la secuencia consenso Fur box. Esto se daba en la especie estudiada, mientras que en *E.coli*, Fur no se unía a esta secuencia. Cabe añadir que este estudio no es algo que se repita, por lo que en la mayoría de las especies Fur reconoce la Fur box consenso. Este hallazgo puede sugerir que el modelo de unión a Fur consenso debería ser expandido o modificado (21).

3.5. Regulación del gen *fur*

El gen *fur* de *E.coli* se encuentra en el operón bicistrónico *fldA-fur*. Esta conformación no es exclusiva de *E.coli*, sino que se encuentra en otras bacterias como *Klebsiella pneumoniae* o *Yersinia pestis*. El gen *fldA* tiene la función de codificar la flavodoxina, una proteína relacionada con las reacciones redox. Este operón se induce por el sistema SoxRS como respuesta al estrés oxidativo (15). Además de esto, el gen *fur* tiene su propio promotor que se induce por OxyR en respuesta al estrés oxidativo provocado por el peróxido de hidrógeno. Con el estrés oxidativo, se produce un incremento de los niveles de Fur para aumentar la capacidad de unión a Fe²⁺, una

represión del transporte de hierro y una inducción de sistemas de almacenamiento de hierro. Estos efectos contribuyen a la homeostasis del hierro de forma que se reducen los niveles de hierro libre en el citosol, lo que debería ayudar a disminuir la toxicidad provocada por el hierro en la bacteria (3). Este gen *fur* también se autorregula mediante una secuencia localizada en la región intergénica del operón *fldA-fur*. También se ha reportado un control de la expresión de *fur* por la proteína receptora cAMP (Crp) (22).

La función principal de Fur es la represión de genes encargados de inducir la captación de hierro bajo condiciones ricas en hierro. Esta represión se produce para evitar el uso de sistemas de transporte cuando existe hierro en el ambiente por encima de 5-10 μM . Aunque esa sea su función primaria, existen varios genes diferentes reprimidos por Fur que no están relacionados con el hierro en bacterias como *E.coli*, como genes que inducen funciones como la respiración, la formación de flagelos para la quimiotaxis, la glicólisis, el metabolismo de purinas, la resistencia al estrés redox o la síntesis de metionina (23–25). En el caso de *E.coli*, se conocen 35 genes regulados por la proteína Fur (Ilustración 4). Debido a la alta influencia de Fur en varios procesos en la célula, se está empezando a considerar como un regulador global que ajusta el metabolismo a cambios externos de concentraciones de hierro, donde los genes que controla Fur se considerarían que forman parte de un “modulón” (3). Esta variedad de funciones de la proteína Fur no es exclusiva de *E.coli*, también existen casos similares en otras bacterias como en *Salmonella*, donde la inactivación de *fur* aumenta la sensibilidad al ácido y por lo tanto interviene en la respuesta al estrés producido por la acidez (26).

Fur- and iron-regulated genes in *E. coli* K-12

Gene	Function	+/-
<i>acnA</i>	Aconitase, [Fe-S] protein	+
<i>bfd</i>	Release of iron from Bfr?	-
<i>bfr</i>	Iron storage	+
<i>cir</i>	Ferric dihydroxybenzoate uptake	-
<i>cyoA</i>	Terminal respiratory oxidase subunit	-
<i>entABCDEF</i>	Enterobactin biosynthesis	-
<i>entS</i>	Export of enterobactin	-
<i>exbBC</i>	Siderophore and vitamin B ₁₂ transport	-
<i>fecABCDE</i>	Ferric dicitrate transport	-
<i>fepA</i>	Ferri-enterobactin transport	-
<i>fepBCDEG</i>	Ferri-enterobactin transport	-
<i>fes</i>	Ferri-enterobactin utilisation	-
<i>fhuABCD</i>	Ferric hydroxamate uptake	-
<i>fhuE</i>	Ferric coprogen rhodotorulate uptake	-
<i>fhuF</i>	Ferrioxamine utilisation	-
<i>fiu</i>	Dihydroxybenzoyl serine uptake	-
<i>flbB</i>	Motility	-
<i>flnA</i>	Iron storage	+
<i>fumA</i>	Aerobic fumarase, [Fe-S] protein	+
<i>fumB</i>	Anaerobic fumarase, [Fe-S] protein	+
<i>fumC</i>	Non-[Fe-S] fumarase	+
<i>fur</i>	Ferric uptake regulation	-
<i>nohA</i>	Phage recombinase	-
<i>nrhHIEF</i>	Deoxyribonucleotide reductase 2	-
<i>orf78</i>	Unknown	-
<i>gpmA</i>	Glycolysis	-
<i>metH</i>	Methionine biosynthesis	-
<i>nohB</i>	Phage function	-
<i>purR</i>	Purine regulon regulation	-
<i>ryhB</i>	Small regulatory RNA	-
<i>sdhCDAB</i>	TCA cycle	+
<i>sodA</i>	Mn-superoxide dismutase	-
<i>sodB</i>	Fe-superoxide dismutase	+
<i>tonB</i>	Siderophore and vitamin B ₁₂ transport	-
<i>ygaC</i>	Unknown	-
<i>yhhY</i>	Unknown	+

*+/-' indicates induction (+) or repression (-) by the Fe²⁺-Fur complex.

Ilustración 4. Genes regulados por Fur en *E. coli* (Andrews et al., 2003)

Existen casos donde se ha visto que parece que Fur activa genes o productos más que reprimirlos. Por ejemplo, en varios genes de *Escherichia coli*, como *acnA*, *bfr*, *fumA* o *fumB*, se conoce que dependen de Fur para su regulación. A pesar de esto, no se ha observado que estos genes posean una Fur box y por lo tanto no se ha demostrado que Fur actúe como un activador transcripcional directo. Esto sugiere que el papel que ejerce Fur como activador transcripcional parece que sea causa de motivos indirectos (4). Si que se conoce un caso en el que Fur actúa directamente como inductor. Esto ocurre en el gen de la ferritina (*pfr*) de *Helicobacter pylori*, el cual está directamente controlado por una interacción directa del ADN con Fur en presencia de hierro (27). Esta inducción se produce con la formación de un complejo Fe²⁺-Fur. Esta formación provoca que la forma libre de hierro de la proteína Fur se una al hierro y deje de reprimir la expresión del gen *pfr*, el cual estaba reprimido por la interacción directa de Fur con el promotor de *pfr* (3).

Esta activación indirecta por parte de Fur podría deberse a un gen reprimido por el complejo Fe²⁺-Fur, denominado *ryhB*. Este gen codifica para un fragmento pequeño no codificante de RNA (RyhB), el cual parece que actúa como un regulador

negativo de genes inducidos indirectamente por el complejo Fe^{2+} -Fur en *E.coli*. Aunque no se conoce que mecanismo produce la activación, se han encontrado homólogos de RyhB en otras bacterias que sugieren que esta inducción indirecta con el empleo de RyhB podría ser común en estas (28).

3.6. Fur y otras proteínas relacionadas

Fur da nombre a una superfamilia de proteínas sensibles a metales que regulan la transcripción de genes relacionados con la homeostasis de esos metales y con la respuesta al estrés oxidativo. Además de Fur, esta familia de metaloreguladores está formada por Zur, Mur, Nur, PerR e Irr (29). Estos homólogos de Fur han sido identificados en bacterias gram negativas, como *Yersinia*, *Salmonella* o *Helicobacter pylori*. También han sido identificados en algunas bacterias gram positivas, como *Staphylococcus*, e incluso en cianobacteria. Estos tienen diferentes funciones. Zur controla la captación del Zinc, Mur tiene un papel en la regulación de la toma de manganeso, Nur interfiere en la regulación de la adquisición de níquel, PerR reacciona con H_2O_2 y se encarga de realizar una respuesta al estrés creado por peróxidos e Irr depende del grupo hemo y regula la respuesta a concentraciones de hierro como Fur (29–31).

Esta familia de proteínas presenta ciertas particularidades generales. Por ejemplo, su forma dimérica se une de forma específica a regiones palindrómicas ricas en adeninas y timinas y emplean un metal como cofactor como correpressor (32). En todas ellas, la unión con el correpressor produce un cambio conformacional que incrementa la afinidad del regulador con el ADN. Además, las proteínas Fur comparten un mismo plegamiento y tienen una longitud media de 120 aminoácidos. A esto se le añade que normalmente esta familia se autorregula de acuerdo a la presencia de secuencias diana en sus propios promotores (29).

La proteína Zur (zinc uptake regulator), ha sido descrita en *E.coli* y *Listeria monocytogenes*. Además, esta proteína varía entre bacterias gramnegativas y grampositivas, de modo que las diferentes proteínas tienen un 25% de identidad de media, por lo que se les considera homólogos distantes. Zur se encarga de mantener la homeostasis del zinc mediante la represión de los sistemas de toma de zinc. Este metal es un micronutriente esencial al igual que el hierro, y como este, en concentraciones altas puede ser letal. Zur se une a este metal para regular su absorción, al igual que ocurre con Fur. Además, Fur y Zur tienen una similitud moderada. Zur lleva a cabo la regulación de la homeostasis del zinc actuando como un represor transcripcional que se une a este metal y bloquea el acceso de la ARN polimerasa a los promotores.

También se han reportado casos en los que Zur actúa de forma positiva en la regulación en *Xanthomonas campestris* y en *Neisseria meningitidis* (29,33–35).

Además de esta proteína, también se encuentra PerR, la cual responde al estrés oxidativo. El regulón PerR incluye genes que, aunque pueden ser reprimidos por Mn o hierro, la desrepresión ocurre con la reacción del complejo PerR-Fe²⁺ con peróxido de hidrógeno (4). Este complejo PerR-Fe²⁺ cataliza la oxidación de la histidina, lo que provoca la interrupción de la coordinación adecuada del hierro en el sitio regulador. PerR está autoregulada en la mayoría de los casos y se ve afectada por niveles de hierro, donde se disminuye la transcripción. La expresión de PerR no es afectada por otros metales. PerR tiene la habilidad de asociarse con hierro o manganeso “in vivo” para aumentar su afinidad de unión con el ADN, aunque esta unión en el caso del manganeso le impide catalizar la oxidación de la histidina (29,36). También existen casos donde PerR controla el almacenamiento de hierro además de controlar la resistencia al estrés oxidativo, como en *Staphylococcus aureus* (37), o donde PerR no es esencial para el estrés oxidativo, como en cianobacteria *Synechocystis PCC 6803* (38).

El níquel es un metal que se requiere para la oxidación del hidrógeno molecular. Está presente en muchas enzimas, como las ureasas, las deshidrogenasas de monóxido de carbono o las sintetetasas de acetil coenzima A. Esto provoca que sea necesario en procesos esenciales, aunque el número de reacciones en las que interviene sea muy pequeño en comparación al hierro o al zinc. La homeostasis del zinc en bacterias se controla por los represores NikR y RcnR, los cuales reaccionan a la presencia de níquel. Nur también tiene su papel en la homeostasis del zinc. Esta proteína regula negativamente el grupo de genes putativos de transporte de níquel *nikABCDE*, uniéndose a la región promotora de este gen en presencia de níquel e impidiendo así su transcripción. Además, también se ha detectado que Nur interviene en la respuesta antioxidativa de *S.coelicolor* (29).

La proteína Mur se ha encontrado en algunas alphaproteobacterias, concretamente en *Rhizobiales* y *Rhodobacteraceae*. En *Rhizobiales*, Mur controla un pequeño número de genes encargados de regular la toma de manganeso. Mur presenta una diferencia con los miembros de su familia, y es que no presenta un sitio de unión a Zn²⁺ como sus parálogos, sino que presenta una estructura donde por cada dímero tiene dos iones Mn²⁺ unidos. Mur actúa bajo condiciones de altas concentraciones de manganeso y se ha observado “in vitro” como esta proteína puede emplear Fe²⁺, Co²⁺, o Zn²⁺ como metales de activación para regular secuencias que dependen de Mur o secuencias “Fur box” (29,39).

La proteína Irr regula la biosíntesis de grupos hemo, y, al contrario que Fur, se activa bajo escasez de hierro (Lucía Escolar et al., 1999). Irr actúa como un regulador positivo y negativo de la expresión génica modulando genes relacionados con el metabolismo del hierro. Irr presenta un mecanismo de regulación distinto al de los miembros de su familia. Este se basa en la modificación traduccional llevada a cabo por el grupo hemo. El grupo hemo causa diferentes efectos en Irr, como la degradación de esta al interactuar con la ferroquelatasa en *B.japonicum* (40), o la disminución de la afinidad entre Irr y el operador en *R.leguminosarum* (41). Esta diferencia en la forma de modificar Irr por parte del grupo hemo parece que está relacionado con las diferencias en la interacción de estos (29).

3.7. CollecTF

La unión de los factores de transcripción a sus sitios de unión es muy importante en las redes regulatorias transcripcionales. En el caso de las bacterias, la transcripción recae en gran manera en las interacciones directas entre el ADN y los TF, y de esto ha derivado un interés en crear bases de datos que almacenen estos sitios de unión para explorar las redes regulatorias transcripcionales. El aumento de información sobre estos sitios de unión se debe al desarrollo de herramientas muy eficaces para la identificación de sitios de unión de factores de transcripción. Este aumento ha provocado el surgimiento de técnicas de genómica comparativa y *machine learning* para obtener una solución a la complejidad del reconocimiento de sitios de unión a través del modelaje de interacciones entre factores de transcripción y sitios de unión. Existen varias bases de datos que almacenan esta información, pero a menudo ocurre que las diferentes bases de datos emplean diferentes sistemas de ranking para la evidencia experimental y que a menudo se produce la combinación de sitios obtenidos experimentalmente y de predicciones computacionales. Como solución a esto surge CollecTF, una base de datos que complementa a otras bases de datos existentes y proporciona anotación de gran calidad en sitios de unión de TF a través de todo el dominio de bacteria. Además, proporciona información sobre la descripción del sitio de unión y varias fuentes de evidencia (42,43).

El objetivo principal de CollecTF es contener datos experimentales validados de sitios de unión de TF en su contexto genómico y así obtener una base de datos de alta calidad. Los datos de esta base de datos se obtienen a través de un equipo que obtiene los datos de forma manual revisando artículos científicos o bien por envío directo de los autores de un artículo científico. Las secuencias reportadas se buscan de manera automática en la secuencia de referencia y se verifica manualmente la localización

genómica. Para las secuencias que de forma automática se comprueba que coinciden se emplea información como la localización de la secuencia para verificar el sitio de unión. En cuanto a las secuencias que no coinciden exactamente, un conservador (curator) realiza una búsqueda de con hasta dos desajustes. Luego ordenan las secuencias de acuerdo a una matriz de peso de posición específica y da acceso a información sobre el método de obtención, de forma que el usuario pueda comprobar los estándares de calidad (42,44).

3.8. Genómica comparativa

Una de las características más sorprendentes de las bacterias es su capacidad de adaptarse a condiciones cambiantes. Así pues, se pueden observar adaptaciones para su supervivencia en medios acuáticos o terrestres, asociaciones con un huésped para sobrevivir como en la simbiosis o bien introducirse en un huésped al cual producen consecuencias patogénicas. Además, esas asociaciones o relaciones con huéspedes pueden ocurrir externa o internamente, donde en esta última forma pueden actuar encapsulándose o bien en un contacto directo con el citosol del huésped (45). Esta versatilidad se debe a una gran variedad genómica entre cepas bacterianas relacionadas. Esta evolución de las bacterias suele tener grandes implicaciones para los humanos, sobre todo en el caso de bacterias patógenas, como se puede observar por ejemplo en las cepas multirresistentes a fármacos. Esto ha provocado la necesidad de desarrollar estrategias para comprender y obtener información acerca de las trayectorias evolutivas de diferentes especies bacterianas (46).

Los recientes avances en la secuenciación de nueva generación han permitido obtener muchos datos sobre el ADN de las bacterias, permitiendo generar genomas de gran calidad del reino bacteria. Esta información puede emplearse para entender los mecanismos genéticos que permiten la existencia de las diversas estrategias para sobrevivir que emplean las bacterias, como la producción de elementos funcionales. Esta información puede ser anotada para identificar el papel que tiene en la biología celular y para mapearla en otros genomas, de forma que se puedan identificar genes homólogos. El problema es que el conocimiento sobre la biología procariota es limitado, por lo que interpretar esos datos se complica. Los estudios que han permitido adquirir conocimiento sobre la biología procariota han sido experimentos de pequeña escala en unos pocos organismos modelo, donde el coste es alto y no es factible realizar estas investigaciones para caracterizar genes desconocidos a grande escala. Por esto, se ha creado un especial interés en desarrollar métodos o programas computacionales para anotar funcionalmente genes nuevos e identificar los genes responsables de llevar a

cabo funciones como la posibilidad de causar enfermedad invadiendo a un huésped (47).

La genómica comparativa es una nueva disciplina que estudia dos o más genomas de especies diferentes para buscar adaptaciones genéticas específicas que expliquen funciones del material genético. Así pues, se han creado métodos para predecir la función de genes a través de búsquedas de similitud entre secuencias, en un intento de identificar genes homólogos de otros caracterizados previamente (47). El uso de estos métodos de predicción que emplean la similitud de secuencias tiene contratiempos, pues a través de una divergencia de una secuencia, una copia puede tener otra función diferente a la que tiene la copia con la que se compara (48). Además, puede haber la presencia de un gen xenólogo, donde a través de una transferencia horizontal de genes, un duplicado puede encontrarse en un genoma. Estos problemas son ampliamente conocidos y actualmente se están ideando soluciones. Además de la predicción de genes, también se puede emplear la genómica comparativa para identificar genes que conservan la parte funcional y características específicas de un clado que puede tener relación con una adaptación ambiental (46,49). La investigación de la genómica comparativa abarca el desarrollo de herramientas computacionales y su uso para analizar genomas. Para el estudio de las relaciones de homología entre genes se puede estudiar la similitud de las secuencias con herramientas como BLAST, GLASS, MUMmer o PatternHunter. Esta herramienta se puede emplear para analizar la estructura del genoma, incluyendo repeticiones, reordenamiento al nivel de ADN y gen y puntos de ruptura, para el estudio de regiones codificantes y no codificantes incluyendo la predicción de elementos regulatorios y para estudiar la conservación de elementos a través de genomas (50).

Es posible establecer perfiles filogenéticos para un gen que ayuden a revelar las relaciones entre varios genomas observando la presencia o ausencia de un gen en un genoma. Los genes funcionales relacionados se deberían mantener o desaparecer en diferentes genomas de una forma correlacionada. Las proteínas con perfiles similares están unidas funcionalmente (50,51).

Las regiones no codificantes abarcan el 97% de la longitud del genoma humano. Tiene gran importancia por su papel en la predicción de la regulación de la transcripción y la replicación del ADN entre otros. La genómica comparativa se ha empleado para identificar regiones conservadas comparando secuencias de ADN no codificante de diferentes especies (52). Esto se basa en la presunción de que la presión selectiva provoca que los elementos regulatorios evolucionen a una velocidad inferior que las secuencias no regulatorias en las regiones no codificantes, por lo que esas regiones se

conservarían. Esta idea se ha empleado con buenos resultados para el descubrimiento de elementos regulatorios involucrados en la regulación de la expresión génica de muchos genes. Además, se ha demostrado que la detección de regiones reguladoras aumenta en gran medida al aumentar el número de especies a comparar. Esta teoría se sustenta en la hipótesis de que las mutaciones afectan a las diferentes especies, pero los elementos regulatorios se deberían mantener, por lo que si se conserva una secuencia no codificante en más de dos organismos distintos significaría que esa secuencia tiene cierta funcionalidad para que se haya conservado en estos (53,54).

Otra estrategia para la identificación de sitios regulatorios consiste en combinar la identificación de regiones altamente conservadas con bases de datos para la búsqueda de sitios de unión de factores de transcripción, empleando un alineamiento global de regiones sinténicas, es decir, derivadas por especiación, y realizando una agrupación por grupos. Esta estrategia reduce el tiempo de análisis e incrementa la especificidad significativamente (50).

3.9. CGB

Los factores de transcripción se unen a regiones de una forma específica y pueden reprimir como potenciar la transcripción de operones diana. Los datos de secuencias genómicas pueden ser usados para identificar sitios diana putativos, y así, reconstruir la red transcripcional bajo el control de un factor de transcripción. Los métodos de genómica comparativa para reconstruir regulones bacterianos se basan en la premisa de que solo los sitios de unión de los factores de transcripción funcionales se conservan tras importantes lapsos evolutivos. Además de esto, encontrar un sitio de unión de un TF en la región promotora de operones ortólogos refuerza la premisa de que es un elemento funcional importante. Debido a la aplicación y potencial que tiene la genómica comparativa, se han desarrollado programas para analizar redes regulatorias bacterianas usando la información disponible de motivos para la predicción de operones ortólogos. CGB es una pipeline integrada para la comparación genómica de redes regulatorias bacterianas que no depende de muchas dependencias externas e introduce un marco de probabilidad para la integración y la reinterpretación de los resultados del análisis.

CGB implementa un pipeline para la reconstrucción comparativa de regulones bacterianos usando el conocimiento actual sobre la especificidad de la unión a TF. Para ejecutar CGB es necesario un archivo JSON con información sobre el número de acceso de una proteína a estudiar y una lista con sitios de unión alineados para al menos un ejemplo de TF, números de acceso para *chromids* o un conjunto de contigs para una o

más especies diana, y varios parámetros de configuración para CGB. Las instancias de referencia de TF se emplean para detectar ortólogos en cada genoma que se desee buscar y para crear un árbol filogenético de las instancias, el cual se emplea para combinar información sobre el sitio de unión del TF en una PSWM (Position specific Weight Matrix) para cada especie. Se predicen los operones y posteriormente se calcula su probabilidad de regulación. La probabilidad se calcula usando métodos de reconstrucción del estado ancestral (55).

CGB emplea un marco centrado en genes, donde el análisis comparativo y el informe de los regulones tienen como unidad fundamental de regulación el gen y los operones se convierten en unidades lógicas de regulación. La información de los motivos de unión a factores de transcripción puede ser transferida entre especies, pero la eficacia de esto es menor conforme se aumenta la distancia evolutiva. CGB estima una filogenia de la referencia proporcionada y los ortólogos diana de TF y usa las distancias inferidas entre estos para generar una PSWM para cada especie objetivo, similar al método usado en CLUSTALW (55).

La PSWM puede ser transformada a una *position-specific scoring matrix* (PSSM) para identificar sitios de unión de TF en secuencias genómicas. Debido a que usar una PSSM no es lo más adecuado para un marco de genómica comparativa, CGB implementa un marco probabilístico Bayesiano desarrollado para el análisis de regulones en secuencias metagenómicas, el cual calcula las probabilidades posteriores de regulación que facilita la interpretación y son directamente comparables entre especies (55).

Para predecir operones, CGB emplea una predicción basada en dos niveles. La distancia intergénica es un predictor efectivo y conocido para la predicción de operones. Los genes emparejados que son adyacentes, que tienen la misma orientación y que no tienen genes que intervienen en la hebra contraria, se consideran como parte de un operón si su distancia intergénica es inferior a un umbral establecido previamente. Este umbral varía según la distancia intergénica promedio en todas las direcciones dentro de un genoma. Debido a la posible separación de un operón durante la predicción, este umbral es conservador (55).

En cuanto a la detección de ortólogos, CGB emplea una detección automática de grupos ortólogos en las especies proporcionadas en el archivo de formato *json*, evitando emplear grupos de ortólogos precompilados, pues estos restringen la cantidad de especies bacterianas que pueden ser analizadas. Esta detección se realiza usando tBLASTX con los parámetros por defecto además de un e-valor de 10^{-10} . Con los hits

encontrados por BLAST se crea un gráfico donde se indican las mejores relaciones recíprocas de hits de BLAST, los productos de los genes y los grupos ortólogos de genes. Para reducir la intensidad computacional y debido a que la reconstrucción de regulones a través de genómica comparativa no requiere que se identifiquen todos los grupos de ortólogos, se define un parámetro en el archivo de entrada para la detección de ortólogos que determina un valor para el cual la probabilidad de regulación posterior de los genes debe ser mayor que este, donde en caso contrario no se detecta el ortólogo. Para la reconstrucción del estado ancestral, se implementa un bootstrapping para inferir la probabilidad de regulación de un operón en un árbol filogenético. También se emplea el paquete BayesTraits para inferir el estado de regulación discreto en los nodos ancestrales para cada replicado del bootstrap. Con esta información, se promedian los estados ancestrales sobre las réplicas creadas con el bootstrapping para finalmente obtener las probabilidades posteriores ancestrales de regulación (55).

4. Materiales y métodos

Para llevar a cabo el análisis comparativo se ha tenido que dividir en dos secciones, las cuales siguen los objetivos específicos de este trabajo. La primera sección, trata la obtención de los datos sobre la secuencia de los motivos reportados de estudios experimentales, la obtención de secuencias extendidas de esos motivos y la búsqueda de motivos que ayuden a clarificar cuales parece ser el modelo que parece que más se ajusta, para luego emplearlo en el estudio de la siguiente sección. Los microorganismos a estudiar han sido todos para los que había información sobre los motivos de Fur en la base de datos. La segunda sección, consiste en emplear el modelo del motivo escogido previamente y realizar el análisis comparativo.

4.1. Obtención modelo de unión de proteína Fur

4.1.1. Obtención y manipulación de los datos

El primer paso para comenzar el análisis ha sido obtener los datos sobre los motivos de unión al ADN de Fur. Los datos se han obtenido de la base de datos descrita previamente, CollecTF, accesible desde <http://www.collectf.org/browse/home/>. En esta, se ha buscado la proteína Fur y se ha empleado la función “combined report” que permite obtener todos los resultados en una misma página. A continuación, se ha creado un archivo con formato *json* donde se han creado secciones para cada especie de las bacterias encontradas en la base de datos. Para cada especie, se han creado subsecciones que contienen el nombre de la cepa de esa especie y el o los “accession numbers” correspondientes del *National Center for Biotechnology Information* (NCBI). A continuación, se han copiado en cada sección las secuencias encontradas en la pestaña “Aligned binding sites” de CollecTF para cada especie.

Una vez obtenido el archivo *json*, se creó un script de Python empleando el paquete *Biopython* para descargar los datos de cada especie (Anexo 4). En concreto, se empleó el módulo Entrez, con el cual se obtuvieron los datos de NCBI en formato GenBank para cada número de acceso del archivo *json* para obtener la secuencia completa de cada uno.

Una vez obtenidos los archivos, se crearon dos diccionarios, uno con los números de acceso y las secuencias completas correspondientes y otro con los números de acceso” y las secuencias reversa complementaria. Con estos diccionarios y el archivo *json* original, se buscó el motivo en la secuencia completa, se anotó la posición donde se había encontrado el motivo y se tomaron cien pares de bases por la

izquierda y por la derecha del motivo, creando una secuencia extendida del motivo. Estas secuencias extendidas se almacenaron en otro archivo *json*.

Finalmente, para tener las secuencias en un formato reconocible por el programa MEME, se crearon archivos con formato fasta a partir de los datos del archivo *json* con las secuencias extendidas.

4.1.2. Búsqueda del modelo del sitio de unión de Fur

Una vez tuvimos las secuencias extendidas en formato fasta que contienen los sitios de unión reportados de estudios experimentales, se buscaron diferentes motivos en estas secuencias para los tres modelos que se habían reportado (14). Para la búsqueda de los motivos, se empleó el software MEME suite y para comprobar los resultados se empleó FIMO.

MEME suite es un software que tiene un conjunto de herramientas que permite realizar distintos análisis de motivos, como motif discovery, motif-motif data base searching o motif-sequence database searching. Ofrece varios programas para llevar a cabo estas funciones, como MEME y GLAM2 para el descubrimiento de motivos, o FIMO, GLAM2SCAN y MAST para buscar ocurrencias de motivos en bases de datos de secuencias de motivos conocidos.

4.1.2.1. MEME

A continuación, se realizó una búsqueda de motivos con el programa MEME para los archivos en formato FASTA de las secuencias de las especies. Para esto, se instaló MEME suite de forma local. Tras esto, se buscaron motivos para el modelo original (a partir de ahora modelo A), el modelo de hexámeros (a partir de ahora modelo B) y el modelo de heptámeros (a partir de ahora modelo C). Para el modelo A, dado que este modelo sostiene que el motivo es palindrómico y está formado por 19 pares de bases, se buscaron motivos palindrómicos de entre 18 y 20 pares de bases, de forma que el programa tuviese cierto margen. Para el modelo B, se realizaron dos análisis distintos. En primer lugar, se realizó la búsqueda de motivos de longitud de entre 5 y 7 pares de bases para buscar el fragmento mínimo del modelo, el hexámero (Modelo B1). En esta primera búsqueda se estableció el método "Any Number of Repetitions" para ver si encontraban hexámeros unidos que diesen lugar a la estructura del modelo B. También se realizó la búsqueda de palíndromos formados por dos hexámeros con la estructura 6-1-6 con el modo "Any Number of Repetitions", para comprobar el modelo y si encontraba el palíndromo correspondiente a la parte final del modelo (Modelo B2). Para ello, se buscaron motivos palindrómicos de entre 12 y 14 pares de bases. Por

último, para el modelo C, se realizó una búsqueda similar al B, donde se buscó el palíndromo 7-1-7 (Modelo C1) y la unidad que lo forma, el heptámero (Modelo C2). La búsqueda de la estructura 7-1-7 se realizó determinando la búsqueda de motivos de una longitud de entre 14 y 16 pares de bases, con el parámetro para buscar palíndromos y con la opción “Any Number of Repetitions”, dado que se buscaban varios motivos. Por último, se buscaron motivos de una longitud de entre 6 y 8 pares de bases con el modo “Any number of Repetitions” para comprobar si se encontraban varios heptámeros en posiciones cercanas, de forma que visualmente se pudiese observar la estructura de heptámeros del modelo C.

Para esclarecer los resultados y debido a que los resultados de MEME para cada una de las especies eran confusos pues variaban entre cada especie y en algunas encontraba un motivo que cuadraba con la secuencia de unión de Fur y en otras especies no encontraba ningún motivo significativo, se decidió agrupar los organismos por clases empleando el buscador taxonómico de NCBI, accesible desde: <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>. En total, se encontraron seis clases: *acidithiobacillia*, *alphaproteobacteria*, *bacilli*, *betaproteobacteria*, *epsilonproteobacteria* y *gammaproteobacteria*. Una vez se tenían las clases, se creó un archivo fasta por cada clase, menos para *acidithiobacillia*, pues solo contenía un organismo. En estos archivos se introdujeron todas las secuencias FASTA de las especies que pertenecían a esa determinada clase. Después se volvieron a ejecutar todos los MEME que se han detallado previamente para cada modelo con estos nuevos archivos para observar si MEME encontraba algún motivo que se adecuase a algún modelo al aumentar el número de secuencias a comparar.

4.1.2.2. FIMO

Debido a que los motivos están solapados en el modelo C, MEME no muestra todos los motivos no muestra motivos superpuestos. Para evitar esto, se empleó FIMO para explorar mejor los motivos de cada modelo, además de poder comprobar las posiciones de los motivos encontrados.

FIMO forma parte del conjunto de herramientas MEME suite. Se emplea para buscar en bases de datos de secuencias, que puede ser suministrada por el usuario, coincidencias para los motivos descubiertos con MEME. Las siglas de FIMO significan “find individual motif occurrences”. Para ejecutar FIMO se necesita un archivo en formato meme, el cual contiene información sobre los resultados obtenidos en MEME. FIMO puntúa la coincidencia de cada motivo para cada posición en la base de datos de secuencias. El resultado de FIMO es una lista de coincidencias para las cuales la

puntuación obtenida es inferior a un límite definido por el usuario. Esta lista contiene las posiciones de los motivos encontrados, de forma que es posible observar que estructura forman o si se produce un solapamiento entre los motivos.

Para realizar el análisis con FIMO se creó un archivo de configuración para cada modelo que los resultados de MEME mostraban evidencias o sugerían que ese modelo era posible. Estos archivos se crearon con los datos para las frecuencias de las letras de cada base y con la matriz de probabilidad específica de posición, a partir de los resultados de MEME para análisis de las clases bacterianas de cada modelo en particular. Se crearon archivos para los modelos C1 y C2, pues eran los modelos que se tenía sospecha donde se quería buscar un posible solapamiento. Estos luego se emplearon para analizar por clases las secuencias fasta de las especies bacterianas. En concreto se analizaron las clases *bacilli* y *gammaproteobacteria* pues eran las que mostraban más resultados.

El resultado obtenido tras el análisis con FIMO es una lista de motivos donde se indica la cadena donde se ha encontrado el motivo, la posición de inicio y final, el p-valor de la coincidencia de ese motivo, el q-valor, que indica la especificidad del motivo respecto del conjunto de coincidencias y finalmente, la secuencia que ha coincidido. De esta forma, se puede inspeccionar para cada modelo las posiciones para ver si se sobreponen o si cuadran con algún modelo. Por ejemplo, para el modelo C1 se buscó si los motivos de 15 pares de bases se solapaban dando la estructura de sacacorchos mencionada en la literatura (14). Esto se buscó observando las posiciones de inicio de dos motivos encontrados en la misma secuencia, de forma que, si la segunda posición de inicio comenzaba 6 pares de bases hacia la derecha, el solapamiento ocurrido sería el del modelo C. Para el modelo C2 se buscó si los heptámeros encontrados se encontraban en posiciones cercanas.

4.2. Análisis de genómica comparativa sobre los datos

Una vez obtenido el modelo que según los datos parecía cuadrar más, se procedió a realizar el análisis de genómica comparativa. Para llevarlo a cabo, se empleó la librería de Python CGB. Como se ha descrito previamente, CGB se emplea para realizar análisis de genómica comparativa de regulación transcripcional en bacterias. CGB es distribuido como un paquete de conda, por lo que se instaló miniconda para crear un entorno donde instalar los paquetes necesarios para ejecutar CGB. En primer lugar, se clonó el repositorio de CGB de github accesible desde este enlace: <https://github.com/ErillLab/cgb>. También se instaló BLAST, CLUSTALO, BayesTraits y

HMMER debido a que CGB los emplea para realizar los alineamientos, además de diversas bases de datos.

4.2.1. Miniconda

Como se ha explicado antes, es necesario crear un entorno de conda para instalar los paquetes necesarios para poder ejecutar CGB.

Conda es un sistema de código abierto para la gestión de paquetes y la gestión de entornos que se puede ejecutar en Windows macOS y Linux. Conda permite crear entornos que contienen una configuración determinada que pueden ser guardados y enviados. Fue creado con la idea original de ser usado para programas que empleasen Python, aunque puede emplear software de cualquier lenguaje. Además, funciona como un gestor de paquetes y como un gestor de entornos, permitiendo cambiar de versión de Python o elegir los paquetes instalados simplemente cambiando el entorno. Existe un instalador mínimo de conda, llamado Miniconda. Este solo incluye conda, Python, los paquetes de los que dependen y un número reducido de paquetes útiles, como pip o zlib. Se accede mediante comandos en el terminal y permite instalar paquetes de conda. Para instalar Miniconda, se accedió al sitio web de conda para descargar desde este enlace: <https://docs.conda.io/en/latest/miniconda.html> el archivo con formato sh correspondiente a la versión de Python 2.7, que es la que emplea Python. Una vez instalado, se creó un entorno para cgb con Python 2.7.

Una vez creado el entorno, se instalaron los paquetes que emplea CGB, los cuales se listan a continuación:

- Biopython
- Cached-property
- Cyclcr
- Scipy
- Matplotlib
- Pyparsing
- Python-dateutil
- Six
- Pytz
- Tqdm
- Ete3
- Biocondaweblogo
- Networkx

4.2.2. Programas externos

CGB emplea cuatro programas externos para realizar el análisis, BLAST, CLUSTALO, BayesTraits y HMMER. A continuación, se va a describir cada uno de estos brevemente y su instalación.

BLAST (basic local alignment searchtool) es una herramienta que permite encontrar regiones de similitud local entre secuencias. Se encuentra accesible en: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Este programa compara secuencias de nucleótidos o de proteínas de una base de datos de secuencias con otras secuencias de consulta y calcula la significancia estadística de los alineamientos realizados. Existen diferentes variantes de BLAST que se emplean para comparar todas las combinaciones de secuencias de nucleótidos o de proteínas contra bases de datos de nucleótidos o proteínas (56,57). CGB emplea BLAST para inferir relaciones funcionales y evolutivas entre secuencias. Se instaló la versión de comandos de BLAST de forma local.

Para el alineamiento múltiple de secuencias, CGB emplea CLUSTALO, también llamado Clustal Omega, accesible en: <https://www.ebi.ac.uk/Tools/msa/clustalo/>. Clustal Omega es una versión reescrita y revisada del conjunto de programas Clustal para el alineamiento de secuencias múltiple, en concreto de las antiguas implementaciones ClustalX y ClustalW. CLUSTALO puede alinear un gran número de secuencias de ADN, ARN o proteínas rápidamente debido al uso del algoritmo mBed para calcular árboles-guía, el cual se construye a partir de distancias por pares entre las secuencias (58,59). Se instaló la versión local de CLUSTALO.

BayesTraits se encuentra en la carpeta que se localiza en el repositorio de GitHub de cgb. BayesTraits es un paquete que permite llevar a cabo análisis de rasgos evolutivos entre grupos de especies para el que se tiene una filogenia. Permite testear hipótesis evolucionarias como calcular los valores del estado ancestral o establecer correlaciones entre rasgos (60).

HMMer es una implementación de código abierto de perfiles de modelos ocultos de Markov (HMMs), los cuales se emplean gracias a la combinación de heurísticas de aceleración y optimizaciones computacionales y matemáticas. Los perfiles HMMs son modelos probabilísticos empleamos para la inferencia estadística de homología, creados a partir de un conjunto de secuencias representativas. Estos perfiles se emplean para modelar la estructura consenso de una familia de secuencias de proteínas. Existen dos métodos de búsqueda con HMMer, hmmsearch y hmmpfam, aunque ambos emplean la comparación de una secuencia a un perfil HMM como base. Con esta comparación se consigue obtener la probabilidad de que esa secuencia sea

creada por el perfil HMM proporcionado usando el algoritmo Viterbi. Esto provoca que el uso de HMMer sea lento, aunque el tiempo de ejecución depende del tamaño de la base de datos o de la secuencia a buscar (61–63).

HMMer se instaló de forma local mediante comandos en el terminal de Linux. Las bases de datos que se emplearon con HMMer son COG, PFAM y eggNOG. Para poder consultarlas, se descargaron los archivos correspondientes a los modelos ocultos de Markov (.hmm) para la versión actual de las bases de datos de las bacterias y se les asignó una ruta específica para luego usarlas con el análisis de CGB.

Los Clusters of Orthologous Groups (COGs) de proteínas permiten predecir funciones de proteínas a partir de la caracterización de proteínas relacionadas evolutivamente en grupos más específicos de ortólogos con una función similar. Debido a su función predictiva, la base de datos COG almacena estos datos de COGs de proteínas (64).

EggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) es una base de datos pública para anotaciones funcionales, historias evolutivas de genes y para relaciones de ortólogos. Tiene como función analizar miles de genomas a la vez para establecer relaciones de ortología entre los genes de estos genomas. EggNog proporciona anotaciones para los ortólogos inferidos, predicciones sobre los tres dominios de la vida y los virus, las cuales son creadas a partir de miles de genomas de estos, y relaciones basadas en análisis filogenéticos. Para llevar a cabo esto, se emplea un algoritmo de agrupamiento de las especies basado en el concepto de triangulación de los mejores aciertos recíprocos (65).

Por último, PFAM es una colección de dominios de proteínas y familias con varios usos como la anotación genómica. Estas familias son conjuntos de regiones proteicas que tienen cierta homología, es decir, comparten cierta similitud de secuencia. Cada familia es definida por 2 alineamientos y un perfil “hidden Markov model” (HMM) o modelo de Markov oculto. En esta base de datos, el perfil HMM se busca contra una gran colección de secuencias para encontrar todas las instancias de la familia. Estas secuencias son obtenidas de UNiProtKB, una base de datos que almacena información funcional de proteínas con una anotación precisa y consistente (66,67).

4.2.3. Análisis comparativo. CGB

Una vez preparado el entorno de conda e instalados las dependencias de CGB, se puede realizar el análisis comparativo. Como se ha detallado previamente, se quiere estudiar las diferencias entre la regulación en bacterias patógenas y no patógenas, para

estudiar el funcionamiento de Fur en cada una de estas. Por esto, dado que se tenían más especies bacterianas (y por lo tanto más datos) en la clase de las *gammaproteobacteria* al realizar la agrupación de los datos y que las *gammaproteobacteria* obtuvieron unos resultados concluyentes para MEME, se decidió realizar un análisis en *Enterobacteria*, en concreto en la familia *Yersiniaceae*, debido al conocimiento que hay en esta familia de bacterias sobre si son patógenas y no patógenas, ya que esto permitiría poder observar diferencias entre estas. Una vez elegida la familia sobre la cual hacer los análisis, se obtuvo el número de acceso y el nombre de las cepas representativas de cada especie de *Yersiniaceae*, los cuales se añadieron al archivo de configuración de CGB. Para que CGB ejecute sus análisis, es necesario proporcionarle un número de acceso de la base de datos NCBI para un genoma de referencia. Se escogió el genoma de referencia de *Yersinia pestis CO92* pues es la cepa de referencia y es donde se encuentran la mayoría de los datos experimentales para esa familia, lo que indica que ha sido muy estudiado. También se debe proporcionar un número de acceso para una proteína de referencia para el análisis, por lo que se escogió la proteína Fur para *Yersinia pestis CO92*. Por último, como sitios para el análisis comparativo, se introdujeron las secuencias correspondientes a los motivos obtenidos en MEME del modelo C que tras evaluar los resultados era el más consistente con los datos, además de algunos parámetros de configuración para la ejecución, como obtener como resultado un gráfico de los motivos y un mapa de calor, o la ruta donde se encontraban los datos de las diferentes bases de datos.

De manera que se pudiese estudiar y comparar los resultados con los resultados de CGB para el modelo C, se creó otro archivo para ejecutar CGB donde en el apartado de los sitios se introdujo solamente el motivo heptamérico del modelo C, dado que como se vio en el estudio de (14), estos sitios era protegidos en su estudio de DNase I footprinting, lo que sugería que eran sitios de gran afinidad.

También se creó un archivo con el número de acceso de la proteína RecA de *Yersinia pestis CO92*. Esto se realizó para obtener un árbol filogenético “puro”, pues se encuentra en todas las bacterias y obtendríamos resultados para todas, obteniendo así un árbol filogenético para poder compararlo con el obtenido para la proteína Fur y comprobar sus similitudes y diferencias.

Además de estos, también se crearon dos archivos más para *Yersiniaceae*. Uno con las secuencias correspondientes a las secuencias de 21 pares de bases del modelo C obtenidas con MEME para el modelo A, TGATAATGATAATCATTATCA. Esto se creó para poder analizar las diferencias que hubiese en la secuencia completa del modelo C y uno

de los motivos palindrómicos que lo forman. El otro, se creó sustituyendo el ID de la proteína Fur para *Yersinia Pestis* por el id de Fur correspondiente para enterobacteria.

Dados los buenos resultados obtenidos para *Bacilli* en MEME y hay estudios de Fur en *Bacillus subtilis*, por último, se crearon dos archivos más para dos géneros de *Bacilli*, *Bacillus* y *Listeria*. Para *Bacillus*, dado que hay más de 200 especies, se decidió escoger los primeros 24 registros, teniendo en cuenta que hubiese el mismo número de patógenos y no patógenos aproximadamente. En el archivo se incluyó las secuencias palindrómicas del modelo C1 encontradas con MEME para *Bacilli* y el número de acceso para la proteína Fur y el genoma de referencia para *Bacillus subtilis*. En cuanto a *Listeria*, se escogieron todas las especies disponibles dado que el número no era muy grande y para varias no existía un genoma completo por lo que no se incluirían. Al igual que en el caso de *Bacillus*, se introdujeron las secuencias del modelo C1 del resultado de MEME para *Bacilli*. En cuanto a los números de acceso, se introdujo el número para la proteína Fur de *Listeria monocytogenes* y el número de identificación del genoma de la misma bacteria.

Una vez obtenidos los archivos de configuración, se ejecuta con Python el script de CGB.

5. Resultados y discusión

En este apartado se expondrá en primer lugar los resultados de MEME y FIMO que han dado lugar a la elección del modelo que aparentemente constituye la disposición de los motivos de la Fur box para luego continuar los análisis empleando los motivos de ese modelo para el análisis de genómica comparativa, es decir, con CGB. Por esto, este apartado se dividirá en tres secciones, MEME, FIMO y CGB.

5.1. MEME

El primer resultado obtenido con MEME fue el conjunto de análisis de motivos de las 27 especies obtenidas al agrupar por especie las 35 cepas que disponían de secuencias con sitios de unión reportados en la literatura que fueron obtenidos experimentalmente. Para anotar las coincidencias se creó una tabla para anotar los resultados para cada modelo (tabla 1). En esta tabla se indicó con un uno si el motivo encontrado por MEME coincidía con el modelo que se estaba analizando y con un cero en caso contrario. Por ejemplo, en el modelo C1 si se encontraba el motivo de 15 pares de bases en una determinada especie se indicaba con un uno para esa especie. Esto se realizó para los cinco modelos propuestos y finalmente se agregó una última fila que sumase los valores 1 para ver cuántas especies seguían el modelo para cada especie.

Se obtuvo un total de una especie que tenía un motivo que parecía seguir el modelo A, 0 coincidencias para el modelo B1 y B2, 9 para el modelo C1 y 3 para el modelo C2. Aunque se esperaba que se encontrase al menos un motivo en alguno de los cinco modelos en cada especie esto no ocurrió. De hecho, el modelo con más coincidencias, el C1, obtuvo 9 especies con coincidencias del total de 27, es decir, solo en un 52,94% de las especies para ese modelo se encontraron motivos. Además, estas 18 especies que no tienen un motivo para el modelo C1 tampoco lo tienen para los demás modelos. Esto pudo deberse a que algunas especies bacterianas no disponían de muchas secuencias en los archivos fasta, por lo que MEME no fue capaz de reconocer e identificar el motivo de la Fur box que se había reportado en la literatura. Además de esto, se deben detallar algunos aspectos sobre los motivos encontrados en los modelos. En el modelo A, se encontraron motivos en las especies *Bacillus subtilis subsp. Subtilis* y *Caulobacter crescentus*, aunque el motivo encontrado no correspondía con el palíndromo de 19 pares de bases del modelo A, sino que coincidía más con un motivo de 17 pares de bases, AATGATAATCATTATCA, por lo que no se marcó como coincidencia del modelo A. Tan solo se encontró el motivo palindrómico de 19 pares de

bases en *E. coli*, aunque MEME obtenía el motivo desplazado. En cuanto al modelo B1 no se encontraron motivos hexaméricos en ninguna especie. Los motivos significativos encontrados tenían una longitud de 7 pares de bases y se correspondían con el motivo heptamérico buscado en el modelo C2, por lo que no se anotaron como coincidencias. Para el modelo B2 no se obtuvieron coincidencias con el palíndromo hexamérico. En cuanto al modelo C1, se encontraron 9 especies donde sí que se obtenía un palíndromo de 15 pares de bases, que solía corresponder con el segundo palíndromo del modelo C, TGATAATCATTATCA. Cabe mencionar que en algunos casos MEME detectó motivos de mayor longitud, donde añadía una o dos bases por la izquierda o por la derecha. Además, se podía observar como el motivo encontrado correspondía con la parte central de la secuencia (Ilustración 5), que era lo esperado pues las secuencias a analizar eran las secuencias donde se había reportado la Fur box más 100 pares de bases añadidos al extremo izquierdo y derecho, dejando la secuencia en la parte intermedia. Finalmente, para el modelo C2 se encontraron solamente 3 especies con motivos, donde se esperaba encontrar el mismo número al menos que el modelo C1 ya que el modelo C2 es un análisis de una parte del motivo del modelo C1. Esto ha podido deberse a que al ser un motivo tan pequeño MEME tuviese problemas para encontrarlo. Cabe destacar que, aunque los resultados no fueron muy buenos pues no se encontraron secuencias en todas las especies, sí que permitieron esclarecer que el modelo C1 era el que más se ajustaba por una amplia diferencia.

MOTIF LOCATIONS

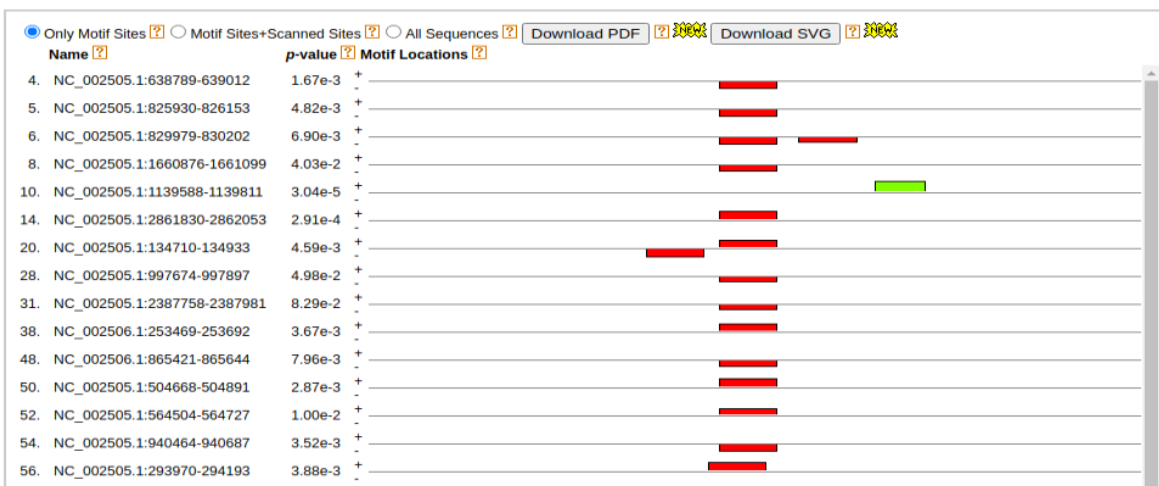


Ilustración 5. Resultado de MEME obtenido para *Vibrio cholerae* donde se muestra en rojo el motivo encontrado que coincide con el modelo C1.

La segunda parte de los resultados de MEME corresponde al descubrimiento de motivos en las secuencias agrupadas por clases. Al igual que en el caso anterior se contabilizó con una tabla los modelos con coincidencias (Anexo 2). Se obtuvieron coincidencias de nuevo para el modelo A, C1 y C2. No se encontraron coincidencias en

ningún modelo para las clases *betaproteobacteria* ni para *epsilonproteobacteria*, lo que refuerza el resultado anterior de que solo a 9 especies se les localizaron motivos. En cuanto al modelo A, MEME dio como resultado una secuencia de 21 pares de bases en las clases *alphaproteobacteria*, *bacilli* y *gammaproteobacteria*. Esta secuencia no contiene el palíndromo correspondiente al modelo A, aunque cabe destacar que la secuencia de 21 pares de bases encontrada corresponde a la secuencia completa del modelo C, tGATAATGATAATCATTATCa, donde la unión de los dos palíndromos de 15 pares de bases tiene esa misma longitud, pero desplazada 3 pares de bases. Al no encontrarse el motivo como tal, se contabilizó como no encontrado. El modelo C1 obtuvo coincidencias para las clases *alphaproteobacteria*, *bacilli* y *gammaproteobacteria*, donde, al igual que en el resultado del análisis por especies, se obtuvo, en la mayoría de los casos, el segundo motivo de la estructura de sacacorchos, TGATAATCATTATCA. Por último, el modelo C2 mostró coincidencias en *bacilli* y *gammaproteobacteria*, donde en ambos casos se obtiene un heptámero que forma parte del modelo C. Además, se observó como los heptámeros encontrados se alineaban formando la estructura palindrómica del modelo C1 (Ilustración 6).

MOTIF LOCATIONS

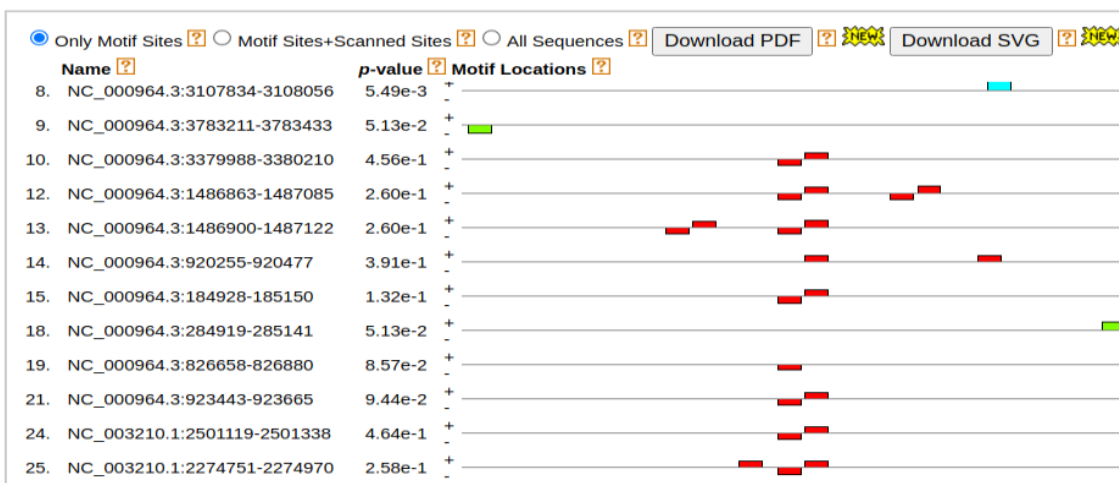


Ilustración 6. Motivos encontrados para el Modelo C2 en *Bacilli*, donde se observa cómo los motivos que coinciden con el heptámero del modelo C2 (en rojo) se agrupan para formar la estructura del modelo C1.

Tras la observación de todos los resultados obtenidos en conjunto, se concluyó que el modelo C1 era el más plausible.

5.2. FIMO

Con los resultados obtenidos tras realizar el análisis con FIMO se comprobaron las posiciones de los motivos encontrados a ver si se superponían. Para *Bacilli*, de tres especies a analizar, se encontraron motivos superpuestos en las tres especies, lo que

estructuras relacionadas con este. A continuación, se van a exponer los resultados encontrados.

Para el análisis realizado en *Yersiniaceae* se encontraron muchos genes con diferencias en la regulación entre patógenas y no patógenas. Comenzando desde arriba, por los genes con una probabilidad posterior de regulación más alta, se encuentra en la fila 14 de los resultados regulación para “MFS transporter”. Los transportadores MFS son portadores secundarios de un solo polipéptido capaces de transportar pequeños solutos como iones en respuesta al gradiente de iones quimiosmótico. Se encontró regulación para la mayoría de las especies de *Serratia* y en concreto, estaba más regulado en las patógenas. En cuanto a *Yersinia*, se encontró regulación para *Y.pestis* y *Y.pseudotuberculosis*. Esta regulación podría deberse al papel que tienen estos transportadores en la transferencia de iones como el hierro. Se han encontrado estudios donde en *Leishmania*, un género de protistas responsables de la enfermedad de las leishmaniosis, una proteína similar a las “MFS transporter”, *LIR1*, tiene un papel fundamental en la homeostasis del hierro, donde esta proteína media la exportación de hierro y así previene la acumulación de niveles tóxicos de este en los parásitos. Además, se encontró que esta proteína era un factor de virulencia que era requerido para la replicación intracelular (68).

Al igual que para “MFS transporter”, se encontró la misma regulación para “aerobactin synthase lucA” (COG4264). Las siglas de lucA significan “iron uptake chelate A” debido a que intervienen en la quelación del hierro. Esta la llevan a cabo a través de la biosíntesis de sideróforos, en este caso de la aerobactina. Los sideróforos son uno de los sistemas de adquisición de hierro más prominentes empleados en bacteria. Estos mecanismos de adquisición de hierro son ampliamente conocidos como factores de virulencia en muchas bacterias patógenas, permitiendo que estas se multipliquen y que causen infecciones. La habilidad de sintetizar ciertos sideróforos se ha asociado con la virulencia en varias bacterias patogénicas (69,70). Esto da indicios de la regulación de la proteína Fur en factores de virulencia y la relaciona con procesos infectivos, confirmando lo que se había descrito en el apartado del papel de Fur en la patogénesis. Además, también se encuentra regulación en “lucA/lucC family siderophore biosynthesis protein”, aunque este operón se encuentra en una posición más inferior, lo que indica que tiene una probabilidad de regulación posterior mucho menor. Este elemento se encuentra regulado en las mismas especies que “aerobactin synthase lucA”, es decir, en casi todas las especies patógenas de *Yersinia* y en todas las *Serratia* patógenas.

También se encontró regulación para “enterobactin transporter EntS” en *Serratia*. La enterobactina es un sideróforo de alta afinidad que se emplea para secuestrar el hierro en sistemas microbianos. Esta se encuentra principalmente en bacterias gram-negativas. El transportador de enterobactina EntS es un grupo de proteínas que exportan al sideróforo enterobactina fuera de la célula y pertenece a la familia MFS (Major Facilitator Superfamily) (71). Este grupo de proteínas está relacionado con la captación de hierro y con la virulencia provocada por los sideróforos.

También se encuentra “MurR/RpiR family transcriptional regulator”. Esta familia de reguladores transcripcionales se encuentra regulada solo en las bacterias pertenecientes a *Serratia* patógenas, donde del total de seis especies de *Serratia* donde se ha encontrado una regulación, cinco de ellas son patógenas. MurR es un miembro de la familia RpiR de reguladores transcripcionales, la cual está muy poco caracterizada. MurR reprime la expresión del operón murQ involucrado en la toma y degradación del ácido N-acetylmurámico. Además, este operón interviene en el crecimiento de las bacterias y en la formación de paredes celulares (72).

En la fila 39 del mapa de calor se encuentra “TonB-dependent receptor”. Este está muy regulado en el grupo de *Serratia* patógenas y solo se encuentra regulado en las no patógenas en *Serratia inhibens* PRI-2C, la cual tiene una función antifúngica. Además, en *Yersinia* solo se encuentra regulado en las dos cepas de *Yersinia pestis* y en *Yersinia pseudotuberculosis*, donde las dos especies son patógenas. Los receptores dependientes de Ton-B son proteínas de membrana que intervienen en el transporte activo de complejos sideróforo-hierro en bacterias gram-negativas. Existen varios estudios donde se describe que la expresión de los genes que codifican para estos receptores se encuentra bajo el control del represor Fur y activada en condiciones bajas de hierro (73). Esto refuerza el resultado obtenido en el mapa de calor sobre la regulación de Fur sobre estos receptores. Al igual que se ha mencionado previamente, que se encuentre regulado en patógenas encajaría con la relación que tiene el mecanismo de captación de hierro a través de sideróforos con la patogenia de estas bacterias. Esto ha sido descrito en varios estudios, donde se describe el papel de los sistemas dependientes de TonB como necesarios para la virulencia en cepas de *E.coli* (74), o en *Acinetobacter baumannii*, donde se sugiere que el receptor de Ton-B tiene un papel crítico en la virulencia (75).

También se encuentra regulación para “hemin-degrading factor” en *Yersinia* y *Serratia* para casi todas las cepas patógenas. Debido al secuestro del hierro en los tejidos del huésped, las bacterias patogénicas emplean a menudo mecanismos de

degradación de grupos hemo para obtener este hierro (76). La presencia de regulación en los resultados por parte de Fur del factor de degradación de grupos hemo indicaría que esta proteína tomaría un papel fundamental en los mecanismos de infección del huésped. De esta forma, Fur regularía la degradación de grupos hemo en el huésped y favorecería la captación del hierro (77).

Otro elemento encontrado regulado solo en bacterias patógenas es “peroxide/acid stress response protein YhcN”. Este elemento se ha encontrado en *Yersinia pestis*, *Yersinia pseudotuberculosis* y en *Gibbsiella quercinecans*. YhcN es una familia de proteínas conservadas. Se ha demostrado que representantes de esta familia tienen un papel importante en diferentes procesos celulares entre los cuales se encuentra la patogénesis y la respuesta al estrés (78). Un ejemplo de esto es la respuesta ante el estrés provocado por el peróxido de hidrógeno, el cadmio y el ácido en *E.coli*, donde esta proteína interviene en estos procesos (79).

Por último, encontramos la “yersiniabactin polyketide synthase HMWP1” en *Yersinia pestis* y en *Yersinia pseudotuberculosis*. La yersiniabactina es un sideróforo secretado por microbios patogénicos en entornos con deficiencia de hierro. Esta es un factor de virulencia empleado para obtener iones férricos del huésped, al igual que otros sideróforos mencionados previamente. Los genes de la producción de yersiniabactina en las cepas de *Y.pestis* se encuentran agrupadas en una isla de patogenicidad (80). Este sideróforo está formado por dos complejos denominados HMWP1 y HMWP2 (81). Este resultado sugiere una regulación de Fur en la biosíntesis de yersiniabactinas, implicando a su vez una relación con la virulencia en bacterias.

En cuanto a *Bacillus*, de los dos estudios de los diferentes modelos realizados solo se empleó el análisis del modelo de 15 pares de bases, ya que el resultado para el modelo de 21 pares de bases empleado para *Bacillus* era muy similar al obtenido con la secuencia del modelo C1 de 15 pares de bases, al igual que ocurrió para el modelo de *Yersiniaceae*. En este género, se encuentra como resultado la expansina “ExIX” predominantemente en bacterias patógenas. Las expansinas son proteínas no hidrolíticas de las que hay evidencia de que son importantes para la patogénesis en algunas especies bacterianas (82). La estructura de la expansina ExIX de *Bacillus subtilis* fue caracterizada en 2011, lo que refuerza el resultado obtenido. Esta parece ser requerida para la colonización o para causar enfermedades a tejidos de plantas (83). Esta relación con la regulación por parte de Fur reforzaría la idea del papel de Fur en varios procesos infectivos. Además, al igual que en *Yersiniaceae*, se encuentra

“lucA/lucC family siderophore biosynthesis protein”. Como se ha descrito previamente, interviene en la patogenicidad a través de la biosíntesis de sideróforos.

Por último, para el análisis de *Listeria*, encontramos en primer lugar el ortólogo “oligopeptidase f” regulado solamente para las especies patógenas *Listeria ivanovii* y *Listeria monocytogenes*. La oligopeptidasa f se ha descrito en *Bacillus* donde realiza una actividad inhibidora de la iniciación de la esporulación (84). Además, también se ha encontrado la proteína de competencia “CoiA” regulada de la misma manera que la oligopeptidasa f en las bacterias patógenas. La competencia es la habilidad de coger ADN y transformarse genéticamente. CoiA es una proteína que se induce su biosíntesis durante la competencia y se requiere para la transformación bacteriana. Se ha descrito *coiA*, el gen que codifica para la proteína CoiA, como un gen importante para la infección de huéspedes en *Streptococcus pneumoniae* (85). Este hallazgo que se encuentra solo para bacterias patógenas en *Listeria* podría indicar una similitud en el comportamiento de esta proteína para este género.

A pesar de que no ha sido posible analizar exhaustivamente todos los elementos obtenidos en los análisis debido a gran extensión de los resultados, se han descrito varios elementos regulados que permiten responder a la problemática planteada. En conjunto, se observan varias similitudes entre los distintos análisis llevados a cabo. Se encuentran algunos genes regulados mayoritariamente en bacterias patógenas y no regulados o regulados en pocas especies de bacterias no patógenas, sugiriendo un papel regulatorio de la patogenicidad en *Fur*. Dentro de los genes relacionados con la patogenicidad, se ha observado una gran regulación en genes relacionados con la biosíntesis de sideróforos. Relacionado con esto, se ha observado una coincidencia en la presencia de regulación de la familia *lucA* e *lucC* en la mayoría de las especies bacterianas analizadas pertenecientes a *Yersiniaceae* y a *Bacillus*, las cuales presentan grandes diferencias.

6. Conclusiones

En este trabajo de fin de máster se ha detallado la proteína Fur y varias cuestiones relacionadas con esta que han proporcionado un marco teórico sobre esta proteína y su regulación. Se ha seguido una estructura de un trabajo descriptivo en su mayor parte. En cuanto a los resultados, se han cumplido los objetivos específicos y generales propuestos en el trabajo.

Por lo que concierne al primer objetivo, se han analizado muchas secuencias con programas de descubrimiento de motivos, tras lo que se ha identificado un motivo con una estructura similar a uno de los propuestos en la literatura. Este motivo se ha supuesto como el regulón central de Fur al observar la presencia de este en varias especies. Este objetivo general se dividía a su vez en objetivos secundarios o específicos. De estos, se han obtenido las secuencias que contenían el sitio de unión de Fur al ADN y se han expandido para emplearlos en posteriores análisis. Con estos motivos, se ha llevado a cabo el segundo objetivo específico realizando análisis de descubrimiento de motivos aplicando diferentes opciones que se ajustasen a los diferentes modelos que se querían estudiar.

En cuanto al segundo objetivo, se han realizado varios análisis de genómica comparativa con la herramienta CGB donde se han encontrado diversos elementos regulados por Fur. Estos análisis se han realizado sobre diversos clados bacterianos, donde se ha podido encontrar diferencias y similitudes entre estos. Además, se han localizado diferencias para las especies bacterianas patógenas y no patógenas entre los elementos encontrados, lo que podría sugerir un papel activo de Fur en la patogenicidad de las bacterias a través de la regulación de diversos genes. Finalmente, se ha buscado en la literatura los elementos regulados encontrados en los análisis o similares y se han comparado con estos para dilucidar posibles funciones de estos en las bacterias encontradas o regulaciones de Fur no descritas sobre estos genes.

Cabe destacar que se requiere un mayor análisis de los resultados obtenidos en el análisis de genómica comparativa, los cuales no se han llevado a cabo debido a la gran extensión que tendrían. Se han encontrado muchos elementos regulados que podrían ayudar al análisis de Fur en otros clados bacterianos. En cuanto al motivo de unión de Fur al ADN, es necesario un mayor análisis. Esto se debe a que el motivo encontrado, a pesar de que era el que más similitudes obtenía con las secuencias experimentales, seguía sin encontrarse en muchas de ellas. Además, se obtuvieron otras secuencias de 17 pares de bases que podrían sugerir que existe otra conformación

estructural para el motivo de Fur, por lo que sería recomendable el estudio de la secuencia de unión a Fur de forma experimental.

7. Bibliografía

1. Panina EM, Mironov AA, Gelfand MS. Comparative analysis of FUR regulons in gamma-proteobacteria. *Nucleic Acids Res.* 2001;29(24):5195–206.
2. Archibald F. *Lactobacillus plantarum*, an organism not requiring iron. *FEMS Microbiol Lett.* 1983 Jun 1;19(1):29–32.
3. Andrews SC, Robinson AK, Rodríguez-Quiñones F. Bacterial iron homeostasis. *FEMS Microbiol Rev.* 2003;27(2–3):215–37.
4. Escolar L, Pérez-Martín J, De Lorenzo V. Opening the iron box: Transcriptional metalloregulation by the fur protein. *J Bacteriol.* 1999;181(20):6223–9.
5. Carpenter BM, Whitmire JM, Merrell DS. This is not your mother's repressor: The complex role of fur in pathogenesis. *Infect Immun.* 2009;77(7):2590–601.
6. Calderwood SB, Mekalanos JJ. Iron regulation of Shiga-like toxin expression of *Escherichia coli* is mediated by the fur locus. *J Bacteriol.* 1987;169(10):4759–64.
7. Stoebner JA, Payne SM. Iron-regulated hemolysin production and utilization of heme and hemoglobin by *Vibrio cholerae*. *Infect Immun* [Internet]. 1988 [cited 2021 Jun 7];56(11):2891–5. Available from: /pmc/articles/PMC259667/?report=abstract
8. Sebastian S, Agarwal S, Murphy JR, Genco CA. The gonococcal Fur regulon: Identification of additional genes involved in major catabolic, recombination, and secretory pathways. *J Bacteriol* [Internet]. 2002 [cited 2021 Jun 7];184(14):3965–74. Available from: <https://pubmed.ncbi.nlm.nih.gov/12081969/>
9. Gancz H, Censini S, Merrell DS. Iron and pH homeostasis intersect at the level of fur regulation in the gastric pathogen *Helicobacter pylori*. *Infect Immun* [Internet]. 2006 Jan [cited 2021 Jun 7];74(1):602–14. Available from: /pmc/articles/PMC1346641/
10. Bury-Moné S, Thiberge JM, Contreras M, Maitournam A, Labigne A, De Reuse H. Responsiveness to activity via metal ion regulators mediates virulence in the gastric pathogen *Helicobacter pylori*. *Mol Microbiol* [Internet]. 2004 Jul [cited 2021 Jun 7];53(2):623–38. Available from: <https://pubmed.ncbi.nlm.nih.gov/15228539/>
11. Rea RB, Gahan CGM, Hill C. Disruption of Putative Regulatory Loci in *Listeria monocytogenes* Demonstrates a Significant Role for Fur and PerR in Virulence. *Infect Immun* [Internet]. 2004 Feb [cited 2021 Jun 7];72(2):717–27. Available from: <https://pubmed.ncbi.nlm.nih.gov/14742513/>
12. Palyada K, Threadgill D, Stintzi A. Iron acquisition and regulation in *Campylobacter jejuni*. *J Bacteriol* [Internet]. 2004 Jul [cited 2021 Jun 7];186(14):4714–29. Available from: <https://pubmed.ncbi.nlm.nih.gov/15231804/>
13. Escolar L, Perez-Martín J, De Lorenzo V. Evidence of an unusually long operator for the Fur repressor in the aerobactin promoter of *Escherichia coli*. *J Biol Chem.* 2000;275(32):24709–14.
14. Baichoo N, Helmann JD. Recognition of DNA by Fur: A reinterpretation of the Fur box consensus sequence. *J Bacteriol.* 2002;184(21):5826–32.
15. Zheng M, Doan B, Schneider TD, Storz G. OxyR and SoxRS regulation of fur. *J Bacteriol.* 1999;181(15):4639–43.

16. Watnick PI, Eto T, Takahashi H, Calderwood SB. Purification of vibrio cholerae fur and estimation of its intracellular abundance by antibody sandwich enzyme-linked immunosorbent assay. *J Bacteriol.* 1997;179(1):243–7.
17. De Lorenzo V, Wee S, Herrero M, Neilands JB. Operator sequences of the aerobactin operon of plasmid colV--K30 binding the ferric uptake regulation (fur) repressor. *J Bacteriol.* 1987;169(6):2624–30.
18. Benevides JM, Thomas GJ, Weiss MA. DNA Recognition by the Helix-Turn-Helix Motif: Investigation by Laser Raman Spectroscopy of the Phage λ Repressor and Its Interaction with Operator Sites OL1 and Or3. *Biochemistry.* 1991;30(24):5955–63.
19. Le Cam E, Frechon D, Barray M, Fourcade A, Delain E. Observation of binding and polymerization of Fur repressor onto operator- containing DNA with electron and atomic force microscopes. *Proc Natl Acad Sci U S A.* 1994;91(25):11816–20.
20. Escolar L, Pérez-Martín J, De Lorenzo V. Binding of the Fur (ferric uptake regulator) repressor of *Escherichia coli* to arrays of the GATAAT sequence. *J Mol Biol.* 1998 Oct 30;283(3):537–47.
21. Friedman YE, O’Brian MR. A novel DNA-binding site for the ferric uptake regulator (Fur) protein from *Bradyrhizobium japonicum*. *J Biol Chem [Internet].* 2003;278(40):38395–401. Available from: <http://dx.doi.org/10.1074/jbc.M306710200>
22. DE LORENZO V, HERRERO M, GIOVANNINI F, NEILANDS JB. Fur (ferric uptake regulation) protein and CAP (catabolite-activator protein) modulate transcription of fur gene in *Escherichia coli*. *Eur J Biochem [Internet].* 1988 [cited 2021 Jun 7];173(3):537–46. Available from: <https://pubmed.ncbi.nlm.nih.gov/2836193/>
23. Stojiljkovic I, Bäumlér AJ, Hantke K. Fur regulon in gram-negative bacteria: Identification and characterization of new iron-regulated *Escherichia coli* genes by a Fur titration assay. *J Mol Biol [Internet].* 1994 Feb 17 [cited 2021 Jun 7];236(2):531–45. Available from: <https://pubmed.ncbi.nlm.nih.gov/8107138/>
24. Park SJ, Gunsalus RP. Oxygen, iron, carbon, and superoxide control of the fumarase fumA and fumC genes of *Escherichia coli*: Role of the arcA, fnr, and soxR gene products. *J Bacteriol [Internet].* 1995 [cited 2021 Jun 7];177(21):6255–62. Available from: <https://pubmed.ncbi.nlm.nih.gov/7592392/>
25. Vassinova N, Kozyrev D. A method for direct cloning of Fur-regulated genes: Identification of seven new Fur-regulated loci in *Escherichia coli*. *Microbiology [Internet].* 2000 [cited 2021 Jun 7];146(12):3171–82. Available from: <https://pubmed.ncbi.nlm.nih.gov/11101675/>
26. Foster JW, Hall HK. Effect of *Salmonella typhimurium* ferric uptake regulator (fur) mutations on iron- and pH-regulated protein synthesis. *J Bacteriol [Internet].* 1992 [cited 2021 Jun 7];174(13):4317–23. Available from: </pmc/articles/PMC206215/?report=abstract>
27. Delany I, Spohn G, Rappuoli R, Scarlato V. The Fur repressor controls transcription of iron-activated and -repressed genes in *Helicobacter pylori*. *Mol Microbiol [Internet].* 2001 Dec 1 [cited 2021 Jun 7];42(5):1297–309. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-2958.2001.02696.x>
28. Massé E, Gottesman S. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A [Internet].* 2002 Apr 2 [cited

- 2021 Jun 7];99(7):4620–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/11917098/>
29. Fillat MF. The fur (ferric uptake regulator) superfamily: Diversity and versatility of key transcriptional regulators. *Arch Biochem Biophys* [Internet]. 2014;546:41–52. Available from: <http://dx.doi.org/10.1016/j.abb.2014.01.029>
 30. Bsat N, Herbig A, Casillas-Martinez L, Setlow P, Helmann JD. *Bacillus subtilis* contains multiple Fur homologues: Identification of the iron uptake (Fur) and peroxide regulon (PerR) repressors. *Mol Microbiol*. 1998;29(1):189–98.
 31. Kim JH, Ji CJ, Ju SY, Yang YM, Ryu SH, Kwon Y, et al. *Bacillus licheniformis* contains two more PerR-like proteins in addition to PerR, Fur, and Zur orthologues. *PLoS One*. 2016;11(5):1–19.
 32. Bags A, Neilands JB. Ferric Uptake Regulation Protein Acts as a Repressor, Employing Iron(II) as a Cofactor To Bind the Operator of an Iron Transport Operon in *Escherichia coli*. *Biochemistry* [Internet]. 1987 [cited 2021 Jun 13];26(17):5471–7. Available from: <https://pubs.acs.org/doi/pdf/10.1021/bi00391a039>
 33. Huang DL, Tang DJ, Liao Q, Li HC, Chen Q, He YQ, et al. The Zur of *Xanthomonas campestris* functions as a repressor and an activator of putative zinc homeostasis genes via recognizing two distinct sequences within its target promoters. *Nucleic Acids Res* [Internet]. 2008 Aug 1 [cited 2021 Jun 13];36(13):4295–309. Available from: <https://academic.oup.com/nar/article/36/13/4295/2410001>
 34. Pawlik M, Christin, Hubert K, Joseph B, Claus H, Schoen C, Vogel U. The zinc-responsive regulon of *Neisseria meningitidis* comprises 17 genes under control of a Zur element. *J Bacteriol* [Internet]. 2012 Dec [cited 2021 Jun 13];194(23):6594–606. Available from: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
 35. Eckelt E, Jarek M, Frömke C, Meens J, Goethe R. Identification of a lineage specific zinc responsive genomic island in *Mycobacterium avium* ssp. *paratuberculosis*. *BMC Genomics*. 2014;15(1):1–15.
 36. Jacquamet L, Traoré DAK, Ferrer JL, Proux O, Testemale D, Hazemann JL, et al. Structural characterization of the active form of PerR: Insights into the metal-induced activation of PerR and fur proteins for DNA binding. *Mol Microbiol* [Internet]. 2009 Jul 1 [cited 2021 Jun 13];73(1):20–31. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2958.2009.06753.x>
 37. Horsburgh MJ, Ingham E, Foster SJ. In *Staphylococcus aureus*, Fur is an interactive regulator with PerR, contributes to virulence, and is necessary for oxidative stress resistance through positive regulation of catalase and iron homeostasis. *J Bacteriol* [Internet]. 2001 [cited 2021 Jun 13];183(2):468–75. Available from: <https://pubmed.ncbi.nlm.nih.gov/11133939/>
 38. Singh AK, McIntyre LM, Sherman LA. Microarray analysis of the genome-wide response to iron deficiency and iron reconstitution in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol* [Internet]. 2003 Aug 1 [cited 2021 Jun 13];132(4):1825–39. Available from: www.plantphysiol.org/cgi/doi/10.1104/pp.103.024018.
 39. Bellini P, Hemmings AM. In vitro characterization of a bacterial manganese uptake regulator of the fur superfamily. *Biochemistry* [Internet]. 2006 Feb 28 [cited 2021 Jun 13];45(8):2686–98. Available from: <https://pubs.acs.org/doi/abs/10.1021/bi052081n>
 40. Qi Z, O'Brian MR. Interaction between the bacterial iron response regulator and

- ferrochelatase mediates genetic control of heme biosynthesis. *Mol Cell*. 2002 Jan 1;9(1):155–62.
41. Singleton C, White GF, Todd JD, Marritt SJ, Cheesman MR, Johnston AWB, et al. Heme-responsive DNA binding by the global iron regulator *Irr* from *Rhizobium leguminosarum*. *J Biol Chem*. 2010 May 21;285(21):16023–31.
 42. Kiliç S, White ER, Sagitova DM, Cornish JP, Erill I. CollecTF: A database of experimentally validated transcription factor-binding sites in Bacteria. *Nucleic Acids Res*. 2014;42(D1):156–60.
 43. Kiliç S, Sagitova DM, Wolfish S, Bely B, Courtot M, Ciufo S, et al. From data repositories to submission portals: Rethinking the role of domain-specific databases in CollecTF. *Database*. 2016;2016:1–10.
 44. Erill I. Every site counts: Submitting transcription factor-binding site information through the CollecTF portal. *J Bacteriol*. 2015;197(15):2454–7.
 45. Silva MT. Classical labeling of bacterial pathogens according to their lifestyle in the host: Inconsistencies and alternatives. *Front Microbiol*. 2012;3(FEB):1–7.
 46. Monk J, Bosi E. Integration of comparative genomics with genome-scale metabolic modeling to investigate strain-specific phenotypical differences. *Methods Mol Biol*. 2018;1716:151–75.
 47. Bundalovic-Torma C, Parkinson J. Comparative genomics and evolutionary modularity of prokaryotes. *Adv Exp Med Biol*. 2015;883:77–96.
 48. Taylor JS, Raes J. Duplication and divergence: The evolution of new genes and old ideas. *Annu Rev Genet*. 2004;38:615–43.
 49. Gabaldón T, Koonin E V. Functional and evolutionary implications of gene ontology. *Physiol Behav*. 2016;176(1):100–106.
 50. Wei L, Liu Y, Dubchak I, Shon J, Park J. Comparative genomics approaches to study organism similarities and differences. *J Biomed Inform*. 2002;35(2):142–50.
 51. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A* [Internet]. 1999 Apr 13 [cited 2021 Jun 5];96(8):4285–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10200254>
 52. Hardison RC. Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet*. 2000;16(9):369–72.
 53. Dubchak I, Brudno M, Loots GG, Pachter L, Mayor C, Rubin EM, et al. Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res* [Internet]. 2000 [cited 2021 Jun 6];10(9):1304–6. Available from: </pmc/articles/PMC310906/>
 54. Frazer KA, Sheehan JB, Stokowski RP, Chen X, Hosseini R, Cheng JF, et al. Evolutionarily conserved sequences on human chromosome 21. *Genome Res*. 2001;11(10):1651–9.
 55. Kiliç S, Sánchez-Osuna M, Collado-Padilla A, Barbé J, Erill I. Flexible comparative genomics of prokaryotic transcriptional regulatory networks. *BMC Genomics*. 2020;21(Suppl 5):1–11.
 56. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J*

- Mol Biol. 1990;215(3):403–10.
57. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: Architecture and applications. *BMC Bioinformatics*. 2009;10:1–9.
 58. Sievers F, Higgins DG. The Clustal Omega Multiple Alignment Package. In: *Methods in Molecular Biology* [Internet]. Humana Press Inc.; 2021 [cited 2021 Jun 3]. p. 3–16. Available from: https://link.springer.com/protocol/10.1007/978-1-0716-1036-7_1
 59. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci*. 2018;27(1):135–45.
 60. Meade A, Pagel M. BayesTraits V3.0.2 November. Vol. 53, *Journal of Chemical Information and Modeling*. 2013. p. 1689–99.
 61. Prakash A, Jeffryes M, Bateman A, Finn RD. The HMMER Web Server for Protein Sequence Similarity Search. *Curr Protoc Bioinforma*. 2017;60(1):3.15.1-3.15.23.
 62. Oliver T, Yeow LY, Schmidt B. High performance database searching with HMMer on FPGAs. *Proc - 21st Int Parallel Distrib Process Symp IPDPS 2007; Abstr CD-ROM*. 2007;(January).
 63. Takagi T, Maruyama T. Accelerating hmmer search using FPGA. In: *FPL 09: 19th International Conference on Field Programmable Logic and Applications*. 2009. p. 332–7.
 64. Dibrova D V., Konovalov KA, Perekhvatov V V., Skulachev K V., Mulkidjanian AY. COGcollator: A web server for analysis of distant relationships between homologous protein families. *Biol Direct*. 2017;12(1):1–11.
 65. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47(D1):D309–14.
 66. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: The protein families database. *Nucleic Acids Res*. 2014;42(D1):222–30.
 67. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. *Nucleic Acids Res*. 2012;40(D1):290–301.
 68. Laranjeira-Silva MF, Wang W, Samuel TK, Maeda FY, Michailowsky V, Hamza I, et al. A MFS-like plasma membrane transporter required for *Leishmania* virulence protects the parasites from iron toxicity. *PLoS Pathog* [Internet]. 2018 Jun 1 [cited 2021 Jun 22];14(6). Available from: </pmc/articles/PMC6021107/>
 69. Bailey DC, Drake EJ, Grant TD, Gulick AM. Structural and functional characterization of aerobactin synthetase *lucA* from a hypervirulent pathotype of *Klebsiella pneumoniae*. *Biochemistry* [Internet]. 2016 Jun 28 [cited 2021 Jun 22];55(25):3559–70. Available from: </pmc/articles/PMC4928626/>
 70. Russo TA, Olson R, MacDonald U, Metzger D, Maltese LM, Drake EJ, et al. Aerobactin mediates virulence and accounts for increased siderophore production under iron-limiting conditions by hypervirulent (hypermucoviscous) *Klebsiella pneumoniae*. *Infect Immun* [Internet]. 2014 [cited 2021 Jun 22];82(6):2356–67. Available from: </pmc/articles/PMC4019165/>
 71. Furrer JL, Sanders DN, Hook-Barnard IG, McIntosh MA. Export of the siderophore

- enterobactin in *Escherichia coli*: Involvement of a 43 kDa membrane exporter. *Mol Microbiol.* 2002;44(5):1225–34.
72. Jaeger T, Mayer C. The transcriptional factors MurR and catabolite activator protein regulate N-acetylmuramic acid catabolism in *Escherichia coli*. *J Bacteriol* [Internet]. 2008 Oct [cited 2021 Jun 22];190(20):6598–608. Available from: <https://journals.asm.org/journal/jb>
 73. Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denancé N, et al. Plant carbohydrate scavenging through TonB-dependent receptors: A feature shared by phytopathogenic and aquatic bacteria. *PLoS One* [Internet]. 2007 Feb 21 [cited 2021 Jun 22];2(2). Available from: [/pmc/articles/PMC1790865/](https://pubmed.ncbi.nlm.nih.gov/1790865/)
 74. Torres AG, Redford † Peter, Welch RA, Payne SM. TonB-Dependent Systems of Uropathogenic *Escherichia coli*: Aerobactin and Heme Transport and TonB Are Required for Virulence in the Mouse. *Infect Immun.* 2001;69(10):6179–85.
 75. Abdollahi S, Rasooli I, Mousavi Gargari SL. The role of TonB-dependent copper receptor in virulence of *Acinetobacter baumannii*. *Infect Genet Evol.* 2018 Jun 1;60:181–90.
 76. Lyles K V., Eichenbaum Z. From host heme to iron: The expanding spectrum of heme degrading enzymes used by pathogenic bacteria [Internet]. Vol. 8, *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A.; 2018 [cited 2021 Jun 22]. Available from: [/pmc/articles/PMC6018153/](https://pubmed.ncbi.nlm.nih.gov/3218153/)
 77. Richard KL, Kelley BR, Johnson JG. Heme uptake and utilization by gram-negative bacterial pathogens [Internet]. Vol. 9, *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A.; 2019 [cited 2021 Jun 22]. p. 81. Available from: www.frontiersin.org
 78. Eletsky A, Michalska K, Houlston S, Zhang Q, Daily MD, Xu X, et al. Structural and functional characterization of DUF1471 domains of *Salmonella* proteins SrfN, YdgH/SssB, and YahO. *PLoS One* [Internet]. 2014 Jul 10 [cited 2021 Jun 22];9(7). Available from: [/pmc/articles/PMC4092069/](https://pubmed.ncbi.nlm.nih.gov/2609269/)
 79. Lee J, Hiibel SR, Reardon KF, Wood TK. Identification of stress-related proteins in *Escherichia coli* using the pollutant cis-dichloroethylene. *J Appl Microbiol.* 2010 Jun;108(6):2088–102.
 80. Miller DA, Luo L, Hillson N, Keating TA, Walsh CT. Yersiniabactin synthetase: A four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*. *Chem Biol.* 2002 Mar 1;9(3):333–44.
 81. Gehring AM, DeMoll E, Fetherston JD, Mori I, Mayhew GF, Blattner FR, et al. Iron acquisition in plague: Modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem Biol* [Internet]. 1998 [cited 2021 Jun 22];5(10):573–86. Available from: <https://pubmed.ncbi.nlm.nih.gov/9818149/>
 82. Narváez-Barragán DA, Tovar-Herrera OE, Torres M, Rodríguez M, Humphris S, Toth IK, et al. Expansin-like Exl1 from *Pectobacterium* is a virulence factor required for host infection, and induces a defence plant response involving ROS, and jasmonate, ethylene and salicylic acid signalling pathways in *Arabidopsis thaliana*. [cited 2021 Jun 23]; Available from: <https://doi.org/10.1038/s41598-020-64529-9>
 83. Martinez-Anaya C. Understanding the structure and function of bacterial expansins: a prerequisite towards practical applications for the bioenergy and agricultural industries

[Internet]. Vol. 9, Microbial Biotechnology. John Wiley and Sons Ltd; 2016 [cited 2021 Jun 23]. p. 727–36. Available from: /pmc/articles/PMC5072189/

84. Kanamaru K, Stephenson S, Perego M. Overexpression of the PepF oligopeptidase inhibits sporulation initiation in *Bacillus subtilis*. *J Bacteriol* [Internet]. 2002 [cited 2021 Jun 23];184(1):43–50. Available from: /pmc/articles/PMC134765/
85. Zhu L, Lin J, Kuang Z, Vidal JE, Lau GW. Deletion analysis of *Streptococcus pneumoniae* late competence genes distinguishes virulence determinants that are dependent or independent of competence induction. *Mol Microbiol* [Internet]. 2015 Jul 1 [cited 2021 Jun 23];97(1):151–65. Available from: /pmc/articles/PMC4536566/

8. Anexos

Anexo 1. Bacterias de las cuales se han obtenido secuencias para el análisis.

Acidithiobacillia:

- *Acidithiobacillus ferrooxidans*

Alphaproteobacteria:

- *Bradyrhizobium diazoefficiens*
- *Caulobacter crescentus*

Bacilli:

- *Bacillus subtilis*
- *Listeria monocytogenes*
- *Staphylococcus aureus*

Betaproteobacteria:

- *Bordetella pertussis*.
- *Neisseria gonorrhoeae*
- *Neisseria meningitidis*
- *Nitrosomonas europaea*

Epsilonproteobacteria:

- *Helicobacter hepaticus*
- *Helicobacter pylori*.

Gammaproteobacteria:

- *Actinobacillus pleuropneumoniae* serovar
- *Aliivibrio salmonicida*
- *Dickeya dadantii*
- *Escherichia coli*
- *Pseudomonas aeruginosa*
- *Pseudomonas syringae*
- *Salmonella entérica*
- *Shewanella oneidensis*
- *Shewanella piezotolerans*
- *Vibrio anguillarum*
- *Vibrio cholerae*
- *Vibrio fischeri*
- *Vibrio vulnificus*
- *Yersinia pestis*

Anexo 2. Resultados de MEME para los diferentes modelos en las diferentes especies.

Especies bacterianas	Modelos				
	A	B1	B2	C1	C2
<i>Acidithiobacillus ferrooxidans</i>	0	0	0	0	0
<i>Actinobacillus pleuropneumoniaes</i> serovar	0	0	0	0	0
<i>Aliivibrio salmonicida</i>	0	0	0	0	0
<i>Bacillus subtilis</i> subsp. <i>Subtilis</i>	0	0	0	1	0
<i>Bordetella pertussis</i>	0	0	0	0	0
<i>Bradyrhizobium diazoefficiens</i>	0	0	0	0	0
<i>Caulobacter crescentus</i>	0	0	0	1	1
<i>Dickeya dadantii</i>	0	0	0	0	0
<i>Escherichia coli</i>	1	0	0	1	0
<i>Helicobacter hepaticus</i>	0	0	0	0	0
<i>Helicobacter pylori</i>	0	0	0	0	0
<i>Listeria monocytogenes</i>	0	0	0	1	0
<i>Neisseria gonorrhoeae</i>	0	0	0	0	0
<i>Neisseria meningitidis</i>	0	0	0	0	0
<i>Nitrosomonas europaea</i>	0	0	0	0	0
<i>Nostoc</i> sp	0	0	0	0	0
<i>Pseudomonas aeruginosa</i>	0	0	0	1	1
<i>Pseudomonas syringae</i>	0	0	0	1	0
<i>Salmonella enterica</i>	0	0	0	1	0
<i>Shewanella oneidensis</i>	0	0	0	0	0
<i>Shewanella piezotolerans</i>	0	0	0	0	0
<i>Staphylococcus aureus</i>	0	0	0	0	0
<i>Vibrio anguillarum</i>	0	0	0	0	0
<i>Vibrio cholerae</i>	0	0	0	1	1
<i>Vibrio fischeri</i>	0	0	0	0	0
<i>Vibrio vulnificus</i>	0	0	0	0	0
<i>Yersinia pestis</i>	0	0	0	1	0
TOTAL	1	0	0	9	3

Clases bacterianas	Modelos				
	A	B1	B2	C1	C2
Alphaproteobacteria	0*	0	0	1	0
Bacilli	0*	0	0	1	1
Betaproteobacteria	0	0	0	0	0
Epsilonproteobacteria	0	0	0	0	0
Gammaproteobacteria	0*	0	0	1	1

*encuentra motivo de 21 desplazado con la secuencia del modelo C.

Anexo 3. Mapas de calor.

Se adjunta un archivo que contiene los mapas de calor en formato html obtenidos para todos los análisis realizados de los distintos clados bacterianos con CGB.

Anexo 4. Script de Python.

Se adjunta el script de Python empleado para la obtención de las secuencias ampliadas para la búsqueda del motivo de unión al ADN de Fur.