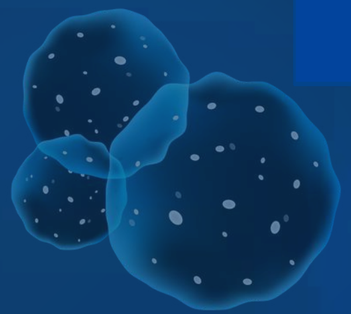
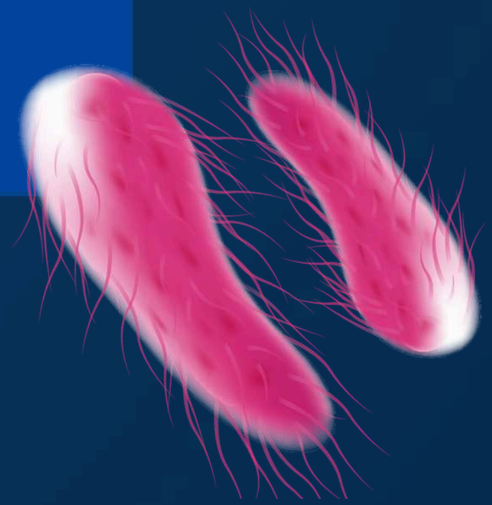
Diferencias y similitudes entre cepas de E. coli

Martínez Oviedo Guillermo
Sánchez Cruz Norma Selene
Sosa Romo Juan Mario

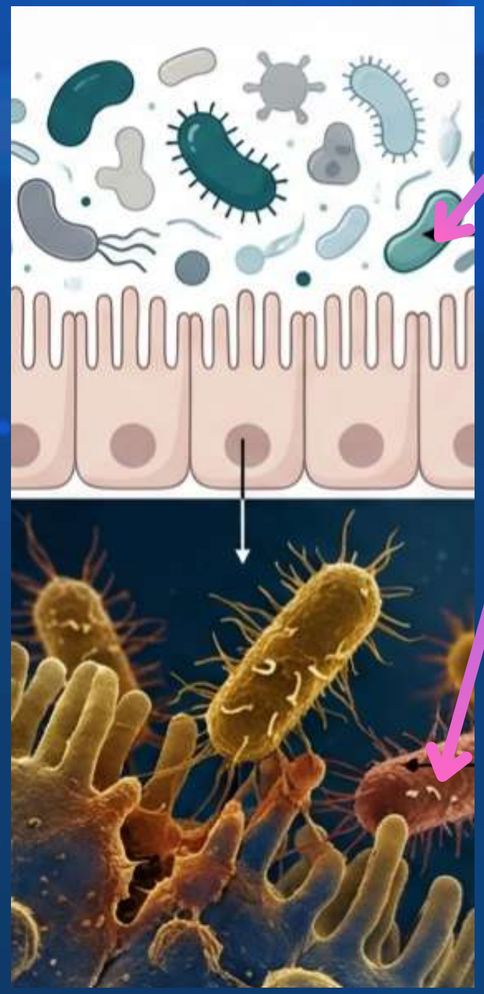




01

Introducción

Contexto biológico



Rol Comensal: Es el anaerobio facultativo más abundante del intestino humano, viviendo en equilibrio y beneficio mutuo.

Rol Patogénico (DEC): Existen linajes (E. coli Diarreagénico) que causan alta morbilidad y mortalidad infantil en países en desarrollo.

Mecanismo de Cambio: La transformación de comensal a patógeno ocurre por la adquisición de factores de virulencia (toxinas, adhesinas) mediante transferencia horizontal de genes.

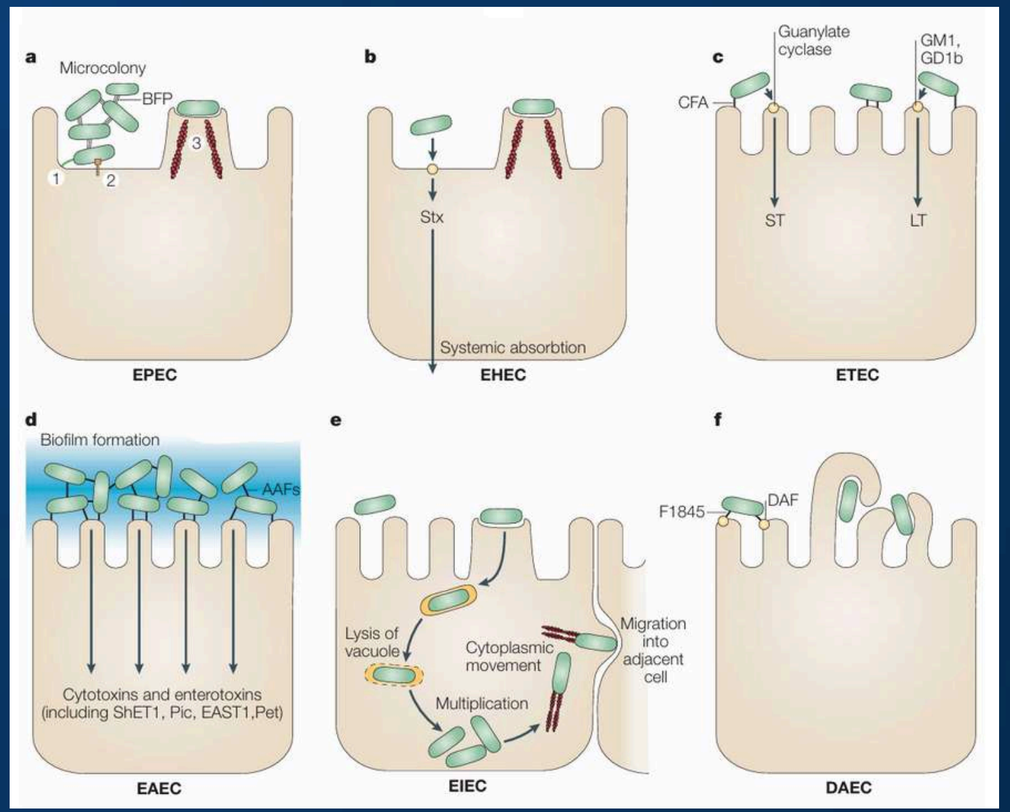
Clasificación por Patotipos y Mecanismos de Virulencia

Los patotipos son grupos clasificados según su combinación específica de factores de virulencia y cuadros clínicos.

Principales Patotipos Diarreagénicos (DEC):

- **EPEC (Enteropatógeno)**
- **EHEC (Enterohemorrágico)**
- **Otros:** ETEC, EAEC, EIEC.

Relevancia: Cada patotipo tiene una "huella" molecular distinta que buscaremos detectar computacionalmente.



El Desafío Genómico: Un Pangenoma Abierto

E. coli tiene un genoma altamente dinámico.

Genoma Núcleo (Core): ~2,200 genes compartidos por todas las cepas (funciones básicas).

Genoma Accesorio: >13,000 familias de genes en la población. Aquí es donde residen la mayoría de los factores de virulencia y adaptación.

La patogenicidad emerge de combinaciones específicas dentro de este vasto genoma accesorio.

Minería de datos para encontrar patrones en esta fracción accesoria.

¿Es posible diferenciar computacionalmente las cepas patógenas de *E. coli* de las comensales basándose únicamente en la presencia o ausencia de marcadores genéticos de virulencia específicos?

Objetivos

Objetivo general

Analizar comparativamente el contenido genómico de cepas patógenas y comensales de *Escherichia coli* mediante herramientas bioinformáticas, con el fin de identificar patrones de presencia y ausencia de genes que permitan discriminar los principales patotipos diarreagénicos.

Objetivo Específicos

Establecer un perfil de referencia de los genes de virulencia canónicos (adhesinas, toxinas y sistemas de secreción) consultando bases de datos especializadas.

Determinar *in silico* **la prevalencia y distribución** de dichos marcadores en un conjunto de genomas representativos.

Evaluar el agrupamiento de las cepas para verificar si la presencia de estos genes permite reconstruir la clasificación patogénica conocida.

02

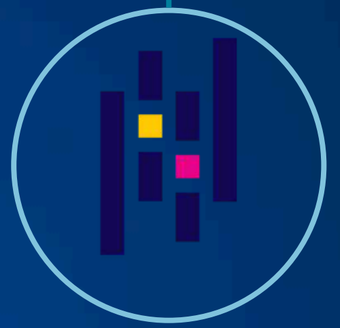
Material y método

El flujo de trabajo: Un enfoque de minería de datos genómicos.

(v1.0.1) para búsquedas locales tipo BLAST de alto rendimiento

Visualización: Mapas de calor (heatmaps) con la librerías Seaborn.

Fuente: Base de datos NCBI RefSeq



Base de datos: Virulence Factor Database

Procesamiento: Scripts personalizados en Python con la librería Pandas.

Adquisición de genomas

Fuente: Base de datos NCBI RefSeq

Conjunto de datos:

- Un grupo control de cepas comensales/no patógenas
- Cepas representantes de los patotipos DEC (principalmente EPEC y EHEC)
- Cepa de referencia *E. coli* K-12 MG1655

Análisis y visualización

Procesamiento: Scripts personalizados en Python con la librería Pandas.

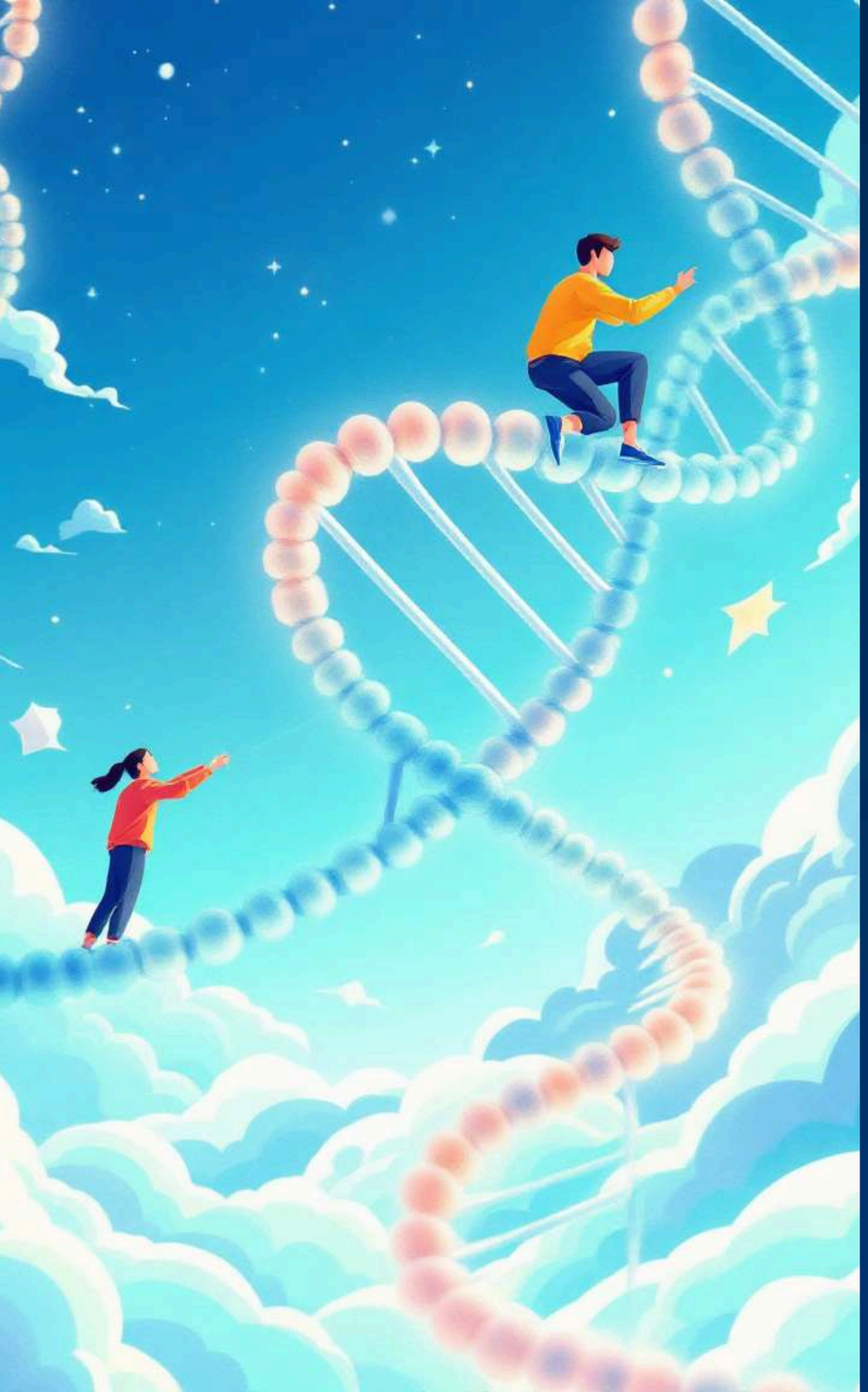
Visualización: Mapas de calor (heatmaps) con la librerías Seaborn.

Detección de factores de virulencia

Herramienta: ABRicate (v1.0.1) para búsquedas locales tipo BLAST de alto rendimiento

Base de datos: Virulence Factor Database (VFDB)

Parámetros estrictos: Identidad mínima del 90% y cobertura mínima 80% para reducir falso positivos.



Traduciendo genomas a datos: La matriz binaria de presencia/ausencia

Proceso:

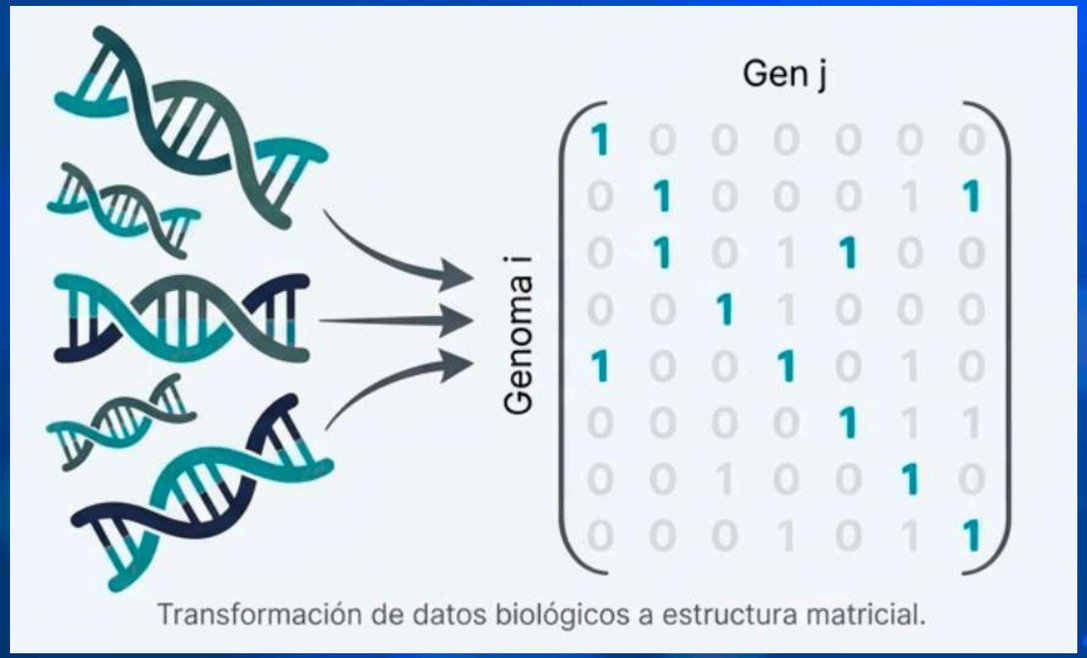
- Los resultados del cribado con ABRicate se transformaron en una matriz binaria.
- Las filas representan cada genoma de *E. coli* analizado.
- Las columnas representan cada gen de virulencia único identificado en el estudio.
- El valor en la matriz M_{ij} indica si el gen está presente 1 o ausente 0

Fórmula:

$$M_{ij} = \begin{cases} 1 & \text{si el gen 'j' esta presente en el genoma 'i'} \\ 0 & \text{si el gen 'j' esta ausente} \end{cases}$$

Importancia:

- Es la estructura de datos central que permite el análisis comparativo y el agrupamiento jerárquico.
- El análisis de agrupamiento se realizó sobre esta matriz utilizando la **distancia de Jaccard**, una métrica ideal para datos binarios aimétricos.



Transformación de datos biológicos a estructura matricial.

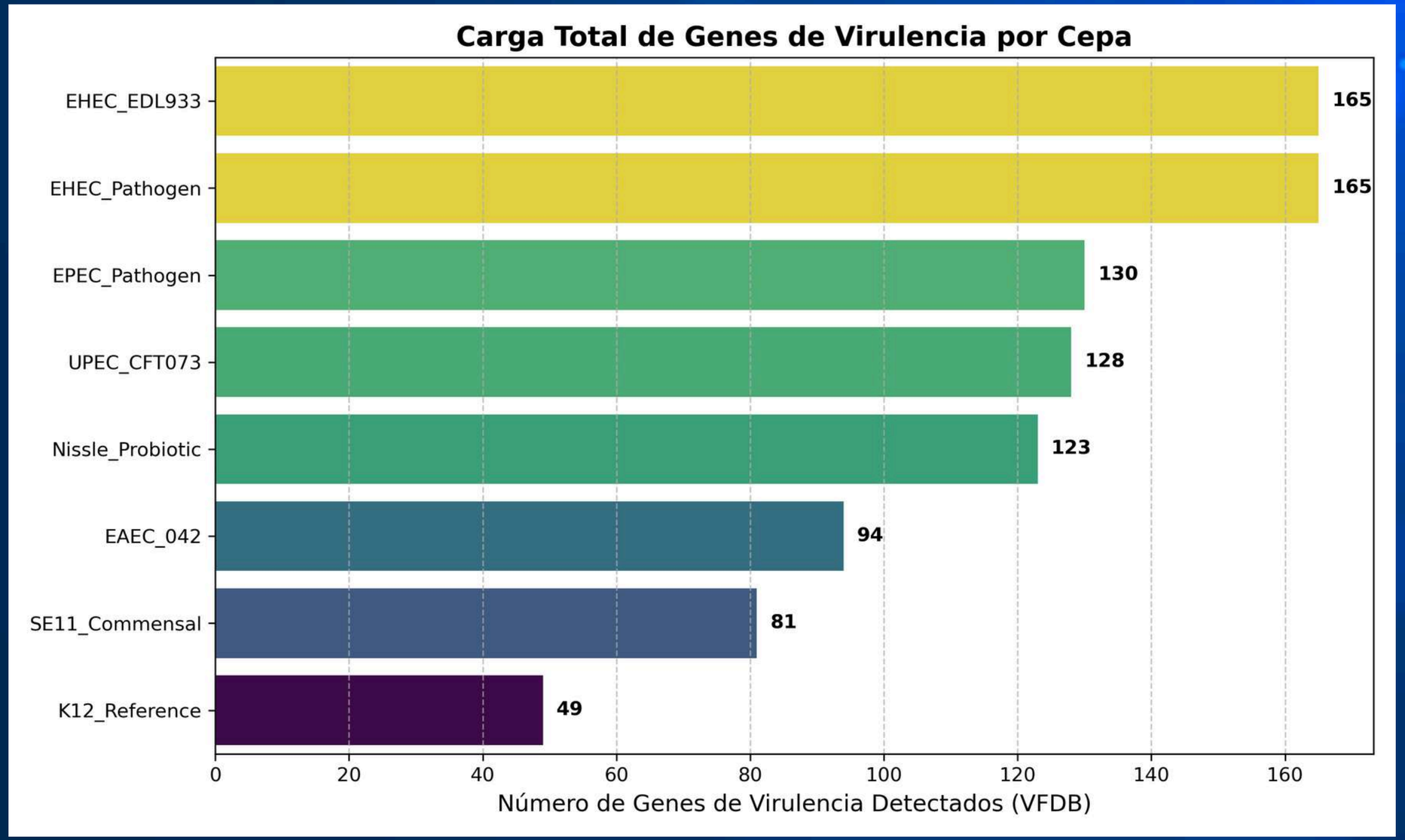
Fórmula

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

J = distancia de Jaccard
A = conjunto 1
B = conjunto 2

03

Resultados y discusión


Carga Total de Genes de Virulencia por Cepa
EHEC_EDL933
EHEC_Pathogen
EPEC_Pathogen
UPEC_CFT073
Nissle_Probiotic
EAEC_042
SE11_Commensal
K12_Reference
165
165
130
128
123
94
81
49
0
20
40
60
80
100
120
140
160
Número de Genes de Virulencia Detectados (VFDB)

Observación principal

El cribado genómico reveló una disparidad significactiva en la carga de factores de virulencia entre los grupos.

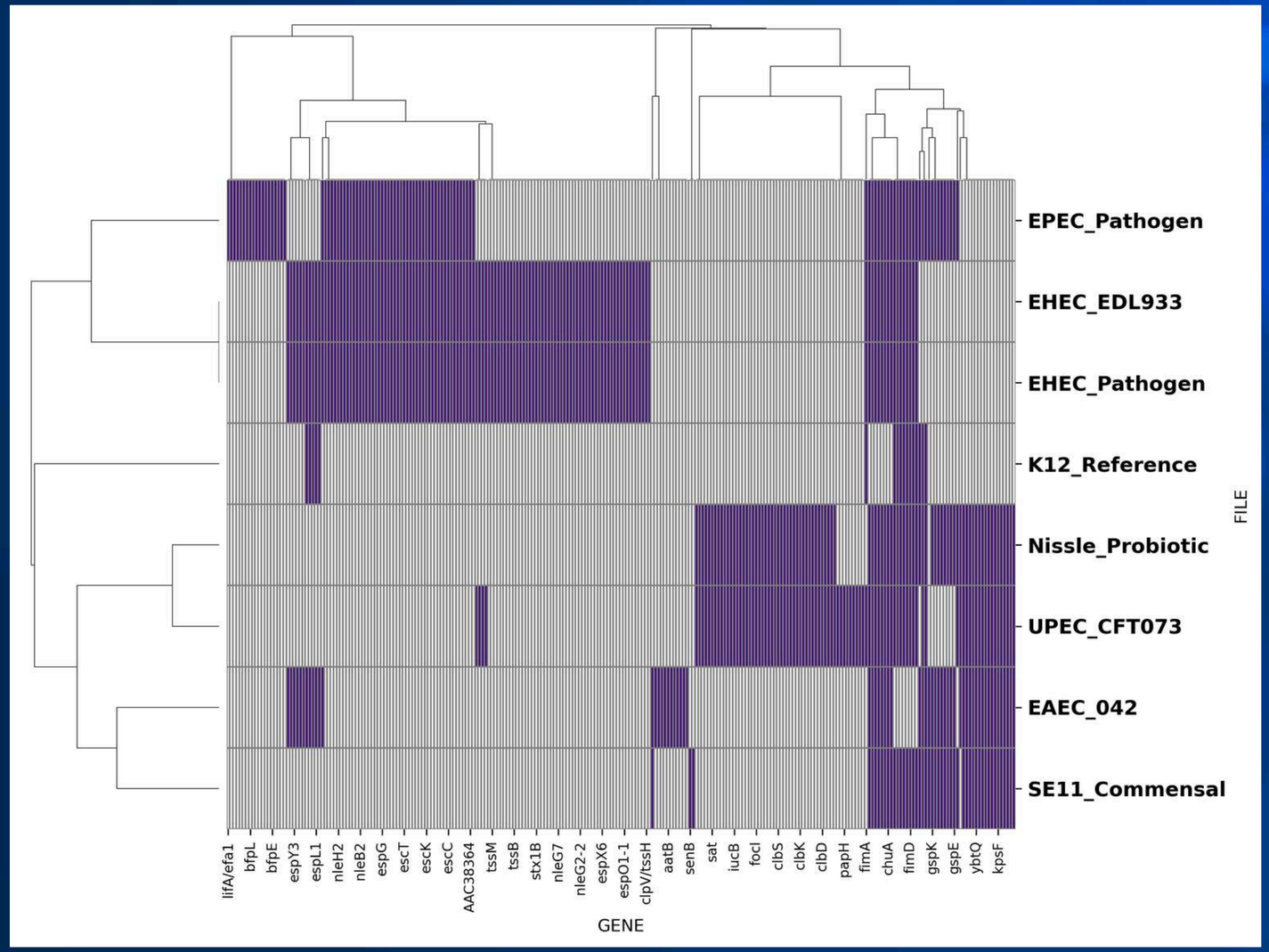
Datos clave

- **Grupo EHEC** (Sakai, EDL933): **165 genes**
- **Cepa EPEC: 130 genes**
- C**epa Uropatógena (UPEC CFT073): 128 genes**
- **Cepa de referencia K-12**: solo **49 genes** basales.

Hallazgo sorprendente

La cepa probiótica ***E. coli*** **Nissle 1917** exhibió **123 genes**, un número comparable al de los patógenos, sugiriendo que su capacidad de colonización depende de mecanismos agresivos.

Descifrando los clústeres: Firmas genómicas que definen a cada grupo.



Los marcadores genéticos que actúan como discriminantes computacionales.

Cepa / Patotipo	Isla LEE (eae)	Toxina Shiga (stx)	Pili BFP	Fimbria P (pap)	Hemolisina (hly)
EHEC (Sakai/EDL933)	+	+	-	-	+
EPEC (E2348/69)	+	-	+	-	-
UPEC (CFT073)	-	-	-	+	+
EAEC (042)	-	-	-	-	-
Nissle 1917	-	-	-	-	-
K-12 Reference	-	-	-	-	-

Matriz resumen de marcadores de virulencia detectados *in silico* que permiten la diferenciación automática de cepas.

Recuperación de la Estructura Biológica de *E. coli*

EHEC y EPEC

Formaron un clúster definido (presencia de isla LEE y secreción Tipo III).

UPEC (CFT073)

Agrupamiento independiente por factores uropatógenos (fimbrias P, hemolisinas).

Comensales (K-12, SE11)

Base del dendrograma (repertorio de virulencia reducido)

Nissle 1917 (Probiótico)

Alta carga de genes de virulencia (potencial colonizador), similar a patógenos.

Limitaciones en la Detección: El Caso de EAEC 042

Resultado

No se agrupó coherentemente (rama aislada/ambigua).

Causa Biológica

Carece de marcadores clásicos (LEE, Stx).

Causa Técnica

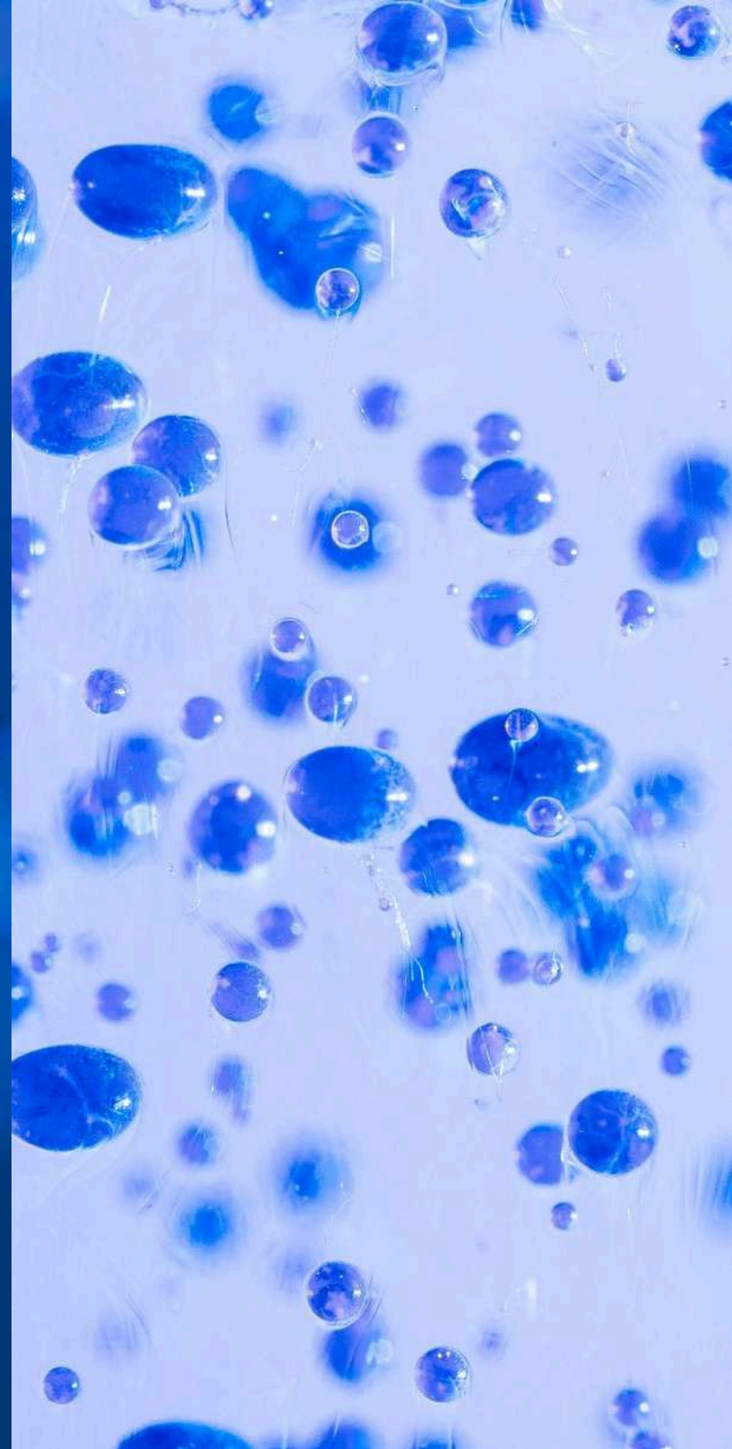
- Factores clave (reguladores plasmídicos, adhesinas) no detectados.
- Umbrales de identidad demasiado estrictos o ausencia en VFDB.

Conclusión

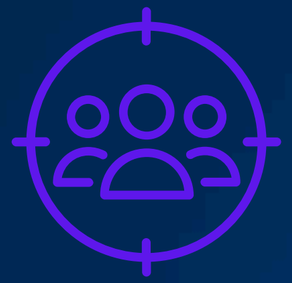
Un subconjunto limitado de genes es insuficiente para patotipos complejos.

Métrica Utilizada: Distancia de Jaccard

- **Idoneidad:** Ideal para matrices binarias dispersas.
- **Enfoque:** Mide la proporción de genes compartidos presentes.
- **El factor (0,0):**
  - Se ignoran las ausencias compartidas.
  - Razón: Computacionalmente, no tener un gen no implica similitud funcional activa.
- **Objetivo:** Resaltar el impacto de los genes de virulencia efectivamente adquiridos.



Limitaciones del Estudio

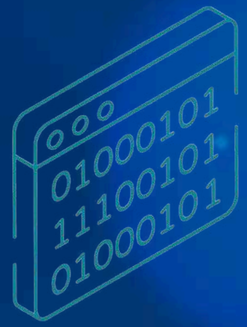


Muestra

Tamaño reducido y poca variación intra-patotipo.

Datos

Dependencia exclusiva de VFDB y parámetros de búsqueda estrictos (riesgo de subdetección).



Codificación Binaria

Pérdida de información sobre:

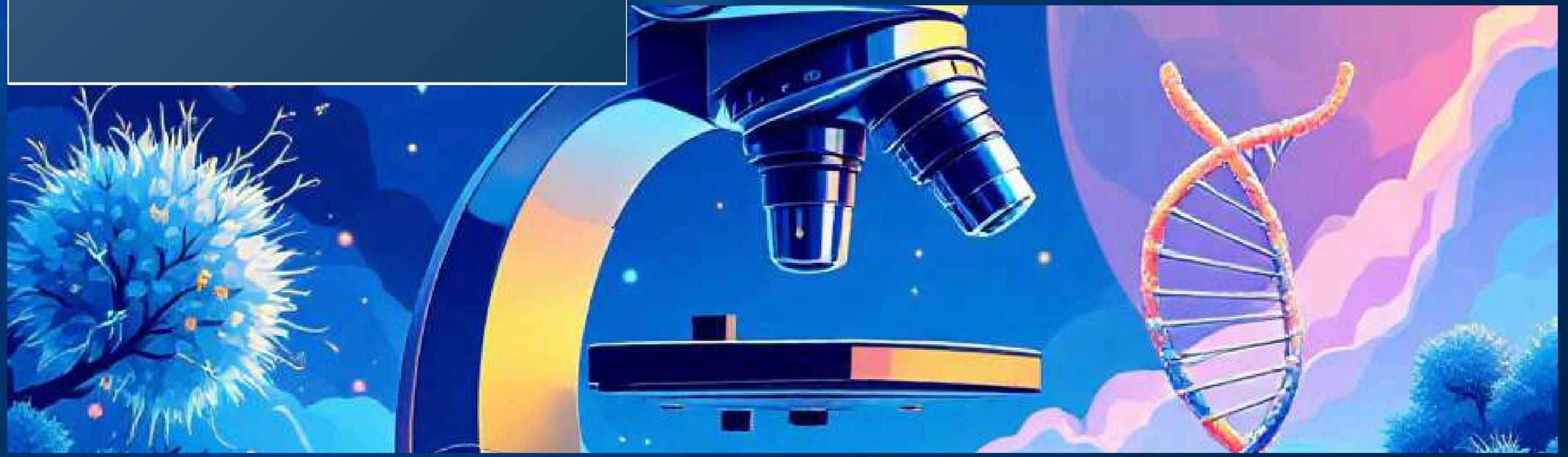
- Variación puntual.
- Número de copias.
- Regulación de expresión.

Algoritmo

Métodos no supervisados insuficientes para señales sutiles (como en EAEC).

04

Conclusión



Conclusión General: Validación de la Metodología

¿Es posible diferenciar computacionalmente las cepas patógenas de *E. coli* de las comensales basándose únicamente en la presencia o ausencia de marcadores genéticos de virulencia específicos?

Sí, la minería computacional es una herramienta eficaz; logra distinguir patotipos principales (EHEC, EPEC, UPEC) de cepas comensales.

Logro del objetivo general

- ☑ El agrupamiento jerárquico sobre matrices binarias fue exitoso.
- ☑ Se reconstruyó con alta fidelidad la clasificación biológica esperada.
- ☑ Separación clara entre cepas entéricas, extraintestinales y comensales.

Cumplimiento de Objetivos Específicos

Perfil *In Silico*

Confrontación exitosa contra la base de datos VFDB.

Diferencias Cuantitativas

- **Patógenos (EHEC): ~165 genes de virulencia.**
- **Comensales: ~49 genes de virulencia.**
- **Marcadores clave identificados: LEE (EPEC/EHEC) y toxina Shiga (stx).**

Evaluación del Agrupamiento

Formación de clústeres definidos:

1. **Grupo LEE-positivo: EHEC y EPEC.**
2. **Grupo Uropatógeno: UPEC.**
3. **Grupo Basal: Cepas comensales (K-12, SE11).**

Hallazgo Principal y Relevancia

Firma Genómica

El perfil de presencia/ausencia actúa como una firma predictiva robusta para la mayoría de los linajes.

Limitación Crítica (El caso EAEC)

- Falla en la clasificación de EAEC 042.
- Lección: Patotipos complejos requieren marcadores específicos y marcos analíticos más amplios.

Conclusión Final:

El análisis de virulencia es una estrategia fundamental para descifrar la patogenicidad en el pangenoma dinámico de *E. coli.*

Bibliografía

Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., & Jin, Q. (2005). VFDB: a reference database for bacterial virulence factors. Nucleic Acids Research, 33 (suppl_1), D325-D328. https://doi.org/10.1093/nar/gki008

Gomes, T. A., Elias, W. P., Scaletsky, I. C., Guth, B. E., Rodrigues, J. F., Piazza, R. M., Ferreira, L. C., & Martinez, M. B. (2016). Diarrheagenic Escherichia coli. Brazilian Journal of Microbiology, 47, 3-30. https://doi.org/10.1016/j.bjm.2016.10.015

Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B. L., Mori, H., et al. (2006). Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. Molecular Systems Biology, 2 (1), 2006-0007. https://doi.org/10.1038/msb4100049

Kaper, J. B., Nataro, J. P., & Mobley, H. L. (2004). Pathogenic Escherichia coli. Nature Reviews Microbiology, 2 (2), 123-140. https://doi.org/10.1038/nrmicro818

Rasko, D. A., Rosovitz, M. J., Myers, G. S., Mongodin, E. F., Fricke, W. F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N. R., Chaudhuri, R., Henderson, I. R., Sperandio, V., & Ravel, J. (2008). The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. Journal of Bacteriology, 190 (20), 6881-6893. https://doi.org/10.1128/JB.00619-08

Gracias