

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 724 874**

21 Número de solicitud: 201830112

51 Int. Cl.:

C12Q 1/68

(2008.01)

12

PATENTE DE INVENCION CON EXAMEN

B2

22 Fecha de presentación:

08.02.2018

43 Fecha de publicación de la solicitud:

17.09.2019

Fecha de concesión:

17.11.2020

45 Fecha de publicación de la concesión:

24.11.2020

73 Titular/es:

**UNIVERSITAT DE GIRONA (100.0%)
PLAZA SANT DOMENEC, NUM. 3
17004 GIRONA (Girona) ES**

72 Inventor/es:

**MARTÍNEZ MEDINA, Margarita;
LÓPEZ SILES, Mireia y
CAMPRUBÍ FONT, Carla**

74 Agente/Representante:

FÚSTER OLAGUIBEL, Gustavo Nicolás

Observaciones:

La lista de secuencias está accesible al público en la página web de la OEPM para su descarga en formato electrónico.

54 Título: **Nuevos marcadores moleculares para el patotipo de Escherichia coli adherente invasiva (AIEC), métodos relacionados y kits**

57 Resumen:

Nuevos marcadores moleculares para el patotipo de Escherichia coli adherente-invasiva (AIEC), métodos relacionados y kits.

La presente invención se refiere a nuevos polimorfismos de un solo nucleótido (SNP) útiles como marcadores moleculares para diferenciar entre cepas de Escherichia coli adherente-invasiva (AIEC) y cepas no-AIEC. Se refiere además a métodos para usar los mismos, tales como métodos para detectar o predecir el fenotipo de AIEC. Además, se refiere a kits que comprenden oligonucleótidos para detectar la presencia de dichos marcadores moleculares y a su uso en los métodos de la invención.

ES 2 724 874 B2

Aviso: Se puede realizar consulta prevista por el art. 41 LP 24/2015. Dentro de los seis meses siguientes a la publicación de la concesión en el Boletín Oficial de la Propiedad Industrial cualquier persona podrá oponerse a la concesión. La oposición deberá dirigirse a la OEPM en escrito motivado y previo pago de la tasa correspondiente (art. 43 LP 24/2015).

DESCRIPCIÓN

NUEVOS MARCADORES MOLECULARES PARA EL PATOTIPO DE *ESCHERICHIA COLI* ADHERENTE INVASIVA (AIEC), MÉTODOS RELACIONADOS Y KITS

5 CAMPO DE LA INVENCIÓN

La presente invención se refiere al campo de la medicina y la microbiología. Específicamente, se refiere a nuevos polimorfismos de un solo nucleótido (SNP) útiles como marcadores moleculares para diferenciar entre cepas de *Escherichia coli* adherente-invasiva (AIEC) y cepas distintas de AIEC (no-AIEC). Se refiere además a métodos para usar los mismos, tales como métodos para detectar o predecir el fenotipo AIEC. Además, se refiere a kits que comprenden oligonucleótidos para detectar la presencia de dichos marcadores moleculares y a su uso en los métodos de la invención.

ANTECEDENTES DE LA INVENCIÓN

- 15 La enfermedad de Crohn (CD) es un tipo de enfermedad inflamatoria intestinal (IBD) crónica idiopática en la que se ha demostrado que están implicados la genética, los factores medioambientales y el microbioma intestinal del huésped (Sartor, R.B. *et al*, 2006). Ha sido descrito que en el intestino de los pacientes con CD se producen alteraciones del microbioma, conocidas como disbiosis, en las que hay una disminución de los microbios
- 20 beneficiosos y un posterior aumento de las especies proinflamatorias (Martinez-Medina, M. *et al*, 2009). Se ha descrito repetidamente que *Escherichia coli* está aumentada en la CD (Martinez-Medina, M. *et al*, 2009; Baumgart, M. *et al*, 2007; Darfeuille-Michaud, A. *et al*, 2004) y su abundancia se correlaciona con la actividad de la enfermedad (Lopez-Siles, M. *et al*, 2014). Boudeau *et al*. (Boudeau *et al*, 1999) definieron un nuevo patotipo de esta especie
- 25 bacteriana denominado *Escherichia coli* adherente-invasiva (AIEC), y este patotipo se ha asociado con la inflamación del intestino en pacientes con CD mediante varios grupos independientes (Martinez-Medina, M. *et al*, 2009; Darfeuille-Michaud, A. *et al*, 2004; Baumgart, M. *et al*, 2007; Martin, H. M. *et al*, 2004; Sasaki, M. *et al*, 2007; Eaves-Piles, T. *et al*, 2008; Dogan, B. *et al*, 2013). El patotipo AIEC posee la capacidad de adherirse e invadir
- 30 las células epiteliales intestinales y de sobrevivir y replicarse dentro de macrófagos (Boudeau *et al*, 1999).

En la actualidad, la identificación del patotipo AIEC se realiza mediante el examen fenotípico de bacterias en cultivo, que es una técnica que requiere un tiempo extremadamente

35 prolongado. En resumen, las cepas de AIEC se identifican evaluando experimentalmente: (i)

la capacidad para adherirse a células epiteliales intestinales; (ii) la capacidad para invadir células epiteliales intestinales, y (iii) la capacidad para sobrevivir y replicarse dentro de macrófagos, tal como se describe por ejemplo en Martínez-Medina, M. *et al*, 2009 o Darfeuille-Michaud, 2004. Por tanto, la existencia de una herramienta molecular para la

5 detección específica de las AIEC sería de gran importancia para facilitar, por ejemplo, estudios de distribución de AIEC que definieran cuales són los reservorios del patotipo y sus vías de transmisión. Además, se han descrito nuevas terapias personalizadas dirigidas contra la microbiota para tratar la CD, en particular, para el tratamiento de pacientes con CD portadoras del fenotipo de AIEC (Sivignon *et al*. 2015; Sivignon *et al*. 2017). La existencia

10 de un marcador molecular para identificar el fenotipo de AIEC será útil en la selección de aquellos sujetos que presentan cepas de AIEC en su tracto intestinal (es decir, portadores de AIEC) y por tanto que se benefician de terapias dirigidas contra AIEC.

Aunque se han dilucidado muchos de los mecanismos moleculares de la patogenicidad de

15 AIEC (Cieza, R. J. *et al*, 2015; Dogan, B. *et al*, 2014; Gibold, L. *et al*, 2015; Martínez-Medina, M. y García-Gil, L., 2014; Vazeille, E. *et al*, 2016), todavía no hay ningún gen ni secuencia distintiva que permita la identificación molecular de cepas AIEC. Se han encontrado algunos genes más frecuentemente en AIEC pero éstos también están presentes en cepas no-AIEC; por ejemplo, el gen *lpfA* se ha detectado en el 68,4-71% de

20 las cepas de AIEC y en el 20-28,1% de las cepas no-AIEC (Vazeille, E. *et al*, 2016; Chassaing, B. *et al*, 2011). Asimismo, la adhesina *fimH* es ubicua en *E. coli*, y aunque se ha encontrado que algunas de sus variantes de secuencia están asociadas con AIEC, dichas variantes también se han detectado en cepas no-AIEC (Iebba, V. *et al*, 2013; Dreux, N. *et al*, 2013; Desilets, M. *et al*, 2015). Además, se han descrito mutaciones puntuales en genes

25 implicados en la patogenicidad de AIEC, pero estas se han identificado exclusivamente en la cepa LF82 mediante comparación con la cepa comensal K-12. Por ejemplo, ha sido descrito que variantes génicas específicas de *ompA* (Rolhion, N. *et al*, 2010) y *chiA* (Low, D. *et al*, 2013) confieren a la cepa LF82 mayores índices de invasión y adhesión respectivamente , pero es necesaria la confirmación de este hallazgo en un conjunto más

30 amplio de cepas. Finalmente, las diferencias entre AIEC y *E. coli* no patogénicas pueden incluir la expresión de genes específicos, tal como se ha propuesto para las fimbrias de tipo 1 (Dreux, N. *et al*, 2013).

Se han realizado varios estudios genómicos con el objetivo de caracterizar el genoma de

35 cepas AIEC y para identificar un biomarcador genético para las mismas. Hasta la fecha se

han secuenciado treinta y cinco cepas de AIEC, la mayoría de las cuales pertenecen al grupo filogenético B2 (Miquel, S. *et al*, 2010; Nash, J. H. *et al*, 2010; Conte, M. P. *et al*, 2014; Dogan, B. *et al*, 2014; Desilets, M. *et al*, 2015; O'Brien, C. L. *et al*, 2016). Sin embargo, todavía no se ha identificado ningún gen ni secuencia exclusivos del patotipo AIEC. Estudios genómicos han confirmado que las AIEC se asemeja a las ExPEC (Dogan, B. *et al*, 2014; Desilets, M. *et al*, 2015; Deshpande, N. P. *et al*, 2015). Estos estudios han sido útiles en la detección de genes de virulencia que se hallan más frecuentemente en cepas de AIEC aisladas de individuos con CD en relación con cepas de AIEC aisladas de pacientes sin IBD. En particular, los genes *fyuA* e *ibeA*, que están implicados en la producción de sideróforos y en la invasividad, respectivamente, se encontraron en más del 70% de cepas de AIEC procedentes de CD y estaban ausentes de cepas de AIEC aisladas de individuos sin IBD (Conte, M. P. *et al*, 2014). Además, estudios genómicos también han demostrado que genes de virulencia tales como *pduC*, que están implicados en la invasividad y la persistencia, son más frecuentes en AIEC que en cepas no- AIEC (el 50% y el 20%, respectivamente) (Dogan, B. *et al*, 2014).

Las dificultades en el descubrimiento de rasgos específicos de AIEC se han debido probablemente al hecho de que los primeros estudios compararon cepas de AIEC y no-AIEC filogenéticamente distantes, y por lo tanto, las diferencias encontradas entre ellas están relacionadas con el origen filogenético de las cepas en lugar de con el fenotipo AIEC (Miquel, S. *et al*, 2010; Nash, J. H. *et al*, 2010). Recientemente, Desilets, *et al*. (Desilets, M. *et al*, 2015) encontraron, mediante la comparación de cepas de AIEC y no-AIEC del mismo grupo filogenético, tres regiones genómicas presentes en todas las cepas de AIEC del grupo filogenético B2 y ausentes de las cepas de AIEC de otros grupos filogenéticos y de cepas comensales de cualquier origen filogenético (incluyendo B2). Sin embargo, se desconoce si estas regiones son específicas para las cepas de AIEC de B2 o también están presentes en otros grupos patogénicos que comparten el mismo origen filogenético, tales como las cepas de ExPEC de B2. Deshpande *et al*. (Deshpande, N. P. *et al*, 2015) describieron 29 polimorfismos diagnósticos de un solo nucleótido (SNP) que producen cambios de aminoácido sinónimos o no sinónimos en una secuencia distintiva (*signature sequence*) que diferencia un grupo de cepas patógenas de B2 (que consiste en 4 cepas de AIEC, 3 de ExPEC, 47 cepas de *E. coli* uropatógenas y 1 cepa de *E. coli* patógena aviar) de otras cepas de *E. coli* presentes en la base de datos del NCBI. Sin embargo, no se encontró ninguna característica específica que discrimine el patotipo AIEC. Finalmente, O'Brien *et al*. (O'Brien, C. L. *et al*, 2016) redujeron la variabilidad del contenido génico realizando el

análisis genómico de un conjunto de cepas de *E. coli* del grupo filogenético B2 con una secuencia tipo idéntica (ST95), disminuyendo de ese modo la posibilidad de detección de elementos genéticos diferenciales delimitados por el trasfondo filogenético. No obstante, la evaluación de la prevalencia génica y de la composición de bases en genes fundamentales no dio como resultado la identificación de un biomarcador específico de AIEC ni siquiera la identificación de un marcador común para la mayoría de las cepas de AIEC examinadas en este estudio.

Además, ha sido descrito que genes que se encuentran más frecuentemente en AIEC también están presentes en otros patotipos de *E. coli* e incluso en *E. coli* no patogénicas (Dogan *et al.*, 2014; Conte *et al.*, 2014; Céspedes *et al.*, 2017; Desilets *et al.*, 2015, Desphande *et al.*, 2015; Nash *et al.*, 2010). Sin embargo, se han sugerido diferencias en la expresión de genes particulares según el patotipo (Dreux, N. *et al.*, 2013; Rolhion, N. *et al.*, 2007).

Por consiguiente, todavía existe la necesidad de desarrollar nuevos métodos moleculares para diferenciar entre cepas con el fenotipo de AIEC y no- AIEC, con sensibilidad y/o especificidad mejoradas. Además, existe la necesidad de desarrollar un método para detectar o predecir el fenotipo de AIEC que sea de aplicación general, independientemente del grupo filogenético de la cepa de *E. coli*.

SUMARIO DE LA INVENCION

Los inventores propusieron que en lugar de las diferencias en el contenido génico, mutaciones puntuales en genes fundamentales (*core genes*) y/o diferencias en la expresión génica puedan ser responsables del fenotipo de AIEC.

Por consiguiente, los inventores estudiaron las diferencias en el contenido génico y los SNP de cepas AIEC y no-AIEC aisladas del intestino humano. Con el objetivo de encontrar un biomarcador específico de AIEC, secuenciaron los genomas de seis cepas de *E. coli* que consistían en tres pares de cepas AIEC/no-AIEC que pertenecían a diferentes grupos filogenéticos (B1, B2 y D). Cada par de cepas se aisló del mismo paciente y tenía un grupo filogenético, tipo de secuencia, perfil génico de virulencia y huella electroforética en gel de campo pulsante (PFGE) idénticos (Tabla 1 y Tabla 2). La hipótesis de los inventores fue que comparando pares de cepas próximas genéticamente, aumentaría la posibilidad de encontrar elementos genéticos específicos característicos del fenotipo de AIEC.

En particular, no se encontraron diferencias significativas en la estructura genómica o incluso en el contenido génico entre cepas de AIEC y no-AIEC, lo que confirma su identidad próxima mediante PFGE (Figura 1). Además, no se encontró ningún gen que estuviera presente en al menos dos cepas de AIEC y ausente de todas las cepas no-AIEC (Figura 3a). Además, el análisis no reveló asociación entre la presencia de genes asociados con AIEC previamente (*lpfA₁₅₄*, *gipA*, *pduC*, *fyuA*, *afaC*, *chuA* y *ibeA*) (Conte, M. P. *et al*, 2014; Cieza, R. J. *et al*, 2015; Dogan, B. *et al*, 2014; Vazeille, E. *et al*, 2016; Chassaing, B. *et al*, 2011; Prorok-Hamon, M. *et al*, 2014; Céspedes, S. *et al*, 2017) y el fenotipo de AIEC (Tabla 1). Este resultado está en línea con observaciones previas (Dogan, B. *et al*, 2014; O'Brien, C. L. *et al*, 2016). Se encontró una situación similar con respecto a variantes genéticas de *fimH*, *ompA* y *chiA*. Aunque se han asociado cambios específicos en las secuencias de aminoácidos de proteínas codificadas por estos genes con mayor capacidad de adhesión/invasión (Dreux, N. *et al*, 2013; Rolhion, N. *et al*, 2010; Low, D. *et al*, 2013), los inventores no observaron tales diferencias entre los aislados de los pares de cepas estudiados. Por tanto, las diferencias en las secuencias de estos genes no determinarían el patotipo de las cepas.

En un intento adicional por explicar las diferencias fenotípicas observadas entre pares, los inventores se propusieron identificar SNP que estuvieran presentes de manera diferencial en cepas de AIEC y en sus homólogas no-AIEC. A partir del número total de SNP identificados en los pares de cepas de AIEC/no-AIEC secuenciadas, se seleccionó un subgrupo más prometedor (es decir, que se ajustaba a los criterios indicados en la Tabla 5). Los SNP seleccionados se validaron adicionalmente mediante secuenciación de Sanger lo que dio como resultado la identificación de un subgrupo de veinte SNP denominados "SNP confirmados" (Tablas 5 y 6).

Además, los inventores exploraron la utilidad de las diferencias genéticas encontradas como marcadores moleculares para la identificación de AIEC. En particular, estudiaron la variabilidad de nucleótidos dentro de los SNP identificados en un conjunto de cepas de AIEC y no-AIEC para validar o refutar la hipótesis que confirmaba que los SNP representan posibles firmas moleculares para la identificación específica del patotipo de AIEC. En total, se estudiaron dieciséis SNP en 22 cepas de AIEC y 28 cepas no-AIEC aisladas de sujetos sanos y pacientes con CD y que pertenecían a varios grupos filogenéticos, concretamente A, B1, B2 y D (Tabla 7). Más específicamente, se incluyeron doce SNP confirmados que no

eran específicos de cepa (característica que se identificó mediante análisis *in silico*) , concretamente los comprendidos en cualquiera de los genes E1-E2_3.6, E3-E4_4.3, E3-E4_4.4, E5-E6_3.12 y E5-E6_3.16=3.22. Además se incluyeron cuatro de los SNP confirmados considerados específicos de cepa mediante el análisis *in silico* (E1-E2_3.4, E1-E2_5, E1-E2_3.7 y E5-E6_3.1), que también fueron analizados para confirmar que eran específicos de cepa.

La distribución de los polimorfismos en el conjunto de cepas analizadas indicó que no había un nucleótido particular para ninguna de las posiciones de SNP que estuviera presente en todas las cepas de AIEC y que difiriera del nucleótido encontrado en esa misma posición en cepas no-AIEC. Sin embargo, los inventores encontraron posiciones de SNP que presentaban diferencias, no sólo en la distribución de variantes de nucleótido según el patotipo (E3-E4_4.3(2), E3-E4_4.4 y E5-E6_3.16=3.22(2)), sino también en su asociación con los índices de adhesión e invasión (E3-E4_4.3(2), E3-E4_4.4, E5-E6_3.16=3.22(2), E5-E6_3.16=3.22(3)), véanse la Tabla 8 y la Figura 4.

Resulta interesante el hecho de que se demostró que dos de los SNP identificados (E3-E4_4.4 y E5-E6_3.16=3.22(2)) eran adecuados para la predicción del fenotipo de AIEC tal como se determinó mediante el modelo de regresión logística binaria (véase la Tabla 9). En particular, el SNP en E3-E4_4.4 ha demostrado clasificar las cepas como AIEC o no-AIEC con un éxito global del 73% y en el caso del SNP E5-E6_3.16=3.22(2), el éxito global fue similar (68,9%). Además, los inventores han ideado un algoritmo de identificación de cepas de AIEC que combina estos dos SNP (es decir, E3-E4_4.4, E5-E6_3.16=3.22(2)) y otro de los SNP no específicos de cepa confirmados (es decir, E5-E6_3.12), que diferencia el patotipo AIEC con una alta tasa de exactitud. Tal como se muestra en la Figura 5, este algoritmo puede predecir el fenotipo de AIEC con una exactitud del 84%.

Basándose en estos hallazgos, los inventores proponen nuevos métodos moleculares para determinar o predecir si una cepa de *E. coli* tiene un fenotipo caracterizado por altos índices de adhesión e invasión, y más específicamente si presenta un fenotipo de AIEC.

Por consiguiente, en un primer aspecto, la invención se refiere a un método para determinar o predecir si una cepa de *E. coli* tiene un fenotipo caracterizado por altas propiedades de adhesión e invasión, en el que dicho método comprende determinar el genotipo de uno, dos, tres, cuatro o más polimorfismos de un solo nucleótido (SNP) seleccionados del grupo

que consiste en: E1-E2_3.6, E3-E4_4.3(1), E3-E4_4.3(2), E3-E4_4.3(3), E3-E4_4.4, E5-E6_3.12 y de E5-E6_3.16=3.22(1) a E5-E6_3.16=3.22(6).

5 En una realización particular, la presente invención proporciona un método para determinar o predecir si una cepa de *E. coli* tiene un fenotipo de AIEC, en el que dicho método comprende:

10 a) determinar la presencia o ausencia del gen E3-E4_4.4 en el genoma de la cepa de *E. coli*, y cuando está presente, determinar el genotipo en el sitio polimórfico E3-E4_4.4 correspondiente a la posición 1433063 pares de bases (pb) de la cepa AIEC de referencia UM146 (NCBI NC_017632.1).

15 En otra realización particular, la invención se refiere a un método para determinar o predecir si una cepa de *E. coli* tiene un fenotipo de AIEC, en el que dicho método comprende:

20 i. determinar la presencia o ausencia del gen E5-E6_3.16=3.22 en el genoma de la cepa de *E. coli*, y cuando está presente, determinar el genotipo en el sitio polimórfico E5-E6_3.16=3.22(2) correspondiente a la posición 2212845 pb de la cepa AIEC de referencia UM146 (NCBI NC_017632.1).

25 En un aspecto relacionado, la invención se refiere a un método para diferenciar o identificar de manera diferencial cepas de AIEC de cepas de *E. coli* no-AIEC, normalmente similares desde el punto de vista genético, en el que dicho método comprende las etapas tal como se describe según el primer aspecto.

30 En otro aspecto, la invención proporciona un método para clasificar un sujeto como portador de AIEC según la presencia de *E. coli* del fenotipo AIEC en su tracto intestinal, en el que dicho método comprende:

30 a. aislar cepas de *E. coli* de una muestra intestinal obtenida de un sujeto;
b. opcionalmente, extraer ADN de dichas cepas de *E. coli*;
c. determinar o predecir la presencia de *E. coli* del fenotipo AIEC usando un método según un método tal como se describe en el presente documento.

En un aspecto relacionado, la invención se refiere a un método para el examen de sujetos portadores de AIEC o la monitorización del estado de portador de AIEC en un sujeto, en el que dicho método comprende las etapas a) a c) tal como se describió en el aspecto anterior.

5 En un aspecto adicional, la invención se refiere a un método para tratar a un sujeto que tiene una enfermedad intestinal, en el que dicho método comprende:

- 10** i) seleccionar un sujeto que tiene una enfermedad intestinal en el que dicho sujeto se ha clasificado como portador de AIEC mediante un método de clasificación de la invención tal como se describió anteriormente en el presente documento;
- ii) administrar a dicho sujeto una cantidad terapéuticamente eficaz de un tratamiento destinado a eliminar o disminuir los niveles de bacterias AIEC de la mucosa intestinal tal como se describe en el presente documento.

15 En un aspecto relacionado, la invención se refiere a compuestos que eliminan o disminuyen los niveles de bacterias de AIEC para su uso en un método de tratamiento de una enfermedad intestinal en un sujeto portador de AIEC, en los que dicho sujeto se ha identificado mediante el uso de un método de clasificación de la invención tal como se describe en el presente documento.

20

Un aspecto adicional de la invención se refiere a un método implementado por ordenador tal como se describe en el presente documento.

25 Todavía en un aspecto adicional, la presente invención se refiere a un kit para determinar el genotipo de los SNP diana en el genoma de una cepa de *E. coli* y a su uso en un método tal como se describe en el presente documento.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

30

Figura 1. Dendrograma mediante UPGMA consenso generado a partir de los coeficientes de correlación de Pearson de perfiles de PFGE con *XbaI* de los tres pares de cepas seleccionados para secuenciación de genoma. La barra indica la similitud de porcentaje de los perfiles.

35

Figura 2. Comparación del mapa de genoma completo de cepas AIEC/ no-AIEC analizados usando MAUVE 2.3. Los recuadros del mismo color indican segmentos de ADN homólogos entre pares. Los límites entre los diferentes bloques de color indican cambio en la reorientación de la secuencia..

5

Figura 3. Similitudes genómicas entre las seis cepas y dentro de cada par mediante análisis de agrupaciones ortólogas de genes (OCG). a: Diagrama de Edward Vennque indica el número de OCG. Las áreas sombreadas corresponden a agrupaciones compartidas exclusivamente entre al menos dos cepas de AIEC. b: Porcentaje de OCG entre pares de cepas filogenéticamente distintas y entre cepas de AIEC. Los porcentajes se calculan en relación con el número de OCG variables para cada cepa. Aquellas OCG que no están presentes en las seis cepas se consideran variables. Otras combinaciones incluyen agrupaciones de genes compartidos por 5 cepas o 3 cepas de dos o tres grupos filogenéticos diferentes. No hubo OCG comunes entre cepas de AIEC.

10

15

Figura 4. Capacidades de adhesión (a) e invasión (b) de las cepas según variantes de nucleótido específicas de SNP. Sólo se muestran los SNP asociados con diferencias significativas ($p < 0,05$ usando la prueba de la U de Mann-Whitney) en las capacidades de adhesión o invasión de las variantes. Los subgrupos homogéneos ($p > 0,05$) dentro de cada panel se indican mediante los mismos superíndices. La mediana de los datos se indica mediante la línea horizontal en cada caja, las cajas cubren los cuantiles del 25% y el 75%, y las barras muestran los percentiles del 10% y el 90%. Los resultados atípicos se marcan como puntos.

20

25

Figura 5. Algoritmo de clasificación para la identificación de AIEC. Los porcentajes representan la proporción de cepas que se clasifican correctamente como AIEC o no-AIEC basándose en el resultado para cada combinación de SNP. Se indica el número de cepas totales correspondientes a cada condición. (-): ausencia de amplificación; otro: un nucleótido diferente de guanina (G) o picos solapantes.

30

DESCRIPCIÓN DETALLADA DE LA INVENCION

Definiciones

El término “cepa bacteriana” tal como se usa en el presente documento se refiere a una variante o subtipo genético de una especie bacteriana. En una realización particular, dicha cepa bacteriana es un aislado de cultivo bacteriano.

- 5** Los términos “sujeto” o “individuo” se usan en el presente documento de manera intercambiable para referirse a todos los animales clasificados como mamíferos e incluyen, pero no se limitan a, animales domésticos y de granja, primates y humanos, por ejemplo, seres humanos, primates no humanos, vacas, caballos, cerdos, ovejas, cabras, perros, gatos o roedores. Preferiblemente, el sujeto es un ser humano hombre o mujer de cualquier
- 10** edad o raza.

- El término “polimorfismo” o “variante alélica” significa una variación de secuencia común de un gen. Pueden encontrarse variantes alélicas en los exones, regiones no traducidas del gen, o en las secuencias que controlan la expresión del gen. La secuenciación génica
- 15** completa identifica a menudo numerosas variantes alélicas (en ocasiones cientos) para un gen dado. La importancia de las variantes alélicas a menudo no está clara hasta que se realiza el estudio adicional del genotipo y el fenotipo correspondiente en una población suficientemente grande.

- 20** El término “polimorfismo de un solo nucleótido” o “SNP” se refiere a un tipo de polimorfismo de ADN que implica la variación de un solo par de bases. Lo más frecuente, es que estas variaciones se encuentren en secuencias codificantes de genes, regiones no codificantes de genes o en regiones intergénicas entre genes. Cuando se producen SNP dentro de un gen o en una región reguladora cerca de un gen, pueden afectar a la función del gen.

- 25**
- El término “polinucleótido,” tal como se usa en el presente documento, se refiere a una cadena de ácido nucleico. A lo largo de toda esta solicitud, los ácidos nucleicos se designan del extremo 5'-terminal al extremo 3'-terminal. Los ácidos nucleicos convencionales, por ejemplo, el ADN y el ARN, normalmente se sintetizan “de 5' a 3' (es decir, mediante la
- 30** adición de nucleótidos en el extremo 3'-terminal de un ácido nucleico en crecimiento).

- El término “oligonucleótido,” tal como se usa en el presente documento incluye oligómeros lineales o circulares de monómeros o ligamientos naturales y/o modificados, incluyendo desoxirribonucleósidos, ribonucleósidos, formas alfa-anoméricas y sustituidas de los
- 35** mismos, ácidos nucleicos peptídicos (PNA), ácidos nucleicos bloqueados (LNA),

fosforotioato, metilfosfonato, y similares. Los oligonucleótidos son capaces de unirse específicamente a un polinucleótido diana mediante un patrón regular de interacciones de monómero con monómero, tales como el apareamiento de bases de tipo Watson-Crick, el apareamiento de bases de tipo Hoogsteen o Hoogsteen invertido, o similares.

5

El término “ensayo de detección basado en ácido nucleico” tal como se usa en el presente documento, se refiere a un ensayo para la detección de una secuencia diana (por ejemplo un SNP) dentro de un ácido nucleico diana y utilizando uno o más oligonucleótidos que se hibridan específicamente con la secuencia diana.

10

El término “secuencia diana,” tal como se usa en el presente documento, se refiere a la secuencia de nucleótidos particular de un ácido nucleico diana que va a detectarse. La “secuencia diana” incluye las secuencias de complejación a la que se complejan oligonucleótidos (por ejemplo, oligonucleótido de sonda, oligonucleótidos de cebado y/o

15

oligonucleótidos de promotor) durante un procedimiento de detección (por ejemplo, un ensayo de detección basado en amplificación tal como, por ejemplo, amplificación mediada por transcripción (TMA) o reacción en cadena de la polimerasa (PCR), o un ensayo de detección no basado en amplificación tal como, por ejemplo, un ensayo basado en 5'-endonucleasa). Cuando el ácido nucleico diana originalmente es monocatenario, el término

20

“secuencia diana” también se referirá a la secuencia complementaria a la “secuencia diana” tal como está presente en el ácido nucleico diana. Cuando el ácido nucleico diana originalmente es bicatenario, el término “secuencia diana” puede referirse a ambas cadenas sentido (+) y antisentido (-).

25

El término “específico” tal como se usa en el presente documento en relación con una secuencia de nucleótidos significa que una secuencia de nucleótidos se hibridará con y/o amplificará una secuencia diana predeterminada y no se hibridará sustancialmente con y/o amplificará una secuencia no diana en las condiciones de ensayo, usándose generalmente condiciones de astringencia.

30

El término “hibridación” tal como se usa en el presente documento se refiere a un procedimiento mediante el cual, en condiciones de reacción predeterminadas, se permite que dos cadenas de ácido nucleico parcial o completamente complementarias se unan de manera antiparalela para formar un ácido nucleico bicatenario con enlaces de hidrógeno

estables y específicos, siguiendo normas explícitas en relación con las cuales las bases de ácido nucleico pueden emparejarse entre sí.

5 El término “hibridación sustancial” significa que la cantidad de hibridación observada será de manera que uno que observe los resultados considerará el resultado positivo con respecto a los datos de hibridación en controles positivos y negativos. Los datos que se consideran “ruido de fondo” no constituyen hibridación sustancial.

10 El término “condiciones de hibridación astringentes” significa aproximadamente de 35°C a 65°C en una disolución salina de NaCl aproximadamente 0,9 molar. La astringencia también puede regirse por parámetros de reacción tales como la concentración y el tipo de especies iónicas presentes en la disolución de hibridación, los tipos y las concentraciones de agentes desnaturizantes presentes, y la temperatura de hibridación. Generalmente, a medida que las condiciones de hibridación se hacen más astringentes, se prefieren sondas más largas **15** para la formación de híbridos estables. Como norma, la astringencia de las condiciones en las que tiene lugar la hibridación determinará ciertas características de los oligonucleótidos preferidos a emplearse, así como su concentración.

20 El término “identidad,” tal como se conoce en la técnica, es una relación entre dos o más secuencias de polipéptido o dos o más secuencias de polinucleótido, tal como se determina mediante la comparación de las secuencias. En la técnica, “identidad” también significa el grado de relación entre secuencias de polipéptido o de polinucleótido, tal como se determina mediante la coincidencia entre cadenas de tales secuencias. El término “similitud” tal como se usa en el presente documento se refiere a un grado de identidad de al menos **25** 90%, preferiblemente de al menos el 95%, el 96%, el 97%, el 98% o el 99% de identidad. La “identidad” y la “similitud” pueden calcularse fácilmente mediante métodos conocidos, incluyendo, pero sin limitarse a, los descritos en Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, Nueva York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, Nueva York, 1993; Computer Analysis of **30** Sequence Data, Parte I, Griffin, A. M., y Griffin, H. G., eds., Humana Press, Nueva Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; y Sequence Analysis Primer, Gribskov, M. y Devereux, J., eds., M Stockton Press, Nueva York, 1991; y Carillo, H., y Lipman, D., Siam J. Applied Math., 48:1073 (1988). Además, pueden obtenerse valores para la identidad en porcentaje de alineaciones de secuencias de **35** aminoácidos y de nucleótidos generadas usando los parámetros por defecto para el

componente AlignX del software Vector NTI Suite 8.0 (Informax, Frederick, Md.). Los métodos preferidos para determinar la identidad están diseñados para producir la mayor coincidencia entre las secuencias sometidas a prueba. Los métodos para determinar la identidad y la similitud se codifican en programas informáticos disponibles públicamente.

- 5** Los métodos mediante programas informáticos preferidos para determinar la identidad y la similitud entre dos secuencias incluyen, pero no se limitan a, el paquete de programas GCG (Devereux, J., *et al.*, Acid Nucleic Research 12(1): 387 (1984)), BLASTP, BLASTN y FASTA (Atschul, S. F. *et al.*, J. Molec. Biol. 215:403-410 (1990)). El programa BLAST X está disponible públicamente en el NCBI y otras fuentes (BLAST Manual, Altschul, S., *et al.*, NCBINLM NIH Bethesda, Md. 20894: Altschul, S., *et al.*, J. Mol. Biol. 215:403-410 (1990)).
- 10** También puede usarse el algoritmo de Smith Waterman bien conocido para determinar la identidad.

- El término “aptámero” tal como se usa en el presente documento se refiere a moléculas de oligonucleótido o péptido que se unen a una molécula diana específica. Los aptámeros se crean habitualmente seleccionándolos de un gran conjunto de secuencia al azar, pero los aptámeros naturales también pueden existir en *riboswitches*. Más específicamente, los aptámeros pueden clasificarse como: aptámeros de ADN y ARN y comprenden cadenas de oligonucleótidos. Los aptámeros peptídicos comprenden dominios de péptidos variables
- 15**
- 20** cortos, unidos en ambos extremos a una estructura de soporte proteica.

- El término “tratamiento” engloba tanto un tratamiento profiláctico como terapéutico. El término “tratamiento terapéutico” o “terapia” tal como se usa en el presente documento se refiere a llevar un cuerpo de un estado patológico o enfermedad de vuelta a su estado sano,
- 25** normal. El término “tratamiento profiláctico” tal como se usa en el presente documento se refiere a prevenir un estado patológico.

- El término “cantidad terapéuticamente eficaz” tal como se usa en el presente documento se refiere a una cantidad que es eficaz, tras la administración de una dosis individual o de
- 30** dosis múltiples a un sujeto (tal como un paciente humano) en el tratamiento profiláctico o terapéutico de una enfermedad, trastorno o estado patológico.

Descripción detallada

- 35** Métodos de la invención

El primer aspecto de la invención se refiere a un método para determinar o predecir si una cepa de *E. coli* tiene un fenotipo caracterizado por alta capacidad de adhesión e invasión (tal como el fenotipo de AIEC), en el que dicho método comprende determinar el genotipo de uno, dos, tres, cuatro o más polimorfismos de un solo nucleótido (SNP) seleccionados del grupo que consiste en: E1-E2_3.6, E3-E4_4.3(1), E3-E4_4.3(2), E3-E4_4.3(3), E3-E4_4.4, E5-E6_3.12, y de E5-E6_3.16=3.22(1) a E5-E6_3.16=3.22(6), en el que la asociación de variantes alélicas específicas de estos SNP con un fenotipo de AIEC altamente invasivo y adhesivo es tal como se describe en la Tabla 6.

5 Una cepa de *E. coli* con fenotipo de alta adhesión se caracteriza por presentar un índice de adhesión (I_{ADH}) ≥ 1 en células epiteliales intestinales. El I_{ADH} puede determinarse calculando el número medio de bacterias por célula, tal como se describe en el ejemplo 1.

Una cepa de *E. coli* con un fenotipo altamente invasivo se caracteriza por presentar un índice de invasión (I_{INV}) $\geq 0,1\%$ en células epiteliales intestinales, que implique el reclutamiento de microtúbulos y la polimerización de actina de las células huésped. El I_{INV} puede calcularse como el porcentaje de bacterias intracelulares tras 1 h de tratamiento con gentamicina en relación con el inóculo (es decir, $I_{INV} (\%) = (\text{bacterias intracelulares} / \text{bacterias inoculadas}) \times 100$), tal como se describe en el ejemplo 1. Además, **15** puede someterse a ensayo la implicación del citoesqueleto de la célula huésped incubando monocapas de las células con inhibidores del citoesqueleto (tales como citocalasina o colchicina) y puede determinarse el porcentaje de reducción de los índices de invasión, tal como se describe en el ejemplo 1. Normalmente, existe implicación del citoesqueleto de la célula cuando hay una reducción de al menos el 10%, preferiblemente, al menos el 20%, al menos el 30%, al menos el 40%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80% o al menos el 90% de reducción del I_{INV} cuando las células se tratan con inhibidores del citoesqueleto. **20**

Células epiteliales intestinales que pueden usarse en la determinación del I_{ADH} y/o I_{INV} incluyen, pero no se limitan a, las líneas celulares I-407 (ATCC[®] CCL-6[™]), Caco-2 (ATCC[®] HTB-37[™]), Hep-2 (ATCC[®] CCL23[™]), T84 (ATCC[®] CCL-248[™]) y HCT-8 (ATCC[®] CCL-244[™]). Preferiblemente, para el I_{ADH} estas se seleccionan del grupo que consiste en I-407 y Caco-2 y para el I_{INV} estas se seleccionan del grupo que consiste en las líneas celulares I-407 y Hep-2 (Darfeuille-Michaud *et al* 2004). **25**

35

En una realización particular, dicho método comprende determinar los genotipos de uno, dos, tres y/o cuatro SNP seleccionados del grupo que consiste en E3-E4_4.3(2), E3-E4_4.4, E5-E6_3.16=3.22(2), y E5-E6_3.16=3.22(3).

5 Preferiblemente, dicho fenotipo altamente invasivo y altamente adhesivo es un fenotipo de AIEC. Boudeau *et al.* (Boudeau *et al.*, 1999) definieron un nuevo patotipo de *Escherichia coli* denominado *Escherichia coli* adherente-invasiva (AIEC), y este patotipo se ha asociado con pacientes con enfermedad de Crohn (CD) por varios grupos independientes (Martínez-Medina, M. *et al.*, 2009; Darfeuille-Michaud, A. *et al.*, 2004; Baumgart, M. *et al.*, 2007; Martin, H. M. *et al.*, 2004; Sasaki, M. *et al.*, 2007; Eaves-Piles, T. *et al.*, 2008; Dogan, B. *et al.*, 2013). El fenotipo AIEC posee la capacidad de adherirse a e invadir células epiteliales intestinales y sobrevivir y replicarse dentro de macrófagos (Boudeau *et al.*, 1999). AIEC constituye un grupo patogénico intestinal distinto de *E. coli* en comparación con patotipos anteriormente descritos. En lo que se refiere a sus genes de virulencia, el patotipo AIEC se asemeja a las

10 *E. coli* patógenas extraintestinales (ExPEC), que en su mayoría no son invasivas (Baumgart, M. *et al.*, 2007; Miquel, S. *et al.*, 2010; Nash, J. H. *et al.*, 2010; Martinez-Medina, M. *et al.*, 2009). Asimismo, el patotipo AIEC es clonalmente diverso, comprendiendo todos los grupos filogenéticos (A, B1, B2 y D), habiéndose identificado los dos últimos como los más abundantes en pacientes con IBD (Martínez-Medina, M. *et al.*, 2009; Kotlowski, R. *et al.*, 2007). Conte M.P. *et al.* (Conte, M.P. *et al.*, 2014) también han descrito que los grupos filogenéticos más prevalentes en pacientes pediátricos con CD son los grupos filogenéticos A y D.

La cepa UM146 es una cepa de AIEC aislada de tejido de biopsia de CD ileal por Krause D.O *et al.* (Krause D.O *et al.*, 2011). Su genoma se ha usado en el presente documento como genoma de referencia (NCBI NC_017632.1; https://www.ncbi.nlm.nih.gov/nuccore/NC_017632).

25

Los métodos actuales para la caracterización del patotipo AIEC se basan en la determinación de la capacidad de la cepa de *E. coli* para adherirse a e invadir células epiteliales intestinales y sobrevivir y replicarse dentro de macrófagos (Boudeau *et al.*, 1999). Más específicamente, una cepa bacteriana de *E. coli* se caracteriza fenotípicamente como AIEC cuando presenta un fenotipo altamente adhesivo y altamente invasivo tal como se describió anteriormente en el presente documento y se caracteriza adicionalmente por

30 presentar un índice de replicación $\geq 100\%$ en una línea celular de macrófagos. Por ejemplo,

35

el índice de replicación puede calcularse como el porcentaje de bacterias intracelulares presentes a las 24 h tras la infección en relación con el número de bacterias intracelulares presentes tras 1 h de tratamiento con gentamicina (es decir, $I_{REPL} (\%) = (UFC \text{ ml}^{-1} \text{ a las } 24 \text{ h} / UFC \text{ ml}^{-1} \text{ a } 1 \text{ h}) \times 100$), tal como se describe en el ejemplo 1. Las líneas celulares de macrófagos que pueden usarse en la determinación del índice de replicación incluyen, pero no se limitan a, las líneas celulares de ratón J774A.1 (ATCC® TIB-67™) y RAW264.7 (ATCC® TIB-71™) y la línea celular humana THP-1 (ATCC® TIB-202™).

5

10

En una realización preferida, una cepa de *E. coli* se clasifica como perteneciente al fenotipo de AIEC cuando cumple los siguientes criterios, según la definición facilitada por Darfeuille-Michaud *et al* 2004:

- (1) la capacidad de las bacterias para invadir células epiteliales Intestine-407 y Hep-2 con un índice de invasión igual a o superior al 0,1% del inóculo original,
- (2) la implicación del reclutamiento de microtúbulos y la polimerización de actina de las células huésped en la captación bacteriana,
- (3) la ausencia de determinantes invasivos conocidos,
- (4) la capacidad para adherirse a células epiteliales intestinales Caco-2 diferenciadas y/o Intestine-407 no diferenciadas con un índice de adhesión igual o superior a 1 bacteria por célula, y
- (5) la capacidad para sobrevivir y replicarse dentro de macrófagos J774.A1.

15

20

Tal como se mencionó anteriormente, se demostró que el genotipado de los SNP E3-E4_4.4 y E5-E6_3.16=3.22(2) es útil para diferenciar entre los fenotipos AIEC y no-AIEC (véanse la Tabla 9 y la Figura 5).

25

Por consiguiente, la invención también se refiere a un método para determinar o predecir si una cepa de *E. coli* tiene un fenotipo de AIEC (en otras palabras, un método para detectar o predecir el fenotipo AIEC), en el que dicho método comprende:

30

- a) determinar la presencia o ausencia del gen E3-E4_4.4 en el genoma de la cepa de *E. coli*, y cuando está presente, determinar el genotipo en el sitio polimórfico E3-E4_4.4 correspondiente a la posición 1433063 pb de la cepa AIEC de referencia UM146 (NCBI NC_017632.1).

En una realización particular, la presencia de un nucleótido de G en el sitio polimórfico de a) es indicativa de un fenotipo no-AIEC. Tal como se muestra en la Figura 5, la presencia de un nucleótido de G en el sitio polimórfico E3-E4_4.4 da como resultado un 85,7% de cepas no-AIEC predichas correctamente.

5

En otra realización, dicho método comprende:

10

a) determinar la presencia o ausencia del gen E3-E4_4.4 en el genoma de la cepa de *E. coli*, y cuando está presente, determinar el genotipo en el sitio polimórfico E3-E4_4.4 correspondiente a la posición 1433063 pb de la cepa AIEC de referencia UM146 (NCBI NC_017632.1); y

15

b) determinar la presencia o ausencia del gen E5-E6_3.16=3.22 en el genoma de la cepa de *E. coli*, y cuando está presente determinar el genotipo en el sitio polimórfico E5-E6_3.16=3.22(2) correspondiente a la posición 2212845 pb de la cepa AIEC de referencia UM146 (NCBI NC_017632.1);

20

en el que la ausencia del gen E3-E4_4.4 en a), y la presencia de un nucleótido de G en el sitio polimórfico de b) son indicativas de un fenotipo de AIEC, y/o

en el que la ausencia del gen E3-E4_4.4 en a) y la ausencia del gen E5-E6_3.16=3.22 o la presencia de un nucleótido distinto de G en el sitio polimórfico de b) son indicativas de fenotipo no-AIEC; y/o

25

en el que la presencia de un nucleótido distinto de G en el sitio polimórfico de a), y la ausencia del gen E5-E6_3.16=3.22 o la presencia de G en el sitio polimórfico de b) son indicativas de fenotipo de AIEC.

30

La expresión “un nucleótido distinto de G” tal como se usa en el presente documento puede referirse a aquellos casos en que se determina la presencia de otro nucleótido o no fue posible establecer un nucleótido específico para esta posición (por ejemplo, se encuentra un pico solapante en esta posición cuando se realiza una secuenciación de Sanger).

En una realización adicional, dicho método comprende:

35

a) determinar la presencia o ausencia del gen E3-E4_4.4 en el genoma de la cepa de *E. coli*, y cuando está presente, determinar el genotipo en el sitio polimórfico

E3-E4_4.4 correspondiente a la posición 1433063 pb de la cepa AIEC de referencia UM146 (NCBI NC_017632.1);

5 b) determinar la presencia o ausencia del gen E5-E6_3.16=3.22 en el genoma de la cepa de *E. coli*, y cuando está presente determinar el genotipo en el sitio polimórfico E5-E6_3.16=3.22(2) correspondiente a la posición 2212845 pb de la cepa AIEC de referencia UM146 (NCBI NC_017632.1); y

10 c) determinar la presencia o ausencia del gen E5-E6_3.12 en el genoma de la cepa de *E. coli*, y cuando está presente determinar el genotipo en el sitio polimórfico E5-E6_3.12 correspondiente a la posición 4271710 pb de la cepa AIEC de referencia UM146 (NCBI NC_017632.1);

15 en el que la presencia de un nucleótido distinto de G en el sitio polimórfico de a), la presencia de un nucleótido distinto de G en el sitio polimórfico de b) y la presencia de G en el sitio polimórfico de c) son indicativas de un fenotipo de AIEC, y/o

en el que la presencia de un nucleótido distinto de G en el sitio polimórfico de a), la presencia de un nucleótido distinto de G en el sitio polimórfico de b) y la presencia de A en el sitio polimórfico de c) son indicativas de un fenotipo no-AIEC.

20

El método según cualquiera de las realizaciones anteriores puede comprender además determinar el genotipo de uno, dos, tres o más de los SNP seleccionados del grupo que consiste en: E1-E2_3.6, E3-E4_4.3(1), E3-E4_4.3(2), E3-E4_4.3(3), E5-E6_3.16=3.22(1), E6_3.16=3.22(3), E6_3.16=3.22(4), E6_3.16=3.22(5) y E5-E6_3.16=3.22(6), tal como se describe en el presente documento.

25

Todavía en una realización adicional, la invención se refiere a un método para determinar o predecir si una cepa de *E. coli* tiene un fenotipo de AIEC, en el que dicho método comprende:

30

- i. determinar la presencia o ausencia del gen E5-E6_3.16=3.22 en el genoma de la cepa de *E. coli*, y cuando está presente determinar el genotipo en el sitio polimórfico E5-E6_3.16=3.22(2) correspondiente a la posición 2212845 pb de la cepa AIEC de referencia UM146 (NCBI NC_017632.1);

en el que la presencia de un nucleótido de G en el sitio polimórfico de b) es indicativa de un fenotipo de AIEC.

- El método según la realización anterior, puede comprender además determinar el genotipo de uno, dos, tres o más SNP seleccionados del grupo que consiste en: E1-E2_3.6, E3-E4_4.3(1), E3-E4_4.3(2), E3-E4_4.3(3), E3-E4_4.4, E5-E6_3.12, E5-E6_3.16=3.22(1), E5-E6_3.16=3.22(3), E5-E6_3.16=3.22(4), E5-E6_3.16=3.22(5) y E5-E6_3.16=3.22(6), tal como se describe en el presente documento.
- 5**
- 10** Un experto en la técnica entenderá que también puede ser posible realizar la determinación de los SNP mencionados anteriormente a nivel de proteínas cuando la sustitución de nucleótidos no es sinónima y da como resultado un cambio de aminoácido. Estos se han especificado en la Tabla 10 a continuación para los SNP confirmados no específicos de cepa tal como se describe en el presente documento.
- 15**

Tabla 10. Variantes de nucleótido en los SNP confirmados no específicos de cepa y que corresponden a cambios de aminoácido.

ID	Variante de nucleótido* ¹	Variante de nucleótido detectado mediante Sanger* ¹	Cambio de aminoácido
E1-E2_3.6	C/T	C/T	Alanina/Treonina
E3-E4_4.3(1)	C/T	C/Y	Asparagina/Aspartato
E3-E4_4.3(2)	C/T	Y/Y	Serina/Glicina
E3-E4_4.3(3)	T/G	T/K	Glutamato/Lisina
E3-E4_4.4	A/G	R/R	Glutamato/ Glicina
E5-E6_3.12	A/G	A/G	Treonina/Alanina
E5-E6_3.16=3.22(1)	T/C	Y/Y	Histidina/Alanina
E5-E6_3.16=3.22(2)	C/G	S/S	Arginina/Prolina
E5-E6_3.16=3.22(3)	G/A	R/R	Arginina/Cisteína
E5-E6_3.16=3.22(4)	C/A	M/M	Leucina/Fenilalanina
E5-E6_3.16=3.22(5)	C/T	Y/Y	Cisteína/Tirosina
E5-E6_3.16=3.22(6)	C/G	S/S	Valina/Leucina

*¹ A: adenina; C: citosina; G: guanina; K: guanina o timina; M: adenina o citosina; R: adenina o guanina; S:

5 guanina o citosina; T: timina; W: adenina o timina; Y: citosina o timina.

Por ejemplo, para el SNP E3-E4_4.4 en el que el cambio de nucleótido A/G se ha descrito en el presente documento que da lugar a un cambio de aminoácido E (Glu)/G (Gly); para el SNP E5-E6_3.16=3.22(2) en el que el cambio de nucleótido C/G se ha descrito en el presente documento que da lugar a un cambio de aminoácido R (Arg)/P(Pro); y el SNP E5-E6_3.12 en el que el cambio de nucleótido A/G se ha descrito en el presente documento que da lugar a un cambio de aminoácido T (Thr)/A (Ala).

Las proteínas codificadas por cada uno de los 5 genes que comprenden SNP no específicos de cepa tal como se describe en el presente documento se facilitan en la Tabla 11 a continuación.

Tabla 11. Proteínas correspondientes a los genes de *E. coli* que comprenden SNP confirmados no específicos de cepa.

ID de gen	Nombre de proteína	Cobertura	Identidad	Número de registro de NCBI
E1-E2_3.6	Proteína hipotética	100%	93%	CAE55806.1
E3_E4_4.3	Proteína hipotética	99%	97%	ERC39664.1
E3_E4_4.4	dGTPasa	100%	99%	WP_021531029.1
E5-E6_3.12	sucC, cadena de succinil-CoA sintetasa beta	100%	99%	EQX39887.1
E5-E6_3.16=3.22	Proteína de la familia Ail/Lom	100%	100%	WP_053902784.1

- 5** Se conocen bien en la técnica métodos para detectar y/o cuantificar específicamente proteínas. Tradicionalmente, la detección y/o la cuantificación de proteínas en disolución se han llevado a cabo mediante inmunoensayos sobre un soporte sólido. Dichos inmunoensayos pueden ser por ejemplo un ensayo de inmunoabsorción ligado a enzimas (ELISA), un ensayo de inmunoabsorción fluorescente (FIA), un inmunoensayo de quimioluminiscencia (CIA) o un radioinmunoensayo (RIA), un inmunoensayo enzimático
- 10** homogéneo, un radioinmunoensayo en fase sólida (SPROA), un ensayo de polarización fluorescente (FP), un ensayo de transferencia de energía por resonancia de fluorescencia (FRET), un ensayo de transferencia de energía por resonancia de fluorescencia con resolución en el tiempo (TR-FRET), un ensayo de resonancia de plasmón superficial (SPR).
- 15** Se engloban específicamente las versiones multiplex y de nueva generación de cualquiera de los anteriores, tales como inmunoensayos de citometría de flujo basado en perlas (*bead-based flow-cytometry*) (por ejemplo, basados en la tecnología Luminex xMAP).

- Otros métodos que pueden usarse para la cuantificación de proteínas son técnicas basadas
- 20** en espectrometría de masas (EM) tal como cromatografía de líquidos acoplada a espectrometría de masas (CL/EM), descrita por ejemplo en el documento US2010/0173786, o CL-EM/EM en tándem (documentos WO2012/155019, US2011/0039287, M. Rauh, J Chromatogr B Analyt Technol Biomed Life Sci, 1 de febrero de 2012, 883-884, 59-67) y el uso de alineamientos de péptidos, proteínas o anticuerpos y versiones multiplex de las

técnicas anteriores, así como la siguiente generación de tales técnicas y combinaciones de las mismas.

El orden cronológico de etapas en los métodos tal como se describe en el presente documento no está limitado particularmente. Estas pueden producirse en el orden indicado, o en cualquier orden, o simultáneamente. En una realización particular, puede usarse la determinación del genotipo en los SNP según los métodos de la invención para obtener una puntuación según un algoritmo matemático dado, en el que el genotipo en cada uno de los SNP usados en los métodos de la invención son las variables de dicho algoritmo matemático.

No se reivindica que un método de la invención, tal como entiende un experto en la técnica, sea correcto en el 100% de las muestras analizadas. Sin embargo, requiere que se clasifique correctamente una cantidad estadísticamente significativa de las muestras analizadas. La cantidad que es estadísticamente significativa puede establecerse por un experto en la técnica por medio del uso de diferentes herramientas estadísticas; ejemplos ilustrativos, no limitativos de dichas herramientas estadísticas incluyen determinar intervalos de confianza, determinar el valor de p, las funciones de discriminación de la prueba de la Chi cuadrado, etc. Intervalos de confianza preferidos son al menos el 90%, al menos el 95%, al menos el 97%, al menos el 98%, al menos el 99%. Los valores de p son, preferiblemente de menos de 0,1, menos de 0,05, menos de 0,01, menos de 0,005 o menos de 0,001. Las enseñanzas de la presente invención preferiblemente permiten determinar o predecir correctamente el fenotipo diana en al menos el 60%, en al menos el 70%, en al menos el 80%, en al menos el 90% o en al menos el 95% del grupo o la población analizada.

La sensibilidad, la especificidad y/o la exactitud son parámetros usados normalmente para describir la validez o el rendimiento de una prueba. En particular, se usan para cuantificar lo bueno y fiable que es el método de discriminación. Una prueba se calibra habitualmente en lo que se refiere a la especificidad y sensibilidad deseadas según el uso específico de la prueba en la práctica clínica. Alta sensibilidad corresponde a un valor predictivo negativo alto y es una propiedad ideal para una prueba de "exclusión". Alta especificidad corresponde a un valor predictivo positivo alto y es una propiedad ideal para una prueba de "inclusión".

En realizaciones preferidas, los métodos de la invención tienen valores de sensibilidad, especificidad y/o exactitud de al menos aproximadamente el 60%, preferiblemente de al menos aproximadamente el 70%, y pueden ser, por ejemplo, de al menos el 75%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o el 100% en al menos el 60% del grupo o la población sometida a ensayo, o preferiblemente en al menos el 65%, el 70%, el 75%, el 80%, el 85%, el 90%, el 95% o el 100% del grupo o la población sometida a ensayo.

La determinación del genotipo en el SNP, que tal como se usa en el presente documento incluye determinar la presencia o ausencia del gen que contiene el polimorfismo de un solo nucleótido (SNP) y/o la variante alélica en cada una de las posiciones de SNP según los métodos de la invención, puede realizarse mediante cualquiera de los métodos conocidos en la técnica, que incluyen, pero no se limitan a, secuenciación de ADN, un ensayo de detección basado en ácido nucleico o cualquier otro método de genotipado conocido en la técnica. En realidad, se dispone de numerosas estrategias para el análisis del genotipo. Los métodos de ejemplo incluyen, pero no se limitan a, secuenciación directa, análisis de polimorfismo de longitud de fragmentos de restricción (RFLP); hibridación con oligonucleótidos específicos de alelos (ASO) que son sondas sintéticas cortas que se hibridan sólo con una secuencia perfectamente coincidente en condiciones de astringencia de hibridación adecuadas; PCR específica de alelo; PCR y secuenciación, PCR usando cebadores mutagénicos; ligasa-PCR, escisión mediante HOT; electroforesis en gel con gradiente de desnaturalización (DGGE), electroforesis en gel con gradiente de desnaturalización por temperatura (TGGE), polimorfismo conformacional de cadena sencilla (SSCP), cromatografía de líquidos de alto rendimiento con desnaturalización (Kuklin *et al*, 1997), microalineamientos de ADN, metagenómica y reconocimiento de aptámeros. La secuenciación directa puede llevarse a cabo mediante cualquier método, incluyendo sin limitación secuenciación química, usando el método de Maxam-Gilbert; mediante secuenciación enzimática, usando el método de Sanger; secuenciación por espectrometría de masas; pirosecuenciación, secuenciación usando un microalineamiento de ADN y PCR cuantitativa.

El término oligonucleótidos específicos para la secuencia diana puede referirse a un cebador, una sonda, un par de cebadores, un par de cebadores y una sonda, o un cebador y una sonda. Un experto en la técnica conocerá los oligonucleótidos más apropiados o la combinación de los mismos según el método molecular usado para propósitos de

genotipado. A modo de ejemplo, para propósitos de amplificación, generalmente se usan dos cebadores, mientras que para propósitos de secuenciación se usa a menudo un solo cebador.

- 5** En algunas realizaciones, la determinación se realiza mediante un ensayo de detección basado en ácido nucleico. En una realización particular, dicho ensayo de detección basado en ácido nucleico es un ensayo basado en amplificación. Un ensayo basado en amplificación de este tipo comprende una etapa de amplificación que comprende poner en contacto una muestra (preferiblemente una muestra de ADN aislada) con dos o más oligonucleótidos de amplificación específicos para una secuencia diana en un ácido nucleico diana para producir un producto amplificado si la secuencia de ácido nucleico diana está presente en la muestra. Los métodos de amplificación adecuados incluyen por ejemplo, amplificación mediada por replicasa, reacción en cadena de ligasa (LCR), amplificación por desplazamiento de cadena (SDA), amplificación mediada por transcripción (TMA) y reacción en cadena de la polimerasa (PCR), que incluye PCR cuantitativa.

- Por ejemplo, el genotipado de los SNP según los métodos de la invención puede detectarse en una muestra de ADN, preferiblemente tras la amplificación. Por ejemplo, el ADN aislado puede someterse a amplificación mediante PCR, usando cebadores de oligonucleótido específicos que son específicos para el polimorfismo o que permiten la amplificación de una región que contiene el polimorfismo. En un primer ejemplo, las condiciones para el apareamiento del cebador pueden elegirse para garantizar la amplificación específica; de modo que la aparición de un producto de amplificación será indicativa de la presencia de la secuencia diana (por ejemplo la forma alélica diana en el SNP). De forma alternativa, el ADN puede amplificarse, tras lo cual puede detectarse la secuencia diana en la secuencia amplificada mediante hibridación con una sonda adecuada o mediante secuenciación directa, o cualquier otro método apropiado conocido en la técnica. Preferiblemente, la determinación del genotipo se realiza mediante PCR específica de alelo.

- 30** Estos oligonucleótidos pueden ser ribonucleótidos o desoxirribonucleótidos. En realizaciones particulares, los oligonucleótidos pueden tener al menos una modificación química. Por ejemplo, los oligonucleótidos adecuados pueden estar compuestos por una o más modificaciones de nucleósido de azúcar bícíclico o “restringidas conformacionalmente”, por ejemplo, “ácidos nucleicos bloqueados”. Los “ácidos nucleicos bloqueados” (LNA) son ribonucleótidos modificados que contienen un puente extra entre los carbonos 2' y 4' del

resto de azúcar de ribosa dando como resultado una conformación “bloqueada” que confiere estabilidad térmica potenciada a los oligonucleótidos que contienen los LNA. En otras realizaciones, los oligonucleótidos pueden comprender ácidos nucleicos peptídicos (PNA), que contienen una estructura principal basada en péptidos en lugar de una estructura principal de azúcar-fosfato. Otras modificaciones químicas que pueden contener los oligonucleótidos incluyen, pero no se limitan a, modificaciones de azúcar, tales como modificaciones de 2'-O-alkilo (por ejemplo 2'-O-metilo, 2'-O-metoxietilo), 2'-fluoro y 4'-tio, y modificaciones de la estructura principal, tales como una o más uniones de fosforotioato, morfolino o fosfonocarboxilato. Por ejemplo, estos oligonucleótidos, particularmente los de longitudes más cortas (por ejemplo, menos de 15 nucleótidos) pueden comprender una o más modificaciones de potenciación de la afinidad, tales como, pero sin limitarse a, LNA, nucleósidos bicíclicos, fosfonoformiatos, 2' O-alkilo y similares. En algunas realizaciones, los oligonucleótidos pueden modificarse clínicamente, por ejemplo para mejorar su resistencia a la degradación por nucleasas (por ejemplo, mediante ocupación de extremos), para llevar ligandos de detección (por ejemplo, fluoresceína) o para facilitar su captura sobre un soporte sólido (por ejemplo, “colas” de poli-desoxiadenosina).

Los oligonucleótidos preferibles útiles en los métodos de la invención tienen de aproximadamente 5 a aproximadamente 50 nucleótidos de longitud, de aproximadamente 10 a aproximadamente 30 nucleótidos de longitud, o de aproximadamente 20 a aproximadamente 25 nucleótidos de longitud. En determinadas realizaciones, los oligonucleótidos que se hibridan específicamente con la secuencia diana tienen de aproximadamente 19 a aproximadamente 21 nucleótidos de longitud. Estos oligonucleótidos pueden comprender en algunas realizaciones una secuencia que es al menos parcialmente complementaria a la secuencia de ácido nucleico diana que comprende la forma alélica diana del SNP, por ejemplo complementaria en al menos aproximadamente el 75%, el 80%, el 85%, el 90%, el 95%, el 96%, el 97%, el 98% o el 99% a la secuencia de ácido nucleico diana.

En la presente invención, la secuencia de ácido nucleico diana se refiere a la región de la secuencia de ácido nucleico del genoma de referencia de AIEC (es decir, de la cepa de AIEC de referencia UM146) que comprende la forma alélica diana del SNP. La secuencia de ácido nucleico diana puede incluir por ejemplo desde 5 hasta 600 nucleótidos en el sentido de 5' y/o en el sentido de 3' del SNP diana, preferiblemente desde 10 hasta 400 nucleótidos en el sentido de 5' y/o en el sentido de 3' del SNP diana, más preferiblemente desde 20

hasta 200 nucleótidos en el sentido de 5' y/o en el sentido de 3' del SNP diana, tal como al menos 20, 30, 40, 50, 60, 70, 80, 90, 100 o al menos 150 nucleótidos en el sentido de 5' y/o en el sentido de 3' del SNP diana. Para propósitos ilustrativos, se muestran secuencias diana de ejemplo para los sitios polimórficos E3-E4_4.4, E5-E6_3.16=3.22(2) y E5-E6_3.12 a continuación en el presente documento.

SEQ ID NO: 27 que comprende el SNP E3-E4_4.4 (resaltado en negrita y subrayado), correspondiente a CP002167.1: 1433652-1432471 del genoma completo de *Escherichia coli* UM146:

10

ATGCACGAAAAGAACATCGCCCTGCTTTGTGATGAAGCCGACCGACTTTTGCAACTGAACATTAA
TCTGCTCCGGCAAATGGTTGAGGAGCCAGATGTGTTATCTGACAGTAAGAACGAAAACAGACTG
CTTTTTGATAAACAGAAAGCACTGAAAAGAATTGAGGAGCTGGAGGGCGAACAAATCAAAACCGC
CCGCAGGGAGATGGTGCTGGCTGTTGTCGGCAGATGAAAGCAGGCAAATCAACCACCATCAA
15 CGCCATTGTGGGGCAGGAAATTCTGCCTAACCGTAACCGCCCCATGACCTCTGTACCGACGCTC
ATCCGCCACGTTCCCGGAAAACTGAGCCGGTTCTCCATCTGGAACATATTGAGCCTGTCCGCA
ATTTATTAATCACACTGCAGGAAAACTCGCCACCCCGGCAGGACAGCAGGTGCGACAGACCCT
GCAGCAAACCGGGGATACCCGCGAACTGCTGGATATTCTGACGGATGATGGCTGGCTCAAAAAT
GAATACCACGGGGAGGAGGAAATCTTTACCGGACTGGCATCGTTAAACGATCTGGTTCGTCTTG
20 CTGCGGCAATGGGGACTGAATTTCTTTTGATGAATACGCAGAAGTGCAGAACTGCCGGTGAT
CGACGTGGAATTCAGCCATCTGGTGGGGATGGATGCATGCCAGGGAACACTCACACTGCTGGAT
ACCCCGGCCCTAATGAGGCCGGACAACCGCAGATGGAAGTGATGATGCGGGATCAACTGCAG
AAAGCCTCTGCGGTTCTGGCTGTGATGGATTACCCAGATGAACTCAAAAGCGGATGAAGACG
TCCGTAAAGAGCTTAATGCCATTGCTGACGTATCAACCGGCCGCCTGTTTGTACTGGTCAATAAA
25 TTTGATGAGAAAGACCGCAATGGCGATGGGGCAGATGCCGTACGCCAGAAAGTTCCGGCAATG
CTGAACAGCGATGTGCTGCCCGCCTCCCGCGTTTATCCCGGATCCTCACGCCAGGCATACCTGG
CTAACCGTGCGCTTCATGAGTTACGGAAAAACGGAACCCTTCCTGTTGATGAAGCCTGGGTGCA
TGATTTTGTGAGGGAAGCCTTCGGTGCATGAAAAAAGATTACGTCTGTAAAGACAGCGAACTGG
CAACTGAAGGGGCGACAGACCTGTGGGAA

30

SEQ ID NO: 28 que comprende el SNP E5-E6_3.16=3.22(2) (resaltado en negrita y subrayado), correspondiente a CP002167.1:2212514-2212960 del genoma completo de *Escherichia coli* UM146:

35

ATCACTTCCGGTCAGCACATCGTGCGTTTTCCCTTGTTGTCAGTTACGCGGAGATAATCCCCGG
AGAAAGTCGACACACGGCTGTAAGCCACACCCGCCATCGAATACGCGCTGAACCATTCATTAC
GCGTACAGACGGCCCCGCCATCACGCTGAACCAGCGGTTACGCACGGAATCTTCATGCCAGCG
GGTATCGCTGTAGTGCGTTTTTGTCTCATCCTCAGCATTGGCATAACTGAAGGACGTAATCAGCC
CCAGCGCGTCCGTAACTCATAACGGTATTTACGTTAATCCCGTTTCAGATTATCGCTGCCGGGA
40 GCGTTTCGTACGGGGCATGAAGATACCCCGCGCTCAGTGTGGACTGATGTTTCAGACGCCCATGCA
GGCGCACCGGATACGGACAGACAGATGGCTGCGGACAAAATGACTGCACAACTTTACGCAT

SEQ ID NO: 29 que comprende el SNP E5-E6_3.12 (resaltado en negrita y subrayado), correspondiente a CP002167.1:4271302-4271808 del genoma completo de *Escherichia coli* UM146:

5

AAAGGCGCGGTACTTAAAGCGCAGGTACATGCTGGTGGGCGCGGTAAAGCTGGCGGCGTAAAA
GTGCTTAAGCAGCTGCCGGAAGCACAGGCTTTTGTACAGCAAATGTTGGGATCGCAACTGGTGA
CTTATCAGACCGGGCCAGAAGGGCAGTATGTCAGCAGTATTTTGTCTGTGCGAAAACATCTATCC
GGTACGCCAGGAAGTCTATTTTGGCATGGTGGTGGATCGTGAAAGCCAGCGGGTCACGTTTATT
10 GTCAGCCCGGAAGGTGGCGTGGAATTGAAAAAGTCGCCCATGAGACACCGGAGAAAATCAGC
AGTGTGAGCATTGATCCACTGACAGGCGTGCAGCCTTGCCATATCCGCGAAATGTTTGTCTGTTCT
GCAACTGGAACACGGGCTATTTGCT**A**CCTTTAGTCGCCTGGTTAATCAGGCGTGGAAGCCTTTA
ACGAACTGGATTTGCCCTGCTGGAAATTAACCCTCTGGTGTTCGCGGAAACGGGCGAA

15 En realizaciones preferidas, el oligonucleótido puede ser sustancialmente complementario a la secuencia de ácido nucleico diana, que es complementaria en al menos aproximadamente el 95%, el 96%, el 97%, el 98% o el 99% a una secuencia de polinucleótido diana. En aquellas realizaciones en las que los oligonucleótidos son parcial o sustancialmente complementarios, es perfectamente complementario para el SNP diana y

20 preferiblemente, complementario en el 97%, el 98% o el 99% para los 1, 2, 3, 4 ó 5 nucleótidos en el sentido de 5' y/o en el sentido de 3' con respecto al SNP diana. En una realización, el oligonucleótido comprende una secuencia que es complementaria en el 100% a la secuencia de ácido nucleico diana que comprende la forma alélica diana del SNP.

25 Los oligonucleótidos que hibridan específicamente con la secuencia diana (por ejemplo, oligonucleótidos de cebado y/o de sonda) pueden comprender una secuencia seleccionada del grupo que consiste en las secuencias de oligonucleótido divulgadas en la Tabla 13 o una secuencia con una identidad de al menos aproximadamente el 90%, preferiblemente al menos aproximadamente el 95%, el 96%, el 97%, el 98% o el 99%.

30

En una realización particular, opcionalmente en combinación con una o más de las características o realizaciones descritas anteriormente o a continuación, el método según la invención es útil para la determinación del fenotipo diana (por ejemplo el fenotipo de AIEC) en una cepa de *E. coli* de cualquier grupo filogenético, preferiblemente de los grupos

35 filogenéticos A, D, B2 o B1. De hecho, la Tabla 9 muestra que la ausencia de G en el sitio polimórfico E3-E4_4.4 y la presencia de G en el sitio E5-E6_3.16=3.22(2) han demostrado

ser predictivas del fenotipo de AIEC dentro de un conjunto de cepas de los grupos filogenéticos A, D, B2 o B1 (véase la Tabla 7).

5 La cepa de *E. coli* cuyo fenotipo se determina según un método de la invención puede obtenerse de una muestra biológica de un sujeto.

10 El término “muestra” o “muestra biológica”, tal como se usa en el presente documento, se refiere a material biológico aislado de un sujeto. La muestra biológica puede contener cualquier material biológico adecuado para detectar la secuencia diana y puede comprender material celular y/o no celular del sujeto. La muestra puede aislarse de cualquier tejido o fluido biológico adecuado tal como, por ejemplo, sangre, plasma sanguíneo, suero, líquido cefalorraquídeo (LCR), orina, líquido amniótico, líquidos linfáticos, secreciones externas de los aparatos respiratorio, digestivo, genitourinario, lágrimas, saliva, glóbulos blancos. En algunas realizaciones, las muestras usadas para la detección de la secuencia diana en los **15** métodos de la invención son muestras que pueden obtenerse usando procedimientos mínimamente invasivos. En una realización preferida, las muestras son muestras intestinales. Estas pueden ser por ejemplo una biopsia del tejido mucoso del aparato intestinal, tal como el íleon, colon y/o recto. Preferiblemente, esta muestra intestinal es una muestra de heces.

20

Estos tipos de muestras se usan de manera rutinaria en la práctica clínica y un experto en la técnica conocerá cómo identificar los medios más apropiados para su obtención y conservación. Una vez que se ha obtenido una muestra, puede usarse recién preparada, puede congelarse, liofilizarse o conservarse usando medios apropiados. Preferiblemente, **25** dicha muestra es una muestra recién preparada.

Normalmente, la determinación molecular o la predicción del fenotipo de AIEC según un método de la invención se realiza en células de una cepa aislada de *E. coli*. Se conocen bien en la técnica métodos para el aislamiento de cepas de *E. coli*. Estos métodos **30** comprenderán generalmente una primera etapa de selección mediante cultivo en medios selectivos (tales como agar TBX o agar MacConkey) seguido por un ensayo de confirmación bioquímica, tal como el ensayo de indol o sistemas de biotipificación tales como el sistema Analytab Enteric (API 20E) o el sistema automatizado Vitek (BioMérieux, Marcy l'Etoile, Francia). En una realización preferida, el aislamiento de la cepa de *E. coli* se **35** realiza mediante cultivo selectivo en agar TBX seguido por ensayos de indol y/o Vitek

(véase por ejemplo, Martínez-Medina *et al.* 2009). Posteriormente, pueden amplificarse y/o secuenciarse regiones diana del genoma de *E. coli* para el genotipado de los SNP diana según los métodos de la invención

- 5** En algunas realizaciones, los métodos de la invención pueden comprender una etapa de extracción de ADN antes de la determinación del genotipo de los SNP diana. En una realización particular, se lisan células bacterianas y se obtienen extractos con cloroformo/metanol. Por ejemplo, puede extraerse el ADN genómico de células bacterianas usando el kit de purificación de ADN genómico Wizard® (Promega) según las instrucciones del fabricante; y preferiblemente las muestras se tratan con ARNasa A. En otras realizaciones, la determinación del genotipo puede llevarse a cabo en el material obtenido además de la lisis de las células bacterianas (por ejemplo, calentando hasta 95°C durante 3 minutos). Por ejemplo, se realiza amplificación por PCR con oligonucleótidos específicos diana en el producto de la lisis celular seguido por secuenciación de los productos de PCR
- 10**
- 15** para la determinación de la forma alélica en el SNP diana.

- En un aspecto relacionado, la invención se refiere a un método para diferenciar o identificar de manera diferencial cepas de AIEC de, cepas no-AIEC, normalmente similares desde el punto de vista genético, en el que dicho método comprende las etapas tal como se describió anteriormente en el presente documento según el primer aspecto. Por ejemplo dicho método permitirá distinguir molecularmente una cepa de *E. coli* con un fenotipo AIEC de una cepa de *E. coli* con un fenotipo ExPEC. El patotipo de AIEC se asemeja a ExPEC en sus genes de virulencia, pero estos patovares difieren fenotípicamente en que ExPEC es principalmente no invasivo (Baumgart, M. *et al.*, 2007; Miquel, S. *et al.*, 2010; Nash, J. H. *et al.*, 2010; Martínez-Medina, M. *et al.*, 2009). Se ha descrito que las cepas de ExPEC pertenecen a diferentes grupos filogenéticos. Para propósitos de ilustración, las cepas de ExPEC incluyen, pero no se limitan a, las siguientes: UMNK88 (número de registro de NCBI CP002729.1) del grupo filogenético A, 55989 (número de registro de NCBI CU928145.2) del grupo filogenético B1, S88 (número de registro de NCBI CU928163.2) del grupo filogenético
- 20**
- 25** B2, 042 (número de registro de NCBI NC_017626.1) del grupo filogenético D y EDL933 (número de registro de NCBI AE005174.2) del grupo filogenético E.

- En otro aspecto, la invención proporciona un método para clasificar un sujeto como portador de AIEC según la presencia de *E. coli* del fenotipo AIEC en su tracto intestinal, en el que
- 30**
- 35** dicho método comprende:

- a. aislar cepas de *E. coli* de una muestra intestinal obtenida de un sujeto;
 - b. opcionalmente, extraer ADN de dichas cepas de *E. coli*;
 - c. determinar o predecir la presencia de *E. coli* del fenotipo AIEC usando un método
- 5** según un método tal como se describe en el presente documento.

En un aspecto relacionado, la invención se refiere a un método para el examen de sujetos portadores de AIEC o la monitorización del estado de portador de AIEC en un sujeto, en el que dicho método comprende las etapas a) a c) tal como se describió en el aspecto anterior.

10

En algunas realizaciones, puede ser deseable identificar un subgrupo de sujetos portadores de AIEC que tienen niveles de AIEC por encima de un umbral dado. Esto puede implicar la determinación de la abundancia de cada uno de los SNP diana en una población de *E. coli* aislada de dicho sujeto.

15

En una realización particular, dicho sujeto tiene una enfermedad intestinal. En una realización preferida, dicha enfermedad es la enfermedad inflamatoria intestinal.

20

El término “enfermedad inflamatoria intestinal” (IBD) tal como se usa en el presente documento se refiere a un grupo de estados intestinales inflamatorios crónicos idiopáticos. Las dos categorías de enfermedad principales que cubre el término son CD y colitis ulcerosa (UC), ambas con características clínicas y patológicas solapantes y diferenciadas.

25

El diagnóstico de IBD requiere un examen físico completo y una revisión de los antecedentes del paciente. Diversas pruebas, incluyendo análisis de sangre, examen de heces, endoscopia, biopsias y estudios de obtención de imágenes ayudan a excluir otras causas y a confirmar el diagnóstico. (World Gastroenterology Organisation Global Guidelines, Inflammatory bowel disease: a global perspective, junio de 2009; y Silverberg *et al.*, Can J Gastroenterol. 2005, 19 Supl. A:5-36). Con una comprensión cada vez mayor de la epidemiología y la genética de la IBD, ha sido evidente para los médicos que la UC y la

30

CD pueden representar en realidad formas distintas de IBD. Por tanto, el término “IBD” tal como se usa en el presente documento incluye fenotipos de las mismas. Normalmente, la CD se distingue de la UC por ser una enfermedad proximal al colon, enfermedad perineal, con fístulas, granulomas histológicos y grosor completo en contraposición a enfermedad limitada a la mucosa. Generalmente, en la CD, los granulomas son evidentes en hasta el

35

50% de los pacientes y las fístulas en el 25%.

Se ha descrito la presencia de AIEC en la mucosa intestinal de un subgrupo de pacientes con CD (Darfeuille-Michaud A, *et al.* 2004; Martínez-Medina M. *et al.*, 2009, Sivignon A, *et al.* 2017). La abundancia de AIEC, definida como el porcentaje de AIEC dentro de la población de *E. coli*, generalmente es baja y variable, oscilando entre el 1% y el 50% en pacientes con CD. Por ejemplo, los inventores encontraron previamente que como promedio, los aislados de AIEC representaban el 9,3%, el 3,7% y el 3,1% de los aislados de *E. coli* en pacientes con CD ileal, ileocolónica y colónica, respectivamente (Martínez-Medina M. y García-Gil J.L., 2014). En una realización particularmente preferida dicha enfermedad intestinal es CD, incluyendo CD ileal, ileocolónica y/o colónica.

Las cepas de AIEC también se han asociado con enfermedad intestinal en animales, particularmente en gatos, perros y cerdos (Martínez Medina *et al.*, 2009). En un estudio previo, Simpson *et al.* (Simpson *et al.*, 2006) detectaron AIEC en perros bóxer que padecían colitis granulomatosa, una enfermedad con características patológicas que solapan con la CD, lo que apoya el papel de AIEC en la CD humana y en enfermedades análogas en animales.

En una realización preferida, dicho sujeto es un sujeto humano. En otra realización preferida, dicho sujeto es un mamífero distinto de un humano. Preferiblemente, dicho animal es un gato, perro o cerdo.

En una realización particular, el método de monitorización del estado de portador de AIEC tal como se describió anteriormente en el presente documento se realiza en un sujeto clasificado como portador de AIEC que tiene una enfermedad intestinal, en el que dicho método comprende la determinación de la presencia y/o la abundancia del fenotipo AIEC tras el tratamiento terapéutico.

Además, se ha notificado que las cepas de AIEC pueden promover o perpetuar la inflamación crónica. Por tanto, se han desarrollado diversas estrategias dirigidas contra estas cepas de *E. coli* para promover su eliminación en el intestino. Dichas estrategias constituyen por tanto tratamientos personalizados para sujetos portadores de AIEC que tienen una enfermedad intestinal (también denominados en el presente documento “pacientes portadores de AIEC”), que tienen como diana la microbiota.

Por consiguiente, en un aspecto adicional, la invención se refiere a un método para tratar a un sujeto que tiene una enfermedad intestinal tal como se describe en el presente documento, en el que dicho método comprende:

- 5** i) seleccionar un sujeto que tiene una enfermedad intestinal en el que dicho sujeto se ha clasificado como portador de AIEC mediante un método de clasificación de la invención tal como se describió anteriormente en el presente documento;
- ii) administrar a dicho sujeto una cantidad terapéuticamente eficaz de un
- 10** tratamiento destinado a eliminar o disminuir los niveles de bacterias AIEC de la mucosa intestinal tal como se describe en el presente documento.

Una cantidad terapéuticamente eficaz disminuirá los niveles de AIEC en al menos el 10%, preferiblemente al menos el 20%, el 30%, el 40%, el 50%, el 60%, el 70%, el 80%, el 90% o al menos el 95% en un método funcional o molecular tal como se describe en el presente

15 documento. La determinación de los niveles de AIEC puede realizarse en una muestra intestinal de dicho sujeto tal como se describió anteriormente en el presente documento.

Las estrategias terapéuticas dirigidas a eliminar o disminuir los niveles de bacterias AIEC de la mucosa intestinal pueden dividirse en dos grupos: i) estrategias que eliminan AIEC, tales

20 como antibióticos o terapia con fagos; y ii) estrategias que bloquean la adhesión bacteriana a las células en cepillo intestinales tal como se describe a continuación en el presente documento (Sivignon A, *et al.* 2017).

El tratamiento con antibióticos puede comprender administrar un solo agente antibiótico o

25 diversos agentes antibióticos en combinación. Por ejemplo, uno o más antibióticos seleccionados de la lista que consiste en ciprofloxacino, rifampicina, sulfametoxazol, tetraciclina y trimetoprim. En una realización particular, dicho tratamiento puede comprender una combinación de ciprofloxacino, tetraciclina y trimetoprim (Subramanian S., *et al.*, 2008).

30 Puede usarse terapia con bacteriófagos como estrategia para seleccionar como diana específicamente AIEC. Esta incluye por ejemplo tratamiento con uno o más bacteriófagos, preferiblemente los que pertenecen a la familia de virus *Myoviridae*. En una realización particular, dicho tratamiento con bacteriófagos puede ser un cóctel de tres bacteriófagos virulentos, concretamente LF82_P2, LF82_P6 y LF82_P8, que pertenecen a la familia

Myoviridae, el cual se describió como particularmente eficaz en la eliminación de AIEC en el intestino de ratones infectados (Galtier M. *et al.*, 2017).

5 Las estrategias que se dirigen a bloquear la adhesión bacteriana pueden promover la descolonización bacteriana sin alterar la microbiota intestinal. Se espera que este enfoque terapéutico ejerza una débil presión selectiva y por tanto que limite la aparición de resistencia bacteriana.

10 Por una parte, puede obtenerse una reducción de la adhesión de bacterias AIEC mediante estrategias probióticas que comprenden la administración de levaduras, ya que las levaduras tienen numerosas estructuras manosiladas en su pared celular que pueden interaccionar con bacterias de *E. coli*. También se ha descrito que las levaduras protegen la mucosa intestinal de lesiones inducidas por bacterias AIEC y que impiden la alteración de la barrera. En una realización particular, dicha terapia puede comprender la administración de 15 *Saccharomyces cerevisiae*. Por ejemplo, ha sido descrito que la cepa de *Saccharomyces cerevisiae* CNCM I-3856 es capaz de inhibir la adhesión de AIEC a células epiteliales intestinales en cultivo y a enterocitos ileales aislados de pacientes con CD (Sivignon *et al.* 2015).

20 Otras estrategias dirigidas a inhibir la adhesión de bacterias AIEC se basan en antagonistas de FimH (por ejemplo, manósidos como análogos de receptor). Ha sido descrito que la adhesión de AIEC a células del huésped está mediada principalmente por fimbrias de tipo 1, que son orgánulos con forma de varilla cubiertos por la adhesina FimH, que se une a glucoproteínas manosiladas del huésped. Se ha descrito que las cepas AIEC se adhieren 25 preferiblemente al borde en cepillo de los enterocitos ileales de pacientes con CD en comparación con enterocitos de individuos control, de manera dependiente de fimbrias de tipo 1 (Barnich *et al.* 2007). Se ha desarrollado una amplia variedad de antagonistas de FimH que bloquean el dominio de reconocimiento de hidratos de carbono (CRD) de FimH, con el fin de impedir que las bacterias se unan a las células epiteliales intestinales y por 30 tanto estas se eliminen del intestino (Sivignon *et al.* 2017).

Se han descrito antagonistas de FimH con estructuras químicas variadas como agentes anti-adhesivos potenciales contra UTI promovidas por *E. coli* (Gouin *et al.*, 2014; Mydock-McGrane *et al.*, 2016; Grabosch *et al.*, 2011 y Firon *et al.*, 1987). Estos incluyen, pero no se 35 limitan a, derivados monovalentes de heptil-manosa, tiazolilamino-manósidos monovalentes (TazMan) y glucopolímeros basados en HM (por ejemplo basados en N-7-(α -D-

manopirosiloxio)heptil-metacrilamida), y pueden usarse en los métodos de la presente invención. En los métodos de la invención también puede usarse la estrategia multivalente en la que se injertan ligandos de FimH en múltiples copias en una estructura principal común (Bernardi *et al.*, 2013).

5

Además, las estrategias dirigidas a eliminar AIEC pueden combinarse con tratamientos usados tradicionalmente para tratar terapéuticamente la CD. El tratamiento más apropiado se seleccionará generalmente según la localización, gravedad y actividad de la enfermedad. Las farmacoterapias comunes usadas actualmente son productos químicos antiinflamatorios

10

derivados de ácido salicílico (es decir mesalazina y sulfasalazina), corticosteroides (es decir prednisona, metil-prednisona y budesonida), antibióticos (es decir metronidazol y ciprofloxacino), inmunosupresores (es decir azatioprina y mercaptopurina), el antimetabolito y antifolato metotrexato, y los denominados fármacos “biológicos” que consisten en anticuerpos contra el factor de necrosis tumoral α (TNF α), tal como infliximab, adalimumab,

15

cetolizumab pegol, etanercept y golimumab. También está indicada la resección intestinal en aquellos pacientes con CD fulminante o con fístulas y para aquellos pacientes que no responden a ninguna de las medicaciones mencionadas anteriormente (casos resistentes). También se contemplan otras terapias tales como trasplante de células madre hematopoyéticas autólogas, o el uso de probióticos y/o prebióticos.

20

En un aspecto relacionado, la invención se refiere a compuestos que eliminan o disminuyen los niveles de bacterias AIEC para su uso en un método de tratamiento de una enfermedad intestinal en un sujeto portador de AIEC que lo necesita, en el que dicho sujeto se ha identificado mediante el uso de un método de clasificación de la invención tal como se describe en el presente documento. Las características y realizaciones preferidas son tal

25

como se describió anteriormente en el presente documento.

Los métodos de la presente invención podrían implementarse por un ordenador. Por tanto, un aspecto adicional de la invención se refiere a un método implementado por ordenador,

30

en el que el método es cualquiera de los métodos divulgados en el presente documento o cualquier combinación de los mismos.

Se indica que también forma parte de la presente invención cualquier programa informático capaz de implementar cualquiera de los métodos de la presente invención o de usarse para

35

implementar cualquiera de estos métodos o cualquier combinación de los mismos.

También se indica que se incluye cualquier dispositivo o aparato que comprenda medios para llevar a cabo las etapas de cualquiera de los métodos de la presente invención o cualquier combinación de los mismos, o que lleve un programa informático capaz de, o para
5 implementar cualquiera de los métodos de la presente invención o cualquier combinación de los mismos, formando parte de la presente memoria descriptiva.

Los métodos de la invención también pueden comprender el almacenamiento de los resultados del método en un soporte de datos, preferiblemente en el que dicho soporte de
10 datos es un medio legible por ordenador. La presente invención se refiere además a un medio de almacenamiento legible por ordenador que tiene almacenado en el mismo un programa informático de la invención o los resultados de cualquiera de los métodos de la invención.

Tal como se usa en el presente documento, “un medio legible por ordenador” puede ser cualquier aparato que puede incluir almacenar, comunicar, propagar o transportar los resultados de la determinación del método de la invención. El medio puede ser un sistema (o aparato o dispositivo) electrónico, magnético, óptico, electromagnético, por infrarrojos o semiconductor o un medio de propagación.

20 Kit y uso de un kit

Todavía en un aspecto adicional, la presente invención se refiere a un kit para determinar el genotipo de los SNP diana en el genoma de una cepa de *E. coli* y a su uso en un método tal como se describe en el presente documento. El kit también puede contener instrucciones
25 que indiquen cómo pueden usarse los materiales dentro del kit.

El término “kit” o “kit de pruebas” indica combinaciones de reactivos y adyuvantes requeridos para un análisis. Aunque un kit de prueba consiste en la mayoría de los casos en varias unidades, también se dispone de elementos de análisis de una pieza, que deben
30 considerarse asimismo como kit de pruebas.

En una realización particular, dicho kit es adecuado para determinar en el genoma de una cepa de *E. coli*, el genotipo de uno, dos, tres, cuatro o más SNP individuales seleccionados del grupo que consiste en: E1-E2_3.6, E3-E4_4.3(1), E3-E4_4.3(2), E3-E4_4.3(3), E3-
35 E4_4.4, E5-E6_3.12, y de E5-E6_3.16=3.22(1) a E5-E6_3.16=3.22(6), en el que dicho kit

comprende uno, dos, tres, cuatro o más de los siguientes:

- un oligonucleótido específico para una secuencia diana que comprende el SNP E1-E2_3.6;
- 5** – un oligonucleótido específico para una secuencia diana que comprende el SNP E3-E4_4.3(1);
- un oligonucleótido específico para una secuencia diana que comprende el SNP E3-E4_4.3(2);
- un oligonucleótido específico para una secuencia diana que comprende el SNP E3-E4_4.3(3);
- 10** – un oligonucleótido específico para una secuencia diana que comprende el SNP E3-E4_4.4;
- un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.12;
- un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(1);
- 15** – un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(2);
- un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(3);
- 20** – un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(4);
- un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(5);
- un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(6); y
- 25** – opcionalmente, que comprende además instrucciones para el uso de dichos reactivos en un método para determinar el genotipo del/de los sitio(s) polimórfico(s) diana.
- 30** En una realización preferida, dicho kit comprende al menos uno, dos, tres o cuatro de los siguientes:
 - un oligonucleótido específico para una secuencia diana que comprende el SNP E3-E4_4.3(2);

- un oligonucleótido específico para una secuencia diana que comprende el SNP E3-E4_4.4;

- un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(2);

5 – un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(3); y

- opcionalmente, que comprende además instrucciones para el uso de dichos reactivos en un método para determinar el genotipo del/de los sitio(s) polimórfico(s) diana.

10

En otra realización particular, dicho kit es adecuado para determinar en el genoma de una cepa de *E. coli*, al menos el genotipo en el sitio polimórfico E3-E4_4.4, y opcionalmente para determinar el genotipo de los sitios polimórficos E5-E6_3.16=3.22(2) y E5-E6_3.12, en el que dicho kit comprende:

15 – un oligonucleótido específico para una secuencia diana que comprende el SNP E3-E4_4.4;

- opcionalmente, que comprende además un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(2);

- opcionalmente, que comprende además un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.12; y

20

- opcionalmente, que comprende además instrucciones para el uso de dichos reactivos en un método para determinar el genotipo del/de los sitio(s) polimórfico(s) diana.

25 En una realización particular adicional, dicho kit es adecuado para determinar en un genoma de la cepa de *E. coli*, al menos el genotipo en el sitio polimórfico E6_3.16=3.22(2), en el que dicho kit comprende:

- un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(2);

30 – opcionalmente, que comprende además un oligonucleótido específico para una secuencia diana que comprende el SNP E3-E4_4.4;

- opcionalmente, que comprende además un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.12; y

– opcionalmente, que comprende además instrucciones para el uso de dichos reactivos en un método para determinar el genotipo del/de los sitio(s) polimórfico(s) diana.

5 En una realización adicional, opcionalmente en combinación con una o más de las características o realizaciones descritas anteriormente o a continuación, este kit puede comprender además uno o más oligonucleótidos específicos para una secuencia diana que comprende uno de los otros SNP no específicos de cepa confirmados descritos en el presente documento.

10

En una realización particular, opcionalmente en combinación con una o más de las características o realizaciones descritas anteriormente o a continuación, dicho(s) oligonucleótido(s) específico(s) para una secuencia diana híbrida(n) con una secuencia que comprende la forma alélica diana del SNP, tal como se describió anteriormente en el

15 presente documento.

En otra realización particular, opcionalmente en combinación con una o más de las características o realizaciones descritas anteriormente o a continuación, dicho(s) oligonucleótido(s) específico(s) para una secuencia diana amplifica(n) una secuencia que

20 comprende el SNP diana. En determinadas realizaciones, los mismos oligonucleótidos se usan para amplificar una secuencia que comprende otros SNP que están ubicados dentro del mismo gen. Por consiguiente, por ejemplo, pueden diseñarse secuencias de oligonucleótidos para amplificar una región del gen E3-E4_4.3 que comprende los SNP E3-E4_4.3(1), E3-E4_4.3(2) y E3-E4_4.3(3) (véanse las secuencias SEQ ID NO:11 y SEQ ID

25 NO:12 en la Tabla 13, que se usaron en los ejemplos como cebadores *forward* y *reverse*, respectivamente para la amplificación del gen E3-E4_4.3).

Preferiblemente, dichos oligonucleótidos comprenden o consisten en una secuencia seleccionada del grupo que consiste en los oligonucleótidos enumerados en la Tabla 13 o

30 una secuencia con una identidad de al menos aproximadamente el 90%, preferiblemente al menos aproximadamente el 95%, el 96%, el 97%, el 98% o el 99%, para las secuencias diana respectivas.

Por consiguiente, en una realización preferida, un oligonucleótido específico para una

35 secuencia diana que comprende el SNP E3-E4_4.4 comprende o consiste en SEQ ID

NO:13, SEQ ID NO:14, o una secuencia con una identidad de al menos aproximadamente el 90%, preferiblemente al menos aproximadamente el 95%, el 96%, el 97%, el 98% o el 99%.

5 En otra realización preferida, un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(2) comprende o consiste en SEQ ID NO:23, SEQ ID NO:24, o una secuencia con una identidad de al menos aproximadamente el 90%, preferiblemente al menos aproximadamente el 95%, el 96%, el 97%, el 98% o el 99%.

10 En una realización preferida adicional, un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(2) comprende o consiste en SEQ ID NO:30 (GCTGAACCATTCATTCACG), SEQ ID NO:31 (TTATTGCAGAAAAGCGAGAGG), o una secuencia con una identidad de al menos aproximadamente el 90%, preferiblemente al menos aproximadamente el 95%, el 96%, el 97%, el 98% o el 99%.

15

Todavía en otra realización preferida, un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.12 comprende o consiste en SEQ ID NO:21, SEQ ID NO:22 o una secuencia con una identidad de al menos aproximadamente el 90%, preferiblemente al menos aproximadamente el 95%, el 96%, el 97%, el 98% o el 99%.

20

En una realización particular, opcionalmente en combinación con una o más de las realizaciones descritas anteriormente o a continuación, dicho kit comprende reactivos para realizar una reacción de PCR, que normalmente contiene una ADN polimerasa, tal como Taq ADN polimerasa (por ejemplo, Taq ADN polimerasa *hot start*), tampón, magnesio, dNTP y opcionalmente otros agentes (por ejemplo, agentes estabilizantes tales como gelatina y albúmina sérica bovina). Además, mezclas de reactivos para llevar a cabo la reacción de PCR cuantitativa o en tiempo real también contienen reactivos para la detección en tiempo real y la cuantificación de productos de amplificación, que se conocen bien en la técnica.

30

Opcionalmente, el kit también puede incluir tubos y disolventes apropiados para la extracción de ADN, por ejemplo, una disolución de cloroformo/metanol. Otras características y realizaciones preferidas del kit de la invención son tal como se describe en el presente documento a lo largo de toda la memoria descriptiva.

35

Se contempla que cualquier característica descrita en el presente documento puede combinarse opcionalmente con cualquiera de las realizaciones de cualquier aspecto de la invención, por ejemplo, cualquier método, uso médico, método de tratamiento, kit de partes y uso de un kit de la invención; y cualquier realización comentada en esta memoria descriptiva puede implementarse con respecto a cualquiera de ellos. Se entenderá que las realizaciones particulares descritas en el presente documento se muestran a modo de ilustración y no como limitaciones de la invención. Las principales características de esta invención pueden emplearse en diversas realizaciones sin apartarse del alcance de la invención. Los expertos en la técnica reconocerán o podrán determinar mediante el uso de únicamente experimentación de rutina, numerosos equivalentes a los procedimientos específicos descritos en el presente documento. Se considera que tales equivalentes están dentro del alcance de esta invención y quedan cubiertos por las reivindicaciones.

Todas las publicaciones y solicitudes de patente se incorporan en el presente documento como referencia en el mismo grado que si se indicara que cada publicación o solicitud de patente individual se incorpora como referencia.

El uso de la palabra “un” o “una” puede significar “uno”, pero también es compatible con el significado de “uno o más”, “al menos uno” y “uno o más de uno”. El uso del término “otro” también puede referirse a uno o más. El uso del término “o” en las reivindicaciones se usa para significar “y/o” al menos que se indique explícitamente que se refiere únicamente a alternativas o que las alternativas son excluyentes entre sí.

Tal como se usa en esta memoria descriptiva y en la reivindicación/las reivindicaciones, las palabras “que comprende” (y cualquier forma de que comprende, tal como “comprenden” y “comprende”), “que tiene” (y cualquier forma de que tiene, tal como “tienen” y “tiene”), “que incluye” (y cualquier forma de que incluye, tal como “incluye” e “incluyen”) o “que contiene” (y cualquier forma de que contiene, tal como “contiene” y “contienen”) son inclusivas o abiertas y no excluyen elementos o etapas de método adicionales, no citados. El término “comprende” también engloba y divulga expresamente los términos “consiste en” y “consiste esencialmente en”. Tal como se usa en el presente documento, la expresión “que consiste esencialmente en” limita el alcance de una reivindicación a las etapas o materiales especificados y a aquello que no afecta materialmente a la(s) característica(s) básica(s) y nueva(s) de la invención reivindicada. Tal como se usa en el presente documento, la expresión “que consiste en” excluye cualquier elemento, etapa o componente no

especificado en la reivindicación excepto por, por ejemplo, las impurezas asociadas habitualmente con el elemento o limitación.

El término “o combinaciones de los mismos” tal como se usa en el presente documento se refiere a todas las permutaciones y combinaciones de los elementos enumerados que preceden al término. Por ejemplo, se pretende que “A, B, C, o combinaciones de los mismos” incluya al menos uno de: A, B, C, AB, AC, BC o ABC, y si el orden es importante en un contexto particular, también BA, CA, CB, CBA, BCA, ACB, BAC o CAB. Continuando con este ejemplo, se incluyen expresamente combinaciones que contienen repeticiones de uno o más elementos o términos, tales como BB, AAA, AB, BBC, AAABCCCC, CBBAAA, CABABB, etc. El experto en la técnica entenderá que normalmente no hay límite en el número de elementos o términos en cualquier combinación, a menos que resulte evidente otra cosa del contexto.

Tal como se usa en el presente documento, las palabras de aproximación tales como, “sin limitación”, “alrededor de”, “en torno a”, “aproximadamente” se refieren a una condición que cuando se modifica de ese modo se entiende que no es necesariamente absoluta o perfecta, sino que debe considerarse suficientemente próxima como para que los expertos habituales en la técnica justifiquen que designa que la condición está presente. Puede establecerse el grado en que la descripción podrá variar dependiendo de lo grande que sea un cambio y todavía un experto habitual en la técnica reconocerá que la característica modificada tiene todavía las características y capacidades requeridas de la característica no modificada. En general, pero sujeto al comentario anterior, un valor numérico en el presente documento que se modifica por una palabra de aproximación tal como “aproximadamente” puede variar con respecto al valor establecido en $\pm$ el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 11, el 12, el 13, el 14 o el 15%. Por consiguiente, el término “aproximadamente” puede significar el valor indicado $\pm$ el 5% de su valor, preferiblemente el valor indicado $\pm$ el 2% de su valor, lo más preferiblemente el término “aproximadamente” significa exactamente el valor indicado ($\pm$ el 0%).

Los siguientes ejemplos sirven para ilustrar la presente invención y no deben considerarse limitativos del alcance de la misma.

EJEMPLOS

Ejemplo 1.- Material y métodos**Selección y caracterización de cepas**

Se seleccionaron tres pares de cepas de *E. coli* aisladas en un estudio anterior (Martinez-Medina, M. *et al*, 2009), que consistía cada par en una AIEC y una no-AIEC de pulsotipo idéntico y que pertenecían a un grupo filogenético distinto (B1, B2 o D). El estudio del que se obtuvieron las cepas de AIEC se aprobó el 22 de mayo de 2006 por el Comité Ético de Investigación Clínica del Hospital Josep Trueta de Girona. La información sobre los pacientes de los que se aislaron las cepas se presenta en la Tabla 12. En esta tabla también se muestra el origen de las cepas del conjunto usado para validar la utilidad de las diferencias genéticas encontradas entre pares de cepas como posibles marcadores moleculares.

Tabla 12. Información sobre los pacientes de los que se aislaron las cepas estudiadas. Los pares de cepas de AIEC y no-AIEC para el estudio genómico comparativo se indican en negrita.

Id paciente	Edad	Sexo	Estado patológico; origen de la muestra	Años desde el primer diagnóstico	Tratamiento	Cepas
53	33	M	C-CD; colon	nd	nd	ECG01, ECG02, AIEC25
55	42	F	CD, inactivo; colon	12	mesalazina	ECG18
61	31	F	IC-CD, activo; colon	0	ninguno	ECG09, ECG65
69	67	M	Control (rectorragia, antecedentes familiares de cáncer colorrectal y poliposis); colon	na	na	ECG41
70	83	M	Control (diverticulosis); colon	na	na	ECG22
71	67	M	Control (micropólipos); colon	na	na	ECG12
72	27	F	C-CD, activo; colon	0	ninguno	ECG05, ECG64
77	48	F	I-CD, activo; colon	1	6-mercaptopurina	AIEC21
79	25	F	I-CD, inactivo; ileon	0	mesalazina	ECG19, ECG42
80	70	M	Control (estreñimiento, síndrome del intestino irritable); colon	na	na	ECG43
81	21	F	Control (diarrea y antecedentes familiares de cáncer colorrectal); colon	na	na	ECG46
82	29	F	Control (rectorragia); colon	na	na	ECG16
83	61	F	Control (rectorragia); colon	na	na	ECG17
84	59	F	Control (antecedentes familiares de cáncer colorrectal); colon	na	na	ECG49
89	31	M	IC-CD, activo; colon	0	ninguno	AIEC12
109	21	M	I-CD, inactivo; colon	8	medicina alternativa	ECG21, ECG34
110	37	F	IC-CD, inactivo; colon	18	ninguno	ECG63, ECG57, AIEC20

111	49	F	I-CD, inactivo; colon	15	azatioprina	ECG28, AIEC17
112	35	F	C-CD, inactivo; colon	19	azatioprina	ECG26
113	43	F	I-CD, inactivo; colon	22	ninguno	ECG15
114	31	F	IC-CD; colon	>2	nd	AIEC02, AIEC05
119	35	M	Control (rectorragia y antecedentes familiares de cáncer colorrectal); colon	na	na	ECG23, AIEC19
120	46	F	I-CD, inactivo; íleon	13	azatioprina	ECG11 , AIEC01
122	31	M	IC-CD, inactivo, fístula perianal; colon	10	azatioprina	AIEC09, AIEC24
123	36	M	C-CD, inactivo; íleon	7	ninguno	AIEC23
124	37	M	Control (rectorragia); íleon	na	na	ECG04, ECG08, AIEC07
125	33	F	Control (rectorragia); íleon	na	na	AIEC04
126	34	M	Control (rectorragia); íleon	na	na	AIEC10
127	30	M	Control (rectorragia); colon	na	na	ECG13
128	37	M	Control (estreñimiento); colon	na	na	AIEC06
132	48	F	I-CD; íleon	9	nd	AIEC11, AIEC14, AIEC15.1, AIEC16.1
142	41	M	Control; colon	na	na	AIEC08

M: Hombre; F: Mujer; na: no aplicable; nd: sin datos; C-CD: enfermedad de Crohn colónica; I-CD: enfermedad de Crohn ileal; IC-CD: enfermedad de Crohn ileocolónica. Activo cuando CDAI>150. También se estudió la cepa LF82².

5

Las principales características de estas cepas se muestran en la Tabla 1. Los criterios de selección, el tipado de secuencias de múltiples loci (MLST) y el protocolo de PFGE se proporcionan a continuación en el presente documento.

10 Selección de cepas

Se seleccionaron tres cepas de AIEC aisladas en un estudio anterior (Martinez-Medina, M. *et al*, 2009) basándose en los criterios siguientes: i) poseer diferentes orígenes

filogenéticos, ii) mostrar altos índices de adhesión (>15,9 bacteria/célula I-407) e invasión (sobreviviendo > el 0,266% del inóculo tras 1 h de tratamiento con gentamicina) en células Intestine-407 (I-407; ATCC CCL-6) y iii) poseer un genotipo similar a ExPEC. Las principales características de estas cepas se muestran en la Tabla 1. Las cepas no-AIEC con pulsotipos idénticos a los de cada una de las AIEC seleccionadas se buscaron en el conjunto de *E. coli* obtenido para cada paciente en el mismo estudio anterior (Martínez-Medina, M. *et al*, 2009).

Tipado de secuencias de múltiples loci

Se realizó tipado de secuencias de múltiples loci (MLST) *in silico* consultando para cada genoma, las secuencias de 7 genes *housekeeping* (*adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA* y *recA*) extraídos de la base de datos MLST de *E. coli* (Universidad de Warwick). Cada alelo se identifica mediante un marcador numérico. Se recogió la combinación de los 7 marcadores numéricos para cada cepa y se usó para obtener los tipos de secuencia (ST) (Tabla 1).

Electroforesis en gel de campo pulsante

Se comprobó si las cepas pertenecían al mismo clon mediante electroforesis en gel de campo pulsante (PFGE) tal como se describe en CDCPulseNetUSA, 2004. Se digirió el ADN, previamente incluido en bloques de agarosa, con XbaI 0,2 U/μl (Takara Bio) según las instrucciones del fabricante. Se analizó el ADN genómico digerido por XbaI en un gel de agarosa al 1% en tampón Tris-ácido borónico-EDTA 0,5X a 14°C usando el sistema CHEF-DR III (Bio-Rad). Se corrió el gel durante 19 h a 6 V/cm, con tiempos de conmutación inicial y final de 2,2 s y 54,2 s, respectivamente. Se tiñó el gel con bromuro de etidio (1 μg/ml) y se normalizaron y calibraron las imágenes en formato TIFF usando el software GelComparII (Applied Maths). Se crearon dendogramas basados en curva densitométricas obtenida de la imagen del gel, usando coeficientes de correlación de Pearson, aplicando una optimización del 0,5% y un suavizado de curva del 0,5% y el método de agrupación UPGMA. El dendograma de las cepas se muestra en la Figura 2.

Caracterización fenotípica de las cepas seleccionadas

Se realizó caracterización fenotípica de las cepas seleccionadas para determinar la adhesión e invasión de la línea de células epiteliales Intestine-407 (I-407 (ATCC CCL-6)) así como la supervivencia y la replicación dentro de dos líneas celulares de macrófagos (J774A.1 (ATCC TIB-67) y THP-1 (ATCC TIB-202) de ratón y humano, respectivamente) tal como se detalla a continuación en el presente documento. Las cepas con un índice de

adhesión ≥ 1 , un índice de invasión $\geq 0,1$ que se redujeron en del 90% al 99,9% cuando se añadieron el inhibidor de microfilamentos citocalasina D (C8273, Sigma-Aldrich) y el inhibidor de microtúbulos colchicina (C9754, Sigma-Aldrich), y con un índice de replicación $\geq 100\%$ en J774 y THP-1 se clasificaron como cepas de AIEC en el presente estudio.

5

Estudios de adhesión e invasión usando células epiteliales Intestine-407

Se usó la línea de células epiteliales Intestine-407 (ATCC CCL-6) para los ensayos de adhesión e invasión. Se realizaron ensayos de cultivo celular, adhesión e invasión por triplicado tal como se describió anteriormente (Boudeau, J. *et al*, 2009). Brevemente, se infectaron dos placas de 24 pocillos que contenían 4×10^5 células/pocillo que se incubaron durante 20 h a una multiplicidad de infección de 10. Se incubaron placas por duplicado, una para el ensayo de adhesión y una para el ensayo de invasión, durante 3 h a 37°C en CO_2 al 5%.

10

15

Para los ensayos de adhesión bacteriana, se lavaron las monocapas de células cinco veces con solución salina tamponada con fosfato (PBS) y luego se lisaron con Triton X-100 al 1% (Sigma-Aldrich, St Louis, MO, EE.UU.). Se cuantificaron las bacterias adherentes sembrándolas en placa en agar Luria-Bertani (LB) (Liofilchem Srl, Italia). El sembrado en placas se realizó en un periodo máximo de 30 min para evitar una lisis bacteriana inducida por Triton X-100. Se determinó la capacidad de adhesión (I_{ADH}) mediante el cálculo del número medio de bacterias por célula. Los aislados se consideraron adherentes cuando $I_{\text{ADH}} \geq 1$.

20

25

Para los ensayos de invasión bacteriana, se lavaron las monocapas dos veces con PBS tras 3 h de infección, y se añadió medio de cultivo celular recién preparado que contenía gentamicina $100 \mu\text{g/ml}$ y se dejó durante 1 h para eliminar las bacterias extracelulares. Tras la lisis celular con Triton X-100 al 1%, se determinó el número de bacterias intracelulares mediante sembrado en placa. La capacidad invasiva se expresó como el porcentaje del inóculo inicial que llegó a ser intracelular: $I_{\text{INV}} (\%) = (\text{bacterias intracelulares} / 4 \times 10^6 \text{ bacteria inoculadas}) \times 100$. Los aislados se consideraron invasivos cuando $I_{\text{INV}} \geq 0,1\%$.

30

35

Se evaluó la implicación del citoesqueleto de las células huésped tal como se describe por Baumgart *et al*. (Baumgart, M. *et al*, 2007). Se sembraron células I-407 a una densidad de 4×10^5 células/pocillo; tras 24 horas, se incubaron las monocapas con citocalasina D ($0,5 \mu\text{g/ml}$) o colchicina ($1 \mu\text{g/ml}$) durante 30 minutos para despolimerizar microfilamentos y

microtúbulos, respectivamente. A continuación, se procedió con las monocapas de células I-407 tal y como se describe para los ensayos de invasión. Finalmente, se determinó el efecto inhibitor y se presentó como el porcentaje de reducción de los índices de invasión.

5 Supervivencia y replicación dentro de macrófagos

Se usaron la línea celular murina de macrófago J774A.1 (ATCC TIB-67) y la línea celular humana THP-1 (ATCC TIB-202) en ensayos de supervivencia y replicación de las seis cepas secuenciadas de *E. coli*. Se realizó el cultivo de las células J774 y se determinó la capacidad de los aislados de *E. coli* individuales para sobrevivir y replicarse dentro de los macrófagos tal como se describió anteriormente (Glasser, A. *et al*, 2001). Brevemente, se sembraron los macrófagos J774 a 2×10^5 células por pocillo en dos placas de 24 pocillos. Se incubaron las placas durante 20 h en medio completo (RPMI 1640 (Lonza, Suiza) complementado con FBS inactivado por calor al 10% (Gibco BRL) y L-glutamina al 1% (Gibco BRL)). Tras la incubación, se reemplazó el medio por medio recién preparado y se sembraron bacterias a una multiplicidad de infección de 10. Para promover la internalización de las bacterias por los macrófagos, se centrifugaron las placas a 900 rpm durante 10 min y se incubaron durante 10 min adicionales a 37°C en CO₂ al 5%. Las bacterias que no se fagocitaron se eliminaron mediante la inclusión de gentamicina (20 µg/ml) en el medio. La línea celular humana THP-1 (ATCC TIB-202) se mantuvo en medio RPMI 1640 (Lonza, Verviers, Bélgica) complementado con suero bovino fetal (Linus) al 10% (vol/vol). La incubación se realizó a 37°C y en una atmósfera que contenía CO₂ al 5%. Se sembraron las células THP-1 en dos placas de 24 pocillos a una densidad de 5×10^5 células/ml y se hicieron crecer en medio completo que contenía 20 ng/ml de 12-miristato-13-acetato de forbol (PMA; Sigma-Aldrich) durante 24 h para promover la diferenciación monocítica. Tras la incubación, se reemplazó el medio por medio recién preparado (RPMI + FBS inactivado por calor al 10%), y se sembraron bacterias a una multiplicidad de infección de 100. De manera similar a las monocapas de J774, se centrifugaron las placas de THP-1 a 900 rpm durante 10 min y se incubaron durante 10 min adicionales a 37°C en CO₂ al 5%. Se lavaron las monocapas de células dos veces con PBS y se añadió medio de cultivo celular recién preparado que contenía gentamicina 100 µg/ml para eliminar las bacterias extracelulares. Tras 40 min de incubación, se lavó una placa dos veces con PBS y se añadieron 0,5 ml de Triton X-100 al 1% (Sigma-Aldrich) a cada pocillo durante 5 min para lisar las células eucariotas. Para determinar el número de bacterias intracelulares recuperadas, se diluyeron las muestras y se sembraron en placa sobre placas de agar LB. Se reemplazó el medio de la segunda placa con medio de medio de cultivo celular recién preparado que contenía gentamicina

20 µg/ml y se incubó durante 23 h. A continuación, se lavó la monocapa y se trató con Triton X-100 al 1% y la suspensión celular se diluyó y se sembró en placa tal como se describió anteriormente.

- 5** Para ambas líneas celulares, se cuantificaron las bacterias intracelulares tal como se describió para el ensayo de invasión tras 1 y 24 h de infección. Los resultados se expresan como el porcentaje medio de bacteria recuperadas a 1 h y 24 h tras la infección: $I_REPL (\%) = (UFC \text{ ml}^{-1} \text{ a } 24 \text{ h} / UFC \text{ ml}^{-1} \text{ a } 1 \text{ h}) * 100$. Las cepas con I_ADH mayor de 1 bacteria/célula, I_INV del 0,1% e I_REPL del 100% o mayor se clasificaron como cepas de **10** AIEC en el presente estudio.

- Se examinó una selección de genes asociados previamente con el fenotipo de AIEC o bien debido a su mayor prevalencia en el patotipo (genes *lpfA*₁₅₄ (Dogan, B. *et al*, 2014; Chassaing, B. *et al*, 2011), *gipA* (Vazeille, E. *et al*, 2016), *chuA* (Dogan, B. *et al*, 2014), *fyuA* **15** (Spurbeck, R. R. *et al*, 2012), *afaC* (Prorok-Hamon, M. *et al*, 2014), *pduC* (Dogan, B. *et al*, 2014) e *ibeA* (Cieza, R. J. *et al*, 2015) o bien debido a la presencia de variantes de aminoácido relevantes para el patotipo (FimH (Iebba, V. *et al*, 2012), OmpA (Rolhion, N. *et al*, 2015) y ChiA (Low, D. *et al*, 2013)) usando la alineación con BLASTn (Altschul, S. F. *et al*, 1997) y con ClustalW (Thompson, J. D. *et al*, 1994).

20

Extracción y secuenciación de ADN genómico

- Se extrajo ADN genómico de células bacterianas cultivadas en medio Luria-Bertani (LB), usando el kit de purificación de ADN genómico Wizard® (Promega) según las instrucciones del fabricante; se trataron las muestras con ARNasa A facilitada con el kit. Se determinó la **25** pureza del ADN usando un espectrofotómetro NanoDrop ND-100 (NanoDrop Technologies), se midió la cantidad de ADN usando un fluorómetro Qubit® 2.0 (Life Technologies) y se examinó la integridad del ADN y la eliminación del ARN en geles de agarosa. Se identificaron bandas únicas de aproximadamente 23 Kb en los geles de agarosa. La proporción de absorbancias de 260/280 de las preparaciones de ADN osciló entre 1,8 y 2, y **30** la cantidad de ADN obtenida osciló entre 15 y 30 µg, lo que indica calidad suficiente del ADN genómico para la secuenciación del genoma. Se usaron dos plataformas de secuenciación: Illumina HiSeq y PacBio Biosciences tal como se describe a continuación en el presente documento.

Para la secuenciación con Illumina, se prepararon a partir de las muestras de ADN bibliotecas de secuenciación usando el kit Illumina TruSeq en EA Quintiles. Brevemente, se fragmentó 1 µg de ADN genómico para dar fragmentos de ~200 pb usando un ultrasonificador Covaris E210. Entonces se transformó en romo los extremos del ADN

5 fragmentado y se añadió una única “cola de A” al extremo 3’ de cada fragmento para facilitar la ligación de adaptadores de secuenciación que contenían una sola base T sobresaliente. Se amplificó el ADN ligado por adaptadores mediante la reacción en cadena de la polimerasa para aumentar la cantidad de ADN listo para secuenciación. Se analizaron las bibliotecas de ADN finales para determinar la distribución de tamaño y la calidad usando

10 un bioanalizador Agilent (kit DNA 1000, Agilent n.º 5067-1504), se cuantificaron usando Picogreen (Life Tech n.º P11496) y se normalizaron hasta una concentración de 2 nM. Se combinaron volúmenes iguales de las bibliotecas de ADN normalizadas y se usó el ADN combinado para preparar una célula de flujo usando el kit V3 de agrupación de extremos emparejados Illumina TruSeq (Illumina n.º PE-401-3001). Las combinaciones se

15 desnaturalizaron usando NaOH 0,1 N recién preparado y se diluyeron hasta 20 pM en tampón de hibridación previamente enfriado. Se diluyeron adicionalmente las combinaciones hasta 9 pM, y se colocó una alícuota de cada una en un instrumento Illumina cBot para producir agrupaciones a través de amplificación en puente. La secuenciación se realizó en un dispositivo Illumina HiSeq 2000 usando una secuenciación *paired-end* de 100

20 bases más un ciclo de índice de 7 bases.

Para la secuenciación con PacBio, se prepararon las bibliotecas de ADN siguiendo las directrices de PacBio y se secuenciaron en chips SMRT usando la tecnología de secuenciación Pacific Biosciences RS (Pacific Biosciences, Menlo Park, CA, EE.UU.) en EA

25 Quintiles. Se purificaron 10 µg de ADN genómico usando el kit de limpieza de ADN PowerClean® (MO BIO Laboratories) y luego se cortaron a 2 kb usando un instrumento Covaris® Adaptive Focused Acoustics. El ADN cortado se purificó usando perlas magnéticas y se verificó en un bioanalizador. La preparación de la biblioteca se realizó usando el kit 2.0 de preparación de ADN molde de Pacific Biosciences (3 Kb – 10 Kb). La

30 selección del tamaño y la purificación de la biblioteca se realizaron usando perlas AMPure 0,6X (Beckman-Coulter Genomics). Cada biblioteca se unió a C2 ADN polimerasa, fue cargada en un chip SMRT, y se observó la secuenciación usando dos películas de 45 min para cada chip. El análisis de calidad de los datos sin procesar se realizó con PRINSEQ.

35 Ensamblaje del genoma *de novo*

Los genomas propuestos se ensamblaron *de novo* (combinando ambas plataformas) usando el software SPAdes (ABL) (Bankevich, A. *et al*, 2012) y se anotaron usando el sistema de anotación de genoma bacteriano BG7 (Pareja-Tobes, P. *et al*, 2012). Para asignar la función génica, se escogió el resultado positivo con el menor E-value obtenido tras el análisis contra la base de datos UniProt. Los genomas obtenidos se han depositado en el archivo europeo de nucleótidos con los números de registro ERS1456453 (AIEC17), ERS1456454 (ECG28), ERS1456455 (AIEC01), ERS1456456 (ECG11), ERS1456457 (AIEC07) y ERS1456458 (ECG04).

10 Genómica comparativa de pares de cepas

Diferencias en estructura genómica y contenido génico

Se usó MAUVE 2.3 (Darling, A. E. *et al*, 2010) para identificar reordenamientos estructurales e inversiones en todo el genoma de la cepa. Se usó CRISPRFinder (<http://crispr.i2bc.paris-saclay.fr/>) (Grissa, I. *et al*, 2007) para estudiar repeticiones palindrómicas agrupadas y regularmente interespaciadas (CRISPR). Para encontrar diferencias de contenido génico, se realizó una comparación mediante BLASTP (punto de corte de E-value 1e-5) y agrupación de Markov (factor de inflación de 2,0) mediante ORTHOVENN (Wang, Y. *et al*, 2015) usando secuencias de proteínas. Se realizó una corrección con BLASTn local (Altschul, S. F. *et al*, 1997) para incluir genes ubicados al comienzo o al final de los cóntigos.

20

Polimorfismos de un solo nucleótido en regiones codificantes

Se usó el programa Harvest para una alineación de genoma central rápida (Treangen, T. J. *et al*, 2014) a fin de detectar SNP entre cepas del mismo par. Para maximizar la sensibilidad, se usó el genoma de AIEC mejor anotado en el momento del análisis, UM146 (NC_017632.1) (Krause, D. O. *et al*, 2011), como genoma de referencia. Para el propósito de este estudio, sólo se eligieron para su examen los SNP entre los pares de cepas de AIEC y no-AIEC que produjeron un cambio de aminoácidos no sinónimo en las regiones codificantes y que no estaban presentes en regiones altamente variables o en los extremos de los cóntigos. Estos SNP se denominan “SNP seleccionados”.

30

Los SNP seleccionados se validaron mediante una nueva secuenciación de Sanger. Usando Primer3, se diseñaron cebadores para flanquear al menos 100 pb en el sentido de 5' y en el sentido de 3' de la posición del SNP para lograr buena calidad de secuencia para asignar nucleótidos en la posición del SNP. Se tuvo cuidado en diseñar cebadores, dirigidos a regiones conservadas entre varias cepas de AIEC y no-AIEC. Todos los

35

cebadores se analizaron adicionalmente con NetPrimer para seleccionar los pares de cebadores óptimos. Los polimorfismos validados se denominan “SNP confirmados”. Los cebadores y las condiciones de PCR usados se presentan en la Tabla 13 a continuación.

Tabla 13. Cebadores y condiciones de PCR usados para amplificar fragmentos de los genes en los que se ubicaron los SNP confirmados.

Cepas	ID	SEQ ID NO	Cebador <i>forward</i>	SEQ ID NO	Cebador <i>reverse</i>	Temperatura de hibridación (°C)
AIEC17 frente a ECG28	E1-E2_3.4	SEQ ID NO: 1	TCCTCAATGAATCGCAGTCTC	SEQ ID NO: 2	TCAAAAGATTGCCCGCTTAC	57
	E1-E2_3.6	SEQ ID NO: 3	CTCATCAGCCGGACATACG	SEQ ID NO: 4	CACCTGTTTTCACTTTTATCTCTG	56
	E1-E2_3.7	SEQ ID NO: 5	GGTAACCCATTGGCCTTG	SEQ ID NO: 6	CAACACTTCGCTGACAAACG	57
	E1-E2_5	SEQ ID NO: 7	CGCTATAACGGCGAACTGAT	SEQ ID NO: 8	TCAGTGTCCGGTATCAAAA	56
AIEC01 frente a ECG11	E3-E4_4.2	SEQ ID NO: 9	GCCAGTAACCTTCGCCATT	SEQ ID NO: 10	TCAGGACAGCGACAAAAGC	57
	E3-E4_4.3	SEQ ID NO: 11	GTTTTCTCCTTTGCCGAACA	SEQ ID NO: 12	TGATGGTGATAATGCTGCTCA	57
	E3-E4_4.4	SEQ ID NO: 13	ATATTACGCTGTCCGCAAT	SEQ ID NO: 14	CGCATCATCACTTCCATCTG	57
	E3-E4_4.5	SEQ ID NO: 15	GCGTTGCCTGATGATACTGA	SEQ ID NO: 16	CGTCGGGGACATCTGACTTA	57
AIEC07 frente a ECG04	E3-E4_4.7	SEQ ID NO: 17	GGAAGAGCTGGAGACAATGC	SEQ ID NO: 18	CACTACCGCCACTCTCCTGT	57
	E5-E6_3.1	SEQ ID NO: 19	CCCTGTTTGTGTACTGCTG	SEQ ID NO: 20	CTGCTCACAGGGCGTCAAATA	56
	E5-E6_3.12	SEQ ID NO: 21	GAAAAAGTCGCCCATGAGAC	SEQ ID NO: 22	CGCAACACCCAGAGGGTTAAT	57
	E5-E6_3.16=3.22	SEQ ID NO: 23	CATCACTTCGGTCAACAC	SEQ ID NO: 24	ATTGCAGAAAAGCGAGAGGT	56
		SEQ ID NO: 30	GCTGAACCATTCATTACG	SEQ ID NO: 31	TTATTGCAGAAAAGCGAGAGG	54
	E5-E6_3.17	SEQ ID NO: 25	TTTTACWCGAAGTCTGATG	SEQ ID NO: 26	GATGTGCTGCTGTCTGYTT	56

Programa de PCR: 1 ciclo a 95°C durante 5 minutos, 30 ciclos de 15 segundos a 95°C y 45 segundos a la temperatura de hibridación del cebador, finalmente un ciclo a 72°C durante 10 minutos. Todos los cebadores se usaron a 0,2 µM; tampón II de PCR a 1x; MgCl₂ a 1,5 mM; dNTP a 200 µM y AmpliTaq Gold polimerasa 1,25 unidades/reacción.

Debe indicarse que se encontraron algunos SNP con picos de nucleótidos ambiguos (denominados “SNP con picos solapantes”) que representan una mezcla de dos nucleótidos en una posición dada. Se analizó la posible causa de estos picos de nucleótidos ambiguos *in silico* usando una combinación de búsquedas de genes en BLASTn (Altschul, S. F. *et al*, 1997) e inspección de lecturas a través de Tablet (Milne, I. *et al*, 2013). La siguiente etapa en la selección de SNP fue la identificación *in silico* de SNP específicos de cepa; estos SNPs se descartaron del análisis.

Para determinar la especificidad de cepa de los “SNP confirmados”, se alinearon las secuencias que contenían la posición variable de las seis cepas secuenciadas en este estudio así como en 3 cepas de AIEC (UM146, LF82, NRG857c), 9 de ExPEC (CFT073, 536, UMN026, S88, APEC01, 042, EDL933, O157 Sakai, E24377A) y 3 comensales (HS, K12 MG1655, ED1a) y se determinó adicionalmente la distribución de SNP entre las cepas. Se analizó la distribución dentro del conjunto de cepas de SNP que presentaban variabilidad en la base en estudio entre las cepas y que se producían dentro de genes que se distribuían ampliamente entre la mayoría de las cepas. Los genes que albergaban “SNP confirmados” se clasificaron usando el consorcio de ontología génica (Gene Ontology Consortium, 2015) y las bases de datos Pfam (Finn, R. D. *et al*, 2016).

Distribución de SNP entre un conjunto de cepas

Para analizar la posible aplicación de los “SNP confirmados” como biomarcadores para la identificación específica de AIEC, se examinó un total de 16 SNP presentes en 9 genes en un conjunto de cepas más amplio de 22 cepas de AIEC y 28 no-AIEC (Martinez-Medina, M. *et al*, 2009) (Tablas 7 y 10) mediante secuenciación de Sanger.

Análisis estadístico

Se calcularon las diferencias en la distribución de nucleótidos presentes en cada sitio polimórfico entre patotipo y grupos filogenéticos usando la prueba de X^2 . Se usó la prueba no paramétrica de Kruskal-Wallis para comparar los índices medios de adhesión e invasión entre más de dos variantes de nucleótido, y se realizó la prueba de la U de Mann-Whitney para analizar comparaciones dos a dos. Se empleó un análisis de regresión logística binaria como modelo para predecir el patotipo AIEC según el nucleótido presente en una posición de SNP particular. Un valor de $p \leq 0,05$ se consideró estadísticamente significativo en todos los casos.

Ejemplo 2.- Caracterización de los tres pares de cepas seleccionados

5 En el presente estudio, se seleccionaron tres cepas de AIEC y tres homólogas no-AIEC correspondientes con las que las cepas de AIEC compartían patrones de PFGE (Figura 1), secuencias tipo (ST), orígenes filogenéticos y genes de virulencia idénticos (Tabla 1) para la secuenciación de sus genomas. Tal como se esperaba, las cepas de AIEC presentaban mayores índices de adhesión e invasión que sus homólogas no-AIEC. Sin embargo, las cepas no-AIEC también tenían la capacidad de sobrevivir y replicarse dentro de macrófagos murinos J774 pero no en macrófagos humanos THP-1, con la excepción de la cepa ECG28.

Tabla 1. Características de los tres pares de cepas de AIEC/no-AIEC secuenciadas.

Cepa	Grupo filogenético	Serotipo	ST ^a	Genes de virulencia	Índice de adhesión ^{b,c}	Índice de invasión ^{c,d}	Índice de replicación intramacrófagos en J774 ^{g,h}	Índice de replicación intramacrófagos en THP-1 ^{h,i}
AIEC17	D	ONT:HNT	569	<i>fimC, mat, ompA, eal, sitA, sitD_ch, irp2, fyuA, chuA, vat, ibeA, kpsMTII, neuC, traT, csgA, fimH, chiA, gipA, pduC</i>	21,6±17,5	0,266±0,055	1053±75	213±60
ECG28					0,6±0,3	0,005±0,001	774±129	228±68
AIEC01	B2	O6:H1	73	<i>focG, mat, iha, ompA, pic, sitA, sitD_ch, irp2, iucD, iutA, fyuA, vat, pks, kpsMTII, traT, fimH, chiA, gipA, chuA</i>	15,9±9,3	0,284±0,106	1567±1060	173±99
ECG11					1,1±0,8	0,004±0,002	716±315	74±29
AIEC07	B1	O22:H7	3232	<i>fimC, mat, csgA, ompA, fimH, lpfA₁₅₄, gipA, chuA, fyuA</i>	20,0±13,4	0,565±0,392	1693±297	189±71
ECG04					1,8±0,7	0,036±0,029	527±194	77±17

^a Tipo de secuencia. ^b Número de bacterias por célula I-407. ^c Porcentaje de bacterias intracelulares tras 1 h de tratamiento con gentamicina en relación con el inóculo. Los porcentajes de reducción de invasión para AIEC17, AIEC01 y AIEC07 fueron del 99,8%, el 99,4% y el 99,8%, respectivamente, en presencia de citocalasina D y del 90,4%, el 99% y el 95%, respectivamente, en presencia de colchicina. ^d Porcentaje de bacterias intracelulares presentes a las 24 h tras la infección en relación con el número de bacterias intracelulares presentes tras 1 h de tratamiento con gentamicina. Resultados de las cepas LF82 y K-12 como cepas de control, respectivamente: ^e 25,66±15,7 y 0,70±0,02. ^f 2,26±1,349 y 0,019±0,020. ^g 777±304,8 y 11±5. ^h 121±59 y 10±7.

Se obtuvieron longitudes similares de genoma entre cepas, y que oscilaban entre 4,825 y 5,213 Kb para las cepas de AIEC y no-AIEC (Tabla 2), y no se encontraron diferencias estructurales significativas entre los pares de cepas (Figura 2). Se detectó una inversión en todas las comparaciones dos a dos, pero sin ser compartida entre los tres pares de cepas.

5

Tabla 2. Características de ensamblaje de genomas secuenciados de AIEC/no-AIEC.

Cepa	Tamaño (kb)	Cóntigos (n.º)	GC (%) ^a	N50 (kb) ^b	N.º de registro
AIEC17	4,958	400	50,42	247	ERS1456453
ECG28	4,981	464	50,38	214	ERS1456454
AIEC01	5,213	333	50,48	186	ERS1456455
ECG11 ^c	5,212	25	50,56	555	ERS1456456
AIEC07	4,825	374	50,62	187	ERS1456457
ECG04	5,013	836	50,46	261	ERS1456458

^a GC(%), contenido definido como (G+C)/(A+T+G+C)×100.

^b N50, la longitud del cóntigo más corto al 50% del ensamblaje.

^c Esta cepa se secuenció mediante PacBio (biblioteca con inserto de 10 kb) y se ensambló con una herramienta HGAP 3.

10

Se buscaron en los genomas de las seis cepas aquellos genes que se habían descrito previamente que estaban asociados con AIEC. Se encontró que *gipA*, *chuA* y *fyuA* estaban presentes en las tres cepas de AIEC pero también en sus homólogas no-AIEC (Tabla 1).

15

Otros genes no estaban presentes en ninguna cepa (*afaC*) o estaban presentes sólo en un solo par de cepas (*lpfA*₁₅₄, *pduC* e *ibeA*).

Se encontró una situación similar con respecto a las variantes génicas de *fimH*, *ompA* y *chiA*. Aunque se han asociado cambios específicos en las secuencias de aminoácidos de proteínas codificadas por estos genes con una mayor capacidad de adhesión/invasión (Dreux, N. *et al*, 2013; Rolhion, N. *et al*, 2010; Low, D. *et al*, 2013), no se observaron tales diferencias entre los aislados de los pares de cepas (Tabla 3). También se realizó análisis de CRISPR, y se obtuvieron perfiles iguales para los pares AIEC17-ECG28 y AIEC07-ECG04; se reconocieron CRISPR no confirmadas en AIEC01 o en ECG11. Por tanto, el

20

patotipo de las cepas no estaría determinado por diferencias en las secuencias de estos genes .

Tabla 3. Distribución de sustituciones de aminoácidos en tres genes asociados previamente con patogénesis de AIEC en cepas de AIEC/no-AIEC. Las sustituciones de aminoácido que se asociaron previamente con AIEC están marcadas en negrita. La primera letra corresponde al aminoácido presente en la cepa estudiada; la última letra indica el aminoácido encontrado en la cepa comensal K-12.

Genes asociados con AIEC	Cepas comparadas		
	AIEC17-ECG28	AIEC01-ECG11	AIEC07-ECG04
<i>fimH</i> ²³	A27V, S70N , N78S	A27V, H166R	A27V, K32N*
<i>ompA</i> ²⁵	D46N, V114I , V196A, T224N , A272G	P46N, D47S, N48V, I49E, A125S, P128Y, GASF130-, D135N, V196A, T224N , A272G	P46N, D47S, N48V, I49E, V114I , A125S, P128Y, GASF130-, D135N, N172H, L182M,
<i>chiA</i> ²⁶	T100N, G166S, M182T, A200S, T286S, ETPV311, N326S, S335V, Q362K , E370K , V378A , V388E , M396L, I414V, N427D, T517A, E548V , A681D, R696K, S804S, Y810H, G811P	G166S, A200S, T286S, ETPV311-, N326A, S335V, Q362K , E370K , V378A , V388E , M396L, I414V, N416D, N427D, T517A, E548V , R696K	Ausente

*SNP no sinónimos descritos en este estudio mediante genómica comparativa, no asociados con patotipo de AIEC sólo presentes en AIEC07.

Ejemplo 3.- Genómica comparativa de pares de cepas de AIEC/no-AIEC

Se compararon los genomas de los tres pares de cepas de AIEC/no-AIEC dentro de los pares para identificar diferencias de contenido génico y SNP que podrían estar implicados en el fenotipo de AIEC.

3.1 Evaluación de diferencias de contenido génico

Se identificaron secuencias que codificaban para proteínas homólogas. Se obtuvieron un total de 5208 agrupaciones ortólogas, 3327 (63,9%) de las cuales eran comunes entre las seis cepas (Figura 3a). Estos genes representaban el 80%, el 77% y el 81% de los genomas en los pares AIEC17-ECG28, AIEC01-ECG11 y AIEC07-ECG04, respectivamente. Los pares de cepas que pertenecían a los grupos filogenéticos B2 y D compartían una mayor proporción de agrupaciones ortólogas de genes (el 9,2–9,6% de sus genomas) en comparación con la proporción que compartían las cepas de B1 con las de B2

(el 2,5–2,6%) o con las del grupo filogenético D (el 2,2–2,3%) (Figura 3b). La alta similitud entre pares de cepas también se puso de manifiesto en lo que se refiere al contenido génico; ya que las cepas que conformaban de un mismo par compartían más del 99,2% de agrupaciones ortólogas de genes.

5

Ninguna de las agrupaciones era compartida por las tres cepas AIEC, ni siquiera exclusivamente por dos de ellas (Figura 3a), lo que indica la ausencia de genes específicos de AIEC que se hallen presentes en todas las cepas AIEC. Sin embargo, dentro de cada par de cepas, se encontraron genes exclusivamente en la cepa AIEC (Tabla 4). AIEC17

10

contiene dos genes que codifican para proteínas no caracterizadas (YgiZ y YeeW), un gen que codifica para un transportador de cianato y un gen que codifica para una proteína de la familia TraR que está implicada en un proceso de *quorum sensing*³⁰. En el caso de AIEC01,

33 de los genes presentes en su genoma estaban ausentes del genoma de ECG11. De estos 33 genes, 20 codifican para proteínas de función desconocida o genérica, seis están

15

relacionados con el transporte transmembrana, cuatro están implicados en la regulación de la transcripción, y tres contribuyen al ensamblaje de flagelos. Finalmente, en comparación con ECG04, AIEC07 alberga tres proteínas de función desconocida (una de las cuales es una proteína específica de *E. coli* uropatógena), una proteína relacionada con el transporte de hierro intracelular (TonB) y el autotransportador UpaH. Esta última proteína media la

20

formación de biopelícula en la cepa uropatógena CFT073 (Allsopp, L. P. *et al*, 2010). La formación de biopelícula también es un rasgo fenotípico de cepas de AIEC (Martínez-Medina, M. *et al*, 2009).

Tabla 4. Número de agrupaciones ortólogas de genes para cada cepa y agrupaciones específicas presentes en cada cepa en comparación con su homólogo.

AIEC17			ECG28			Compartidos			Sólo en AIEC17	Sólo en ECG28
4139			4137			4135			34661;D7XCN7-Transportador de cianato ^b	32006;B3Y1R0- Proteína prevista ^a
									39495;G0FG36- Proteína de la familia TraR ^e	36779;E6BSL1- Proteína no caracterizada ^a
									57563;P64527-Proteína no caracterizada, YeeW ^a	
									59454-, Q46867-Proteína de membrana interna, YgiZ ^a	
AIEC01			ECG11			Compartidos			Sólo en AIEC01	Sólo en ECG11
4293			4268			4260			3007;B1VC65-Hidrolasa ^e	1036;Q8FCT4-Éster metílico de pimeloil-[proteína portadora de acilo] esterasa ^e
									25404;P0AG05-Estreptomicina 3"-adenililtransferasa ^e	12287;C3SKM7-αTDP-glucosa-4,6-deshidratasa ^e
									30306;B1EKJ8-Proteína no caracterizada ^a	23485;C3TC51-Aconitato hidratasa ^e
									34057;D7XGD6-Proteína no caracterizada ^a	24113;A0A137B781-Antitoxina HipB ^e
									36463;E7BTG4-Proteína no caracterizada, YubJ ^a	27497;A0A128B3E2-Posible metiltransferasa dependiente de adoMet, profago DLP12 ^e
									41088;H4JWB9-Proteína no caracterizada ^a	35556;Q47330-Posible glucosiltransferasa ^e
									46848;I2B9P9-Posible proteína de transporte de metabolitos de membrana interna, YdfJ ^b	97289;Q0Q086UTM4-Componente IIB de enzima fosfotransferasa específica de ascorbato ^b
									49208;J7QMD2-Proteína no caracterizada, cepa ECONIH1 ^a	116809;W8ZZ27- Proteína no caracterizada ^a
									50771;I2B9X8-Histidina cinasa ^e	

Tabla 4 (Cont.)

51401;P19319-Nitrato reductasa respiratoria 2 ^o	
55284;P03851-Proteína de 9,4 kDa no caracterizada ^a	
55674;P0A908 -Proteína que interacciona con MltA ^c	
55723;P0A9G4-CueR regulador de la transcripción de tipo HTH ^c	
56335;P0ACX1-Proteína de membrana interna, YdgC ^a	
56339;P0ACY5-Proteína no caracterizada, YeaG ^a	
56474;P0AE32-Proteína permeasa de transportador ABC de arginina, ArtM ^b	
56475;P0AE35- Proteína permeasa de transportador ABC de arginina, ArtQ ^b	
57533;P24216-Proteína asociada con gancho flagelar 2 ^d	
57640;P26608-Proteína flagelar, FliS ^d	
57776;P30859-Posible proteína de unión a arginina de transportador ABC, 2 ^b	
57780;P30860-Proteína de unión a arginina de transportador ABC, 1 ^b	
58516;P38105-Proteína de detección de inanición, RspB ^c	
59044;P52108-Proteína reguladora de la transcripción, RstA ^c	
59768;P76097-Proteína no caracterizada, YdcJ ^a	
59126;P56259-Posible proteína no caracterizada, YifN ^a	

Tabla 4 (Cont.)

		AIEC07		ECG04	Compartidos		
		4105	4121	4100			
		60232;P77400-Proteína de transporte de membrana interna, YbaT^b 60737;Q1R1V6-Proteína no caracterizada^a 61126; Q325B0-Glutaminasa^e 61671;Q57LR7-Cisteína sintasa^e 63434;Q9S4W7-Posible proteína antirrestricción, YubI^a 63745;C4ZQK4-Proteína flagelar, Flit^d 63829;C4ZZI6-Proteasa HtpX^e 64415;C4ZWZ4-Proteína de membrana de UPF0060, YnfA^a					
		Sólo en AIEC07					Sólo en ECG04
		33786;E2QKQ2-Proteína TonB^b 37356;G0FCH4- Proteína no caracterizada^a 45186;I3RYC0- Proteína no caracterizada^a 45841;JSY690-Autotransportador, UpaH^b 58698;Q8VV25-Proteína específica de <i>Escherichia coli</i> uropatógena^a					24721;P0CF06-Proteína de elemento de inserción IS1 2, Insa^e 30689;B7JCB1- sistema de transporte ABC, componente de membrana interna, SitC^b 31723;B7UPH0- Proteína no caracterizada^a 32043;C5J5A4-Componente de enzima fosfotransferasa IIB de tipo fructosa, 2^b 32153;C8CGP6-Proteína reelaxosoma, TraM^e 33090;D7X8P5- Proteína de dominio de chaperona de ensamblaje de fimbrias Gram negativas^e

Tabla 4 (Cont.)

34269;D9Z590-YfdA ^a	
34690;E2QEX8-Antitoxina ^e	
38492;G0FCH6-Proteína no caracterizada ^a	
38494;G0FCI7-Proteína de entrada ^e	
39795;H3MK58- Proteína que contiene dominio de metilación/escisión de N-terminal de tipo prepilina ^e	
39833;H4I2H5-Proteína no caracterizada ^a	
42792;H5V1F8-Posible proteína no caracterizada ^a	
47160;J1QEW4-Proteína de dominio central de integrasa ^e	
50439;I2B5V7-Proteína de dominio de cupina de barril conservado 2 ^e	
55545;P26841-Regulador de la transcripción de tipo HTH, XapR ^c	
57052;P45757-Posible proteína de sistema de secreción de tipo II, C ^b	
57299;P57018-Proteína de 9,0 KDa no caracterizada ^a	
59092;Q1RFR5- Proteína no caracterizada ^a	
59116;Q1RPL7-Proteína ECs1337 ^a	
59330;Q46687-Gen regulador de las fimbrias ^e	

^a Función desconocida; ^b Relacionado con transporte; ^c Implicado en la regulación de la transcripción; ^e Contribuyen al ensamblaje de flagelos; ^e Otras funciones. El código de anotación de BG7 y el número de registro de UniProt se indican para cada gen y se asignan según la base de datos, julio de 2015.

3.2 Detección de SNP asociados con AIEC

La aproximación utilizada para identificar SNP asociados con AIEC, se caracterizó por incluir tan sólo aquellas posiciones de nucleótido que variaban entre las cepas de un par y a la vez que también se ubicaban en secuencias homólogas de un genoma de referencia de AIEC (cepa UM146). Se encontraron un total de 286 polimorfismos; la mayoría (213) de los SNP estaban ubicados en secuencias codificantes de genes (tabla 5).

Tabla 5. Número de de SNP en los pares de cepas de AIEC/no-AIEC secuenciados.

	AIEC17-ECG28	AIEC01-ECG11	AIEC07-ECG04
SNP totales	51	126	109
SNP totales en genes	40	91	82
SNP seleccionados^{*1}	20 (7)	10 (5)	30 (19)
SNP confirmados ^{*2}	4 (4)	7 (3)	9 (4)
SNP estudiados en un conjunto de cepas ^{*3}	4 (4)	4 (2)	8 (3)

* El número de genes en los que están ubicados los SNP se indica entre paréntesis.

¹ SNP que se adaptan a los siguientes criterios: (I) producen un cambio de aminoácido no sinónimo; (II) no están ubicados en los extremos de un cóntigo; y (III) se validaron bioinformáticamente mediante ClustalW.

²Validado mediante el método de Sanger.

³SNP confirmados que no eran específicos de cepa (y cuatro SNP específicos de cepa para validación).

Sólo 20 de las 60 diferencias genómicas que se volvieron a secuenciar se confirmaron mediante secuenciación de Sanger. La ubicación de los SNP confirmados, las variantes de nucleótido y las funciones génicas se muestran en la Tabla 6.

Tabla 6. Ubicación de los SNP confirmados, variantes de nucleótido y funciones génicas.

ID	Posición en UM146(NC_017 632)	Ubicación de SNP en el genoma de AIEC (posición de cóntigo)	Variante de nucleótido* ¹ (AIEC/distinto de AIEC)	Nombre de proteína	Familia de proteína	Ontología de genes
E1-E2_3.4	1960688	8: 204059	C/T	Regulador de la transcripción de la familia GntR	PF00392; PF07702	GO:0003677; GO:0003700; GO:0006351
E1-E2_3.6	2503226	105: 325	C/T	Proteína de fago	PF06174	
E1-E2_3.7	51330	3: 50414	T/C	Serina peptidasa, DegQ	PF13365; PF13180; PF00595	GO:0004252
E1-E2_5	4195813	51: 69	G/T	Transportador de vitamina B12, BtuB	PF07715	GO:0009279; GO:0015235; GO:0006811; GO:0046872; GO:0046930; GO:0015288; GO:0004872
E3-E4_4.3	1425714; 1425708; 1425672	3: 167, 173, 209	C/Y, Y/Y, T/K	Posible proteína no caracterizada	PF06174	
E3-E4_4.4	1433063	84: 1126	R/R	dGTPasa	PF00350	GO:0005525
E3-E4_4.7	1430771; 1431049; 1431061	80: 920, 932, 1013	S/S, S/S, Y/Y	Proteína de quimiotaxia	PF13990	
E5-E6_3.1	4654977	3: 6356	A/C	FimH	PF00419; PF09160	GO:0007155; GO:0009289
E5-E6_3.12	4271710	83: 442	A/G	Subunidad beta de succinil-CoA ligasa	PF08442	GO:0005524; GO:0000287; GO:0030145; GO:0004775; GO:0006099
E5-E6_3.16=3.22	2212719; 2212845; 2212846; 2212934; 2212947; 2212951	51: 418, 544, 545, 633, 646, 650	Y/Y, S/S, R/R M/M, Y/Y, S/S	Proteína de la familia Ail/Lom de enterobacterias	PF06316	GO:0009279; GO:0016021
E5-E6_3.17	2217953	62:583	R/R	Posible componente de profago	PF00877	

*¹ A: adenina; C: citosina; G: guanina; K: guanina o timina; M: adenina o citosina; R: adenina o guanina; S: guanina o citosina; T: timina; W: adenina o timina; Y: citosina o timina.

- Cabe indicar que 14 de estos SNP (70%) presentaban picos solapantes en los cromatogramas de Sanger (Tabla 6). Se planteó la hipótesis de que i) las cepas con bases
- 5** ambiguas pudieran presentar más de una copia del gen relevante en sus genomas o ii) la existencia de una variabilidad intraclonal en el sitio polimórfico. Para identificar cuáles de estas posibilidades daban lugar a los SNP con bases ambiguas, en primer lugar se realizaron búsquedas mediante BLASTn (Altschil, S. F. *et al*, 2009) en los genomas de las cepas y en el genoma de la cepa de referencia de AIEC para buscar genes duplicados, y en
- 10** segundo lugar, se analizaron los *reads* obtenidos de la secuenciación desiguiente generación (NGS) usando Tablet (Milne, I. *et al*, 2013) para confirmar las observaciones. Los genes duplicados mostraban un nucleótido único o principal en el SNP que difería del nucleótido presente en la otra copia del gen (éste era el caso para los SNP hallados en los genes E3-E4_4.3, E3-E4_4.4 y E3-E4_4.7). Los genes que no estaban duplicados
- 15** mostraban un único resultado en BLASTn y dos nucleótidos diferentes que tenían una frecuencia casi igual en las lecturas de secuenciación (éste era el caso de los SNP hallados en los genes E5-E6_3.16=3.22 y E5-E6_3.17). Por tanto, se concluyó que los picos solapantes no se debían a artefactos técnicos.
- 20** En este estudio, se ha realizado por primera vez un análisis de genómica comparativa entre cepas de AIEC y no- AIEC que se consideran clones con respecto a sus patrones de PFGE. En comparación con estudios de genómica comparativa previos, la metodología excluye la detección de variaciones genómicas que son inherentes, por ejemplo, en el origen filogenético de las cepas. Este enfoque lo han propuesto los inventores con el objetivo de
- 25** aumentar la posibilidad de identificar firmas moleculares que sean específicas de AIEC. De hecho, las dificultades previas en el descubrimiento de un rasgo específico para las cepas de AIEC puede explicarse por el hecho de que habitualmente se han comparado cepas de AIEC y no-AIEC no relacionadas filogenéticamente (Dogan, B. *et al*, 2014; Desilets, M. *et al*, 2015). Además, la mayoría de las AIEC estudiadas han pertenecido a un grupo filogenético particular (O'Brien, C. L. *et al*, 2016; Deshpande, N. P. *et al*, 2015). Los investigadores han
- 30** elegido secuenciar pares de cepas que pertenecen a grupos filogenéticos diferentes (B1, B2 y D) y se ha estudiado la distribución de las diferencias halladas en una colección que incluía aislados de origen filogenético diferente para determinar si eran universales entre las cepas de AIEC y estaban ausentes de las cepas no-AIEC.

No se encontraron diferencias significativas en la estructura genómica o incluso en el contenido génico entre los pares de cepas AIEC y no-AIEC estudiadas, lo que confirma su estrecha identidad tal y como se había visto mediante PFGE. No se encontró ningún gen que estuviera presente en al menos dos cepas de AIEC y ausente de todas las cepas -no-AIEC. Sin embargo, se observaron pequeñas diferencias en el contenido génico entre cepas del mismo par. No obstante, este resultado debe confirmarse porque podría ser una consecuencia de un ensamblaje incompleto del genoma (no se habrían hallado aquellos genes que aparecen fragmentados en el ensamblaje). En general, los resultados apoyan la idea de que el fenotipo de AIEC no está determinado por la presencia o ausencia de un gen particular, tal como observaron O'Brien *et al.* (O'Brien, C. L. *et al.*, 2016) en cepas del tipo de secuencia ST95.

En un intento adicional de explicar las diferencias fenotípicas observadas entre pares, se buscó identificar SNP que estaban presentes de manera diferencial en cepas de AIEC y sus homólogos no-AIEC. La tasa de aparición de tales SNP que se validaron adicionalmente mediante secuenciación de Sanger era baja. Esto puede explicarse por la acumulación de pequeños errores durante la construcción de la biblioteca y la secuenciación provocados por la fidelidad imperfecta de ADN polimerasas y la tasa de error intrínseco de la plataforma de secuenciación, y por errores derivados de los parámetros usados para el ensamblaje de las lecturas (Olson, N. D. *et al.*, 2015; McElroy, K. *et al.*, 2014). Ocasionalmente se observó que hay variabilidad y el nucleótido que aparece en el *cóntig* consenso resultante del ensamblado no representa todas las lecturas (*reads*).

Ejemplo 4.- Distribución de SNP en una colección de cepas de AIEC/no-AIEC

Se estudió la variabilidad de nucleótidos dentro de los SNP identificados en una colección de cepas de AIEC y no-AIEC para validar o refutar la hipótesis de que los SNP confirmados representan firmas moleculares para aplicar en la identificación específica del patotipo AIEC. Se estudiaron dieciséis SNP en 22 cepas de AIEC y 28 cepas no-AIEC aisladas de sujetos sanos y pacientes con CD y que pertenecen a varios grupos filogenéticos (A (n=9), B1 (n=7), B2 (n=28) y D (n=5)) (Tabla 7). Se seleccionaron sólo los "SNP confirmados" que no eran específicos de cepa tras realizar la evaluación *in silico*, porque los SNP específicos de cepa no serían útiles para la identificación de las AIEC. Se analizaron también cuatro SNP considerados específicos de cepa según el resultado obtenido de las pruebas *in silico*

(E1-E2_3.4, E1-E2_5, E1-E2_3.7 y E5-E6_3.1) para confirmar que eran específicos de cepa.

5 Tabla 7. Colección de cepas de *E. coli* usada para la validación de la idoneidad de las variantes de SNP como supuestas firmas para la detección de AIEC. Se indican el origen de grupo filogenético, el tipo de paciente a partir del que se aisló la cepa (sujeto con enfermedad de Crohn (CD) o sano (H)) y los índices de adhesión e invasión. Puede encontrarse información adicional sobre las características de las cepas en Martínez-Medina, *et al.* (2009)⁵.

Cepas	Grupo filogenético	Fenotipo	Origen	Índice adhesión ^a	Índice invasión ^b	Índice de replicación intramacrófagos en J774 ^c
AIEC10	A	AIEC	H	5,9	0,226	1414
AIEC19	A	AIEC	H	2,4	0,111	1568
AIEC23	A	AIEC	CD	9,7	0,568	2362
AIEC24	A	AIEC	CD	2	0,309	1626
ECG16	A	no- AIEC	H	0	0,002	na
ECG22	A	no- AIEC	H	0	0,006	na
ECG18	A	no- AIEC	CD	0,1	0,001	na
ECG19	A	no- AIEC	CD	0	0,023	na
ECG65	A	no- AIEC	CD	0,1	0,001	na
AIEC07	B1	AIEC	H	20	0,565	1693
ECG04	B1	no- AIEC	H	0,5	0,023	813
ECG46	B1	no- AIEC	H	0,0	0,001	na
ECG02	B1	no- AIEC	CD	0	0,001	na
ECG21	B1	no- AIEC	CD	6,1	0,033	na
ECG63	B1	no- AIEC	CD	0	0,003	na
ECG64	B1	no- AIEC	CD	0,3	0,019	na
AIEC04	B2	AIEC	H	21,6	0,32	585
AIEC06	B2	AIEC	H	10,2	0,177	1718
AIEC08	B2	AIEC	H	1,1	0,172	105
LF82	B2	AIEC	CD	25,7	2,261	777
AIEC01	B2	AIEC	CD	15,9	0,284	1567
AIEC02	B2	AIEC	CD	0,9	0,802	2188
AIEC05	B2	AIEC	CD	9,4	0,202	705
AIEC09	B2	AIEC	CD	5,4	0,216	2562
AIEC11	B2	AIEC	CD	4,4	0,508	848
AIEC12	B2	AIEC	CD	22,3	0,142	94
AIEC14	B2	AIEC	CD	9,8	0,238	801
AIEC15,1	B2	AIEC	CD	10	0,305	660
AIEC16,1	B2	AIEC	CD	9,7	1,400	921
AIEC21	B2	AIEC	CD	17	0,109	1297
AIEC25	B2	AIEC	CD	2,8	0,482	776
ECG08	B2	no- AIEC	H	0,3	0,004	na
ECG12	B2	no- AIEC	H	0,1	0,003	na
ECG13	B2	no- AIEC	H	0,2	0,004	na
ECG17	B2	no- AIEC	H	0,2	0,019	na
ECG41	B2	no- AIEC	H	0,3	0,003	na
ECG43	B2	no- AIEC	H	0,5	0,018	na
ECG49	B2	no- AIEC	H	0,3	0,008	na

ECG01	B2	no- AIEC	CD	0,2	0,002	na
ECG05	B2	no- AIEC	CD	0	0,039	na
ECG09	B2	no- AIEC	CD	0,2	0,001	na
ECG15	B2	no- AIEC	CD	5,3	0,038	na
ECG26	B2	no- AIEC	CD	0,2	0,005	na
ECG42	B2	no- AIEC	CD	0	0,032	na
AIEC17	D	AIEC	CD	21,6	0,266	1053
AIEC20	D	AIEC	CD	14,2	0,125	344
ECG28	D	no- AIEC	CD	2,2	0,057	1060
ECG34	D	no- AIEC	CD	0,4	0,019	na
ECG57	D	no- AIEC	CD	11,8	0,013	na
ECG23	Atípico	no- AIEC	H	0,9	0,052	na

a Número de bacterias por célula I-407. b Porcentaje de bacterias intracelulares tras 1 h de tratamiento con gentamicina en relación con el inóculo. c Porcentaje de bacterias intracelulares a las 24 h tras la infección en relación con el número tras 1 h de tratamiento con gentamicina. na; no analizado.

Resulta interesante que se encontraron algunas variantes de nucleótido que aparecían más frecuentemente en cepas de AIEC que en cepas no-AIEC (Tabla 8).

Tabla 8. Prevalencia de genes que albergan SNP en una colección de cepas de AIEC/no-AIEC y frecuencia de variantes de nucleótidos particulares con respecto al fenotipo de AIEC, y origen de grupo filogenético de las cepas. Sólo se presentan los SNP que presentan diferencias estadísticamente significativas con respecto a patotipo. Los valores se proporcionan en porcentajes con respecto al número total de cepas de AIEC o no-AIEC. Se presentan diferencias estadísticamente significativas para cada variante en negrita. *Para el análisis de grupo filogenético, se descartó la cepa no-AIEC atípica.

ID	Amplificación por PCR		Base de SNP vs fenotipo de AIEC			Base de SNP vs grupo filogenético					
	AIEC (n=22)	no-AIEC (n=28)	Ntd (N cepas)	AIEC	no-AIEC	valor de p	A (n=9)	B1 (n=7)	B2 (n=28)	D (n=5)	valor de p
E3-E4_4.3(2)	100	100	C (33)	45,45	82,14		77,78	71,43	60,71	60,00	
			T (9)	40,91	0,00	<0,001	22,22	0,00	21,43	20,00	0,667
			Y (8)	13,64	17,86		0,00	28,57	17,86	20,00	
E3-E4_4.4	77,28	71,43	A (5)	13,64	7,14		0,00	14,29	14,29	0,00	
			G (14)	9,09	42,86	0,010	33,33	28,57	32,14	0,00	0,636
			R (18)	54,55	21,43		33,33	14,29	42,86	40,00	
E5-E6_3.16=3.22(2)	90,91	89,28	C (14)	13,64	39,29		44,45	28,57	17,86	40,00	
			G (8)	31,82	3,57		22,22	0,00	14,29	40,00	
			T (13)	36,36	17,86		22,22	0,00	39,29	0,00	0,026
			S (5)	9,09	10,71	0,012	0,00	42,86	7,14	0,00	
			K (3)	0,00	10,71		0,00	0,00	10,71	0,00	
			Y (2)	0,00	7,14		11,11	14,29	0,00	0,00	

A: adenina; C: citosina; G: guanina; K: guanina o timina; M: adenina o citosina; R: adenina o guanina; S: guanina o citosina; T: timina; W: adenina o timina; Y: citosina o timina.

*Para el análisis de grupo filogenético, se descartó la cepa no-AIEC atípica

En particular, sólo en cepas AIEC se encontraron variantes con timidina en el SNP E3-E4_4.3(2), mientras que la variante que contenían citosina era más frecuentes dentro de cepas no-AIEC. Se obtuvieron resultados similares cuando el análisis se restringió a las

5 cepas del grupo filogenético B2 ($p=0,037$); allí, las cepas de AIEC ($n=6$) eran de nuevo las únicas que presentaban timidina, y la citosina era también más prevalente en cepas no-AIEC que en cepas de AIEC ($n=10$ y $n=7$, respectivamente). Este gen estaba presente en todas las cepas estudiadas, y las variantes de SNP no estaban asociadas con el origen de grupo filogenético de las cepas sino sólo con el fenotipo de AIEC. Otro SNP interesante era

10 el presente en el gen E3-E4_4.4, en el que el 42,86% de las cepas no-AIEC presentaban guanina mientras que menos del 10% de las cepas AIEC presentaban esta variante. Sin embargo, en este caso, no todas las cepas albergaban el gen, tal como es el caso de la cepa LF82. Finalmente, una guanina en SNP E5-E6_3.16=3.22(2) se hallaba más frecuentemente en cepas de AIEC, mientras que una citosina en esta posición estaba

15 asociada con cepas no-AIEC. Si sólo se consideran cepas B2, la variante de citosina se encuentra exclusivamente en cepas no-AIEC ($n=5$), mientras que la variante de guanina es específica para cepas AIEC ($n=4$) ($p=0,007$); por tanto, podría ser de interés como biomarcador para las cepas de AIEC de grupo filogenético B2. Sin embargo, los porcentajes de cepas que presentan estas dos variantes son bajos (el 41,7% de cepas no-AIEC B2 y el

20 30,8% de cepas AIEC B2).

Desde un punto de vista funcional, el gen E5-E6_3.16=3.22 codifica para una proteína de la familia Ail/Lom de enterobacterias. Esta familia de proteínas incluye proteínas de la membrana externa que están implicadas en virulencia bacteriana, tales como OmpX (en *E. coli* y *Enterobacter cloacae*) (Meng, X. *et al*, 2016) y PagC (en *Salmonella typhimurium*) (Pulkkinen, W. S. y Miller, S. I., 1991). Se ha notificado que estas proteínas desempeñan papeles en adhesión celular y en supervivencia intramacrófagos, respectivamente. El gen que contiene SNP E3-E4_4.4 codifica para una dGTPasa que es un miembro de la familia de proteínas de tipo dinamina (PF00350), cuya función en bacterias aún se desconoce.

25 Finalmente, la función de E3E4_4.3 es desconocida, aunque comparte el 97,2% (cobertura del 74,5%) de homología de secuencia de aminoácidos con la proteína hipotética yeeT. La baja homología de secuencia de estos genes con genes en las bases de datos disponibles actualmente hace difícil identificar con confianza las proteínas codificadas por estos genes. Por tanto, será necesaria la creación de mutantes isogénicos para entender la función

30

biológica de estas proteínas y demostrar los efectos de sus posibles variaciones de aminoácidos.

Ejemplo 5.- SNP en relación con la capacidad de adhesión e invasión

- 5 Tal como se esperaba, las cepas que portaban variantes de SNP asociadas con el patotipo AIEC mostraban un mayor índice de adhesión e invasión (Figura 4). La única excepción fue SNP E5-E6_3.16=3.22(2), en el que a pesar de ser mayor el índice de adhesión para la variante asociada a AIEC, no alcanzó la significación estadística. A su vez, cepas con guanina en SNP E3-E4_4.4 mostraron menores índices de adhesión e invasión. Un
- 10 polimorfismo adicional, SNP E5-E6_3.16=3.22(3), mostró diferencias significativas; cepas con adenina en este SNP presentaban una capacidad de invasión aumentada.

- Los SNP E5-E6_3.16=3.22(2) y E5-E6_3.16=3.22(3) son consecutivos y dan como resultado el mismo cambio de aminoácido. Para visualizar sus efectos sobre la secuencia
- 15 de aminoácidos, se prestó atención a las posibles combinaciones de nucleótidos encontradas a través de la colección de cepas de *E. coli*. Las combinaciones de las dos posiciones cambiantes condujeron a la posibilidad de 13 variantes de SNP que pueden traducirse en 6 aminoácidos diferentes. Tal como se esperaba, la combinación guanina-adenina estaba asociada con los valores de invasión más altos ($0,63\% \pm 0,76$). Se
- 20 observaron diferencias estadísticamente significativas según el patotipo cuando se comparó la combinación guanina-adenina ($n=7$) con la combinación citosina-guanina ($n=11$) ($p=0,009$). La primera conduce a un aminoácido básico a pH=7 (serina), mientras que la última codifica para un aminoácido neutro (alanina) que puede afectar a la función de la proteína.

25

Ejemplo 6.- Utilidad de los SNP como firmas moleculares para la identificación de AIEC

El uso de un modelo de regresión logística binario reveló dos SNP que eran predictivos del fenotipo de AIEC (Tabla 9).

30

Tabla 9. Modelo de regresión logística binario para los SNP asociados con el patotipo AIEC. Se indican las variables de la ecuación, el riesgo de ser AIEC (razón de probabilidades), el valor de p del modelo de regresión y el porcentaje de cepas clasificadas satisfactoriamente.

	Variables de la ecuación				Observado	Predicho			
	B	Valor de p	Razón de probabilidades	IC del 95%		Distinta de AIEC	AIEC	% correcto	% global
No tiene G en E3-E4_4.4	2,420	0,006	11,250	2,004-63,168	no-AIEC	12	8	60,0	73,0
Constante	-1,792	0,019	0,167		AIEC	2	15	88,2	
Tiene G en E5-E6_3.16=3.22(2)	2,559	0,023	12,923	1,430-116,785	no-AIEC	24	1	96,0	68,9
Constante	-0,613	0,075	0,542		AIEC	13	7	35,0	

5

El SNP en E3-E4_4.4 puede clasificar las cepas como AIEC o no-AIEC con un éxito global del 73%; las cepas de *E. coli* con una base de nucleótido distinta de guanina en esta posición tienen un 65,2% de probabilidad de presentar el fenotipo de AIEC, mientras que las cepas que la presentan tienen sólo un 14,3% de probabilidad de presentar el fenotipo AIEC.

10 En el caso del SNP E5-E6_3.16=3.22(2), el éxito global era similar (68,9%), pero sólo el 35% de las cepas AIEC se clasificaron correctamente.

Aunque el éxito global en la predicción de AIEC basada en el SNP en E3-E4_4.4 era alto, el 40% de las cepas no-AIEC se clasificaron erróneamente como AIEC. Para mejorar la especificidad de predicción, se diseñó un algoritmo de clasificación basado en la identificación de los nucleótidos presentes en tres SNP (Figura 5). En este algoritmo, se determina en primer lugar la variante en SNP E3-E4_4.4, y las cepas que contienen una guanina se clasifican como no-AIEC con una probabilidad de éxito del 85,7%. A continuación, en aquellas cepas con otro resultado en SNP E3-E4_4.4 se analizan para el gen E5-E6_3.16=3.22(2). Los aislados resultantes se clasifican como AIEC o no-AIEC dependiendo de los resultados combinados obtenidos para ambos genes. Con este método, la probabilidad de éxito oscila entre el 71,4% y el 100%, con la excepción de aislados que presentan nucleótidos diferentes de guanina en ambos genes, que siguen sin poder clasificarse. Por este motivo, se incluyó un tercer gen (E5-E6_3.12). A pesar de no

15

20

encontrarse más frecuentemente en AIEC (concretamente el 39,3% de no-AIEC y el 22,7% de AIEC presentaban adenina, mientras que el 60,7% de no-AIEC y el 77,3% de AIEC presentaban guanina; $p=0,197$), este gen era útil en la clasificación de este grupo de cepas. Globalmente, el algoritmo de clasificación presenta un 82,1% de especificidad, un 86,4% de sensibilidad y un 84% de exactitud dentro de la colección de cepas.

Estos resultados sugieren colectivamente que aunque no se podían encontrar firmas moleculares específicas para AIEC y presentes en todas las AIEC, con el presente enfoque se han identificado algunos genes con polimorfismos que pueden representar una ventaja o desventaja para las capacidades de adhesión e invasión de *E. coli* y por tanto pueden estar implicados en el fenotipo de AIEC. Además, se ha diseñado una estrategia molecular novedosa basada en la identificación de los nucleótidos presentes en tres sitios polimórficos que puede implementarse como herramienta para la identificación de AIEC.

15 BIBLIOGRAFÍA

1. Sartor, R. B. Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nat Clin Pr. Gastroenterol Hepatol* **3**, 390–407 (2006).
2. Martínez-Medina, M. *et al.* Molecular diversity of *Escherichia coli* in the human gut: New ecological evidence supporting the role of adherent-invasive *E. coli* (AIEC) in Crohn's disease. *Inflamm. Bowel Dis.* **15**, 872–882 (2009).
3. Darfeuille-Michaud, A. *et al.* High prevalence of adherent-invasive *Escherichia coli* associated with ileal mucosa in Crohn's disease. *Gastroenterology* **127**, 412–421 (2004).
4. Baumgart, M. *et al.* Culture independent analysis of ileal mucosa reveals a selective increase in invasive *Escherichia coli* of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. *ISME J.* **1**, 403–418 (2007).
5. López-Siles, M. *et al.* Mucosa-associated *Faecalibacterium prausnitzii* and *Escherichia coli* co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. *Int. J. Med. Microbiol.* **304**, 464–475 (2014).
6. Boudeau, J., Glasser, A. L., Masseret, E., Joly, B. & Darfeuille-Michaud, A. Invasive ability of an *Escherichia coli* strain isolated from the ileal mucosa of a patient with Crohn's disease. *Infect. Immun.* **67**, 4499–509 (1999).
7. Martin, H. M. *et al.* Enhanced *Escherichia coli* adherence and invasion in Crohn's disease and colon cancer. *Gastroenterology* **127**, 80–93 (2004).

8. Sasaki, M. *et al.* Invasive *Escherichia coli* are a feature of Crohn's disease. *Lab. Investig.* **87**, (2007).
9. Eaves-Pyles, T. *et al.* *Escherichia coli* isolated from a Crohn's disease patient adheres, invades, and induces inflammatory responses in polarized intestinal epithelial cells. *Int. J. Med. Microbiol.* **298**, 397–409 (2008).
10. Dogan, B. *et al.* Multidrug resistance is common in *Escherichia coli* associated with Ileal Crohn's Disease. *Inflamm. Bowel Dis.* **19**, 141–150 (2013).
11. Miquel, S. *et al.* Complete genome sequence of crohn's disease-associated adherent-invasive *E. coli* strain LF82. *PLoS One* **5**, 1–16 (2010).
12. Nash, J. H. *et al.* Genome sequence of adherent-invasive *Escherichia coli* and comparative genomic analysis with other *E. coli* pathotypes. *BMC Genomics* **11**, 667 (2010).
13. Martínez-Medina, M. *et al.* Similarity and divergence among adherent-invasive *Escherichia coli* and extraintestinal pathogenic *E. coli* strains. *J. Clin. Microbiol.* **47**, 3968–79 (2009).
14. Kotlowski, R., Bernstein, C. N., Sepehri, S. & Krause, D. O. High prevalence of *Escherichia coli* belonging to the B2+D phylogenetic group in inflammatory bowel disease. *Gut* **56**, 669–75 (2007).
15. Conte, M. P. *et al.* Adherent-invasive *Escherichia coli* (AIEC) in pediatric Crohn's disease patients: phenotypic and genetic pathogenic features. *BMC Res. Notes* **7**, 748 (2014).
16. Cieza, R. J., Hu, J., Ross, B. N., Sbrana, E. & Torres, A. G. The IbeA invasin of adherent-invasive *Escherichia coli* mediates interaction with intestinal epithelia and macrophages. *Infect. Immun.* **83**, 1904–18 (2015).
17. Dogan, B. *et al.* Inflammation-associated Adherent-invasive *Escherichia coli* are enriched in pathways for use of propanediol and iron and M-cell. *Inflamm. Bowel Dis.* **20**, 1919–1932 (2014).
18. Gibold, L. *et al.* The Vat-AIEC protease promotes crossing of the intestinal mucus layer by Crohn's disease-associated *Escherichia coli*. *Cell. Microbiol.* n/a-n/a (2015). doi:10.1111/cmi.12539
19. Martínez-medina, M. & Garcia-gil, L. J. *Escherichia coli* in chronic inflammatory bowel diseases : An update on adherent invasive *Escherichia coli* pathogenicity. **5**, 213–227 (2014).
20. Vazeille, E. *et al.* GipA factor supports colonization of Peyer's Patches by Crohn's Disease-associated *Escherichia coli*. *Inflamm. Bowel Dis.* **22**, 68–81 (2016).

21. Chassaing, B. *et al.* Crohn disease-associated adherent-invasive *E. coli* bacteria target mouse and human Peyer's patches via long polar fimbriae. *J. Clin. Invest.* **121**, 966–75 (2011).
22. Iebba, V. *et al.* Microevolution in fimH gene of mucosa-associated *Escherichia coli* strains isolated from pediatric patients with inflammatory bowel disease. *Infect. Immun.* **80**, 1408–17 (2012).
23. Dreux, N. *et al.* Point mutations in FimH Adhesin of Crohn's Disease-associated Adherent-invasive *Escherichia coli* enhance intestinal inflammatory response. *PLoS Pathog.* **9**, 1–17 (2013).
24. Desilets, M. *et al.* Genome-based definition of an Inflammatory Bowel Disease-associated Adherent-invasive *Escherichia coli* pathovar. *Inflamm. Bowel Dis.* **0**, 1–12 (2015).
25. Rolhion, N. *et al.* Abnormally expressed ER stress response chaperone Gp96 in CD favours adherent-invasive *Escherichia coli* invasion. *Gut* **59**, 1355–62 (2010).
26. Low, D. *et al.* Chitin-binding domains of *Escherichia coli* ChiA mediate interactions with intestinal epithelial cells in mice with colitis. *Gastroenterology* **145**, 602–12.e9 (2013).
27. O'Brien, C. L. *et al.* Comparative genomics of Crohn's disease-associated adherent-invasive *Escherichia coli*. *Gut* gutjnl-2015-311059 (2016). doi:10.1136/gutjnl-2015-311059
28. Deshpande, N. P., Wilkins, M. R., Mitchell, H. M. & Kaakoush, N. O. Novel genetic markers define a subgroup of pathogenic *Escherichia coli* strains belonging to the B2 phylogenetic group. *FEMS Microbiol. Lett.* 1–7 (2015). doi:10.1093/femsle/fnv193
29. Rolhion, N., Carvalho, F. A. & Darfeuille-Michaud, A. OmpC and the σ E regulatory pathway are involved in adhesion and invasion of the Crohn's disease-associated *Escherichia coli* strain LF82. *Mol. Microbiol.* **63**, 1684–1700 (2007).
30. Vannini, A. *et al.* The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *EMBO J.* **21**, 4393–401 (2002).
31. Allsopp, L. P. *et al.* UpaH is a newly identified autotransporter protein that contributes to biofilm formation and bladder colonization by uropathogenic *Escherichia coli* CFT073. *Infect. Immun.* **78**, 1659–69 (2010).
32. Martínez-Medina, M. *et al.* Biofilm formation as a novel phenotypic feature of adherent-invasive *Escherichia coli* (AIEC). *BMC Microbiol.* **9**, 202 (2009).
33. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–402 (1997).

34. Milne, I. *et al.* Using Tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* **14**, 193–202 (2013).
35. Prorok-Hamon, M. *et al.* Colonic mucosa-associated diffusely adherent *afaC+* *Escherichia coli* expressing *lpfA* and *pks* are increased in inflammatory bowel disease and colon cancer. *Gut* **63**, 761–70 (2014).
- 5 36. Céspedes, S. *et al.* Genetic diversity and virulence determinants of *Escherichia coli* strains isolated from patients with Crohn's Disease in Spain and Chile. *Front. Microbiol.* **8**, 639 (2017).
37. Olson, N. D. *et al.* Best practices for evaluating single nucleotide variant calling methods for microbial genomics. *Front. Genet.* **6**, 235 (2015).
- 10 38. McElroy, K., Thomas, T. & Luciani, F. Deep sequencing of evolving pathogen populations: applications, errors, and bioinformatic solutions. *Microb. Inform. Exp.* **4**, 1 (2014).
39. den Bakker, H. C. *et al.* Comparative genomics of the bacterial genus *Listeria*: Genome evolution is characterized by limited gene acquisition and limited gene loss. *BMC Genomics* **11**, 688 (2010).
- 15 40. Herring, C. D. *et al.* Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat. Genet.* **38**, 1406–1412 (2006).
- 20 41. Meng, X. *et al.* Virulence characteristics of extraintestinal pathogenic *Escherichia coli* deletion of gene encoding the outer membrane protein X. *J. Vet. Med. Sci* **78**, 1261–1267 (2016).
42. Pulkkinen, W. S. & Miller, S. I. A *Salmonella typhimurium* virulence protein is similar to a *Yersinia enterocolitica* invasion protein and a bacteriophage lambda outer membrane protein. *J. Bacteriol.* **173**, 86–93 (1991).
- 25 43. Spurbeck, R. R. *et al.* *Escherichia coli* isolates that carry *vat*, *fyuA*, *chuA*, and *yfcV* efficiently colonize the urinary tract. *Infect. Immun.* **80**, 4115–22 (2012).
44. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–80 (1994).
- 30 45. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–77 (2012).
46. Pareja-Tobes, P., Manrique, M., Pareja-Tobes, E., Pareja, E. & Tobes, R. BG7: a new approach for bacterial genome annotation designed for next generation sequencing
- 35

- data. *PLoS One* **7**, e49239 (2012).
47. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**, e11147 (2010).
 48. Grissa, I., Vergnaud, G. & Pourcel, C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **35**, W52–W57 (2007).
 49. Wang, Y., Coleman-Derr, D., Chen, G. & Gu, Y. Q. OrthoVenn: A web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* **43**, W78–W84 (2015).
 50. Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **15**, 524 (2014).
 51. Krause, D. O., Little, A. C., Dowd, S. E. & Bernstein, C. N. Complete genome sequence of adherent invasive *Escherichia coli* UM146 isolated from ileal crohn's disease biopsy tissue. *J. Bacteriol.* **193**, 583 (2011).
 52. Gene Ontology Consortium, T. G. O. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049-56 (2015).
 53. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279-85 (2016).
 54. Sivignon A, *et al.* Development of heptylmannosidebased glycoconjugate antiadhesive compounds against adherent-invasive *Escherichia coli* bacteria associated with Crohn's disease. *mBio* **6**(6):e01298-15 (2015).
 55. Sivignon A, *et al.* The potential of FimH as a novel therapeutic target for the treatment of Crohn's disease. *Expert Opinion on Therapeutic Targets*, **21**:9, 837-847, (2017).
 56. Subramanian S, *et al.* Replication of colonic Crohn's disease mucosal *Escherichia coli* isolates within macrophages and their susceptibility to antibiotics. *Antimicrob Agents Chemother.*; **52**(2):427–434 (2008).
 57. Galtier M, *et al.* Bacteriophages targeting adherent invasive *Escherichia coli* strains as a promising new treatment for Crohn's disease. *J Crohns Colitis*.;jjw 224 (2017).
 58. Barnich N, *et al.*, CEACAM6 acts as a receptor for adherent-invasive *E. coli*, supporting ileal mucosa colonization in Crohn disease. *J Clin Invest.*; **117**(6):1566–1574 (2007).
 58. Sivignon A, *et al.* *Saccharomyces cerevisiae* CNCM I-3856 prevents colitis induced by AIEC bacteria in the transgenic mouse model mimicking Crohn's disease. *Inflamm Bowel Dis.*; **21**(2):276–286 (2015).

59. Krause, D.O., *et al.*, Complete genome sequence of adherent invasive *Escherichia coli* UM146 isolated from ileal Crohn's disease biopsy tissue, *J. Bacteriol.* **193** (2), 583 (2011)
- 5 60. Gouin SG, Roos G, Bouckaert J. Carbohydrates as drugs: Discovery and application of FimH antagonists. Seeberger PH, Rademacher C, editors. Springer International Publishing (Basilea, Suiza). págs. 123–168 (2014).
61. Mydock-McGrane LK, Cusumano ZT, Janetka JW. Mannose-derived FimH antagonists: a promising anti-virulence therapeutic strategy for urinary tract infections and Crohn's disease. *Expert Opin Ther Pat.*; **26**(2):175–197 (2016).
- 10 62. Grabosch C, Hartmann M, Schmidt-Lassen J, *et al.* Squaric acid monoamide mannosides as ligands for the bacterial lectin FimH: covalent inhibition or not? *ChemBioChem.*; **12**(7):1066–1074 (2011).
63. Firon N, Ashkenazi S, Mirelman D, *et al.* Aromatic alpha-glycosides of mannose are powerful inhibitors of the adherence of type 1 fimbriated *Escherichia coli* to yeast and intestinal epithelial cells. *Infect Immun.*; **55**(2):472–476 (1987).
- 15 64. Bernardi A, Jiménez-Barbero J, Casnati A, *et al.* Multivalent glycoconjugates as anti-pathogenic agents. *Chem Soc Rev.*; **42** (11):4709–4727 (2013).
65. Simpson KW, Dogan B, Rishniw M, *et al.* Adherent and invasive *Escherichia coli* is associated with granulomatous colitis in boxer dogs, *Infect Immun.*; **74**(8):4778-92 (2006).
- 20

REIVINDICACIONES

1. Método para determinar o predecir si una cepa de *E. coli* tiene un fenotipo de *E. coli* adherente-invasiva (AIEC), en el que dicho método comprende:

5

- a) determinar la presencia o ausencia del gen E3-E4_4.4 en el genoma de la cepa de *E. coli*, y cuando está presente, determinar el genotipo en el sitio polimórfico E3-E4_4.4 correspondiente a la posición 1433063 pb del genoma de la cepa AIEC de referencia UM146 (SNP E3-E4_4.4);

10

en el que la presencia de un nucleótido de G en el sitio polimórfico de a) es indicativa de un fenotipo distinto de AIEC.

2. Método según la reivindicación 1, en el que dicho método comprende además:

15

- b) determinar la presencia o ausencia del gen E5-E6_3.16=3.22 en el genoma de la cepa de *E. coli*, y cuando está presente determinar el genotipo en el sitio polimórfico E5-E6_3.16=3.22(2) correspondiente a la posición 2212845 pb del genoma de la cepa AIEC de referencia UM146 (SNP E5-E6_3.16=3.22(2));

20

en el que la ausencia del gen E3-E4_4.4 en a), y la presencia de un nucleótido de G en el sitio polimórfico de b) son indicativas de un fenotipo de AIEC.

3. Método según la reivindicación 1, en el que dicho método comprende además:

25

- b) determinar la presencia o ausencia del gen E5-E6_3.16=3.22 en el genoma de la cepa de *E. coli*, y cuando está presente determinar el genotipo en el SNP E5-E6_3.16=3.22(2);

30

- c) determinar la presencia o ausencia del gen E5-E6_3.12 en el genoma de la cepa de *E. coli*, y cuando está presente determinar el genotipo en el sitio polimórfico E5-E6_3.12 correspondiente a la posición 4271710 pb del genoma de la cepa AIEC de referencia UM146 (SNP E5-E6_3.12);

en el que la presencia de un nucleótido distinto de G en el sitio polimórfico de a), la presencia de un nucleótido distinto de G en el sitio polimórfico de b) y la presencia de una A en el sitio polimórfico de c) son indicativas de un fenotipo distinto de AIEC.

- 5** 4. Método según cualquiera de las reivindicaciones 1 a 3, en el que dicho método comprende además determinar el genotipo de uno, dos, tres o más de los SNP seleccionados del grupo que consiste en: E1-E2_3.6, E3-E4_4.3(1), E3-E4_4.3(2), E3-E4_4.3(3), E5-E6_3.16=3.22(1), E5-E6_3.16=3.22(3), E5-E6_3.16=3.22(4), E5-E6_3.16=3.22(5), y E5-E6_3.16=3.22(6).
- 10** 5. Método según cualquiera de las reivindicaciones 1 a 4, en el que dicha cepa de *E. coli* es de cualquier grupo filogenético, preferiblemente de los grupos filogenéticos A, D, B2 o B1.
- 15** 6. Método según cualquiera de las reivindicaciones 1 a 5, en el que se realiza genotipado mediante un método seleccionado del grupo que consiste en secuenciación directa, análisis de polimorfismo de longitud de fragmentos de restricción (RFLP), hibridación con oligonucleótidos específicos de alelos (ASO), reacción en cadena de la polimerasa (PCR) específica de alelos; PCR y secuenciación, PCR usando cebadores mutagénicos;
- 20** ligasa-PCR, escisión mediante HOT; electroforesis en gel con gradiente de desnaturalización (DGGE), electroforesis en gel con gradiente de desnaturalización por temperatura (TGGE), polimorfismo conformacional de cadena sencilla (SSCP), cromatografía de líquidos de alto rendimiento con desnaturalización, microalineamientos de ADN, metagenómica y reconocimiento de aptámeros.
- 25** 7. Método según cualquiera de las reivindicaciones 1 a 6, en el que se realiza genotipado mediante un método que comprende PCR y secuenciación.
- 30** 8. Método según cualquiera de las reivindicaciones 1 a 7, en el que dicho método comprende una etapa de extracción de ADN antes de la determinación del genotipo del SNP diana.
- 35** 9. Método según cualquiera de las reivindicaciones 1 a 8, en el que dicha cepa de *E. coli* se ha obtenido de una muestra intestinal de un sujeto, preferiblemente de una muestra de heces.

10. Método según cualquiera de las reivindicaciones 1 a 9, en el que dicho sujeto es un sujeto humano.
- 5** 11. Método según cualquiera de las reivindicaciones 1 a 10, en el que dicho método comprende además almacenar los resultados del método en un soporte de datos, preferiblemente en el que dicho soporte de datos es un medio legible por ordenador.
- 10** 12. Método implementado por ordenador, en el que el método es tal como se define en cualquiera de las reivindicaciones 1 a 11.
- 10** 13. Método para clasificar un sujeto como portador de AIEC según la presencia de *E. coli* del fenotipo AIEC en su tracto intestinal, en el que dicho método comprende:
- 15** a. aislar cepas de *E. coli* de una muestra intestinal obtenida de un sujeto;
- 15** b. opcionalmente, extraer ADN de dichas cepas de *E. coli*;
- c. determinar o predecir la presencia de *E. coli* del fenotipo AIEC usando un método según cualquiera de las reivindicaciones 1 a 12.
- 20** 14. Método para el examen de sujetos portadores de AIEC o la monitorización del estado de portador de AIEC en un sujeto, en el que dicho método comprende las etapas a) a c) según la reivindicación 13.
- 25** 15. Método para clasificar un sujeto como portador de AIEC según la reivindicación 13, o método para el examen o la monitorización de un sujeto según la reivindicación 14, en el que dicho sujeto tiene la enfermedad de Crohn (CD).
- 30** 16. Kit adecuado para determinar en el genoma de una cepa de *E. coli* al menos el genotipo del SNP E3-E4_4.4, y opcionalmente para determinar el genotipo del SNP E5-E6_3.16=3.22(2) y SNP E5-E6_3.12, en el que dicho kit comprende:
- 30** – un oligonucleótido específico para una secuencia diana que comprende el SNP E3-E4_4.4;
- opcionalmente, que comprende además un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(2);
- 35** – opcionalmente, que comprende además un oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.12; y

- opcionalmente, que comprende además instrucciones para el uso de dicho(s) reactivo(s) en un método para determinar el genotipo del SNP diana.

17. Kit según la reivindicación 16, en el que:

- 5** – dicho oligonucleótido específico para una secuencia diana que comprende el SNP E3-E4_4.4 es una secuencia de ácido nucleico que comprende SEQ ID NO:13, SEQ ID NO:14 o una secuencia con una identidad de al menos aproximadamente el 90%, preferiblemente al menos aproximadamente el 95%, el 96%, el 97%, el 98% o el 99% con ella; y/o
- 10** – dicho oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.16=3.22(2) es una secuencia de ácido nucleico que comprende SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:30, SEQ ID NO:31, o una secuencia con una identidad de al menos aproximadamente el 90%, preferiblemente al menos aproximadamente el 95%, el 96%, el 97%, el 98% o el 99% con ella; y/o
- 15** – dicho oligonucleótido específico para una secuencia diana que comprende el SNP E5-E6_3.12 es una secuencia de ácido nucleico que comprende SEQ ID NO:21, SEQ ID NO:22 o una secuencia con una identidad de al menos aproximadamente el 90%, preferiblemente al menos aproximadamente el 95%, el 96%, el 97%, el 98% o el 99% con ella.

20

18. Uso de un kit según cualquiera de las reivindicaciones 16 ó 17, para determinar o predecir si una cepa de *E. coli* tiene un fenotipo de AIEC.

Figura 1

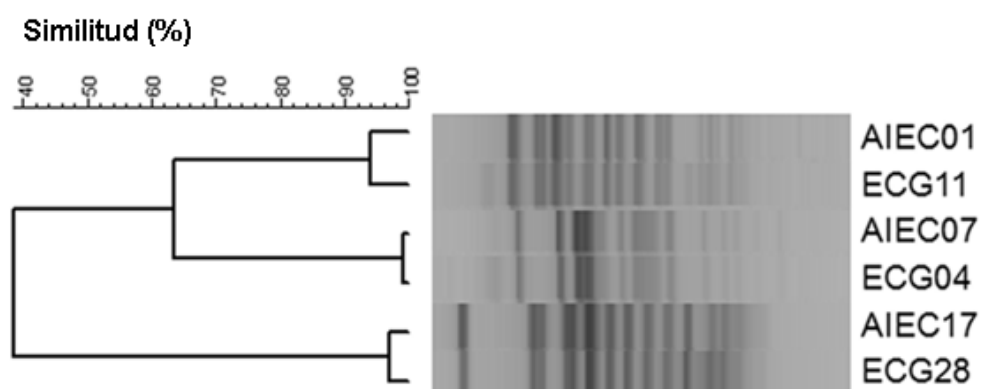


Figura 2

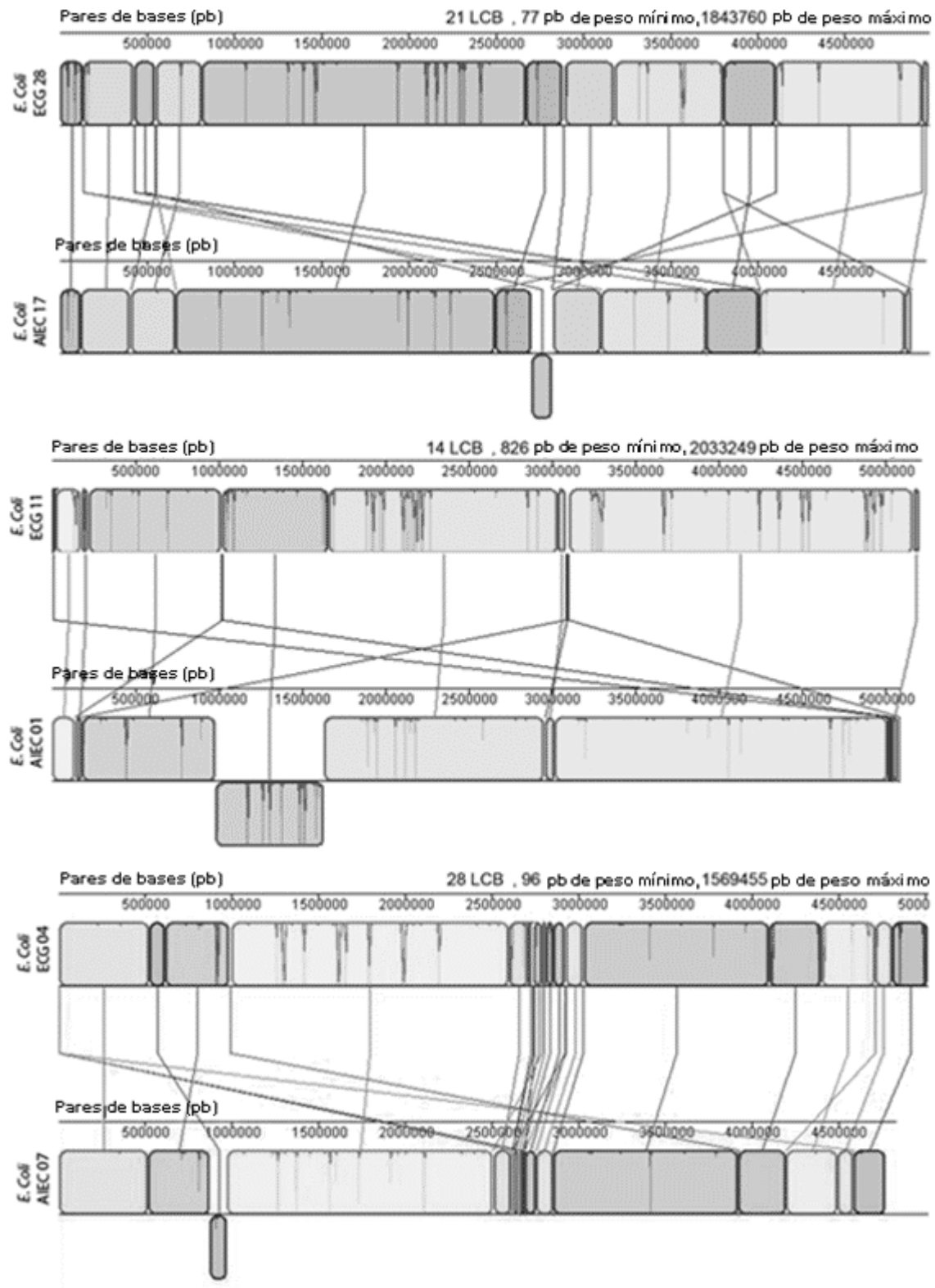


Figura 3

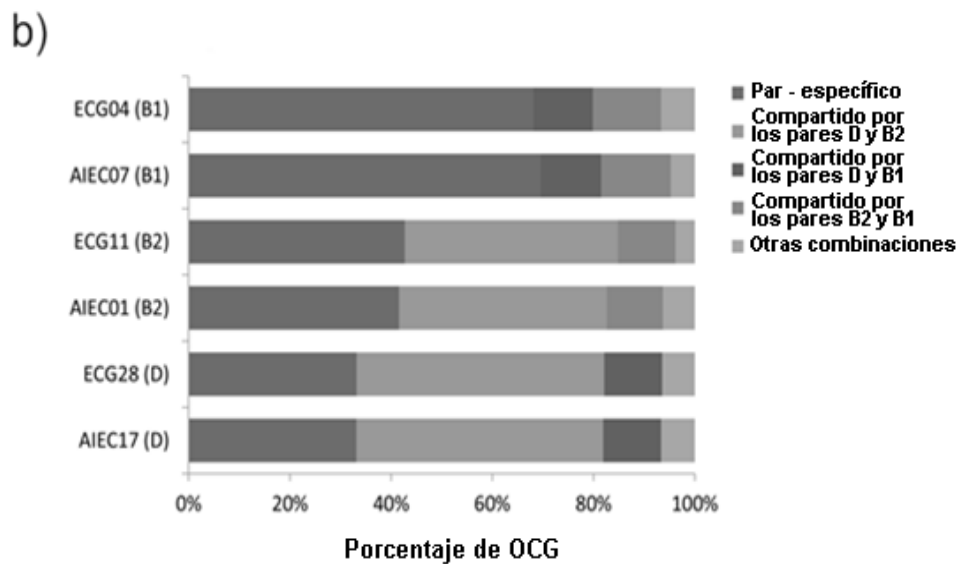
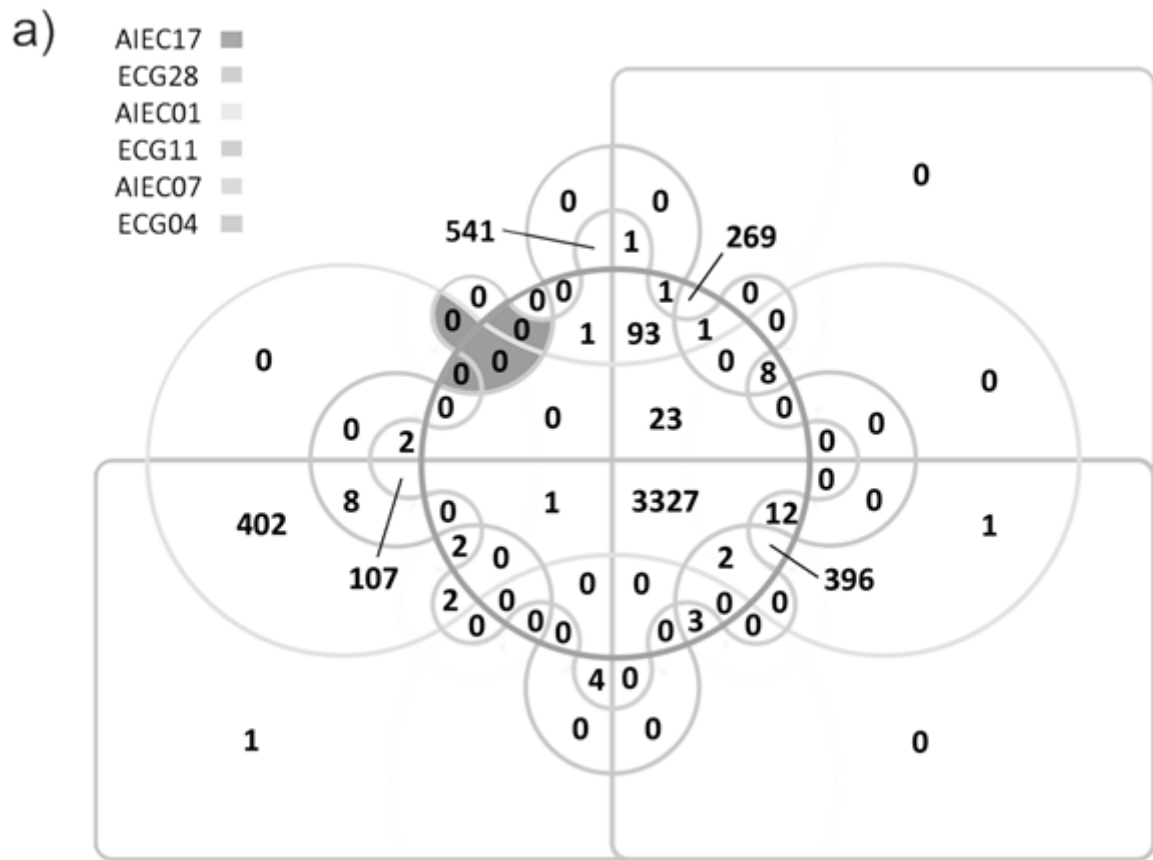


Figura 4

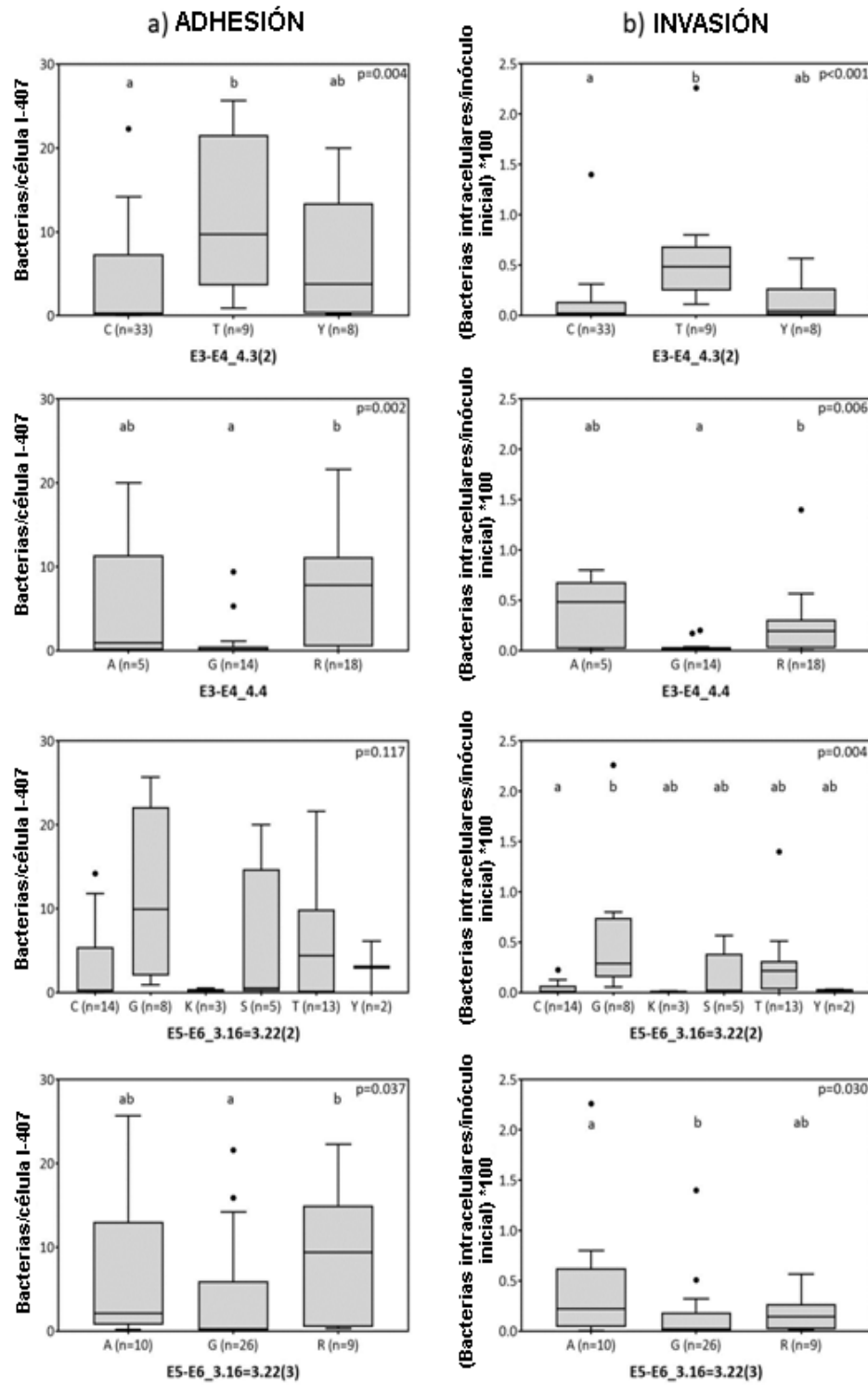


Figura 5

