

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 767 527**

51 Int. Cl.:

**C12Q 1/6883** (2008.01)

**C12Q 1/689** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.06.2015 PCT/CN2015/082044**

87 Fecha y número de publicación internacional: **07.01.2016 WO16000539**

96 Fecha de presentación y número de la solicitud europea: **23.06.2015 E 15814937 (7)**

97 Fecha y número de publicación de la concesión europea: **20.11.2019 EP 3161167**

54 Título: **Detección de taxones bacterianos para predecir resultados adversos del embarazo**

30 Prioridad:

**30.06.2014 US 201462018920 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**17.06.2020**

73 Titular/es:

**THE CHINESE UNIVERSITY OF HONG KONG  
(100.0%)  
Shatin, N.T.  
Hong Kong, CN**

72 Inventor/es:

**CHIM, STEPHEN SIU-CHUNG;  
CHEUNG, CHEE-YIN;  
CHEUNG, WAN-CHEE;  
MENG, MENG;  
LEUNG, TAK-YEUNG y  
LEE, KEUN-YOUNG**

74 Agente/Representante:

**SÁEZ MAESO, Ana**

ES 2 767 527 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Detección de taxones bacterianos para predecir resultados adversos del embarazo

Antecedentes de la invención

5 En un embarazo normal sin complicaciones, el cuello uterino es largo y cerrado hasta el final del tercer trimestre, cuando eventualmente se acortará y se dilatará cuando el feto se desarrolle completamente y esté listo para el nacimiento. Por el contrario, en un embarazo complicado por acortamiento cervical o dilatación cervical avanzada, el cuello uterino progresa a acortamiento o dilatación, respectivamente, muy por delante del ritmo normal anterior. En consecuencia, ciertas mujeres con estas afecciones darán lugar a un parto antes de las treinta y siete semanas de gestación, cuando el desarrollo fetal es incompleto. Por lo tanto, esto puede conducir a mortalidad y morbilidad neonatal. Para prolongar dicho embarazo, los médicos pueden colocar un cerclaje o un pesario cervical para sostener el cuello uterino. Con base principalmente en el consenso y la opinión de expertos, se recomienda que la colocación del cerclaje pueda ser beneficiosa si se descarta la infección intraamniótica. Se sugieren pautas similares para la colocación de pesarios. Para descartar una infección efectiva, se requiere un método de detección altamente sensible.

15 En la actualidad, la detección de bacterias en una muestra se basa en ensayos de cultivo, microscopía y reacción en cadena de la polimerasa (PCR) específica de especies bacterianas, que solo ofrecen sensibilidad de baja a moderada. Dicha sensibilidad podría mejorarse si podemos enfocarnos específicamente en los taxones bacterianos que son diferencialmente abundantes en el cuello uterino anormal (es decir, corto/dilatado), pero no en aquellos que son solo parte de la flora "normal" que reside en un cuello uterino "normal". Sin embargo, hasta la fecha, hay datos limitados sobre esos taxones, porque los taxones bacterianos en el cuello uterino anormal no se han perfilado sistemáticamente y se han comparado con aquellos en el cuello uterino "normal" de mujeres con parejas adecuadas. Dada la prevalencia y las implicaciones del parto prematuro, existe la necesidad de nuevos métodos para detectar con mayor precisión un mayor riesgo de un resultado adverso del embarazo en las mujeres, de modo que se puedan tomar medidas preventivas a tiempo para reducir o eliminar las posibilidades de parto prematuro o complicaciones neonatales. Esta invención satisface esta y otras necesidades relacionadas.

25 Breve resumen de la invención

La presente descripción se basa, en parte, en el descubrimiento de que el nivel de bacterias que pertenecen a un grupo específico de taxones bacterianos (por ejemplo, una especie o géneros bacterianos) en el cuello uterino de una mujer aumenta en correlación con la probabilidad de un resultado adverso del embarazo, como un parto prematuro a menos de 34 semanas o 37 semanas de edad gestacional, o una afección neonatal adversa, como un puntaje de Apgar de menos de 7 a 1 o 5 minutos, corioamnionitis, síndrome de dificultad respiratoria, displasia broncopulmonar, hemorragia intraventricular, sepsis neonatal y muerte neonatal dentro de los 7 días posteriores al nacimiento. Como tal, en un primer aspecto, la presente invención proporciona un método para determinar el riesgo de un embarazo o un resultado neonatal adverso para un sujeto femenino, por ejemplo, una mujer embarazada o una mujer no embarazada. El método de esta divulgación incluye los pasos de (a) detectar en una muestra biológica tomada del sujeto femenino el nivel de bacterias de al menos un taxón bacteriano seleccionado del grupo que consiste en *Megasphaera cerevisiae*, *Alloscardovia omnicolens*, *Ureaplasma urealyticum*, *Ureaplasma parvum*, *Atopobium vaginae*, *Parvibacter caecicola*, *Lactobacillus casei*, *Veillonella montpellierensis*, *Anaerococcus senegalensis*, *Bulleidia extracta*, *Mycoplasma hominis*, *Propionimicrobium lymphophilum*, bacterias no cultivadas que tienen una secuencia de nucleótidos de ARNr 16S con al menos 92% de identidad de secuencia con la secuencia de nucleótidos del número de acceso de GenBank No. JQ781443.1, *Corynebacterium pyruviciproducens*, *Megasphaera cerevisiae*, *Acidipila rosea*, *Murdochiella asaccharolytica*, bacterias no cultivadas que tienen una secuencia de nucleótidos ARNr 16S con al menos un 90% de identidad de secuencia con la secuencia de nucleótidos de GenBank Accession No. JF295520.1, *Howardella ureilytica*, *Actinobaculum schaalii*, *Peptoniphilus duerdenii*, *Fastidiosipila sanguinis*, *Sneathia sanguinegens*, *Parvimonas micra*, *Peptoniphilus lacrimalis*, un taxón bacteriano especificado por una secuencia de nucleótidos ARNr 16S en las Tablas 4A, 4C y 8, y un taxón bacteriano especificado por la clasificación de especies más cercana basada en la alineación de nucleótidos BLAST en las Tablas 4E, 4G y 9; y (b) determinar que el sujeto femenino tiene un mayor riesgo de un embarazo o un resultado neonatal adverso si el nivel de bacterias de al menos un género bacteriano es mayor que el de un nivel de control estándar.

50 En algunas realizaciones, el embarazo o el resultado neonatal adverso incluyen parto prematuro a <34 semanas, parto prematuro a <37 semanas, parto dentro de aproximadamente 1-196 días después de la toma de la muestra biológica, parto dentro de aproximadamente 1-196 días después de que se realiza una intervención clínica, una puntuación de Apgar a 1 minuto de <7, una puntuación de Apgar a 5 minutos de <7, corioamnionitis, síndrome de dificultad respiratoria, displasia broncopulmonar, hemorragia intraventricular, muerte neonatal dentro de los 7 días posteriores al nacimiento o sepsis neonatal. En algunos casos, el método incluye determinar que el sujeto femenino tiene un riesgo de dilatación cervical avanzada o acortamiento cervical prematuro si el nivel de bacterias que pertenecen al al menos un taxón bacteriano es mayor que un nivel de control estándar.

En las realizaciones de la invención, el método incluye detectar el nivel de *Parvimonas micra*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Peptoniphilus lacrimalis*, *Megasphaera cerevisiae*, *Parvibacter caecicola* en una muestra biológica tomada del sujeto femenino, y determinar que el sujeto femenino tiene un mayor riesgo de

un resultado adverso del embarazo, como el parto dentro de los 7 días posteriores a la intervención cervical (es decir, intervención de cerclaje/pesario), si el nivel total de la bacteria aumenta en comparación con un nivel de control estándar. En algunos casos, el método también incluye determinar que el sujeto femenino está en riesgo de tener una infección en la cavidad amniótica, la cavidad uterina, el cuello uterino o la vagina.

- 5 Se puede identificar una clase de *Megasphaera cerevisiae* que tiene una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del número de acceso de GenBank NR\_113307.1. En algunas realizaciones, se detecta que las bacterias del taxón *Megasphaera cerevisiae* tienen una secuencia de nucleótidos de ARNr 16S con al menos un 93% o 94% de identidad de secuencia con la secuencia del número de acceso GenBank NR\_113307.1. En algunas realizaciones, se detecta que las bacterias del taxón *Megasphaera cerevisiae* tienen una secuencia de nucleótidos de ARNr 16S que se muestra en la SEQ ID NO: 5.

- 10 La bacteria *Alloscardovia omnicoles* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del No. de acceso GenBank NR\_042583.1. En algunas realizaciones, se detecta que las bacterias del taxón *Alloscardovia omnicoles* tienen una secuencia genómica 16S ARNr con un 100% de identidad de secuencia con la secuencia de nucleótidos del número de acceso GenBank NR\_042583.1.

- 15 La bacteria *Ureaplasma urealyticum* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia genómica 16S ARNr del No. de acceso GenBank NR\_102836.1. En algunas realizaciones, se detecta que las bacterias del taxón *Ureaplasma urealyticum* tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 100% con la secuencia de nucleótidos del número de acceso GenBank NR\_102836.1. La bacteria *Ureaplasma parvum* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia genómica 16S ARNr del No. de acceso GenBank NR\_074176.1. En algunas realizaciones, se detecta que las bacterias del taxón *Ureaplasma parvum* tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 100% con la secuencia de nucleótidos del No. de acceso GenBank NR\_074176.1. En algunas realizaciones, se detecta que las bacterias del taxón *Ureaplasma parvum* tienen una secuencia de nucleótidos de ARNr 16S que se muestra en la SEQ ID NO: 7.

- 20 La bacteria *Atopobium vaginae* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del número de acceso GenBank NR\_117757.1. En algunas realizaciones, se detecta que las bacterias del taxón *Atopobium vaginae* tienen una secuencia de nucleótidos de ARNr 16S con un 97% de identidad de secuencia con la secuencia de nucleótidos del número de acceso GenBank NR\_117757.1. En algunas realizaciones, se detecta que las bacterias del taxón *Atopobium vaginae* tienen una secuencia de nucleótidos de ARNr 16S que se muestra en la SEQ ID NO: 8.

- 25 La bacteria *Parvibacter caecicola* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del No. de acceso de GenBank NR\_117374.1. En algunas realizaciones, se detecta que las bacterias del taxón *Parvibacter caecicola* tienen una secuencia de nucleótidos de ARNr 16S con al menos 92% de identidad de secuencia con la secuencia de nucleótidos del No. de acceso GenBank NR\_117374.1. En algunas realizaciones, se detecta que las bacterias del taxón *Parvibacter caecicola* tienen una secuencia de nucleótidos de ARNr 16S que se muestra en la SEQ ID NO: 9.

- 30 La bacteria *Lactobacillus casei* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del No. de acceso GenBank NR\_075032.1. En algunas realizaciones, se detecta que las bacterias del taxón *Lactobacillus casei* tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 100% con la secuencia de nucleótidos del número de acceso GenBank NR 075032.1.

- 35 La bacteria *Veillonella montpellierensis* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del No. de acceso GenBank NR 028839.1. En algunas realizaciones, se detecta que las bacterias del taxón *Veillonella montpellierensis* tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 100% con la secuencia de nucleótidos del No. de acceso GenBank NR 028839.1.

- 40 La bacteria *Anaerococcus senegalensis* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del No. de acceso de GenBank NR\_118220.1. En algunas realizaciones, se detecta que las bacterias del taxón *Anaerococcus senegalensis* tienen una secuencia de nucleótidos

de ARNr 16S con una identidad de secuencia del 100% con la secuencia de nucleótidos del No. de acceso GenBank NR\_118220.1.

5 Las bacterias *Bulleidia extracta* pueden identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del No. de acceso de GenBank NR\_028773.1. En algunas realizaciones, se detecta que las bacterias del taxón *Bulleidia extracta* tienen una secuencia de nucleótidos de ARNr 16S con un 97% de identidad de secuencia con la secuencia de nucleótidos del No. de acceso GenBank NR\_028773.1.

10 La bacteria *Mycoplasma hominis* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos 16S ARNr del número de acceso GenBank NR\_113679.1. En algunas realizaciones, se detecta que las bacterias del taxón *Mycoplasma hominis* tienen una secuencia genómica de ARNr 16S con una identidad de secuencia del 100% con la secuencia de nucleótidos del número de acceso GenBank NR\_113679.1.

15 La bacteria *Propionimicrobium lymphophilum* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del No. de acceso GenBank NR\_114337.1. En algunas realizaciones, se detecta que las bacterias del taxón *Propionimicrobium lymphophilum* tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 100% con la secuencia de nucleótidos del No. de acceso GenBank NR\_114337.1.

20 Una clase de bacterias no cultivadas relevantes para la presente descripción puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del No. de acceso de GenBank JQ781443.1. En algunas realizaciones, las bacterias de este taxón tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 92% con respecto a la secuencia de nucleótidos del No. de acceso de GenBank JQ781443.1.

25 Las bacterias *Corynebacterium pyruviciproducens* pueden identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del No. de acceso GenBank NR\_116569.1. En algunas realizaciones, se detecta que las bacterias del taxón *Corynebacterium pyruviciproducens* tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 100% con la secuencia de nucleótidos del No. de acceso GenBank NR\_116569.1.

30 Se puede identificar otra clase de *Megasphaera cerevisiae* que tiene una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del número de acceso de GenBank NR\_113307.1. En algunas realizaciones, se detecta que las bacterias del taxón *Megasphaera cerevisiae* tienen una secuencia de nucleótidos de ARNr 16S con al menos un 93% o 94% de identidad de secuencia con la secuencia del número de acceso GenBank NR\_113307.1. En algunas realizaciones, se detecta que las bacterias del taxón *Megasphaera cerevisiae* tienen una secuencia de nucleótidos de ARNr 16S que se muestra en la SEQ ID NO: 29.

35 La bacteria *Acidipila rosea* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos 16S ARNr del número de acceso GenBank NR\_113179.1. En algunas realizaciones, se detecta que las bacterias del taxón *Acidipila rosea* tienen una secuencia de nucleótidos de ARNr 16S con un 97% de identidad de secuencia con la secuencia de nucleótidos del número de acceso GenBank NR\_113179.1.

40 La bacteria *Murdochiella asaccharolytica* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del número de acceso GenBank NR\_116331.1. En algunas realizaciones, se detecta que las bacterias del taxón *Murdochiella asaccharolytica* tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 99% con respecto a la secuencia de nucleótidos del número de acceso GenBank NR\_116331.1.

45 Se puede identificar otra clase de bacterias no cultivadas relevantes para la presente descripción que tienen una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del número de acceso de GenBank JF295520.1. En algunas realizaciones, se detecta que las bacterias de este taxón tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 90% con respecto a la secuencia de nucleótidos del número de acceso de GenBank JF295520.1.

50 La bacteria *Howardella ureilytica* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del número de acceso de GenBank NR\_044022.2. En algunas realizaciones, se detecta que las bacterias del taxón *Howardella ureilytica* tienen una secuencia de nucleótidos de

ARNr 16S con una identidad de secuencia del 93% con respecto a la secuencia de nucleótidos del número de acceso GenBank NR\_044022.2.

5 Las bacterias *Actinobaculum schaalii* pueden identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos 16S ARNr del número de acceso GenBank NR\_116869.1. En algunas realizaciones, se detecta que las bacterias del taxón *Actinobaculum schaalii* tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 100% con la secuencia de nucleótidos del número de acceso GenBank NR\_116869.1.

10 La bacteria *Peptoniphilus duerdenii* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del número de acceso GenBank NR\_116346.1. En algunas realizaciones, se detecta que las bacterias del taxón *Peptoniphilus duerdenii* tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 100% con la secuencia de nucleótidos del número de acceso GenBank NR\_116346.1.

15 La bacteria *Fastidiosipila sanguinis* puede identificarse al tener una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del número de acceso GenBank NR\_042186.1. En algunas realizaciones, se detecta que las bacterias del taxón *Fastidiosipila sanguinis* tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 100% con la secuencia de nucleótidos del número de acceso de GenBank NR\_042186.1.

20 Se puede identificar que *Sneathia sanguinegens* tiene una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del número de acceso de GenBank AJ344093.1. En algunas realizaciones, se detecta que las bacterias del taxón *Sneathia sanguinegens* tienen una secuencia de nucleótidos de ARNr 16S con al menos 93% o 94% de identidad de secuencia con la secuencia de GenBank No. de acceso AJ344093.1.

25 Se puede identificar que *Parvimonas micra* tiene una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos de ARNr 16S del número de acceso de GenBank NR\_114338.1. En algunas realizaciones, se detecta que las bacterias del taxón *Parvimonas micra* tienen una secuencia de nucleótidos de ARNr 16S con al menos el 93%, por ejemplo, 93%, 94%, 95%, 96%, 97% o más, identidad de secuencia con la secuencia de GenBank Accession No. NR\_114338.1. En algunas realizaciones, se detecta que las bacterias del taxón *Parvimonas micra* tienen una secuencia de nucleótidos de ARNr 16S que se muestra en la SEQ ID NO: 43.

30 Se puede identificar que *Peptoniphilus lacrimalis* tiene una secuencia de nucleótidos de ARNr 16S con al menos 90%, por ejemplo, al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia de nucleótidos 16S ARNr de los números de acceso de GenBank AB971812.1 y NR\_041938.1. En algunas realizaciones, se detecta que las bacterias del taxón *Peptoniphilus lacrimalis* tienen una secuencia de nucleótidos de ARNr 16S con una identidad de secuencia del 100% con la secuencia de los números de acceso a GenBank AB971812.1 y NR\_041938.1. En algunas realizaciones, se detecta que las bacterias del taxón *Peptoniphilus lacrimalis* tienen una secuencia de nucleótidos de ARNr 16S que se muestra en SEQ ID NO: 44. En algunas realizaciones, el método incluye detectar al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 taxones bacterianos diferentes.

35 En algunas realizaciones, el sujeto femenino es una mujer embarazada de aproximadamente 13 semanas a aproximadamente 37 semanas de gestación. En otras realizaciones, la mujer embarazada tiene entre aproximadamente 13 y aproximadamente 25 semanas de gestación. En algunos casos, la mujer embarazada tiene un cuello uterino abierto prematuramente. La mujer puede correr el riesgo de tener una infección en la cavidad amniótica, la cavidad uterina, el cuello uterino o la vagina. En otras realizaciones, el sujeto femenino es una mujer no embarazada. En algunas realizaciones, la mujer no embarazada está planeando su futuro embarazo. En algunos casos, el sujeto femenino tiene antecedentes de parto prematuro, muerte fetal o aborto espontáneo. En algunos casos, el sujeto femenino planea recibir una intervención clínica, como una intervención de cerclage o suplementos de progesterona antes o después del embarazo.

40 En algunas realizaciones, el método también incluye extraer ácidos nucleicos de la muestra biológica tomada del sujeto femenino antes de realizar la etapa (a). En algunas realizaciones, la muestra es un hisopo cervical (que incluye una muestra de hisopo del orificio externo), un hisopo vaginal (que incluye una muestra de hisopo del fórnix), una muestra de orina, una muestra de líquido amniótico, una muestra de sangre materna (muestra de sangre entera materna), una muestra de suero materno, una muestra de plasma materno o una muestra de moco cervical. En algunas realizaciones, la muestra es un hisopo placentario, un hisopo umbilical o cualquier muestra tomada directa o indirectamente del sistema reproductivo. Esto incluye cualquier muestreo de la superficie del tracto reproductivo femenino a través de raspado, corte, enjuague, ducha vaginal, aplicación de una corriente de gas, un líquido, un vacío, una fuerza de succión, una forma de energía (por ejemplo, campo electrostático, LÁSER) o un gradiente de productos

químicos (por ejemplo, quimioatrayentes para inducir la quimiotaxis). En alguna realización, la muestra se toma directa o indirectamente del sistema gastrointestinal de la mujer embarazada, incluyendo un hisopo bucal, un hisopo de garganta, un hisopo anal, un hisopo rectal o una muestra de heces.

5 En algunas realizaciones, la etapa de detección incluye un ensayo de amplificación de polinucleótidos. En algunos casos, el ensayo de amplificación es un ensayo de reacción en cadena de la polimerasa (PCR). Opcionalmente, el ensayo de PCR puede ser un ensayo de PCR cuantitativo. En algunos casos, el paso de detección incluye la hibridación sonda/cebador específica de secuencia, que puede ocurrir en ausencia de amplificación de polinucleótidos. En otras realizaciones, la etapa de detección incluye la determinación de la secuencia de polinucleótidos, tal como, pero sin limitación, secuenciación paralela masiva.

10 Por ejemplo, si el nivel de bacterias que pertenecen a un taxón bacteriano que incluye *Megasphaera cerevisiae*, *Alloscardovia omnicoles*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Parvibacter caecicola*, *Lactobacillus casei*, *Veillonella montpellierensis*, *Anaerococcus senegalensis*, *Bulleidia extracta*, *Mycoplasma hominis*, *Propionimicrobium lymphophilum*, bacterias no cultivadas que tienen una secuencia de nucleótidos 16S ARNr con al menos 92% de identidad de secuencia con la secuencia de nucleótidos de GenBank No. de acceso JQ781443.1, *Corynebacterium pyruviciproducens*, *Megasphaera cerevisiae*, *Acidipila rosea*, *Murdochiella asaccharolytica*,  
 15 secuencia de bacterias no cultivadas con un nucleótido de secuencia 16S ARNr con al menos un 90% de identidad de secuencia con la secuencia de nucleótidos de GenBank No. de acceso JF295520.1, *Howardella ureilytica*, *Actinobaculum schaalii*, *Peptoniphilus duerdenii*, *Fastidiosipila sanguinis*, *Sneathia sanguinegens*, *Parvimonas micra*, *Peptoniphilus lacrimalis*, un taxón bacteriano especificado por una secuencia genómica 16S ARNr en las Tablas 4A, 4C y 8, un taxón bacteriano especificado por la clasificación de especies más cercana basada en la alineación de nucleótidos BLAST en las Tablas 4E, 4G y 9, o cualquier combinación de las mismas está elevada en comparación con un control estándar, la embarazada está en riesgo de tener un embarazo o un resultado neonatal adverso. Cuando se determina un aumento en el nivel de bacterias de uno o más de los taxones bacterianos seleccionados en comparación con el control estándar, indica que la mujer tiene un mayor riesgo de parto prematuro o de dar a luz a un niño con una complicación neonatal. Por ejemplo, cuando el nivel de bacterias de uno o más de los taxones, tales como *Megasphaera cerevisiae*, *Alloscardovia omnicoles*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Parvibacter caecicola*, *Lactobacillus casei*, *Veillonella montpellierensis*, *Anaerococcus senegalensis*, *Bulleidia extracta*, *Mycoplasma hominis*, *Propionimicrobium lymphophilum*, bacterias no cultivadas que tienen una secuencia de nucleótidos de ARNr 16S con al menos un 92% de identidad de secuencia con la secuencia de nucleótidos de GenBank No. de acceso JQ781443.1, *Corynebacterium pyruviciproducens*, *Megasphaera cerevisiae*, *Acidipila rosea*, *Murdochiella asaccharolytica*, bacterias no cultivadas que tienen una secuencia de nucleótidos de ARNr 16S con al menos un 90% de identidad de secuencia con la secuencia de nucleótidos del número de acceso de GenBank JF295520, *Howardella ureilytica*, *Actinobaculum schaalii*, *Peptoniphilus duerdenii*, *Fastidiosipila sanguinis*, *Sneathia sanguinegens*, *Parvimonas micra*, *Peptoniphilus lacrimalis*, un taxón bacteriano especificado por una secuencia genómica 16S ARNr en las Tablas 4A, 4C y 8, y un taxón bacteriano especificado por la clasificación de especies más cercana basada en la alineación de nucleótidos BLAST en las Tablas 4E, 4G y 9, en una muestra de hisopo cervical o vaginal de una mujer embarazada se detecta que es más alta en comparación según el control estándar, la mujer tiene una mayor probabilidad de tener un embarazo adverso y/o un resultado neonatal. Los posibles embarazos adversos o resultados neonatales incluyen, entre otros, parto prematuro a <34 semanas, parto prematuro a <37 semanas, puntaje de Apgar a 1 minuto de <7, puntaje de Apgar a 5 minutos de <7, corioamnionitis, síndrome de dificultad respiratoria, displasia broncopulmonar, hemorragia intraventricular, muerte neonatal dentro de los 7 días posteriores al nacimiento y/o sepsis neonatal. Además, si el sujeto femenino es una mujer embarazada, el posible embarazo adverso o los resultados neonatales incluyen además el parto dentro de un período de aproximadamente 1-196 días (por ejemplo, 1 día, 7 días, 14 días, 21 días, 28 días, 56 días, 84 días, 112 días, 140 días, 168 días o 196 días) después de la toma de la muestra biológica, parto dentro de un período de aproximadamente 1-196 días (por ejemplo, 1 día, 7 días, 14 días, 21 días, 28 días, 56 días, 84 días, 112 días, 140 días, 168 días o 196 días) después de que se realiza la intervención clínica.

El método proporcionado en este documento puede usarse para determinar que la mujer embarazada tiene un riesgo de tener dilatación cervical avanzada o acortamiento cervical avanzado. Por ejemplo, si el nivel de bacterias de al menos un taxón bacteriano, tales como *Megasphaera cerevisiae*, *Alloscardovia omnicoles*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Parvibacter caecicola*, *Lactobacillus casei*, *Veillonella montpellierensis*, *Anaerococcus senegalensis*, *Bulleidia extracta*, *Mycoplasma hominis*, *Propionimicrobium lymphophilum*, bacterias no cultivadas que tienen una secuencia de nucleótidos de ARNr 16S con al menos un 92% de identidad de secuencia con la secuencia de nucleótidos de GenBank Accession No. JQ781443.1, *Corynebacterium pyruviciproducens*, *Megasphaera cerevisiae*, *Acidipila rosea*, *Murdochiella asaccharolytica*, bacterias no cultivadas que tienen una secuencia de nucleótidos de ARNr 16S con al menos 90% de identidad de secuencia con la secuencia de nucleótidos de GenBank No. de acceso JF295520.1, *Howardella ureilytica*, *Actinobaculum schaalii*, *Peptoniphilus duerdenii*, *Fastidiosipila sanguinis*, *Sneathia sanguinegens*, *Parvimonas micra*, *Peptoniphilus lacrimalis*, un taxón bacteriano especificado por una secuencia de nucleótidos de ARNr 16S en las Tablas 4A, 4C y 8, y un taxón bacteriano especificado por la clasificación de especies más cercana basada en la alineación de nucleótidos BLAST en las Tablas 4E, 4G y 9 es mayor que la de un nivel de control estándar, la mujer tiene una mayor probabilidad de tener dilatación cervical avanzada o acortamiento cervical prematuro. Se predice que dicho sujeto exhibirá dilatación cervical de, por

ejemplo,  $\geq 2$  cm, sin trabajo de parto activo entre 13 semanas y hasta 37 semanas de gestación, o acortamiento cervical antes del trabajo de parto a término completo y el parto.

En algunos casos, el método puede incluir además detectar en la muestra biológica el nivel de bacterias de al menos un taxón bacteriano seleccionado del grupo que consiste en *Jonquetella anthropi*, *Aerococcus urinae*, un taxón bacteriano especificado por una secuencia de nucleótidos de 16S ARNr en Tablas 4B y 4D, y un taxón bacteriano especificado por la clasificación de especies más cercana basada en la alineación de nucleótidos BLAST en las Tablas 4F y 4H; y determinar que el sujeto femenino tiene un mayor riesgo de un resultado adverso del embarazo si el nivel de bacterias del al menos un taxón bacteriano es más bajo que el de un nivel de control estándar.

Si se predice que la mujer embarazada tendrá un embarazo o un resultado neonatal adverso, el método puede incluir repetir los pasos (a) y (b) en un momento posterior usando un tipo de muestra de la muestra biológica del sujeto femenino, en donde un aumento en el nivel de bacterias del al menos un taxón bacteriano en el momento posterior en comparación con el nivel determinado en el paso original (a) indica un mayor riesgo de tener un embarazo o un resultado neonatal adverso. Además, una vez que se indica que una mujer embarazada tiene un mayor riesgo de experimentar un embarazo o un resultado neonatal adverso, un médico puede proporcionar tratamiento para que la mujer minimice el riesgo de dicho resultado adverso. Por ejemplo, la mujer puede ser monitorizada de cerca durante el embarazo o ser transferida oportunamente a una unidad terciaria con cuidados intensivos neonatales.

En otro aspecto, la presente invención proporciona un kit para determinar el riesgo de tener un embarazo o un resultado neonatal adverso en un sujeto femenino. El kit de esta divulgación puede incluir (a) un control estándar que proporciona una muestra biológica tomada de una mujer embarazada y que contiene bacterias que pertenecen a al menos un taxón bacteriano seleccionado del grupo que consiste en *Megasphaera cerevisiae*, *Alloscardovia omnicoles*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Parvibacter caecicola*, *Lactobacillus casei*, *Veillonella montpellierensis*, *Anaerococcus senegalensis*, *Bulleidia extracta*, *Mycoplasma hominis*, *Propionimicrobium lymphophilum*, bacterias no cultivadas que tienen una secuencia de nucleótidos de ARNr 16S con al menos 92% de identidad de secuencia con la secuencia de nucleótidos de GenBank No. de acceso JQ781443.1, *Corynebacterium pyruviciproducens*, *Megasphaera cerevisiae*, *Acidipila rosea*, *Murdochiella asaccharolytica*, bacterias no cultivadas que tienen una secuencia de nucleótidos ARNr 16S con al menos un 90% de identidad de secuencia con la secuencia de nucleótidos de GenBank Accession No. JF295520.1, *Howardella ureilytica*, *Actinobaculum schaalii*, *Peptoniphilus duerdenii*, *Fastidiosipila sanguinis*, *Sneathia sanguinegens*, *Parvimonas micra*, *Peptoniphilus lacrimalis*, un taxón bacteriano especificado por una secuencia de nucleótidos de ARNr 16S en las Tablas 4A, 4C y 8; y un taxón bacteriano especificado por la clasificación de especies más cercana basada en la alineación de nucleótidos BLAST en las Tablas 4E, 4G y 9 y (b) uno o más agentes que identifican específica y cuantitativamente al menos un taxón bacteriano seleccionado del grupo que consiste en *Megasphaera cerevisiae*, *Alloscardovia omnicoles*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Parvibacter caecicola*, *Lactobacillus casei*, *Veillonella montpellierensis*, *Anaerococcus senegalensis*, *Bulleidia extracta*, *Mycoplasma hominis*, *Propionimicrobium lymphophilum*, bacteria no cultivada que tiene una secuencia de nucleótidos de ARNr 16S con al menos 92% de identidad de secuencia con la secuencia de nucleótidos de GenBank No. de acceso JQ781443.1, *Corynebacterium pyruviciproducens*, *Megasphaera cerevisiae*, *Acidipila rosea*, *Murdochiella asaccharolytica*, bacterias no cultivadas que tienen una secuencia de nucleótidos ARNr 16S con al menos un 90% de identidad de secuencia con la secuencia de nucleótidos de GenBank Accession No. JF295520.1, *Howardella ureilytica*, *Actinobaculum schaalii*, *Peptoniphilus duerdenii*, *Fastidiosipila sanguinis*, *Sneathia sanguinegens*, *Parvimonas micra*, *Peptoniphilus lacrimalis* y un taxón bacteriano especificado por una secuencia de nucleótidos de ARNr 16S en las Tablas 4A, 4C y 8, y un taxón bacteriano especificado por la clasificación de especies más cercana basada en la alineación de nucleótidos BLAST en las Tablas 4E, 4G y 9. En algunas realizaciones, el uno o más agentes pueden incluir uno o más pares de cebadores oligonucleotídicos que hibridan específicamente y amplifican un polinucleótido del al menos un género bacteriano en un ensayo de amplificación. En algunas realizaciones, el uno o más agentes pueden incluir además una sonda polinucleotídica que hibrida específicamente con una secuencia polinucleotídica del al menos un taxón bacteriano. En algunos casos, el kit incluye un manual de instrucciones.

En otro aspecto, la presente invención proporciona un método para determinar si una mujer embarazada tiene un mayor riesgo de tener dilatación cervical avanzada o acortamiento cervical prematuro. El método de esta divulgación incluye los pasos de (a) extraer ácidos nucleicos de una muestra biológica tomada del sujeto femenino; (b) detectar en los ácidos nucleicos el nivel de al menos un taxón bacteriano seleccionado del grupo que consiste en *Megasphaera cerevisiae*, *Alloscardovia omnicoles*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Parvibacter caecicola*, *Lactobacillus casei*, *Veillonella montpellierensis*, *Anaerococcus senegalensis*, *Bulleidia extracta*, *Mycoplasma hominis*, *Propionimicrobium lymphophilum*, bacterias no cultivadas que tienen una secuencia de nucleótidos 16S ARNr con al menos un 92% de identidad de secuencia con la secuencia de nucleótidos de GenBank Accession No JQ781443.1, *Corynebacterium pyruviciproducens*, *Megasphaera cerevisiae*, *Acidipila rosea*, *Murdochiella asaccharolytica*, bacterias no cultivadas que tienen una secuencia de nucleótidos de ARNr 16S con al menos un 90% de identidad de secuencia con la secuencia de nucleótidos del número de acceso de GenBank JF295520.1, *Howardella ureilytica*, *Actinobaculum schaalii*, *Peptoniphilus duerdenii*, *Fastidiosipila sanguinis*, *Sneathia sanguinegens*, *Parvimonas micra*, *Peptoniphilus lacrimalis*, un taxón especificado por una secuencia de nucleótidos de ARNr 16S en las Tablas 4A, 4C y 8, y un taxón bacteriano especificado por la clasificación de especies más cercana basada en la alineación de nucleótidos BLAST en las Tablas 4E, 4G y 9; y (c) determinar que el sujeto femenino tiene un mayor riesgo de tener dilatación cervical avanzada o acortamiento cervical prematuro si el nivel del al menos un

- taxón bacteriano es mayor que el de un nivel de control estándar. Por ejemplo, se predice que una mujer embarazada corre el riesgo de tener un cérvix dilatada prematuramente y/o un cuello uterino acortado prematuramente, si se determina que tiene un mayor nivel de bacterias de uno o más de los taxones bacterianos seleccionados en su cuello uterino o vagina. Si ella tiene más bacterias del grupo, como *Megasphaera cerevisiae*, *Alloscardovia omnicolens*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Parvibacter caecicola*, *Lactobacillus casei*, *Veillonella montpellierensis*, *Anaerococcus senegalensis*, *Bulleidia extracta*, *Mycoplasma hominis*, *Propionimicrobium lymphophilum*, bacterias no cultivadas que tienen una secuencia de nucleótidos 16S ARNr con al menos 92% de identidad de secuencia con la secuencia de nucleótidos de GenBank Accession No. JQ781443.1, *Corynebacterium pyruviciproducens*, *Megasphaera cerevisiae*, *Acidipila rosea*, *Murdochiella asaccharolytica*, bacterias no cultivadas que tienen una secuencia de nucleótidos 16S ARNr con al menos 90% de secuencia de nucleótidos ARNr con Identidad de secuencia con la secuencia de nucleótidos de GenBank No. de acceso JF295520.1, *Howardella ureilytica*, *Actinobaculum schaalii*, *Peptoniphilus duerdenii*, *Fastidiosipila sanguinis*, *Sneathia sanguinegens*, *Parvimonas micra*, *Peptoniphilus lacrimalis*, un taxón bacteriano especificado por una secuencia de nucleótidos de ARNr 16S en las Tablas 4A, 4C y 8, un taxón bacteriano especificado por la clasificación de especies más cercana basada en la alineación de nucleótidos BLAST en las Tablas 4E, 4G y 9, o cualquier combinación de las mismas en su cuello uterino o vagina en, por ejemplo, aproximadamente 13 a aproximadamente 25 semanas de gestación, es probable que ella tenga un cérvix dilatada o un cuello uterino acortado antes de aproximadamente 34 a aproximadamente 37 semanas de gestación. Se pueden detectar al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 taxones bacterianos diferentes en la muestra.
- En algunas realizaciones, la muestra biológica es un hisopo cervical, un hisopo vaginal, una muestra de orina, una muestra de líquido amniótico, una muestra de sangre materna, una muestra de suero materno, una muestra de plasma materno, una muestra de moco cervical, un hisopo de placenta, un hisopo del cordón umbilical o cualquier muestra tomada directa o indirectamente del sistema reproductivo o del sistema gastrointestinal. El paso de detección puede ser un ensayo de amplificación de polinucleótidos, un ensayo que implica la determinación de la secuencia de polinucleótidos o un ensayo que implica la hibridación sonda/cebador específico de la secuencia. En algunos casos, el ensayo de amplificación es un ensayo de reacción en cadena de la polimerasa (PCR). Opcionalmente, el ensayo de PCR es un ensayo de PCR cuantitativo o un ensayo de PCR de transcriptasa reversa.

#### Breve descripción de los dibujos

- Las Figuras 1A-1E muestran la asociación entre los resultados clínicos, como el parto prematuro espontáneo (sPTB) después de la intervención y el parto a término (TB) después de la intervención, y los taxones bacterianos. La Figura 1A muestra los resultados clínicos de 25 pacientes con insuficiencia cervical (IC) después de la intervención. Cada columna representa a un paciente (P1-P25) en orden ascendente de edad gestacional (GA) al momento del parto. Rectángulos negros: P1-P10 que resultan en "parto prematuro espontáneo (sPTB) <34 semanas después de la intervención". Rectángulos blancos: P11-P25 resultando en "parto a término (TB) ≥37 semanas después de la intervención". Latencia, el intervalo entre el tratamiento y el parto. SDR, síndrome de dificultad respiratoria. DBP, displasia broncopulmonar. VIH, hemorragia intraventricular. ROP, retinopatía del prematuro. Muerte neonatal, muerte dentro de los 28 días posteriores al parto. Y Sí n, no; -, no determinado. Las Figuras 1B y 1C muestran los 10 taxones bacterianos más abundantes en los cérvices CI "sPTB después de la intervención" y "TB después de la intervención", respectivamente. Los valores mostrados son el log<sub>10</sub> de los valores de abundancia (recuentos de lectura de secuenciación normalizada, el método de Escala de suma acumulativa (CSS)) de cada taxón bacteriano (fila) en la muestra de hisopo cervical de cada paciente (columna). Los recuentos de lectura normalizados se transforman en log (abundancia) como se indica en la Figura 1E. Cada fila representa una unidad taxonómica operativa (Otu) formada por secuencias de agrupamiento de identidad ≥97%. Cada Otu se clasifica taxonómicamente a nivel de género utilizando el Clasificador de ARNr Bayesiano Naive del Ribosomal Database Project (RDP) (Versión 2.9, septiembre de 2014, RDP 16S ARNr conjunto de entrenamiento 10). Los *Lactobacillus* se comparan con la base de datos 16S ARNr (GenBank) utilizando BLAST (puntaje más alto) y MOLE-BLAST (la mejor alineación múltiple de coincidencias BLAST) para obtener la información de la especie. La Figura 1D muestra los taxones bacterianos diferencialmente abundantes entre los grupos "sPTB después de la intervención" (n=10) y "TB después de la intervención" (n=15) (prueba de suma de rango de Mann-Whitney). Siete taxones permanecen como diferencialmente abundantes después del ajuste para múltiples pruebas por el método de False Discovery Rate (FDR) (p <0.05 y valor q <0.05, es decir, FDR <5%, en asterisco). Los últimos 6 taxones con p <0.01 se seleccionan adicionalmente para calcular una puntuación resumida, es decir, la abundancia total log (base 10) del valor de 6 taxones seleccionados (LA6) para cada muestra de hisopo cervical. La abundancia total es la suma aritmética de las abundancias de los taxones seleccionados en escala común (lineal).
- Las figuras 2A-2C proporcionan análisis estadístico y valores de LA6 para los dos grupos, por ejemplo, pacientes con parto prematuro espontáneo (sPTB) después de la intervención y pacientes con parto a término (TB) después de la intervención. La figura 2A proporciona valores LA6 (la abundancia total de los 6 taxones bacterianos seleccionados en escala logarítmica (base 10)) en dos grupos de pacientes con insuficiencia cervical (IC) que reciben intervención clínica pero que dan como resultado resultados diferentes. Se tomaron muestras de hisopo cervical para medir el LA6 de pacientes con IC antes de la intervención de cerclaje/pesario. Después del tratamiento, 10 pacientes resultaron en parto prematuro espontáneo <34 semanas (el grupo "sPTB después de la intervención", círculos), y 15 pacientes resultaron en parto a término ≥37 semanas (el grupo "TB después del tratamiento", triángulos). Los 6 taxones se seleccionaron en función de sus abundancias significativamente diferentes (p <0.01) entre estos dos grupos en los

datos de secuenciación masivamente paralelos (Figura ID). Las líneas horizontales largas y cortas de la barra de error se dibujan en el rango medio e intercuartil, respectivamente. La figura 2B muestra la proporción de embarazos no culminados en parto en diferentes períodos de gestación en pacientes con IC con LA6 >1.15 (LA6-positivo) versus aquellas con LA6 ≤ 1.15 (LA6-negativo). Pacientes con LA6 positivos entregados antes de la intervención clínica que los pacientes con LA6 (edad gestacional media al parto de 23.7 semanas vs. 38.4 semanas; intervalo de confianza del 95%, 20.6 semanas - 25.4 semanas vs. 38.0 semanas - 38.7 semanas; Chi-cuadrado, 32.352; df, 1; prueba de Logrank, p <0.0001; razón de riesgo, 6.24; intervalo de confianza del 95%, 1.50 a 25.9). Los intervalos de confianza del 95% de los porcentajes de embarazos no culminados en parto para pacientes LA6-positivos y LA6-negativas se muestran como líneas alrededor de la línea continua en negrita y la línea de puntos en negrita, respectivamente. La figura 2C muestra la proporción de embarazos no culminados en parto en diferentes días después del tratamiento en pacientes con IC LA6 positivo versus LA6 negativo. Pacientes con LA6 positivos entregados por un período más corto después de la intervención que sus homólogos con LA6 negativos (mediana número de días entre la intervención y el parto, 10 días frente a 126 días; intervalo de confianza del 95%, 8 días - 32 días frente a 112 días - 134 días; Chi-cuadrado, 32.520; df, 1; Prueba de Logrank, p <0.00001; razón de riesgo, 6.34; intervalo de confianza del 95%, 1.51 a 26.6).

#### Definiciones

En esta divulgación, el término "o" se emplea generalmente en su sentido que incluye "y/o" a menos que el contenido indique claramente lo contrario.

El término "embarazo adverso o resultado neonatal" se refiere a una afección que reduce la posibilidad de dar a luz/dar a luz a un bebé sano. Ejemplos no limitantes de un resultado adverso del embarazo incluyen abortos espontáneos múltiples en el primer trimestre, pérdida del embarazo en el segundo trimestre, parto prematuro (por ejemplo, espontáneo o indicado), preclampsia prematura, clampsia prematura, restricción del crecimiento fetal, desprendimiento de placenta, muerte fetal/muerte fetal, defectos de nacimiento, puntaje de Apgar a 1 minuto de <7, puntaje de Apgar a 5 minutos de <7, corioamnionitis clínica, corioamnionitis patológica, síndrome de dificultad respiratoria neonatal, displasia broncopulmonar neonatal, sepsis neonatal, hemorragia intraventricular neonatal, etc.

El término "insuficiencia cervical" se refiere a una afección del cuello uterino, tal como debilitamiento o dilatación avanzada del cuello uterino que puede conducir a la pérdida del embarazo o al nacimiento en el segundo trimestre. La debilidad cervical, el acortamiento cervical prematuro, la dilatación cervical prematura o avanzada, el traumatismo cervical, una anomalía estructural del cuello uterino o cualquier combinación de estos pueden contribuir a la insuficiencia cervical. Las intervenciones clínicas para controlar la insuficiencia cervical incluyen, pero no se limitan a, suplementos de progesterona, cerclaje cervical y pesario cervical.

El término "taxón bacteriano" se refiere a la taxonomía, es decir, la clasificación de bacterias basada en el rango. La clasificación biológica jerárquica incluye vida, dominio, reino, filum, clase, orden, familia, género y especie.

El término "muestra biológica" o "muestra" incluye secciones de tejidos tales como muestras de biopsia y autopsia, y secciones congeladas tomadas con fines histológicos, o formas procesadas de cualquiera de tales muestras. Las muestras biológicas incluyen un hisopo cervical, un hisopo vaginal, un hisopo uterino, fracciones de sangre y sangre o productos (por ejemplo, suero, plasma, plaquetas, glóbulos rojos y similares), esputo o saliva, tejido linfático y de la lengua, células cultivadas, por ejemplo, cultivos primarios, explantes y células transformadas, heces, orina, un tejido de biopsia, etc. Una muestra biológica se obtiene típicamente de un organismo eucariota, que puede ser un mamífero, un primate y un sujeto femenino humano.

El término "biopsia" se refiere al proceso de extracción de una muestra de tejido para evaluación diagnóstica o pronóstica, y a la muestra de tejido en sí. Cualquier técnica de biopsia conocida en la técnica se puede aplicar a los métodos de diagnóstico y pronóstico de la presente descripción. La técnica de biopsia aplicada dependerá del tipo de tejido a evaluar (por ejemplo, cuello uterino, vagina, lengua, colon, próstata, riñón, vejiga, ganglios linfáticos, hígado, médula ósea, células sanguíneas, tejido estomacal, etc.) entre otros factores. Las técnicas representativas de biopsia incluyen, pero no se limitan a, una biopsia con hisopo, biopsia por escisión, biopsia por incisión, biopsia con aguja, biopsia quirúrgica y biopsia de médula ósea y pueden comprender colonoscopia. Los expertos en la materia conocen una amplia gama de técnicas de biopsia que elegirán entre ellas y las implementarán con una mínima experimentación.

En esta divulgación, el término molécula de ácido nucleico "aislada" significa una molécula de ácido nucleico que está separada de otras moléculas de ácido nucleico que normalmente están asociadas con la molécula de ácido nucleico aislada. Por lo tanto, una molécula de ácido nucleico "aislada" incluye, sin limitación, una molécula de ácido nucleico que está libre de secuencias de nucleótidos que flanquean naturalmente uno o ambos extremos del ácido nucleico en el genoma del organismo del que se deriva el ácido nucleico aislado (por ejemplo, un ADNc o un fragmento de ADN genómico producido por PCR o digestión con endonucleasa de restricción). Dicha molécula de ácido nucleico aislada se puede introducir en un vector (por ejemplo, un vector de clonación o un vector de expresión) por conveniencia de manipulación o para generar una molécula de ácido nucleico de fusión. Además, una molécula de ácido nucleico aislada puede incluir una molécula de ácido nucleico diseñada, tal como una molécula de ácido nucleico recombinante o sintética. Una molécula de ácido nucleico que existe entre cientos y millones de otras moléculas de ácido nucleico dentro de, por ejemplo, una biblioteca de ácido nucleico (por ejemplo, una biblioteca de ADNc o genómica) o un gel

(por ejemplo, agarosa o poliacrilamina) que contiene ADN genómico digerido por restricción, no es un ácido nucleico "aislado".

El término "ácido nucleico", "nucleótido o" polinucleótido "se refiere a ácidos desoxirribonucleicos (ADN) o ácidos ribonucleicos (ARN) y sus polímeros en forma monocatenaria o bicatenaria. A menos que esté específicamente limitado, el término abarca ácidos nucleicos que contienen análogos conocidos de nucleótidos naturales que tienen propiedades de unión similares al ácido nucleico de referencia y se metabolizan de manera similar a los nucleótidos naturales. A menos que se indique lo contrario, una secuencia particular de ácido nucleico también abarca implícitamente variantes modificadas de manera conservadora (por ejemplo, sustituciones de codones degenerados), alelos, ortólogos, polimorfismos de un solo nucleótido (SNP) y secuencias complementarias, así como la secuencia indicada explícitamente. Específicamente, las sustituciones de codones degenerados se pueden lograr mediante la generación de secuencias en las que la tercera posición de uno o más codones seleccionados (o todos) se sustituye con residuos de base mixta y/o desoxiinosina (Batzler et al., *Nucleic Acid Res.* 19: 5081 (1991); Ohtsuka et al., *J. Biol. Chem* 260: 2605-2608 (1985); y Rossolini et al., *Mol. Célula. Probes* 8: 91-98 (1994)). El término ácido nucleico se usa indistintamente con gen, ADNc y ARNm codificado por un gen.

Los términos "polipéptido", "péptido" y "proteína" se usan indistintamente en el presente documento para referirse a un polímero de residuos de aminoácidos. Los términos se aplican a los polímeros de aminoácidos en los que uno o más residuos de aminoácidos son un mimético químico artificial de un aminoácido natural correspondiente, así como a los polímeros de aminoácidos naturales y a los polímeros de aminoácidos no naturales. Como se usa en el presente documento, los términos abarcan cadenas de aminoácidos de cualquier longitud, incluidas proteínas de longitud completa (es decir, antígenos), en donde los residuos de aminoácidos están unidos por enlaces peptídicos covalentes.

El término "aminoácido" se refiere a aminoácidos naturales y sintéticos, así como a análogos de aminoácidos y miméticos de aminoácidos que funcionan de manera similar a los aminoácidos naturales. Los aminoácidos de origen natural son los codificados por el código genético, así como los aminoácidos que luego se modifican, por ejemplo, hidroxiprolina,  $\gamma$ -carboxiglutamato y O-fosfoserina. Para los fines de esta solicitud, los análogos de aminoácidos se refieren a compuestos que tienen la misma estructura química básica que un aminoácido natural, es decir, un carbono que está unido a un hidrógeno, un grupo carboxilo, un grupo amino y un grupo R, por ejemplo, homoserina, norleucina, metionina sulfóxido, metionina metil sulfonio. Dichos análogos tienen grupos R modificados (por ejemplo, norleucina) o esqueletos de péptidos modificados, pero conservan la misma estructura química básica que un aminoácido natural. Para los propósitos de esta solicitud, los miméticos de aminoácidos se refieren a compuestos químicos que tienen una estructura que es diferente de la estructura química general de un aminoácido, pero que funciona de manera similar a un aminoácido natural.

Los aminoácidos pueden incluir aquellos que tienen D-quiralidad no natural, como se describe en el documento WO01/12654, que puede mejorar la estabilidad (por ejemplo, vida media), biodisponibilidad y otras características de un polipéptido que comprende uno o más de tales D-aminoácidos. En algunos casos, uno o más, y potencialmente todos los aminoácidos de un polipéptido terapéutico tienen D-quiralidad.

Se puede hacer referencia a los aminoácidos en este documento por los símbolos de tres letras comúnmente conocidos o por los símbolos de una letra recomendados por la IUPAC-IUB Biochemical Nomenclature Commission. Los nucleótidos, del mismo modo, pueden ser referidos por sus códigos de una sola letra comúnmente aceptados.

Los términos "idéntico" o porcentaje de identidad, en el contexto de la descripción de dos o más secuencias de polinucleótidos o aminoácidos, se refieren a dos o más secuencias o subsecuencias que son iguales o tienen un porcentaje específico de residuos de aminoácidos o los nucleótidos que son iguales (por ejemplo, una variante de una proteína bacteriana o interés utilizado en el método de esta invención (por ejemplo, para predecir resultados adversos del embarazo) tiene al menos 80% de identidad de secuencia, preferiblemente 85%, 90%, 91%, 92%, 93, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad, a una secuencia de referencia, por ejemplo, una proteína bacteriana de interés de tipo salvaje correspondiente), cuando se compara y se alinea para obtener una correspondencia máxima sobre una ventana de comparación, o región designada o medida, usando uno de los siguientes algoritmos de comparación de secuencia o mediante alineación manual e inspección visual. Dichas secuencias se dice que son "sustancialmente idénticas". Con respecto a las secuencias de polinucleótidos, esta definición también se refiere al complemento de una secuencia de prueba. Preferiblemente, la identidad existe sobre una región que tiene al menos aproximadamente 50 aminoácidos o nucleótidos de longitud, o más preferiblemente sobre una región que tiene una longitud de 75-100 aminoácidos o nucleótidos.

Para la comparación de secuencias, típicamente una secuencia actúa como una secuencia de referencia, con la cual se comparan las secuencias de prueba. Cuando se usa un algoritmo de comparación de secuencias, las secuencias de prueba y referencia se ingresan en una computadora, se designan las coordenadas de subsecuencia, si es necesario, y se designan los parámetros del programa del algoritmo de secuencia. Se pueden usar los parámetros predeterminados del programa o se pueden designar parámetros alternativos. Luego, el algoritmo de comparación de secuencias calcula el porcentaje de identidades de secuencia para las secuencias de prueba en relación con la secuencia de referencia, en función de los parámetros del programa. Para la comparación de secuencias de ácidos nucleicos y proteínas, se utilizan los algoritmos BLAST y BLAST 2.0 y los parámetros predeterminados que se analizan a continuación.

Una "ventana de comparación", como se usa en el presente documento, incluye referencia a un segmento de cualquiera de la cantidad de posiciones contiguas seleccionadas del grupo que consiste de 20 a 600, usualmente de aproximadamente 50 a aproximadamente 200, más usualmente de aproximadamente 100 a aproximadamente 150 en donde se puede comparar una secuencia con una secuencia de referencia del mismo número de posiciones contiguas después de que las dos secuencias estén alineadas de manera óptima. Los métodos de alineación de secuencias para comparación son bien conocidos en la técnica. La alineación óptima de secuencias para comparación se puede realizar, por ejemplo, mediante el algoritmo de homología local de Smith & Waterman, *Adv. Appl. Matemáticas*. 2: 482 (1981), mediante el algoritmo de alineación de homología de Needleman & Wunsch, *J. Mol. Biol.* 48: 443 (1970), mediante la búsqueda del método de similitud de Pearson and Lipman, *Proc. Nat'l. Acad. Sci. USA* 85: 2444 (1988), mediante implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Paquete de Software de Genética de Wisconsin, Genetics Computer Group, 575 Science Dr., Madison, WI), o mediante alineación manual e inspección visual (ver, por ejemplo, *Current Protocols in Molecular Biology* (Ausubel et al., eds. 1995 supplement)).

Ejemplos de algoritmos que son adecuados para determinar el porcentaje de identidad de secuencia y la similitud de secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul et al., (1990) *J. Mol. Biol.* 215: 403-410 and Altschul et al. (1977) *Nucleic Acids Res.* 25: 3389-3402, respectivamente. El software para realizar análisis BLAST está disponible públicamente en el sitio web del National Center for Biotechnology Information, [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov). El algoritmo implica primero identificar pares de secuencias de alta puntuación (HSP) mediante la identificación de palabras cortas de longitud  $W$  en la secuencia de consulta, que coinciden o satisfacen una puntuación de umbral de valor positivo  $T$  cuando se alinean con una palabra de la misma longitud en una secuencia de base de datos.  $T$  se conoce como el umbral de puntuación de palabras vecinas (Altschul et al., supra). Estas coincidencias iniciales de palabras vecinas actúan como semillas para iniciar búsquedas para encontrar HSP más largos que las contengan. Luego, las coincidencias de palabras se extienden en ambas direcciones a lo largo de cada secuencia hasta que se pueda aumentar la puntuación de alineación acumulativa. Los puntajes acumulativos se calculan utilizando, para secuencias de nucleótidos, los parámetros  $M$  (puntaje de recompensa para un par de residuos coincidentes; siempre  $>0$ ) y  $N$  (puntaje de penalización para residuos que no coinciden; siempre  $<0$ ). Para las secuencias de aminoácidos, se usa una matriz de puntuación para calcular la puntuación acumulativa. La extensión de los aciertos de palabras en cada dirección se detiene cuando: la puntuación de alineación acumulativa cae en la cantidad  $X$  de su valor máximo alcanzado; el puntaje acumulativo va a cero o menos, debido a la acumulación de uno o más alineamientos de residuos de puntuación negativa; o se alcanza el final de cualquiera de las secuencias. Los parámetros del algoritmo BLAST  $W$ ,  $T$  y  $X$  determinan la sensibilidad y la velocidad de la alineación. El programa BLASTN (para secuencias de nucleótidos) utiliza por defecto un tamaño de palabra ( $W$ ) de 28, una expectativa ( $E$ ) de 10,  $M=1$ ,  $N=-2$  y una comparación de ambas cadenas. Para las secuencias de aminoácidos, el programa BLASTP utiliza por defecto un tamaño de palabra ( $W$ ) de 3, una expectativa ( $E$ ) de 10 y la matriz de puntuación BLOSUM62 (ver, por ejemplo, Henikoff and Henikoff, *Proc. Nat'l. Acad. Sci. Estados Unidos* 89: 10915 (1989)).

El algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin and Altschul, *Proc. Nat'l. Acad. Sci. USA*, 90: 5873-5787 (1993)). Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña ( $P(N)$ ), que proporciona una indicación de la probabilidad de que se produzca una coincidencia entre dos secuencias de nucleótidos o aminoácidos por casualidad. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de suma más pequeña en una comparación del ácido nucleico de prueba con el ácido nucleico de referencia es menor que aproximadamente 0.2, más preferiblemente menor que aproximadamente 0.01, y lo más preferiblemente menor que aproximadamente 0.001.

Una indicación de que dos secuencias de ácido nucleico o polipéptidos son sustancialmente idénticas es que el polipéptido codificado por el primer ácido nucleico es reactivo cruzado inmunológicamente con los anticuerpos generados contra el polipéptido codificado por el segundo ácido nucleico, como se describe a continuación. Por lo tanto, un polipéptido es típicamente sustancialmente idéntico a un segundo polipéptido, por ejemplo, donde los dos péptidos difieren solo por sustituciones conservadoras. Otra indicación de que dos secuencias de ácido nucleico son sustancialmente idénticas es que las dos moléculas o sus complementos se hibridan entre sí en condiciones estrictas, como se describe a continuación. Otra indicación más de que dos secuencias de ácido nucleico son sustancialmente idénticas es que pueden usarse los mismos cebadores para amplificar la secuencia.

Los términos "condiciones de hibridación restrictivas" y "alta restricción" se refieren a condiciones bajo las cuales una sonda se hibridará con su subsecuencia diana, típicamente en una mezcla compleja de ácidos nucleicos, pero no con otras secuencias. Las condiciones estrictas dependen de la secuencia y serán diferentes en diferentes circunstancias. Las secuencias más largas se hibridan específicamente a temperaturas más altas. Una guía extensa para la hibridación de ácidos nucleicos se encuentra en Tijssen, *Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993) y será fácilmente comprendido por los expertos en la materia. Generalmente, las condiciones restrictivas se seleccionan para que sean aproximadamente  $5-10^{\circ}\text{C}$  más bajas que el punto de fusión térmica ( $T_m$ ) para la secuencia específica a un pH de fuerza iónica definido. La  $T_m$  es la temperatura (bajo una fuerza iónica definida, pH y concentración nucleica) a la que el 50% de las sondas complementarias al objetivo se hibridan con la secuencia objetivo en equilibrio (ya que las secuencias objetivo están presentes en exceso, a  $T_m$ , 50% de las sondas están ocupadas en equilibrio). También se pueden lograr condiciones estrictas con la adición de agentes desestabilizadores como la formamida. Para la hibridación selectiva o específica, una señal positiva es al menos dos veces el fondo,

preferiblemente 10 veces la hibridación de fondo. Las condiciones de hibridación estrictas ejemplares pueden ser las siguientes: 50% de formamida, 5 x SSC y 1% de SDS, incubando a 42°C, o 5 x SSC, 1% de SDS, incubando a 65°C, con lavado en 0.2 x SSC y SDS al 0.1% a 65°C.

5 Los ácidos nucleicos que no se hibridan entre sí en condiciones restrictivas siguen siendo sustancialmente idénticos si los polipéptidos que codifican son sustancialmente idénticos. Esto ocurre, por ejemplo, cuando se crea una copia de un ácido nucleico utilizando la máxima degeneración de codones permitida por el código genético. En tales casos, los ácidos nucleicos típicamente hibridan en condiciones de hibridación moderadamente estrictas. Ejemplos de "condiciones de hibridación moderadamente estrictas" incluyen una hibridación en un regulador de formamida al 40%, NaCl 1 M, SDS al 1% a 37°C, y un lavado en 1x SSC a 45°C. Una hibridación positiva es al menos dos veces de fondo. Los expertos en la materia reconocerán fácilmente que pueden utilizarse condiciones de hibridación y lavado 10 alternativas para proporcionar condiciones de restrictividad similar. Se proporcionan pautas adicionales para determinar los parámetros de hibridación en numerosas referencias, por ejemplo, Current Protocols in Molecular Biology, ed. Ausubel et al.

15 La expresión "se une específicamente" cuando se usa en el contexto de referirse a una secuencia polinucleotídica que forma un complejo bicatenario con otra secuencia polinucleotídica describe la "hibridación polinucleotídica" basada en el emparejamiento de bases de Watson-Crick, como se proporciona en la definición para el término "método de hibridación de polinucleótidos".

20 Como se usa en esta solicitud, un "aumento" o una "disminución" se refiere a un cambio detectable positivo o negativo en la cantidad de un control de comparación, por ejemplo, un control estándar establecido (tal como un nivel de expresión promedio de un ARNm bacteriano o proteína que se encuentra en el tejido cervical o vaginal normal de un sujeto femenino de control embarazada). Un aumento es un cambio positivo que generalmente es al menos 10%, o al menos 20%, o 50%, o 100%, y puede ser tan alto como al menos 2 veces o al menos 5 veces o incluso 10 veces del valor de control. De manera similar, una disminución es un cambio negativo que típicamente es al menos 10%, o al menos 20%, 30% o 50%, o incluso tan alto como al menos 80% o 90% del valor de control. Otros términos que indican 25 cambios cuantitativos o diferencias desde una base comparativa, como "más", "menos", "más alto" y "más bajo", se usan en esta solicitud de la misma manera que se describe anteriormente. En contraste, el término "sustancialmente igual" o "sustancialmente falta de cambio" indica poco o ningún cambio en la cantidad del valor de control estándar, típicamente dentro de  $\pm 10\%$  del control estándar, o dentro de  $\pm 5\%$ , 2%, o incluso menos variación del control estándar.

30 Un "método de hibridación de polinucleótidos", como se usa en el presente documento, se refiere a un método para detectar la presencia y/o cantidad de una secuencia de polinucleótidos predeterminada basada en su capacidad para formar pares de bases Watson-Crick, en condiciones de hibridación apropiadas, con una sonda polinucleotídica de una secuencia conocida. Ejemplos de tales métodos de hibridación incluyen transferencia Southern, transferencia Northern e hibridación in situ.

35 Los "cebadores", como se usan en el presente documento, se refieren a oligonucleótidos que se pueden usar en un método de amplificación, tal como una reacción en cadena de la polimerasa (PCR), para amplificar una secuencia de nucleótidos basada en la secuencia de polinucleótidos correspondiente a un gen de interés, por ejemplo, el ADNc o secuencia genómica para un gen bacteriano específico o una porción del mismo. Típicamente, al menos uno de los cebadores de PCR para la amplificación de una secuencia de polinucleótidos es específico de secuencia para esa secuencia de polinucleótidos. La longitud exacta del cebador dependerá de muchos factores, incluida la temperatura, 40 la fuente del cebador y el método utilizado. Por ejemplo, para aplicaciones de diagnóstico y pronóstico, dependiendo de la complejidad de la secuencia diana, el cebador oligonucleotídico generalmente contiene al menos 10, 15, 20, 25 o más nucleótidos, aunque puede contener menos nucleótidos o más nucleótidos. Los factores implicados en la determinación de la longitud apropiada del cebador son fácilmente conocidos por un experto en la materia. En esta divulgación, el término "par de cebadores" significa un par de cebadores que hibridan con cadenas opuestas de una molécula de ADN diana o con regiones del ADN diana que flanquean una secuencia de nucleótidos para amplificar. 45 En esta divulgación, el término "sitio de cebador" significa el área del ADN objetivo u otro ácido nucleico con el que se hibrida un cebador.

50 Una "etiqueta", "etiqueta detectable" o "unidad estructural detectable" es una composición detectable por medios espectroscópicos, fotoquímicos, bioquímicos, inmunoquímicos, químicos u otros medios físicos. Por ejemplo, las etiquetas útiles incluyen  $^{32}\text{P}$ , colorantes fluorescentes, reactivos densos en electrones, enzimas (por ejemplo, como se usan comúnmente en un ELISA), biotina, digoxigenina o haptenos y proteínas que pueden hacerse detectables, por ejemplo, incorporando un componente radiactivo en el péptido o usado para detectar anticuerpos específicamente reactivos con el péptido. Normalmente, un marcador detectable se une a una sonda o una molécula con características de unión definidas (por ejemplo, un polipéptido con una especificidad de unión conocida o un polinucleótido), para 55 permitir la presencia de la sonda (y por lo tanto su objetivo de unión) a ser fácilmente detectable.

60 El "control estándar" tal como se usa en el presente documento se refiere a una cantidad o concentración predeterminada de bacterias que pertenecen a un género bacteriano específico, un polinucleótido bacteriano o un polipéptido bacteriano que está presente en una muestra de tejido normal establecida, por ejemplo, una muestra de tejido cervical normal. El valor de control estándar es adecuado para el uso de un método de la presente invención, para servir como base para comparar la cantidad de un género bacteriano específico, ARNm o proteína que está

presente en una muestra de prueba. Una muestra establecida que sirve como control estándar proporciona una cantidad promedio del género bacteriano, ARNm o proteína que es típica para una muestra de tejido cervical de un ser humano promedio y sano con, por ejemplo, un cuello uterino cerrado o un cuello uterino de longitud normal, como se define convencionalmente. Un valor de control estándar puede variar dependiendo de la naturaleza de la muestra, la forma de recolección de la muestra y otros factores como el sexo, la edad, el origen étnico de los sujetos (y en el caso de las mujeres embarazadas, la edad gestacional) según a quien se establece dicho valor de control.

El término "promedio", como se usa en el contexto de la descripción de una humana que está embarazada y no está en riesgo de tener un embarazo o un resultado neonatal adverso, como se define convencionalmente, se refiere a ciertas características, especialmente la cantidad de bacterias de uno o taxones bacterianos más específicos, encontrados en el cuello uterino de la persona que son representativos de un grupo de humanos embarazadas seleccionados al azar que están libres de cualquier riesgo de tener un embarazo o un resultado neonatal adverso. Este grupo seleccionado debe comprender un número suficiente de humanos de modo que la cantidad promedio de bacterias de los taxones específicos en el cuello uterino entre estos individuos refleje, con una precisión razonable, la cantidad correspondiente de bacterias de los taxones en la población general de humanas sanas, normales, embarazadas. Además, el grupo seleccionado de humanos embarazadas generalmente tiene una edad gestacional similar a la de un sujeto femenino cuya muestra de tejido cervical se analiza para indicar un riesgo de tener un embarazo o un resultado neonatal adverso. Además, también se consideran otros factores como la edad, el estado de recibir el mismo tipo de intervención o similar (por ejemplo, intervención de pesario/cerclaje), origen étnico, historial médico y, de preferencia, coinciden estrechamente entre los perfiles del sujeto femenino de prueba y el grupo seleccionado. de individuos que establecen el valor "promedio".

El término "cantidad" o "nivel" como se usa en esta solicitud se refiere a la cantidad de un taxón bacteriano de interés, un polinucleótido bacteriano de interés o un polipéptido bacteriano de interés presente en una muestra. Dicha cantidad puede expresarse en términos absolutos, es decir, la cantidad total del taxón, polinucleótido o polipéptido bacteriano en la muestra, o en términos relativos, es decir, la concentración del taxón, polinucleótido o polipéptido bacteriano en la muestra.

El término "sujeto" incluye individuos que buscan atención médica debido a un riesgo potencial de tener un resultado adverso del embarazo o un resultado neonatal, por ejemplo, cualquier persona embarazada. Los sujetos también incluyen individuos que han tenido un embarazo o un resultado neonatal adverso durante un embarazo anterior.

Descripción detallada de la invención

## I. Introducción

La invención se basa, en parte, en el descubrimiento de taxones bacterianos diferencialmente abundantes en las muestras de hisopo cervical de mujeres con dilatación cervical avanzada/acortamiento cervical, en comparación con aquellas en muestras apropiadamente controladas de mujeres emparejadas apropiadamente sin la correspondiente condición. El nivel elevado de bacterias de taxones particulares (por ejemplo, *Megasphaera cerevisiae*, *Alloscardovia omnicoles*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Parvibacter ciecicola*, *Lactobacillus casei*, *Veillonella montpellierensis*, *Anaerococcus senegalensis*, *Bulleidia extracta*, *Mycoplasma hominis*, *Propionimicrobium lymphophilum*, bacterias no cultivadas que tienen al menos 92% de identidad de secuencia con la secuencia de nucleótidos de GenBank Accession No. JQ781443.1, *Corynebacterium pyruviciproducens*, *Megasphaera cerevisiae*, *Acidipila rosea*, *Murdochiella asaccharolytica*, bacterias no cultivadas que tienen al menos 90% de identidad de secuencia con la secuencia de nucleótidos de GenBank Accession No. JF295520.1, *Howardella ureilytica*, *Actinobaculum schaalii*, *Peptoniphilus duerdenii*, *Fastidiosipila sanguinis*, *Sneathia sanguinegens*, *Parvimonas micra*, *Peptoniphilus lacrimalis*, un taxón bacteriano especificado por una secuencia de nucleótidos de ARNr 16S en las Tablas 4A, 4C y 8, y un taxón bacteriano especificado por la clasificación de especies más cercana basada en la alineación de nucleótidos BLAST en las Tablas 4E, 4G y 9) también es predictivo de un riesgo de tener un embarazo adverso o resultado neonatal, como parto prematuro (por ejemplo, parto prematuro espontáneo) <34 semanas, parto prematuro (por ejemplo, parto prematuro espontáneo) <37 semanas, parto dentro de un período posterior a la toma de la muestra biológica (por ejemplo, 1 día, 7 días, 14 días, 21 días, 28 días, 56 días, 84 días, 112 días, 140 días, 168 días, 196 días), parto dentro de un período posterior a la realización de la intervención clínica (por ejemplo, 1 día, 7 días, 14 días, 21 días, 28 días, 56 días, 84 días, 112 días, 140 días, 168 días, 196 días), una puntuación de Apgar a 1 minuto <7, una puntuación de Apgar a 5 minutos <7, corioamnionitis (por ejemplo, clínica o patológica), síndrome de dificultad respiratoria, displasia broncopulmonar, hemorragia intraventricular, muerte neonatal dentro de los 7 días posteriores al nacimiento y sepsis neonatal .

## II Metodología general

La práctica de esta invención utiliza técnicas de rutina en el campo de la biología molecular. Los textos básicos que describen los métodos generales de uso en esta invención incluyen Sambrook and Russell, *Molecular Cloning, A Laboratory Manual* (3rd ed. 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel et al., eds., 1994)).

Para los ácidos nucleicos, los tamaños se dan en kilobases (kb) o pares de bases (pb). Estas son estimaciones derivadas de electroforesis en gel de agarosa o acrilamida, de ácidos nucleicos secuenciados o de secuencias de ADN publicadas. Para las proteínas, los tamaños se dan en kilodaltons (kDa) o números de residuos de aminoácidos. Los tamaños de las proteínas se estiman a partir de la electroforesis en gel, de proteínas secuenciadas, de secuencias de aminoácidos derivadas o de secuencias de proteínas publicadas.

Los oligonucleótidos que no están disponibles comercialmente pueden sintetizarse químicamente, por ejemplo, de acuerdo con el método de triéster de fosforamidita en fase sólida descrito por primera vez por Beaucage and Caruthers, *Tetrahedron Lett.* 22: 1859-1862 (1981), usando un sintetizador automatizado, como se describe en Van Devanter et al., *Nucleic Acids Res.* 12: 6159-6168 (1984). La purificación de oligonucleótidos se realiza utilizando cualquier estrategia reconocida en la técnica, por ejemplo, electroforesis en gel de acrilamida nativa o cromatografía líquida de alto rendimiento de intercambio aniónico (HPLC) como se describe en Pearson and Reanier, *J. Chrom.* 255: 137-149 (1983).

### III. Adquisición de muestras de tejido y análisis de taxones bacterianos

La presente invención se refiere a medir la cantidad de bacterias de un taxón bacteriano específico que se encuentra en el cuello uterino o la vagina de una mujer embarazada, especialmente en una muestra de hisopo cervical o vaginal, como un medio para evaluar el riesgo de tener un resultado adverso del embarazo, o resultado neonatal, tal como la labor de parto prematuro y el alumbramiento prematuro. Por lo tanto, los primeros pasos antes de practicar esta invención son obtener una muestra de tejido cervical o vaginal de un sujeto femenino de prueba, de tal manera que se puedan analizar los ácidos nucleicos, por ejemplo, ARN o ADN, contenidos en la muestra. En algunas realizaciones, la cantidad de bacterias de un taxón de bacterias específico encontradas en el cuello uterino o la vagina de una mujer embarazada puede representarse por la cantidad de bacterias específicas en una muestra biológica que no proviene del cuello uterino o la vagina.

#### A. Adquisición y preparación de muestras biológicas.

Una muestra biológica, tal como tejido cervical o vaginal, moco cervical, líquido amniótico o sangre materna se obtiene de una persona para ser analizada o controlada usando un método de la presente invención. La recolección de células epiteliales cervicales o vaginales, moco cervical, líquido amniótico o sangre materna (por ejemplo, sangre entera, suero materno y/o plasma materno) de un individuo se realiza de acuerdo con el protocolo estándar que generalmente siguen los hospitales o clínicas, como durante un examen cervical. Se recolecta una cantidad apropiada de epitelio cervical o vaginal, células raspadas, moco y/o líquido biológico y se puede almacenar de acuerdo con los procedimientos estándar antes de la preparación posterior.

El análisis de las bacterias encontradas en la muestra de una paciente embarazada según la presente invención puede realizarse usando, por ejemplo, células, tejidos, mucosas o fluidos encontrados en la muestra. Los métodos para preparar muestras de células, tejidos o fluidos para la extracción de ácido nucleico son bien conocidos entre los expertos en la materia. Por ejemplo, la muestra de mucosa cervical o vaginal de un sujeto femenino puede tratarse de tal manera que se pueda analizar el ADN o ARN bacteriano en la muestra.

#### B. Extracción y cuantificación de ADN.

Existen numerosos métodos para extraer ADN bacteriano de una muestra biológica. Los métodos para extraer ADN de una muestra biológica son bien conocidos y se practican habitualmente en el arte de la biología molecular, véase, por ejemplo, Sambrook and Russell, *supra*. La contaminación por ARN debe eliminarse para evitar interferencias con el análisis de ADN. El tratamiento previo de la muestra biológica con regulador de lisis y enzimas, incluidas la mutanolisina y la proteinasa K, también se puede utilizar antes de la extracción. Los métodos para detectar el ADN objetivo incluyen análisis por PCR, análisis cuantitativo con marcado de fluorescencia o análisis de transferencia Southern. El ADN objetivo puede ser el gen que codifica el ARN ribosómico 16S (el gen 16S ARNr) u otros genes o secuencias genómicas de interés que posee un taxón bacteriano específico.

Una variedad de métodos de amplificación de polinucleótidos están bien establecidos y se usan con frecuencia en la investigación. Por ejemplo, los métodos generales de reacción en cadena de la polimerasa (PCR) para la amplificación de la secuencia de polinucleótidos son bien conocidos en la técnica y, por lo tanto, no se describen en detalle aquí. Para una revisión de los métodos, protocolos y principios de PCR en el diseño de cebadores, ver, por ejemplo, Innis, et al., *PCR Protocols: A Guide to Methods and Applications*, Academic Press, Inc. NY, 1990. Los reactivos y protocolos de PCR también son disponible de proveedores comerciales, como Roche Molecular Systems.

Aunque la amplificación por PCR se usa típicamente en la práctica de la presente invención, un experto en la materia reconocerá que la amplificación de la secuencia genómica relevante puede lograrse mediante cualquier método conocido, tal como la reacción en cadena de la ligasa (LCR), la transcripción, amplificación mediada y replicación de secuencia autosostenida o amplificación basada en secuencia de ácido nucleico (NASBA), cada una de las cuales proporciona amplificación suficiente. La tecnología de ADN ramificado desarrollada más recientemente también puede usarse para determinar cuantitativamente la cantidad de marcadores de ARNm bacterianos específicos. Para una descripción detallada de la amplificación de la señal de ADN ramificado para la cuantificación directa de secuencias de ácido nucleico en muestras clínicas, véase, por ejemplo, Nolte, *Adv. Clin. Chem.* 33: 201-235, 1998.

Las técnicas para la determinación de la secuencia de polinucleótidos también están bien establecidas y se practican ampliamente en el campo de investigación relevante. Por ejemplo, los principios básicos y las técnicas generales para la secuenciación de polinucleótidos se describen en varios informes de investigación y tratados sobre biología molecular y genética recombinante, como Wallace et al., *Supra*; Sambrook and Russell, *supra*, y Ausubel et al., *supra*.  
 5 Los métodos de secuenciación de ADN practicados rutinariamente en laboratorios de investigación, ya sean manuales o automatizados, pueden usarse para practicar la presente invención. Los medios adicionales adecuados para detectar una secuencia de polinucleótidos para practicar los métodos de la presente invención incluyen, pero no se limitan a, espectrometría de masas, extensión de cebadores, hibridación de polinucleótidos, PCR en tiempo real, análisis de curvas de fusión, análisis de fusión de alta resolución, análisis de heterodúplex, secuenciación en paralelos masiva (por ejemplo, secuenciación de próxima generación) y electroforesis.

C. Extracción y cuantificación de ARN

Un experto en la materia reconoce que existen numerosos métodos para extraer ARN bacteriano de una muestra biológica. Se pueden seguir los métodos generales de preparación de ARN (por ejemplo, descritos por Sambrook and Russell, *Molecular Cloning: A Laboratory Manual* 3d ed., 2001); diversos reactivos o kits disponibles comercialmente, como el reactivo Trizol (Invitrogen, Carlsbad, CA), los kits de ARNm directo Oligotex (Qiagen, Valencia, CA), los minikits RNeasy (Qiagen, Hilden, Alemania) y PolyATtract® Series 9600™ (Promega, Madison, WI), también se puede utilizar para obtener ARNm de una muestra biológica de un sujeto femenino de prueba. También se pueden usar combinaciones de más de uno de estos métodos.

Es esencial que todo el ADN contaminante se elimine de las preparaciones de ARN. Por lo tanto, se debe utilizar un manejo cuidadoso de las muestras, un tratamiento exhaustivo con DNasa y controles negativos adecuados en los pasos de amplificación y cuantificación.

D. Determinación cuantitativa basada en PCR del nivel de ARN

Una vez que se extrae el ARN de una muestra, se puede cuantificar la cantidad de cualquier transcripción de ARN de interés que expresen las bacterias de un taxón bacteriano específico. Por ejemplo, la cantidad de ARN ribosómico 16S (ARNr) para un taxón bacteriano en particular, como, entre otros, *Megasphaera cerevisiae*, *Alloscardovia omnicoles*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Parvibacter caecicola*, *Lactobacillus casei*, *Veillonella montpellierensis*, *Anaerococcus senegalensis*, *Bulleidia extracta*, *Mycoplasma hominis*, *Propionimicrobium lymphophilum*, bacterias no cultivadas que tienen al menos un 92% de identidad de secuencia con la secuencia de nucleótidos de GenBank Accession No. JQ781443.1, *Corynebacterium pyruviciproducens*, *Megasphaera cerevisiae*, *Acidipila rosea*, *Murdochiella asaccharolytica*, bacterias no cultivadas que tienen al menos un 90% de identidad de secuencia con la secuencia de nucleótidos del número de acceso de GenBank JF295520.1, *Howardella ureilytica*, *Actinobaculum schaalii*, *Peptoniphilus duerdenii*, *Fastidiosipila sanguinis*, *Sneathia sanguinegens*, *Parvimonas micra*, *Peptoniphilus lacrimalis*, un taxón bacteriano especificado por un secuencia de nucleótidos Tablas 4A, 4C y 8, y se puede detectar y medir un taxón bacteriano especificado por la clasificación de especies más cercana basada en la alineación de nucleótidos BLAST en las Tablas 4E, 4G y 9. El método preferido para determinar el nivel de transcripción de ARN es un método basado en la amplificación, por ejemplo, por reacción en cadena de la polimerasa (PCR), especialmente la reacción en cadena de la polimerasa por transcripción reversa (RT-PCR).

Antes de la etapa de amplificación, se debe sintetizar una copia de ADN (ADNc) de una transcripción de ARN bacteriano de interés. Esto se logra mediante la transcripción reversa, que puede llevarse a cabo como un paso separado, o en una reacción en cadena homogénea de la polimerasa de la transcripción reversa (RT-PCR), una modificación de la reacción en cadena de la polimerasa para amplificar el ARN. Romero y Rotbart describen los métodos adecuados para la amplificación por PCR de ácidos ribonucleicos en *Diagnostic Molecular Biology: Principles and Applications* págs. 401-406; Persing et al., Eds., Mayo Foundation, Rochester, MN, 1993; Egger et al., *J. Clin. Microbiol* 33: 1442-1447, 1995; y Patente de los Estados Unidos No. 5.075.212.

Los métodos generales de PCR son bien conocidos en la técnica y, por lo tanto, no se describen en detalle aquí. Para una revisión de los métodos, protocolos y principios de PCR en el diseño de cebadores, ver, por ejemplo, Innis, et al., *PCR Protocols: A Guide to Methods and Applications*, Academic Press, Inc. N.Y., 1990. Los reactivos y protocolos de PCR también son disponible de proveedores comerciales, como Roche Molecular Systems.

La PCR se lleva a cabo generalmente como un proceso automatizado con una enzima termoestable. En este proceso, la temperatura de la mezcla de reacción se cicla a través de una región desnaturalizante, una región de fusión de cebador y una región de reacción de extensión automáticamente. Las máquinas específicamente adaptadas para este propósito están disponibles comercialmente.

Aunque la amplificación por PCR del ARN objetivo se usa típicamente en la práctica de la presente invención. Sin embargo, un experto en la materia reconocerá que la amplificación de estas especies de ARN bacterianas en la muestra se puede lograr mediante cualquier método conocido, como la reacción en cadena de la ligasa (LCR), la amplificación mediada por la transcripción y la replicación de secuencia autosostenida o amplificación basada en secuencia de ácido nucleico (NASBA), cada una de las cuales proporciona amplificación suficiente. La tecnología de

ADN ramificado desarrollada más recientemente también se puede utilizar para determinar cuantitativamente la cantidad de marcadores de ARN bacterianos específicos. Para una revisión de la amplificación de la señal de ADN ramificado para la cuantificación directa de secuencias de ácido nucleico en muestras clínicas, ver Nolte, *Adv. Clin. Chem* 33: 201-235, 1998.

#### 5 E. Otros métodos cuantitativos para ADN y ARN

Las transcripciones bacterianas de ADN o ARN de interés también se pueden detectar utilizando otras técnicas estándar, bien conocidas por los expertos en la materia. Aunque la etapa de detección está precedida típicamente por una etapa de amplificación, la amplificación no es necesaria en los métodos de la invención. Por ejemplo, el ADN o el ARN pueden identificarse por fraccionamiento de tamaño (por ejemplo, electroforesis en gel), ya sea que se proceda o no por una etapa de amplificación. Después de analizar una muestra en un gel de agarosa o poliacrilamida y marcar con bromuro de etidio de acuerdo con técnicas bien conocidas (ver, por ejemplo, Sambrook and Russell, *supra*), la presencia de una banda del mismo tamaño que la comparación estándar es una indicación de la presencia de un ADN o ARN objetivo, cuya cantidad puede compararse con el control en función de la intensidad de la banda. Alternativamente, las sondas de oligonucleótidos específicas para el ADN o ARN de interés pueden usarse para detectar la presencia de tales especies de ADN o ARN e indicar la cantidad de ADN o ARN en comparación con la comparación estándar, en función de la intensidad de la señal impartida por la sonda.

La hibridación de sonda específica de secuencia es un método bien conocido para detectar un ácido nucleico particular que comprende otras especies de ácidos nucleicos. En condiciones de hibridación suficientemente estrictas, las sondas hibridan específicamente solo con secuencias sustancialmente complementarias. La restrictividad de las condiciones de hibridación se puede relajar para tolerar cantidades variables de desajuste de secuencia.

Una serie de formatos de hibridación bien conocidos en la técnica, que incluyen, pero no se limitan a, ensayos de hibridación en fase de solución, fase sólida o en fase mixta. Los siguientes artículos proporcionan una visión general de los diversos formatos de ensayo de hibridación: Singer et al., *Biotechniques*, 4: 230, 1986; Haase et al., *Methods in Virology*, pp. 189-226, 1984; Wilkinson, *Hibridación in situ*, Ed. Wilkinson, IRL Press, Oxford University Press, Oxford; y Hames and Higgins eds., *Nucleic Acid Hybridization: A Practical Approach*, IRL Press, 1987.

Los complejos de hibridación se detectan según técnicas bien conocidas. Las sondas de ácido nucleico capaces de hibridarse específicamente con un ácido nucleico diana, es decir, el ARN o el ADN amplificado, pueden marcarse mediante cualquiera de los varios métodos utilizados típicamente para detectar la presencia de ácidos nucleicos hibridados. Un método común de detección es el uso de autorradiografía usando sondas etiquetadas con  $^3\text{H}$ ,  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$  o  $^{32}\text{P}$ , o similares. La elección del isótopo radiactivo depende de las preferencias de investigación debido a la facilidad de síntesis, la estabilidad y la vida media de los isótopos seleccionados. Otros marcadores incluyen compuestos (por ejemplo, biotina y digoxigenina), que se unen a antiligandos o anticuerpos marcados con fluoróforos, agentes quimioluminiscentes y enzimas. Alternativamente, las sondas se pueden conjugar directamente con marcadores tales como fluoróforos, agentes quimioluminiscentes o enzimas. La elección de la etiqueta depende de la sensibilidad requerida, la facilidad de conjugación con la sonda, los requisitos de estabilidad y la instrumentación disponible.

Las sondas y los cebadores necesarios para practicar la presente invención pueden sintetizarse y marcarse usando técnicas bien conocidas. Los oligonucleótidos utilizados como sondas y cebadores pueden sintetizarse químicamente de acuerdo con el método del triéster de fosoramidita en fase sólida descrito por primera vez por Beaucage and Caruthers, *Tetrahedron Letts.*, 22: 1859-1862, 1981, utilizando un sintetizador automático, como se describe en Needham-VanDevanter et al., *Nucleic Acids Res.* 12: 6159-6168, 1984. La purificación de oligonucleótidos es por electroforesis en gel de acrilamida nativa o por HPLC de intercambio aniónico como se describe en Pearson and Regnier, *J. Chrom.*, 255: 137-149, 1983.

#### F. Amplificación y análisis de secuencia

Se puede realizar una reacción de amplificación antes del análisis de secuencia. Una variedad de métodos de amplificación de polinucleótidos están bien establecidos y se utilizan con frecuencia en la investigación. Por ejemplo, los métodos generales de reacción en cadena de la polimerasa (PCR) para la amplificación de la secuencia de polinucleótidos son bien conocidos en la técnica y, por lo tanto, no se describen en detalle aquí. Para una revisión de los métodos, protocolos y principios de PCR en el diseño de cebadores, ver, por ejemplo, Innis, et al., *PCR Protocols: A Guide to Methods and Applications*, Academic Press, Inc. NY, 1990. Los reactivos y protocolos de PCR también son disponible de proveedores comerciales, como Roche Molecular Systems.

Las técnicas para la determinación de la secuencia de polinucleótidos también están bien establecidas y se practican ampliamente en el campo de investigación relevante. Por ejemplo, los principios básicos y las técnicas generales para la secuenciación de polinucleótidos se describen en varios informes de investigación y tratados sobre biología molecular y genética recombinante, como Wallace et al., *supra*; Sambrook and Russell, *supra*, and Ausubel et al., *supra*. Los métodos de secuenciación de ADN practicados rutinariamente en laboratorios de investigación, ya sean manuales o automatizados, pueden usarse para practicar la presente invención. Los medios adicionales adecuados para detectar una secuencia de polinucleótidos para practicar los métodos de la presente invención incluyen, pero no

se limitan a, espectrometría de masas, extensión de cebadores, hibridación de polinucleótidos, PCR en tiempo real, análisis de curvas de fusión, análisis de fusión de alta resolución, análisis de heterodúplex, secuenciación en paralelos masiva y electroforesis.

#### IV. Establecer un control estándar

5 Para establecer un control estándar para un tipo de muestra particular (por ejemplo, hisopo cervical o hisopo vaginal) para practicar el método de esta invención, se selecciona primero un grupo de mujeres embarazadas sanas, mujeres embarazadas que no corren el riesgo de tener un efecto adverso en el resultado del embarazo o el resultado neonatal, o las mujeres embarazadas que luego se confirman que dan a luz dentro del marco de tiempo normal de su embarazo, como se define convencionalmente. Por ejemplo, el grupo puede incluir un grupo de mujeres embarazadas que han tenido un parto y parto a término. En algunas realizaciones, el grupo de mujeres embarazadas puede haber tenido un parto y parto a término después de la intervención clínica, como cerclaje o pesario. Estas personas se encuentran dentro de los parámetros apropiados, si corresponde, con el fin de detectar y/o controlar el riesgo de resultados adversos del embarazo utilizando los métodos de la presente invención. Por ejemplo, los individuos pueden tener una edad gestacional similar y un estado de salud comparable. Opcionalmente, los individuos tienen una edad similar o un origen étnico similar.

El tiempo de parto normal de las personas seleccionadas se confirmará más adelante, y cualquiera de las personas seleccionadas que resulte dar a luz antes o después del plazo de parto normal será excluido del grupo para proporcionar datos como "control estándar".

20 El estado saludable de los individuos seleccionados se confirma mediante métodos bien establecidos y empleados habitualmente, que incluyen, pero no se limitan a, el examen físico general de los individuos y la revisión general de su historial médico.

Además, el grupo seleccionado de individuos sanos debe ser de un tamaño razonable, de modo que la cantidad/concentración promedio de bacterias de uno o más taxones bacterianos en la muestra de tejido cervical obtenida del grupo pueda considerarse razonablemente representativa de la nivel normal o promedio entre la población general de mujeres embarazadas sanas. Preferiblemente, el grupo seleccionado comprende al menos 10 sujetos humanos embarazadas.

Una vez que se establece un valor promedio para las bacterias de uno o más taxones en base a los valores individuales encontrados en cada sujeto del grupo de control saludable seleccionado, este valor o perfil promedio o medio o representativo se considera un control estándar. También se determina una desviación estándar durante el mismo proceso. En algunos casos, se pueden establecer controles estándar separados para grupos definidos por separado que tienen características distintas como edad, edad gestacional o origen étnico.

#### V. Predicción del riesgo de embarazo adverso o resultado neonatal según la puntuación del microbio cervical

Usando los métodos descritos en este documento, se puede predecir si una mujer embarazada tiene una probabilidad de tener un resultado adverso del embarazo, por ejemplo, un parto prematuro espontáneo o un parto antes de las 34 semanas en el embarazo actual. Del mismo modo, se puede predecir si un sujeto femenino no embarazada tiene una probabilidad de tener un resultado adverso del embarazo en el futuro embarazo. En algunas realizaciones, para cada sujeto, los niveles de abundancia de bacterias que pertenecen a al menos 1, por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más taxones bacterianos se pueden utilizar para calcular una puntuación microbiana cervical. El puntaje del sujeto femenino puede compararse con un valor de corte establecido por los puntajes del grupo de control estándar, y usarse para determinar si es probable que la embarazada tenga un resultado adverso del embarazo. En algunas realizaciones, las bacterias pertenecen al menos a 3, por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más taxones de bacterias. En algunos casos, las bacterias incluyen *Sneathia sanguinegens*, *Parvimonas micra*, *Ureaplasma urealyticum* (o *Ureaplasma parvum*), *Atopobium vaginae*, *Peptoniphilus lacrimalis*, *Megasphaera cerevisiae* y *Paravibacter caecicola*. En otros casos, los niveles de abundancia de *Parvimonas micra*, *Ureaplasma urealyticum* (o *Ureaplasma parvum*), *Atopobium vaginae*, *Peptoniphilus lacrimalis*, *Megasphaera cerevisiae* y *Paravibacter caecicola* se utilizan para calcular la puntuación de microbios cervicales para el sujeto femenino. En algunas realizaciones, cada valor de  $\log_{10}$  (abundancia) para el taxón bacteriano seleccionado puede transformarse en la escala lineal y luego sumarse, después de lo cual el valor total se transforma con  $\log$  y se expresa como una puntuación de microbio cervical en la escala  $\log_{10}$ . En algunos casos, una puntuación de microbios cervicales mayor de 1.15 puede indicar que la embarazada está en riesgo de tener un parto prematuro espontáneo. En algunos casos, la embarazada está en riesgo de tener un parto prematuro espontáneo después de la intervención, como cerclaje o pesario. En algunas realizaciones, una puntuación cervical de más de 1.15 puede indicar que el sujeto femenino embarazada tiene una mayor probabilidad de dar a luz en menos de aproximadamente 37.0 semanas, por ejemplo, aproximadamente 36.5 semanas, aproximadamente 36.0 semanas, aproximadamente 35.0 semanas, aproximadamente 34.0 semanas, aproximadamente 33.0 semanas, aproximadamente 32.0 semanas, aproximadamente 31.0 semanas, aproximadamente 30.0 semanas, aproximadamente 29.0 semanas, aproximadamente 28.0 semanas, aproximadamente 27.0 semanas, aproximadamente 26.0 semanas, aproximadamente 25.0 semanas, aproximadamente 24.0 semanas, aproximadamente 23.0 semanas, aproximadamente 22.0 semanas, aproximadamente 21.0 semanas, o menos.

En algunas realizaciones, para cada muestra biológica, el nivel de abundancia de un taxón bacteriano se determina mediante un ensayo de PCR cuantitativa o un ensayo de hibridación específico de secuencia dirigido a secuencias de polinucleótidos específicas para el taxón en cuestión. En otras realizaciones, para cada muestra biológica, el nivel de abundancia de un taxón bacteriano se determina mediante secuenciación paralela masiva de las secuencias del gen marcador que sirven como proxy del taxón respectivo. En algunos casos, el nivel de abundancia de un taxón bacteriano en una muestra son los recuentos de lectura normalizados de la secuencia del gen del ARN ribosómico 16S (ARNr) que representa el taxón en cuestión. En algunos casos, los recuentos de lectura normalizados se calculan a partir de los recuentos de lectura sin procesar utilizando métodos de normalización establecidos para minimizar la variación técnica entre muestras, como una profundidad de secuencia diferente por muestra o un tamaño de biblioteca diferente por muestra. En algunos casos, el nivel de abundancia de un taxón bacteriano se expresa en la escala  $\log_{10}$ .

En algunas realizaciones, la abundancia de un grupo seleccionado de taxones bacterianos (por ejemplo, taxones que son significativamente más abundantes en las cérvix dilatadas que las cérvix cerradas) se combinan por adición o multiplicación para dar la puntuación microbiana cervical, como se ilustra en el valor LA6 en el Ejemplo 2. En otras realizaciones, la puntuación se calcula en 3 pasos: (a) los niveles de abundancia de un grupo sub-seleccionado de taxones bacterianos (por ejemplo, taxones que son significativamente más abundantes en las cérvix dilatadas que los cerrados cervices) se combinan por adición o multiplicación; (b) los niveles de abundancia de otro grupo de taxones subseleccionados (por ejemplo, taxones que son significativamente más abundantes en las cérvix cerradas que en las cérvix dilatadas) también se combinan por adición o multiplicación; y (c) la puntuación del microbio cervical se calcula restando o dividiendo entre la suma o producto del primer grupo de taxones seleccionado y la suma o producto del segundo grupo de taxones subseleccionado. En otras realizaciones, la puntuación del microbio cervical se calcula en función del número de tipos de taxones bacterianos seleccionados que están presentes en una muestra, como se ilustra en la prueba DIBT1, la prueba DIBT2 y la prueba DIBT3 en el Ejemplo 1. En las pruebas DIBT1, DIBT2 y DIBT3, la puntuación del microbio cervical es el número de taxones bacterianos seleccionados, y un sujeto femenino con una puntuación mayor o igual a 1 indica un mayor riesgo de embarazo adverso o resultado neonatal. En otras realizaciones, la puntuación del microbioma cervical se calcula en base a la clasificación de taxones bacterianos seleccionados entre todos los taxones detectados en una muestra.

#### VI. Kits

La invención proporciona kits para practicar los métodos descritos en este documento para evaluar el nivel de bacterias de taxones específicos en una mujer embarazada, que pueden usarse para diversos fines, tales como determinar el riesgo de tener un embarazo o un resultado neonatal adverso.

Los kits para llevar a cabo ensayos para determinar el nivel de ARN de bacterias de un taxón bacteriano de interés incluyen típicamente al menos un oligonucleótido útil para la hibridación específica con al menos un segmento de una secuencia codificante de interés o su secuencia complementaria. Opcionalmente, este oligonucleótido se marca con una unidad estructural detectable. En algunos casos, los kits pueden incluir al menos dos cebadores de oligonucleótidos que pueden usarse en la amplificación de al menos un segmento de un transcrito de ADN o ARN bacteriano de interés por PCR, particularmente por RT-PCR.

Los kits para llevar a cabo ensayos para determinar el nivel de proteína de bacterias de un taxón bacteriano de interés incluyen típicamente al menos un anticuerpo útil para la unión específica a la secuencia de aminoácidos de la proteína diana. Opcionalmente, este anticuerpo está marcado con una unidad estructural detectable. El anticuerpo puede ser un anticuerpo monoclonal o un anticuerpo policlonal. En algunos casos, los kits pueden incluir al menos dos anticuerpos diferentes, uno para la unión específica a la proteína objetivo (es decir, el anticuerpo primario) y el otro para la detección del anticuerpo primario (es decir, el anticuerpo secundario), que a menudo se une a una unidad estructural detectable.

Típicamente, los kits también incluyen un control estándar apropiado. Los controles estándar indican el valor promedio de una proteína objetivo o un ARNm objetivo expresado por bacterias de un taxón bacteriano específico en el epitelio cervical de sujetos sanos y embarazadas que no corren el riesgo de tener un embarazo o un resultado neonatal adverso. En algunos casos, dicho control estándar puede proporcionarse en forma de un valor establecido. Además, los kits de esta invención pueden proporcionar manuales de instrucciones para guiar a los usuarios en el análisis de muestras de prueba y evaluar el riesgo de tener un evento adverso del embarazo, como el parto prematuro, en un sujeto femenino de prueba.

#### Ejemplos

Los siguientes ejemplos se proporcionan solo a modo de ilustración y no a modo de limitación. Los expertos en la materia reconocerán fácilmente una variedad de parámetros no críticos que podrían cambiarse o modificarse para producir esencialmente los mismos resultados o resultados similares.

Ejemplo 1. Taxones bacterianos diferencialmente abundantes en el cuello uterino de mujeres con complicaciones asociadas al embarazo

#### Antecedentes

Para apoyar el cuello uterino en un intento de prolongar el embarazo de mujeres con dilatación cervical avanzada o acortamiento cervical prematuro, los médicos pueden colocar un cerclaje (Owen, Hankins et al. 2009) o un pesario cervical. Sin embargo, se demostró que las mujeres con infección, como la infección intraamniótica, a pesar de la colocación del cerclaje, se asociaron con resultados adversos del embarazo, incluida la pérdida del embarazo, el parto prematuro, la ruptura de la membrana (Romero, Gonzalez et al., 1992). Específicamente, entre 33 mujeres con dilatación cervical  $> 2$  cm, membrana intacta, y sin trabajo de parto activo entre las 14 y 24 semanas de gestación, 17 (51.5%) tenían invasión microbiana de la cavidad amniótica. Todos los pacientes con invasión microbiana de la cavidad amniótica tuvieron complicaciones. Los pacientes que se sometieron a cerclaje cervical en presencia de un cultivo de líquido amniótico positivo tuvieron ruptura de membranas, corioamnionitis clínica o pérdida del embarazo (Romero, Gonzalez et al., 1992). Por lo tanto, el cerclaje puede no ser beneficioso para este subconjunto de pacientes con dilatación cervical avanzada. Basado principalmente en el consenso y la opinión de expertos, se ha recomendado que la colocación de cerclaje puede ser beneficiosa solo si se descarta la infección intraamniótica (American College of Obstetricians and Gynecologists. 2014. Cerclage for the management of cervical insufficiency. Practice Bulletin No. 142. *Obstet Gynecol* 2014; 123: 372-9). Del mismo modo, también se ha sugerido descartar infección antes de la colocación del pesario, ya que atrapar al feto dentro de un entorno altamente infeccioso o inflamatorio puede conducir a resultados neonatales adversos, incluida la parálisis cerebral o el daño cerebral del feto.

Actualmente, la detección de bacterias en una muestra se basa en cultivos, microscopía y ensayos de reacción en cadena de polimerasa (PCR) específicos de especies bacterianas, que solo ofrecen sensibilidad baja a moderada y, por lo tanto, no son buenos para descartar infección. Por ejemplo, al usar un cultivo de líquido amniótico para bacterias aeróbicas y anaeróbicas, así como micoplasmas genitales y PCR dirigidos a *Ureaplasma urealyticum* y *Ureaplasma parvum*, 44 pacientes (76%) de los 58 pacientes con dilatación cervical avanzada obtuvieron resultados negativos de bacterias tanto para cultivo como para PCR (Oh, Lee et al., 2010). Sin embargo, eventualmente, entre estas mujeres con resultado negativo de bacterias, 63% (15/24) dieron positivo para coriodeciduitis, 30% (7/23) positivo para amnionitis, 25% (6/24) positivo para funisitis, 62% (24/39) dio a luz a un recién nacido con una puntuación de Apgar de 1 min  $< 4$ , 41% (18/44) dio a luz a un recién nacido que murió dentro de 1 día de nacimiento (Oh, Lee et al., 2010). La prevalencia de inflamación y el mal resultado neonatal en estos 44 pacientes con cultivo negativo y resultados de PCR sugirieron que muchos de estos resultados negativos para bacterias podrían ser falsamente negativos.

Además, también se detectaron ureaplasmas en embarazos de resultados normales sin complicaciones (Gray, Robinson et al., 1992, Gerber, Vial et al., 2003, Perni, Vardhana et al., 2004). Es posible que los Ureaplasmas sean solo parte de la bacteria comensal, o flora normal, que residen también en el tracto reproductivo de las mujeres, lo que resulta en un embarazo a término y un recién nacido normal. Esto resalta la limitación de detectar solo unas pocas especies bacterianas, seleccionadas por un enfoque candidato, en muchos de los estudios previos.

En contraste, se ha perfilado sistemáticamente la presencia y abundancia relativa de esencialmente todos ( $> 9.500$  tipos/taxones de bacterias con secuencias genómicas conocidas) en cada una de las muestras de hisopo cervical, obtenidas de mujeres con dilatación/acortamiento cervical avanzado y sin esas condiciones, usando 16S secuenciación genómica paralela masiva basada en ARN ribosómico (ARNr). La presente metodología ha superado varias limitaciones de esos estudios previos. Primero, la presente metodología puede detectar universalmente esencialmente todos los taxones bacterianos, pero no solo una lista seleccionada de candidatos. En segundo lugar, la presente metodología es independiente del cultivo y, por lo tanto, también se pueden detectar bacterias fastidiosas. En tercer lugar, la presente metodología no requiere bacterias vivas, ya que se basa en una PCR que puede amplificar los fragmentos de ADN genómico de bacterias vivas o muertas. Cuarto, la presente metodología investiga un tipo de muestra no invasiva, el hisopo cervical, que se puede obtener fácilmente de cualquier mujer embarazada con o sin complicaciones. Esto facilita el emparejamiento apropiado con un grupo de control normal y la comparación de datos entre la enfermedad y los grupos de control, a diferencia del líquido amniótico que requiere procedimientos invasivos con riesgo de pérdida fetal y no se recomienda para el embarazo normal a menos que haya una indicación. Quinto, la presente metodología investiga el hisopo cervical, que puede obtenerse fácilmente de cualquier mujer embarazada a cualquier edad gestacional. Esto facilita no solo el emparejamiento con un grupo de control, sino también un muestreo posiblemente temprano y, por lo tanto, la detección y el tratamiento, a diferencia del líquido amniótico o la placenta, que generalmente se obtienen después de 14 o 37 semanas de gestación, respectivamente.

Hay muchas publicaciones que utilizan secuenciación genómica paralela masiva basada en ARN ribosómico 16S (ARNr) para perfilar comunidades bacterianas en muchos sitios anatómicos de humanos "normales y sanos" sin complicaciones. Estos incluyen publicaciones sobre las comunidades bacterianas en muestras de frotis vaginales de mujeres no embarazadas o mujeres embarazadas "normales y sanas" (Ravel, Gajer et al., 2011; Aagaard, Riehle et al., 2012, Gajer, Brotman et al., 2012). Sin embargo, pocos han analizado muestras de mujeres con complicaciones del embarazo (Hummelen, Fernandes et al., 2010).

Específicamente, uno fue un estudio sobre las comunidades bacterianas en la vagina de mujeres VIH positivas, que no proporcionó un conjunto de datos aplicable a la población obstétrica principal sin VIH (Hummelen, Fernandes et al., 2010). Otro fue un estudio sobre las comunidades bacterianas en las muestras placentarias entregadas antes de las 37 semanas o después de las 37 semanas, lo que obviamente no facilitó una comparación gestacional por edad y, por lo tanto, implicó un conjunto de datos confundido por la diferencia de edad gestacional (Aagaard, Ma et al., 2014). Por lo tanto, hasta la fecha, no existe una publicación sobre el perfil sistemático de los taxones bacterianos en el cuello

uterino de mujeres con tales resultados adversos y la comparación con aquellos en el cuello uterino de mujeres gestacionales de edad similar sin complicaciones.

El presente estudio actual proporciona el primer perfil sistemático y relativamente completo de las comunidades bacterianas en las muestras de hisopo cervical de mujeres embarazadas con complicaciones (dilatación o acortamiento cervical avanzado). Además, por primera vez, comparamos sistemáticamente ese perfil de comunidades bacterianas en las mujeres con complicaciones, con mujeres gestacionales de la misma edad sin tales complicaciones. Además, por primera vez, proporcionamos una lista de taxones bacterianos que son diferencialmente abundantes en el cuello uterino de las mujeres con tales complicaciones, en comparación con el de las mujeres sin tales complicaciones. La provisión de esta lista es una base importante para el diseño de ensayos de PCR para apuntar específicamente a la "flora anormal" en un cuello uterino grávido, pero no a la "flora normal". Por lo tanto, razonamos que el presente estudio proporciona datos para mejorar la detección de bacterias avanzadas asociadas a la dilatación cervical.

Notablemente, la presencia de miembros seleccionados de esta lista de taxones bacterianos entre 13 y 25 semanas de gestación está asociada o es predictiva de resultados adversos, incluyendo el nacimiento antes de las 34 semanas y la hemorragia intraventricular. Por lo tanto, también razonamos que nuestra lista de taxones bacterianos también puede proporcionar un pronóstico temprano de estos resultados adversos, a fin de facilitar una monitorización cercana temprana o la transferencia a un hospital terciario con la atención neonatal adecuada.

El propósito del estudio descrito en este documento era perfilar sistemáticamente todas las bacterias e identificar una lista de taxones bacterianos y sus secuencias genómicas parciales que abundan de manera diferencial en el cuello uterino de las mujeres con dilatación cervical avanzada o acortamiento cervical. Específicamente, el estudio fue diseñado para (i) perfilar sistemáticamente los taxones bacterianos en el cuello uterino de las mujeres con dilatación cervical avanzada ("el cuello uterino dilatado"), y compararlos con los del cuello uterino de mujeres con esta enfermedad ("el cuello uterino cerrado"); (ii) perfilar sistemáticamente los taxones bacterianos en el cuello uterino de las mujeres con acortamiento cervical ("el cuello uterino acortado"), y compararlos con los del cuello uterino de las mujeres con esta combinación ("el cuello uterino de longitud normal"); (iii) identificar sistemáticamente una lista de taxones bacterianos diferencialmente abundantes en el cérvix dilatada, en relación con los del cuello uterino cerrado utilizando los datos en (i); y (iv) identificar sistemáticamente una lista de taxones bacterianos diferencialmente abundantes en el cuello uterino acortado, en relación con los del cuello uterino de longitud normal utilizando los datos en (ii).

Se tiene aquí la hipótesis de que las comunidades bacterianas que colonizan el cérvix dilatada son diferentes de las que colonizan el cuello uterino normalmente cerrado. Para probar esta hipótesis, se ha perfilado sistemáticamente los taxones bacterianos en el cérvix dilatada avanzado ("el cérvix dilatada", n=19) y los del cuello uterino normalmente cerrado ("el cuello uterino cerrado", n=13) usando ARN ribosómico 16S secuenciación genómica paralela masiva basada en (ARNr). Para probar sistemáticamente taxones diferencialmente abundantes entre los dos grupos, se ha comparado la abundancia relativa de todos los taxones perfilados en el grupo de cérvix dilatada y aquellos en el grupo de cuello uterino cerrado utilizando procedimientos estadísticos apropiados. Para evaluar si los miembros de esta lista de taxones bacterianos pueden predecir cualquier embarazo adverso o resultados neonatales, se ha seguido estos embarazos desde el reclutamiento para este estudio hasta 28 días después del nacimiento, y se ha calculado la sensibilidad, especificidad, valores predictivos positivos, valores predictivos negativos. para predecir esos resultados siempre que estuvieran disponibles.

Del mismo modo, también se ha perfilado sistemáticamente los taxones bacterianos en el cuello uterino acortado (n=11), y en comparación con los de los controles gestacionales de edad similar con un cuello uterino de longitud normal (n=11). También se ha identificado una lista de taxones diferencialmente abundantes entre el cuello uterino acortado y los grupos de longitud normal mediante procedimientos estadísticos y se siguieron los resultados clínicos. Siempre que los resultados estuvieron disponibles, también se evaluó el desempeño predictivo de los miembros seleccionados de esa lista.

#### Métodos y Resultados

Contratación de participantes. Este estudio se realizó con la aprobación ética de la junta de revisión institucional respectiva y se obtuvieron muestras de mujeres embarazadas con consentimiento informado. En el grupo de prueba (el grupo de cérvix dilatada, n=19), solo se reclutaron mujeres embarazadas con dilatación cervical avanzada indolora >0.8 cm.

Mientras que, en el grupo de referencia (el grupo de cuello uterino cerrado, n=13), solo se reclutaron mujeres embarazadas sin dilatación cervical avanzada indolora >0.8 cm. Además, para permitir una comparación equitativa de los resultados y los resultados clínicos, tanto en el grupo de prueba como en el de referencia, solo se reclutaron mujeres (i) sin regular y frecuente antes de las 34 semanas de gestación y (ii) con una indicación de someterse a colocación de cerclaje cervical. Las características clave de estos 32 participantes (19 en el grupo de cérvix dilatada y 13 en el grupo de cuello uterino cerrado) se enumeran en la Tabla 1.

Para evitar complicar el fenotipo (es decir, dilatación cervical avanzada), tanto en la prueba como en los grupos de referencia, se excluyeron los embarazos que implicaban preeclampsia, embarazos múltiples, sufrimiento fetal, restricción del crecimiento, anomalías cromosómicas o estructurales en el momento de la recogida de muestras. Para minimizar algunos factores de confusión importantes que afectan a las comunidades bacterianas en el cuello uterino, los participantes que tuvieron actividades sexuales o aplicaron cualquier otra aplicación vaginal (por ejemplo, medicamentos o supositorios vaginales, ducha vaginal) 48 horas antes de la recolección de la muestra o con antibióticos o antimicóticos 30 días antes de la recolección de la muestra, también se excluyeron los tumores de ovario. Participantes que tuvieron un aborto espontáneo o un parto previo entre 14 y 36 semanas (n=3, todos en el grupo de

5 cervix dilatada, prueba exacta de Fisher, p=0.253), evacuación quirúrgica previa por aborto espontáneo o interrupción del embarazo (n=16, 9 dilatados y 7 cerrados, p=1.00), cirugía cervical (n=6, 2 dilatados y 4 cerrados, p=0.194), anomalía uterina (n=6, 3 dilatados y 3 cerrados, p=0.666), flujo vaginal anormal reciente (n=5, 4 cervix dilatada y 1 cerrado, p=0.625) se documentaron claramente en nuestra base de datos. Estos factores no fueron estadísticamente diferentes entre los dos grupos, pero se incluirían para interpretar los datos.

Recolección y extracción de ADN de muestras de hisopo cervical: para minimizar la posibilidad de contaminación por el medio ambiente, el personal clínico u otras partes del tracto reproductivo femenino, la muestra de hisopo cervical se recolectó utilizando un Calgiswab Tipo III (Puritan, Guilford, Maine, Estados Unidos) Antes de cualquier otro procedimiento inmediatamente después de la apertura del aparato reproductor femenino por el espéculo. Para asegurarse de que se muestrearon y compararon las mismas ubicaciones anatómicas, se recogió cada muestra de hisopo cervical desde una posición fija en el lado periférico (la posición de las 12 en punto frente al médico) del orificio externo. Para mantener la coherencia para una comparación equitativa entre todas las muestras, un solo clínico recolectó todas las muestras del cervix dilatada y los grupos de cuello uterino cerrado. Para minimizar las variaciones en la recolección, cada hisopo se recogió girando 360 grados una vez. Para minimizar cualquier riesgo aumentado de infectar a las participantes o al feto en el útero, se recogieron los hisopos sin tocar el tapón de moco cervical y fueron estériles. Durante la recolección de la muestra de hisopo, también se tuvo especial cuidado de no tocar los labios ni ninguna parte del tracto reproductor femenino que no sea el orificio externo. Para controlar la contaminación de bacterias en la sala de operaciones, los reactivos y los procedimientos de recolección, se recogió otro hisopo de control negativo en paralelo con cada hisopo cervical, pero sin tocar a los pacientes.

El hisopo cervical y los hisopos de control negativo se sumergieron en agua estéril y libre de nucleasas y se almacenaron a -80°C hasta la extracción. Los hisopos se extrajeron para ADN genómico utilizando un método establecido (Método 1 en Yuan, Cohen et al., 2012), que garantizaría una representación equitativa de las comunidades bacterianas que se encuentran comúnmente en los tractos reproductivos femeninos. Este método implicaba el pretratamiento de la muestra mediante la mutanolisina (Sigma-Aldrich) y un método de extracción de ADN basado en columnas (QiaAmp DNA Mini Kit, Qiagen). Para minimizar cualquier variación de lote, todas las muestras se extrajeron el mismo día.

Amplificación por PCR y secuenciación en paralelo masiva: dado que las muestras de hisopo cervical inevitablemente comprenderían ADN genómico humano entre el ADN genómico bacteriano, se ha amplificado específicamente el gen 16S ARNr, que comúnmente poseen todas las bacterias, pero no los humanos. Para facilitar la amplificación de las secuencias de ADN genómico de esencialmente todas las bacterias, se ha optado por utilizar un par de cebadores de PCR, a saber, V4 y V5, que eran complementarios al gen 16S ARNr de regiones altamente conservadas (Claesson, Wang et al.2010). Se ha probado el uso del Ribosomal Database Project (RDP) (Wang, Garrity et al. 2007), la base de datos pública más grande que contiene secuencias de ARNr 16S, se ha comprobado que nuestro par elegido de cebadores de PCR podría amplificar teóricamente >97% de >9.200 tipificados (establecido) taxones bacterianos de secuencias genómicas de 16S ARNr conocidas. Por lo tanto, este par de cebadores de PCR es aplicable para un perfil sistemático y no sesgado de comunidades bacterianas en este estudio.

Se amplificó el ADN genómico extraído de cada muestra de hisopo usando el par de cebadores de PCR V4-V5 (Claesson, Wang et al. 2010), que flanquea las regiones hipervariables V4 y V5 del gen 16S ARNr. Las secuencias de los cebadores directo e inverso son 5'-[Secuencia clave del cebador A] [Secuencia MID] AYT GGG YDT AAA GNG-3' (SEQ NO ID: 1) y 5'-[Clave B del cebador] CCG TCA ATT YYT TTR AGT TT-3' (SEQ ID NO: 2), respectivamente, donde la secuencia de cebador A, la secuencia de cebador B y la secuencia MID se describen en las "454 Sequencing System Guidelines for Amplicon Experimental Design July 2011" plataforma de secuencia paralela GX-FLX 454 Titanium (Roche). Cada PCR se realizó como una reacción de 50 µL con 2.5 unidades de la ADN polimerasa FastStart Taq (Sistema FastStart HiFi PCR dNTPack, Roche), MgCl<sub>2</sub> 4 mM, 100 nM de cada cebador y dNTP 200 µM. Todas las PCR se realizaron en un termociclador PTC-100 (Bio-Rad) utilizando las siguientes condiciones de termociclado: 95°C durante 2 minutos, seguido de 33 ciclos de 95°C durante 30 segundos, 40°C durante 30 segundos y 72°C durante 1 minuto, con una extensión final a 72°C durante 5 minutos y 25°C durante 5 minutos. Luego sometimos el producto de PCR a electroforesis. Confirmamos un solo amplicón de PCR del tamaño esperado para las 32 muestras de hisopo cervical, y ningún amplicón de PCR para los 32 controles de reactivos negativos correspondientes. Por lo tanto, el medio ambiente, los reactivos y los procedimientos estaban libres de cualquier contaminación de secuencias genómicas de ARNr 16S bacterianas no deseadas.

Posteriormente, purificamos todos los 32 productos de PCR, que se derivaron de las 32 muestras de hisopo cervical y que se unieron con las secuencias de identificador multiplex (MID) y secuencias de adaptador incorporadas a través de los extremos 5' de los cebadores de PCR anteriores, de acuerdo con las instrucciones recomendadas de Roche.

Los productos purificados se sometieron a una secuenciación genómica masivamente paralela utilizando el GX-FLX 454 Titanium (Roche), de acuerdo con las instrucciones del fabricante, apuntando a un promedio de alrededor de 10.000 lecturas de secuenciación sin procesar por muestra.

5 Para cada muestra, los datos de secuenciación sin procesar se eliminaron a nivel de diagrama de flujo, utilizando una implementación de Pyronoise (Quince, Lanzen et al., 2011) en Mothur (Schloss, Westcott et al., 2009). Las lecturas sin procesar se desnovilizaron por flujograma, se filtraron por calidad y longitud, se eliminaron quimera, se alinearon, se pre-agruparon y se agruparon en unidades taxonómicas operativas, que luego se clasificaron taxonómicamente, según el conjunto de capacitación del Ribosomal Database Project (RDP) (v9, 2012).

10 Identificación sistemática de taxones bacterianos diferencialmente abundantes entre el grupo de cérvix dilatada (n=19) y el grupo cerrado de cuello uterino (n=13): después de todos los pasos analíticos anteriores, se ha observado 342 taxones bacterianos en las 32 muestras de hisopo cervical. Para normalizar los recuentos de lectura variables en diferentes muestras, se realizó submuestreo aleatorio utilizando Mothur (Schloss, Westcott et al. 2009) para que cada una de las 32 muestras contuviera 7.594 lecturas procesadas para su posterior análisis. Alternativamente, también se ha normalizado al representar los recuentos de lectura de cada taxón en una muestra como una relación de los  
15 recuentos de lectura totales de esa muestra.

Para identificar los taxones diferencialmente abundantes entre el grupo de cérvix dilatada y el grupo de cuello uterino cerrado, se realizó una prueba estadística, a saber, Metastats, que está especialmente diseñado para este tipo de datos de secuenciación (White, Nagarajan et al. 2009). En esencia, Metastats presenta una prueba T no paramétrica y una prueba heurística para usar la prueba exacta de Fisher si un determinado taxón aparece en un promedio de  
20 menos de 1 lectura por muestra (el llamado problema de conteo disperso que plantea un desafío para detectar cambios significativos en este tipo de datos).

Después de la eliminación de secuencias de taxones que aparecieron solo una vez (secuencias de un solo tono) y el ajuste para múltiples pruebas usando el método de la tasa de descubrimiento falso (FDR) a FDR <5% (Storey and Tibshirani, 2003), 16 taxones permanecieron como estadísticamente significativos diferente ( $q < 0.05$ ). Entre ellos, 9  
25 taxones (Tabla 2A) aumentaron significativamente y 7 taxones (Tabla 2B) disminuyeron significativamente en el grupo de cérvix dilatada, en relación con el grupo de cuello uterino cerrado. La clasificación taxonómica más cercana de estas 16 secuencias genómicas a nivel de reino a género se enumeran en las Tablas 3A y 3B. Las secuencias genómicas de ARNr 16S se enumeran en las Tablas 4A y 4B. La clasificación de especies más cercana de estas 16 secuencias genómicas basada en la alineación de nucleótidos BLAST contra la base de datos de ARN ribosómico 16S (realizada utilizando el sitio web NCBI BLAST en junio de 2014) se enumeran en las Tablas 4E y 4F.  
30

Identificación sistemática de taxones bacterianos diferencialmente abundantes entre el grupo de cérvix dilatada (n=10) y el grupo de cuello uterino cerrado (n=10) que fueron emparejados a 1:1 por la semana gestacional más cercana en la recolección de la muestra: La edad gestacional en la recolección de muestras entre los dos grupos no es estadísticamente significativamente diferente. Sin embargo, para minimizar cualquier efecto de la edad gestacional en  
35 nuestro análisis, se ha emparejado a cada mujer en el grupo de cérvix dilatada con otras mujeres en el grupo de cuello uterino cerrado en la semana de gestación más cercana (dentro de 2 semanas) en la recolección de la muestra. Además, para minimizar cualquier efecto de diferentes tamaños de muestra de los dos grupos, se ha realizado la coincidencia anterior en una proporción 1:1. Finalmente, se pudo unir a 10 mujeres en el grupo de cérvix dilatada con 10 mujeres en el grupo de cuello uterino cerrado para la semana de gestación en la recolección de muestras.

40 Después del ajuste para múltiples pruebas usando el método de la tasa de descubrimiento falso (FDR) a FDR <5% (Storey and Tibshirani, 2003), 20 taxones permanecieron como estadísticamente significativamente diferentes ( $q < 0.05$ ). Entre ellos, 15 taxones (Tabla 2C) aumentaron significativamente y 5 taxones (Tabla 2D) disminuyeron significativamente en el grupo de cérvix dilatada, en relación con el grupo de cuello uterino cerrado. La clasificación taxonómica más cercana de estas 20 secuencias genómicas a nivel de reino a género se enumeran en las Tablas 3D y 3C. Las secuencias genómicas de 16S ARNr se enumeran en las Tablas 4C y 4D. La clasificación de especies más cercana de estas 16 secuencias genómicas basada en la alineación de nucleótidos BLAST contra la base de datos de ARN ribosómico 16S (realizada utilizando el sitio web NCBI BLAST en junio de 2014) se enumeran en las Tablas 4G y 4H.  
45

Rendimiento predictivo por miembros seleccionados de las listas de taxones bacterianos diferencialmente abundantes en diversos resultados. Se seleccionaron los 9 taxones de la Tabla 2A y los 15 taxones de la Tabla 2C para construir dos pruebas, a saber, la Prueba de Bacterias con Aumento Diferencial (DIBT) 1 y DIBT2, respectivamente.  
50

DIBT1 y DIBT2 pueden usar datos de secuenciación genómica masivamente paralelos o, más preferiblemente, datos de PCR específica de especie para detectar la presencia de un taxón bacteriano dado. Si alguno de estos 9 o 15 taxones aumentados diferencialmente estaba presente en esa muestra, se define como DIBT1 positivo o DIBT2 positivo, respectivamente. De lo contrario, si todos estos 9 o 15 taxones están ausentes en una muestra, se define como DIBT1 negativo o DIBT2 negativo, respectivamente.  
55

Para explorar la asociación entre DIBT1 o DIBT2 y el embarazo adverso o los resultados neonatales, se realizó la prueba exacta de Fisher. Para explorar el potencial de DIBT1 o DIBT2 en la predicción de resultados clave de

embarazo y neonatal, se calcularon los verdaderos positivos, falsos positivos, falsos negativos y verdaderos negativos de DIBT1 o DIBT2 en la predicción de estos resultados (Tablas 5A y 5B, respectivamente). Además, se calculó la sensibilidad, especificidad, valores predictivos positivos y negativos de DIBT1 y DIBT2 para predecir estos resultados (Tablas 5A y 5B, respectivamente).

5 Existen asociaciones significativas entre DIBT1 y dilatación cervical avanzada (prueba exacta de Fisher,  $p=0.028$ ), parto prematuro espontáneo <28 semanas ( $p=0.0023$ ), parto prematuro espontáneo <34 semanas ( $p=0.00065$ ), parto prematuro <28 semanas ( $p=0.0057$ ), parto prematuro <34 semanas ( $p=3.2 \times 10^{-6}$ ) y hemorragia intraventricular ( $p=0.01$ ). Es de destacar que los 7 casos de parto prematuro espontáneo se detectaron como DIBT1 positivo (sin falsos negativos). También es de destacar que, entre los 13 casos clasificados como DIBT1 positivos, todos se sometieron a parto prematuro <34 semanas (sin falsos positivos). Todos estos nacimientos prematuros no se debieron a preeclampsia, restricción del crecimiento intrauterino, anomalía cromosómica fetal). Por lo tanto, especulamos que estos 13 nacimientos prematuros pueden desencadenarse por la infección de una o más de estas "flora anormal" que se identificaron en este estudio y se seleccionaron para su inclusión en el DIBT1.

10 En cuanto a DIBT2, se observan asociaciones significativas entre este y el parto prematuro espontáneo <28 semanas ( $p=0.0023$ ), el parto prematuro espontáneo <34 semanas ( $p=0.00065$ ), el parto prematuro <28 semanas ( $p=0.0057$ ), el parto prematuro nacimiento <34 semanas ( $p=0.0032$ ). Es de destacar que los 7 casos de parto prematuro espontáneo <34 semanas fueron detectados por DIBT2 como positivos (sin falsos negativos).

15 En otro estudio (estudio B), se ha realizado un análisis similar en muestras de hisopo cervical recolectadas a una edad gestacional similar, pero de mujeres con un cuello uterino corto o con un cuello uterino de longitud normal (Tabla 6). Para cada muestra, se extrajo el ADN y se amplificó por PCR mediante los cebadores V4V5 y se sometió a análisis GS-FLX 454. Después del ajuste para múltiples pruebas usando el método de la tasa de descubrimiento falso (FDR) a FDR <5% (Storey and Tibshirani, 2003), 24 taxones permanecieron como estadísticamente significativamente diferentes ( $q < 0.05$ ). Entre ellos, 17 taxones (Tabla 7) aumentaron significativamente y 7 taxones (no mostrados) disminuyeron significativamente en el grupo de cuello uterino corto, en relación con el grupo de cuello uterino de longitud normal. Las secuencias genómicas de ARNr 16S se enumeran en la Tabla 8. La clasificación de especies más cercana de estas 17 secuencias genómicas basada en la alineación de nucleótidos BLAST contra la base de datos de ARN ribosómico 16S (realizada utilizando el sitio web NCBI BLAST en junio de 2014) se enumera en la Tabla 9.

20 Rendimiento predictivo de miembros seleccionados de las listas de taxones bacterianos diferencialmente abundantes en diversos resultados. Se seleccionaron los 17 taxones de la Tabla 7 para construir otra prueba, a saber, la Prueba de Bacterias con Aumento Diferencial (DIBT) 3.

25 DIBT3 puede usar datos de secuenciación genómica masivamente paralelos o, más preferiblemente, datos de PCR específica de especie para detectar la presencia de un taxón bacteriano dado. Si alguno de estos 17 taxones aumentados diferencialmente estaba presente en esa muestra, se define como DIBT3 positivo. De lo contrario, si todos estos 17 taxones están ausentes en una muestra, se define como DIBT3 negativo, respectivamente. La asociación de los resultados positivos DIBT3 y el embarazo o el resultado neonatal adverso se tabula en la Tabla 10.

30 Existen asociaciones significativas entre DIBT3 y dilatación cervical prematura (prueba exacta de Fisher,  $p=0.00022$ ), parto prematuro espontáneo <34 semanas ( $p=0.049$ ), parto prematuro <34 semanas ( $p=0.049$ ). Es de destacar que los 3 casos de parto prematuro espontáneo se detectaron como DIBT3 positivo (sin falsos negativos).

35 Los marcadores bacterianos proporcionados en este documento se usaron para predecir con precisión los resultados adversos del embarazo y los resultados neonatales basados en una prueba molecular realizada tan pronto como 13 semanas de gestación (Tablas 5A, 5B y 10). El método descrito en este documento facilita la intervención temprana, como la monitorización cercana o la transferencia oportuna a una unidad de tratamiento terciario con cuidados intensivos neonatales.

40 Ejemplo 2: Firma de microbioma cervical para la identificación de pacientes con insuficiencia cervical que resulta en parto prematuro espontáneo después de la intervención clínica

45 No se han investigado sistemáticamente los taxones bacterianos que colonizan los cervixes de pacientes con insuficiencia cervical (IC) que responden de manera diferente a la intervención clínica (cerclaje/pesario). Utilizando secuenciación paralela masiva, interrogamos las abundancias de más de 9.600 taxones por muestra de hisopo cervical obtenida antes de la intervención de pacientes con IC reclutadas en serie con embarazo único y mujeres con IC adecuadamente emparejadas. Se observó que los microbiomas cervicales estaban alterados en los pacientes con IC, en comparación con los de los controles sin IC. Notablemente, se identificaron 6 taxones diferencialmente abundantes en pacientes que resultaron en "parto prematuro espontáneo (<34 semanas, sPTB) después de la intervención", en comparación con los que resultaron en "parto a término ( $\geq 37$  semanas) después de la intervención". Usando el  $\log_{10}$  (abundancia total de estos 6 taxones), LA6,  $> 1.15$  para definir un resultado positivo, clasificamos correctamente a todos los pacientes menos uno que resultaron en "sPTB después de la intervención" (9/10=90%), sin falso positivo (0/15=0%). Los pacientes con LA6 positivo permanecieron sin recibir durante un periodo más corto después de la intervención [número medio de días entre la intervención y el parto, 10 días frente a 126 días; Prueba de Logrank, p

<0.00001; cociente de riesgos, 6.34; Intervalo de confianza del 95%, 1.51 a 26.6], en comparación con pacientes con LA6 negativo. Además, los pacientes positivos para LA6 dieron a luz antes que sus contrapartes negativas para LA6 [edad gestacional media en el momento del parto de 23.7 semanas frente a 38.4 semanas; Prueba de Logrank, p <0.0001; cociente de riesgos, 6.24; Intervalo de confianza del 95%, 1.50 a 25.9]. El presente estudio destaca el uso potencial del microbioma cervical previo a la intervención para proporcionar información pronóstica del embarazo después de la intervención de cerclaje/pesario.

#### Introducción

La insuficiencia cervical (IC) es un factor de riesgo para el parto prematuro (PTB), que se asocia con la morbilidad neonatal y la muerte perinatal. Se manifiesta en las mujeres afectadas con un cérvix dilatada prematuramente (dilatación cervical, 1 cm-5 cm) o acortado (longitud cervical <25 mm) en el segundo trimestre, en lugar del tercero. Se ha demostrado que la intervención clínica mediante la colocación de cerclaje quirúrgico o pesario cervical (Shirodkar, *Antiseptic*, 52, 299-300 (1955); Cross, *Lancet*, 274, 127 (1959)) en el cuello uterino debilitado de pacientes con IC disminuye la tasa de PTB <28 semanas (Pereira et al., *Am J Obstet Gynecol* 197, 483 e481-488 (2007)) o PTB <34 semanas (Althuisius et al., *Am J Obstet Gynecol* 189, 907-910 (2003); Goya et al., *Lancet* 379, 1800-1806 (2012)), aumentan la tasa de supervivencia neonatal (3) y el intervalo entre la presentación y el parto (Pereira et al., *Am J Obstet Gynecol* 197, 483 e481-488 (2007)).

Sin embargo, la intervención cervical no beneficia a todos los pacientes con IC. En particular, los pacientes con IC con infección intraamniótica (IAI) resultaron en una tasa 4 veces mayor de PTB <34 semanas después de la intervención de cerclaje (Romero et al., *Am J Obstet Gynecol* 167, 1086-1091 (1992)), en comparación con aquellos recibiendo la misma intervención, pero sin IAI. Los pacientes que recibieron intervención de cerclaje solo si se descartó IAI resultaron en una tasa más baja de PTB <34 semanas, en comparación con aquellos que recibieron la intervención sin pruebas de IAI (Mays et al., *Obstet Gynecol* 95, 652-655 (2000)). Dado que el IAI es altamente prevalente (38% -51%) en pacientes con IC (Romero et al., *Am J Obstet Gynecol* 167, 1086-1091 (1992); Mays et al., *Obstet Gynecol* 95, 652-655 (2000)), los expertos han sugerido a los médicos que consideren descartar que el IAI use la amniocentesis previa al cerclaje para detectar microorganismos (Berghella et al., *Am J Obstet Gynecol* 209, 181-192 (2013); Airoidi et al., *Am J Perinatol* 26, 63-68 (2009)). Sin embargo, la amniocentesis es invasiva y se asocia con una posibilidad pequeña pero finita de pérdida fetal. Por lo tanto, investigamos la colonización cervical en pacientes con IC y se observó correlación la expresión del microbioma con los resultados de la intervención cervical. Los resultados muestran que el muestreo de hisopo cervical puede ser una alternativa relativamente no invasiva a la amniocentesis en pacientes con IC.

La secuenciación en paralelo masiva (MPS) ha facilitado una cultura independiente del cultivo y, por lo tanto, ha proporcionado una visión más sensible e integral de los microorganismos que colonizan diferentes sitios del cuerpo. Convencionalmente, se ha pensado que la placenta que se encuentra dentro de la cavidad amniótica es estéril. Contrariamente a esto, un estudio de microbioma y metagenómico basado en MPS publicado en esta revista ha revelado que la placenta alberga un microbioma de baja biomasa que varía en asociación con una historia remota de infección prenatal materna y parto prematuro (Aagaard et al., *Sci Transl Med* 6, 237ra265 (2014)). El presente estudio se realizó para investigar la asociación entre el microbioma cervical prenatal y el resultado del parto prematuro de pacientes con IC.

Los datos de alta resolución generados por MPS y la información cuantitativa que se puede inferir de ellos también han ampliado nuestro conocimiento sobre microorganismos en la salud de las mujeres. Se han observado cinco clases principales de comunidades bacterianas (grupos comunitarios) en el tracto vaginal de mujeres no embarazadas en edad reproductiva. En mujeres cuya proporción vaginal de no *Lactobacillus* sp. aumentó como se ve comúnmente en el grupo comunitario IV, sus puntajes Nugent, un factor de diagnóstico comúnmente utilizado para identificar a las mujeres con vaginosis bacteriana, también aumentaron (Ravel et al., *Proc Natl Acad Sci USA* 108, 4680-4687 (2011)). Notablemente, las mujeres con vaginosis bacteriana tenían más probabilidades de tener un cérvix dilatada (relación ajustada de probabilidades, 4.9; intervalo de confianza del 95% 2.2-10.9) (Kilpatrick et al., *Am J Obstet Gynecol* 194, 1168-1176 (2006)). El estudio longitudinal del microbioma vaginal en mujeres no embarazadas ha revelado que algunas comunidades bacterianas cambian notablemente en períodos cortos de tiempo, mientras que otras son relativamente estables (Gajer et al., *Sci Transl Med* 4, 132ra152 (2012)). Según la teoría ecológica, las comunidades menos estables son más susceptibles a la invasión de organismos patógenos (Dunstan et al., *Ecology* 87, 2842-2850 (2006)). En este estudio, los microbiomas cervicales en pacientes con IC se compararon con los de mujeres sin IC adecuadamente emparejadas.

#### Resultados

Reclutamos a 34 pacientes con insuficiencia cervical (IC) y se obtuvo una muestra de hisopo cervical de cada uno antes del tratamiento con cerclaje/pesario. Las mujeres de embarazo único en el segundo trimestre que participaron en este estudio tuvieron: (i) dilatación cervical avanzada indolora (1.5 cm-5.0 cm) y/o acortamiento cervical (longitud cervical <25 mm), y; (ii) membrana intacta, y; (iii) sin contracciones laborales; durante el tiempo del muestreo de hisopo cervical.

Todos estos pacientes con IC recibieron intervención de cerclaje/pesario debido al cérvix dilatada/acortado prematuramente, respectivamente. Los pacientes fueron seguidos hasta un mes después del parto. Se excluyeron nueve embarazos complicados por PTB iatrogénica debido a preeclampsia, sufrimiento fetal, restricción del crecimiento, cromosomas fetales o anomalías estructurales. Entre los 25 pacientes restantes con IC, 15 mujeres tuvieron partos a término (TB, entregados en o después de las 37 semanas de gestación) después de la intervención y 10 resultaron en nacimientos prematuros espontáneos (sPTB, entregados en menos de 34 semanas) después de la intervención (Figura 1A). De estos, 7 involucraron morbilidad neonatal, incluido el síndrome de dificultad respiratoria (SDR), displasia broncopulmonar (DBP), hemorragia intraventricular (IVH) y retinopatía del prematuro (ROP) o mortalidad perinatal (Figura 1A).

#### 10 Perfilado de taxones bacterianos por paralelo masivo

Cada muestra de hisopo cervical recogida antes de la intervención clínica se sometió a extracción de ADN bacteriano, amplificación por PCR del gen de ARN ribosómico 16S (ARNr) y secuenciación masivamente paralela del amplicón (Titanio, GS-FLX 454, Roche). Los cebadores de PCR dirigidos a las regiones V4 y V5 del gen 16S ARNr podrían amplificar más de 9.600 bacterias bien establecidas de secuencias conocidas de 16S ARNr para su análisis (Claesson et al., *Nucleic Acids Res* 38, e200 (2010)). Este alcance fue más amplio que la mayoría de los estudios de microbioma publicados del tracto reproductivo femenino.

Para minimizar la detección espuria de "nuevos" taxones que surgen de errores de secuenciación, las lecturas de secuenciación sin procesar (0.68 millones de lecturas) fueron eliminadas y filtradas por calidad en lecturas procesadas (0.54 millones de lecturas) utilizando métodos bien establecidos (Schloss et al., *Appl Environ Microbiol* 75, 7537-7541 (2009); Quince et al., *Nat Methods* 6, 639-641 (2009)). En promedio, cada una de las 25 muestras fue secuenciada a una profundidad de 22.000 lecturas procesadas de alta calidad. Tal profundidad de secuenciación fue mayor que la mayoría de los estudios de microbioma publicados del tracto reproductivo femenino.

Además, las lecturas procesadas de todas las muestras con al menos un 97% de identidad de secuencia se agruparon como una unidad taxonómica operativa (Otu, es decir, un taxón bacteriano). En total, se detectaron 152 taxones bacterianos en las 25 cervices. Para permitir una comparación más justa entre las muestras secuenciadas en diferentes recuentos de lectura, se ha realizado la normalización de la Escala de suma acumulativa (CSS) (Paulson et al., *Nat Methods* 10, 1200-1202 (2013)) y expresamos el valor de abundancia (unidad: recuentos de lectura) para cada taxón como su recuento de lectura normalizado por CSS en la escala  $\log_{10}$  (es decir, el valor de registro (abundancia)).

#### 30 Taxones bacterianos altamente abundantes en pacientes con insuficiencia cervical

La Figura 1B muestra los 10 taxones bacterianos más abundantes observados en las 10 cervices "sPTB después de la intervención". Contrariamente a un tracto reproductivo femenino sano predominante por Lactobacilli, un miembro del género *Gardnerella* (Otu 4) ha sido identificado como el taxón bacteriano más abundante en el cuello uterino "sPTB después del tratamiento" [Figura 1B, fila #1, es decir, el taxón con los mayores valores de  $\log(\text{abundancia})$  totales en los 10 cérvices "sPTB después de la intervención"]. De hecho, 7 de los 10 taxones bacterianos más abundantes en este grupo se han clasificado como géneros no Lactobacillus: *Gardnerella*, dos *Sneathias*, *Aerococcus*, *Megasphaera*, *Pseudomonas* y *Anaerococcus* (Figura 1B, filas #1 a #10 y columnas debajo de "sPTB").

En comparación, solo 5 de los 10 taxones bacterianos más abundantes en el cuello uterino "TB después del tratamiento" se han clasificado como géneros no Lactobacillus (Figura 1C, filas #1 a #10 y columnas bajo "TB"). En particular, las 3 bacterias más abundantes se han identificado como *Lactobacillus crispatus*, *L. iners* y *L. jensenii*, que se sabe que predominan en el tracto reproductivo femenino sano (Ravel et al., *Proc Natl Acad Sci USA* 108, 4680-4687 (2011)).

#### Taxones diferencialmente abundantes en pacientes con diferentes respuestas a la intervención

Es importante destacar que se ha identificado 7 taxones bacterianos que son diferencialmente abundantes entre los grupos "sPTB después del tratamiento" y "TB después del tratamiento" (Figura ID). Los valores logarítmicos (abundancia) de los 7 taxones, a saber, *Sneathia* (Otu 11), *Parvimonas* (Otu 16), *Ureaplasma* (Otu 56), *Atopobium* (Otu 42), *Peptoniphilus* (Otu 28), *Megasphaera* (Otu 47) y *Paraeggerthella* (Otu 40), fueron mayores en el primer grupo (prueba de suma de rango de Mann-Whitney,  $p < 0.05$ ; se realizó una corrección de prueba múltiple usando el método de la tasa de descubrimiento falso (FDR),  $FDR < 5\%$ ) (Figura ID, filas #1 a #7, última columna). Sorprendentemente, de estos, los últimos 6 taxones se observaron exclusivamente en el grupo "sPTB después del tratamiento", pero no en el grupo "TB después del tratamiento" [Figura ID, filas #2 a #7, muchos valores de registro (abundancia) en las columnas "sPTB" son  $\geq 0$ , pero todos los valores de registro (abundancia) en las columnas "TB" son 0].

#### Taxones diferencialmente abundantes y resultado después de la intervención

Para resumir las abundancias de los últimos 6 taxones diferencialmente abundantes (Mann-Whitney,  $p < 0.01$ ; Tabla 11), se ha calculado para cada muestra de hisopo cervical un valor LA6, que se refiere al  $\log_{10}$  (abundancia total de los 6 taxones diferencialmente abundantes). En resumen, para cada muestra, transformamos cada uno de sus valores de  $\log$  (abundancia) para *Parvimonas* (Otu 16), *Ureaplasma* (Otu 56), *Atopobium* (Otu 42), *Peptoniphilus* (Otu 28),

Megasphaera (Otu 47) y Paraeggerthella (Otu 40) de vuelta a la escala lineal común. Después de agregar esos 6 valores de abundancia en la escala común, transformamos esa abundancia total y expresamos el valor LA6 en la escala  $\log_{10}$ .

5 Los valores medios de LA6 fueron 2.61 y 0.78 en los grupos "sPTB después del tratamiento" y "TB después del tratamiento", respectivamente (Figura 2A). Se muestra que los valores medios de LA6 aumentan 3.36 veces en el primer grupo (Mann-Whitney,  $p < 0.0001$ ). Para encontrar el umbral óptimo para identificar el grupo "sPTB después del tratamiento" entre los pacientes con IC que reciben tratamiento, se ha trazado la curva de características operativas del receptor (ROC) [área bajo la curva ROC (intervalo de confianza del 95%), 0.95 (0.84-1.06);  $p = 0.0002$ ]. Usando el LA6  $> 1.15$  como umbral para definir un resultado positivo, se pudo identificar a todos los pacientes con IC "sPTB después de la intervención" excepto uno (9/10 = 90% de sensibilidad) sin falsos positivos (1 - 0/15 = 100% de especificidad).

15 Es importante destacar que los pacientes positivos para LA6 dieron a luz antes de la intervención clínica que los pacientes positivos para LA6 (Figura 2B, mediana de edad gestacional en el momento del parto de 23.7 semanas frente a 38.4 semanas; intervalo de confianza del 95%, 20.6 semanas - 25.4 semanas frente a 38.0 semanas - 38.7 semanas; Chi-cuadrado, 32.352; df, 1; prueba de Logrank,  $p < 0.0001$ ; razón de riesgo, 6.24; intervalo de confianza del 95%, 1.50 a 25.9; MedCalc, versión 14.12). También es importante destacar que los pacientes LA6 positivos dieron a luz por un período más corto después de la intervención que sus contrapartes LA6 negativas (Figura 2C número medio de días entre la intervención y el parto, 10 días frente a 126 días; intervalo de confianza del 95%, 8 días - 32 días frente a 112 días - 134 días; Chi-cuadrado, 32.520; df, 1; prueba de Logrank,  $p < 0.00001$ ; razón de riesgo, 6.34; intervalo de confianza del 95%, 1.51 a 26.6).

## Discusión

### Firma de microbioma cervical identificada en este estudio

25 Se ha demostrado que las abundancias de 7 taxones bacterianos en pacientes con IC aumentaron significativamente en las mujeres dando como resultado "sPTB después de la intervención clínica", en comparación con las que resultaron en "TB después de la intervención clínica". Además, se ha calculado el valor LA6, que representa los valores de abundancia total de los 6 taxones más significativamente incrementados, para cada paciente con IC. Con base en los valores LA6 de las 25 muestras de hisopo cervical obtenidas antes de la intervención de cerclaje/pesario, se identificaron correctamente 10 de 11 pacientes con IC "sPTB después de la intervención" (9/10 = 90% de sensibilidad) sin ningún falso positivo en la TB de 15 pacientes "después de la intervención" (1 - 0/15 = 100% de especificidad).

30 Este estudio estuvo limitado por el tamaño de la muestra de mujeres embarazadas asiáticas e implicó un diseño retrospectivo. Sin embargo, ha proporcionado un panel de taxones bacterianos no obvio y enfocado que está asociado con IC, un riesgo importante para el parto prematuro. La firma del microbioma cervical identificada en este estudio, concretamente LA6, se ha ilustrado para proporcionar información pronóstica después de la intervención de cerclaje/pesario. Aparentemente, los pacientes con IC que dieron positivo para LA6 tienen un mayor riesgo de parto a una edad gestacional significativamente más temprana después de la intervención clínica que los que dieron negativo (23.7 semanas versus 38.4 semanas; razón de riesgo, 6.24; intervalo de confianza del 95%, 1.50 a 25.9). No solo eso, los pacientes positivos para LA6 también tienen un mayor riesgo de parto mucho antes de la intervención que sus contrapartes negativas para LA6 (10 días frente a 126 días; razón de riesgo, 6,34; intervalo de confianza del 95%, 1,51 a 26,6).

### 40 Materiales y métodos

#### Reclutamiento de participantes.

45 La aprobación de ética para realizar este estudio se obtuvo de la Institutional Review Board respectiva. Se obtuvo el consentimiento informado de mujeres embarazadas asiáticas que asistieron al Department of Obstetrics and Gynaecology, Prince of Wales Hospital, The Chinese University of Hong Kong o the Department of Obstetrics and Gynaecology, Hallym University, Seúl, Corea del Sur, Corea del Sur y cumplieron los siguientes criterios de inclusión: (i) dilatación cervical indolora (1.0 cm-5.0 cm) en el segundo trimestre; y (ii) membrana intacta y; (iii) sin contracciones laborales (una vez cada 10 minutos). Las mujeres fueron excluidas de este estudio si tenían: (a) un embarazo múltiple ( $\geq 2$  fetos); o (b) coito o aplicación de cualquier aplicación vaginal 48 horas antes de la recolección de la muestra de hisopo cervical; o (c) usó antibióticos/antimicóticos 30 días antes de la recolección de la muestra de hisopo cervical.

### 50 Recolección de hisopos cervicales

55 Antes del tratamiento con cerclaje, se recolectó una muestra de hisopo cervical del paciente con IC girando una hisopo Dacron estéril 360° una vez en el orificio externo. Para el cérvix dilatada, el hisopo se obtuvo del sistema operativo en la posición de las 12 en punto frente al médico. La muestra de hisopo se recogió inmediatamente después de la apertura del tracto reproductivo por espéculo, y antes de cualquier otro procedimiento. Se aplicaron técnicas antisépticas. Se tuvo especial cuidado para evitar que el hisopo entre en contacto con cualquier parte del tracto reproductivo (por ejemplo, vagina, labios), que no sea la ubicación designada del cuello uterino.

## Extracción de ADN bacteriano

Se extrajo el ADN genómico bacteriano en cada muestra de acuerdo con un protocolo publicado, que se optimizó para mantener una buena representación de taxones bacterianos en una muestra (Yuan et al., PLoS One 7, e33865 (2012)).

## Clasificación taxonómica

- 5 Cada Otu se clasifica taxonómicamente a nivel de género utilizando Naive Bayesian ARNr Classifier del Ribosomal Database Project (RDP) (Versión 2.9, septiembre de 2014, RDP 16S ARNr conjunto de entrenamiento 10). Los *Lactobacillus* se comparan con la base de datos 16S ARNr (GenBank) utilizando BLAST (puntaje más alto) y MOLE-BLAST (la mejor alineación múltiple de coincidencias BLAST) para obtener la información de la especie.

## Referencias

- 10
- Aagaard, K., J. Ma, K. M. Antony, R Ganu, J. Petrosino and J. Versalovic (2014). "The placenta harbors a unique microbiome." *Sci Transl Med*, 6(237): 237-265.
- 15 Aagaard, K., K. Riehle, J. Ma, N. Segata, T. A Mistretta, C. Coarfa, S. Raza, S. Rosenbaum, I. Van den Veyver, A. Milosavljevic, D Gevers, C. Huttenhower, J. Petrosino and J. Versalovic (2012). "A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy." *PLoS One* 7(6): e36466.
- Claesson, M. J., Q. Wang, O. O'Sullivan, R. Greene-Diniz, J. R. Cole, R. P. Ross and P. W. O'Toole (2010). "Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S ARNr gene regions." *Nucleic Acids Res* 38(22): e200.
- 20 Gajer, P., R. M. Brotman, G. Bai, J. Sakamoto, U. M. Schutte, X. Zhong, S. S. Koenig, L. Fu, Z. S. Ma, X. Zhou, Z. Abdo, L. J. Forney and J. Ravel (2012). "Temporal dynamics of the human vaginal microbiota." *Sci Transl Med* 4(132): 132-152.
- Gerber, S., Y. Vial, P. Hohlfeld and S. S. Witkin (2003). "Detection of *Ureaplasma urealyticum* in second-trimester amniotic fluid by polymerase chain reaction correlates with subsequent preterm labor and delivery." *J Infect Dis* 187(3): 518-521.
- 25 Gray, D. J., H. B. Robinson, J. Malone and R. B Thomson, Jr. (1992). "Adverse outcome in pregnancy following amniotic fluid isolation of *Ureaplasma urealyticum*." *Prenat Diagn* 12(2): 111-117.
- Hummelen, R., A. D. Fernandes, J. M. Macklaim, R. J. Dickson, J. Changalucha, G. B. Gloor and G. Reid (2010). "Deep sequencing of the vaginal microbiota of women with HIV." *PLoS One* 5(8): e12078.
- 30 Oh, K. J., S. E. Lee, H. Jung, G. Kim, R. Romero and B. H. Yoon (2010). "Detection of ureaplasmas by the polymerase chain reaction in the amniotic fluid of patients with cervical insufficiency." *J Perinat Med* 38(3): 261-268.
- Owen, J., G. Hankins, J.D. Iams, V. Berghella, J. S. Sheffield, A. Perez-Delboy, R. S. Egennan, D. A. Wing, M. Tomlinson, R. Silver, S.M. Ramin, E. R. Guzman, M. Gordon, H. Y. How, E. J. Knudtson, J. M Szychowski, S. Cliver and J. C. Hauth (2009). "Multicenter randomized trial of cerclage for preterm birth prevention in high-risk women with shortened midtrimester cervical length." *Am J Obstet Gynecol* 201(4): 375 e371-378.
- 35 Perni, S.C., S. Vardhana, I. Korneeva, S. L. Tuttle, L. R. Paraskevas, S. T. Chasen, R. B. Kalish and S. S. Witkin (2004). "*Mycoplasma hominis* and *Ureaplasma urealyticum* in midtrimester amniotic fluid: association with amniotic fluid cytokine levels and pregnancy outcome." *Am J Obstet Gynecol* 191(4): 1382-1386.
- Quince, C., A. Lanzen, R. J. Davenport and P. J. Turnbaugh (2011). "Removing noise from pyrosequenced amplicons." *BMC Bioinformatics* 12: 38.
- 40 Ravel, J., P. Gajer, Z. Abdo, G. M. Schneider, S. S. Koenig, S. L. McCulle, S. Kadebach, R. Gorle, J. Russell, C. O. Tacket, R. M. Brotman, C. C. Davis, K. Ault, L. Peralta and L. J. Forney (2011). "Vaginal microbiome of reproductive-age women." *Proc Natl Acad Sci USA* 108 Suppl 1: 4680-4687.
- Romero, R., R. Gonzalez, W. Sepulveda, F. Brandt, M Ramirez, Y. Sorokin, M. Mazor, M. C. Treadwell and D. B. Cotton (1992). "Infection and labor. VIII. Microbial invasion of the amniotic cavity in patients with suspected cervical incompetence: prevalence and clinical significance." *Am J Obstet Gynecol* 167(4 Pt 1): 1086-1091.
- 45 Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G., Thallinger, D. J. Van Horn and C. F. Weber (2009). "Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities." *Appl Environ Microbiol* 75(23): 753 7-7541.

Storey, J.D. and R. Tibshirani (2003). "Statistical significance for genomewide studies." Proc Natl Acad Sci USA 100(16): 9440-9445.

Wang, Q., G. M. Garrity, J. M. Tiedje and J. R. Cole (2007). "Naive Bayesian classifier for rapid assignment of ARNr sequences into the new bacterial taxonomy." Appl Environ Microbiol 73(16):5261-5267.

5 White, J.R, N. Nagarajan and M. Pop. (2009) "Statistical methods for detecting differentially abundant features in clinical metagenomic samples." PLoS Comput Biol 5(4):e1000352.

Yuan, S., D. B. Cohen, J. Ravel, Z. Abdo and L. J. Forney (2012). "Evaluation of methods for the extraction and purification of DNA from the human microbiome." PLoS One 7(3):e33865.

10 Aunque la invención anterior se ha descrito con cierto detalle a modo de ilustración y ejemplo con fines de claridad de comprensión, un experto en la materia apreciará que se pueden practicar ciertos cambios y modificaciones dentro del alcance de las reivindicaciones adjuntas.

Tabla 1. Características clave de todos los participantes.

	Cérvix dilatada (n=19)	Cérvix cerrada (n=19)	valor p*
Edad gestacional en la recolección de hisopos (semanas) [media ±DE]	20.7 ± 3.3	18.8 ± 3.8	p=0.283
Historia previa de aborto espontáneo de parto prematuro (n, %)	3 (16%)	0 (0%)	p=0.253
Longitud cervical en la recolección de hisopos (mm) [media ± DE]	7.9 ± 11.9	23.9 ± 8.8	p<0.001
Número de mujeres con cérvix dilatada (n, %) [Dilatación media ± DE]	19 (100%) [1.9 cm ± 1.25 cm]	0 (0%)	p<0.001
Número de mujeres sometidas a cerclaje cervical (n, %)	19 (100%)	13 (100%)	p=1.00
Número de mujeres con resultados básicos del embarazo** (n, %)	18 (95%)	13 (100%)	p=1.00
* Prueba de suma de rango de Mann-Whitney para variables continuas. Prueba exacta de Fisher para variables categóricas.			
** Edad gestacional al momento del parto, modo de parto y peso al nacer.			

15 Tabla 2A. Datos de secuenciación de los taxones bacterianos diferencialmente abundantes (significativamente incrementados) identificados al comparar los 342 taxones en los 19 cervices dilatados versus los 13 cerrados (normalizados por submuestreo aleatorio).

Taxón	Abundancia relativa media (dilatada)	Error estándar (dilatada)	Abundancia relativa media (cerrada)	Error estándar (cerrada)	Cambio de eventos (aumentado)	Valor p	Valor p ajustado *
d-c-019	1.10E-03	1.10E-03	0.00E+00	0.00E+00	33.2	9.99E-04	6.85E-03
d-c-030	3.67E-04	3.67E-04	0.00E+00	0.00E+00	11.1	9.99E-04	6.85E-03
d-c-037	1.25E-04	8.10E-05	0.00E+00	0.00E+00	3.8	1.16E-04	2.65E-03

ES 2 767 527 T3

Taxón	Abundancia relativa media (dilatada)	Error estándar (dilatada)	Abundancia relativa media (cerrada)	Error estándar (cerrada)	Cambio de eventos (aumentado)	Valor p	Valor p ajustado *
d-c-040	1.25E-04	1.25E-04	0.00E+00	0.00E+00	3.8	1.16E-04	2.65E-03
d-c-043	1.25E-04	8.50E-05	0.00E+00	0.00E+00	3.8	1.16E-04	2.65E-03
d-c-045	1.11E-04	7.80E-05	0.00E+00	0.00E+00	3.4	3.93E-04	6.74E-03
d-c-038	1.04E-04	9.00E-05	0.00E+00	0.00E+00	3.1	7.37E-04	6.85E-03
d-c-047	8.30E-05	7.00E-05	0.00E+00	0.00E+00	2.5	2.29E-03	1.21E-02
d-c-054	8.30E-05	8.30E-05	0.00E+00	0.00E+00	2.5	2.29E-03	1.21E-02

\* Ajustado para múltiples pruebas por el método False Discovery Rate.

Tabla 2B. Los datos de secuenciación de los taxones bacterianos diferencialmente abundantes (significativamente disminuidos) identificados comparando los 342 taxones en los 19 cervices dilatados versus los 13 cerrados (normalizados por submuestreo aleatorio).

Taxón	Abundancia relativa media (dilatada)	Error estándar (dilatada)	Abundancia relativa media (cerrada)	Error estándar (cerrada)	Cambio de eventos (aumentado)	Valor p	Valor p ajustado *
d-c-067	7.00E-06	7.00E-06	9.10E-05	6.20E-05	13.0	1.91E-03	1.19E-02
d-c-039	0.00E+00	0.00E+00	2.43E-04	2.43E-04	7.4	9.99E-04	6.85E-03
d-c-052	0.00E+00	0.00E+00	1.82E-04	1.82E-04	5.5	9.99E-04	6.85E-03
d-c-074	0.00E+00	0.00E+00	8.10E-05	8.10E-05	2.5	7.42E-04	6.85E-03
d-c-082	0.00E+00	0.00E+00	6.10E-05	6.10E-05	1.8	4.50E-03	2.20E-02
d-c-092	0.00E+00	0.00E+00	5.10E-05	5.10E-05	1.5	1.11E-02	4.74E-02
d-c-098	0.00E+00	0.00E+00	5.10E-05	5.10E-05	1.5	1.11E-02	4.74E-02

\*Ajustado para múltiples pruebas por el método False Discovery Rate.

5

Tabla 2C. Datos de secuenciación de los taxones bacterianos diferencialmente abundantes (significativamente incrementados) identificados al comparar los 349 taxones en las 10 cérvix dilatadas versus las 10 cerradas (1:1 emparejado por edad gestacional en la recolección de la muestra, emparejado a <2 semanas, normalizado como proporción de recuento de lectura total para cada muestra).

ES 2 767 527 T3

Taxón	Abundancia relativa media (dilatada)	Error estándar (dilatada)	Abundancia relativa media (cerrada)	Error estándar (cerrada)	Cambio de eventos (aumentado)	Valor p	Valor p ajustado*
d-c-012	1.75E-03	1.75E-03	0.00E+00	0.00E+00	53.0	9.99E-04	7.38E-03
d-c-030	5.93E-04	5.93E-04	0.00E+00	0.00E+00	18.0	9.99E-04	7.38E-03
d-c-040	1.49E-04	1.49E-04	0.00E+00	0.00E+00	4.5	9.99E-04	7.38E-03
d-c-047	1.20E-04	8.50E-05	0.00E+00	0.00E+00	3.6	9.99E-04	7.38E-03
d-c-050	1.07E-04	8.50E-05	0.00E+00	0.00E+00	3.2	9.99E-04	7.38E-03
d-c-081	7.80E-05	7.80E-05	0.00E+00	0.00E+00	2.4	6.09E-04	7.38E-03
d-c-053	7.40E-05	6.70E-05	0.00E+00	0.00E+00	2.2	9.99E-04	7.38E-03
d-c-015	7.00E-05	7.00E-05	0.00E+00	0.00E+00	2.1	1.39E-03	7.68E-03
d-c-088	7.00E-05	7.00E-05	0.00E+00	0.00E+00	2.1	1.39E-03	7.68E-03
d-c-072	5.60E-05	4.40E-05	0.00E+00	0.00E+00	1.7	9.99E-04	7.38E-03
d-c-105	5.20E-05	5.20E-05	0.00E+00	0.00E+00	1.6	7.19E-03	3.35E-02
d-c-068	3.60E-05	3.60E-05	0.00E+00	0.00E+00	1.1	9.99E-04	7.38E-03
d-c-071	3.60E-05	3.60E-05	0.00E+00	0.00E+00	1.1	9.99E-04	7.38E-03
d-c-083	2.60E-05	2.60E-05	0.00E+00	0.00E+00	0.8	1.39E-03	7.68E-03
d-c-087	2.60E-05	2.60E-05	0.00E+00	0.00E+00	0.8	1.39E-03	7.68E-03

\*Ajustado para múltiples pruebas por el método False Discovery Rate.

Tabla 3A. Clasificación a nivel de género de los taxones bacterianos diferencialmente abundantes (significativamente incrementados) identificados al comparar los 342 taxones en los 19 cervices dilatados versus los 13 cerrados (normalizados por submuestreo aleatorio).

Taxón	Clasificación taxonómica del taxón identificado a nivel de reino, filum, clase, familia, orden y género con puntaje de clasificación (entre paréntesis).
d-c-019	Bacterias (100); Firmicutes (100); Bacilos (100); Lactobacillales (100); Lactobacillaceae (100); Lactobacillus (100);
d-c-030	Bacterias (100); Firmicutes (100); Negativicutes (100); Selenomonadales (100); Veillonellaceae (100); Veillonella (100);
d-c-037	Bacterias (100); Firmicutes (100); Negativicutes (100); Selenomonadales (100); Veillonellaceae (100); Sin clasificación (93);

ES 2 767 527 T3

Taxón	Clasificación taxonómica del taxón identificado a nivel de reino, filum, clase, familia, orden y género con puntaje de clasificación (entre paréntesis).
d-c-040	Bacterias (100); "Actinobacterias" (100); Actinobacterias (100); Bifidobacteriaceae (100); Bifidobacteriaceae (100); Alloscardovia (100);
d-c-043	Bacterias (100); "Tenericutes" (100); Mollicutes (100); Micoplasmatales (100); Mycoplasmataceae (100); Ureaplasma (100);
d-c-045	Bacterias (100); "Actinobacterias" (100); Actinobacterias (100); Coriobacteriaceae (100); Coriobacteriaceae (100); Atopobio (100);
d-c-038	Bacterias (100); "Actinobacterias" (100); Actinobacterias (100); Coriobacteriaceae (100); Coriobacteriaceae (100); sin clasificar (100);
d-c-047	Bacterias (100); Firmicutes (100); Clostridia (100); Clostridiales (100); Clostridiales_Incertae_Sedis_XI (100); Anaerococo (100);
d-c-054	Bacterias (100); Firmicutes (100); Erysipelotrichia (100); Erysipelotrichales (100); Erysipelotrichaceae (100); Bulleidia (100);
Nota: El género "sin clasificar" representa un género novedoso no registrado previamente en la base de datos utilizada para la clasificación taxonómica, pero recientemente identificado por secuenciación en este estudio.	

Tabla 3B. Clasificación a nivel de género de los taxones bacterianos diferencialmente abundantes (significativamente disminuidos) identificados al comparar los 342 taxones en los 19 dilatados versus los 13 cerrados

Taxón	Clasificación taxonómica del taxón identificado en el nivel de reino, filum, clase, familia, orden y género con puntaje de clasificación (entre paréntesis).
d-c-067	Bacterias (100); sin clasificar (100);
d-c-039	Bacterias (100); "Sinergistetes" (100); Sinergia (100); Sinergistales (100); Synergistaceae (100); Jonquetella (100);
d-c-052	Bacterias (100); Firmicutes (100); Bacilos (100); Lactobacillales (100); Aerococcaceae (100); Aerococcus (100);
d-c-074	Bacterias (100); Firmicutes (100); Clostridia (100); Clostridiales (100); Clostridiales_Incertae_Sedis_XI (100); sin clasificar (100);
d-c-082	Bacterias (100); "Actinobacterias" (100); Actinobacterias (100); Actinomicetales (100); Corynebacteriaceae (100); Corynebacterium (100);
d-c-092	Bacterias (100); sin clasificar (100);
d-c-098	Bacterias (100); sin clasificar (100);
Nota: El género "sin clasificar" representa un género novedoso no registrado previamente en la base de datos utilizada para la clasificación taxonómica, pero recientemente identificado por secuenciación en este estudio.	

ES 2 767 527 T3

Tabla 3C. Datos de clasificación a nivel de género de los taxones bacterianos diferencialmente abundantes (significativamente incrementados) identificados al comparar los 349 taxones en las 10 cérvix dilatadas versus las 10 cerradas (1:1 emparejado por edad gestacional en la recolección de la muestra, emparejado a <2 semanas, normalizado como relación del recuento total de lecturas para cada muestra).

Taxón	Clasificación taxonómica del taxón identificado a nivel de reino, filum, clase, familia, orden y género con puntaje de clasificación (entre paréntesis).
d-c-012	Bacterias (100); "Tenericutes" (100); Mollicutes (100); Micoplasmatales (100); Mycoplasmataceae (100); Micoplasma (100);
d-c-030	Bacterias (100); Firmicutes (100); Negativicutes (100); Selenomonadales (100); Veillonellaceae (100); Veillonella (100);
d-c-040	Bacterias (100); "Actinobacterias" (100); Actinobacterias (100); Bifidobacteriaceae (100); Bifidobacteriaceae (100); Alloscardovia (100);
d-c-047	Bacterias (100); Firmicutes (100); Clostridia (100); Clostridiales (100); Clostridiales_Incertae_Sedis_XI (100); Anaerococo (100);
d-c-050	Bacterias (100); "Actinobacterias" (100); Actinobacterias (100); Actinomicetales (100); Propionibacteriaceae (100); Propionimicrobium (100);
d-c-053	Bacterias (100); "Actinobacterias" (100); Actinobacterias (100); Actinomicetales (100); Corynebacteriaceae (100); Corynebacterium (100);
d-c-068	Bacterias (100); Firmicutes (100); Clostridia (91); Clostridiales (91); sin clasificar (91); sin clasificar (91);
d-c-071	Bacterias (100); "Actinobacterias" (100); Actinobacterias (100); Actinomicetales (100); Actinomycetaceae (100); Actinobaculum (100);
d-c-072	Bacterias (100); Firmicutes (100); Clostridia (100); Clostridiales (100); Incertae_Sedis_XI (100); Murdochiella (100);
d-c-081	Bacterias (100); sin clasificar (100);
d-c-015	Bacterias (100); Firmicutes (100); Negativicutes (100); Selenomonadales (100); Veillonellaceae (100); Megasphaera (98);
d-c-083	Bacterias (100); Firmicutes (100); Clostridia (100); Clostridiales (100); sin clasificar (100); sin clasificar (100);
d-c-087	Bacterias (100); Firmicutes (100); Clostridia (100); Clostridiales (100); Ruminococcaceae(100); Fastidiosipila (100);
d-c-088	Bacterias (100); "Acidobacterias" (100); Acidobacteria_Gp1 (100); Gp1 (100); sin clasificar (100); sin clasificar (100);
d-c-105	Bacterias (100); sin clasificar (100);

Nota: El género "sin clasificar" representa un género novedoso no registrado previamente en la base de datos utilizada para la clasificación taxonómica, pero recientemente identificado por secuenciación en este estudio.

ES 2 767 527 T3

Tabla 3D. Clasificación a nivel de género de los taxones bacterianos diferencialmente abundantes (significativamente disminuidos) identificados al comparar los 349 taxones en las 10 cérvix dilatadas frente a las 10 cerradas (1:1 emparejado por edad gestacional en la recolección de la muestra, emparejado a <2 semanas, normalizado como proporción del recuento total de lecturas para cada muestra).

Taxón	Clasificación taxonómica del taxón identificado a nivel de reino, filum, clase, familia, orden y género con puntaje de clasificación (entre paréntesis).
d-c-018	Bacterias (100); Firmicutes (100); Bacilos (100); Lactobacillales (100); Lactobacillaceae (100); Lactobacillus (99);
d-c-052	Bacterias (100); Firmicutes (100); Bacilos (100); Lactobacillales (100); Aerococcaceae (100); Aerococcus (100);
d-c-074	Bacterias (100); Firmicutes (100); Clostridia (100); Clostridiales (100); Clostridiales_Incertae_Sedis_XI (100); sin clasificar (100);
d-c-076	Bacterias (100); Firmicutes (100); Bacilos (100); Lactobacillales (100); Lactobacillaceae (100); Lactobacillus (100);
d-c-082	Bacterias (100); "Actinobacterias" (100); Actinobacterias (100); Actinomycetales (100); Corynebacteriaceae (100); Corynebacterium (100);

Nota: El género "sin clasificar" representa un género novedoso no registrado previamente en la base de datos utilizada para la clasificación taxonómica, pero recientemente identificado por secuenciación en este estudio.

5

Tabla 4A. Secuencias del gen 16S ARNr de taxones bacterianos de los taxones bacterianos diferencialmente abundantes (significativamente incrementados) identificados al comparar los 342 taxones en los 19 cervices dilatados versus los 13 cerrados (normalizados por submuestreo aleatorio).

Taxón	Secuencia genómica de ARN ribosómico 16S
d-c-019	GGAGCGCAGGCGGTTTTTTAAGTCTGATGTGAAAGCCCTCGGCTTAACCGAGGAAGCGCA TCGGAAACTGGGAACTTGAGTGACAGAAGAGGACAGTGGAACCTCATGTGTAGCGGTGAAA TGCGTAGATATATGGAAGAACACCCAGTGGCGAAGGCGGCTGTCTGGTCTGTAAGTACTGACGCT
d-c-030	CGCGCGCAGGCGGACTAGCCAGTCAGTCTTAAAAGTTCGGGGCTTAACCCCGTGATGGGA TTGAACTACTAGTCTAGAGTATCGGAGAGGAAAGTGAATTCCTAGTGTAGCGGTGAAAT GCGTAGATATTAGGAAGAACACCCAGTGGCGAAGGCGACTTTCTGGACGAACACTGACGCT
d-c-037	GGCGCGCAGGCGGTTTCGGTAAGTCTGTCTTAAAGTGCAGGCTTAACCCCGTGAGGGACGG AAACTGTGCAACTTGAGTGTGCGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCG TAGATATTAGGAGGAACACCGGTGGCGAAGGCGGCTTTCTGGACGAACACTGACGCT
d-c-040	AGATGAGATGGCGGTTTTGTGCGCTCTGGTGTGAAAGTCCATCGCTTAACGGTGGATTGGC GCTGGGTACGGGCAGGCTAGAGTGTAGTAGGGGAGACTGGAATTCCTCGGTGTAACGGTGG AATGTGTAGATATCGGGAAGAACACCAATGGCGAAGGCAGGTCTCTGGGCTATTACTGACG CT
d-c-043	CGAGCGCAGGCGGTTTTGTAAGTTTGGTATTAATCTAGATGCTTAACGTCTAGCTGTATCA AAAAGTGTAAACCTAGAGTGTAGTAGGGAGTTGGGGAACCTCATGTGGAGCGGTAAAATGC GTAGATATATGGAAGAACACCGGTGGCGAAGGCGCAACTTGGACTATCACTGACGCT

ES 2 767 527 T3

Taxón	Secuencia genómica de ARN ribosómico 16S
d-c-045	CGCGCGTAGGCGGTCTGTTAGGTCAGGAGTCAAATCTGGGGGCTCAACCCCTATCCGCTC CTGATACCGGCAGGCTTGAGTCTGGTATGGGAAGGTGGAATTCOAAGTGTAGCGGTGAAAT GCGCAGATATTTGGAAGAACACCAGTGGCGAAGGCGGCCTTCTGGGCCATGACTGACGCT
d-c-038	CGCGCGCAGGCGGTTGCTCAAGCGGAACCTCTAATCTCGGGGCTTAACCTCGAGCCGGGT TCCGAAGTGGACGACTCGAGTGCAGTGGAGGAGGAGATGGAATTCOAGTGTAGCGGTGAA TGCGCAGATATCGGGAAGAACACCAACGGCGAAGGCAGTCTGCTGGGCCGTCCTGACGC T
d-c-047	AGTACGTAGGCGGCCTAGTAAGTTAGAAGTAAAAGAATATAGCTCAACTATATAAAGCTTTT AAAAGTGTAGGCTTGAGAGATGAAAGGGAAAGTGAATTCCTAGTGTAGCGGTGAAATGC GCAGATATTAGGAAGAATACCGGTGGCGAAGGCGACTTTCTGGTCATCATCTGACGCT
d-c-054	GGTGCGTAGGCGGTCTGTTAAGTTCATGGTTAAATTTGGGGCTCAACCCCATGAGCCAT GGATACTGGCAGACTAGAGTATTGGAGAGGCAAGCGGAATTCATGTGTAGCGGTAAAATG CGTAGATATATGGAGGAACACCGGTGGCGAAGGCGGCTTGCTAGCCAAAGACTGACGCT
d-c-019 (SEQ ID NO: 3); d-c-030 (SEQ ID NO: 4); d-c-037 (SEQ ID NO: 5); d-c-040 (SEQ ID NO: 6); d-c-043 (SEQ ID NO: 7); d-c-045 (SEQ ID NO: 8); d-c-038 (SEQ ID NO: 9); d-c-047 (SEQ ID NO: 10); d-c-054 (SEQ ID NO: 11).	

Tabla 4B. Secuencias del gen 16S ARNr de taxones bacterianos de los taxones bacterianos diferencialmente abundantes (significativamente disminuidos) identificados al comparar los 342 taxones en los 19 cervices dilatados versus los 13 cerrados.

Taxón	Secuencia genómica de ARN ribosómico 16S
d-c-067	GGTGCGCAGGCGGCTTTACAAGTTGGATGTGAAATATTGTGGCTCAACCACAAACGTGCAT CCAAAAGTGCAAAGCTTGAGTTAAGGAGAGGTAAGTGAATTCCTGGTGTAGCGGTGGAAT GCGTAGATATCAGGAGGAATACCGGTGGCGAAGGCGACTTACTGGACTTAAACTGACGCT
d-c-039	GGCGCGTAGGCGGAATGGCAAGTCAGCAAGTAAAAGCGTGGGGCTCAACCCCATGATGCG GCTGAAACTGTTATTCTAGAGGCATGGAGAGGCAAACCGGAATTCOAGTGTAGCGGTGAA TGCGTAGATATCGGGAAGAACACCAGTGGCGAAGGCGGTTTTGCTGGCCATGAACTGACGC T
d-c-052	TGGGCGCAGGCGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTTAACCGTGAAGTGCAT TGGAAGTGGGGAAGTGGAGTACAGAAGAGGAAAGTGAAGTCCATGTGTAGCGGTGGAAT GCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTCACTGACGCT
d-c-074	CGTTTCGAGGCGGCAATGCAAGTCTCGTGTGAAAGGCAAGGGCTCAACCCCTTGTAAAGCAC AAGAAAGTGCATAGCTTGAGTAGTGGAGAGGCAAGTGAATTCCTAGTGTAGCGGTGAAAT GCGTAGATATTAGGAGGAATACCGGTGGCGAAGGCGACTTGCTGGACACAAACTGACGCT
d-c-082	GGCTCGTAGGTGGTTTGTGCGTCGTCTGTGAAATTCOAGGCTTAACCTCCGGGCGTGCA GGCGATACGGGCATAACTTGAGTACTGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGA AATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTCTCTGGGCAGTAACTGAC GCT

ES 2 767 527 T3

Taxón	Secuencia genómica de ARN ribosómico 16S
d-c-092	CGCGTGTAGGCGGATTCTCAAGTTGGATGTGAAACCCCTTGGCTAAACTGAGGGCTTGCAT TCAAACTGAGGACCTTGAGTATCAGAGGGGAAAGTGGAAATCCTGGTGGAGCGGTAAAAT GCGTAGAGATCAGGAGGAACACCGGTGGCGAAGGCGACTTTCTGGCTGACAACTGACGCT
d-c-098	TGTGTGTAGGTGGCGTGATTAGTCGTTTGTGAAAGATCCGAGCTTAACTTGGAAAACGCGA ACGAAACGGTCATGCTTGAGTATGTGAGAGGTAAGCAGAACTCATGGTGTAGGGGTGAAAT CCGTTGATATCATGGGGAATACCAAAGCGAAGGCAGCTTACTGGCACATTACTGACACT
d-c-067 (SEQ ID NO: 12); d-c-039 (SEQ ID NO: 3); d-c-052 (SEQ ID NO: 14); d-c-074 (SEQ ID NO: 15); d-c-082 (SEQ ID NO: 16); d-c-092 (SEQ ID NO: 17); d-c-098 (SEQ ID NO: 18).	

Tabla 4C. Secuencia del gen 16S ARNr de taxones bacterianos de los taxones bacterianos diferencialmente abundantes (aumento significativo de taxones identificados al comparar los 349 taxones en las 10 cérvix dilatadas frente a las 10 cerradas (1:1 emparejado por edad gestacional en la recolección de la muestra, emparejado a <2 semanas, normalizado como proporción del recuento total de lecturas para cada muestra).

5

Taxón	Secuencia genómica de ARN ribosómico 16S
d-c-012	GGTTCGTAGGCTGTTTGTAAAGTCTGGAGTTAAATCCCGGGGCTCAACCCCGGCTCGCTTT GGATACTAGCAAAGTCTAGAGTTAGATAGAGGTAAGCGGAATTCATGTGAAGCGGTGAAATG CGTAGATATATGGAAGAACACCAAAGGCGAAGGCAGCTTACTGGGTCTATACTGACGCT
d-c-030	CGCGCGCAGGCGGACTAGCCAGTCAGTCTTAAAAGTTCGGGGCTTAACCCCGTGATGGGA TTGAACTACTAGTCTAGAGTATCGGAGAGGAAAGTGGAAATCCTAGTGTAGCGGTGAAAT GCGTAGATATTAGGAAGAACACCAAGTGGCGAAGGCGACTTTCTGGACGAACACTGACGCT
d-c-040	AGATGAGATGGCGGTTTGTGCGTCTGGTGTGAAAGTCCATCGCTTAACGGTGGATTGGC GCTGGGTACGGGCAGGCTAGAGTGTAGTAGGGGAGACTGGAATTCCTGGTGTAAACGGTGG AATGTGTAGATATCGGGAAGAACACCAATGGCGAAGGCAGGTCTCTGGGCTATTACTGACG CT
d-c-047	AGTACGTAGGCGGCCTAGTAAGTTAGAAGTGAAGAATATAGCTCAACTATATAAAGCTTTT AAAAGTGTAGGCTTGGAGATGAAAGGGAAAGTGGAAATCCTAGTGTAGCGGTGAAATGC GCAGATATTAGGAAGAATACCGGTGGCGAAGGCGACTTTCTGGTCATCATCTGACGCT
d-c-050	GGCTTGTAGGCGGTTTGTGCGTCGAAAGTGTAAACTCAGTGCTTAACGCTGAGCCTGCTT TCGATACGGGCTGACTAGAGGAAGGTAGGGGAGAATGGAATTCGGGTGGAGCGGTGGAA TGCGCAGATATCGGGAGGAACACCAGTGGCGAAGGCGGTTCTCTGGACCTTCTCTGACGC
d-c-053	CGCTCGTAGGTGGTGTGTTGCGTCTGTGTAATCCAGGGGCTTAACTTTTGGTTGGCAG GCGATACGGGCATTGCTTGAGTGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAA ATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCGAGGTCTCTGGGCAGTTACTGACG
d-c-068	TGAGCGTAGGCGGCAGTACAAGTCTGGGAGTGAAGTCTGGGGCTCAACCCCGAGACTGCT CTCGAAACTGTACAGCTAGAGTGCAGGATGGGCGAGGCGGAATTCCTGGTGTAGCGGTGAA ATGCGTAGATATCAGGAGGAACACCGGTGGCGAAGGCGGCCTGCTGGACTGTAAGTACG CT

ES 2 767 527 T3

Taxón	Secuencia genómica de ARN ribosómico 16S
d-c-071	GGCTTGTAGGCGGCTTGTGCGCCTGCTGTGAAAACGCGGGGCTTAACCTTCGCGCGTGCA GTGGGTACGGCAGGCTTGAGTGTGGTAGGGGTGACTGGAATTCAGGTGTAGCGGTGG AATGCGCAGATATCTGGAGGAACACCGATGGCGAAGGCAGGTCAGTGGGCCATTACTGAC GCT
d-c-072	GGTACGTAGGCGGTTTGTAAAGTTTGGCGTTAAATCACGGGGCTCAACCCCGTTCAGCGTT GAAAACCTGGCAAACCTTGAGTAGTAGAGGGGACAGTGAATTCCTAGTGTAGCGGTGAAATG CGTAGAGATTAGGAAGAATACCGGTGGCGAAGGCAGTGTCTGGATACATACTGACGCT
d-c-081	GGTGTGTAGGCGGTTATATTAGTCATTTGTTAAATCCTCGGGCTTAACCCGAGAATCGCGA GCGAAACGGTATAACTAGAAAGTGTGAGGGGTGTACAGAACTCATGGTGTAGGGGTGAAAT CCGTTGATATCATGGGGAATACCAAAGCGAAGGCAGTACACTGGCACATATTTGACGCT
d-c-015	TGCGCGCAGGCGGTTTCGGTAAGTCTGTCTTAAAAGTGCAGGGGCTTAACCCCGTGAAGGGGA CGGAAACTGTCAACTTGAGTGTCCGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAAT GCGTAGATATTAGGAGGAACACCGGTGGCGAAAGCGGCTTTCTGGACGACAACCTGACGCT
d-c-083	GGTTCGCAGGCGGAATAACAAGTCAGATGTGAAAGGCATGGGCTCAACCCATGTAAGCATT TGAAACTGTAATTCTTGAGAAGTGGAGAGGTAAGTGAATTCCTAGTGTAGCGGTGAAATG CGTAGATATTAGGAGGAATACCTGTGGCGAAGGCAGTACTGGACACAAATCTGACGCT
d-c-087	CGCGTGTAGGCGGCACTGTAAGTCAGATGTGAAATCTCCCGGCTCAACCCGGAGCGTGCA TCTGATACTGCAATGCTTGAGTGTAGAGGGGAAAGCGGAATTCCTAGTGTAGCGGTAAAA TGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCAGGCTTTCTGGCTATTAACCTGACGCT
d-c-088	GGTGCGTAGGCGGTTTCGGTAAGTCTTGTGTGAAATCTTCAGGCTCAACTTGAAGTCTGCAC GAGAACTGCCGGGCTTGAGTGTGGGAGAGGTGAGTGAATTTCCGGTGTAGCGGTGAAA TGCGTAGATATCGGAAGGAACACCTGTGGCGAAAGCGGCTCACTGGACCACAACCTGACGC T
d-c-105	GGTGTGTAGGTGGTTATGTTAGTCTCCTTTCAAAGCTCCCGGCCTAACCCGGGAAAAGGGAG GGGAAACGGCACAACCTAGAGGATGCGAGGGGTCTGTGGAACCTCATGGAGTAGGGGTGAAA TCCGTTGATATCATGGGGAACACCAAAGCGAAGGCAGCAGACTGGCGCATTCTGACAC T
d-c-012 (SEQ ID NO: 19); d-c-030 (SEQ ID NO: 20); d-c-040 (SEQ ID NO: 21); d-c-047 (SEQ ID NO: 22); d-c-050 (SEQ ID NO: 23); d-c-053 (SEQ ID NO: 24); d-c-068 (SEQ ID NO: 25); d-c-071 (SEQ ID NO: 26); d-c-072 (SEQ ID NO: 27); d-c-081 (SEQ ID NO: 28); d-c-015 (SEQ ID NO: 29); d-c-0837 (SEQ ID NO: 30); d-c-087 (SEQ ID NO: 31); d-c-088 (SEQ ID NO: 32); d-c-105 (SEQ ID NO: 33).	

Tabla 4D. Secuencias del gen 16S ARNr de taxones bacterianos taxones bacterianos diferencialmente abundantes (significativamente disminuidos) identificados al comparar los 349 taxones en las 10 cérvix dilatadas frente a las 10 cerradas (1:1 emparejado por edad gestacional en la recolección de la muestra, emparejado a <2 semanas, normalizado como proporción del recuento total de lecturas para cada muestra).

5

Taxón	Secuencia genómica de ARN ribosómico 16S
d-c-018	CGAGTGCAGGCGGTTTTCTAAGTCTGATGTGAAAGCCTTCGGCTTAACCCGGAGAAGTGCAT CGGAAACTGGATAACTTGAGTGCAGAAGAGGGTAGTGAACCTCCATGTGTAGCGGTGGAAT GCGTAGATATATGGAAGAACACCAGTGGCGAAGGCAGGCTACCTGGTCTGCAACTGACGCT

ES 2 767 527 T3

Taxón	Secuencia genómica de ARN ribosómico 16S
d-c-052	TGGGCGCAGGCGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTTAACCGTGGAAGTGCAT TGGAAACTGGGGAACCTTGAGTACAGAAGAGGAAAGTGGAACCTCCATGTGTAGCGGTGGAAT GCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTCACTGACGCT
d-c-074	CGTTCGCAGGCGGCAATGCAAGTCTCGTGTGAAAGGCAAGGGCTCAACCCTTGTAAGCAC AAGAACTGCATAGCTTGAGTAGTGGAGAGGCAAGTGGAATTCCTAGTGTAGCGGTGAAAT GCGTAGATATTAGGAGGAATACCGGTGGCGAAGGCGACTTGCTGGACACAACTGACGCT
d-c-076	TGAGCGCAGGCGGTTTTTAAAGTCTGATGTGAAAGCCTTCGGCTTAACCGGAGAAGTGCAT CGGAACTGGGAGACTTGAGTGCAGAAGAGGACAGTGGAACTCCATGTGTAGCGGTGGAA TGCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGGCTGTCTAGTCTGTAAGTGCAGCT
d-c-082	GGCTCGTAGGTGGTTTGTGCGTCTGTGAAATTCGGGGCTTAACCTCCGGGCGTGCA GGCGATACGGGCATAACTTGAGTACTGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGA AATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTCTCTGGGCAGTAACTGAC GCT
d-c-018 (SEQ ID NO: 34); d-c-052 (SEQ ID NO: 35); d-c-074 (SEQ ID NO: 36); d-c-076 (SEQ ID NO: 37); d-c-082 (SEQ ID NO: 38).	

Tabla 4E. Clasificación de nivel de especie por alineación BLAST a la base de datos 16S ARNr de NCBI. Los taxones bacterianos diferencialmente abundantes (significativamente incrementados) identificados al comparar los 342 taxones en los 19 cervices dilatados versus los 13 cerrados (normalizados por submuestreo aleatorio) se alinearon.

5

Taxón	Mejor coincidencia de BLAST	Número de acceso de NCBI GenBank	Puntuación de alineación BLAST	Valor esperado	Identidad de secuencia alineada	Brechas
d-c-019	Lactobacillus casei ATCC 334 cepa ATCC 334 ARN ribosómico 16S, secuencia completa	NR_075032.1	335	4.00E-92	181/181(100%)	0/181(0%)
	Lactobacillus paracasei subsp. paracasei cepa NBRC 15889 16S gen de ARN ribosómico, secuencia parcial	NR_113337.1	335	4.00E-92	181/181(100%)	0/181(0%)
	Lactobacillus casei cepa NBRC 15883 16S gen de ARN ribosomal, secuencia parcial	NR_113333.1	335	4.00E-92	181/181(100%)	0/181(0%)
	Lactobacillus paracasei cepa ATCC 25302 16S gen de ARN ribosomal, secuencia parcial	NR_117987.1	335	4.00E-92	181/181(100%)	0/181(0%)

ES 2 767 527 T3

Taxón	Mejor coincidencia de BLAST	Número de acceso de NCBI GenBank	Puntuación de alineación BLAST	Valor esperado	Identidad de secuencia alineada	Brechas
	Lactobacillus paracasei subsp. tolerans cepa NBRC 15906 16S gen de ARN ribosomal, secuencia parcial	NR_041054.1	335	4.00E-92	181/181(100%)	0/181(0%)
	Lactobacillus casei cepa ATCC 393 16S gen de ARN ribosómico, secuencia parcial	NR_041893.1	335	4.00E-92	181/181(100%)	0/181(0%)
	Lactobacillus casei cepa JCM 1134 16S gen de ARN ribosómico, secuencia completa	NR_115534.1	335	4.00E-92	181/181(100%)	0/181(0%)
	Lactobacillus paracasei subsp. paracasei, cepa R094 16S, gen de ARN ribosómico, secuencia parcial	NR_025880.1	335	4.00E-92	181/181(100%)	0/181(0%)
	Lactobacillus zeae cepa RIA 482 16S gen de ARN ribosómico, secuencia parcial	NR_037122.1	335	4.00E-92	181/181(100%)	0/181(0%)
d-c-030	Cepa de Veillonella mentpelilerensis ADV 281.99 Gen de ARN ribosómico 16S, secuencia parcial	NR_028839.1	335	4.00E-92	181/181(100%)	0/181(0%)
d-c-037*	Megasphaera cerevisiae cepa DSM 20462 16S gen de ARN ribosómico, secuencia parcial	NR_113307.1	265	6.00E-71	169/181(93%)	3/181(1%)
	Megasphaera cerevisiae cepa VTT-E-85230 16S gen de ARN ribosómico, secuencia completa	NR_044650.1	265	6.00E-71	169/181(93%)	3/181(1%)
d-c-040	Alloscardovia omnicolens cepa CCUG 31649 16S ARN ribosómico qeno, secuencia completa	NR_042583.1	322	3.00E-88	174/174(100%)	0/174(0%)
d-c-043	Ureaplasma urealyticum serovar 10 str. ATCC 33699 cepa ATCC 33699 16S ARN ribosómico, secuencia completa	NR_102836.1	335	4.00E-92	181/181(100%)	0/181(0%)

ES 2 767 527 T3

Taxón	Mejor coincidencia de BLAST	Número de acceso de NCBI GenBank	Puntuación de alineación BLAST	Valor esperado	Identidad de secuencia alineada	Brechas
	Ureaplasma parvum serovar 3 str. ATCC 700970 cepa ATCC 700970 16S ARN ribosómico, secuencia completa	NR_074176.1	335	4.00E-92	181/181(100%)	0/181(0%)
	Ureaplasma parvum serovar 3 cepa ATCC 27815 16S gen de ARN ribosómico, secuencia completa	NR_074762.1	335	4.00E-92	181/181(100%)	0/181(0%)
	Ureaplasma parvum serovar 3 cepa ATCC 27815 16S gen de ARN ribosómico, secuencia parcial	NR_027532.1	335	4.00E-92	181/181(100%)	0/181(0%)
	Ureaplasma urealyticum serovar 8 cepa ATCC 27618 16S gen de ARN ribosómico, secuencia parcial	NR_041710.1	335	4.00E-92	181/181(100%)	0/181(0%)
d-c-045	Cepa de Atopobium vaginae DSM 15829 16S gen de ARN ribosómico, secuencia parcial	NR_117757.1	307	1.00E-83	176/181(97%)	0/181(0%)
	Atopobium vaginae cepa 961*00022/98 16S gen de ARN ribosómico, secuencia parcial	NR_029349.1	307	1.00E-83	176/181(97%)	0/181(0%)
d-c-038*	Parvibacter caecicola cepa NR06 16S gen de ARN ribosómico, secuencia parcial	NR_1173741	257	1.00E-68	167/181(92%)	0/181(0%)
d-c-047	Anaerococcus senegalensis cepa JC48 16S gen de ARN ribosómico, secuencia parcial	NR_118220.1	333	2.00E-91	180/180(100%)	0/180(0%)
d-c-054	Bulleidia extracta cepa W1219 16S gen de ARN ribosómico, secuencia parcial	NR_028773.1	302	4.00E-82	175/181(97%)	0/181(0%)

Notas:

\* representa un taxón con identidad de secuencia <97% como cualquier especie conocida y puede representar especies nuevas, no reportadas anteriormente, que ahora se identifican por secuenciación en este estudio.

ES 2 767 527 T3

Tabla 4F. Clasificación de nivel de especie por alineación BLAST a la base de datos 16S ARNr de NCBI. Taxones bacterianos diferencialmente abundantes (significativamente disminuidos) identificados comparando los 342 taxones en los 19 cervices dilatados versus los 13 cerrados.

Taxón	Mejor coincidencia de BLAST	Número de acceso de NCBI GenBank	Puntuación de alineación BLAST	Valor esperado	Identidad de secuencia alineada	Brechas
d-c-074*	Eubacterium aggregans cepa SR12 16S gen de ARN ribosómico, secuencia parcial	NR_024926.1	176	3E-44	112/120(93%)	2/120(1%)
	Acetobacterium bakii cepa DSM 8239 16S gen de ARN ribosómico, secuencia parcial	NR_026329.1	176	3E-44	112/120(93%)	2/120(1%)
	Eubacterium barkeri cepa ATCC 25849 16S gen de ARN ribosómico, secuencia completa	NR_044661.1	176	3E-44	112/120(93%)	2/120(1%)
d-c-039	Jonquetella anthropi cepa ADV 126 16S gen de ARN ribosómico, secuencia parcial	NR_044215.1	337	1E-92	182/182(100%)	0/182(0%)
d-c-052	Aerococcus urinae ACS-120-V-Col10a cepa ACS-120-V-Col10a 16S ARN ribosómico, secuencia completa	NR_074879.1	335	4E-92	181/181(100%)	0/181(0%)
d-c-067	[Bacteroides] coagulans cepa EUH 581-73 16S ribosomal gen de ARN, secuencia parcial	NR_104900.1	337	1E-92	182/182(100%)	0/182(0%)
	[Bacteroides] coagulans cepa JCM 12528 16S gen de ARN ribosómico, secuencia parcial	NR_113066.1	337	1E-92	182/182(100%)	0/182(0%)
d-c-082	Corynebacterium cepa lactis RW2-5 16S ribosomal gen de ARN, secuencia parcial	NR_104996.1	337	1E-92	182/182(100%)	0/182(0%)
	Corynebacterium amycolatum cepa CCUG 35685 16S gen de ARN ribosómico, secuencia parcial	NR_117605.1	337	1E-92	182/182(100%)	0/182(0%)
	Corynebacterium amycolatum 16S, gen de ARN ribosómico, secuencia parcial	NR_1170301	337	1E-92	182/182(100%)	0/182(0%)
	Corynebacterium freneyi cepa ISPB 6695110 16S gen de ARN ribosómico, secuencia parcial	NR_042025.1	337	1E-92	182/182(100%)	0/182(0%)

ES 2 767 527 T3

Taxón	Mejor coincidencia de BLAST	Número de acceso de NCBI GenBank	Puntuación de alineación BLAST	Valor esperado	Identidad de secuencia alineada	Brechas
	Corynebacterium xerosis cepa DSM 20743 16S gen de ARN ribosómico, secuencia parcial	NR_119180.1	337	1E-92	182/182(100%)	0/182(0%)
	Corynebacterium xerosis cepa ATCC 373 16S gen de ARN ribosómico, secuencia parcial	NR_026213.1	337	1E-92	182/182(100%)	0/182(0%)
	Corynebacterium amycolatum cepa S160 16S gen de ARN ribosómico, secuencia completa	NR_026215.1	337	1E-92	182/182(100%)	0/182(0%)
d-c-092*	Desulfococcus oleovorans Hxd3 cepa Hxd3 16S ARN ribosómico, secuencia completa	NR_075012.1	126	3E-29	84/92(91%)	0/92(0%)
	Desulfosalsimonas propionica cepa PropA 16S gen de ARN ribosómico, secuencia parcial	NR_1156781	126	3E-29	84/92(91%)	0/92(0%)
d-c-098 *	Bacteria no cultivada clon F64Q2U401BWA69 16S gen de ARN ribosómico, secuencia parcial	HQ145117.1	252	1E-36	166/181(92%)	0/181(0%)
Notas:						
* representa un taxón con identidad de secuencia <97% como cualquier secuencia conocida y puede representar especies nuevas, no reportadas previamente, que ahora se identifican por secuenciación en este estudio.						

5 Tabla 4G. Clasificación de nivel de especie por alineación BLAST a la base de datos 16S ARNr de NCBI. Se alinearon taxones bacterianos diferencialmente abundantes (significativamente incrementados) identificados al comparar los 349 taxones en las 10 cérvix dilatadas versus las 10 cerradas (1:1 emparejado por edad gestacional en la recolección de la muestra, emparejado a <2 semanas, normalizado como proporción del recuento total de lecturas para cada muestra).

Taxón	Mejor coincidencia de BLAST	Número de acceso de NCBI GenBank	Puntuación de alineación BLAST	Valor esperado	Identidad de secuencia alineada	Brechas
d-c-012	Mycoplasma hominis cepa NBRC 14850 16S gen de ARN ribosómico, secuencia parcial	NR_113679.1	333	2.00E-91	180/180(100%)	0/180(0%)

ES 2 767 527 T3

Taxón	Mejor coincidencia de BLAST	Número de acceso de NCBI GenBank	Puntuación de alineación BLAST	Valor esperado	Identidad de secuencia alineada	Brechas
	<i>Mycoplasma hominis</i> cepa PG21 16S gen de ARN ribosómico, secuencia completa	NR_041881.1	333	2.00E-91	180/180(100%)	0/180(0%)
d-c-030	<i>Veillonella montpellierensis</i> cepa ADV 281.99 16S gen de ARN ribosómico, secuencia parcial	NR_028839.1	335	4.00E-92	181/181(100%)	0/181(0%)
d-c-040	<i>Alloscardovia omnicolens</i> cepa CCUG 31649 16S gen de ARN ribosómico, secuencia completa	NR_042583.1	322	3.00E-88	174/174(100%)	0/174(0%)
d-c-047	<i>Anaerococcus senegalensis</i> cepa JC48 16S gen de ARN ribosómico, secuencia parcial	NR_118220.1	333	2.00E-91	180/180(100%)	0/180(0%)
d-c-050	<i>Propionimicrobium lymphophilum</i> cepa JCM 582916S gen de ARN ribosómico, secuencia parcial	NR_114337.1	335	4.00E-92	181/181(100%)	0/181(0%)
d-c-081*	Bacteria no cultivada clon MTS53 16S gen de ARN ribosómico, secuencia parcial	JQ781443.1	250	4.00E-63	165/180(92%)	0/180(0%)
d-c-053	<i>Corynebacterium pyruviciproducens</i> cepa 06-1773O 16S gen de ARN ribosómico, secuencia parcial	NR_116569.1	337	1.00E-92	182/182(100%)	0/182(0%)
d-c-015*	<i>Megasphaera cerevisiae</i> cepa DSM 20462 16S gen de ARN ribosómico, secuencia parcial	NR_113307.1	278	7.00E-75	170/180(94%)	0/180(0%)
	<i>Megasphaera cerevisiae</i> cepa VTT-E-85230 16S gen de ARN ribosómico, secuencia completa	NR_044650.1	278	7.00E-75	170/180(94%)	0/180(0%)
d-c-088	<i>Acidipila rosea</i> cepa AP8 16S gen de ARN ribosómico, secuencia parcial	NR_113179.1	309	3.00E-84	177/182(97%)	0/182(0%)
d-c-072	<i>Murdochiella asaccharolytica</i> cepa WAL 1855C 16S gen de ARN ribosómico, secuencia parcial	NR_116331.1	331	6.00E-91	180/181(99%)	0/181(0%)
d-c-105*	Bacteria no cultivada clon Ovdat63c11 16S gen de ARN ribosómico, secuencia parcial	JF295520.1	237	3.00E-59	165/183(90%)	1/183(0%)

ES 2 767 527 T3

Taxón	Mejor coincidencia de BLAST	Número de acceso de NCBI GenBank	Puntuación de alineación BLAST	Valor esperado	Identidad de secuencia alineada	Brechas
	Bacteria no cultivada clon GASP-WB2W3_D02 16S gen de ARN ribosómico, secuencia parcial	EF074306.1	237	3.00E-59	165/183(90%)	1/183(0%)
	Clon bacteriano OD1 de división candidato no cultivado AKYH1067 16S gen de ARN ribosómico, secuencia parcial	AY922093.1	237	3.00E-59	163/180(91%)	1/180(0%)
d-c-068*	Howardella ureilytica cepa GPC 589 16S gen de ARN ribosómico, secuencia completa	NR_044022.2	268	5.00E-72	169/181(93%)	0/181(0%)
d-c-071	Actinobaculum schaalii cepa CCUG 27420 16S gen de ARN ribosómico, secuencia parcial	NR_116869.1	337	1.00E-92	182/182(100%)	0/182(0%)
	Actinobaculum schaalii cepa CCUG 27420 16S gen de ARN ribosómico, secuencia completa	NR_119244.1	337	1.00E-92	182/182(100%)	0/182(0%)
	Actinobaculum schaalii cepa B 5329 16S gen de ARN ribosómico, secuencia parcial	NR_029339.1	337	1.00E-92	182/182(100%)	0/182(0%)
d-c-083	Peptoniphilus duerdenii cepa WAL 18896 16S gen de ARN ribosómico, secuencia parcial	NR_116346.1	337	1.00E-92	182/182(100%)	0/182(0%)
d-c-087	Fastidiosipila sanguinis cepa CCUG 47711 16S gen de ARN ribosómico, secuencia completa	NR_042186.1	318	4.00E-87	179/182(98%)	1/182(0%)
Notas:						
* representa un taxón con identidad de secuencia <97% como cualquier secuencia conocida y puede representar especies nuevas, no reportadas previamente, que ahora se identifican por secuenciación en este estudio.						

5 Tabla 4H. Clasificación de nivel de especie por alineación BLAST a la base de datos 16S ARNr de NCBI. Se alienaron taxones bacterianos diferencialmente abundantes (significativamente disminuidos) identificados al comparar los 349 taxones en las 10 cérvix dilatadas versus las 10 cerradas (1:1 emparejado por edad gestacional en la recolección de la muestra, emparejado a <2 semanas, normalizado como proporción del recuento total de lecturas para cada muestra.).

ES 2 767 527 T3

Taxón	Mejor coincidencia de BLAST	Número de acceso de NCBI GenBank	Puntuación de alineación BLAST	Valor esperado	Identidad de secuencia alineada	Brechas
d-c-018	Lactobacillus fermentum IFO 3956 cepa IFO 3956 16S ARN ribosómico, secuencia completa	NR_075033.1	335	4E-92	181/181(100%)	0/181(0%)
	Lactobacillus fermentum cepa CIP 102980 16S gen de ARN ribosómico, secuencia completa	NR_104927.1	335	4E-92	181/181(100%)	0/181(0%)
	Lactobacillus fermentum cepa NBRC 15885 16S gen de ARN ribosómico, secuencia parcial	NR_113335.1	335	4E-92	181/181(100%)	0/181(0%)
d-c-052	Aerococcus urinae ACS-120-V-Col10a cepa ACS-120-V-Col10a 16S ARN ribosómico, secuencia completa	NR_074879.1	335	4E-92	181/181(100%)	0/181(0%)
d-c-074	Eubacterium aggregans cepa SR12 16S gen de ARN ribosómico, secuencia parcial	NR_024926.1	176	3E-44	112/120(93%)	2/120(1%)
	Acetobacterium bakii cepa DSM 8239 16S gen de ARN ribosómico, secuencia parcial	NR_026329.1	176	3E-44	112/120(93%)	2/120(1%)
	Eubacterium barkeri cepa ATCC 25849 16S gen de ARN ribosómico, secuencia completa	NR_044661.1	176	3E-44	112/120(93%)	2/120(1%)
d-c-076	Lactobacillus brevis ATCC 367 cepa ATCC 367 16S ARN ribosómico, secuencia completa	NR_075024.1	335	4E-92	181/181(100%)	0/181(0%)
	Lactobacillus senmaizukei cepa NBRC 103853 16S gen de ARN ribosómico, secuencia parcial	NR_114251.1	335	4E-92	181/181(100%)	0/181(0%)
	Lactobacillus brevis cepa ATCC 14869 16S gen de ARN ribosómico, secuencia parcial	NR_116238.1	335	4E-92	181/181(100%)	0/181(0%)
d-c-082	Corynebacterium cepa lactis RW2-5 16S ribosomal RNA genes, secuencia parcial	NR_104996.1	337	1E-92	182/182(100%)	0/182(0%)

ES 2 767 527 T3

Taxón	Mejor coincidencia de BLAST	Número de acceso de NCBI GenBank	Puntuación de alineación BLAST	Valor esperado	Identidad de secuencia alineada	Brechas
	Corynebacterium amycolatum cepa CCUG 35685 16S gen de ARN ribosómico, secuencia parcial	NR_117605.1	337	1E-92	182/182(100%)	0/182(0%)
	Corynebacterium amycolatum 16S, gen de ARN ribosómico, secuencia parcial	NR_117030.1	337	1E-92	182/182(100%)	0/182(0%)
	Corynebacterium freneyi cepa ISPB 6695110 16S gen de ARN ribosómico, secuencia parcial	NR_042025.1	337	1E-92	182/182(100%)	0/182(0%)
	Corynebacterium xerosis cepa DSM 20743 16S gen de ARN ribosómico, secuencia parcial	NR_119180.1	337	1E-92	182/182(100%)	0/182(0%)
	Corynebacterium xerosis cepa ATCC 373 16S gen de ARN ribosómico, secuencia parcial	NR_026213.1	337	1E-92	182/182(100%)	0/182(0%)
	Corynebacterium amycolatum cepa S160 16S gen de ARN ribosómico, secuencia completa	NR_026215.1	337	1E-92	182/182(100%)	0/182(0%)

Notas:

NA representa una no coincidencia BLAST significativa en la base de datos de ARN ribosómico 16S, y puede representar especies nuevas, previamente no reportadas, que ahora se identifican por secuenciación en este estudio.

\* representa un taxón con identidad <97% y puede representar especies nuevas, previamente no reportadas, que ahora se identifican por secuenciación en este estudio.

Tabla 5A. Asociación entre la prueba propuesta, DIBT1, y embarazo adverso o resultados neonatales. La muestra con uno o más taxones seleccionados en el DIBT1 se considera positiva.

Resultado	Verdadero positivo	Falso positivo	Falso negativo	Verdadero negativo	Valor p exacto de Fisher	Sensibilidad	Especificidad	Valor predictivo positivo	Valor predictivo negativo
Parto prematuro espontáneo <28 semanas*	6	7	0	18	2.33E-03	100.0%	72.0%	46.2%	100.0%
Parto prematuro espontáneo <34 semanas*	7	6	0	18	6.53E-04	100.0%	75.0%	53.8%	100.0%
Nacimiento prematuro <28 semanas*	8	5	2	16	5.72E-03	80.0%	76.2%	61.5%	88.9%
Nacimiento prematuro <34 semanas*	13	0	3	15	3.22E-06	81.3%	100.0%	100.0%	83.3%
Puntuación de Apgar a 1 minuto <7	7	1	2	3	2.17E-01	77.8%	75.0%	87.5%	60.0%
Puntuación de Apgar a 5 minuto <7	6	2	1	4	1.03E-01	85.7%	66.7%	75.0%	80.0%
Corioamnionitis (clínica o patológica)	5	3	2	3	5.92E-01	71.4%	50.0%	62.5%	60.0%
Corioamnionitis (clínica)	2	6	1	4	1.00E+00	66.7%	40.0%	25.0%	80.0%
Corioamnionitis (patológica)	1	7	1	4	1.00E+00	50.0%	36.4%	12.5%	80.0%
Síndrome de dificultad respiratoria	5	3	1	4	2.66E-01	83.3%	57.1%	62.5%	80.0%
displasia broncopulmonar	2	4	1	4	1.00E+00	66.7%	50.0%	33.3%	80.0%
Hemorragia intraventricular*	7	0	1	4	1.01E-02	87.5%	100.0%	100.0%	80.0%
Muerte neonatal dentro de los 7 días posteriores al nacimiento	2	6	0	5	4.87E-01	100.0%	45.5%	25.0%	100.0%
Sepsis neonatal	3	4	1	4	5.76E-01	75.0%	50.0%	42.9%	80.0%

\* p < 0.05

Tabla 5B. Asociación entre la prueba propuesta, DIBT2 y embarazo adverso o resultados neonatales. La muestra con uno o más taxones seleccionados en el DIBT2 se considera positiva.

Resultado	Verdadero positivo	Falso positivo	Falso negativo	Verdadero negativo	Valor p exacto de Fisher	Sensibilidad	Especificidad	Valor predictivo positivo	Valor predictivo negativo
Parto prematuro espontáneo <28 semanas*	6	7	0	18	2.3E-03	100.0%	72.0%	46.2%	100.0%
Parto prematuro espontáneo <34 semanas*	7	6	0	18	6.5E-04	100.0%	75.0%	53.8%	100.0%
Parto prematuro <28 semanas*	8	5	2	16	5.7E-03	80.0%	76.2%	61.5%	88.9%
Parto prematuro <34 semanas*	11	2	5	13	3.2E-03	68.8%	86.7%	84.6%	72.2%
Puntuación de Apgar a 1 minuto <7	6	1	3	3	2.7E-01	66.7%	75.0%	85.7%	50.0%
Puntuación de Apgar a 5 minuto <7	5	2	2	4	2.9E-01	71.4%	66.7%	71.4%	66.7%
Corioamnionitis (clínica o patológica)	5	2	2	4	2.9E-01	71.4%	66.7%	71.4%	66.7%
Corioamnionitis (clínica)	2	5	1	5	1.0E+00	66.7%	50.0%	28.6%	83.3%
Corioamnionitis (patológica)	1	6	1	5	1.0E+00	50.0%	45.5%	14.3%	83.3%
Síndrome de dificultad respiratoria	5	2	1	5	1.0E-01	83.3%	71.4%	71.4%	83.3%
Displasia broncopulmonar	2	3	1	5	5.5E-01	66.7%	62.5%	40.0%	83.3%
Hemorragia intraventricular*	6	0	2	4	6.1E-02	75.0%	100.0%	100.0%	66.7%
Muerte neonatal dentro de los 7 días posteriores al nacimiento	2	5	0	6	4.6E-01	100.0%	54.5%	28.6%	100.0%
Sepsis neonatal	3	3	1	5	5.5E-01	75.0%	62.5%	50.0%	83.3%

\* P< 0.05

Tabla 6. Características de los participantes en el estudio B.

	Grupo de cuello uterino corto	Grupo de cuello uterino de longitud normal	Valor P*
	n = 11	n = 11	
Edad gestacional en la recolección de la muestra (semanas)	20.8 +/- 2.37	19.4 +/- 3.07	0.2358
Longitud cervical en la recolección de muestra (mm)	20.2 +/- 9.16	34.5 +/- 5.04	0.0002
Resultado de parto disponible	10	10	
Edad gestacional al momento del parto (semanas)	35.8 +/- 6.00	39.7 +/- 1.34	0.0619
Peso al nacer (g)	2421 +/- 1012	3231 +/- 348	0.0277
Parto por cesárea (n, %)	1 (10%)	6 (60%)	0.0573
Nacimiento prematuro espontáneo <28 semanas (n, %)	2 (20%)	0 (0%)	0.4737
Parto prematuro espontáneo <34 semanas (n, %)	3 (30%)	0 (0%)	0.2105
Nacimiento prematuro <28 semanas (n, %)	2 (20%)	0 (0%)	0.4737
Nacimiento prematuro <34 semanas (n, %)	3 (30%)	0 (0%)	0.2105

\*Prueba T y prueba exacta de Fisher para variables continuas y discontinuas, respectivamente.

5

Tabla 7. Taxones bacterianos diferencialmente abundantes (significativamente aumentados) en el grupo de cuello uterino corto, en comparación con el grupo de cuello uterino de longitud normal. (normalizado por submuestreo aleatorio). Los valores P se ajustan para múltiples pruebas mediante el método de False Discovery Rate

Taxón	Abundancia relativa media (grupo de cuello uterino corto)	Error estándar (grupo de cuello uterino corto)	Abundancia relativa media (grupo de cuello uterino de longitud normal)	Error estándar (grupo de cuello uterino de longitud normal)	Cambio de eventos (aumentado)	Valor p	Valor p ajustado*
s-n-006	4.2E-02	4.1E-02	6.0E-06	6.0E-06	6960	1.5E-02	4.4E-02
s-n-007	9.1E-02	9.1E-02	0	0	6726	1.0E-03	4.5E-03

ES 2 767 527 T3

Taxón	Abundancia relativa media (grupo de cuello uterino corto)	Error estándar de grupo de cuello uterino corto)	Abundancia relativa media (grupo de cuello uterino de longitud normal)	Error estándar de grupo de cuello uterino de longitud normal)	Cambio de eventos (aumentado)	Valor p	Valor p ajustado*
s-n-008	4.1E-02	4.1E-02	0	0	3064	1.0E-03	4.5E-03
s-n-012	6.2E-03	6.2E-03	0	0	462	1.0E-03	4.5E-03
s-n-014	3.9E-03	3.9E-03	0	0	289	1.0E-03	4.5E-03
s-n-022	7.9E-04	7.7E-04	0	0	58	1.0E-03	4.5E-03
s-n-024	3.5E-04	3.5E-04	0	0	26	1.0E-03	4.5E-03
s-n-025	2.0E-04	2.0E-04	0	0	15	1.0E-03	4.5E-03
s-n-027	2.5E-04	2.3E-04	0	0	18	1.0E-03	4.5E-03
s-n-028	3.4E-04	3.4E-04	0	0	25	1.0E-03	4.5E-03
s-n-029	2.0E-04	2.0E-04	0	0	15	1.0E-03	4.5E-03
s-n-030	1.3E-04	1.3E-04	0	0	10	1.0E-03	4.5E-03
s-n-046	6.8E-05	6.8E-05	0	0	5	1.0E-03	4.5E-03
s-n-054	8.0E-05	8.0E-05	0	0	6	1.0E-03	4.5E-03
s-n-063	6.1E-05	6.1E-05	0	0	5	2.0E-03	8.2E-03
s-n-068	5.5E-05	5.5E-05	0	0	4	3.9E-03	1.4E-02
s-n-071	4.3E-05	3.7E-05	0	0	3	1.6E-02	4.4E-02

Tabla 8. Secuencia genómica del gen 16S ARNr en los taxones bacterianos diferencialmente abundantes (significativamente aumentados) en el grupo de cuello uterino corto, en comparación con el grupo de cuello uterino de longitud normal. (normalizado por submuestreo aleatorio).

ES 2 767 527 T3

Taxón	Secuencia de nucleótidos del gen de ARN ribosómico 16S
s-n-006	AGGGCGCAGGCTGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTTAACCGTGGAA GTGCATTGGAACTGGGAAGCTTGAGTACAGAAGAGGAAAGTGGAACCTCCATGTG TAGCGGTGGAATGCGTAGATATATGGAAGAACACCAGTGGCGAAAGCGACTTTCT GGTCTGTCACTGACGCTGA
s-n-007	CGCTCGTAGGCGGTTTCGTGCGGTCCGGTGTGAAAGTCCATCGCTTAACGGTGGAA TCCGCGCCGGGTACGGGCGGGCTTGAGTGCAGTGGGGAGACTGGAATCCCGG GTGTAACGGTGGAAATGTGTAGATATCGGGAAGAACACCAATGGCGAAGGCAGGT CTCTGGGCCGTTACTGACGCTGA
s-n-008	CGCGTGTAGGCGGCTAGATAAGTGTGATGTTTAAATCCAAGGCTTAACCTTGGGG TTCATTACAACTGTTTAGCTTGAGTGTGAGAGGATAGTGGAAATCCTAGTGTA GCGGTAAAATGCGTAGATATTAGGAGGAACACCGGTGGCGAAGGCGGCTATCTG GACAGTAACTGACGCTGA
s-n-012	CGAGCGTAGGCTGTTTGTTAAGCGTGTTGTGAAATGTAGGAGCTCAACTTTTAGAT TGCAGCGCGAACTGGCAGACTTGAGTGCACACAACGTAGGCGGAATTCATGGTG TAGCGGTGAAATGCTTAGATATCATGACGAACTCCGATTGCGAAGGCAGCTTACG GGAGCGCAACTGACGCTAA
s-n-014	GGTACGTAGGCGGTTTTTAAGTCAGGTGTCAAAGCGTGGAGCTTAACTCCATTAA GCACTTGAACTGAAAGACTTGAGTGAAGGAGAGGAAAGTGGAAATCCTAGTGTA GCGGTGAAATGCGTAGATATTAGGAGGAATACCGGTGGCGAAGGCGACTTTCTG GACTTTTACTGACGCTCA
s-n-022	TGTTTCGCAGGCGGCAATGCAAGTCAGATGTAAAAGGCAAAGGCTCAACCTTTGTA AGCATCTGAACTGTATAGCTTGAGAAGTGTAGAGGCAAGTGGAAATTTTGTAGTGA GCGGTGAAATGCGTAGATATTTAAAAGAATACCGGTGGCGAAGGCGACTTGCTGG GCACAATCTGACGCTGA
s-n-024	AGTACGTAGGCGGATAAGCAAGTTAGAAGTGAATCCTATAGCTCAACTATAGTAA GCTTTTAAACTGCTCATCTTGAGGTATGGAAGGGAAAGTGGAAATCCTAGTGTA CGGTGAAATGCGCAGATATTAGGAGGAATACCGGTGGCGAAGGCGACTTTCTGG CCATAAACTGACGCTGA
s-n-025	GGCGCGCAGGCGGTTTTTTAAGTCCGTCTTAAAGTGCAGGGGCTTAACCCCGTG AGGGGACCGAACTGGAAGACTTGAGTGTCCGAGAGGAAAGCGGAATTCCTAGT GTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCGGTGGCGAAGGCGGCTTT CTGGACGACAACCTGACGCTGA
s-n-027	GGCGCGCAGGCGGTCCTTAAGTCCATCTTAGAAGTGCAGGGGCTTAACCCCGTG ATGGGATGGAACTGGGAGACTGGAGTATCCGAGAGGAAAGTGGAAATCCTAGT GTAGCGGTGAAATGCGTAGATATTAGGAAGAACACCGGTGGCGAAGGCGACTTTC TGGACGAAAACCTGACGCTGA

ES 2 767 527 T3

Taxón	Secuencia de nucleótidos del gen de ARN ribosómico 16S
s-n-028	GGTGCGTAGGTGGTCTTTCAAGTCGGTGGTTAAAGGCTACGGCTCAACCGTAGTT AGCCTCCGAAACTGGAAGACTTGAGTGCAGGAGAGGAAAGTGAATTCCCAGTGT AGCGGTGAAATGCGTAGATATTGGGAGGAACACCAGTAGCGAAGGCGGCTTTCT GGA CTGCAACTGACACTGA
s-n-029	GGCGTGTAGGCGGCTAGATAAGTGTGATGTTTAAATCCAAGGCTTAAACCTTGGG GTT CATTACAACTGTTTAGCTTGAGTGTGGAGAGGATAGTGAATTCC TAGTGT AGCGGTAAAATGCGTAGATATTAGGAGGAACACCGGTGGCGAAGGCGGCTATCT GGACAGTAACTGACGCTGA
s-n-030	GGCGCGTAGGCGGTCTGTTAGGTCAGGAGTCAAATCTGGGGGCTCAACCCCTAT CCGCTCCTGATACCGGCAGGCTTGAGTCTGGTATGGGAAGGTGGAATTCCAAGT GTAGCGGTGAAATGCGCAGATATTTGGAAGAACACCAGTGGCGAAGGCGGCCTT CTGGGCCATGACTGACGCTGA
s-n-046	GATGAGATGGCGGTTTCGTCGCGTCCGGTGTGAAAGTCCATCGCTTAACGGTGG TCCGCGCCGGGTACGGGCGGGCTTGAGTCCGGTAGGGGAGACTGGAATTCCCG GTGTAACGGTGAATGTGTAGATATCGGGAAGAACACCAATGGCGAAGGCAGGT CTCTGGGCCGTTACTGACGCTGA
s-n-054	GGCAGGTAGGTGGTCTGATTAGTCAATAGGTGAAATCCTCGGGCTTAACCCGAGA AGTGCCTTTGAAACGGTCAGACTGGAGTACTCTAGAGGGTCGTGGAATTCCCGGT GTAGCGGTGAAATGCGTAGAGATCGGGAGGAACACCAGAGGCGAAGGCGGCGA CCTGGGGAGTACTGACACTCA
s-n-063	CGTGCGTAGGCGGTTCTTTAAGTCAGAGGTGAAAGACGGCAGCTTAACTGTGCG GTGCCTTTGATACTGAAGAACTTGAATTGGGTTGAGGAATGCGGAATGAGACAAG TAGCGGTGAAATGCATAGATATGTCTCAGAACCCCGATTGCGAAGGCAGCATTCC AAGCCTATATTGACGCTGA
s-n-068	GGCAGGTAGGCTGTTTTGTAAGTCCGGCGTGAAATCCCAGAGCTCAACTCTGGAA CTGCGTTGGAACTACATGACTTGAGTATCGGAGAGGTTAGGGGAATTCTCGGTG TAAGGTGAAATCTGTAGATATCGAGAGGAACACCAGTGGCGAAGGCGCCTAACT GGCCGATTACTGACGCTGA
s-n-071	AGAGCGCAGGCGGTTTCTTAAGTCTGATGTGAAAGCCCTCGGCTCAACCGAGGAA GGTCATTGGAACTGGGGA ACTTGAATGCAGAAGAGGAGAGTGAATTCCATGTG TAGCGGTGAAATGCGTAGATATATGGAGGAACACCAGTGGCGAAGGCGACTCTCT GGTCTGTAATTGACGCTGA
s-n-006 (SEQ ID NO: 39); s-n-007 (SEQ ID NO: 40); s-n-008 (SEQ ID NO: 41); s-n-012 (SEQ ID NO: 42); s-n-014 (SEQ ID NO: 43); s-n-022 (SEQ ID NO: 44); s-n-024 (SEQ ID NO: 45); s-n-025 (SEQ ID NO: 46); s-n-027 (SEQ ID NO: 47); s-n-028 (SEQ ID NO: 48); s-n-029 (SEQ ID NO: 49); s-n-030 (SEQ ID NO: 50); s-n-046 (SEQ ID NO: 51); s-n-054 (SEQ ID NO: 52); s-n-063 (SEQ ID NO: 53); s-n-068 (SEQ ID NO: 54); s-n-071 (SEQ ID NO: 55).	

Tabla 9. Clasificación del nivel de especie por alineación BLAST a la base de datos 16S ARNr en NCBI. El gen 16S ARNr en los taxones bacterianos diferencialmente abundantes (significativamente incrementados) en el grupo de

ES 2 767 527 T3

cuello uterino corto, comparado con el grupo de cuello uterino de longitud normal se alinearon. \* indica taxón con <97% de identidad de nucleótidos con secuencias conocidas y puede representar un taxón novedoso, previamente no reportado.

Taxón	Descripción	Adhesión	Puntuación	Esperado	Identidad	Brechas
s-n-006	<i>Aerococcus christensenii</i> cepa CCUG 28831 1 ...	NR_044929.2	326	3E-89	180/183(98%)	0/183(0%)
s-n-007	<i>Bifidobacterium pseudolongum</i> subsp. pseudolo...	NR_043442.2	339	3E-93	183/183(100%)	0/183(0%)
	<i>Bifidobacterium pseudolongum</i> subsp. globosum...	NR_043441.2	339	3E-93	183/183(100%)	0/183(0%)
	<i>Bifidobacterium longum</i> cepa ATCC 15707 16S...	NR_044691.2	339	3E-93	183/183(100%)	0/183(0%)
	<i>Bifidobacterium breve</i> ACS-071-V-Sch8b cepa...	NR_102863.1	339	3E-93	183/183(100%)	0/183(0%)
	<i>Bifidobacterium longum</i> NCC2705 cepa NCC270...	NR_074744.1	339	3E-93	183/183(100%)	0/183(0%)
	<i>Bifidobacterium longum</i> subsp. longum cepa ...	NR_117506.1	339	3E-93	183/183(100%)	0/183(0%)
	<i>Bifidobacterium breve</i> cepa DSM 20213 16S r...	NR_040783.1	339	3E-93	183/183(100%)	0/183(0%)
	<i>Bifidobacterium choerinum</i> cepa Su 806 16S ...	NR_037116.1	339	3E-93	183/183(100%)	0/183(0%)
	<i>Bifidobacterium longum</i> subsp. Infantis strai...	NR_043437.1	339	3E-93	183/183(100%)	0/183(0%)
s-n-008	<i>Saccharofermentans acetigenes</i> cepa P6 16S ...	NR_115340.1	191	1E-48	107/109(98%)	0/109(0%)
s-n-012	<i>Prevotella amnii</i> cepa CCUG 53648 16S riboso...	NR_042587.1	335	4E-92	182/183(99%)	0/183(0%)
s-n-014*	<i>Parvimonas micra</i> cepa JCM 12970 16S riboso...	NR_114338.1	298	6E-81	176/183(96%)	1/183(0%)
	<i>Parvimonas micra</i> cepa 3119B 16S ribosomal ...	NR_036934.1	298	6E-81	176/183(96%)	1/183(0%)
	<i>Parvimonas micra</i> cepa ATCC 33270 16S ribos...	NR_114675.1	298	6E-81	176/183(96%)	1/183(0%)
s-n-022	<i>Peptoniphilus lacrimalis</i> cepa CCUG 31350 1...	NR_041938.1	337	1E-92	182/182(100%)	0/182(0%)

ES 2 767 527 T3

Taxón	Descripción	Adhesión	Puntuación	Esperado	Identidad	Brechas
s-n-024	Anaerococcus prevotii cepa DSM 20548 16S r...	NR_074575.1	337	1E-92	182/182(100%)	0/182(0%)
	Anaerococcus tetradius cepa CCUG 46590 16S...	NR_041941.1	337	1E-92	182/182(100%)	0/182(0%)
	Anaerococcus prevotii cepa CCUG 41932 16S ...	NR_041939.1	337	1E-92	182/182(100%)	0/182(0%)
s-n-025*	Megasphaera elsdenii cepa DSM 20460 16S ri...	NR_102980.1	283	2E-76	174/184(95%)	2/184(1%)
	Megasphaera elsdenii cepa DSM 20460 16S ri...	NR_113306.1	283	2E-76	174/184(95%)	2/184(1%)
	Megasphaera elsdenii cepa DSM 20460 16S ri...	NR_113305.1	283	2E-76	174/184(95%)	2/184(1%)
s-n-027	Dialister micraerophilus cepa JCM 17567 16...	NR_113356.1	337	1E-92	182/182(100%)	0/182(0%)
	Dialister micraerophilus cepa ADV 04.01 16...	NR_041895.1	337	1E-92	182/182(100%)	0/182(0%)
s-n-028	Peptostreptococcus anaerobius cepa NCTC 11...	NR_042847.1	339	3E-93	183/183(100%)	0/183(0%)
	Peptostreptococcus anaerobius cepa ATCC 27...	NR_118652.1	339	3E-93	183/183(100%)	0/183(0%)
s-n-029	Saccharofermentans acetigenes cepa P6 16S ...	NR_115340.1	191	1E-48	107/109(98%)	0/109(0%)
s-n-030	Atopobium vaginae cepa DSM 15829 16S ribos...	NR_117757.1	309	3E-84	177/182(97%)	0/182(0%)
	Atopobium vaginae cepa 961*00022/98 16S ri...	NR_029349.1	309	3E-84	177/182(97%)	0/182(0%)
s-n-046	Bifidobacterium pseudolongum subsp. pseudolo...	NR_043442.2	326	3E-89	176/176(100%)	0/176(0%)
	Bifidobacterium pseudolongum subsp. alobosum...	NR_043441.2	326	3E-89	176/176(100%)	0/176(0%)
	Bifidobacterium longum cepa ATCC 15707 16S...	NR_044691.2	326	3E-89	176/176(100%)	0/176(0%)
	Bifidobacterium breve ACS-071-V-Sch8b cepa...	NR_102863.1	326	3E-89	176/176(100%)	0/176(0%)

ES 2 767 527 T3

Taxón	Descripción	Adhesión	Puntuación	Esperado	Identidad	Brechas
	<i>Bifidobacterium longum</i> NCC2705 cepa NCC270...	NR_074744.1	326	3E-89	176/176(100%)	0/176(0%)
	<i>Bifidobacterium longum</i> subsp. <i>longum</i> cepa ...	NR_117506.1	326	3E-89	176/176(100%)	0/176(0%)
	<i>Bifidobacterium breve</i> cepa DSM 20213 16S r...	NR_040783.1	326	3E-89	176/176(100%)	0/176(0%)
	<i>Bifidobacterium choerinum</i> cepa Su 806 16S ...	NR_037116.1	326	3E-89	176/176(100%)	0/176(0%)
	<i>Bifidobacterium longum</i> subsp. <i>Infantis</i> strai...	NR_043437.1	326	3E-89	176/176(100%)	0/176(0%)
s-n-054*	<i>Spiribacter salinus</i> cepa M19-40 16S riboso...	NR_103952.1	167	2E-41	140/164(85%)	3/164(1%)
s-n-063	<i>Sphingobacterium spiritivorum</i> cepa NBRC 14...	NR_113707.1	327	7E-90	181/183(99%)	0/183(0%)
	<i>Sphingobacterium spiritivorum</i> cepa NCTC 11...	NR_044077.1	327	7E-90	181/183(99%)	0/183(0%)
	<i>Sphingobacterium spiritivorum</i> cepa JCM 127...	NR_115498.1	327	7E-90	181/183(99%)	0/183(0%)
s-n-068*	<i>Ochrobactrum pseudogrignonense</i> cepa CCUG 3...	NR_042589.1	172	4E-43	155/185(84%)	4/185(2%)
s-n-071	<i>Vagococcus fluvialis</i> cepa M-29c 16S riboso...	NR_026489.1	339	3E-93	183/183(100%)	0/183(0%)

Tabla 10. Asociación entre los taxones diferencialmente abundantes (significativamente incrementados) en el Estudio B y los resultados adversos del embarazo o neonatal. La muestra con uno o más taxones seleccionados en el DIBT1 se considera positiva.

Resultados	Verdadero positivo	Falso positivo	Falso negativo	Verdadero negativo	Valor p exacto de Fisher	Sensibilidad	Especificidad	Valor predictivo positivo	Valor predictivo negativo
Longitud cervical <29 mm	9	0	2	11	0.00022	82%	100%	100%	85%
Parto prematuro espontáneo <28 semanas*	2	6	0	12	0.15	100%	67%	25%	100%
Nacimiento prematuro espontáneo <34 semanas*	3	5	0	12	0.049	100%	71%	38%	100%
Nacimiento prematuro <28 semanas*	2	6	0	12	0.15	100%	67%	25%	100%
Nacimiento prematuro <34 semanas*	3	5	0	12	0.049	100%	71%	38%	100%

\* P < 0.05

Tabla 11. Taxones de bacterias que comprenden LA6

Taxón ID	Taxón ID	SEQ ID NO.	Secuencia	Datos de abundancia	Datos de clasificación taxonómica	Clasificación de nivel de especie por alineación BLAST a la base de datos 16S ARNr	Número de acceso de NCBI GenBank	Porcentaje de identidad	de
OTU #16	sn-014	43	Tabla 8	Tabla 7	Tabla 9	<i>Parvimonas micra</i>	NR_114338.1	176/183	96%
OTU #56	d-c-043	7	Tabla 4A	Tabla 2A	Tabla 4E	<i>Ureaplasma urealyticum</i> o <i>Ureaplasma parvum</i>	NR_102836.1 or NR_074176.1	181/181	100%
OTU #42	d-c-045	8	Tabla 4A	Tabla 2A	Tabla 4E	<i>Atopobium vaginae</i>	NR_117757.1	176/181	97%
OTU #28	sn-022	44	Tabla 8	Tabla 7	Tabla 9	<i>Peptoniphilus lacrimalis</i>	NR_041938.1	182/182	100%
OTU #47	d-c-037	5	Tabla 4A	Tabla 2A	Tabla 4E	<i>Megasphaera cerevisiae</i>	NR_113307.1	169/181	93%
OTU #40	d-c-038	9	Tabla 4A	Tabla 2A	Tabla 4E	<i>Parvibacter caecicola</i>	NR_117374.1	167/181	92%

\* Porcentaje de identidad entre la secuencia de nucleótidos del gen 16S ARNr del taxón bacteriano identificado en el presente estudio y la de la coincidencia BLAST más cercana como se representa por el número de acceso de NCBI GenBank.

Listado de secuencia informal

SEQ ID NO: 1

Cebador de avance de PCR V4-V5 (sin secuencia clave de cebador A y secuencia MID en el extremo 5' del cebador)

AYT GGG YDT AAA GNG

5 SEQ ID NO: 2

Cebador reverso de PCR V4-V5 (sin secuencia B del cebador B en el extremo 5' del cebador)

CCG TCA ATT YYT TTR AGT TT

SEQ ID NO: 3

Taxón d-c-019 de la Tabla 4A - secuencia de ARNr 16S.

10 GGAGCGCAGGCGGTTTTTTAAGTCTGATGTGAAAGCCCTCGGCTTAACCGAGGAAGCGCA  
TCGGAAACTGGGAACTTGAGTGCAGAAGAGGACAGTGGAATCCATGTGTAGCGGTGAAA  
TGCGTAGATATATGGAAGAACACCAGTGCGGAAGGCGGCTGTCTGGTCTGTAAGTACTGACGCT

SEQ ID NO: 4

Taxón d-c-030 de la Tabla 4A - secuencia de ARNr 16S.

CGCGCGCAGGCGGACTAGCCAGTCAGTCTTAAAGTTCGGGGCTTAACCCCGTGATGGGA  
TTGAAACTACTAGTCTAGAGTATCGGAGAGGAAAGTGGAAATCCTAGTGTAGCGGTGAAAT  
GCGTAGATATTAGGAAGAACACCAGTGCGGAAGGCGACTTTCTGGACGAACACTGACGCT

SEQ ID NO: 5

15 Taxón d-c-037 (taxón OTU 47) de la Tabla 4A - secuencia de ARNr 16S.

GGCGCGCAGGCGGTTTCGGTAAGTCTGTCTTAAAGTGCGGGCTTAACCCCGTGAGGGACGG  
AAACTGTGCAACTTGAGTGTTCGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCG  
TAGATATTAGGAGGAACACCGGTGGCGAAAGCGGCTTTCTGGACGACAACACTGACGCT

SEQ ID NO: 6

Taxón d-c-040 de la Tabla 4A - secuencia de ARNr 16S.

AGATGAGATGGCGGTTTGTTCGCGTCTGGTGTGAAAGTCCATCGCTTAACGGTGGATTGGC  
GCTGGGTACGGGCAGGCTAGAGTGTAGTAGGGGAGACTGGAATTCCTCGGTGTAACGGTGG  
AATGTGTAGATATCGGGAAGAACACCAATGGCGAAGGCAGGTCTCTGGGCTATTACTGACG  
CT

20 SEQ ID NO: 7

Taxón d-c-043 (taxón OTU 56) de la Tabla 4A - secuencia de ARNr 16S.

CGAGCGCAGGCGGTTTTGTAAGTTTGGTATTAATCTAGATGCTTAACGTCTAGCTGTATCA  
AAAAGTGTAAACCTAGAGTGTAGTAGGGAGTTGGGGAACCTCCATGTGGAGCGGTAAAATGC  
GTAGATATATGGAAGAACACCGGTGGCGAAGGCGCCAACCTTGGACTATCACTGACGCT

SEQ ID NO: 8

Taxón d-c-045 (taxón OTU 42) de la Tabla 4A - secuencia de ARNr 16S.

25 CGCGCGTAGGCGGCTGTAGGTCAGGAGTCAAATCTGGGGGCTCAACCCCTATCCGCTC  
CTGATACCGGCAGGCTTGAGTCTGGTATGGGAAGGTGGAATCCAAGTGTAGCGGTGAAAT  
GCGCAGATATTTGGAAGAACACCAGTGCGGAAGGCGGCCTTCTGGGCCATGACTGACGCT

SEQ ID NO: 9

Taxón d-c-038 (taxón OTU 40) de la Tabla 4A - secuencia de ARNr 16S.

CGCGCGCAGGCGGTTGCTCAAGCGGAACCTCTAATCTCGGGGCTTAACCTCGAGCCGGGT  
TCCGAAGTGGACGACTCGAGTGCAGTAGAGGCAGATGGAATCCCGGTGTAGCGGTGAA  
TGCGCAGATATCGGGAAGAACACCAACGGCGAAGGCAGTCTGCTGGGCCGTCAGTACGCG  
T

SEQ ID NO:10

Taxón d-c-047 de la Tabla 4A - secuencia de ARNr 16S.

5 AGTACGTAGGCGGCCTAGTAAGTTAGAAGTGAAAGAATATAGCTCAACTATATAAAGCTTTT  
AAAAGTGTAGGCTTGAGAGATGAAAGGGAAAGTGAATTCCTAGTGTAGCGGTGAAATGC  
GCAGATATTAGGAAGAATACCGGTGGCGAAGGCGACTTTCTGGTCATCATCTGACGCT

SEQ ID NO:11

Taxón d-c-054 de la Tabla 4A - secuencia de ARNr 16S.

GGTGCGTAGGCGGTCTGTTAAGTTCATGGTTAAATTTGGGGCTCAACCCCATGAGCCAT  
GGATACTGGCAGACTAGAGTATTGGAGAGGCAAGCGGAATTCATGTGTAGCGGTAAAATG  
CGTAGATATATGGAGGAACACCGGTGGCGAAGGCGGCTTGCTAGCCAAAGACTGACGCT

SEQ ID NO:12

10 Taxón d-c-067 de la Tabla 4B - secuencia de ARNr 16S.

GGTGCGCAGGCGGCTTTACAAGTTGGATGTGAAATATTGTGGCTCAACCACAAACGTGCAT  
CCAAAAGTGCAAAGCTTGAGTTAAGGAGAGGTAAAGTGAATTCCTGGTGTAGCGGTGGAAT  
GCGTAGATATCAGGAGGAATACCGGTGGCGAAGGCGACTTACTGGACTTAAACTGACGC

SEQ ID NO:13

Taxón d-c-039 de la Tabla 4B - secuencia de ARNr 16S.

GGCGCGTAGGCGGAATGGCAAGTCAGCAAGTGAAAGCGTGGGGCTCAACCCCATGATGCG  
GCTGAAACTGTTATTCTAGAGGCATGGAGAGGCAAACCGAATTCCTGGTGTAGCGGTGAAA  
TGCGTAGATATCGGGAAGAACACCAGTGGCGAAGGCGGTTTCTGGCCATGAACTGACGC  
T

15 SEQ ID NO:14

Taxón d-c-052 de la Tabla 4B - secuencia de ARNr 16S.

TGGGCGCAGGCGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTTAACCGTGGAAGTGCAT  
TGAAAAGTGGGAACTTGAGTACAGAAGAGGAAAGTGAAGTCCATGTGTAGCGGTGGAAT  
GCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTCACTGACGCT

SEQ ID NO:15

Taxón d-c-074 de la Tabla 4B - secuencia de ARNr 16S.

20 CGTTTCGCAGGCGGCAATGCAAGTCTCGTGTGAAAGGCAAGGGCTCAACCCTTGTAAAGCAC  
AAGAAACTGCATAGCTTGAGTAGTGGAGAGGCAAGTGAATTCCTAGTGTAGCGGTGAAAT  
GCGTAGATATTAGGAGGAATACCGGTGGCGAAGGCGACTTGCTGGACACAAACTGACGCT

SEQ ID NO:16

Taxón d-c-082 de la Tabla 4B - secuencia de ARNr 16S.

GGCTCGTAGGTGGTTTGTGCGCGTCTGTGAAATTCGGGGCTTAACTCCGGGCGTGCA  
GGCGATACGGGCATAACTTGAGTACTGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGA  
AATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTCTCTGGGCAGTAACTGAC  
GCT

SEQ ID NO:17

25 Taxón d-c-092 de la Tabla 4B - secuencia de ARNr 16S.

CGCGTGTAGGCGGATTCTCAAGTTGGATGTGAAACCCCTTGGCTAAACTGAGGGCTTGCAT  
TCAAACTGAGGACCTTGAGTATCAGAGGGGAAAGTGAATTCCTGGTGGAGCGGTAAAT  
GCGTAGAGATCAGGAGGAACACCGGTGGCGAAGGCGACTTTCTGGCTGACAACTGACGCT

SEQ ID NO:18

Taxón d-c-098 de la Tabla 4B - secuencia de ARNr 16S.

TGTGTGTAGGTGGCGTGATTAGTCGTTTTGTGAAAGATCCGAGCTTAACTTGGAAAACGCGA  
ACGAAACGGTCATGCTTGAGTATGTGAGAGGTAAGCAGAACTCATGGTGTAGGGGTGAAAT  
CCGTTGATATCATGGGGAATACCAAAGCGAAGGCAGCTTACTGGCACATTACTGACACT

5 SEQ ID NO:19

Taxón d-c-012 de la Tabla 4C - secuencia de ARNr 16S.

GGTTCGTAGGCTGTTTGTAAAGTCTGGAGTTAAATCCCGGGGCTCAACCCCGGCTCGCTTT  
GGATACTAGCAAAGTACTAGAGTTAGATAGAGGTAAGCGGAATTCATGTGAAGCGGTGAAATG  
CGTAGATATATGGAAGAACACCAAAGGCGAAGGCAGCTTACTGGGTCTATACTGACGCT

SEQ ID NO:20

Taxón d-c-030 de la Tabla 4C - secuencia de ARNr 16S.

CGCGCGCAGGCGGACTAGCCAGTCAGTCTTAAAGTTCCGGGGCTTAACCCCGTGTATGGGA  
TTGAAACTACTAGTCTAGAGTATCGGAGAGGAAAGTGAATTCCTAGTGTAGCGGTGAAAT  
10 GCGTAGATATTAGGAAGAACACCAGTGGCGAAGGCGACTTTCTGGACGAACACTGACGCT

SEQ ID NO:21

Taxón d-c-040 de la Tabla 4C - secuencia de ARNr 16S.

AGATGAGATGGCGGTTTTGTCGCGTCTGGTGTGAAAGTCCATCGCTTAAACGGTGGATTGGC  
GCTGGGTACGGGCAGGCTAGAGTGTAGTAGGGGAGACTGGAATTCCTGGTGTAAACGGTGG  
AATGTGTAGATATCGGGAAGAACACCAATGGCGAAGGCAGGTCTCTGGGCTATTACTGACG  
CT

SEQ ID NO:22

15 Taxón d-c-047 de la Tabla 4C - secuencia de ARNr 16S.

AGTACGTAGGCGGCCTAGTAAGTTAGAAGTGAAGAATATAGCTCAACTATATAAAGCTTTT  
AAAAGTGTAGGCTTGAGAGATGAAAGGGAAAGTGAATTCCTAGTGTAGCGGTGAAATGC  
GCAGATATTAGGAAGAATACCGGTGGCGAAGGCGACTTTCTGGTCATCATCTGACGCT

SEQ ID NO:23

Taxón d-c-050 de la Tabla 4C - secuencia de ARNr 16S.

GGCTTGTAGGCGGTTTTGTCGCGTTCGAAAGTGTAAACTCAGTGCTTAAACGCTGAGCCTGCTT  
TCGATACGGGCTGACTAGAGGAAGGTAGGGGAGAATGGAATTCCTGGTGGAGCGGTGGAA  
TGCGCAGATATCGGGAGGAACACCAGTGGCGAAGGCGGTTCTCTGGACCTTTCCTGACGC

20 SEQ ID NO:24

Taxón d-c-053 de la Tabla 4C - secuencia de ARNr 16S.

CGCTCGTAGGTGGTGTGTTGCGTCTGTGTAATCCAGGGGCTTAACTTTTGGTTGGCAG  
GCGATACGGGCATTGCTTGAGTGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAA  
ATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCAGTTACTGACG

SEQ ID NO:25

Taxón d-c-068 de la Tabla 4C - secuencia de ARNr 16S.

TGAGCGTAGGCGGCAGTACAAGTCGGGAGTAAAACTCGGGGCTCAACCCCGAGACTGCT  
CTCGAAACTGTACAGCTAGAGTGCAGGATGGGCAGGCGGAATTCCTGGTGTAGCGGTGAA  
ATGCGTAGATATCAGGAGGAACACCGGTGGCGAAGGCGGCCTGCTGGACTGTAAGTACG  
CT

SEQ ID NO:26

Taxón d-c-071 de la Tabla 4C - secuencia de ARNr 16S.

5 GGCTTGTAGGCGGCTTGTGCGCCTGCTGTGAAAACGCGGGGCTTAACTTCGCGCGTGCA  
GTGGGTACGGGCAGGCTTGAGTGTGGTAGGGGTGACTGGAATTCAGGTGTAGCGGTGG  
AATGCGCAGATATCTGGAGGAACACCGATGGCGAAGGCAGGTCACTGGGCCATTACTGAC  
GCT

SEQ ID NO:27

Taxón d-c-072 de la Tabla 4C - secuencia de ARNr 16S.

GGTACGTAGGCGGTTTGTAAAGTTTGGCGTTAAATCACGGGGCTCAACCCCGTTCAGCGTT  
GAAACTGGCAAACCTTGAGTAGTAGAGGGGACAGTGAATTCCTAGTGTAGCGGTGAAATG  
CGTAGAGATTAGGAAGAATACCGGTGGCGAAGGCGACTGTCTGGATACATACTGACGCT

SEQ ID NO:28

10 Taxón d-c-081 de la Tabla 4C - secuencia de ARNr 16S.

GGTGTGTAGGCGGTTATATTAGTCATTTGTTAAATCCTCGGGCTTAAACCCGAGAATCGCGA  
GCGAAACGGTATAACTAGAAAAGTGTGAGGGGTGTACAGAACTCATGGTGTAGGGGTGAAAT  
CCGTTGATATCATGGGGAATACCAAAGCGAAGGCAGTACACTGGCACATATTTGACGCT

SEQ ID NO:29

Taxón d-c-015 de la Tabla 4C - secuencia de ARNr 16S.

TGCGCGCAGGCGGTTCCGGTAAGTCTGTCTTAAAAGTGCGGGGCTTAAACCCCGTGAGGGGA  
CGGAAACTGTGCAACTTGAGTGTCCGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAAT  
GCGTAGATATTAGGAGGAACACCGGTGGCGAAGGCGGCTTTCTGGACGACAACTGACGCT

15 SEQ ID NO:30

Taxón d-c-083 de la Tabla 4C - secuencia de ARNr 16S.

GGTTCGCAGGCGGAATAACAAGTCAGATGTGAAAGGCATGGGCTCAACCCATGTAAGCATT  
TGAAACTGTAATTCCTTGAGAAGTGGAGAGGTAAGTGAATTCCTAGTGTAGCGGTGAAATG  
CGTAGATATTAGGAGGAATACCTGTGGCGAAGGCGACTTACTGGACACAAATCTGACGCT

SEQ ID NO:31

Taxón d-c-087 de la Tabla 4C - secuencia de ARNr 16S.

20 CGCGTGTAGGCGGCACTGTAAGTCAGATGTGAAATCTCCCGGCTCAACCGGGAGCGTGCA  
TCTGATACTGCAATGCTTGAGTGATAGAGGGGAAAGCGGAATTCCTAGTGTAGCGGTAAAA  
TGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGCTATTAAGTACGCT

SEQ ID NO:32

Taxón d-c-088 de la tabla 4C - secuencia de ARNr 16S.

GGTGCGTAGGCGGTTCCGGTAAGTCTTGTGTGAAATCTTCAGGCTCAACTTGAAGTCTGCAC  
GAGAAACTGCCGGGCTTGAGTGTGGGAGAGGTGAGTGAATTCCTAGTGTAGCGGTGAAA  
TGCGTAGATATCGGAAGGAACACCTGTGGCGAAGGCGGCTCACTGGACCACAACTGACGC  
T

SEQ ID NO:33

25 Taxón d-c-105 de la Tabla 4C - secuencia de ARNr 16S.

GGTGTGTAGGTGGTTATGTTAGTCTCCTTTCAAAGCTCCCGGCCTAACCGGGAAAAGGGAG  
GGGAAACGGCACAACTAGAGGATGCGAGGGGTCTGTGGAAGTCCATGGAGTAGGGGTGAAA  
TCCGTTGATATCATGGGGAACACCAAAAAGCGAAGGCAGCAGACTGGCGCATTCTGACAC  
T

SEQ ID NO:34

Taxón d-c-018 de la Tabla 4D - secuencia de ARNr 16S.

CGAGTGCAGGCGGTTTTCTAAGTCTGATGTGAAAGCCTTCGGCTTAACCGGAGAAGTGCAT  
CGGAAACTGGATAACTTGAGTGCAGAAGAGGGTAGTGGAAGTCCATGTGTAGCGGTGGAAT  
GCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGGCTACCTGGTCTGCAACTGACGCT

5 SEQ ID NO:35

Taxón d-c-052 de la Tabla 4D - secuencia de ARNr 16S.

TGGGCGCAGGCGGTTTTCTTAAGTCTGATGTGAAAGCCCACGGCTTAACCGTGGAAGTGCAT  
TGGAAGTGGGAACTTGAGTACAGAAGAGGAAAGTGGAACTCCATGTGTAGCGGTGGAAT  
GCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTCACTGACGCT

SEQ ID NO:36

Taxón d-c-074 de la Tabla 4D - secuencia de ARNr 16S.

CGTTTCGCAGGCGGCAATGCAAGTCTCGTGTGAAAGGCAAGGGCTCAACCCTTGTAAGCAC  
AAGAACTGCATAGCTTGAGTAGTGGAGAGGCAAGTGGAAATTCCTAGTGTAGCGGTGAAAT  
10 GCGTAGATATTAGGAGGAATACCGGTGGCGAAGGCGACTTGCTGGACACAACTGACGCT

SEQ ID NO:37

Taxón d-c-076 de la Tabla 4D - secuencia de ARNr 16S.

TGAGCGCAGGCGGTTTTTTAAGTCTGATGTGAAAGCCTTCGGCTTAACCGGAGAAGTGCAT  
CGGAAACTGGGAGACTTGAGTGCAGAAGAGGACAGTGGAACTCCATGTGTAGCGGTGGAA  
TGCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGGCTGTCTAGTCTGTAAGTACTGACGCT

SEQ ID NO:38

15 Taxón d-c-082 de la Tabla 4D - secuencia de ARNr 16S.

GGCTCGTAGGTGGTTTTGTCGCGTCTGTGAAATTCGGGGCTTAACCTCCGGGCGTGCA  
GGCGATACGGGCATAACTTGAGTACTGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGA  
AATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCGGGTCTCTGGGCAGTAACTGAC  
GCT

SEQ ID NO:39

Taxón s-n-006 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

AGGGCGCAGGCTGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTTAACCGTGGA  
GTGCATTGGAAGTGGGAAGCTTGAGTACAGAAGAGGAAAGTGGAACTCCATGTG  
TAGCGGTGGAATGCGTAGATATATGGAAGAACACCAGTGGCGAAAGCGACTTTCT  
GGTCTGTCACTGACGCTGA

20 SEQ ID NO:40

Taxón s-n-007 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

CGCTCGTAGGCGGTTTCGTCGCGTCCGGTGTGAAAGTCCATCGCTTAACGGTGGA  
TCCGCGCCGGTACGGGCGGGCTTGAGTGCGGTAGGGGAGACTGGAATTCGG  
GTGTAACGGTGGAATGTGTAGATATCGGGAAGAACACCAATGGCGAAGGCAGGT  
CTCTGGGCCGTTACTGACGCTGA

SEQ ID NO:41

Taxón s-n-008 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

CGCGTGTAGGCGGCTAGATAAGTGTGATGTTTAAATCCAAGGCTTAACCTTGGGG  
TTCATTACAACTGTTTAGCTTGAGTGCTGGAGAGGATAGTGGAATTCCTAGTGTA  
GCGGTAAAATGCGTAGATATTAGGAGGAACACCGGTGGCGAAGGCGGCTATCTG  
GACAGTAACTGACGCTGA

SEQ ID NO:42

Taxón sn-012 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

CGAGCGTAGGCTGTTTGTAAAGCGTGTTGTGAAATGTAGGAGCTCAACTTTTAGAT  
TGCAGCGCGAACTGGCAGACTTGAGTGCGCACACGTTAGGCGGAATTCATGGTG  
TAGCGGTGAAATGCTTAGATATCATGACGAACTCCGATTGCGAAGGCAGCTTACG  
5 GGAGCGCAACTGACGCTAA

SEQ ID NO:43

Taxón s-n-014 (taxón OTU 16) de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

GGTACGTAGGCGGTTTTTAAGTCAGGTGTCAAAGCGTGGAGCTTAACTCCATTAA  
GCACTTGAACTGAAAGACTTGAGTGAAGGAGAGGAAAGTGGAATTCCTAGTGTA  
GCGGTGAAATGCGTAGATATTAGGAGGAATACCGGTGGCGAAGGCGACTTTCTG  
GACTTTTACTGACGCTCA

SEQ ID NO:44

10 Taxón s-n-022 (taxón OTU 28) de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

TGTTTCGAGGCGGCAATGCAAGTCAGATGTAAAAGGCAAAGGCTCAACCTTTGTA  
AGCATCTGAACTGTATAGCTTGAGAAGTGTAGAGGCAAGTGGAATTTTTAGTGTA  
GCGGTGAAATGCGTAGATATTA AAAAGAATACCGGTGGCGAAGGCGACTTGCTGG  
GCACAATCTGACGCTGA

SEQ ID NO:45

Taxón s-n-024 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

AGTACGTAGGCGGATAAGCAAGTTAGAAGTGAAATCCTATAGCTCAACTATAGTAA  
GCTTTTAAACTGCTCATCTTGAGGTATGGAAGGGAAAGTGGAATTCCTAGTGTA  
CGGTGAAATGCGCAGATATTAGGAGGAATACCGGTGGCGAAGGCGACTTTCTGG  
CCATAAACTGACGCTGA

15 SEQ ID NO:46

Taxón s-n-025 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

GGCGCGCAGGCGGTTTTTTAAGTCGGTCTTAAAAGTGCGGGGCTTAACCCCGTG  
AGGGGACCGAACTGGAAGACTTGAGTGTCGGAGAGGAAAGCGGAATTCCTAGT  
GTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCGGTGGCGAAGGCGGCTTT  
CTGGACGACAACTGACGCTGA

SEQ ID NO:47

Taxón s-n-027 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

GGCGCGCAGGCGGTCACCTTAAAGTCCATCTTAGAAGTGCGGGGCTTAACCCCGTG  
ATGGGATGGAACTGGGAGACTGGAGTATCGGAGAGGAAAGTGGAATTCCTAGT  
GTAGCGGTGAAATGCGTAGATATTAGGAAGAACACCGGTGGCGAAGGCGACTTTC  
20 TGGACGAAACTGACGCTGA

SEQ ID NO:48

Taxón s-n-028 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

GGTGCGTAGGTGGTCTTTCAAGTCGGTGGTTAAAGGCTACGGCTCAACCGTAGTT  
AGCCTCCGAAACTGGAAGACTTGAGTGCAGGAGAGGAAAGTGGAATTCAGTGT  
AGCGGTGAAATGCGTAGATATTGGGAGGAACACCAGTAGCGAAGGCGGCTTTCT  
GGACTGCAACTGACACTGA

SEQ ID NO:49

Taxón s-n-029 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

5 GGCGTGTAGGCGGCTAGATAAGTGTGATGTTTAAATCCAAGGCTTAAACCTTGGG  
GTTCAATACAACTGTTTAGCTTGAGTGTGGAGAGGATAGTGGAATTCCTAGTGT  
AGCGGTAAAATGCGTAGATATTAGGAGGAACACCGGTGGCGAAGGCGGCTATCT  
GGACAGTAACTGACGCTGA

SEQ ID NO:50

Taxón s-n-030 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

10 GGCGCGTAGGCGGTCTGTTAGGTCAGGAGTCAAATCTGGGGGCTCAACCCCTAT  
CCGCTCCTGATACCGGCAGGCTTGAGTCTGGTATGGGAAGGTGGAATTCAGT  
GTAGCGGTGAAATGCGCAGATATTTGGAAGAACACCAGTGGCGAAGGCGGCCTT  
CTGGGCCATGACTGACGCTGA

SEQ ID NO:51

Taxón s-n-046 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

GATGAGATGGCGGTTCTGTCGCGTCCGGTGTGAAAGTCCATCGCTTAAACGGTGGGA  
TCCGCGCCGGGTACGGGCGGGCTTGAGTGCAGGAGGACTGGAATTCGG  
GTGTAACGGTGGAAATGTGTAGATATCGGGAAGAACACCAATGGCGAAGGCAGGT  
CTCTGGGCCGTTACTGACGCTGA

SEQ ID NO:52

Taxón s-n-054 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

15 GGCAGGTAGGTGGTCTGATTAGTCAATAGGTGAAATCCTCGGGCTTAACCCGAGA  
AGTGCCTTTGAAACGGTCAGACTGGAGTACTCTAGAGGGTCGTGGAATTCGGT  
GTAGCGGTGAAATGCGTAGAGATCGGGAGGAACACCAGAGGCGAAGGCGGCGA  
CCTGGGGAGTACTGACACTCA

SEQ ID NO:53

Taxón s-n-063 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

CGTGCGTAGGCGGTTCTTTAAGTCAGAGGTGAAAGACGGCAGCTTAACTGTGCGCA  
GTGCCTTTGATACTGAAGAACTTGAATTGGGTTGAGGAATGCGGAATGAGACAAG  
TAGCGGTGAAATGCATAGATATGTCTCAGAACCCCGATTGCGAAGGCAGCATTCC  
AAGCCTATATTGACGCTGA

SEQ ID NO:54

20 Taxón s-n-068 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

GGCAGGTAGGCTGTTTTGTAAGTCCGGCGTGAAATCCCAGAGCTCAAACCTCTGGAA  
CTGCGTTGGAACTACATGACTTGAGTATCGGAGAGGTTAGGGGAATTCCTCGGTG  
TAAGGGTGAATCTGTAGATATCGAGAGGAACACCAGTGGCGAAGGCGCCTAACT  
GGCCGATTACTGACGCTGA

SEQ ID NO:55

ES 2 767 527 T3

Taxón s-n-071 de la Tabla 8 - secuencia de nucleótidos del gen 16S ARNr.

AGAGCGCAGGCGGTTTCTTAAGTCTGATGTGAAAGCCCTCGGCTCAACCGAGGAA  
GGTCATTGGAAACTGGGGAAGTTGAATGCAGAAGAGGAGAGTGGAATTCCATGTG  
TAGCGGTGAAATGCGTAGATATATGGAGGAACACCAGTGGCGAAGGCGACTCTCT  
GGTCTGTAATTGACGCTGA

SEQ ID NO:56

OTU #11 taxón - *Sneathia sanguinegens* - secuencia de nucleótidos del gen 16S ARNr.

CGCATCTAGGCGGTAAGACAAGTTGAAGGTGAAAACCTGTGGCTCAACCATAGGCT  
TGCCTACAAAAGTGTGAACTAGAGTACTGGAAAGGTGGGTGGAAGTACACGAGTA  
GAGGTGAAATTCGTA

5

**REIVINDICACIONES**

1. Un método para determinar el riesgo de un embarazo o un resultado neonatal adverso para un sujeto femenino, comprendiendo dicho método:
  - 5 (a) detectar en una muestra biológica tomada del sujeto femenino el nivel de bacterias de *Parvimonas micra*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Peptoniphilus lacrimalis*, *Megasphaera cerevisiae* y *Parvibacter caecicola*; y
  - b) determinar que el sujeto femenino tiene un mayor riesgo de un embarazo o un resultado neonatal adverso si el nivel de cada una de las bacterias es mayor que un nivel de control estándar.
- 10 2. El método de la reivindicación 1, en donde la muestra biológica es un hisopo cervical, un hisopo vaginal, una muestra de orina, una muestra de líquido amniótico, una muestra de sangre materna, una muestra de suero materno, una muestra de plasma materno, una muestra de moco cervical, un hisopo placentario, un hisopo del cordón umbilical o cualquier muestra tomada directa o indirectamente del sistema reproductivo o del sistema gastrointestinal.
- 15 3. El método de la reivindicación 1, en donde la etapa de detección comprende un ensayo de amplificación de polinucleótidos, un ensayo que implica la determinación de la secuencia de polinucleótidos, o un ensayo que implica hibridación de sonda/cebador específica de secuencia, en donde el ensayo de amplificación es preferiblemente un ensayo de reacción en cadena de la polimerasa (PCR) y en donde el ensayo de PCR es más preferiblemente un ensayo de PCR cuantitativo o un ensayo de PCR de transcriptasa reversa.
- 20 4. El método de la reivindicación 1, en donde el embarazo o el resultado neonatal adverso comprende parto prematuro a <34 semanas, parto prematuro a <37 semanas, parto dentro de aproximadamente 1-196 días después de la toma de la muestra biológica, parto dentro de aproximadamente 1-196 días después de que se realiza una intervención clínica, una puntuación de Apgar a 1 minuto de <7, una puntuación de Apgar a 5 minutos de <7, corioamnionitis, síndrome de dificultad respiratoria, displasia broncopulmonar, hemorragia intraventricular, muerte neonatal dentro de los 7 días posteriores al nacimiento o sepsis neonatal.
- 25 5. El método de la reivindicación 1, que comprende además determinar que el sujeto femenino tiene un riesgo de dilatación cervical avanzada o acortamiento cervical prematuro si el nivel de cada una de las bacterias es mayor que un nivel de control estándar.
6. El método de la reivindicación 1, que comprende además determinar que el sujeto femenino está en riesgo de tener una infección en la cavidad amniótica, la cavidad uterina, el cuello uterino o la vagina.
- 30 7. El método de la reivindicación 1, que comprende además extraer ácidos nucleicos de la muestra biológica antes del paso (a).
8. El método de la reivindicación 1, cuando se predice que el sujeto femenino tiene un embarazo o un resultado neonatal adverso, que comprende además repetir los pasos (a) y (b) en un momento posterior usando otra muestra biológica del tipo de muestra del sujeto femenino, en donde un aumento en el nivel de cada una de las bacterias en el momento posterior en comparación con el nivel determinado en el paso original (a) indica un mayor riesgo de tener un
  - 35 embarazo o un resultado neonatal adverso.
9. El método de la reivindicación 1, que comprende además:
  - detectar en la muestra biológica el nivel de bacterias que pertenecen a al menos un taxón bacteriano seleccionado del grupo que consiste en *Jonquetella anthropi*, *Aerococcusurinae*, un taxón bacteriano especificado por una secuencia de nucleótidos de ARNr 16S en las Tablas 4B y 4D; y
  - 40 un taxón bacteriano especificado por la clasificación de especies más cercana basada en la alineación de nucleótidos BLAST en las Tablas 4F y 4H y
  - determinar que el sujeto femenino tiene un mayor riesgo de un embarazo o un resultado neonatal adverso si el nivel de bacterias que pertenecen al al menos uno de los taxones bacterianos adicionales es inferior al nivel de control estándar.
- 45 10. Un kit para determinar el riesgo de tener un embarazo o un resultado neonatal adverso en un sujeto femenino, que comprende
  - (a) un control estándar para determinar el nivel de cada una de las bacterias que proporciona una muestra biológica tomada de un sujeto femenino embarazado que contiene bacterias de *Parvimonas micra*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Peptoniphilus lacrimalis*, *Megasphaera cerevisiae* y *Parvibacter caecicola*; y
  - 50 (b) uno o más agentes que identifican específica y cuantitativamente cada una de las bacterias de *Parvimonas micra*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Peptoniphilus lacrimalis*, *Megasphaera cerevisiae* y *Parvibacter caecicola*.

11. El kit de la reivindicación 10, en donde el agente es uno o más cebadores oligonucleotídicos que hibrida específicamente a y amplifica un polinucleótido de la bacteria en un ensayo de amplificación, o una sonda polinucleotídica que hibrida específicamente con una secuencia polinucleotídica de la bacteria.
- 5 12. Un método para determinar si una mujer embarazada tiene un mayor riesgo de tener dilatación cervical avanzada o acortamiento cervical prematuro, comprendiendo dicho método:
- (a) detectar en una muestra biológica tomada del sujeto femenino el nivel de bacterias de *Parvimonas micra*, *Ureaplasma urealyticum* o *Ureaplasma parvum*, *Atopobium vaginae*, *Peptoniphilus lacrimalis*, *Megasphaera cerevisiae* y *Parvibacter caecicola*; y
- 10 (b) determinar que el sujeto femenino tiene un mayor riesgo de tener dilatación cervical avanzada o acortamiento cervical prematuro si el nivel de cada una de las bacterias es mayor que el de un nivel de control estándar.
13. El método de la reivindicación 12, que comprende además extraer ácidos nucleicos de la muestra biológica antes de realizar la etapa (a).
- 15 14. El método de la reivindicación 12, en donde la muestra biológica es un hisopo cervical, un hisopo vaginal, una muestra de orina, una muestra de líquido amniótico, una muestra de sangre materna, una muestra de suero materno, una muestra de plasma materno, una muestra de moco cervical, un hisopo placentario, un hisopo del cordón umbilical o cualquier muestra tomada directa o indirectamente del sistema reproductivo o del sistema gastrointestinal.
- 20 15. El método de la reivindicación 12, en donde la etapa de detección comprende un ensayo de amplificación de polinucleótidos, un ensayo que implica la determinación de la secuencia de polinucleótidos, o un ensayo que implica la hibridación sonda/cebador específica de la secuencia, en donde el ensayo de amplificación es preferiblemente un ensayo de reacción en cadena de la polimerasa (PCR) y en donde el ensayo de PCR es más preferiblemente un ensayo de PCR cuantitativo o un ensayo de PCR de transcriptasa reversa.

FIG. 1A

Resultados después del tratamiento		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	P25
Parto prematuro espontáneo (sPTB) o parto a término (TB)		sPTB	TB																							
GA en el momento del parto (semanas)		18	20	21	21	24	24	25	26	31	33	38	38	38	38	38	38	38	39	40	40	40	40	40	40	41
Latencia (días)		22	10	7	2	24	2	34	5	91	129	121	126	108	103	149	156	171	119	112	109	126	134	138	104	129
Muerte perinatal		SI	SI	SI	SI	SI	NO	NO	NO	SI	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
RDS/BPD		+	+	+	NY	NY	NY	+	+	NY	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ROP/IVH		+	+	+	NY	NY	NY	+	+	NY	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

FIG. 1B

Las 10 bacterias más abundantes en los cuellos uterinos CI "sPTB después del tratamiento".

Taxón	Unidad taxonómica operacional	sPTB	TB	valor de p (% FDR < 5%)																							
<i>Gardnerella</i>	Clu 4	2.3	3.5	4.1	4.1	0	5.7	2.4	4.0	4.1	5.7	2.7	3.1	2.5	1.7	2.5	2.0	2.5	0	0	2.0	0.8	2.7	2.0	0	0.022	
<i>L. ivorii</i>	Clu 2	2.6	2.7	5.9	2.9	0	3.7	2.4	4.3	5.8	3.0	0	5.7	2.3	0	2.4	6.3	2.6	6.3	0	0	0.3	2.5	2.0	6.5	2.3	0.463
<i>L. crispatus</i>	Clu 1	0.2	1.9	2.7	2.9	6.3	0	1.7	1.7	3.1	2.5	5.3	1.7	5.7	5.3	6.3	0	5.2	3.2	5.8	6.2	2.3	0	6.2	2.9	5.8	0.183
<i>Sneathia</i>	Clu 11	2.8	5.5	0	3.8	0	1.4	1.7	4.8	4.7	0	0	0	2.1	0	0	0	0	0	0	0	0	0	0	2.6	1.7	0.010 *
<i>Sneathia</i>	Clu 10	0	3.4	0	6.9	0	1.4	2.2	5.0	3.8	0	1.8	0	0	1.8	0	0	2.0	0	0	2.4	0	2.0	0	0	0	0.077
<i>Aerococcus</i>	Clu 8	0	0	0	2.7	0	3.4	2.0	3.3	3.9	6.7	0	3.2	0	0	0	0	0	0	0	0	1.5	0	0	3.1	0	0.024
<i>Megasphaera</i>	Clu 15	0	2.3	0	4.3	0	3.7	0	4.5	3.6	0	0	0	0	0	0	0	0	0	0	2.0	0	1.7	0	0	0.023	
<i>Parvimonas</i>	Clu 26	0	2.3	2.2	0	2.1	2.5	2.5	2.3	2.9	1.6	1.8	1.7	1.8	0	0	4.1	0	2.5	1.9	2.0	2.7	2.2	2.4	0	2.5	0.576
<i>Aerococcus</i>	Clu 18	0	3.0	0	4.5	1.8	2.4	0	3.3	3.1	0	1.8	0	2.1	0	0	0	0	1.9	0	0	2.0	0	0	0	2.0	0.067
<i>L. gasseri</i>	Clu 3	0	0	3.0	1.3	0	1.4	6.1	3.2	1.8	0	2.1	1.7	1.8	2.2	2.3	3.3	0	0	0	0	2.5	0	0	2.8	0	0.794

FIG. 1C

Las 10 bacterias más abundantes en los cuellos uterinos CI "TB después del tratamiento".

Taxón	Unidad taxonómica operacional	sPTB	TB	valor de p (% FDR < 5%)																							
<i>L. crispatus</i>	Clu 1	0.2	1.9	2.7	2.9	6.3	0	1.7	1.7	3.1	2.5	5.0	1.2	5.7	5.3	6.3	0	5.2	3.2	5.8	6.2	2.3	0	6.2	2.9	5.9	0.183
<i>L. ivorii</i>	Clu 2	2.6	2.7	5.9	2.9	0	3.7	2.4	4.3	5.8	3.0	0	5.7	2.3	0	2.4	6.3	2.6	6.3	0	0	0.3	2.5	2.0	6.5	2.3	0.463
<i>L. gasseri</i>	Clu 5	2.3	0	4.9	0	0	0	0	0	0	0	4.8	0	3.3	0	2.2	0	4.1	5.0	3.3	5.4	2.6	0	0	0	0.183	
<i>Gardnerella</i>	Clu 4	2.3	3.5	4.1	4.1	0	5.7	2.4	4.8	4.1	5.7	2.7	3.1	2.5	1.7	2.5	2.0	2.5	0	0	2.0	0.8	2.7	2.0	0	0.022	
<i>Parvimonas</i>	Clu 26	0	2.3	2.2	0	2.1	2.5	2.5	2.3	2.6	1.6	1.8	1.7	1.8	0	0	4.1	0	2.5	1.9	2.0	2.7	2.2	2.4	0	2.5	0.576
<i>L. ivorii</i>	Clu 19	0	0	2.7	0	0	0	0	0	0	0	0	0	3.5	0	1.9	0	3.5	0	3.2	2.0	4.0	0	0	0	3.9	0.052
<i>L. gasseri</i>	Clu 3	0	0	3.0	1.3	0	1.4	6.1	3.2	1.8	0	2.1	1.7	1.8	2.2	2.3	3.3	0	0	0	0	2.5	0	0	2.8	0	0.794
<i>Corynebacterium</i>	Clu 44	2.0	0	0	0	0	0	0	0	0	0	0	0	2.1	0	0	0	0	2.3	2.1	2.3	0	2.0	0	0	1.7	0.100
<i>Sneathia</i>	Clu 10	0	3.4	0	6.9	0	1.4	2.2	5.0	3.8	0	1.8	0	0	1.8	0	0	2.0	0	0	2.4	0	2.0	0	0	0.077	
<i>Aerococcus</i>	Clu 18	0	3.0	0	4.5	1.8	2.4	0	3.3	3.1	0	1.8	0	2.1	0	0	0	0	1.9	0	0	2.0	0	0	0	2.0	0.067

FIG. 1D

Bacterias diferencialmente abundantes en los cuellos uterinos CI "sPTB después del tratamiento" y "TB después del tratamiento".

Taxón	Unidad taxonómica operacional	sPTB	TB	TB	TB	TB	TB	TB	TB	TB	TB	TB	TB	TB	TB	TB	valor de p (% FDR < 5%)									
<i>Sneathia</i>	Clu 11	2.0	5.5	0	3.8	0	1.4	1.7	4.8	4.7	0	0	0	2.1	0	0	0	0	0	0	0	0	0	2.6	1.7	0.010 *
<i>Parvimonas</i>	Clu 16	0	0	0	4.5	0	3.7	2.0	4.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.009 *
<i>Ureaplasma</i>	Clu 56	2.3	0	2.4	0	2.8	0	0	0	3.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.009 *
<i>Atopobium</i>	Clu 42	0	0	0	2.7	0	1.9	0	2.9	2.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.008 *
<i>Peptoniphilus</i>	Clu 28	0	0	0	0	0	2.1	2.0	3.7	1.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.009 *
<i>Megasphaera</i>	Clu 47	0	0	0	2.5	0	2.8	0	1.4	2.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.009 *
<i>Farnegegerhella</i>	Clu 40	0	1.4	0	2.0	0	3.8	0	2.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.009 *

FIG. 1E

Recuentos de lectura normalizados:	Log(abundancia):
> 10,000	> 4.0
> 1,000 a 10,000	> 3.0 a 4.0
> 100 a 1,000	> 2.0 a 3.0
> 10 a 100	> 1.0 a 2.0
> 1 a 10	> 0.0 a 1.0
1	0

FIG. 2A

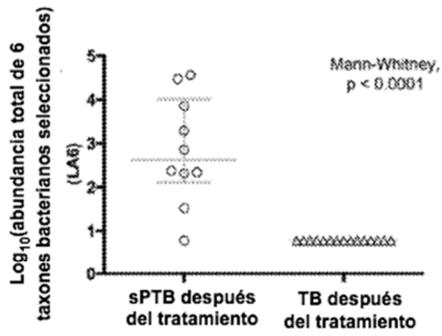


FIG. 2B

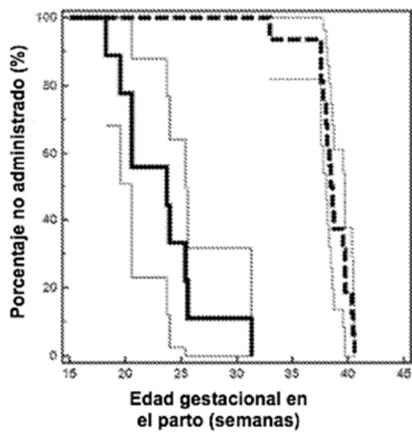


FIG. 2C

