Tesis Doctoral

María Concepción Magno Pérez-Bryan

Junio 2016

2016

MARÍA CONCEPCIÓN PÉREZ-BRYAN

TESIS DOCTORAL

UNIVERSIDAD DE MÁLAGA
Facultad de Ciencias

Tesis Doctoral

Análisis genómico y funcional de las singularidades de dos cepas de *Bacillus amyloliquefaciens* con capacidad de biocontrol

María Concepción Magno Pérez-Bryan

Málaga 2016

Directores
Diego F. Romero Hinojosa
Alejandro Pérez García

Programa de doctorado
Fundamentos celulares y moleculares de los seres vivos

UNIVERSIDAD
DE MÁLAGA

Departamento de Microbiología

Facultad de Ciencias

Análisis genómico y funcional de las singularidades de dos cepas de *Bacillus amyloliquefaciens* con capacidad de biocontrol

TESIS DOCTORAL

María Concepción Magno Pérez-Bryan

Málaga, 2016

AUTOR: María Concepción Magno Pérez-Bryan

http://orcid.org/0000-0002-4559-5269

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

UNIVERSIDAD
DE MÁLAGA

Departamento de Microbiología

Facultad de Ciencias

Análisis genómico y funcional de las singularidades de dos cepas de *Bacillus amyloliquefaciens* con capacidad de biocontrol

Memoria presentada por **Dª María Concepción Magno Pérez-Bryan** para optar al grado de Doctor por la Universidad de Málaga

UNIVERSIDAD
DE MÁLAGA

Departamento de Microbiología

Facultad de Ciencias

D. JUAN JOSÉ BORREGO GARCÍA, Director del Departamento de Microbiología de la Universidad de Málaga.

INFORMA:

Que **Dª MARÍA CONCEPCIÓN MAGNO PÉREZ-BRYAN** ha realizado en los laboratorios de este departamento el trabajo experimental conducente a la elaboración de la presente Memoria de Tesis Doctoral.

Y para que así conste, y tenga los efectos que correspondan, en cumplimiento de la legislación vigente, expido el presente informe,

En Málaga, 28 de Abril de 2016.

DPTO. DE MICROBIOLOGÍA - UNIVERSIDAD DE MÁLAGA

Fdo. D. Juan José Borrego García

UNIVERSIDAD
DE MÁLAGA

Departamento de Microbiología

Facultad de Ciencias

D. ALEJANDRO PÉREZ GARCÍA, Profesor Titular del Departamento de Microbiología de la Universidad de Málaga.

D. DIEGO FRANCISCO ROMERO HINOJOSA, Profesor Contratado Doctor del Departamento de Microbiología de la Universidad de Málaga.

INFORMAN:

Que, **Dª MARÍA CONCEPCIÓN MAGNO PÉREZ-BRYAN** ha realizado bajo nuestra dirección el trabajo experimental conducente a la elaboración de la presente Memoria de Tesis Doctoral.

Y para que así conste, y tenga los efectos que correspondan, en cumplimiento de la legislación vigente, expedimos el presente informe,

En Málaga, 28 de Abril de 2016.

Fdo. D. Alejandro Pérez García

Fdo. D. Diego Francisco Romero Hinojosa

Este trabajo ha sido subvencionado por:

- Plan Nacional de I+D+I del Ministerio de Ciencia e Innovación AGL2010-21848-CO2-01 y AGL-2012-31968 y Proyecto de Excelencia de la Junta de Andalucía P10-AGR-5797, ambos cofinanciados con fondos europeos FEDER (UE).
- Fondos de la empresa holandesa Koppert Biological Systems (8.06/60.4086).
- La estancia realizada en el Centro de Biotecnología y Genómica de Plantas (Universidad Politécnica de Madrid-Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria) de Madrid, ha sido financiada por el Plan Propio de la Universidad de Málaga.

AGRADECIMIENTOS

Y por fin llegó el tan ansiado día de la defensa de mi tesis doctoral…Parece mentira lo rápido que puede pasar el tiempo y es que ya hace unos cinco años que comencé este trabajo de investigación. El camino no ha sido fácil, pero la recompensa sin duda lo merece. Lo que está claro es que a mi lado ha habido muchas personas que me han ayudado, en mayor o menor medida, a afrontar todos los obstáculos que se han ido presentando durante este período de tiempo. Por ello, me gustaría dedicar este espacio y daros las gracias por confiar en mí y haberme apoyado en esta etapa, en muchas ocasiones, difícil de afrontar.

En primer lugar, agradecer a los principales "culpables" de que haya podido realizar la tesis doctoral en el departamento. A Francisco Cazorla, por aceptar que formase parte del grupo al elegirme para hacer el trabajo de Máster con vosotros y a Cayo Ramos por conseguir la beca de la que me he beneficiado durante estos años y, por supuesto, a la Junta de Andalucía y Universidad de Málaga, entidades que financian este trabajo.

En segundo lugar, agradecer a todos los que me habéis enseñado y aportado vuestros conocimientos científicos, así como me habéis demostrado que puedo contar con vosotros, y que por causas laborales os encontrais dispersos por otros países: Houda, Laura y David, me ayudasteis mucho, especialmente, en mis primeras charlas, tanto en los seminarios como en las famosas jornadas del Máster; Claudia, porque me acogiste en seguida y me enseñaste a manejarme con los ensayos de plantas en La Mayora; Guti, por tus bromas y por hacerme sentir una más del grupo; Nuria y Mariki, por vuestra amistad y esas tardes de charlas y desconexión que hacían el trabajo más llevadero; Victor, por tus continuos consejos como compi de trabajo, tu ayuda sin límites, sobretodo cuando no era capaz de manejarme sóla y por tu amistad, que espero conservar durante mucho tiempo.

Por otro lado, me gustaría agradecer a Eva Arrebola su paciencia y comprensión conmigo en mis comienzos; me enseñaste como manejarme en un laboratorio, a gestionar los tiempos, a realizar distintas técnicas experimentales, esa no es tarea fácil. Gracias por tu amistad y cariño. Me alegro mucho de haber compartido contigo el tiempo que estuve por la Mayora. Gracias también a Juan Antonio Torés por dejarme hacer uso de aquellas instalaciones y de sus aportaciones en los seminarios de grupo.

Además, también he podido disfrutar de una estancia en el Centro de Biotecnología y Genómica de Plantas de Madrid bajo la supervisión de Pablo Rodríguez Palenzuela. El junto con Pedro Martínez han sido una pieza clave para dar forma a uno de mis principales bloques de trabajo: el análisis de los genomas de *Bacillus amyloliquefaciens* CECT 8237 y CECT 8238. La verdad es que guardo muy buenos recuerdos de todas las personas con las que compartí aquella etapa tanto con Emilia López, Chechu y Pablo Rodríguez como con mis compis Marta, Alex^2, Isa, Mariela y Pedro. Muchas gracias por facilitarme tanto las cosas y hacerme sentir como en casa.

Todas estas experiencias me han servido, además de como aprendizaje a nivel científico, a madurar como persona y comprender que en el mundo de la ciencia hay que luchar y que sin esfuerzo, dedicación y sacrificio, no hay nada que hacer.

Sin embargo, también ha habido tiempo de desconexión y diversión, como algunos congresos. Aquí por supuesto os tengo que incluir a Jesús Martínez, Joaquín Caro y Carmen Vida, con los que he compartido la mayoría de estos ratos y, de últimas a Álvaro Polonio. Muchas gracias por esos buenos momentos que hemos compartido y que nos hemos reído tanto… Las playas paradisíacas, las excursiones improvisadas y los túneles siniestros de Grecia; los paseos nocturnos y las comilonas post-congreso en Lérida; las caminatas campo a través por los montes de Madrid y posterior recompensa; las bebidas calentitas para entrar en calor en Viena y las cenas de integración. Además de ser muy buenos compis de aventurillas os aprecio mucho como personas y como científicos. Jesús Martínez, eres un ejemplo de que con esfuerzo y tesón se pueden conseguir grandes cosas. Joaquín Caro, eres un transgresor y no te conformas con nada, pero mira eso hasta donde te ha llevado… A Harvard nada menos! Carmen Vida, eres capaz de enfrentarte con cualquier obstáculo y salir airosa; me alegro mucho de ser tu amiga y de haber compartido tantas charlas y momentos contigo… ¡y los que nos quedan! Álvaro Polonio, te has tenido que pelear con *Pseudomonas*, *Bacillus* y ahora *Podosphaera*, pero fíjate en cada rama has dejado tu huella; sólo espero que te vaya muy bien en esta etapa porque eres muy válido y te lo mereces.

También agradecer a dos personitas que son muy importantes, no sólo para mí sino para todo el grupo. A Irene Linares, por tu paciencia infinita en mis peores momentos de

agobio y estrés, porque siempre has estado ahí para consolarme. Gracias por acompañarme en esta etapa y haberme alegrado el día cuando más lo he necesitado. Sin duda, tu amistad es una de las mejores cosas con las que me quedo. A Saray Morales, porque tu comienzo no ha sido nada fácil, y a pesar de eso desde el primer momento has estado dispuesta a ayudarme en lo que ha hecho falta. Muchas gracias por tu esfuerzo y dedicación, desde luego, montar un laboratorio desde cero no es nada fácil y has luchado para que todo saliera bien.

Otra de las personas que ha sido clave para abordar este trabajo ha sido Jesús Hierrezuelo (ó Jesús 2); has participado activamente en mi trabajo, y te agradezco mucho tu colaboración y tu paciencia para explicarme todas las cosas. Sin ti, no habría sido posible obtener tantos resultados derivados de los análisis químicos. En este sentido, también agradecer a María Luisa Pola de la Unidad de Espectrometría de Masas y Mercedes Martín de la Unidad de Proteómica del SCAI, por los servicios prestados en relación a los análisis químicos y de espectrometría de masas. Gracias también a la empresa Koppert biological systems por financiar parte de este trabajo y valorar los resultados obtenidos.

A las nuevas becarias del Lab1, Franche y Sandra, desearos mucha suerte en vuestra etapa investigadora que estais comenzando y que seguro empezáis a ver frutos muy pronto. Aprovecharos ahora que teneis a Eva y Carmen que os van a aportar muchas cosas y os ayudarán en todo lo que necesiteis. A todos mis compis del BacBio desearos una estupenda trayectoria científica. Jesús Cámara (ó Jesús 3) y Marisa, me consta que estais aprendiendo muy rápido y cada vez os manejáis mejor con todas las técnicas, espero que eso se traduzca en muchas publicaciones. Zahira, eres muy trabajadora, todos entendemos que compaginar estudios e investigación no es fácil de llevar a veces; sé fuerte y sigue así que ya verás como muy pronto tu esfuerzo se verá recompensado. Elena, a pesar de ser la que lleva menos tiempo con nosotros, has sido muy importante en mi última etapa. Muchas gracias por tus ánimos, tu apoyo y tu ayuda, ya sabes que te he cogido mucho cariño y también espero poder ayudarte en esta nueva etapa que va a comenzar.

A todos mis compis muchas gracias por alegrarme las comidas, incluso, en algunos casos las cenas después de un día estresante de trabajo. A Juan Carlos Codina muchas gracias por endulzarnos las tardes con esas palmeritas de chocolate y por tu compañía, gracias por tu dedicación y actualizar nuestros currículums en el SICA. A Lola Fernández, a pesar de no conocerte demasiado, me encanta esa energía y buen rollo que desprendes y me alegro mucho por todas esas publicaciones que estás consiguiendo a raíz de tu postdoc.

A los compis de genética, Eloy, Adrián y Alba, me alegro mucho de haber tenido la oportunidad de conoceros mejor durante mi última etapa en la que me he dedicado a los análisis de qRT-PCR. Eloy, muchas gracias por tu paciencia para explicarme la técnica y cómo analizar los datos. Gracias a los tres por hacerme sentir una más y por las risas a pesar del agobio, así da gusto!

Muchas gracias a mis directores de tesis, Alejandro Pérez García y Diego Romero Hinojosa. Alejandro, me has acompañado desde el comienzo de mi trayectoria científica cuando hice el trabajo final de Máster, siempre has estado dispuesto a ayudarme cuando me ha hecho falta y me has enseñado a tener una visión más global y crítica de los datos y resultados. Diego, me alegro mucho de haber compartido contigo esta etapa, de haberte conocido mejor, tanto a nivel personal como científico, y de presenciar como creces y formas tu propio grupo investigador, otro ejemplo de que con esfuerzo se consiguen las cosas. Muchas gracias por tu dedicación en estos últimos meses especialmente porque han sido clave para estar ahora mismo donde estoy. A Antonio de Vicente Moreno, muchas gracias por tu dedicación y tus aportaciones, siempre a tener en cuenta; la verdad es que sin ti el departamento no sería lo mismo. Eres capaz de tenerlo todo bajo control y sin estrés y sabes perfectamente qué consejo dar y a quién.

Muchas gracias a mi familia y amigos, que siempre os preocupais por mí, por compartir tanto buenos como malos momentos, porque al final la vida es así. Sin embargo, es en esos momentos de estrés, agobio, incluso bajón de ánimos, cuando uno se da cuenta de lo que tiene, me siento muy afortunada, os quiero mucho.

Por último, agradecer a las tres personas que son más importantes en mi vida, mis padres, Concha y Arturo, y Tomás, porque siempre estais a mi lado para animarme cuando más lo necesito y darme fuerzas para afrontar cualquier adversidad. Papá y mamá, para mi sois un ejemplo de fortaleza y superación continua, a veces me agotaís hasta a mi. Sois las personas más generosas que conozco, dais lo mejor de vosotros siempre y no esperáis nada a cambio; estais pendientes de todo, me haceis reir y consolais en esos días en los que no me aguanto ni yo. Sois lo más grande que tengo y os echaré mucho de menos cuando comience mi nueva etapa, os quiero mucho. A Tomás, gracias por tu sinceridad y tus buenos consejos en los momentos de flaqueza en los que necesitas razones para encontrar la paz y la tranquilidad. Siempre me apoyas en todo lo que hago, y consigues sacarme una sonrisilla aunque no me encuentre bien. Sé que estos últimos meses no sólo han sido duros para mí, tu también lo has tenido que sufrir, pero a partir de ahora todo será más fácil. Me haces muy feliz y estoy convencida de que pasaré el resto de mis días a tu lado, porque no sé vivir sin tí. Te quiero mucho.

Muchas gracias,

Conchita Magno

A mis padres, Concha y Arturo

A Tomás

INDEX

FIGURES INDEX

TABLES INDEX

INDEX ANNEXES

ANNEX I

ANNEX II

RESUMEN

El uso continuado e irracional de compuestos químicos para el control de plagas y enfermedades en plantas está causando graves daños a distintos niveles: medioambiental, contaminación de acuíferos; salud pública, intoxicaciones; o ecológica, destrucción de la microbiota residente o desarrollo de resistencias por parte de los patógenos. Una estrategia complementaria al control químico es el control biológico basado en la utilización de enemigos naturales de los patógenos causantes de las distintas enfermedades. Actualmente, se apuesta por el empleo de estos agentes de biocontrol en sintonía con unas buenas prácticas culturales, el empleo de variedades resistentes de plantas, así como el uso racional de compuestos químicos, en lo que denominamos programas de control integrado.

El género *Bacillus* representa una fuente de cepas con potencial para ser usados como agentes de control biológico. Bacterias de este género se caracterizan por su capacidad para formar endosporas en condiciones ambientales adversas, asegurando así su persistencia en el medio. De forma específica, algunos miembros de este género, pertenecientes principalmente a las especies de *Bacillus licheniformis*, *B. subtilis* y *B. amyloliquefaciens*, poseen ciertas características altamente deseables como la producción de toda una serie de metabolitos secundarios los cuales pueden contribuir a la salud de planta, bien antagonizando a los patógenos, bien induciendo el crecimiento y los mecanismos de defensa de la planta. Entre los compuestos activos que producen estas bacterias encontramos los péptidos de síntesis no ribosomal, de síntesis ribosomal y de modificación post-traduccional, o compuestos no peptídicos como los poliquétidos, cada uno con distinguibles características estructurales y funcionales.

El desarrollo de tecnologías de secuenciación de nueva generación ha favorecido, en gran medida, la disponibilidad de una gran cantidad de genomas y por tanto de información relacionada con las bases genéticas implicadas en la ecología y actividad de

biocontrol. Con toda esta información y las herramientas bioinformáticas apropiadas se ha conseguido: i) establecer de forma más precisa las relaciones taxonómicas e identificación de especies bacterianas, ii) detectar genes relacionados con mecanismos de acción ya descritos y el descubrimiento de otros posiblemente implicados en nuevas estrategias de control biológico, iii) identificar los factores genéticos relacionados con patogenicidad y, como consecuencia de todo ello, saber de antemano qué cepas podrían ser más eficientes y beneficiosas para su uso y desarrollo como producto comercial para el control de plagas y enfermedades en plantas. En trabajos previos llevados a cabo en nuestro laboratorio se demostró que las cepas *Bacillus* spp. CECT 8237 y CECT 8238, anteriormente nombradas como UMAF6639 y UMAF6614 respectivamente, controlan de forma robusta y consistente enfermedades fúngicas y bacterianas de cucurbitáceas. Además, se comprobó que el principal mecanismo de acción en la filosfera de melón se basaba en el antagonismo directo mediado por la producción de los lipopéptidos iturina y fengicina. Por otro lado, en la rizosfera de melón, ambas cepas contribuyen indirectamente a mejorar la salud de la planta al promover su crecimiento e inducir una respuesta de defensa (Resistencia Sistémica Inducida, ISR, en sus siglas en inglés), que la prepara para el posible ataque de agentes patógenos. Además, todos estos mecanismos de acción parecen estar relacionados con la producción del lipopéptido surfactina, que actúa como: i) autoinductor en la formación de biofilms, asegurando así la persistencia de estos agentes de biocontrol y, por tanto, el antagonismo de forma eficiente frente a los patógenos, y ii) molécula de comunicación *Bacillus*-planta que activa el sistema inmune de las plantas (ISR). Sin embargo, a pesar de la importancia de los lipopéptidos en la capacidad de biocontrol por parte de estas dos cepas de *Bacillus,* pensamos que deben existir factores adicionales que de alguna manera contribuyen a la sorprendente actividad de biocontrol de estas cepas. Para responder a esta cuestión,

definimos una serie de objetivos para llevar a cabo: i) secuenciación y anotación de los genomas de *Bacillus* spp. CECT 8237 y CECT 8238 y clarificación de la situación taxónomica de ambas cepas dentro del género *Bacillus*, ii) detección de todo el arsenal de genes contenidos en los genomas de ambas cepas, que están potencialmente relacionados con su capacidad de biocontrol, así como el estudio de su posible funcionalidad, iii) estudio de la presencia de ciertas regiones genómicas en los cromosomas de las cepas CECT 8237 y CECT 8238, poco conservadas dentro del género *Bacillus* y iv) caracterización de una región detectada exclusivamente en la cepa CECT 8237 y probablemente implicada en la síntesis de un nuevo lantibiótico.

En primer lugar, se llevó a cabo la secuenciación y el ensamblaje de los genomas completos de ambas cepas en el Instituto de Genómica de Beijing. Estas secuencias fueron anotadas primeramente con el servidor *Rast* para obtener una aproximación y poder clasificar el conjunto de genes contenidos en dichas secuencias dentro de diversas categorías funcionales. A continuación, se realizó un análisis filogenético mediante la técnica multigen, basada en el alineamiento de las secuencias asociadas a genes implicados en el procesamiento de la información genética (*housekeeping*), así como en su regulación, como son los denominados factores sigma. Para ello, se seleccionaron una variedad de cepas pertenecientes a especies estrechamente relacionadas, como son, *B. amyloliquefaciens*, *B. subtilis*, *B. atrophaeus* y *B. licheniformis* y algunos representantes de otras especies más alejadas filogenéticamente como *B. thuringiensis* y *B. megaterium*. Como grupo externo se consideró la cepa de *Clostridium cellulolyticum* H10, representante de bacterias esporuladas y Gram-positivas. Las cepas de estudio, CECT 8237 y CECT 8238 se agruparon junto al grupo de *B. amyloliquefaciens*, más concretamente, a aquellas cepas de la especie intrínsecamente asociadas con el entorno de la planta. Por tanto, esta técnica de identificación más robusta ha permitido la

reclasificación de dichas cepas que fueron inicialmente identificadas como *B. subtilis*, en base a la homología de secuencia del gen que codifica el ARNr 16S y a sus perfiles metabólicos. Estas técnicas de identificación cada vez son menos usadas debido a la aparición de técnicas de secuenciación menos costosas y, como consecuencia directa, a la amplia disponibilidad de genomas secuenciados lo que permite abordar este objetivo de una forma más precisa. La distinción de dos grupos divergentes dentro de la especie de *B. amyloliquefaciens*, por un lado cepas relacionadas con el entorno de la planta, y por otro, las más relevantes a nivel industrial, se ha vinculado no sólo a las relaciones filogenéticas establecidas sino también al enriquecimiento en su contenido génico relacionado con ciertas actividades funcionales. Mediante la detección de genes exclusivamente representados en un grupo u otro se observó la ganancia por parte de las cepas de *B. amyloliquefaciens* asociadas a plantas en determinadas categorías funcionales como metabolismo de aminoácidos, carbohidratos o lípidos, síntesis de metabolitos secundarios o mecanismos de defensa. La presencia de genes para la degradación de distintos azúcares se correlacionó con su capacidad para aprovechar diversas fuentes de carbono alternativas disponibles en los exudados de la planta, mientras que genes relacionados con la síntesis de metabolitos secundarios, indicaría la capacidad de competir con patógenos posiblemente presentes en el entorno de la planta.

En nuestro estudio identificamos además regiones genómicas en CECT 8237 o CECT 8238 poco conservadas dentro del género *Bacillus*. Para ello, se diseñaron secuencias de comandos especiales, con la condición de que permitieran la selección de ventanas genómicas de al menos 7 genes consecutivos y cuya medida de conservación no superara el percentil 30. En este análisis se incluyeron un total de 76 genomas completos pertenecientes a diversas especies dentro del género *Bacillus*. Como resultado, se obtuvieron tres escenarios diferentes, regiones poco conservadas dentro del género y

altamente conservadas (1), parcialmente conservadas (2) o ausentes (3) dentro de la especie de *B. amyloliquefaciens*. En el primer caso, se identificaron principalmente regiones relacionadas con la síntesis de metabolitos secundarios, cuya función ya había sido caracterizada en otras cepas de esta especie, como cabía esperar. En el segundo caso, se detectaron genes relacionados con el metabolismo de ácidos grasos, esporulación, síntesis de vitaminas y proteínas relacionadas con fagos. En último lugar, destacó la presencia de genes adicionales relacionados con péptido sintetasas, tioesterasas, malonil coenzima A transacilasa y otros genes relacionados con el metabolismo de ácidos grasos y, posiblemente, con la síntesis de metabolitos secundarios desconocidos. En particular, en la cepa CECT 8238 se identificó también toda la batería de genes necesaria para el funcionamiento del sistema de secreción tipo IV, únicamente detectado, hasta la fecha, en la cepa de *B. subtilis* subsp. *subtilis* 168. Finalmente, este análisis llevado a cabo para la búsqueda de regiones poco conservadas, se validó mediante PCR usando para ello cebadores diseñados sobre algunas de las secuencias identificadas como únicas en las cepas CECT 8237 y CECT 8238, obteniéndose señal únicamente en las cepas específicas y no otras.

En tercer lugar, se ha profundizado en el arsenal de genes posiblemente relacionados con los diversos mecanismos de acción que pueden desarrollar estas bacterias para llevar a cabo su actividad de biocontrol. En cuanto a formación de biofilms y colonización se han identificado todos los genes implicados en dichos procesos descritos hasta la fecha, entre los que se incluyen aquellos relacionados con la comunicación celular, las cascadas de regulación y la síntesis de los principales componentes de la matriz extracelular. En comparación con otras cepas de *Bacillus*, tanto *B. subtilis* como *B. amyloliquefaciens*, se observaron altos niveles de conservación de los loci implicados en el programa de desarrollo que conduce a la formación de biofilms; sin embargo, en los ensayos de biofilm

basados en la morfología de colonia, observamos fenotipos muy diferentes. Pudimos confirmar la existencia de diferencias en algunos sistemas de regulación, como el de *quorum sensing comQXPA* y el tándem de operones *rap-phr*, ambos sistemas relacionados de forma indirecta con la formación de biofilms. Por un lado, el grupo de genes *comQXPA* se identificó en cepas pertenecientes principalmente a las especies de *B. subtilis* y *B. amyloliquefaciens* y se empleó para generar un árbol filogenético que permitiera identificar las divergencias en cuanto a ferotipos. Como resultado, se obtuvo que las cepas CECT 8237 y CECT 8238 poseían distintos ferotipos, al igual que la cepa tipo *B. amyloliquefaciens* subsp. *plantarum* FZB42, que curiosamente compartía ferotipo con el grupo de *B. subtilis*. Esta diversificación permite la comunicación entre cepas pertenecientes a un mismo ferotipo pero no entre diferentes; por ello, se supone que este polimorfismo es fruto de una respuesta adaptativa. Es más, el hecho de que CECT 8237 y CECT 8238 difieran en ferotipo se puede deber a que ambas cepas habitan el mismo nicho y, por tanto, dicha diferenciación puede suponer una ventaja a la hora de evitar interferencias en la expresión de genes regulados por este sistema de comunicación de *quorum sensing*. Por otro lado, en cuanto al tándem *rap-phr*, se trataron de detectar todos los genes previamente descritos para *B. subtilis*, 11 proteínas Rap (RapA a RapK) y 8 péptidos Phr (PhrA, PhrC, PhrE, PhrF, PhrG, PhrH, PhrI and PhrK), siendo éstos últimos los que se encargan de reprimir la expresión de las proteínas reguladoras intracelulares Rap. La mayoría de los genes *rap* y su correspondiente antagonista *phr*, fueron identificados en las cepas de estudio salvo algunas excepciones. Por ejemplo, *rapG* está ausente tanto en CECT 8237 como en CECT 8238. La cepa CECT 8237 presenta dos copias del gen *rapH*, nombradas como *rapH1* y *rapH2*, pero no los *phr* asociados. El tándem *rapI-phrI* sólo se ha identificado en la cepa CECT 8238, dada su relación directa con el sistema de secreción tipo IV ICEBs1, no fue de extrañar puesto que, de igual

manera, la maquinaria génica implicada en este sistema sólo se detectó en dicha cepa. Finalmente, el tándem *rapK-phrK* se identificó en las cepas CECT 8237 y CECT 8238 pero no en la tipo FZB42, mientras que una nueva aspartato fosfatasa, anotada como *rapX*, pero no su correspondiente *phrX*, se identificó en estas tres cepas y otras del género pero no en *B. subtilis*. Todos estos resultados nos llevan a pensar que el sistema de regulación *rap-phr* en *B. amyloliquefaciens* no es tan parecido al caracterizado en *B. subtilis* como cabría esperar. Por tanto, pensamos que todo el programa de desarrollo, en el que se incluyen los sistemas de comunicación indicados anteriormente, que conduce a la formación de biofilms y estudiado en *B. subtilis*, requiere de una revisión experimental, para así tratar de entender qué genera las diferencias morfológicas de colonias de especies relacionadas.

Las cepas CECT 8237 y CECT 8238 son inductoras de los mecanismos de defensa de la planta hospedadora (ISR), y promotoras del crecimiento de la planta, y sabemos que la surfactina está directamente involucrada en el proceso de ISR. En esta tesis hemos identificado otros atributos bacterianos relacionados con la ISR: los denominados MAMPs (patrones moleculares asociados a microorganismos), como diversas flagelinas, esenciales para la movilidad, los genes *tuaA-tagO*, para la síntesis del ácido teicurónico presente en las paredes celulares de bacterias Gram-positivas y *tufA*, que sintetiza el factor de elongación EF-Tu, todos componentes bacterianos esenciales que son reconocidos por la planta. Por otro lado, se han identificado los genes que participan en la ruta dependiente de triptófano para la síntesis de la fitohormona ácido-3-indolacético (AIA), molécula que promueve el crecimiento de la planta y cuya producción por parte de CECT 8237 y CECT 8238 ya se demostró en estudios previos. Además, se han localizado en los genomas de ambas cepas los genes para la producción de acetoína y 2,3-butanodiol, compuestos orgánicos volátiles mediadores en la comunicación bacteria-

planta. Para demostrar la funcionalidad de estos genes se llevaron a cabo diversos análisis, de tipo cualitativo y cuantitativo. En primer lugar, se demostró que CECT 8237 y CECT 8238 producen acetoína, como resultado del test de Voges-Proskauer, mientras que análisis realizados por cromatografía de capa fina (TLC) revelaron la presencia de 2,3-butanodiol en el sobrenadante asociado a dichas cepas. En segundo lugar, se analizaron los sobrenadantes bacterianos a distintos tiempos y se evaluaron cuantitativamente las cinéticas de producción correspondientes a ambas moléculas mediante cromatografía de gases asociada a espectrometría de masas (MS). Entre las cepas estudiadas, se pudo distinguir una mayor producción de acetoína por parte de *B. subtilis* subsp. *subtilis* 168 y NCIB 3610, y *B. amyloliquefaciens* subsp. *plantarum* FZB42; y mayor producción de butanodiol en *B. amyloliquefaciens* CECT 8237 y CECT 8238, llegando a ser hasta un orden de magnitud mayor que en las otras tres cepas. En general, las cinéticas de expresión fueron similares en nuestras condiciones experimentales, en la mayor parte de los casos se alcanzaron los valores máximos de ambos volátiles a las 24 h de crecimiento, salvo FZB42 que acumuló mayor cantidad de acetoína a las 36 h y CECT 8237 cuya mayor cantidad de acetoína apareció a las 12 h y de butanodiol a las 36 h. En paralelo a estos análisis se realizaron ensayos de promoción de crecimiento con semillas de melón y se observó que todas las cepas empleadas fueron capaces de promover la formación de radículas más gruesas y abundantes en comparación con las semillas no tratadas. A pesar de no observarse claras diferencias entre los distintos tratamientos, se puede plantear la participación de estas moléculas volátiles en la actividad de promoción del crecimiento por CECT 8237 y CECT 8238, como ya se ha demostrado anteriormente en otras cepas de *Bacillus*; no obstante, estos resultados se tendrán que confirmar en futuros trabajos. Además de la producción de moléculas inductoras del crecimiento de la planta, las bacterias pueden mejorar la salud de la misma

comportándose como biofertilizantes. En nuestro análisis genómico comparativo aparecieron los genes hipotéticamente implicados en la movilización del hierro, así como el gen *phy*, estudiado en otras cepas por su participación en la eliminación de quelatos asociados al fitato y, por tanto, dejando el fósforo libre para su utilización por parte de la planta.

En muchos casos, la actividad de promoción de crecimiento de la planta se asocia también a la capacidad del agente de biocontrol de combatir a los posibles agentes patógenos, impidiendo su proliferación, principalmente por la producción de antibióticos. Hasta la fecha, se había demostrado experimentalmente la producción por parte de CECT 8237 y CECT 8238 de los lipopéptidos surfactina, iturinas y fengicinas. Sin embargo, las predicciones del software AntiSMASH revelaron la presencia de los grupos de genes relacionados con la síntesis del sideróforo bacilibactina, del dipéptido bacilisina y de los poliquétidos macrolactina, bacillaene y dificidina. Esto se confirmó experimentalmente mediante análisis en cromatografía líquida de alta presión (HPLC) – electrospray (ESI)-espectrometría de masas, de los extractos metanólicos procedentes de los sobrenadantes libres de células de cultivos de CECT 8237 y CECT 8238 crecidos en medio Landy o GA. En el primer caso se detectaron los metabolitos bacilisina, bacillaene y dihidrobacillaene, y en el segundo se identificaron trazas de los compuestos bacilibactina, dificidina y macrolactina. Estos metabolitos secundarios han sido caracterizados con anterioridad en otras cepas utilizadas como agentes de control biológico pertenecientes principalmente a la especie de *B. amyloliquefaciens.* Además de la actividad antimicrobiana frente a patógenos fúngicos y bacterianos, hay un campo emergente en el control biológico relacionado con la capacidad de controlar enfermedades causadas por nematodos patógenos. La amplia distribución dentro del género *Bacillus* de proteasas degradadoras de cutícula sugiere una posible implicación en este otro mecanismo de acción. Así,

buscamos y detectamos los genes, *nprE* que codifica la síntesis del precursor de una proteasa extracelular neutral, y *aprE* que codifica un precursor de proteasa hipotéticamente relacionada con la degradación de cutícula.

Sin embargo, merecen especial atención algunas regiones del genoma de ambas cepas donde confluyeron regiones poco conservadas dentro del género *Bacillus,* que contenían grupos de genes potencialmente implicados en la síntesis de metabolitos secundarios, sugiriendo la posibilidad de tratarse de nuevos antibióticos. En la cepa CECT 8237 se identificó la maquinaria genética completa para la síntesis de un nuevo metabolito de síntesis no ribosomal, coincidiendo con la región atípica 18 (AR18), y únicamente detectada en el genoma de *B. amyloliquefaciens* LFB112. En la cepa CECT 8238, se detectó un grupo de genes con elevada probabilidad de sintetizar un péptido desconocido de síntesis no ribosomal, asociado a la región atípica 3. Este último grupo de genes ya había sido identificado previamente en los genomas de las cepas *B. amyloliquefaciens* subsp. *plantarum* NAU-B3, YAU B9601-Y2 y *B. amyloliquefaciens* Y2, todas asociadas a plantas. No obstante, la funcionalidad y estructuras de las correspondientes moléculas aún no se han demostrado.

En la cepa CECT 8237 además se identificó otro grupo de genes asociado a la síntesis de un metabolito secundario novedoso, ausente en todas las cepas de *B. amyloliquefaciens* incluidas en los análisis comparativos y presente, en cambio, en la cepa *B. licheniformis* 9945A. Esto despertó nuestro interés debido a su exclusividad y, por ello, nos centramos en su caracterización. En primer lugar, el análisis comparativo con las secuencias depositadas en las bases de datos, reveló la presencia de ciertos dominios conservados que nos permitieron relacionar dicha región con la posible síntesis de un nuevo lantibiótico. Por otro lado, esta región, coincidente con la AR21 en CECT 8237, parecía haber sido adquirida por transferencia horizontal de acuerdo con algunos

parámetros estudiados como la variación en el contenido de GC o el patrón trinucleotídico de lectura. En segundo lugar, se demostró la funcionalidad de los genes contenidos en dicha región mediante RT-PCR, obteniéndose como resultado tres unidades policistrónicas. Estas unidades se corroboraron por la presencia de diversos promotores al comienzo de cada unidad. Una vez demostrada la funcionalidad de los genes a nivel transcripcional, quisimos comprobar la capacidad por parte de esta bacteria de producir este nuevo lantibiótico. Para ello, nos centramos en los genes estructurales para la síntesis de los prepéptidos y los comparamos con los correspondientes a los lantibióticos previamente caracterizados, y compuestos también por dos péptidos. Sobre estas secuencias se plantearon las modificaciones postraduccionales, formación de enlaces tioéter o escisión del péptido señal, que podrían tener lugar en el lantibiótico encontrado en CECT 8237, según la información disponible y relacionada con los lantibióticos liquenicidina, haloduracina y plantaricina, producidos por *B. licheniformis*, *B. halodurans* y *Lactobacillus plantarum* respectivamente. En base a esta información previa, se pudieron detectar mediante análisis de espectrometría de masas (desorción/ionización mediante láser asistida por matriz y acoplada a un analizador TOF), dos péptidos fruto del procesamiento de los prepéptidos iniciales, nombrados como LanA1 y LanA2, en los extractos de cultivos de la cepa CECT 8237. Los péptidos LanA1 y LanA2 se asociaron a los valores de masa molecular en el rango de 3042Da y 2500Da respectivamente. A pesar de que se han planteado diversas propuestas para las estructuras de LanA1 y LanA2, sus correspondientes estructuras definitivas quedan pendientes de confirmar por otros métodos, una vez que se haya conseguido poner a punto la estrategia de purificación de ambos péptidos. En este grupo de genes también se han identificado genes relacionados con la inmunidad frente a este lantibiótico, como los que codifican los transportadores, *lanF1G1E1* y *lanF2G2E2*, pero no se identificó ningún gen homólogo a *lanI*, que codifica

una lipoproteína asociada a la membrana y que constituye un segundo sistema de inmunidad. Sin embargo, la presencia de dos sistemas de inmunidad se ha relacionado con la capacidad de un determinado lantibiótico de formar poros en la membrana de la célula diana, como ocurre en el caso de los lantibióticos compuestos por dos péptidos. Dado que esa es la situación ante la cual nos encontramos, se proponen dos hipótesis: que el gen anotado como proteína hipotética, está en realidad relacionado con la síntesis de la lipoproteína que constituye la segunda línea de defensa, o que los dos transportadores suplen dicha función de protección, como ocurre en el caso del lantibiótico haloduracina.

En cuanto a la posible regulación de la expresión de los genes que conforman el lantibiótico de CECT 8237, se analizó por PCR cuantitativa los niveles de expresión relativa de los genes estructurales *lanA1* y *lanA2* a lo largo del tiempo y en dos condiciones de temperatura diferentes, 30ºC y 37ºC. Los valores máximos de expresión se alcanzaron al comienzo de la fase estacionaria de crecimiento aunque a partir de este punto la actividad transcripcional descendió, y no se apreció ninguna influencia de la temperatura en la cinética de expresión de ambos genes. A diferencia de otros lantibióticos que presentan un sistema de regulación de dos componentes, constituido por una histidina quinasa y un regulador de respuesta, en este grupo de genes únicamente detectamos la presencia de un gen regulador, BAMY6639_14535. De acuerdo con los datos obtenidos en los ensayos de PCR cuantitativa suponemos que, además de este gen, que posiblemente regulará positivamente la expresión de los genes relacionados con el lantibiótico, debe haber otros mecanismos dependientes de la densidad celular que regulen de alguna manera la actividad de dicho gen, o incluso que ante la ausencia de células diana se produzca un descenso en la expresión de los genes (Kies et al., 2003; Coburn et al., 2004).

La capacidad de estos agentes de biocontrol de formar biofilms es de gran relevancia debido a su implicación en la colonización, persistencia y, por tanto, protección de la planta hospedadora frente a los agentes patógenos potenciales. Por ello, quisimos estudiar la producción de este nuevo lantibiótico, así como otros metabolitos secundarios, en condiciones estáticas de crecimiento, tanto en la interfaz aire-líquido como en placa, en ensayos de formación de biofilms. En los ensayos realizados en medio M líquido, las muestras se analizaron por espectrometría de masas y, además, se llevaron a cabo ensayos de expresión, a partir de ARN extraído de las células asociadas a la película o en suspensión en el medio bajo la película. En este caso, para la detección de los péptidos mediante espectrometría de masas (MALDI-TOF), no se hizo ningún procesado previo de la muestra, sino que se visualizaron *in situ*. Para llevar a cabo este estudio optimizamos las condiciones de análisis, que incluyó la elección de la matriz, una mezcla de solución de ácido α-ciano-4-hidroxicinámico y ácido 2,5-dihidroxibenzoico (CHCA-DHB), que nos permitió resolver las masas moleculares relacionadas con los péptidos. En cuanto a los niveles de expresión de los genes de síntesis *lanA1* y *lanA2,* éstos fueron mayores en las células asociadas a la película en las condiciones ensayadas; sin embargo, los datos de espectrometría de masas arrojaban una mayor intensidad de lantibiótico en el medio líquido. Este resultado podría explicarse por dos motivos: o bien las células asociadas a la película están produciendo el lantibiótico y éste es liberado rápidamente al sobrenadante, o bien existe un “gap” de tiempo entre la expresión de los genes estructurales y las modificaciones posteriores conducentes a los péptidos maduros.

Cuando estudiamos el comportamiento del gen *ituA* para la síntesis de iturina, pudimos apreciar que, al contrario que en el caso anterior, está más activo en las células planctónicas según los resultados de qRT-PCR. Esto se confirmó en los análisis procedentes del MALDI-TOF, donde encontramos una mayor acumulación de isoformas

de la iturina en las células asociadas a esta fase en las dos condiciones de temperatura consideradas. Por otro lado, cuando evaluamos los niveles de expresión relativos con respecto al gen *srfAA*, implicado en la síntesis de surfactina, observamos un comportamiento dependiente de la temperatura. Así, a 30ºC las células planctónicas están más activas en la transcripción de esta molécula, mientras que a 37ºC lo son las asociadas a la película. Al analizar los espectros de masas derivados del MALDI-TOF, observamos que la surfactina se ajustaba perfectamente al comportamiento dependiente de temperatura que se obtuvo por qRT-PCR. Estos resultados demuestran la complejidad del apasionante sistema de regulación de producción y secreción de diferentes metabolitos secundarios, y la necesidad de seguir investigando la interconexión entre ellos y el comportamiento multicelular de *Bacillus*.

Además, nos preguntamos cuál sería la distribución de los antibióticos en medio sólido. En los ensayos realizados en medio M agar se apreciaron diferencias morfológicas con respecto a las temperaturas de crecimiento ensayadas, 30ºC y 37ºC, un hecho que ya resultó interesante y discrepante en comparación con lo que se había observado en condiciones de cultivo en agitación. Además, pudimos distinguir tres áreas claramente diferenciadas en ambas condiciones de temperatura, especialmente a las 72 h de incubación; por ello, dichas áreas fueron objeto de estudio para la distribución de metabolitos a lo largo de la colonia. Al igual que en el caso anterior, la detección de los péptidos se llevó a cabo mediante espectrometría de masas *in situ*. A diferencia de lo que ocurrió en cultivos en agitación, en estas condiciones se observaron diferencias a 30 o 37ºC, lo que correlacionaba con la diferente morfología de colonia observada. Los péptidos asociados al lantibiótico se localizaron en las tres áreas definidas a 30ºC de temperatura, aunque la mayor intensidad de señal se concentró en la zona intermedia de la colonia, la cual se caracterizaba por una mayor concentración de arrugas, una

característica morfológica típica de biofilms de *Bacillus*. Por el contrario, a 37ºC, el péptido LanA1 estaba presente en las áreas interna e intermedia pero no en la zona más externa de la colonia, y sorprendentemente, el péptido 2 no apareció en ninguna de las zonas de la colonia.

En general, las células móviles se localizan mayoritariamente en la zona más externa de la colonia, presumiblemente a la búsqueda de zonas con disponibilidad de nutrientes, mientras que en las zonas más internas de la colonia, donde los nutrientes están agotados, se disparan otros procesos fisiológicos como el metabolismo secundario y la formación de biofilms. De hecho, al estudiar la distribución de otros metabolitos a lo largo de la colonia encontramos patrones similares a los detectados en el caso del lantibiótico. Así cuando observamos la distribución correspondiente a 30ºC de temperatura, la máxima concentración de los lipopéptidos iturina y fengicina, y de los poliquétidos bacillaene y dificidina se localizó en el área intermedia, mientras que la surfactina apareció mayoritariamente en el área más interna. El comportamiento de estos mismos metabolitos a 37ºC fue más variable, aunque las mayores concentraciones se detectaron en el núcleo (bacillaene, dihidrobacillaene, oxidificidina y surfactina) o zona intermedia (dificidina, iturina y fengicina) de la colonia. En estas condiciones de temperatura, sólo se detectaron trazas de los metabolitos dificidina, surfactina e iturina en la zona más externa de la colonia. Una observación interesante fue el distinto patrón de distribución detectado para las distintas isoformas de los lipopéptidos iturina y surfactina, un hecho que lleva a plantear una posible especialización de estas isoformas relacionada con su distribución espacial. Si bien los ensayos anteriores nos mostraban el comportamiento de los antibióticos en medio sólido, una condición física con implicaciones en la regulación de formación de biofilms (McLoon et al., 2011; Wang et al., 2016).

Finalmente nos preguntamos si la maquinaria génica asociada al lantibiótico y otros metabolitos conocidos también estaba activa en condiciones naturales, es decir, en la planta. Para ello, se inocularon plantas de melón con suspensiones bacterianas de *B. subtilis* subsp. *subtilis* 3610 y B. *amyloliquefaciens* CECT 8237. Como resultado de los ensayos de qRT-PCR, los cuáles se referenciaron con respecto a los datos obtenidos para la cepa de *B. subtilis* subsp. *subtilis* 3610, pudimos observar que tanto los genes estructurales del lantibiótico, como los de síntesis de iturina y surfactina, estaban activos en la planta de melón. Merece especial atención que los niveles de expresión relativa de la surfactina fueron superiores a los siete días después de la inoculación, mientras que en el resto de metabolitos estudiados, ocurría a los catorce días. Esto se podría correlacionar con la necesidad del agente de biocontrol de colonizar primero la planta, mediante la formación de biofilms sobre la hoja, lo cual está vinculado con el papel de la surfactina como molécula señal para la formación de biofilms, y a continuación, sintetizar los metabolitos con actividad antimicrobiana. Sin embargo, está aún por confirmar si los datos de expresión se corresponden con la detección química del lantibiótico sobre las hojas de melón.

A pesar del alto nivel de conservación de las secuencias de los genomas de CECT 8237 y CECT 8238 con respecto a la cepa tipo de la misma especie, en nuestro estudio hemos encontrado un número de singularidades génicas en los genomas de ambas cepas que podrían estar hipotéticamente relacionadas con su actividad de biocontrol. En efecto, en los ensayos de biocontrol sobre hojas de melón frente al hongo fitopatógeno, *Podosphaera xanthii* y el patógeno bacteriano, *Pectobacterium carotovorum*, los mayores niveles de protección se obtuvieron con las cepas CECT 8237 y CECT 8238. Como se ha expuesto anteriormente, la actividad de biocontrol consiste en la combinación de una serie de herramientas o mecanismos de acción, muchas de las cuales son compartidas por

todas las cepas de *Bacillus* usadas, de ahí que las diferencias en la reducción de los síntomas de ambas enfermedades no fuesen significativas.

Considerando todos los resultados expuestos en esta tesis, se refuerza de nuevo los múltiples mecanismos de acción que participan en la capacidad de biocontrol de *B. amyloliquefaciens* CECT 8237 y CECT 8238, basado en: i) multicelularidad, lo que asegura la colonización y persistencia en la superficie de la planta, ii) producción de fitohormonas y MAMPs que promueven el crecimiento de la planta y la ISR y iii) producción de metabolitos secundarios, que tienen una doble funcionalidad, antagonismo directo de los patógenos de planta o adaptabilidad bacteriana al entorno. En especial, la amplia variedad de metabolitos secundarios detectados, cuya producción ha sido demostrada en esta tesis, así como aquellos que se caracterizarán en próximos trabajos, junto con los resultados de estudios previos, pone de manifiesto la importancia del antagonismo basado en la producción de antibióticos en la actividad de biocontrol por parte de CECT 8237 y CECT 8238.

INTRODUCTION

INTRODUCTION

The irrational use of chemical compounds to control plant diseases is causing an appreciable number of problems leading to: i) environmental damage and the appearance of resistances from pathogens to those compounds, decreasing their effectiveness (Swarupa et al., 2014); ii) a detrimental effect in human health, a lead to a consciousness-raising movement for the implementation of alternative ways to combat pests in plants (Paul Chowdhury et al., 2015). The biological control, or the use of natural enemies of pathogens, seems to contribute to relieve part of the chemotherapy awareness.

In order to develop efficient biological control agents, it is important to consider the relation of the antagonistic and the pathogen, the host plant and the environmental conditions where this interaction will occur. In general, the rizosphere of the plants is known as a biologically active zone in soil where diverse interactions, including positive (symbiotic) and negative (pathogenic, parasitic), are taking place, due to the continuous liberation of nitrogen and carbon sources. Rhizobacteria are root-colonizing bacteria which exert beneficial effects to plants using a variety modes of action: promotion of the plant growth, induction of the host systemic resistance or direct antagonism towards the pathogens (Choudhary and Johri, 2009; Wu et al., 2015a). In contrast, the phyllosphere of the plants has been historically considered a hostile habitat due to the continuous fluctuation of temperature, relative humidity and reduced availability of nutrients; therefore the colonization and persistence are more challenging. Nonetheless, there are also bacteria capable of colonizing the phyllosphere and contribute in a multifaceted way to the health of plant (Kim et al., 2015). Among these microbial species, members of the *Bacillus* genus are potential candidates to be used as biological control agents (Pérez-García et al., 2011).

1. The genus *Bacillus* as a source of potential biocontrol agents

The *Bacillus* genus is a large and heterogeneous group of Gram-positive, rod-shaped, sporulating bacteria which can be obligate aerobes or facultative anaerobes. This genus is ubiquitously present in Nature and includes free-living and pathogenic bacteria. Under different stress conditions they produce endospores which led them to stay in dormancy for long periods (Maughan and Van der Auwera, 2011).

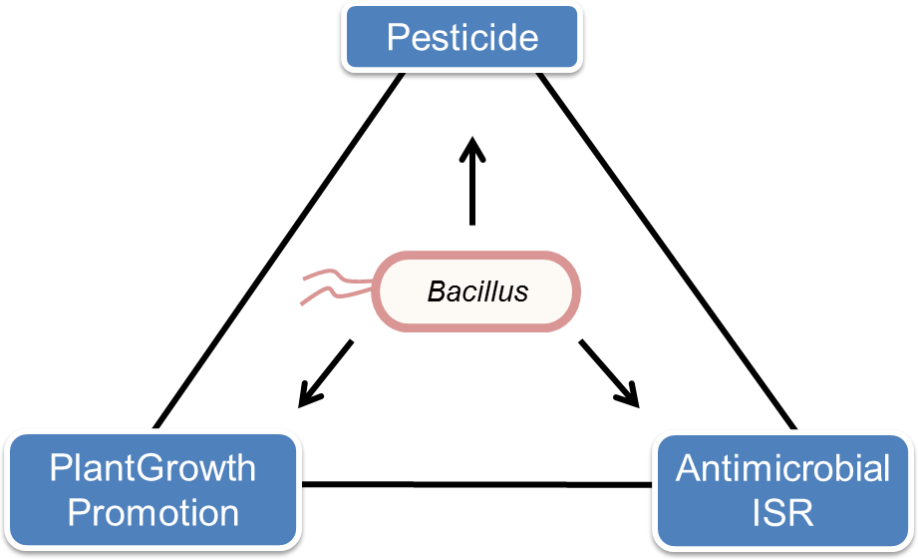


Figure 1. Agricultural application of bacilli-based products: microbial pesticides, fungicides and plant-growth inducers. Some members of the *Bacillus* species also show bactericidal and nematicide capability, but these attributes are still vaguely explored. Adapted from Pérez-García et al., 2011.

Some members of the genus show several characteristics desirable for good biological control agents. It is remarkable the production of a myriad of secondary metabolites, the efficient colonization of diverse habitats, the ability to trigger the plant defences and the growth promotion of the host plant. The resilience of the endospores to physico-chemical stress like dryness or irradiation might ensure the long-term protection of the crops and facilitate the production of reliable formulates and its storage (Cawoy et al., 2015). The

majority of the reported strains with biocontrol skills belong to three species closely related phylogenetically: *Bacillus licheniformis*, *B. subtilis* and *B. amyloliquefaciens* (Figure 2). The phylogenetic analyses of traditional taxonomy relied their classification on the variable sequence of *rrs* gene, encoding the ribosomal RNA 16S. However, the 16S taxonomy resolution is limited and seems to be useful for a quick identification of new isolates or massive sequencing projects where groups of diverse strains are considered (Maughan and Van der Auwera, 2011). However, for phylogenetic analysis within the same or closely related species a more reliable and potent analysis is required to reveal the specific relationships of the strain and other species. For that purpose, several housekeeping genes widely spread in all studied strains are usually considered which result in more accurate classification (Liu et al., 2013).

Multigene and complete genome phylogeny analyses have been mainly done with single isolates of special interest. As an example, the plant-beneficial species *B. amyloliquefaciens* and the closely related *B. methylotrophicus* have been extensively studied because of their applicability in crops as efficient biological control agents against plant pathogens and as inducers of plant-growth. In fact, through comparative genome analysis, it was recently reported that *B. amyloliquefaciens* subsp. *plantarum* strains clustered with the *B. methylotrophicus* clade, revealing that both groups should be considered as *B. methylotrophicus* (Dunlap et al., 2015). From their phylogenetic analysis, based on the complete genome sequences, Dunlap and collaborators proved that the type strains of *B. amyloliquefaciens* subsp. *plantarum* and *B. methylotrophicus* shared around 95% of the genes and contained only minor differences, while *B. amyloliquefaciens* subsp. *plantarum* and other members of the *B. amyloliquefaciens* taxon did not even cluster together.

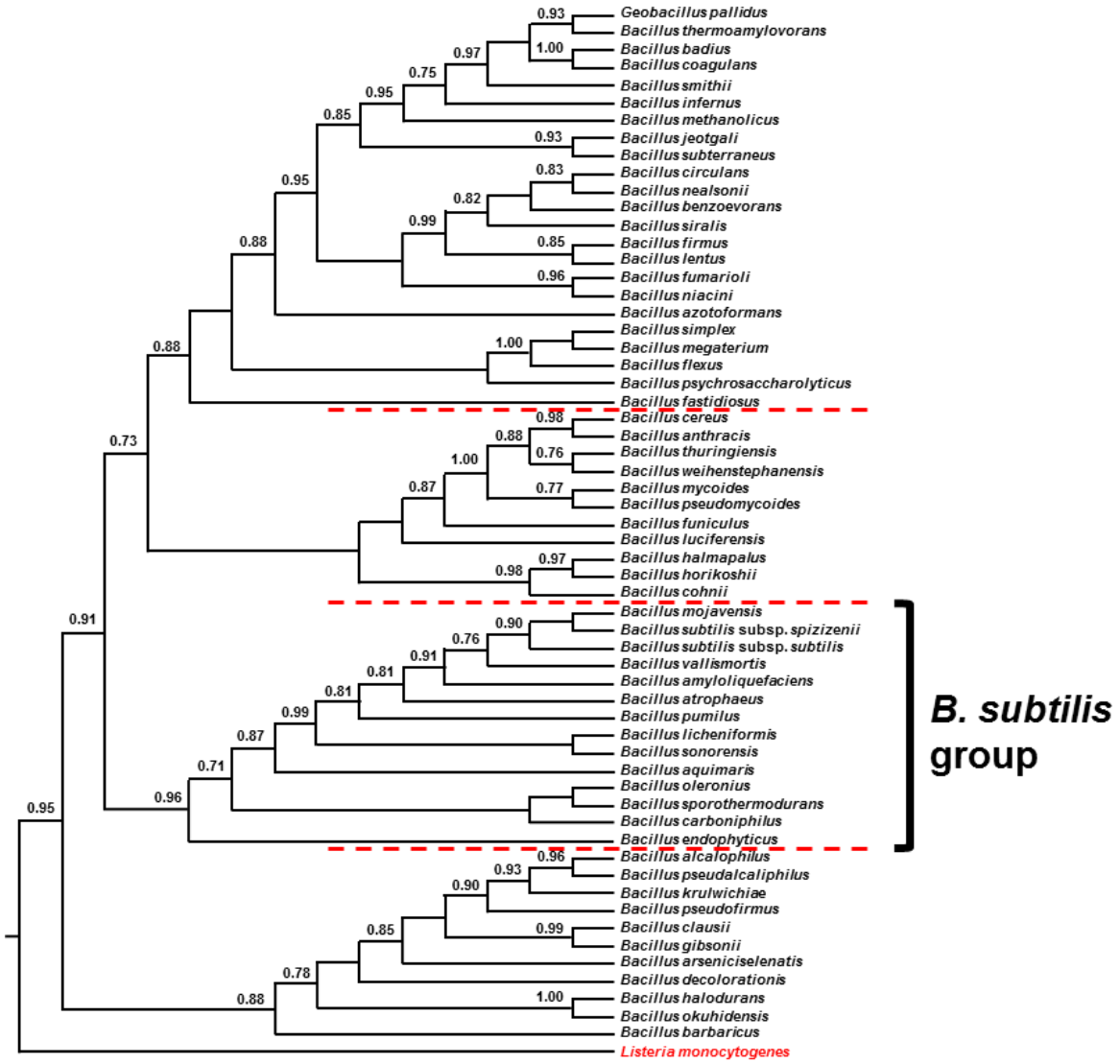


Figure 2. Evolutionary relationships of 59 *Bacillus* species inferred using Maximum Likelihood analysis of the 16S rDNA locus. Adapted from Maughan and Van der Auwera, 2011.

2. Comparative genomic analysis for the detection of specific genetic features related to biocontrol

Besides the complete genome phylogeny analysis, next-generation sequencing systems (NGS) offer a unique opportunity to investigate the bacterial genetic determinants that support the development of the biological control activities (Liu et al., 2012). The acquisition of genome sequences of numerous bacterial species has generated a vast amount of information that in combination with the appropriate bioinformatics tools, has expanded our knowledge of the genetic potential that may contribute to bacterial fitness, adaptation to the environment and interaction with plants (Figure 3) (Earl et al., 2007; Cai et al., 2014). The application of such analyses permits: i) to visualize the similarities and most importantly differences in the genetic attributes of related species and potentially used as biocontrol agents (Paul Chowdhury et al., 2015), and ii) to seed further analyses directed to elucidate the functionality of those differentiable features possibly associated with bacterial fitness, interaction with plants or biocontrol activity.

Several bioinformatics tools have been designed to extract the information from the genomic regions with possible implication in biocontrol. A good example can be found in the specific detection of genes related to the synthesis of new antibiotics as non-ribosomal peptide synthetases which contain widely conserved domains including thioesterase, adenylation and condensation domains among others (Aleti et al., 2015). In parallel, genomes comparison done with related bacteria but pathogenic helps to disregard the presence of any genetic factor related to pathogenicity to either plants or humans and that could compromise their perception as **G**enerally **R**ecognized **A**s **S**afe microbes (GRAS) (Moreno Switt et al., 2012).

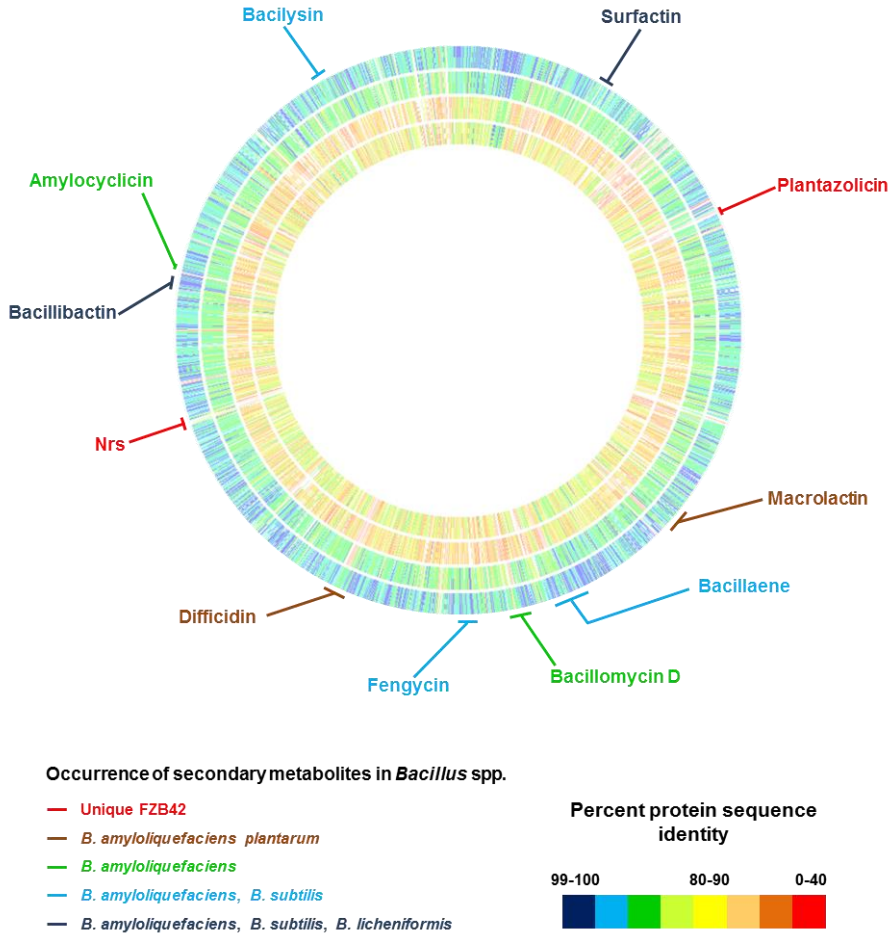


Figure 3. Genome comparison of type strains of *Bacillus amyloliquefaciens*, *B. subtilis* and *B. licheniformis*. Outer circle: *B. amyloliquefaciens* subsp. *plantarum* SQR-9; second circle: *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM7T; third circle: *B. subtilis* subsp. *subtilis* 168T; inner circle: *B. licheniformis* DSM13T. These strains were aligned with the type strain of *B. amyloliquefaciens* subsp. *plantarum* FZB42T using the RAST server. The colour code indicates the percentage of similarity of single gene products. The gene clusters responsible for non-ribosomal synthesis of polyketides, lipopeptides and other secondary metabolites are indicated. Adapted from Paul Chowdhury et al., 2015.

3. The multifaceted contribution of *Bacillus* to the plant health

Nowadays, the use of diverse products based on *Bacillus* cells as fertilizers, fungicides or pesticides is among the most extended practices in sustainable agriculture (Pérez-García et al., 2011). *Bacillus thuringiensis* could be considered a paradigm in the biological fight

especially against insect plagues caused by lepidopterans, dipterans and coleopterans larvae. The spore of *B. thuringiensis* possesses the peculiarity of containing a protein called the parasporal crystal, because of its aspect in the electron microscope, which is toxic to the insects that feed the spores (Peng et al., 2016). *B. amyloliquefaciens*, *B. licheniformis*, *B. pumilus* and *B. subtilis* are species mainly used as biofungicides due to their ability to produce biologically active compounds which directly inhibit the proliferation of plant pathogens (Lahlali et al., 2012).

The plants are continuously exposed to pathogens, and they have developed a versatile and effective defence strategy consisting of: i) pre-existing physical and chemical barriers and ii) inducible responses activated after the interaction of the pathogen with the plant. Some *Bacillus* species fight indirectly against the pathogen by triggering the defence mechanisms of the host plant, and confer immunity to the plant against a wide range of diseases, either locally or systemically. After recognition of certain molecular patterns associated to the beneficial microbe (Choudhary and Johri, 2009), the plant responds triggering distinct signalling pathways involving the salicylic acid (SA), jasmonic acid (JA), or ethylene (ET) that ends in the partial activation of the plant defence machinery, a process called “priming”. This is a physiological stage that enables the plants to activate their defence responses quicker and/or stronger after sensing biotic or abiotic stresses. In addition, the plant keeps the alert and develops, in most of the cases, local and systemic immunity conferring resistance against a wide spectrum of potential pathogens (García-Cristobal et al., 2015).

Apart from those attributes, several *Bacillus* species contribute to the plant health increasing the nutrients availability and promoting plant growth, reasons to call them biofertilizers. For instance, *B. amyloliquefaciens* produces extracellular phytases which reduce the chelates complex and release the essential nutritional minerals to the

environment, or phytohormones as indole-3-acetic acid (IAA) that promote the plant growth (Idris et al., 2007). *B. megaterium* increases the iron availability reducing the metals, and produce citoquinins which exert a beneficial effect in plant growth. Other *Bacillus* species as *B. pumilus* promote the plant growth through the synthesis of other hormones as gibberellin or abscisic acid (ABA), and jasmonic acid. Another mechanism used by strains of *B. amyloliquefaciens* and *B. subtilis* to promote the plant growth consists in the synthesis of volatile compounds, principally 2,3-butanediol and acetoin (Farag et al., 2013).

All those features described above reinforce the concept of *Bacillus*-based products as suitable candidates to be used in integrated pest management programmes. In these programmes, the biological control must be a key element, but need to be complemented with the right cultural practices, the use of resistant varieties of plants and reasonable use of chemical compounds.

3.1 Antagonism activity of *Bacillus* based on the production of antimicrobials

The efficient antagonistic activity of *Bacillus,* based on the production of antimicrobial compounds, has become the most relevant feature related to the biocontrol activity. Strains of this genus are known to secrete cell wall-degrading enzymes such as chitinases, glucanases and proteases, and peptide antibiotics and other small molecules like volatile organic compounds which inhibit to some extent the spread of the pathogen (Suyotha et al., 2013).

It is however noticeable the structural and chemical diversity of active secondary metabolites with antimicrobial activities, which can be divided into two main groups: i) peptides ribosomally synthesized and post-translationally modified (lantibiotics and similar

peptides) and ii) peptides non-ribosomally synthesized and other non-peptide molecules as polyketides, amino sugars and phospholipids (Figure 4).

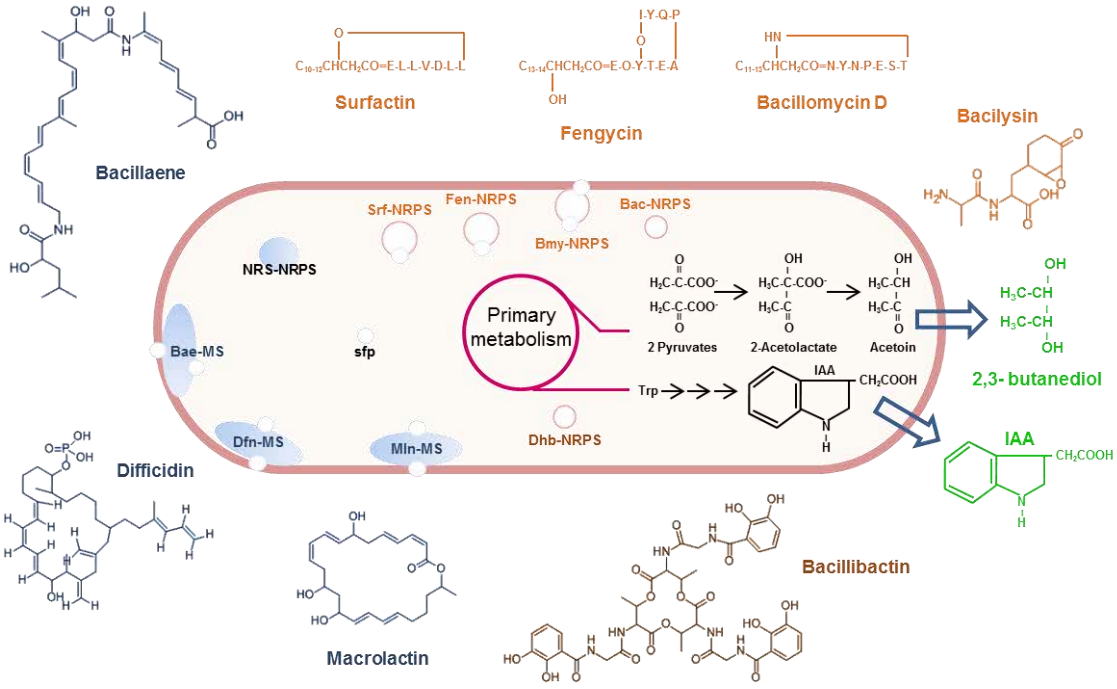


Figure 4. Non-ribosomally synthesized secondary metabolites produced by *B. amyloliquefaciens* subsp. *plantarum* FZB42, and related to biological control. The antimicrobial polyketides are synthetized by membrane-anchored polyketide-megasynthases (PKS-MS). The PKS bacillaene, difficidin and macrolactin are shown in blue. The lipopeptides surfactin, fengycin and bacillomycinD (orange) are nonribosomally synthesized by modularly organized, giant peptide synthetases (NRPS), which are either diffusible or membrane anchored. NRPSs are also involved in synthesis of the dipeptide bacilysin (orange) and the Fe2+ siderophore bacillibactin (brown). Indole-3-acetic acid (IAA) and 2,3-butanediol are shown in green. Adapted from Chen et al., 2007.

3.1.1. The role of lipopeptides in the biocontrol ability of *Bacillus*

Within the arsenal of secondary metabolites, the lipopeptides might be considered the most commonly produced by *Bacillus* species. Lipopeptides are amphipathic molecules, possess a hydrophobic lipid tail bound to a short cyclic oligopeptide, and depending on the composition of peptide ring can be classified into three main families: iturins, fengycins and

surfactins (Cawoy et al., 2015). The iturins and fengycins exert a strong antifungal activity interacting with the lipid bilayer that defines the biological membranes of the pathogen, which causes the permeabilization of the cell leading to the osmotic imbalance and finally the cell death. The surfactins are powerful surfactants that may bind to the lipid membranes in a similar way that the other lipopeptides do, but its effect can be attenuated by the cholesterol content in those membranes. Even that not as robust as the latest, it is believed that surfactin is a potent bactericidal and virucidal, but vaguely fungitoxic. However, its main contribution to the biocontrol appears to be related to the ecology and fitness of *Bacillus* on plant surfaces (Zeriouh et al., 2014; Cawoy et al., 2015).

Some works have pointed to the possible cooperation of these three families, in targeting cell membranes (Maget-Dana et al., 1992; Zeriouh et al., 2014). It was recently published that [Ile7] surfactin homologous retained fungitoxic activity against *Botrytis cinerea* in a synergetic manner in cooperation with bacillomycin D, which suggests a functional specialization depending on the surfactin homologous (Tanaka et al., 2015). In the context of plant surfaces, surfactin seems to act as a signal molecule that triggers the genetic cascade that activates the synthesis of the amyloid-like protein TasA and exopolysaccharides (EPS), the main components of the extracellular matrix, leading to the establishment of robust biofilms (López et al., 2009; Romero, 2013; Zeriouh et al., 2014). In parallel, the production of fengycins and iturins occurs to fight potential pathogens through direct antagonism. Additionally, it has been proven the implication of fengycin and surfactin in eliciting the ISR in plants conferring an alarm status for subsequent pathogens interaction (Cawoy et al., 2013; García-Gutiérrez et al., 2013). Taking together, the interactions between these three families would ensure the persistence of a biocontrol agent in plant and the long-term protection of the plants through interconnected mechanisms of action which participate in the biocontrol processes.

3.1.2. The implication of polyketides and other non-ribosomally synthesized secondary metabolites in antagonism of *Bacillus*

The polyketide difficidin and the dipeptide bacilysin have been demonstrated to efficiently control the fire blight caused by *Erwinia amylovora*, a bacterial disease that affects apple and pear trees (Chen et al., 2009). In addition, both compounds have shown antibacterial activity against *Xanthomonas oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola*, the etiological agents of the bacterial blight and bacterial leaf streak diseases of rice (Wu et al., 2015b). It has been reported that difficidin inhibits the protein synthesis and possibly damages the cell membranes, while bacilysin action relies on its transport into the host cells, where it is hydrolysed by intracellular peptidases to anticapsin which inhibits the activity of glucosamine 6-phospate synthase (GlcN6P) interfering with the cell wall synthesis (Özcengiz and Öğülür, 2015). It was also published the anti-cyanobacterial activity of bacilysin, as a consequence of the damage inflicted to the algal cell wall and cell organelle membranes (Wu et al., 2014).

As the other *Bacillus* polyketides, macrolactins show antibacterial activity and might have the potential to be used in medical application. Macrolactin shows bacteriostatic rather than bactericidal activity against clinical strains of multidrug-resistant Gram-positive bacterial pathogens including methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant enterococci, and a small-colony variant of *Burkholderia cepacia* (Romero-Tabarez et al., 2006). In addition, macrolactin has exhibited significant antagonistic effects against *Ralstonia solanacearum*, the causal agent of Granville wilt of tobacco, other solanaceous plants, bananas and peanut (Yuan et al., 2012).

The bacillaene A compound produced by *Bacillus* strains associated to fungus-growing termite *Macrotermes natalensis* shows selective inhibition of known and putatively antagonistic fungi of *Termitomyces* (Um et al., 2013). The catecholate iron-siderophore

bacillibactin enables *Bacillus* cells to accumulate and take up iron ions from their natural environment under iron limitation, an ecological advantage against competitors in these starving conditions. Instead of targeting directly the pathogens, bacillibactin might contribute to the biocontrol skills of *Bacillus* by competition for micronutrients against the potential pathogens (Miethke et al., 2006).

3.1.3. Peptides ribosomally synthesized: The lantibiotics

Lantibiotics are antimicrobial peptides ribosomally synthesized as inactive prepeptides and post-translationally modified to their biologically active forms. They are characterized for the presence of intramolecular rings formed by the nonproteinogenic thioether amino acids lanthionine and methyllanthionine, a modification that gives the name to this family of molecules (**lant**hionine containing **antibiotics**) (Barbosa et al., 2015).

Lantibiotics are chemically and structurally very diverse but they are exclusively produced by and act mainly against Gram-positive bacteria and, in some cases, also exhibit antimicrobial activity against Gram-negative bacteria. The extensive posttranslational modifications of the peptides are responsible for the heterogeneity in the final structure and antimicrobial spectrum (Willey and Donk, 2007; Hyungjae and Hae-Yeong, 2011). The posttranslational modifications commonly found in all lantibiotics are the dehydration of Ser and Thr residues in the propeptide to yield 2,3-didehydroalanine (Dha) and (Z)-2,3-didehydrobutyrine (Dhb), respectively. This is followed by the stereospecific intramolecular addition of a Cys residue onto Dha or Dhb to form a lanthionine (Lan) or methyllanthionine (MeLan). Further processing of the peptides implies the proteolytic cleavage of the signal peptide which permits their export outside the cells. Based in the specific modifications, lantibiotics can be divided into three different sub-types:

a. **Type I lantibiotics** include the linear peptides that are modified by the enzymes LanB, which dehydrates Thr and Ser residues, and LanC, which mediates the cyclization of the peptide. The peptides are exported by LanT and the signal peptides cleaved by the signal peptidase LanP. Some examples of the class I lantibiotics are subtilin, nisin, and epidermin produced by *B. subtilis, Lactococcus lactis*, and *Staphylococcus epidermidis*, respectively (Götz et al., 2014; Spieß et al., 2015). In particular, the mode of action of nisin, a paradigm of this group, is based on the interaction with the cell wall precursor lipid II inhibiting the polymerization of murein subunits which lead to the formation of pores in the plasmatic membrane.
b. **Type II lantibiotics** are globular peptides modified by LanM enzymes, which are responsible for dehydration and cyclization of the prepeptides. The secretion and processing of the signal peptide are done by LanT, a multifunctional protein with a conserved N-terminal cysteine protease domain. Modified versions of this group are the two-component lantibiotics: each peptide is separately synthesized by its own structural gene and modified by its own modifying enzyme, and further assembled in the mature two-component lantibiotic. Nevertheless, a single LanT enzyme removes the signal peptide that permits the export of both molecules outside the cells. It is remarkable that each peptide separately shows weak or no antibiotic activity, which is however potentiated when acting together. Lantibiotics of this group are mersacidin and cinnamycin, examples of one-component produced by *B. amyloliquefaciens* and *Streptomyces cinnamoneus* respectively, or lacticin, haloduracin and lichenicidin, representative of the two-component version synthesized by *Lactococcus lactis*, *B. halodurans* and *B. licheniformis* respectively (Lawton et al., 2007; Dischinger et al., 2009; Ökesli et al., 2011; He et al., 2012; Samar et al., 2012).

c. Type III lantibiotics constitute a heterogeneous group of lanthionine-containing peptides that have mainly morphogenetic functions and very limited antibiotic activities. They are known to possess different ways of maturation, including a unique structural variation of lanthionine named labionin (Lab), which consists in a carbacyclic ring coupled through a quaternary carbon to lanthionine. In addition, their enzymes have well-defined domains architecture. For example, the multifunctional enzyme LabKC, responsible of Lab occurrence in labyrinthopeptins (labionin-containing lantibiotics), contains three different domains: lyase, kinase and cyclase, like other enzymes from this class, all necessary for the lantibiotic maturation (Meindl et al., 2010). The peptides included in this family are, among others, SapB, SapT, erythreapeptin-7 and labyrinthopeptin A1 produced by *Streptomyces coelicolor*, *Streptomyces tendae*, *Saccharopolyspora erythraea* and *Actinomadura namibiensis* respectively (Kodani et al., 2004; Kodani et al., 2005; Völler et al., 2012; Férir et al., 2013).

A further classification of the lantibiotics attends to their mode of action (Figure 5): i) some bind to the lipid II of the target bacteria, inhibiting the cell wall biosynthesis, ii) some attack the bacterial membrane causing the formation of pores or iii) others combine both modes of action: inhibition of peptidoglycan biosynthesis by binding and dislocation of the lipid II, and induction of pore formation in the bacterial membranes. Both modes of action can be implemented in a single molecule or being a combination of two functionally specialized peptides. Subtilin, representative of the type I lantibiotics, forms pores in the membrane as a result of the interaction of the N-terminal Tryptophan residue and the lipid II. The lantibiotic mersacidin (type II) associates with the lipid II, forming a stable complex that interferes with the biosynthesis of the cell wall at transglycosylation level, blocking the incorporation of glucose and D-alanine into cellular macromolecules. The two-peptide lantibiotics like haloduracin (type II), is fully active by the synergetic action of both

peptides. The first peptide, Halα, binds the lipid II and prevents the peptidoglycan biosynthesis, while the second peptide, Halβ, recognizes the complex Halα-lipid II leading to membrane permeabilization (Barbosa et al., 2015).

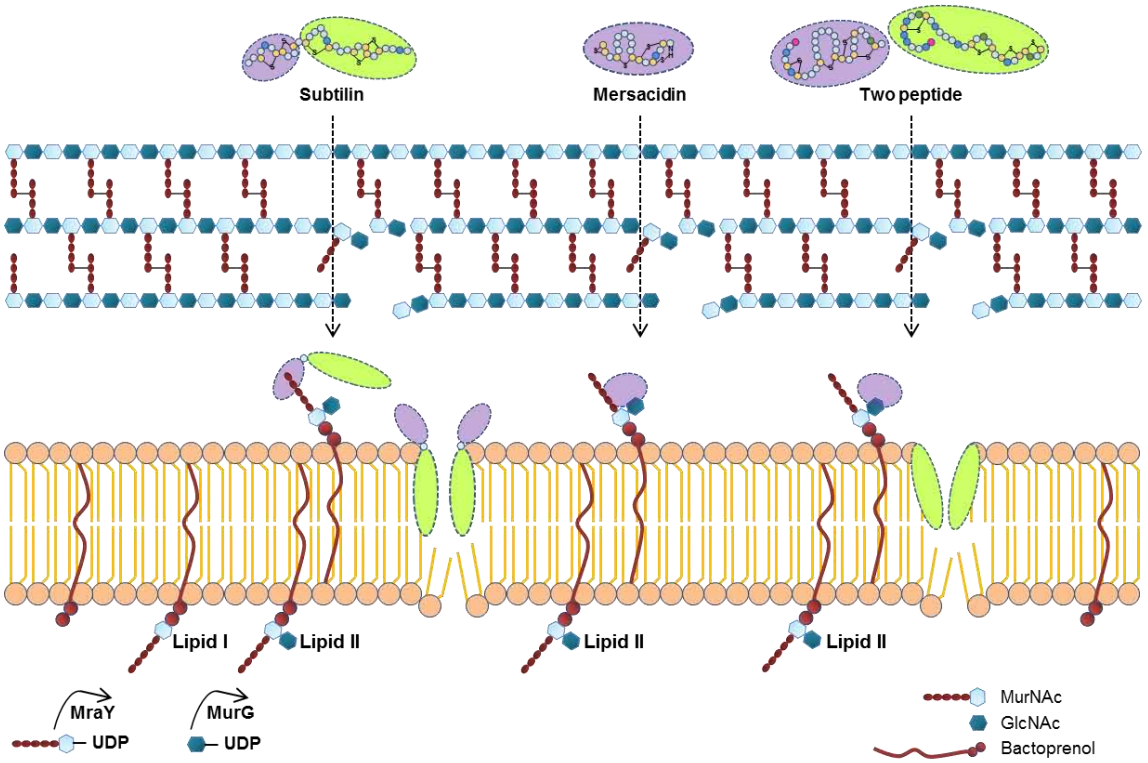


Figure 5. Schematic representation of the mode of action of the lantibiotics synthesized by *Bacillus* spp. This process of interaction involves binding of the target molecule, lipid II, preventing the correct synthesis of the cell wall. Additionally, pore formation may derive from disruption of the cellular membrane. Adapted from Barbosa et al., 2015.

Considering all the information mentioned above, it might be deduced that the selection of a promising biocontrol agent must be preceded by the tight investigation of the arsenal of bacterial features that contribute to its bioactivity to i) ensure the best results in terms of plant protection and ii) promote its use as a trusted tool for disease control either alone or included in integrated pest management programmes (IPMs) (Xu and Jeger, 2012).

OBJECTIVES

In previous research conducted in our laboratory, two *Bacillus* strains, CECT 8237 (formerly UMAF6639) and CECT 8238 (formerly UMAF6614) were proved to be good candidates as biocontrol agents against fungal and bacterial diseases of cucurbits. In the melon phyllosphere, the direct antagonism towards fungal and bacterial pathogens mediated by the production of iturin and fengycin lipopeptides appears to be the main mechanism (Romero et al., 2007b; Zeriouh et al., 2011). In the melon rhizoplane, both strains are able to contribute to plant health by promoting the plant growth and inducing the systemic resistance of the plants (ISR) (García-Gutiérrez et al., 2012; García-Gutiérrez et al., 2013). Either way, this multifaceted mode of action appears to be related to the production of the lipopeptide surfactin, which acts as i) a self-trigger of biofilm formation, which ensures persistence of these biocontrol agents and the efficient antagonism towards the pathogens (Zeriouh et al., 2014), and ii) an inter-kingdom communication molecule that activates the immune system of plants (ISR) (Ongena et al., 2007; García-Gutiérrez et al., 2013). In addition to lipopeptides, we presumed that additional features of these two *Bacillus* strains must be implicated in their remarkable biocontrol activities and adaptation to the environment. In this study, the complete genome sequences analysis of both strains was performed with the hope to reveal other potential bacterial factors implicated in their biological control activity. To do so, we plan the next specific aims:

1. To sequence and annotate the genomes of *Bacillus* spp. CECT 8237 and CECT 8238 and clarify their taxonomic situation within the *Bacillus* genus.

2. To examine the whole arsenal of biocontrol-related features contained in the genome sequences of CECT 8237 and CECT 8238 and evaluate their functionality.

3. To reveal the presence of genomic regions, ideally poorly conserved within the *Bacillus* genus, and potential reservoirs of additional genetic factors related to the biocontrol activity.

4. To characterize a genomic region exclusively detected in CECT 8237 and possibly implicated in the synthesis of a novel lantibiotic.

MATERIALS AND METHODS

MATERIALS AND METHODS

Bacterial strains and growth conditions

The strains used in this study are listed in Table 1. All of these strains were grown in Luria–Bertani medium (LB) at 37°C overnight with agitation, except *Brevibacillus laterosporus* CECT 15, which was grown at 28°C. The medium M (1% of tryptone, 1% of NaCl, 1% KH_2PO_4, 0.5% of yeast extract, pH 6.5) was used for the production of the lantibiotic. For biofilm experiments, 10 µl or 2 µl of the starting culture was spotted onto MSgg, medium M or LB broth in pellicle formation assays, or solidified with agar (1.5%) in colony morphology analyses, respectively, and incubated without agitation at 30°C (Romero et al., 2010).

Table 1. Bacterial genomes and strains used in this study.

Strains	GenBank Code	Genome Size (Mb)
Comparative analysis (Atypical Regions)		
Bacillus amyloliquefaciens		
CECT 8237	CP006058	4,03
CECT 8238	CP006960	4,01
CC178	CP006845	3,92
DSM7	FN597644	3,98
IT-45	CP004065	3,93
LFB112	CP006952	3,94
LL3	CP002634	4,00
TA208	CP002627	3,94
XH7	CP002927	3,94
Y2	CP003332	4,24
subsp. *plantarum* AS43.3	CP003838	3,96
subsp. *plantarum* CAU-B946	HE617159	4,02
subsp. *plantarum* FZB42	CP000560	3,92
subsp. *plantarum* NAU-B3	HG514499	4,20
subsp. *plantarum* UCMB5033	HG328253	4,07
subsp. *plantarum* UCMB5036	HF563562	3,91
subsp. *plantarum* UCMB5113	HG328254	3,89
subsp. *plantarum* YAU B9601-Y2	HE774679	4,24
Bacillus anthracis		
Ames	AE016879	5,23
Ames Ancestor	AE017334	5,23
A0248	CP001598	5,23

A2012	AAAC01000001	5,09
CDC 684	CP001215	5,23
H9401	CP002091	5,22
Sterne	AE017225	5,23
Bacillus atrophaeus		
1942	CP002207	4,17
Bacillus cellulosilyticus		
DSM 2522	CP002394	4,68
Bacillus cereus		
AH187	CP001177	5,27
AH820	CP001283	5,30
ATCC 10987	AE017194	5,22
ATCC 14579	AE016877	5,41
B4264	CP001176	5,42
E33L	CP000001	5,30
FRI-35	CP003747	5,08
F837/76	CP003187	5,22
G9842	CP001186	5,39
NC7401	AP007209	5,22
Q1	CP000227	5,21
03BB102	CP001407	5,27
biovar anthracis str. CI	CP001746	5,20
Bacillus clausii		
KSM-K16	AP006627	4,30
Bacillus coagulans		
2-6	CP002472	3,07
36D1	CP003056	3,55
Bacillus cytotoxicus		
NVH 391-98	CP000764	4,09
Bacillus halodurans		
C-125	BA000004	4,20
Bacillus licheniformis		
ATCC 14580 (=DSM13)	CP000002	4,22
9945A	CP005965	4,38
Bacillus megaterium		
DSM 319	CP001982	5,10
QM B1551	CP001983	5,10
WSH-002	CP003017	4,98
Bacillus pseudofirmus		
OF4	CP001878	3,86
Bacillus pumilus		
SAFR-032	CP000813	3,7
Bacillus selenitireducens		
MLS10	CP001791	3,59
Bacillus sp.		
JS	CP003492	4,12
1NLA3E	CP005586	4,82
Bacillus subtilis		
BSn5	CP002468	4,09
QB928	CP003783	4,15

XF-1	CP004019	4,06
subsp. *natto* str. BEST195	AP011541	4,09
subsp. *spizizenii* str.TUB10	CP002905	4,21
subsp. *spizizenii* str.W23	CP002183	4,03
subsp. *subtilis* str. BAB-1	CP004405	4,02
subsp. *subtilis* str. BSP1	CP003695	4,04
subsp. *subtilis* str. RO-NN-1	CP002906	4,01
subsp. *subtilis* str. 168	CM000487	4,21
subsp. *subtilis* str. 6051-HGW	CP003329	4,22
Bacillus thuringiensis		
Al Hakam	CP000485	5,26
BMB171	CP001903	5,33
Bt407	CP003889	5,50
HD-771	CP003752	5,89
HD-789	CP003763	5,50
MC28	CP003687	5,41
serovar chinensis str. CT-43	CP001907	5,49
serovar finitimus str. YBT-020	CP002508	5,36
serovar konkukian str. 97-27	AE017355	5,24
serovar kurstaki str. HD73	CP004069	5,65
serovar thuringiensis str. IS5056	CP004123	5,49
Bacillus weihenstephanensis		
KBAB4	CP000903	5,26

Strains	**References**	**Purposes**
Experimental application		
Bacillus amyloliquefaciens		
CECT 8237	Laboratory collection	Traceability / Biofilm
CECT 8238	Laboratory collection	Traceability / Biofilm
BGSC 10A1	BGSC*	Traceability
BGSC 10A3	BGSC	Traceability
BGSC 10A5T	BGSC	Traceability
BGSC 10A18	BGSC	Traceability
subsp. *plantarum* FZB42	BGSC	Traceability / Biofilm
Bacillus licheniformis		
BGSC 5A36 (=DSM13)	BGSC	Lantibiotic detection
Bacillus subtilis		
subsp. *subtilis* str. 168	Laboratory collection	Traceability / Biofilm
subsp. *subtilis* str. 3610	Laboratory collection	Biofilm
UMAFBiA758	Laboratory collection	Traceability
UMAF1605	Laboratory collection	Traceability
UMAF1610	Laboratory collection	Traceability
UMAF6619	Laboratory collection	Traceability
UMAF8561	Laboratory collection	Traceability
UMAF8562	Laboratory collection	Traceability
Bacillus spp.		
Bacillus cereus		
ATCC 14579 (CECT 148)	CECT	Traceability

UMAF8564	Laboratory collection	Traceability
Bacillus flexus		
CIP 106928T	CIP	Traceability
Bacillus thuringiensis		
serovar kurstaki str. CECT 4454	CECT	Traceability
Brevibacillus laterosporus		
CECT 15	CECT	Traceability
Paenibacillus polymyxa		
CECT 155	CECT	Traceability

***** BGSC: *Bacillus* Genetic Stock Center; CECT: Spanish Type Culture Collection; CIP: Collection of Institute Pasteur.

Biological control assays

Biological control assays were conducted on detached melon leaves (*Cucumis melo* cv. Rochet) using the double petri plate system described elsewhere (Romero et al., 2004; Romero et al., 2007b; Zeriouh et al., 2011). Briefly, cell suspensions of *Bacillus* strains were sprayed on melon leaves. After the whole leaf surface was completely air dried, the leaves were inoculated with cell suspensions (10^8 CFU/ml) of the phytopathogenic bacteria *Pectobacterium carotovorum* subsp. *carotovorum* NCPPB 2349 or conidial suspensions (10^5 spores/ml) of the cucurbit powdery mildew fungus *Podosphaera xanthii* 2086. After inoculation, the leaves were maintained at 25°C with a 16-h photoperiod. Disease severities were evaluated according to a 0-3 scale of specific values based on the diameter of chlorotic and necrotic symptoms for the bacterial disease and the percentage of the leaf area covered with fungal biomass for powdery mildew (Romero et al., 2004; Romero et al., 2007b; Zeriouh et al., 2011).

The data were analysed using SPSS v.20.0 software (SPSS Inc., Chicago, IL, USA). A one-way analysis of variance (ANOVA) was applied, and the means of each treatment were separated by Fisher's least significant difference test (P=0.05).

Genome sequencing and assembly

Chromosomal DNA was isolated using the Jet-Flex genomic DNA purification commercial kit (Genomed Laboratories), and genome sequencing and assembling were performed at the Beijing Genomic Institute (BGI, Shenzhen, China). A library of randomly sheared DNA fragments 0.5-2 kb in size was subjected to Illumina GA II (Solexa) Sequencing. All the generated reads were qualitatively assessed before assembling with SOAP *de novo*. Primer walking and polymerase chain reaction (PCR) amplification were used to fill the remaining gaps and to solve misassembled regions caused by repetitive sequences.

The genome sequences were deposited in NCBI GenBank under accession numbers CP006058 (CECT 8237) and CP006960 (CECT 8238).

Genomic data and annotation

The assembled genomes of CECT 8237 and CECT 8238 were submitted to the NCBI Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/) for automatic annotation and manually reviewed. Gene locations and protein products were generated from above annotation (ASN.1 file) using the script "asn2all" belonging to the NCBI Toolkit (http://www.ncbi.nlm.nih.gov/toolkit). Genome sequences of the *Bacillus* genus strains used in this work were downloaded from the NCBI complete bacterial genome repository (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria). The corresponding accession numbers are summarised in Table 1. Annotation of COG categories were computed by aligning the set of predicted protein sequences against the COG PSSM of the CDD (http://www.ncbi.nlm.nih.gov/cdd) using rpsblast. Only hits with an E-value ≤ 0.00001 were retained.

Circular layouts were generated using Circos (Krzywinski et al., 2009).

Phylogenetic analyses

For these analyses, some of the genomes of *Bacillus* spp. strains listed in Table 1 were selected including the closely related species *B. amyloliquefaciens*, *B. subtilis*, *B. atrophaeus* and *B. licheniformis* and several representatives of other species as *B. thuringiensis* and *B. megaterium*. The comparison of partial sequences corresponding to several housekeeping, a sporulating and some sigma factor genes was performed. All of these sequences were refined and adjusted to a consensus size by ContigExpress, which belongs to Vector NTI Advance 10 (Life Technologies) and was concatenated in the following order: *nusA-rpoA-dnaA-rpoB-gyrA-gyrB-rpoC-spoVG-sigW-sigH-sigB*. Multiple alignments were conducted with *ClustalW* and a phylogenetic tree was constructed in MEGA 5 using the Neighbour-joining method based on a pairwise distance matrix with the Tamura-Nei nucleotide substitution model (Tamura et al., 2011).

A similar procedure was used in the comparative study with the competence loci *comQXPA*. The sequences of these proteins were separately refined and adjusted to a consensus size. Further, they were concatenated as they are naturally organised in the genome (*comQ-comX-comP-comA*) and aligned to build a phylogenetic tree using the MEGA 5 software and the Neighbour-joining method based on a pairwise distance matrix with the Tamura-3-parameter nucleotide substitution model. The topology of the trees was evaluated by the bootstrap resampling method with 10000 replicates.

For the construction of the phylogeny shown in Annex Figure 1, only plant-associated *B. amyloliquefaciens* strains (Table 3) were considered. Then, CDS ubiquitously present in such strains and absent in industrial *B. amyloliquefaciens* strains were selected. This group initially contained 24 CDS. Among them, 8 were excluded based on their divergent lengths (more than 20 amino acids than the same CDS in the rest of the strains). For each strain, we concatenated the rest of the 16 ubiquitous CDS and used MEGA 5 to align them

and build a phylogenetic tree based on a Neighbour-joining method as in above trees. Information on each of the 24 CDS among plant-associated *B. amyloliquefaciens* strains considered in this analysis is included in Annex Table 1.

Comparative genome analysis

To evaluate the genetic differences between *B. amyloliquefaciens* strains corresponding to plant-associated and non-plant-associated representatives (Table 3), the pan-genome of both groups was determined using an iterative BLASTp approach (E-value ≤ 5×10^{-7}). Starting with the whole set of predicted CDS from CECT 8237, we performed iterative BLASTp versus all the plant-associated *B. amyloliquefaciens* genomes sequentially. The genome of CECT 8238 was used for the first iteration. During each iteration, the CDS missing (BLASTp non-hits) from the current version of the pan-genome were identified and added to the pan-genome before the next strain was considered (Baltrus et al., 2011). A similar procedure was carried out to determine the pan-genome corresponding to non-plant-associated strains. The CDS of each pan-genome were compared with each other using BLASTp (E-value ≤ 5×10^{-7}) and those only present in either pan-genome were retained. Using this method, it was possible to identify all the CDS contained in at least one of the group of strains but not in the other.

For the detection of atypical regions in CECT 8237 and CECT 8238 strains non-conserved in the *Bacillus* genus, windows of at least seven consecutive genes with low conservation coverage were retained (Figure 6). Firstly, the conservation measure (CM) for a given gene product in a specific strain with respect to the studied strain (CECT 8237 or CECT 8238 as it corresponds) was calculated. Then, a matrix was generated consisting of the CMs corresponding to all genes of a given strain. In order to evaluate the conservation coverage that is associated to a given gene product belonging to the studied

strain (CECT 8237 or CECT 8238), the average of all CM for that gene was estimated. Finally, these data are used to select windows of at least 7 consecutive genes whose conservation coverage is under the 30th percentile.

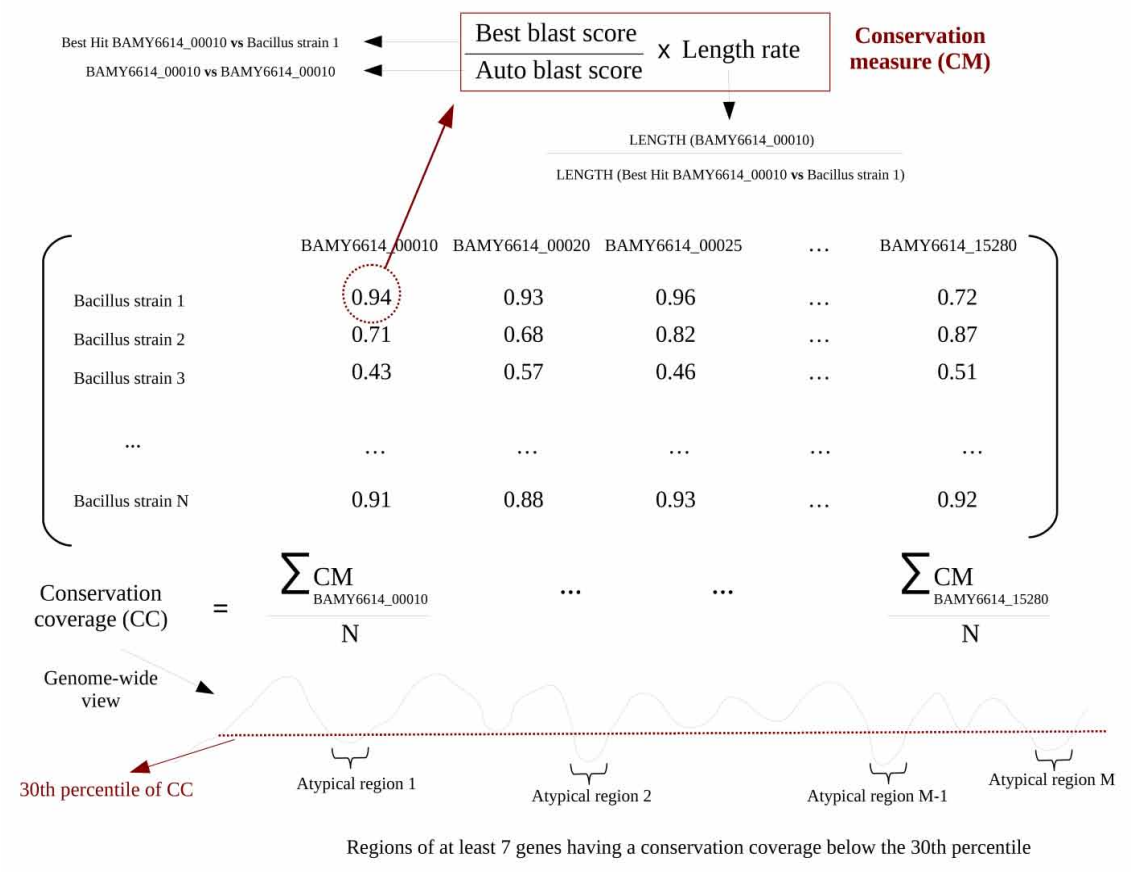


Figure 6. Flow chart of the analysis done to detect atypical regions in *Bacillus amyloliquefaciens* CECT 8237 and CECT 8238. A. Formula to calculate the conservation measure (CM) for a given gene product in a specific strain with respect to the studied strain (CECT 8237 or CECT 8238). B. Matrix consisting of the conservation measures of all genes of a given strain. C. Formula to evaluate the conservation coverage that is associated to a given gene product belonging to the studied strain (CECT 8237 or CECT 8238). D. Selection of atypical regions by selecting windows of at least 7 consecutive genes with low conservation coverage.

Prediction of protein domains and events of horizontal transference of DNA

The prediction of putative horizontal gene transfer (HGT) events was performed using the Alien Hunter software (Vernikos and Parkhill, 2006), and potential prophage loci were detected with the Prophage Finder tool (Bose and Barber, 2006). To evaluate the trinucleotide pattern for each chromosome, the distribution of all 64 trinucleotides was determined for the whole DNA sequence and for 2 kb sub-windows. Then, the χ^2 statistic on the difference between the trinucleotide composition of each window and that of the whole chromosome were computed. Large values for χ^2 in a given window denote different trinucleotide composition from the rest of the chromosome. The probability values were computed assuming uniform distribution of the DNA composition along the genome. Considering the level of error of this assumption, high χ^2 values should be interpreted as indicators of unusual regions on the chromosome that require further investigation.

The prediction of biosynthesis gene clusters dedicated to production of secondary metabolites was performed with the antiSMASH software 2.0 (Blin et al., 2013). Additionally, the prediction of putative promoters associated with transcriptional units in the lantibiotic gene cluster was determined with the use of softberry software (http://linux1.softberry.com/berry.phtml).

PCR analysis for bacterial traceability

DNA was obtained from bacteria grown in LB medium for 12 h at 37ºC using the Ultraclean Microbial DNA Isolation commercial kit (MO BIO Laboratories). The pair of primers CECT 8237-Fw (GGCAGAACAAGAGCAATC) and CECT 8237-Rv (GTCCATGTGAGTGAAATCC) were designed on the loci *BAMY6639_17480* of *B. amyloliquefaciens* CECT 8237, and CECT 8238-Fw (ATTGCCTTTTGGATGATTCG) and CECT 8238-Rv (TCAAGTGGATTTTTGGGAGA) on the loci *BAMY6614_00315* of *B.*

amyloliquefaciens CECT 8238. The pair of primers Bm-Fw (GCGATTTGTATGCCTATTTTACA) and Bm-Rv (GCCGTCATACAATTGAATCAGTT) was designed on the *bmyB* loci of the operon involved in the synthesis of the lipopeptide bacillomycin. The pair of primers Rp-Fw (GCGTGGATATGGTACTAC) and Rp-Rv (CTTCAAGTGATTTGCGTCC) amplified a fragment of the housekeeping gene *rpoA*. Standard PCR was performed using the Go Taq Flexi DNA polymerase with 1-2 ng of DNA in a final reaction volume of 25 µl according to the manufacturer's instructions (Promega). The PCR reaction cycle was performed at 94°C for 2 min, followed by 35-cycle amplification programme (94°C for 1 min, 61°C for 1 min and 72°C for 1 min) and a final extension cycle at 72°C for 7 min. The primers that partially amplified *rpoA* (a housekeeping gene present in all strains) or *bmyB* (gene involved in the biosynthesis of bacillomycin) were included in these studies.

RNA extraction

RNA was isolated from cultures of *B. amyloliquefaciens* CECT 8237 grown overnight at 30ºC or 37ºC in LB agar concurrently. Then, two millilitres of a bacterial suspension in medium M, with a final optical density of 2 (600 nm) was used to inoculate flasks containing 100 millilitres of the same medium. The cultures were incubated for 24 h with orbital shaking at 100 r.p.m., at 30ºC or 37ºC. Bacterial growth curves were evaluated under these conditions and cell pellets from 1 millilitre of culture were collected for RNA extraction.

The isolation of total RNA from bacterial cells was done using the commercial kit NucleoSpin RNA Plant (Macherey-Nagel) with minor modifications. Cell pellets were initially treated in BirnBoim A buffer solution [20% sucrose, 10mM Tris-HCl (pH=8.1), 10mM EDTA, 50mM NaCl], containing lysozyme and incubation of thirty minutes at 37ºC.

Further, the RNA was extracted from the lysates using TRIzol reagent as recommended by the manufacturer (Sigma), adding 10 µl of proteinase K (10 mg/ml) and incubation of 1 hour at 60ºC.

The RNA concentration was estimated using a Nanodrop NanoVue Plus. The integrity and purity of the RNA samples was derived from the ratio of absorbance at 260/280 nm and 260/230 nm and in agarose gel electrophoresis. In all RNA samples, the 260/280 ratio fluctuated between 2.0-2.2 and the 260/230 ratio was higher than 2, indicating that the samples were homogenous.

Reverse Transcription-PCR analysis

The DNA-free RNA was obtained from cultures of B. *amyloliquefaciens* CECT 8237 grown in medium M for 12 hours at 37ºC. The RNA concentration was estimated using a Nanodrop NanoVue Plus. The isolated RNA (3 µg) was converted into cDNA using the SuperScript III reverse transcriptase following the manufacturer's instructions (Invitrogen).

Pair of primers that amplified regions located within and between genes related to the lantibiotic production and processing (Table 2) were designed using the software OligoAnalyzer 3.1 (http://eu.idtdna.com/calc/analyzer) with the next parameters: i) a GC content of 50% approximately, ii) a length of the amplicons around 18 bp, iii) the change in the Gibbs energy (ΔG) for hairpin above -0.5, tending to be more than +1, iv) ΔG for self-dimer higher than -5Kcal/moles preferably positive values and v) ΔG for hetero-dimer more than -5Kcal/moles tending to positive values.

The RT-PCR was performed using the Go Taq Flexi DNA polymerase with 100 ng of cDNA in a final reaction volume of 25 µl (Promega). The RT-PCR programme included: 94ºC during 2 min, followed by PCR amplification using a 35-cycle amplification programme (94ºC for 1 min, 54ºC for 1 min, and 72ºC for 30sec) and a final extension

cycle at 72ºC for 7 min. Positive control reactions were included for each pair of primers containing genomic DNA of *B. amyloliquefaciens* CECT 8237.

Quantitative real-time (qRT)-PCR

Quantitative real-time (qRT)-PCR was performed using the iCycler-iQ system and iQ SYBR Green Supermix kit from Bio-Rad. The primer pairs to amplify these target genes were designed using the software Primer3 (http://bioinfo.ut.ee/primer3/) and Beacon designer (http://www.premierbiosoft.com/qOligo/Oligo.jsp?PID=1) keeping the parameters described elsewhere (Thornton and Basu, 2011) (Table 2). For qRT-PCR assays, the RNA concentration was adjusted to 100 ng/µl. Then, 1 µg of DNA-free total RNA was retro-transcribed to cDNA using the SuperScript III reverse transcriptase in a final reaction volume of 20 µl, according to the manufacturer's instructions (Invitrogen). The qRT-PCR cycle was performed at 95ºC during 3 min, followed by PCR amplification using a 40-cycle amplification programme (95ºC for 20sec, 56ºC for 30sec, and 72ºC for 30sec), a third step of 95ºC during 30sec and a final 40-cycle at 75ºC for 15sec, whose temperature is increasing 0,5ºC each cycle until the 40 cycles have been completed to evaluate the melting-curve.

This analysis was extended to the following experimental conditions:

- Growth kinetics of *B. amyloliquefaciens* CECT 8237 in liquid cultures in agitation and incubation at 30ºC or 37ºC.
- Biofilms of *B. amyloliquefaciens* CECT 8237 in static cultures in medium M and incubation at 30ºC or 37ºC.
- Melon leaves spread with cells suspensions of *B. amyloliquefaciens* CECT 8237 or *B. subtilis* subsp. *subtilis* 3610 (7 or 14 days post inoculation).

In order to normalize the data, the *rpsJ* gene encoding the 30S ribosomal protein S10, was used as a reference gene (Leães et al., 2016). The target genes, *lanA1* and *lanA2*, encoding the two structural genes that participate in the production of the lantibiotic, were amplified using the pairs of primers lanA1-for and lanA1-rev, lanA2-for and lanA2-rev, generating transcripts of 78bp or 126bp respectively (Table 2).

In biofilm growth conditions the following additional target genes were included: *srfAA*, encoding surfactin synthetase A; *ituA*, encoding iturin synthetase A.

The primers specificity was determined by qRT-PCR to be optimal for all pair of primers (approximately 100%). It was considered as a template a linear increase of DNA concentration which should be corresponded with a linear increase of transcript abundance (Matas et al., 2013). The relative transcript abundance was estimated using the ΔΔ cyclethreshold (Ct) method (Livak and Schmittgen, 2001). As mentioned above, transcriptional data associated to the target genes were normalized in relation to the gene *rpsJ* under each experimental condition. Thus, data are shown as the fold change in expression in a specific condition compared with another.

The relative expression ratio was calculated as the difference between qPCR threshold cycles (Ct) in the *rpsJ* gene and the Ct of the target gene (ΔCt= Ct*rpsJ* – Ct*lanA1*). Fold-change values were evaluated as 2^ΔCt, considering that one PCR cycle represents a twofold difference in template abundance (Matas et al., 2013). The qRT-PCR analyses were performed three times (technical replicates) using three independent RNA isolations (biological replicates).

Table 2. Pairs of primers specifically designed for RT-PCR and qRT-PCR assays.

Primers pairs	Primers	Sequence (5'-3')	Amplification region	Amplicon (bp)
RT-PCR Within genes				
1	lanA1-for	GTGGAAAAACCCTGCTTT	gene *lanA1*	184
	lanA1-rev	ACAGCTACTTTGACATTC		
2	lanM1-for	CACGGTAACAGTGGTATC	gene *lanM1*	508
	lanM1-rev	GTAATCCCATCCCTATAC		
3	lanA2-for	GCTATCTATCGTGATGTA	gene *lanA2*	164
	lanA2-rev	GCTGCTGTGAGACAAGTA		
4	lanM2-for	GTGAACGAAAAAGGGTCT	gene *lanM2*	364
	lanM2-rev	CTGAGCAGACAATATCCG		
5	lanT(P)-for	TACAGGAAGTATCACACT	gene *lanT(P)*	501
	lanT(P)-rev	TCAGGATTATGGAGAGTA		
6	lanHp-for	CGGAAATAGTATAGGAGG	gene *lanHP*	208
	lanHp-rev	TGTCAACTCTTGGTACAC		
7	lanF-for	TGGCAGACAGGAAAGTCG	gene *lanF*	407
	lanF-rev	TTCTTCAATCATTAGACC		
8	lanG-for	GCTTCCTGTTTCACATTG	gene *lanG*	419
	lanG-rev	TTCGTGACAGAGATACCA		
9	lanE-for	GCTTTTGTGGACTCAGAT	gene *lanE*	322
	lanE-rev	GCCCGCATTCCCTCCTAT		
10	lanR-for	CGAAAAAGTTGACGAATC	gene *lanR*	202
	lanR-rev	TCCTCATTATATTGGAAC		
11	lanH-for	CAAATCATATTCGCTTGG	gene *lanH*	418
	lanH-rev	TCTATCTCTATCTGCTTC		
12	lanI-for	TACGCCACATTCAAAGTG	gene *lanI*	248
	lanI-rev	TACCTACAGAGATATAAG		
13	lanJ-for	CTTATGATAGGGTTACTG	gene *lanJ*	248
	lanJ-rev	GGATATAACATAGGAATC		
14	lanP-for	GGACAGATGAAAGGAATC	gene *lanN*	421
	lanP-rev	GACCAACAGAGGCTACAG		
Between genes				
15	lanSA1-for	AGACCTCCTATTAACTGG	gene *S-A1*	394
	lanSA1-rev	GATGTCGTACCAAGCACA		
16	lanIA1-for	TGTAGAAAATCGGCTGTT	gene *I-A1*	257
	lanIA1-rev	GATGTCGTACCAAGCACA		
17	lanA1M1-for	TGTGCTTGGTACGACATC	gene *lanA1-M1*	364
	lanA1M1-rev	CAGATTCCAGATCCTTAC		
18	lanM1A2-for	ATGGTGACTTTGGACAGT	gene *lanM1-A2*	312
	lanM1A2-rev	TACATCACGATAGATAGC		
19	lanA2M2-for	TACTTGTCTCACAGCAGC	gene *lanA2-M2*	512
	lanA2M2-rev	CTCGTTCATTTAGGTAAG		
20	lanM2T-for	GAAGGAGGAATAGAAAGC	gene *lanM2-T*	381
	lanM2T-rev	CCATACAACAAAGGGCAC		
21	lanTHp-for	CAACGGTGATGAATTGTG	gene *lanT-Hp*	393
	lanTHp-rev	ACATAAAACATAAAGCAG		
22	lanHpF-for	GTTTGGAACCATCTATCT	gene *lanHp-F*	268
	lanHpF-rev	TGAGTCCGAGCAACATAC		
23	lanFG-for	GAAGAAAGAACGGATAAG	gene *lanF-G*	564
	lanFG-rev	TCAGTGTACTCCCAAATG		

24	lanGE-for	GGTTTGGTCTTCGGGCTG	gene *lanG-E*	559
	lanGE-rev	ACTCCTAAAATATGTTGC		
25	lanER-for	ATAGGAGGGAATGCGGGC	gene *lanE-R*	296
	lanER-rev	AGTGAAGGGGTGTATTTC		
26	lanRH-for	CCCAAGCACTTAGTTGTT	gene *lanR-H*	351
	lanRH-rev	GATGTTCATAGTAGGCAG		
27	lanHI-for	GGATCAGAAGCAGATAGAG	gene *lanH-I*	411
	lanHI-rev	GCGTAGTAAGCAAAGTCG		
28	lanIJ-for	GTTGTCATTGGGATATGG	gene *lanI-J*	397
	lanIJ-rev	CTGAAAAAGAATAGTACC		
29	lanJP-for	GGCTCTTCCTCTTTATGG	gene *lanJ-P*	512
	lanJP-rev	TTCATTGTAACTATCTAC		
30	lanPR2-for	TGATAAATACGATTCCAA	gene *lanP-R2*	381
	lanPR2-rev	GCGATGCGTTTGATTGAT		

Primers pairs	**Primers**	**Sequence (5'-3')**	**Amplification region**	**Amplicon (bp)**
qRT-PCR				
1	lanA1-for	TGCTTGGTACGACATCAGTTGC	*gene lanA1*	78
	lanA1-rev	CAGCTACTTTGACATTCCACAG		
2	lanA2-for	GGGAGGTCAACGAGGAAGAA	*gene lanA2*	126
	lanA2-rev	CGGTAGTTGGGCAAATAGC		
3	srfAA-for	AAGGAAACATCGTCACACAT	*gene srfAA*	145
	srfAA-rev	TTTAACAGCGAACCGAACAT		
4	ituA-for	TGGCAGCAATCGGACCTTCA	*gene ituA*	174
	ituA-rev	CTGTTTCAAGTGTTCGGGTA		
5	rpsJ-for	TCTGGTCCGATTCCGTTGCC	*gene rpsJ*	142
	rpsJ-rev	CAGTTTGTGGTGTTGGGTTCA		

Detection of volatile compounds and secondary metabolites

The detection of acetoin was conducted using the Voges-Proskauer test (Nicholson, 2008). The bacterial strains were grown in LB supplemented with 1% glucose at 37°C with slight agitation. After the addition of reagents, the culture turned red to indicate the production of acetoin. The production of 2,3-butanediol was followed using the protocol previously described (Nicholson, 2008). Briefly, cells were grown in the same conditions described above. After 24 h of growth, the cells were removed by centrifugation and filtration, and 30 µl of the supernatants were developed in thin layer chromatography pre-coated silica gel plates using n-hexane-ethyl acetate (1:5) as the mobile phase. The presence of 2,3-butanediol was revealed with a Cerium-ammonium-molybdate, CAM solution (40 g ammonium pentamolybdate, 1.6 g cerium (IV) sulphate, and 800 ml of

diluted sulphuric acid (1:9, with water, v/v)) and incubated in an oven at 150°C. A standard of 2,3-butanediol migrated at an Rf of 0.5 under these conditions. The identification and quantification of 2,3-butanediol were further performed on a Thermo Scientific GC Ultra coupled to a DSQ simple quadrupole. The gas chromatograph is equipped with a DB 5 column (15 m × 25 mm, at a film thickness of 0.25 µm). The operating conditions included a helium flow rate of 1.2 ml/min and an injection port temperature of 250°C. The temperature of the column was programmed for 1 min isothermal at 4°C, increased by 10°C/min for 2 min and 25°C/ min for 2 min.

For the detection of secondary metabolites, CECT 8237 and CECT 8238 were grown in Landy broth or GA medium (Chen et al., 2009). The cultures in Landy medium were grown 12 h or 40 h at 30°C and 150 rpm. After centrifugation for 40 min at 11,000 rpm and 4°C, the supernatants were extracted by solid phase extraction in a Merck LiChrolut® RP-18 cartridge. After binding and subsequent washing steps with MilliQ water (5 bed volume), the metabolites were eluted with methanol (2 bed volume), dried under a vacuum and resuspended in 100 µL of methanol. The cultures in GA broth were grown at 29°C and 165 rpm for 24 h, and the supernatants were collected after centrifugation for 40 min at 11,000 rpm. The samples were extracted three times with ethyl acetate, evaporated and dissolved in 100 µL of methanol. The resulting samples were analysed by reverse-phase high pressure liquid chromatography (Dionex 3000, Thermo C18, 50 × 2.1 mm, 2.6 µm, Accucore RP-MS) coupled to a ESI-MS (Orbitrap; Q-Exactive, Thermo) in positive mode. The temperature in the HPLC was held constant at 30°C during the experiment. The run was performed with a flow rate of 0.2 ml/min and molecules were eluted in a binary solvent system (solvent A: CH_3CN; solvent B: water-0.1% formic acid) as follows: 80 % B for 1 min, followed by a 7-min gradient from 80 % B to 10 % B for 15 min and a subsequent 5-min gradient from 10 % B to 80 % B for 3 min.

Production, extraction and detection of the two-peptide lantibiotic lichenicidin

For the detection of the lichenicidin-like lantibiotic, the strains *B. amyloliquefaciens* CECT 8237, hypothetical producer of a lichenicidin-like antibiotic, *B. licheniformis* DSM13, the producer strain of lichenicidin, or *B. amyloliquefaciens* CECT 8238, lacking the genomic region dedicated to the production of this lantibiotic, were grown in medium M, previously reported as optimal for the synthesis of lichenicidin (Caetano et al., 2011). All these strains were cultured in 50 ml of medium M at 37ºC with aeration during 48 hours. Then, cell pellets were collected by centrifugation at 10000g and 4ºC for 30 minutes. After that, the cells were washed with 35 ml 70% isopropanol (adjusted to pH 2 with HCl) and maintained with shaking at 4ºC during four hours. Cells were removed by centrifugation and supernatant sterilized by filtration was stored at -20ºC. For further analysis, the samples were concentrated in a vacuum to a final volume of 2 ml, and amended with 2 µl of trifluoroacetic acid (0,1%).

The peptide mass fingerprinting was determined by matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry using a 4700 Proteomics analyzer (ABSCIEX). Each sample, corresponding to washed cells of CECT 8237, DSM13 and CECT 8238, was directly mixed with an equal volume of α-cyanohydroxycinnamic acid solution (CHCA) and spotted onto the MALDI plate. The analyser was used in the reflector positive ion mode at 20 kV Source 1 acceleration voltage.

The fragmentation of 3042 Da peak was carried out in two modes: CID OFF, post source decay (PSD) fragmentation type, in which fragments with higher mass are obtained, and CID ON (collision induced dissociation), that generates a larger fragmentation correlated with peaks of lower mass. The TOF-TOF post source decay (PSD) fragmentation was obtained at 8kV Source 1 acceleration voltage. The analysis of the mass spectrum was performed using the Data Explorer software.

In situ* detection of lichenicidin-like lantibiotic in supernatants and biofilms of *B. amyloliquefaciens

Bacterial colonies were grown on standard LB agar plates and used for the optimization of the efficient detection of the diverse secondary metabolites, produced by *B. amyloliquefaciens* CECT 8237, in a MALDI-TOF-TOF ultraflex analyser. For that purpose, several matrix solutions were tested including α-cyanohydroxycinnamic acid solution (CHCA), sinapinic acid (SPA) or 2,5-dihydroxy benzoic acid (DHB), DHB-SPA or DHB-CHCA mixtures.

For the detection of lantibiotic in cultures of CECT 8237 grown in medium M with agitation conditions, the crude samples of cells and supernatants were directly analysed. The supernatants were mixed with an equal volume of α-cyanohydroxycinnamic acid solution-2,5-dihydroxy benzoic acid (CHCA-DHB) and spotted onto the MALDI plate. A similar procedure was used with bacterial cells, but using diluted bacterial suspensions.

For the specific detection of the lantibiotic in standing conditions, pellicle formation assays were done as described above. Microtiter plates of 24 wells containing 1 ml of medium M were inoculated with 10 µl of bacterial suspension to a final OD of 0.1, and incubated without agitation at 30ºC and 37ºC for 24-72 h. The pellicle and the spent medium were separated and treated to generate two biochemically distinct fractions, cells and supernatant, which were processed as described earlier for mass spectrometry analysis. To analyse these fractions, as the bacterial concentration was different, the cell pellets were resuspended in 75 µl or 750 µl of distilled water containing trifluoroacetic acid 0,1%. Then, 2 µl of this suspension was mixed with an equal volume of matrix and 0,75 µl of this mixture spotted onto the MALDI plate. In parallel, the detection of the lantibiotic was also studied in bacterial colonies grown in biofilm inducible medium M or MSgg agar at 30°C or 37ºC during 72 h as described above. Two independent colonies formed at 72 h at

30 or 37ºC were analysed in three differentiable areas: core, middle and outmost. Four replicates per sample were evaluated. To avoid contamination between samples, the matrix, composed by DHB-CHCA mixture, was firstly placed on the MALDI-plate. Then, the selected area within the colony was directly touched with a thin pipette tip and transferred to the analytical matrix.

Promotion of root growth

Melon seedlings were used to evaluate the potential of bacterial strains to promote root growth (García-Gutiérrez et al., 2012). The strains were grown at 28°C for 12 h. The cells were pelleted by centrifugation, were washed in water twice and resuspended in water to a final cell density of 10^8 CFU/ml. The melon seed were then soaked in the bacterial suspensions for 1 h and incubated in a growth chamber for 7 days at 25°C and a 16-h light photoperiod. The roots were split and the weight measured in a precision balance.

Detection of the two-peptide lantibiotic in plant

The same biocontrol bioassay with melon leaves was used to study the production of the lantibiotic and other antibiotics by *Bacillus amyloliquefaciens* CECT 8237 in plant. Melon leaves from mature plants were inoculated with 1 ml of cell suspensions of *Bacillus* strains (10^8 CFU/ml): *Bacillus subtilis* subsp. *subtilis* 3610 or *Bacillus amyloliquefaciens* CECT 8237 separately. After 7 and 14 days post-inoculation, four leaves for each treatment were collected. The first leave was used to count the number of cells associated to leaves. For these bacterial counts, melon leaves, previously weighed, were homogenized in a stomacher during three minutes at maximum speed (BagMixer – www.interscience.fr). The second, third and fourth leaves were grinded separately in a mortar containing liquid

nitrogen. Aliquots of 0,2g approximately were collected and stored at -80ºC for RNA isolation and qRT-PCR assays as described previously.

RESULTS

RESULTS

Bacillus spp. CECT 8237 and CECT 8238 show better biocontrol abilities in comparison with the type strains of *B. amyloliquefaciens/subtilis*

In previous works we demonstrated the biocontrol activity of the two strains CECT 8237 and CECT 8238, but never contrasted with other biocontrol strains (Romero et al., 2007b; Zeriouh et al., 2011). To initiate this study, we decided to compare the antagonistic activity of *Bacillus* spp. CECT 8237 or CECT 8238 and *B. amyloliquefaciens* subsp. *plantarum* FZB42 and that of *B. subtilis* subsp. *subtilis* 168 and NCIB3610, which are strains that produce different combinations of antimicrobials against *Pectobacterium carotovorum*, the bacterial pathogen responsible for soft rot disease, and *Podosphaera xanthii*, the fungal pathogen responsible for cucurbit powdery mildew disease. *B. amyloliquefaciens* subsp. *plantarum* FZB42 and the studied strains performed better than did any of the *B. subtilis* strains in both pathosystems (Figure 7 and Table 3). The protection provided by the CECT 8237 or CECT 8238 strains against fungal (68% and 73% reduction in severity, respectively) and bacterial diseases (61% and 53% reduction in severity, respectively) was slightly better than FZB42, but not significantly different (Table 3). The *B. subtilis* subsp. *subtilis* NCIB 3610 strain showed restricted biocontrol activity against both diseases, and *B. subtilis* subsp. *subtilis* 168 failed to control *P. carotovorum,* although it provided some protection against *P. xanthii* (a reduction of 39% of the symptoms). These findings seem to indicate the potential of these strains to produce antimicrobial compounds. Indeed, the *B. subtilis* strains included in this analysis do not produce the polyketides macrolactin and difficidin or lipopeptides of the iturin family, which are important bacterial factors in the control of these diseases (Zeriouh et al., 2011; Yuan et al., 2012). In addition, the strain *B. subtilis* subsp. *subtilis* 168 is not able to synthesise the lipopeptide surfactin due to a mutation of the phosphopantheinyl transferase *sfp*, which could explain its limited

capability to control both diseases compared with *B. subtilis* subsp. *subtilis* NCIB 3610 (Marahiel, 2009). In contrast, *B. amyloliquefaciens* subsp. *plantarum* FZB42 and the studied strains provided efficient control of diseases that could be related to the production of a wide variety of secondary metabolites.

Table 3. The severity of disease symptoms caused by *Pectobacterium carotovorum* subsp. *carotovorum* and *Podosphaera xanthii* on melon leaves after treatments with vegetative cells of *Bacillus subtilis*,*B. amyloliquefaciens* or the studied strains.

Treatments	*P. carotovorum* subsp. *carotovorum*		*P. xanthii*	
	Severity[y]	Reduction[z]	Severity	Reduction
Untreated	55 a	-	37 a	-
B. subtilis subsp. *subtilis* 168	49 ab	11	22 b	39
B. subtilis subsp. *subtilis* NCIB 3610	39 bc	29	19 b	47
B. amyloliquefaciens subsp. *plantarum* FZB42	30 cd	45	16 bc	55
Bacillus sp. CECT 8237	21 d	61	12 c	68
Bacillus sp. CECT 8238	26 d	53	10 c	73

[y] Disease severity was assessed for *P. carotovorum* as a percentage of necrotic tissue according to a 0 to 3 scale of symptoms and for *P. xanthii* as a percentage of leaf area covered by powdery mildew. Values followed by the same letter in each column are not significantly different at P=0.05, according to Fisher's least significant difference test.

[z] Percentage of disease reduction achieved by treatments in regard to values of disease severity on untreated leaves.

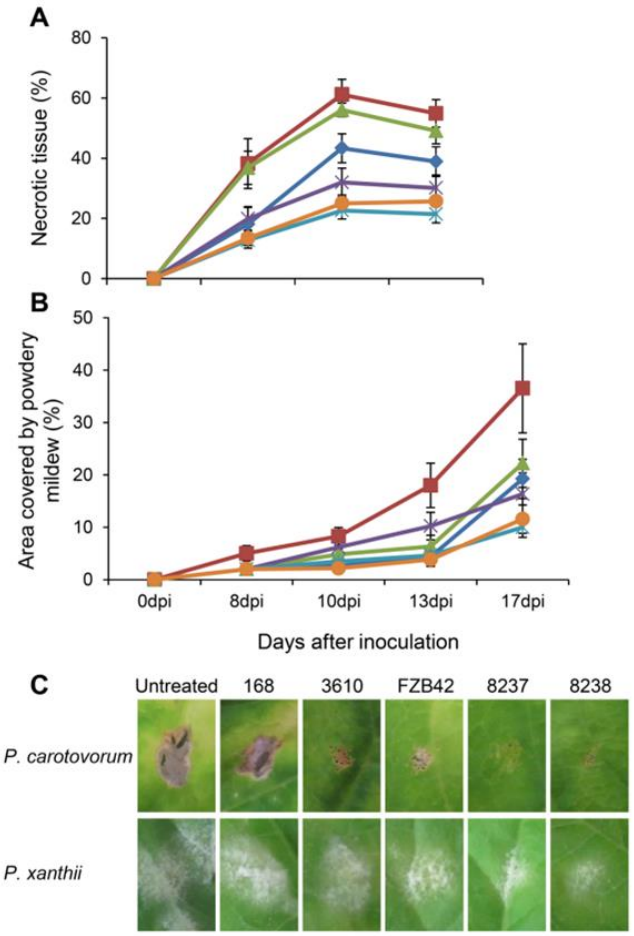


Figure 7. *Bacillus* spp. CECT 8237 and CECT 8238 are more efficient biocontrol agents than other *Bacillus* strains. The biocontrol traits of different strains were tested on melon leaves against: A, the plant pathogenic bacteria *Pectobacterium carotovorum* subsp. *carotovorum* and B, the powdery mildew fungus *Podosphaera xanthii*. The melon leaves were first treated with washed cells of the biocontrol agent and, later, were inoculated with the plant pathogens. The progress of the symptoms was evaluated as the percentage of necrotic leaf induced by *P. carotovorum* or the percentage of leaf area covered by powdery mildew. The treatments assayed were: untreated, non-treated control (red square); 168, *B. subtilis* subsp. *subtilis* 168 (green triangle); 3610, *B. subtilis* subsp. *subtilis* NCIB 3610 (blue diamond); FZB42, *B. amyloliquefaciens* subsp. *plantarum* FZB42 (purple X); 8237, *Bacillus* sp. CECT 8237 (blue X); and 8238, *Bacillus* sp. CECT 8238 (orange circle). C, Representative pictures of each treatment were taken 10 or 17 days after inoculation with *P. carotovorum* or *P. xanthii,* respectively.

Sequencing and assembly of the complete genome of *Bacillus* spp. CECT 8237 and CECT 8238

The previous results indicated that indeed, both strains should contain singularities compared to other Bacillus strains, relevant for their biocontrol skills. To investigate this, their chromosomal DNA were isolated and their genomes were sequenced and assembled at the Beijing Genomic Institute (BGI, Shenzhen, China). Preliminary data were obtained from the assembly of the bacterial genomes, named draft genomes (Table 4, top). The overlapping regions were grouped into adjacent DNA fragments called *contigs* which were ordered and clustered into *scaffolds*, roughly organized but not necessarily linked in a unique continuous fragment of sequence. Further, the local assembly and gap closure was performed with paired-end reads locating in gaps, and additionally PCR gap closure was carried out to generate the complete genome sequences (Table 4, bottom). The initial GC content was similar to that achieved after filling the remaining gaps and solving the misassembled regions. This parameter is generally conserved along the species and even partially along the genus. In fact, the genome GC content may vary between organisms from 17% to 75%. For these strains, the GC percentage, 46,3% and 46,5% in CECT 8237 and CECT 8238, is closer to *B. subtilis* or *B. amyloliquefaciens* which ranges from 44% to 48% respectively. The genome size, another characteristic commonly shared with the genus and species, ranged between 4,03 Mb to 4 Mb in CECT 8237 and CECT 8238 respectively, as other *Bacillus* species.

Table 4. Sequencing and assembling data corresponding to genome sequences of *Bacillus* spp. CECT 8237 and CECT 8238.

SEQUENCING DATA	DRAFT GENOME SEQUENCES			
	CECT 8237		CECT 8238	
	Scaffolds	***Contigs***	***Scaffolds***	***Contigs***
Total number	57	152	40	158
Total length (bp)	3,956.855	3,946.598	3,965.655	3,952.470
Maximum length (bp)	660.955	172.422	637.909	281.942
Minimum length (bp)	752	32	572	31
GC percentage	46,26	46,26	46,37	46,37
	COMPLETE SEQUENCES			
	Scaffolds	***Contigs***	***Scaffolds***	***Contigs***
Total number	1	1	1	1
Total length (bp)	4,034.636	4,034.636	4,005.145	4,005.145
Maximum length (bp)	4,034.636	4,034.636	4,005.145	4,005.145
Minimum length (bp)	4,034.636	4,034.636	4,005.145	4,005.145
GC percentage	46,34	46,34	46,49	46,49

Annotation of the complete genome sequences of *Bacillus* spp. CECT 8237 and CECT 8238

Both complete genome sequences were initially annotated with the *Rast* server attempting the estimation of their genetic content dedicated to diverse cellular functions (Figure 8 and Table 5). In general, *Bacillus* spp. CECT 8237 and CECT 8238 showed the same metabolic strategies in relation to the genetic content. The functional categories in which both strains dedicate more energy (around 14%) are associated with the metabolism of carbohydrates, amino acids and derivatives. In addition, a significant percentage of their genetic machinery, 8% approximately, is involved in the synthesis of cofactors, vitamins, prosthetic groups and pigments, being the majority of these compounds essential to participate in diverse metabolic pathways. We also found the genes related to proteins and

RNA metabolism (5-6% each), followed by the genes associated with fatty acids, lipids and isoprenoids and those involved in the synthesis of the cell wall (4-5%). Besides the essential categories mentioned above, we searched for genes possibly involved in the interactions with pathogens or plants. However, this general annotation derived from *Rast* server missed most of the information associated to those categories and available in the database, forcing to create specific *ad hoc* scripts, that in combination with the pre-existing tools permitted to deal with this misinformation (results are specified later on).

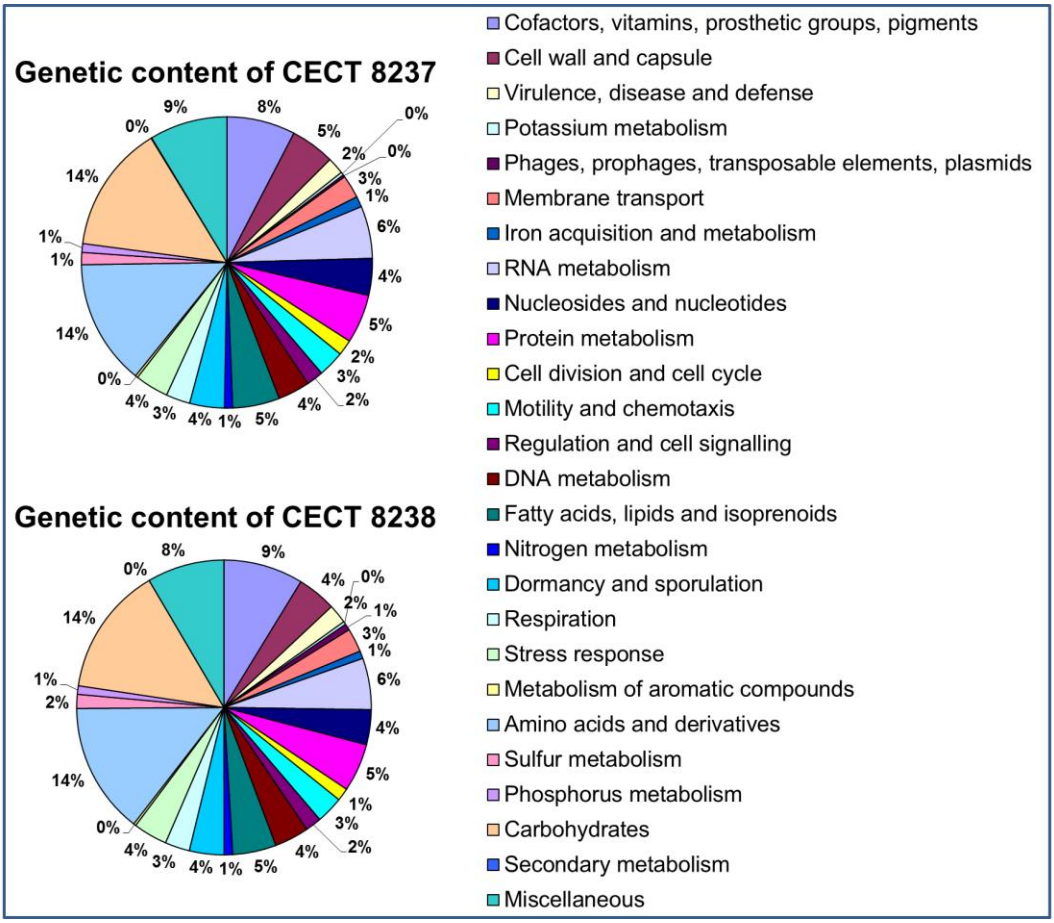


Figure 8. Genetic content of the complete genome sequences of *Bacillus* spp. CECT 8237 and CECT 8238. The functional categories indicated in the attached legend are corresponded with the automatic annotation of the *Rapid Annotation* using *Subsystem Technology* (Rast) server. The percentages indicate the genetic machinery contained in CECT 8237 (Up) and CECT 8238 (Down) genomes and dedicated to each category.

Table 5. Automatic annotation of the complete genomes of CECT 8237 and CECT 8238 from *Rapid Annotation using Subsystem Technology* (*Rast*) server. The main functional categories are divided (coloured differentially) into diverse specific subgroups. The number of genes contained at each category are indicated for both genome sequences.

CATEGORIES AND SUBCATEGORIES	CECT 8237		CECT 8238	
	Number of genes	% of the total	Number of genes	% of the total
Cofactors, Vitamins, Prosthetic Groups, Pigments	**219**	**7,61**	**242**	**8,11**
Biotin biosynthesis	36		38	
Thiamin biosynthesis	22		30	
Quinone cofactors	8		14	
Tetrapyrroles	18		18	
Riboflavin, FMN, FAD	29		31	
Pyridoxin biosynthesis and degradation pathway	10		9	
NAD and NADP	18		22	
Folate and pterines	56		58	
Lipoic acid	4		4	
Coenzyme A	18		18	
Cell Wall	**141**	**4,90**	**136**	**4,56**
Capsular and extracellular polysacchrides	46		43	
Gram-Positive cell wall components	29		25	
No subcategory	66		68	
Virulence, Disease and Defence	**54**	**1,88**	**63**	**2,11**
Adhesion	1		1	
Bacteriocins, ribosomally synthesized antibacterial peptides	16		17	
Resistance to antibiotics and toxic compounds	37		36	
Invasion and intracellular resistance			9	
Potassium metabolism	**10**	**0,35**	**10**	**0,34**
No subcategory	10		10	
Phages, Prophages, Transposable elements	**7**	**0,24**	**21**	**0,70**
Phages, prophages	7		21	
Membrane Transport	**74**	**2,57**	**78**	**2,61**
ABC transporters	19		23	
Protein translocation across cytoplasmic membrane	41		38	
Uni- Sym- and Antiporters	2		2	
TRAP transporters	1		1	
No subcategory	11		14	
Iron acquisition and metabolism	**34**	**1,18**	**25**	**0,84**
Siderophores	16		5	
No subcategory	18		20	
RNA Metabolism	**167**	**5,81**	**169**	**5,67**
RNA processing and modification	131		132	
Transcription	36		37	
Nucleosides and Nucleotides	**118**	**4,10**	**115**	**3,86**
Pyrimidines	36		36	
Purines	63		59	

Detoxification	9		10	
No subcategory	10		10	
Protein Metabolism	**156**	**5,42**	**158**	**5,30**
Protein folding	12		13	
Protein biosynthesis	79		82	
Protein processing and modification	26		24	
Protein degradation	39		39	
Cell Division and Cell Cycle	**50**	**1,74**	**42**	**1,41**
No subcategory	50		42	
Motility and Chemotaxis	**84**	**2,92**	**89**	**2,98**
Flagellar motility in Prokaryota	72		76	
No subcategory	12		13	
Regulation and Cell signaling	**54**	**1,88**	**53**	**1,78**
Programmed cell death and toxin-antitoxin systems	19		17	
No subcategory	35		36	
DNA Metabolism	**102**	**3,55**	**113**	**3,79**
DNA repair	60		63	
DNA replication	27		31	
DNA uptake, competence	6		6	
No subcategory	9		13	
Fatty Acids, Lipids, and Isoprenoids	**150**	**5,22**	**143**	**4,79**
Phospholipids	31		31	
Fatty acids	83		76	
Isoprenoids	19		19	
No subcategory	17		17	
Nitrogen Metabolism	**26**	**0,90**	**28**	**0,94**
No subcategory	26		28	
Dormancy and Sporulation	**111**	**3,86**	**114**	**3,82**
Spore DNA protection	5		6	
No subcategory	106		108	
Respiration	**77**	**2,68**	**81**	**2,72**
ATP synthases	8		9	
Electron accepting reactions	34		34	
Electron donating reactions	16		17	
No subcategory	19		21	
Stress Response	**107**	**3,72**	**112**	**3,75**
Osmotic stress	12		12	
Oxidative stress	42		42	
Cold shock	1		1	
Heat shock	16		18	
Detoxification	10		12	
Periplasmic stress	1			
No subcategory	25		27	
Metabolism of Aromatic Compounds	**8**	**0,28**	**8**	**0,27**
Peripheral pathways for catabolism of aromatic compounds	3		3	
Metabolism of central aromatic intermediates	2		2	
No subcategory	3		3	

Amino Acids and Derivatives	**401**	**13,94**	**431**	**14,45**
Glutamine, glutamate, aspartate, asparagine; ammonia assimilation	42		42	
Histidine Metabolism	16		16	
Arginine; urea cycle, polyamines	46		59	
Lysine, threonine, methionine, and cysteine	114		128	
Branched-chain amino acids	80		81	
Aromatic amino acids and derivatives	47		47	
Proline and 4-hydroxyproline	17		18	
Alanine, serine, and glycine	39		40	
Sulfur Metabolism	**39**	**1,36**	**46**	**1,54**
Inorganic sulfur assimilation	9		14	
Organic sulfur assimilation	21		23	
No subcategory	9		9	
Phosphorus Metabolism	**28**	**0,97**	**28**	**0,94**
No subcategory	28		28	
Carbohydrates	**408**	**14,19**	**424**	**14,21**
Central carbohydrate metabolism	122		123	
Aminosugars	7		7	
Di- and oligosaccharides	66		69	
One-carbon Metabolism	31		32	
Organic acids	25		24	
Fermentation	45		49	
Sugar alcohols	36		38	
Polysaccharides			1	
Monosaccharides	74		79	
No subcategory	2		2	
Secondary Metabolism	**2**	**0,07**		
No subcategory	2			
Miscellaneous (other functions)	**249**	**8,66**	**254**	**8,51**

The strains CECT 8237 and CECT 8238 are identified as *Bacillus amyloliquefaciens*

We further did an accurate multi-locus phylogeny analysis to classify both strains within the *Bacillus* genus. In this study, we considered i) housekeeping genes, which are widely used in bacterial characterization studies due to their conserved distribution along the species and the sequence variation between bacterial populations, ii) sigma factor coding genes, which permitted us to distinguish members from the same genus, species or even the same group (Schmidt et al., 2011), and iii) the gene *spoVG*, which is a representative of the developmental programme that leads to endospore formation in the *Bacillus* genus. We included diverse strains of *B. subtilis* and the closely related *B. amyloliquefaciens*, members of the *Bacillus cereus* group, as the most distantly related strains, and *Clostridium*, as the external Gram-positive and sporulating species that is unrelated to the *Bacillus* group (Figure 9A). The CECT 8237 and CECT 8238 strains clustered with the *B. amyloliquefaciens* group (Figure 9A), which was an unexpected finding and contrary to their initial identification as *B. subtilis*. However, this was not unprecedented because *B. amyloliquefaciens* and *B. subtilis* are phenotypically and genetically related species (Cawoy et al., 2015).

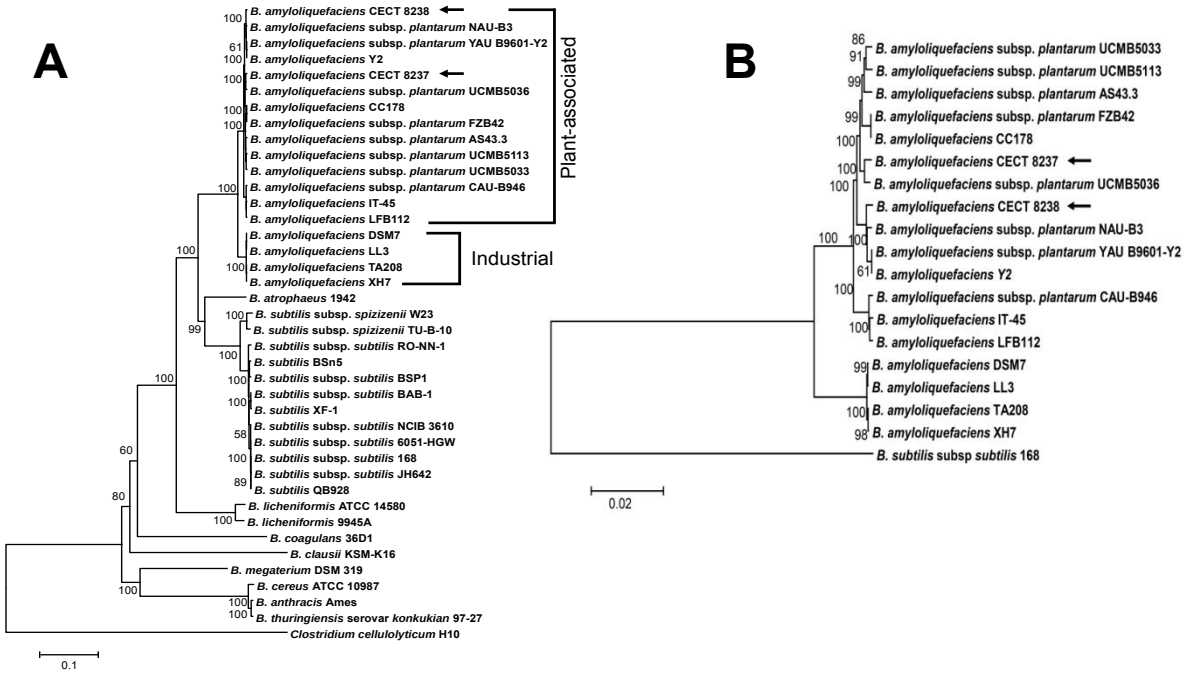


Figure 9. The isolates CECT 8237 and CECT 8238 cluster with the group of plant-associated *Bacillus amyloliquefaciens* subsp *plantarum*. A. Phylogenetic analysis of the isolates CECT 8237 and CECT 8238, and the closely related species *B. amyloliquefaciens*, *B. subtilis*, *B. atrophaeus* and *B. licheniformis*, and other representatives of the *Bacillus* genus such as *B. cereus*, *B. thuringiensis*, *B. coagulans*, *B. megaterium* and *B. anthracis*. The *Clostridium cellulolyticum* H10 was used as an out-group. B. Phylogenetic analysis of the isolates CECT 8237 and CECT 8238 and the plant-associated *B. amyloliquefaciens* strains. The *B. subtilis* subsp. *subtilis* 168 was used as an out-group. In both analyses, eleven concatenated genes (*nusA*, *rpoA*, *dnaA*, *rpoB*, *gyrA*, *gyrB*, *rpoC*, *spoVG*, *sigW*, *sigH* and *sigB* genes) were handled to build a Neighbour-joining tree using MEGA 5 bootstrap values (10,000 repetitions), which are shown on the branches. The topology was identical for the trees produced by the minimum evolution and maximum parsimony methods. The sequences from all of the strains were extracted from published genome sequences. Arrows indicate the location of the CECT 8237 and CECT 8238 strains in the tree.

An additional phylogenetic tree was generated, including *B. amyloliquefaciens* strains and the type strain *B. subtilis* subsp. *subtilis* 168 as the out-group, in order to identify in more details the taxonomic situation of the studied strains within the *B. amyloliquefaciens* species (Figure 9B).

Plant-associated *B. amyloliquefaciens* strains gain skills in competition and environmental adaptability

A previous genomic study proposed that *B. amyloliquefaciens* strains could be sorted into two distinct clades: (i) the clade of the plant-associated strains and (ii) the clade that is associated not with plants but with industrial applicability (Borriss et al., 2011). In our analysis, which included more genome sequences than in the previous study, we found this clear separation of the two clades. The CECT 8237 and CECT 8238 strains appeared to be more closely related to the strains of the plant-associated clade (Figure 9A-B). However, it was noteworthy that both strains evolved in different branches to the strain FZB42, a paradigm of the group of agriculturally relevant *B. amyloliquefaciens* strains (Borriss et al., 2011). This separation between plant-associated and industrially relevant strains would tentatively reflect the divergent evolution of these strains imposed by the habitat they actually occupy, and thus we should ideally find genomic differences. In general, the genome sizes of plant-associated strains were larger than industrial strains, but this variation in size did not correlate with an increase in the number of putative open reading frames or predicted proteins (Table 6).

Table 6. A comparison of the genomic features of *Bacillus amyloliquefaciens* CECT 8237 and CECT 8238 with other plant-associated or industrially relevant *Bacillus amyloliquefaciens* strains.

Strains	**Features**						
	Genome size (bp)	Plasmid size (bp)	G+C content (%)	Number of genes	Protein-coding sequences (CDSs)	tRNAs	rRNAs
Plant-associated							
CECT 8237	4,034,636	-	46,34	4059	3918	82	27
CECT 8238	4,005,145	-	46,49	4049	3894	82	27
AS43.3	3,961,368	-	46,59	3979	3861	89	29
CAU-B946	4,019,861	-	46,51	3948	3823	95	30
CC178	3,916,828	-	46,5	4074	3950	86	27
FZB42	3,918,589	-	46,48	3813	3693	89	30
IT-45	3,928,857	8009	46,62/40,47[a]	3976/9[a]	3851/9[a]	95	30
LFB112	3,942,754	-	46,7	4023	3859	94	30
NAU-B3	4,196,170	8438	46/40,3[a]	4123/5[a]	4001/5[a]	92	30
UCMB5033	4,071,167	-	46,2	4095	3877	86	30
UCMB5036	3,910,324	-	46,6	3842	3636	89	29
UCMB5113	3,889,532	-	46,7	3854	3656	89	29
YAU B9601-Y2	4,242,774	-	45,85	4110	3981	91	30
Y2	4,238,624	-	45,85	4348	4234	85	29
Industrial							
DSM7	3,980,199	-	46,1	4043	3892	94	30
LL3	3,995,227	6758	45,7/42[a]	4346/9[a]	4219/9[a]	72	22
TA208	3,937,511	-	45,8	4177	4089	70	18
XH7	3,939,203	-	45,8	4286	4190	75	21

a. Data from the chromosome sequence (first column) and plasmid sequence (second column).

Thus, the differences between these clades should be noted in the functionalities of their predicted proteins. To test this idea, we first defined the set of genes contained in all members of each clade (in other words, their pan-genomes) and used this information to search specifically for the genes exclusive to each clade. All plant-associated and industrially relevant *B. amyloliquefaciens* strains included in this analysis are listed in Table 6. In the first step of this analysis, we found that 5453 coding sequences (CDS) formed the pan-genome of clade I (plant-associated strains) and that 4753 CDS corresponded to clade II (industrial strains). It was remarkable that 1365 CDS appeared exclusively in any strain of clade I, though only 612 CDS were present in at least one industrial strain but absent in all members of the other clade. From these CDS, we selected those (16 CDS) found in all plant-associated strains (Annex Table 1) and built a phylogenetic tree (Annex Figure 1), which provided a similar distribution of strains as previously seen (Figure 9); the FZB42 strain in separated branches from CECT 8237 or CECT 8238. Then, these exclusive CDS were classified into different clusters of orthologous groups (COG) categories. In order to analyse whether CDS only present in one set of strains were enriched in any COG categories with respect to the other set, an enrichment test of the computed COG categories was performed. After the application of hypergeometric tests, we found that these exclusive genes were not significantly enriched in genes belonging to any COG class.

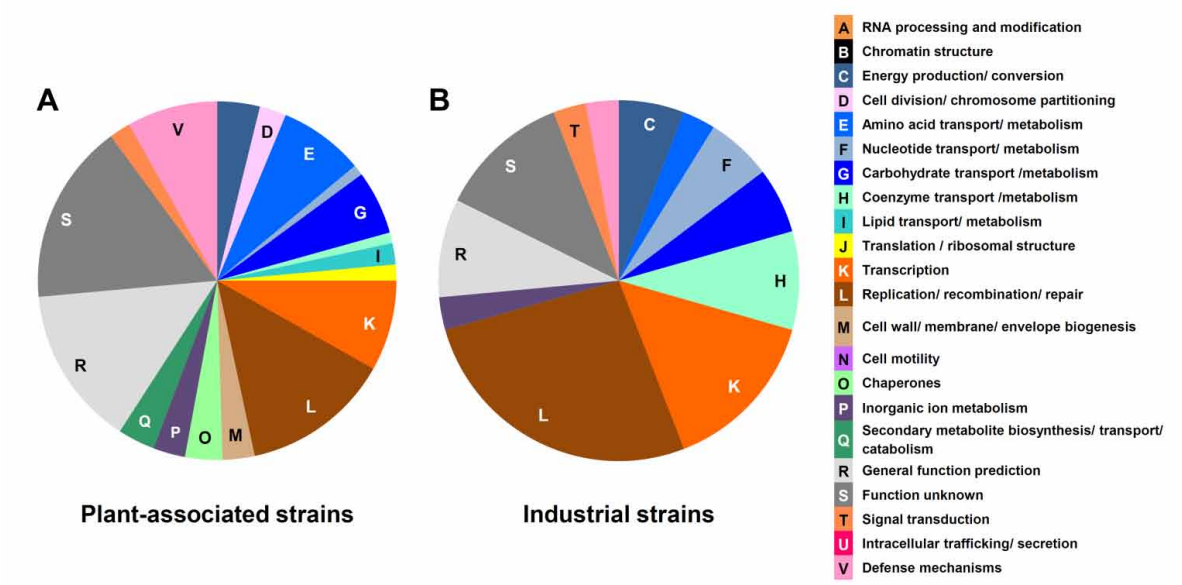


Figure 10. The classification of genes into clusters of orthologous groups (COG) reveals the differences between plant-associated and industrially relevant *Bacillus amyloliquefaciens* strains. Circular charts symbolize the percentages of genes classified into the different COG specifically detected in **A**, *B. amyloliquefaciens* strains associated with plants or **B**, industrially relevant strains. A colour code is specified in the figure legend.

However, we focused on those genes that showed homology to any CDS in the database and permitted their classification into different COG categories, 157 and 28 genes in plant-associated and industrial strains respectively (Annex Table 2). The relative proportion of COG categories represented in one set of strains against the other was plotted in sector graphs (Figure 10). We found that the unique CDS of clade I (plant-associated strains) primarily sorted into the following categories: metabolism of amino acids (E), carbohydrates (G) or lipids (I), synthesis of secondary metabolites (Q), general function prediction (R) or defence mechanisms (V) (Figure 10A). Contrary to this finding, the unique CDS of the clade II (industrial strains) appeared dispersed among the different categories, although some were overrepresented, such as metabolism of coenzymes (H) and

replication, recombination and repair (L) (Figure 10B). In view of these observations, it could be said that the *B. amyloliquefaciens* strains associated with plants possess a specific core of genetic features, which would predict their better performance in the context of plant health than those strains not associated with plants. (The relatedness of these strains is also reflected in Annex Figure 1, where CECT 8238, NAU-B3, YAU B9601-Y2 and Y2 strains clustered together as a result of the alignment corresponding to genes distinctive of plant-associated strains).

B. amyloliquefaciens CECT 8237 and CECT 8238 have acquired gene clusters that support their biocontrol skills

Once we defined the core genes distinctive to strains associated with plants, we investigated whether the CECT 8237 or CECT 8238 strains could bear unique and differentiable features to the other members of the same clade. To answer this question, we compared their genomes with those of other plant-associated *B. amyloliquefaciens* strains (Table 6) that were available in the database, using the MAUVE genome alignment software (Figure 11, outer circle). Specifically, the level of conservation along the chromosomal regions is represented by different tones of blue: darker means highly conserved and softer, which indicates poorly conserved sequences. Despite the level of conservation of their genome sequences, we could identify a number of unique “windows” (white areas), which, by definition, were absent in the other *B. amyloliquefaciens* strains and, thus, a potential reservoir of distinctive bacterial features. Separately, we performed a bioinformatics analysis, based on the methodology shown in Figure 1, to compare the studied strains with 76 *Bacillus* complete genome sequences, including those of plant-associated and industrial *B. amyloliquefaciens* strains (Table 1, strains for comparative analysis). As a result, this analysis permitted us to screen for atypical regions (AR) that

were non-conserved along the *Bacillus* genus (Figure 11) and that were composed of consecutive genes hypothetically related to the same physiological function (Figure 11, green boxes). Using a minimum of seven consecutive genes as the screening window, we found three different scenarios.

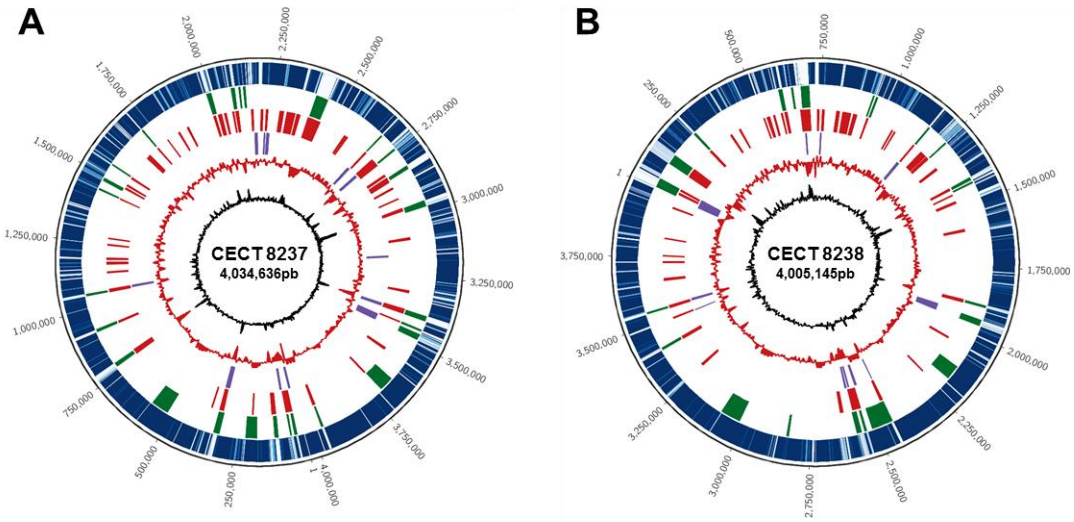


Figure 11. *Bacillus amyloliquefaciens* CECT 8237 and CECT 8238 possess unique genomic regions in comparison to other plant-associated *B. amyloliquefaciens* strains. The genomes of CECT 8237 (A) or CECT 8238 (B) strains were compared to those of other *Bacillus* species available in the database, and the results were organised in circles. The outer scale designates the genome's coordinates in base pairs (bp). The first circle shows the comparative genome analysis (MAUVE Genome Alignment software) with several *B. amyloliquefaciens* potential biological control agents: regions highly conserved are represented in dark blue; regions slightly conserved are represented in light blue; and specific regions detected only in CECT 8237 or CECT 8238 are represented in white. The second circle shows the non-conserved (green) regions of CECT 8237 and CECT 8238 genomes among the *Bacillus* genus (the algorithm used is described in the Material and Methods). The third circle (red) shows the regions of DNA acquired horizontally using Alien Hunter software. The fourth circle (purple) shows the predicted prophages (Prophage finder tool). The fifth circle (red) shows the percentage of G+C in relation to the average G+C content in a 2,000-bp window. The inner circle (black) shows the trinucleotide composition in relation to the average trinucleotide composition in a 2,000-bp window.

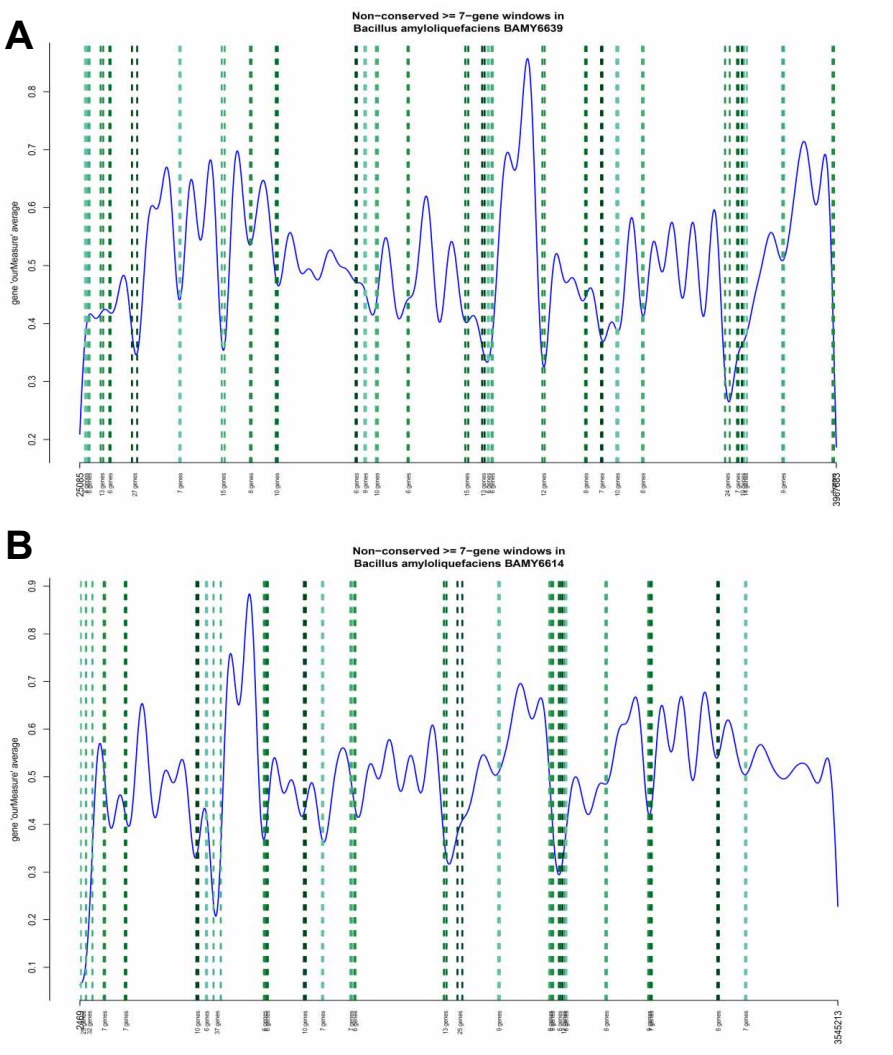
First, we detected putative gene clusters in both strains that coincided with highly conserved regions (Figure 11, outer line, dark blue boxes) and, thus, shared with most of the plant-associated *B. amyloliquefaciens* strains. Unsurprisingly, and as we previously noted, the analysis of these gene clusters revealed that they were involved in the biosynthesis of known secondary metabolites, such as fengycin, difficidin, macrolactin or bacillaene. Second, certain atypical regions overlapped with partially conserved regions (Figure 11, outer line, light blue boxes), indicating that some genes of these AR are shared with any of the *B. amyloliquefaciens* plant-associated strains. For example, some regions harbour genes related to fatty-acid metabolism, cell envelope biogenesis, vitamin biosynthesis proteins, and phage-related proteins. In the atypical regions of CECT 8238, we found genes related to peptide synthetase and thioesterase as well as genes dedicated to sporulation, ABC transporter proteins or N-acetyltransferases. Third, and even more interesting, we found 13 gene clusters coincident with unique genomic windows in CECT 8237 (Figure 11A, outer line, white boxes), indicating that they were absent in other *B. amyloliquefaciens* strains and poorly or not conserved in the *Bacillus* genus. A total of 28 atypical regions were detected in the *B. amyloliquefaciens* CECT 8237 genome, and 20 of them might be considered genomic islands according to the *Alien Hunter* analysis, local deviations in the trinucleotide usage patterns and variation in the percent GC content (Figure 12A and Annex Table 3). These data are suggestive of genetic acquisition after successive events of horizontal transfer. According to this idea, four of these genomic islands also overlapped with putative prophages, as predicted by the Prophage finder algorithm, and two other atypical regions contained genes related to phage proteins or mobile element domains as resolvases and transposases. The size of the genomic islands ranged from 4 to 24 Kb, and most of the genes encoded hypothetical proteins with unknown functionality. In addition, we identified other genes that are tentatively associated

with the interaction of bacteria with the host plants, such as the *BAMY6639_15165* gene encoding a putative adhesion protein with a predicted sorting signal for its covalent anchorage to the bacterial cell envelop, which might hypothetically mediate cell to cell interactions or adherence to the host surfaces (AR22). In the category of essential metabolism, we found xanthine dehydrogenases (AR15), which participate in purine metabolism, and lipases (AR12), *N*-acetyltransferases (AR2, AR5) and malonyl-CoA transacylase (AR18), which is involved in fatty-acid metabolism. Finally, it was interesting to detect genes that were hypothetically involved in the transport of bacitracin (AR3), a branched cyclic dodecylpeptide antibiotic which is nonribosomally produced by *B. licheniformis* and *B. subtilis* strains, (Dintner et al., 2014), but not the homologous biosynthetic genes. This finding led us to hypothesize about the putative mechanism by which this strain might possess immunity to bacitracin and related metabolites.

A similar number of atypical regions (26) were detected in the *B. amyloliquefaciens* CECT 8238 genome, and 15 of these regions appeared to be genomic islands (from 4 to 33 Kb) (Figure 12B and Annex Table 4). Seven of these genomic islands were identified as putative prophages, and an additional region contained genes related to phage proteins or mobile element domains such as peptidases, terminases and integrases. As observed in the other strain, most of the genes contained in these genomic islands are hypothetical proteins, but it was also possible to identify genes coding proteins with known functionalities. For the genes overlapping the unique genomic windows (Figure 11B, outer line, white boxes), we found interesting features, i.e., the enzymes beta-xylosidase and sucrose-6-phosphate hydrolase, which enable this bacterium to utilize sucrose as an alternative carbon source to glucose, and the xanthine dehydrogenases and *N*-acetyltransferases, which (as previously indicated for CECT 8237) are involved in fatty-acid metabolism. We also detected the biosynthetic genes of macrolactin, difficidin or

bacillaene, metabolites related to direct antagonism of pathogenic microbes. It is worth noting the presence of the ICEBs1 excisionase, which has been previously reported to be part of the type IV secretion system in *B. subtilis* subsp. *subtilis* 168, a hypothetical mechanism for horizontal transfer of bacterial goods. Further analysis of this region led us to identify all of the putative genes implied in this process (*ydcQ*, *yddE*, *yddD*, *yddG*, *yddB*, *yddC* and *yddH*) (Alvarez-Martinez and Christie, 2009).

Figure 12. Unique windows detected in the genome of *Bacillus amyloliquefaciens* CECT 8237 (A) or CECT 8238 (B) and non-conserved in the *Bacillus* genus. The blue line denotes the fluctuations of the conservation measurement along the genome sequences. Downward peaks correlate with atypical regions and upward peaks denote highly conserved regions. Each green line represents an atypical region that contains several genes. A total of 28 atypical regions have been identified in CECT 8237 and 26 in CECT 8238.



To confirm the specificity of some of these atypical regions in our strains, we designed pairs of primers on the *BAMY6639_17480* gene in CECT 8237 and on the *BAMY6614_00315* gene in CECT 8238 (see Materials and Methods for details). In this analysis, we included a collection of strains and isolates from our own laboratory collection (Table 1, strains for experimental application). The primers were highly specific and provided a product from the DNA of the corresponding strain CECT 8237 or CECT 8238 but not in the rest of the *Bacillus* strains included in this analysis (Figure 13A-B). As a control of the diagnostic polymerase chain reaction (PCR), we also tested a pair of primers that amplified a fragment of the gene *bmyB,* implicated in the synthesis of bacillomycin. As

expected from the available genome information, a product was obtained in some of the *B. amyloliquefaciens* and *B. subtilis* strains but not in any of the *B. cereus* group or *B. subtilis*-type strains (Figure 13C). Finally, the pair of primers that targeted the *rpoA* gene resulted in a product in all of the strains (Figure 13D).

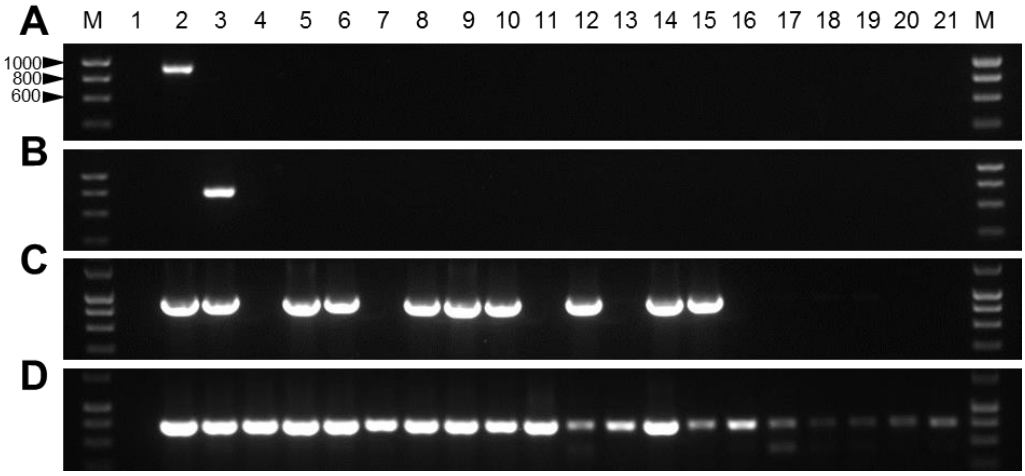


Figure 13. Diagnostic PCR for traceability of *Bacillus amyloliquefaciens* CECT 8237 and CECT 8238. A and B represent amplicons obtained with a specific pair of primers designed on the unique genomic areas of CECT 8238 or CECT 8237, respectively. C. The *bmyB* gene was detected in all of the strains that were producers of the iturins family of lipopeptides (Bacillomycin or Iturin). D. A partial sequence of the *housekeeping* gene *rpoA* was amplified in all of the strains. Lane numbers are (1) Water sample without DNA template, (2) CECT 8238, (3) CECT 8237, (4) *B. subtilis* subsp. *subtilis* 168, (5) *B. subtilis* UMAF8561, (6) *B. subtilis* UMAF8562, (7) *B. subtilis* UMAFBiA758, (8) *B. subtilis* UMAF6619, (9) *B. subtilis* UMAF1605, (10) *B. subtilis* UMAF1610, (11) *B. amyloliquefaciens* BGSC10A1, (12) *B. amyloliquefaciens* BGSC 10A3, (13) *B. amyloliquefaciens* BGSC 10A5T, (14) *B. amyloliquefaciens* subsp. *plantarum* FZB42, (15) *B. amyloliquefaciens* BGSC 10A18, (16) *B. flexus* CIP 106928T, (17) *B. thuringiensis* ssp. *kurstaki* CECT 4454, (18) *B. cereus* ATCC 14579, (19) *B. cereus* UMAF8564, (20) *Paenibacillus polymyxa* CECT 155 and (21) *Brevibacillus laterosporus* CECT 15. M, molecular weight marker HyperLadder^{TM} 1Kb (BIOLINE). Numbers on the left are the size of the amplicons in base pairs.

***B. amyloliquefaciens* CECT 8237 and CECT 8238 have singularities in cellular communication and biofilm formation**

It is assumed that the good performance of a biocontrol agent relies on its effective colonization and persistence (Romero et al., 2004; Zeriouh et al., 2014). An important contributor to this bacterial fitness is the formation of biofilms. Within a biofilm, cells are more protected from external aggressions than planktonic cells, which is due to an extracellular matrix made of exopolysaccharides, proteins, or nucleic acids that provides the entire community with structural robustness (Flemming and Wingender, 2010). An important aspect for the establishment and growth of a bacterial biofilm is motility, and both strains display swimming motility, an independent cell movement in liquid medium, and swarming, a multicellular movement driven by the coordinated action of *swrA*, *swrB* and *swrC* and the lipopeptide surfactin (Annex Table 5) (Kearns et al., 2004). Indeed, we have previously demonstrated that the long-term persistence of these strains on melon leaves and efficient biocontrol activity both rely on the formation of biofilms, which appear to be dependent on the action of surfactin (Zeriouh et al., 2014). In agreement with these findings, we identified all of the genes described until now to participate in formation of biofilms, specifically, i) regulatory genes such as *abrB*, *sigW*, *sinI*, *sinR* and *spo0A* (Vlamakis et al., 2013) and ii) structural genes for the synthesis of the components of the extracellular matrix, including the 15-gene exopolysaccharide operon (*epsA-O*), the operon involved in the formation of TasA amyloid-like protein (*tapA-sipW-tasA*) and the hydrophobin protein BslA (Kearns et al., 2005; Romero et al., 2010; Hobley et al., 2013). We were not able to identify differentiable features between our strains and the type strain *B. amyloliquefaciens* subsp. *plantarum* FZB42, which harboured all of these genes and are almost identical at amino acid sequence level (Figure 14, black boxes). However, the colony architecture was visually different among CECT 8237, CECT 8238 and the type

strain FZB42 and the *B. subtilis* subsp. *subtilis* type strains 168 and NCBI 3610 (Figure 15), which led us to speculate about subtle differences in other complementary factors or regulatory pathways that might influence the final phenotypes.

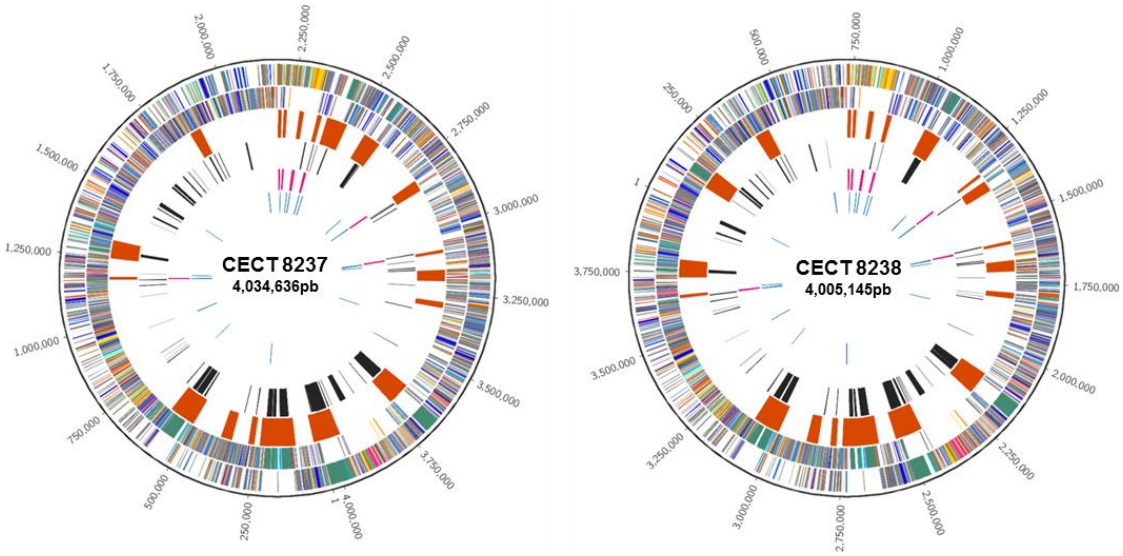


Figure 14. Bacterial features related to the biocontrol activity of *Bacillus amyloliquefaciens* CECT 8237 and CECT 8238. The outer scale of the circles designates coordinates in base pairs (bp). The first and second circles show the predicted coding regions on the plus strand or minus strand, respectively, and colour-coded by Clusters of Orthologous Groups (COGs) (see Figure 10 for details). The third circle (red) shows the predictions of gene clusters involved in the production of secondary metabolites (antiSMASH software). The fourth circle (black) shows genes previously reported as related to biocontrol activity. The fifth circle (pink) shows rRNA genes. The sixth circle (blue) shows tRNA genes.

In terms of microbial ecology, this bacterial communication system plays an important role in defining putative interactions and cooperation between members of the same pherotype, which means that they share the same QS type (Stefanic et al., 2012; Dogsa et al., 2014). Unsurprisingly, we identified these loci in CECT 8237 and CECT 8238 and confirmed that their amino acid sequences are slightly conserved with FZB42; *comQ* and *comX*, in particular, retrieved notoriously low levels of similarity. Thus, we characterized

the possible diversity of the *comQXPA* locus along several strains corresponding to the *B. subtilis* group, including *B. amyloliquefaciens* plant-associated and non-plant-associated strains (Figure 16). At a glance, it can be observed that the *B. subtilis* group predominantly exhibits overlapping *comQ-comX*, that the *B. amyloliquefaciens* group contains a majority of strains that have independent loci, and that a few *B. amyloliquefaciens* strains show overlapping *comQ-comX-comP* (Figure 16). However, it was noteworthy that FZB42 clusters together with the *B. subtilis* strains and that the CECT 8237 and CECT 8238 strains fall into divergent pherotypes within the *B. amyloliquefaciens* group.

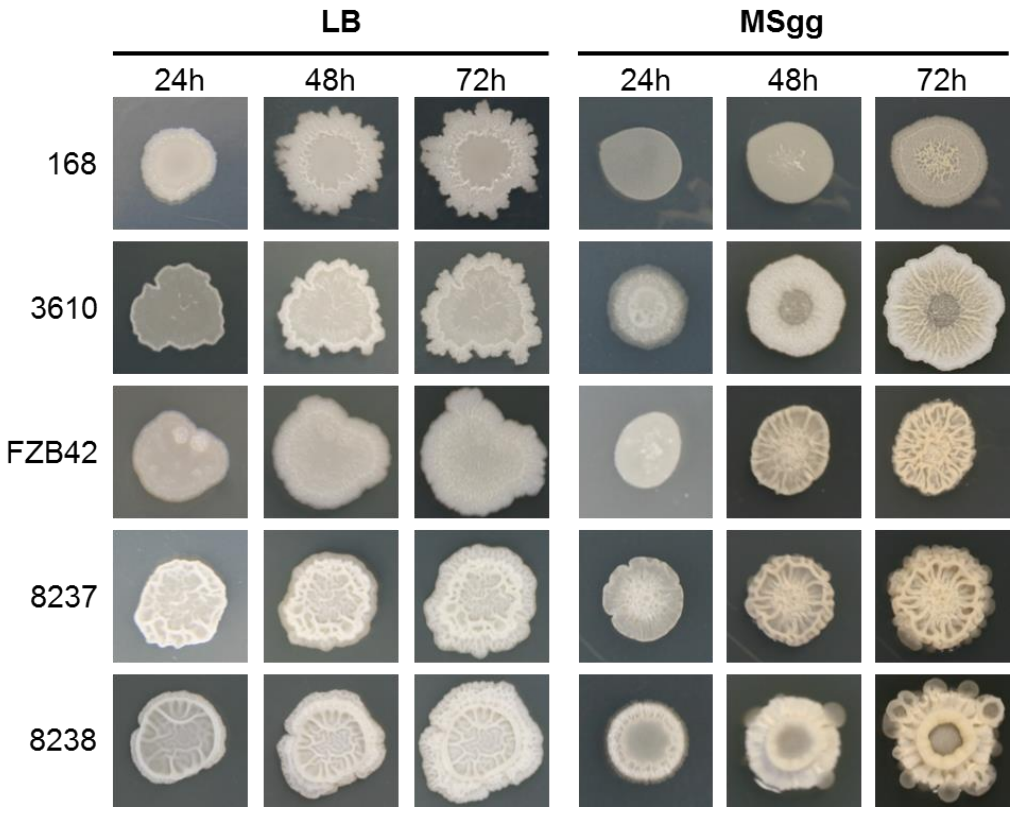


Figure 15. *Bacillus* strains differ in the morphological features of biofilms. Biofilm formation was evaluated as colony morphology in LB or MSgg agar and incubation at 30ºC. The strains are: *B. subtilis* subsp. *subtilis* 168 (168), *B. subtilis* subsp. *subtilis* NCIB 3610 (3610), *B. amyloliquefaciens* subsp. *plantarum* FZB42 (FZB42), *B. amyloliquefaciens* CECT 8237 (8237) and *B. amyloliquefaciens* CECT 8238 (8238).

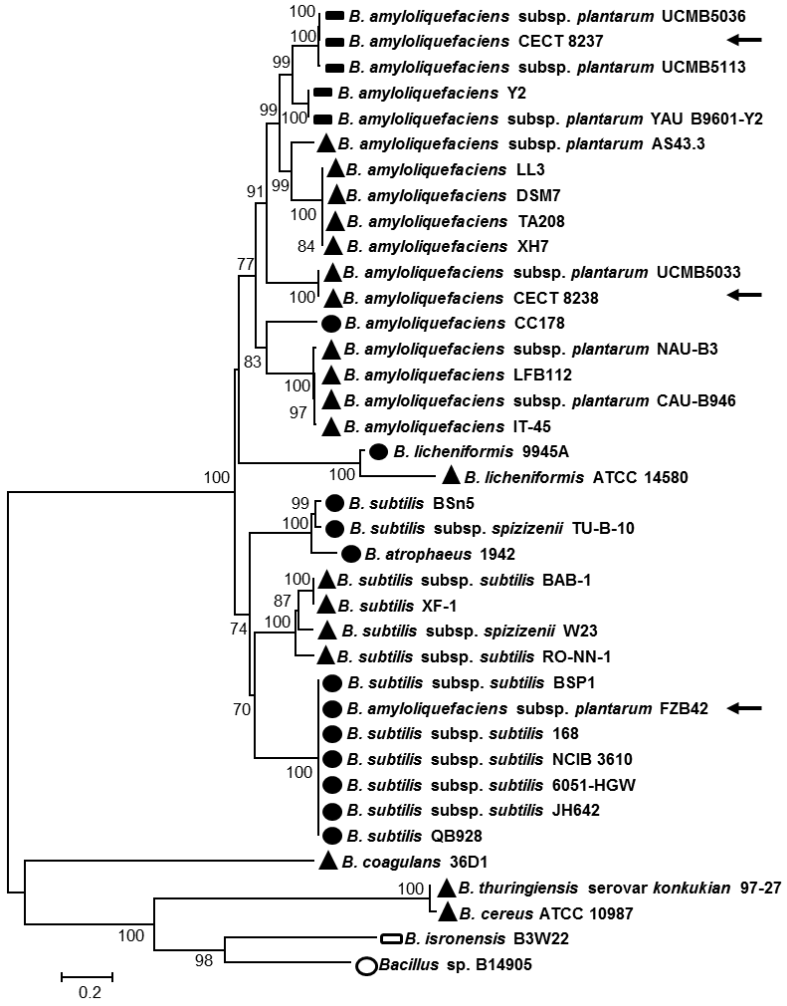


Figure 16. *Bacillus amyloliquefaciens* CECT 8237 and CECT 8238 belongs to different pherogroups based on the analysis of the competence loci *comQXPA*. The sequences of the competence genes *comQ-comX-comP-comA* were refined and manually tested for accuracy and concatenated to build a Neighbour-joining tree using MEGA 5 bootstrap values (10,000 replicates), which are shown on the branches. The topology was identical for the trees produced by the minimum evolution and maximum parsimony methods. The sequences from all of the strains used were extracted from published genome sequences. The symbols represent: no overlapping (▲), *comQ-comX* overlaps (●), *comP-comA* overlaps (○), *comQ-comX-comP* overlaps (■), *comX-comP-comA* (□). Arrows indicate the location of the CECT 8237, CECT 8238 and FZB42 strains in the tree.

In addition to the ComX pheromone, the extracellular Phr peptides regulate relevant biological processes in *Bacillus* spp., such as sporulation, synthesis of antibiotics and biofilm formation. Each Phr peptide is intimately associated with a cognate intracellular regulator Rap protein that is directly suppressed by the expression of those peptides (Gallego del Sol and Marina, 2013). *B. subtilis* encodes eleven Rap proteins (RapA to RapK) and eight Phr peptides (PhrA, PhrC, PhrE, PhrF, PhrG, PhrH, PhrI and PhrK), which inhibit the response regulators of a diverse two-component regulatory system (Gallego del Sol and Marina, 2013). Additional Rap-Phr modules have been found in plasmids of diverse wild *Bacillus* isolates, and examples are Rap60-Phr60 in the plasmid pTA1060, RapP-PhrP in the plasmid pBS32, or RapQ-PhrQ in the plasmid pBSG3 (Yang et al., 2015). Considering the relevance of these *rap-phr* operons in the modulation of biological responses in *Bacillus* spp., we decided to search for the distribution of all of these genes in the genome sequences of CECT 8237, CECT 8238, and other *Bacillus* strains (Annex Table 5). Most of the *rap* genes and their corresponding antagonistic *phr* were identified, but we did find exceptions. For example, *rapG* is absent in both CECT 8237 and CECT 8238. The strain CECT 8237 appears to have two copies of the gene *rapH*, as represented by *rapH1* and *rapH2*, but not the cognates *phr*. The tandem *rapI-phrI* is only present in the strain CECT 8238, an expected finding, given that the presence of the mobile element ICEBs1 is only in this strain (Auchtung et al., 2005). Finally, *rapK-phrK* is present in CECT 8237 and CECT 8238 but absent in FZB42, and an additional aspartate phosphatase annotated as RapX, but not the cognate PhrX, was identified in both strains (i.e., FZB42 and other *B. amyloliquefaciens* strains, but not *B. subtilis*).

B. amyloliquefaciens CECT 8237 and CECT 8238 are triggers of the immune response and growth of plants

We demonstrated in separate studies that *B. amyloliquefaciens* CECT 8237 and CECT 8238 contribute to plant health by triggering plant defence machinery, known as induced systemic resistance (ISR), and promoting plant growth (García-Gutiérrez et al., 2012; García-Gutiérrez et al., 2013). In these studies, we proved the relevant role of the lipopeptide surfactin in triggering the ISR in melon plants (García-Gutiérrez et al., 2013). Studies in other pathosystems have reported that, as surfactin, the lipopeptide fengycin may also mediate communication with plants, eliciting the ISR in addition to their well-known surfactant or antimicrobial activity (Ongena et al., 2007; Zeriouh et al., 2014; Cawoy et al., 2015). Besides these specific ISR triggers, the cumulative study of diverse biocontrol agents permits a number of generic bacterial attributes to be tentatively implicated in this biocontrol activity (Table 7). Thus, it was not surprising to find in the genomes of both strains the genes involved in the synthesis of these molecules that act as elicitors of the nonspecific basal plant immunity and are known as microbe-associated molecular patterns (MAMPs). Flagellin proteins, which are essential in motility, are also known as triggers of the innate plant immune response against potential pathogens. Other known MAMPs present in these strains are encoded by the operon *tuaA-tagO*, which participates in the synthesis of teichuronic acid, a basic component of the cell walls of Gram-positive bacteria, and the elongation factor *tufA* genes (Boller and Felix, 2009; Kierul et al., 2015).

Table 7. Genes involved in the beneficial contribution of *Bacillus amyloliquefaciens* CECT 8237 and CECT 8238 to plant health.

Genes	Protein	Functionality	CECT 8237 % ID[z]	CECT 8238 % ID[z]
Bacterial target molecules for general plant immune response				
flgK	Flagellin HAP1	Involved in elicitation of plant basal defense	99	99
fliD	Flagellin HAP2	Involved in elicitation of plant basal defense	95	94
hag	Flagellin HAG	Involved in elicitation of plant basal defense	93	90
tufA	Elongation factor EF-Tu	Involved in elicitation of plant basal defense	99	100
tuaA-tagO	Operon involved in teichuronic acid/ lipopolysaccharide biosynthesis	Involved in elicitation of plant basal defense	99	99
Biofertilization				
yvrC	Putative iron-binding protein	Putative iron availability	99	99
yvrB	Putative iron permease	Putative iron availability	99	99
yvrA	Putative ABC transport system ATP-binding protein	Putative iron availability	98	99
yusV	Putative iron(III) ABC transport ATPase component	Putative iron availability	99	99
phy	Phytase	Phosphate availability	99	98
yclQ	Ferrichrome ABC transporter	Transport/binding proteins and lipoproteins	99	99
Phytostimulation and induced systemic resistance				
dhaS	Putative indole 3-acet-aldehyde dehydrogenase	Involved in Trp-dependent indole 3-acetic acid (IAA) synthesis	99	99
ysnE	Putative IAA-acetyl-transferase	Involved in Trp-dependent IAA synthesis	98	98
yhcX	Nitrilase	Involved in Trp-dependent IAA synthesis	99	99
alsD	Alpha-acetolactate decarboxylase	Synthesis of acetoin	98	99

alsS	Acetolactate synthase	Synthesis of acetoin	99	99
alsR	LysR transcription regulator	Regulator of the *alsDS* operon	99	99
bdhA	Acetoin reductase/ 2,3-butanediol dehydrogenase	synthesis of 2,3-butanediol	99	99

[z] Percent identity (ID) compared with *B. amyloliquefaciens* subsp. *plantarum* FZB42. The percentage of coverage is 100 %, except for the *hag* gene, which is 70 % or 81 %, respectively, in CECT 8237 or CECT 8238.

In a previous study, we demonstrated that both CECT 8237 and CECT 8238 strains are producers of the auxin phytohormone indole-3-acetic acid (IAA) (García-Gutiérrez et al., 2012), a molecule known to promote plant growth. Similarly, we found the genes of the tryptophan-dependent pathway implicated in the synthesis of this phytohormone (Table 7). In both genomes, we identified the genes implied in the synthesis of acetoin (*alsS* and *alsD*) or 2,3-butanediol (*bdhA*) (Table 7 and Figure 17), two bacterial organic volatile compounds involved in this bacteria-plant communication (Farag et al., 2013). We first attempted to detect the production of both volatiles using a qualitative analysis. The Voges-Proskauer test (Figure 17C), proved that CECT 8237 and CECT 8238 produce acetoin (reddish colour of the medium) as well as the indicator strains *B. subtilis* subsp. *subtilis* 168 and NCIB 3610, or *B. amyloliquefaciens* subsp. *plantarum* FZB42 (Nicholson, 2008; Kierul et al., 2015).

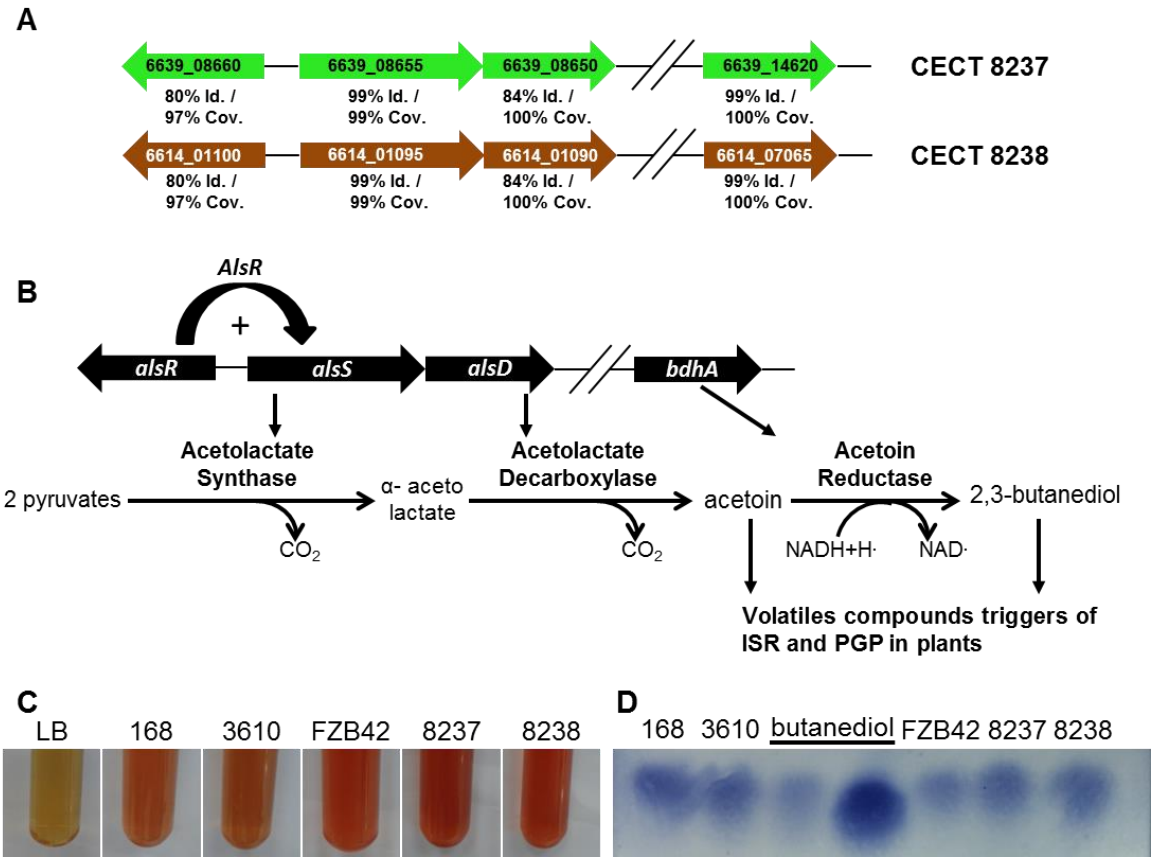


Figure 17. *Bacillus amyloliquefaciens* CECT 8237 and CECT 8238 have the genes of the biosynthetic pathway of the volatile compounds acetoin and 2,3-butanediol. A. The genetic organisation of the genomic regions containing these biosynthetic genes in CECT 8237 and CECT 8238. The codes inside the arrows indicate the reference number of the corresponding gene in the genome sequence from each strain. The percentages of identity and coverage are indicated for each locus compared to *B. subtilis* subsp. *subtilis* 168. B. *alsR* encodes a positive regulator of the operon *alsS-alsD* that encodes the two enzymes that synchronically produce acetoin. The monocistronic *bdhA* encodes the enzyme that reduces acetoin to the final product 2,3-butanediol. *Bacillus* strains were growth in LB supplemented with 1% glucose to determine the production of these volatile compounds. C. The Voges-Proskauer test illustrated the production of acetoin (reddish colour of the culture) after 48 h of growth. D. TLC analysis of the supernatants of *B. amyloliquefaciens* strains and *B. subtilis* subsp. *subtilis* 168 revealed the presence of 2,3-butanediol, which migrated as the standard butanediol.

The analysis of bacterial supernatants in thin-layer chromatography (TLC) plates (Figure 17D), revealed a spot with the retention factor (Rf) similar to a 30mM butanediol standard solution as the control (Nicholson, 2008). Next, we evaluated the kinetics of production of both molecules over time using a gas chromatography-mass spectrometry (MS) analysis of bacterial supernatants, as described in the corresponding material and methods section (see Annex Figure 2 for the characteristic chromatogram and mass spectra associated to each molecule). It was interesting that *B. subtilis* subsp *subtilis* 168 and NCIB3610 or *B. amyloliquefaciens* subsp. *plantarum* FZB42 produced more acetoin than did butanediol, whereas the opposite occurred in CECT 8237 and CECT 8238 (Figure 18A). In addition, the amount of butanediol was surprisingly higher (one order of magnitude) in CECT 8237 and CECT 8238 than in the rest of the strains. In general, the kinetics was similar in these experimental conditions. The amount of both volatiles in the supernatants were picked at 24 h of growth, with the exceptions of FZB42, which accumulates more acetoin at 36 h, and CECT 8237, which accumulates more acetoin at 12 h and butanediol at 36 h. As anticipated from these findings, all of the strains promoted the formation of heavier and more abundant roots of melon seeds than the untreated seeds (Figure 18B and C). However, no clear differences could be determined between treatments.

In addition to the production of plant growth promoters, beneficial bacteria may contribute to plant health as biofertilizers (Pérez-García et al., 2011). Indeed, in the genomes of both strains, we found the genes *yusV* and *yclQ*, putatively involved in iron mobilization, and the group of genes orthologous to *yvrCBA* in *B. subtilis* subsp. *subtilis* 168, which is possibly related to vitamin B12 transport (Table 7). Finally, as mentioned above, the *phy* gene that encodes 3-phytase, an enzyme thoroughly reported as a key element in biofertilization, was also identified.

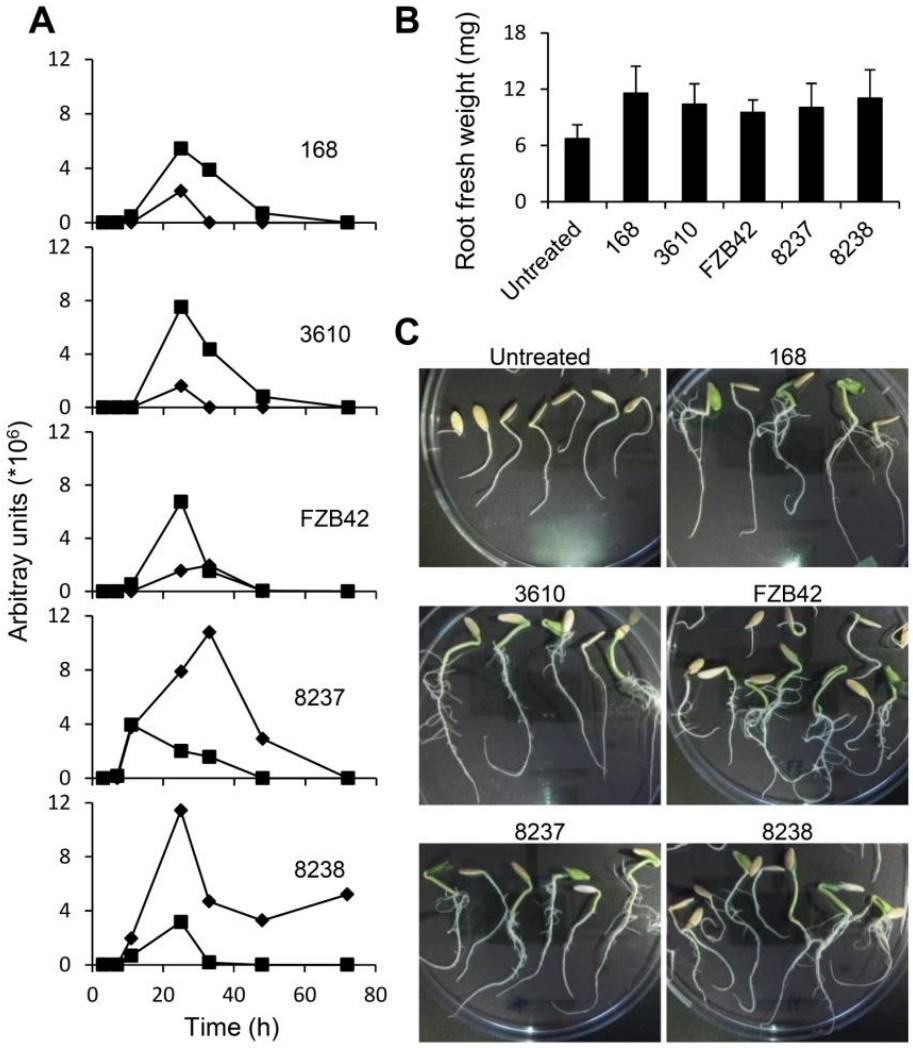


Figure 18. *Bacillus amyloliquefaciens* CECT 8237 and CECT 8238 produce the plant-growth promoter volatile compounds acetoin and 2,3-butanediol. A. Gas chromatography-mass spectrometry analysis of bacterial supernatants was used to determine the kinetics of the production of acetoin (black squares) and 2,3-butanediol (black diamonds) by *Bacillus* strains grown in Luria-Bertani medium supplemented with 1% glucose at 30°C. B. Estimation of the root weight indicates that bacilli induce the growth of the root system of melon seeds. C. Representative pictures of the experiment at 7 days shows more abundant and developed roots in the melon seeds treated with bacilli compared to untreated seeds. The treatments were: untreated, nontreated control; 168, *B. subtilis* subsp. *subtilis* 168; 3610, *B. subtilis* subsp. *subtilis* NCIB 3610; FZB42, *B. amyloliquefaciens* subsp. *plantarum* FZB42; 8237, *B. amyloliquefaciens* CECT 8237; or 8238, *B. amyloliquefaciens* CECT 8238.

B. amyloliquefaciens CECT 8237 and CECT 8238 are biological factories of antimicrobial compounds

Secondary metabolites are important and versatile weapons that bacteria use to fight other microbes and are thus highly valuable in plant protection against microbial pathogens (Chen et al., 2009; Pérez-García et al., 2011; Yuan et al., 2012). We experimentally demonstrated the production of lipopeptides, surfactin, iturins and fengycins (Romero et al., 2007a). Genome data mining using the Anti-Smash platform confirmed these findings and revealed the detection of operons for the production of i) bacillibactin siderophore (Miethke et al., 2006), ii) dipeptide bacilysin, (Steinborn et al., 2005), iii) macrolactin (Yuan et al., 2012), iv) bacillaene (Butcher et al., 2007), and v) difficidin (Chen et al., 2009). High-pressure liquid chromatography (HPLC) –electrospray (ESI)-MS analysis of the methanolic extracts obtained from cell-free supernatants of both strains grown in Landy broth (Chen et al., 2009) revealed the presence of bacilysin, bacillaene and dihydrobacillaene (Figure 19, left charts and Annex Figure 3). The same analysis of the methanolic extracts from the cultures in the GA medium (Chen et al., 2009) showed traces of bacillibactin, difficidin and macrolactin (Figure 19, right charts and Annex Figure 3).

In addition, our comparative analysis showed the presence of additional gene clusters that are hypothetically involved in the synthesis of unknown or uncharacterized secondary metabolites (Figure 20).

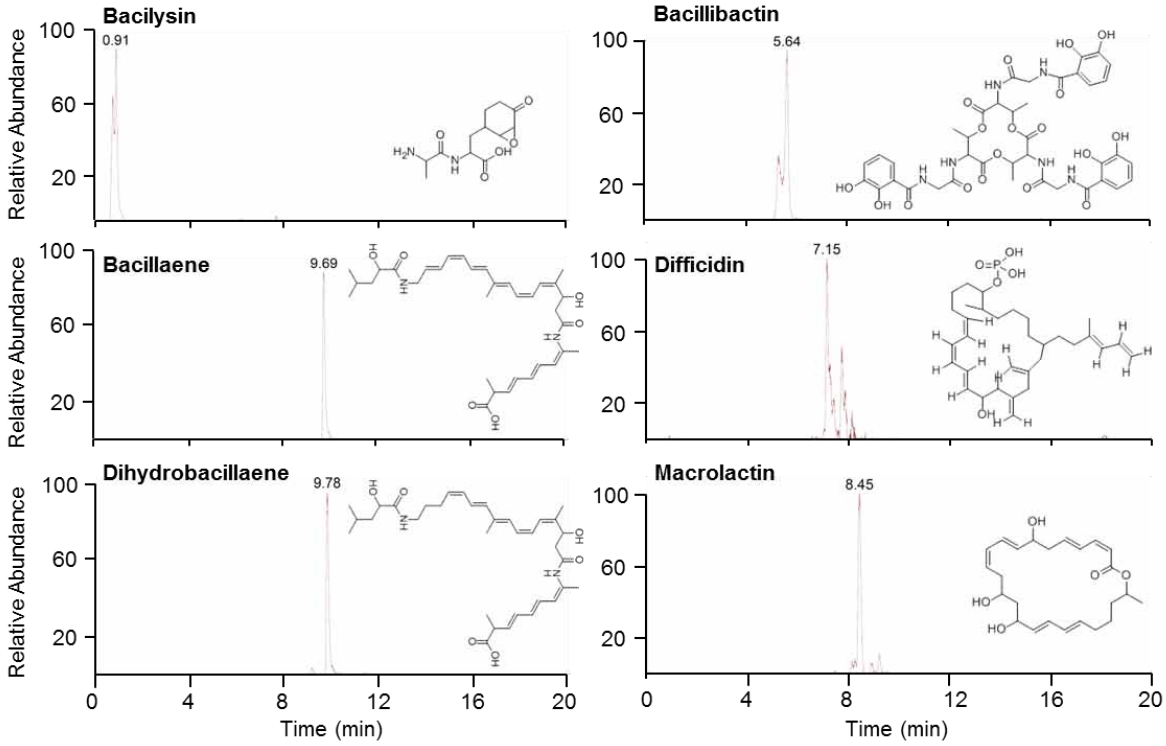


Figure 19. *Bacillus amyloliquefaciens* CECT 8237 and CECT 8238 produce a variety of known secondary metabolites. High-performance liquid chromatography-electrospray mass spectrometry analysis of methanolic extracts from the cell-free supernatants of cultures in Landy medium for 40 h at 30°C revealed the presence of bacilysin, bacillaene, and dihydrobacillaene. The analysis of methanolic extracts from the cell-free supernatants of cultures in gibberellic acid medium for 24 h at 30°C revealed the production of bacillibactin, difficidin and macrolactin. The chemical structures of the molecules are included (insets).

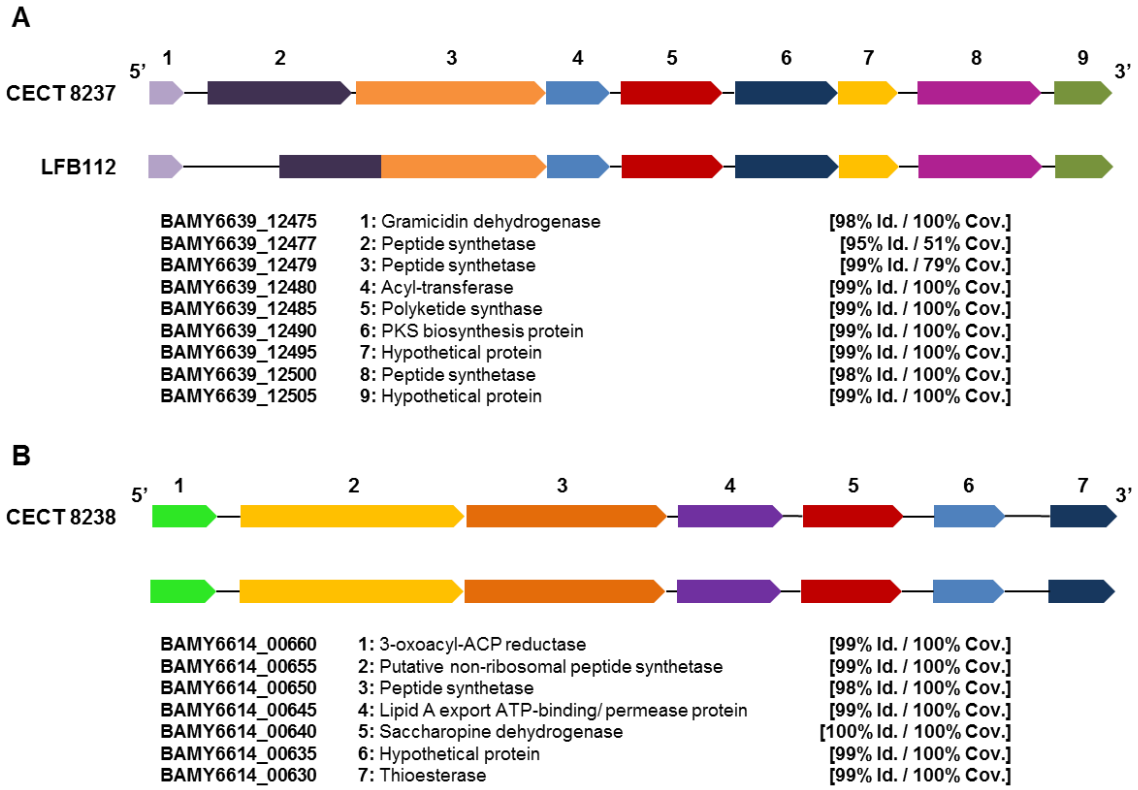


Figure 20. *Bacillus amyloliquefaciens* CECT 8237 and CECT 8238 have putative gene clusters dedicated to production of novel secondary metabolites. A. A gene cluster putatively involved in the synthesis, modification and transport of a new nonribosomal peptide detected in CECT 8237 and *B. amyloliquefaciens* LFB112. B. A gene cluster possibly involved in the synthesis, modification and transport of an uncharacterised nonribosomal peptide and found in CECT 8238 and the *B. amyloliquefaciens* strains NAU-B3, YAU B9601 and Y2.

Interestingly, these novel gene clusters were contained in the atypical regions that were previously described for these strains (Figure 11), indicating that they are nonconserved among the *Bacillus* genus. In the CECT 8237 strain (AR18), we only found a complete gene cluster present in the recently sequenced genome of *B. amyloliquefaciens* LFB112 that was likely involved in the synthesis of a novel non-ribosomal peptide synthetase (Fig. 18A) (Cai et al., 2014). In the CECT 8238 strain, we found a group of genes with high probability to synthesize an unknown nonribosomal peptide synthetase (Fig. 18B), which

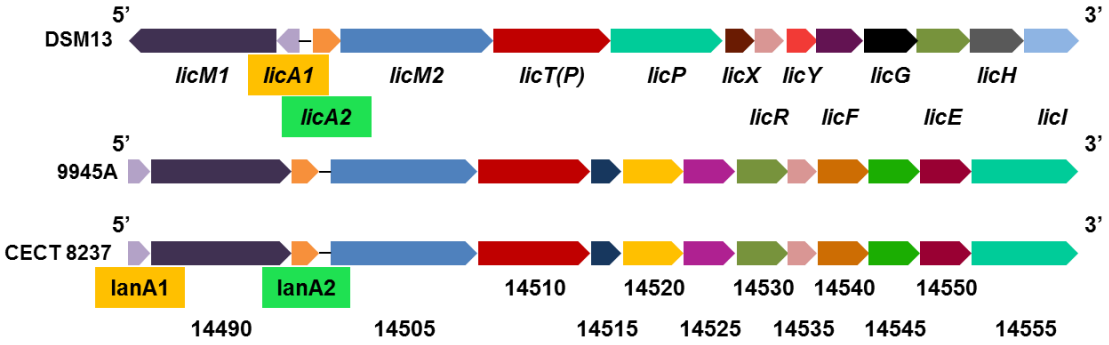
correlated with AR3. This region has been previously reported in *B. amyloliquefaciens* subsp. *plantarum* NAU-B3, YAU B9601-Y2 and *B. amyloliquefaciens* Y2 strains, which are all associated with plants (He et al., 2012). The functionality and structures of the molecules potentially synthesized by these hypothetical gene clusters have not been demonstrated yet.

B. amyloliquefaciens CECT 8237 contains a gene cluster orthologous to the lichenicidin lantibiotic production of _B. licheniformis_ DSM 13

In addition to the atypical regions mentioned above, we also detected the AR21 in *B. amyloliquefaciens* CECT 8237, a unique genomic window (Figure 11A, outer line, white boxes and Figure 12A) poorly conserved within *Bacillus* genus, and a putative genomic island according to the Alien Hunter analysis, local deviations in the trinucleotide usage patterns and variation in the percent GC content (37.6% comparing with the whole genome 46.34%). Even though initially annotated as genes with unknown function, the analysis with antiSMASH (Figure 14, left) predicted their putative involvement in the biosynthesis of a secondary metabolite. Further analysis of this gene cluster of 16731bp, revealed the highly probable dedication to the production of a lantibiotic similar to the lichenicidin lantibiotic produced by *B. licheniformis* DSM13 (Figure 21). The fact that this AR was absent in all *B. amyloliquefaciens* strains except CECT 8237 or *B. licheniformis* 9945A, suggested an acquisition through horizontal transfer.

Our comparative analysis revealed the presence of the entire genetic machinery in *B. amyloliquefaciens* CECT 8237 orthologous to that of *B. licheniformis* and dedicated to the synthesis of lichenicidin: the two structural genes *lanA1* (BAMY6639_14485) and *lanA2* (BAMY6639_14495) orthologous to *licA1* and *licA2* respectively; the two corresponding modifying genes, *lanM1* (BAMY6639_14490) and *lanM2* (BAMY6639_14505); *lanT* gene

(BAMY6639_14510) encoding the protein involved in transport and processing of the signal peptides of the prepeptides; and *lanP* (BAMY6639_14555) which encodes a serine protease supposedly involved in further processing of the peptides. In addition, we found the putative transcriptional regulator *lanR* (BAMY6639_14535), a locus encoding a small hypothetical protein (BAMY6639_14515), and several genes hypothetically implicated in immunity and transport or secretion of molecules (BAMY6639_14520, BAMY6639_14525, BAMY6639_14530, BAMY6639_14540, BAMY6639_14545 and BAMY6639_14550) (Dischinger et al., 2009).



BAMY6639_14485	**1:** Lichenicidin prepeptide *LanA*-like protein	[78% Id. / 100% Cov.]
BAMY6639_14490	**2:** Lantibiotic biosynthesis protein *LanM*	[69% Id. / 98% Cov.]
BAMY6639_14495	**3:** Lichenicidin prepeptide *LanA*-like protein	[56% Id. / 90% Cov.]
BAMY6639_14505	**4:** Lantibiotic biosynthesis protein *LanM*	[67% Id. / 99% Cov.]
BAMY6639_14510	**5:** Lantibiotic ABC exporter, peptidase/ATP-binding protein *LanT*	[81% Id. / 100% Cov.]
BAMY6639_14515	**6:** Hypothetical protein	[72% Id. / 100% Cov.]
BAMY6639_14520	**7:** Putative ABC transporter ATP-binding protein	[82% Id. / 100% Cov.]
BAMY6639_14525	**8:** Lantibiotic-exporting membrane permease/ABC-transporter permease	[73% Id. / 100% Cov.]
BAMY6639_14530	**9:** Lantibiotic-exporting membrane permease/Immunity protein *LanE*	[77% Id. / 100% Cov.]
BAMY6639_14535	**10:** Transcriptional regulator *LanR*	[85% Id. / 94% Cov.]
BAMY6639_14540	**11:** Lantibiotic-exporting membrane permease	[70% Id. / 100% Cov.]
BAMY6639_14545	**12:** Lantibiotic-exporting membrane permease	[56% Id. / 100% Cov.]
BAMY6639_14550	**13:** Lantibiotic-exporting membrane permease	[70% Id. / 97% Cov.]
BAMY6639_14555	**14:** Intracellular serine protease *LanP*	[68% Id. / 96% Cov.]

Figure 21. *Bacillus amyloliquefaciens* CECT 8237 gene cluster possibly implicated in the production of a lichenicidin-like lantibiotic. Arrangement of the genes contained in the genomic cluster dedicated to lichenicidin production in *B. licheniformis* DSM13, *B. licheniformis* 9945A and *B. amyloliquefaciens* CECT 8237. The homology between genes is denoted by the same colour code. Genes implicated in the synthesis of the prepeptides are highlighted in yellow (*lanA1*) and green (*lanA2*). Genes putatively involved in the synthesis, modification and transport of the new antibiotic detected in CECT 8237 and *B. licheniformis* 9945A are also indicated. The percentages of identity and coverage are referred to CECT 8237 in comparison with 9945A at amino acid level.

To demonstrate the functionality of the lantibiotic related genes, a reverse transcription-polymerase chain reaction (RT-PCR) analysis was performed using the pairs of primers that target regions inter or intra loci (Figure 22A). The primers amplified on genomic DNA from *B. amyloliquefaciens* CECT 8237 and gave signal of the expected size, indicative of their specificity and functionality (Figure 22B-C, top pictures and Table 2). The results from the RT-PCR clarified that all the ORFs of the region are transcribed and probably in more than one transcriptional unit (Figure 22B, bottom pictures): *lanA1*, *lanM1*, *lanA2*, *lanM2* and *lanT,* the structural and modifier genes and the gene involved in transport and processing of the signal peptides are transcribed in one polycistronic mRNA; *lanR*, the putative transcriptional regulator and *lanFGE* predicted to be involved in transport, in other mRNA molecule; and *lanP*, encoding a peptidase, and the rest of transporter genes co-transcribe in a third transcriptional unit. Accordingly, we predicted the presence of putative promoters upstream the first ORF of each polycistronic unit (Figure 22A top, red arrows).

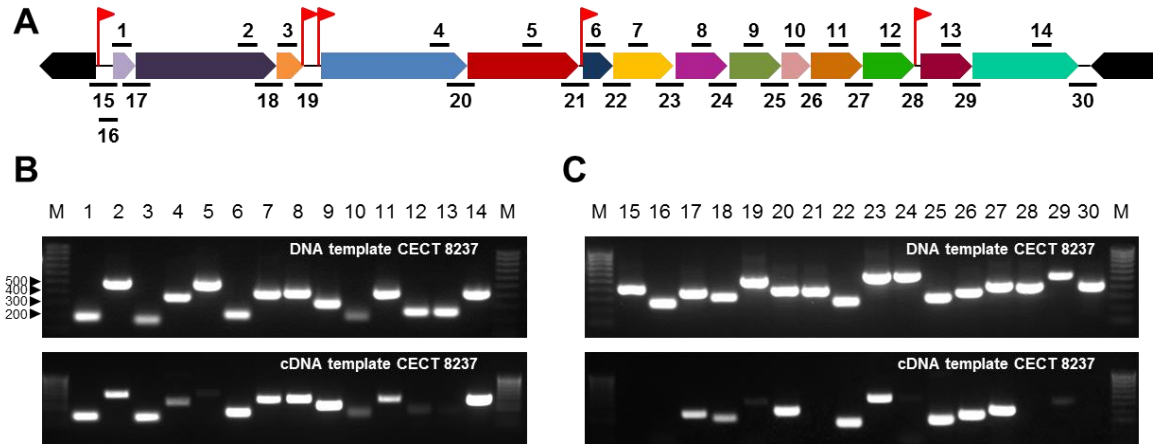


Figure 22. The new lantibiotic gene cluster identified in *Bacillus amyloliquefaciens* CECT 8237 is transcribed to different polycistronic mRNAs. A. Expected amplifications of intra-genic (top lines) or intergenic (bottoms lines) regions with each specific pair of primers, and putative promoters driving the transcription of each polycistronic mRNA (red arrows). B. and C. PCR products obtained within genes or between genes respectively, on genomic DNA (top pictures) or cDNA (bottom pictures). M, the molecular weight marker HyperLadder^{TM} 100bp (BIOLINE). Numbers on the left are the size of the amplicons in base pairs.

B. amyloliquefaciens CECT 8237 produces a two-peptide lantibiotic associated to cells and mostly released to the supernatant

In order to convert the inactive prepeptides into their active forms, diverse post translational modifications including proteolytic cleavage of the signal peptides occur. As stated earlier, dehydration of serine and threonine residues and cyclization of those with cysteine residues constitute the main modification steps leading to lanthionine and methyllanthionine production.

To elucidate the occurrence of these modifications in *B. amyloliquefaciens* CECT 8237, the amino acid sequences of lanA1 and lanA2 prepeptides were aligned with the corresponding prepeptides of other well-characterized two-peptide lantibiotics (Figure 23).

First, we noticed that amino acid sequence is highly conserved between lanA1 core peptide sequences, as indicate the coloured blocks. Second, we could distinguish two regions within the prepeptides: the signal peptide in the N-terminal part of the immature peptide with the predicted recognition site for cleavage (red arrow), followed by the core peptide that will suffer further maturation, which is again conserved in lanA1 peptides (Figure 23A), but not in lanA2 and orthologous, where the first amino acid after cleavage of the signal peptide varies (Figure 23B, red dots).

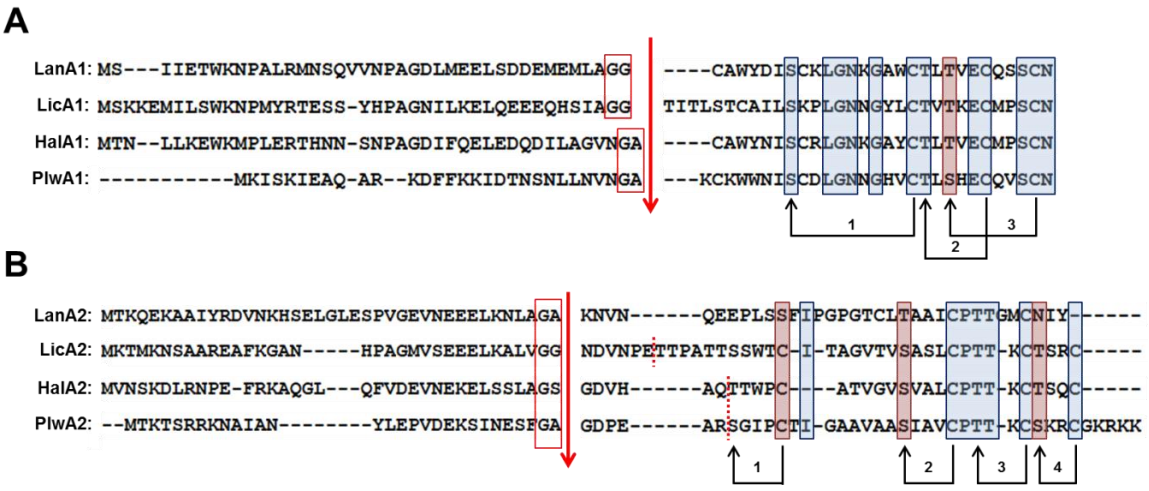


Figure 23. Theoretical processing of the two peptides associated to lichenicidin-like lantibiotic produced by *B. amyloliquefaciens* CECT 8237. A-B. Amino acid alignment of the LanA1 or LanA2 prepeptides with the prepeptides of the lantibiotics lichenicidin (LicA), haloduracin (HalA) and plantaricin (PlwA) respectively. Red arrow indicates the cleavage site of the signal peptide. The red point lines specifically show the N terminus of the mature peptides where an additional proteolysis occurs. Amino acid identities are highlighted in blue (100%) and pink (50-75%). The arrows indicate the thioether bridging pattern corresponding to each peptide.

The four lantibiotics (including LanA1) contained the cleavage sequence GG, GA or GS recognized by the corresponding signal peptidase (LanT in lichenicidin and lichenicidin-like peptide) (Figure 23). All these target sequences are related to the signal peptidase

cleavage site of type-B lantibiotics. Additional proteolysis implicating the removal of the next six amino acids happens in the second peptide of lichenicidin, haloduracin and plantaricin, (Figure 23B, red dotted lines). It is known that removal of this hexapeptide in lichenicidin is carried out by the action of the LicP protease, while for the other lantibiotics remains unidentified. The gene cluster of CECT 8237 contains an orthologous of LanP (Figure 21), thus we might hypothesize a similar processing event of the LanA2 peptide.

Three amino acids, serine, threonine and cysteine are relevant residues for further processing of the peptides. It is known that most of the Ser/Thr residues are dehydrated in lichenicidin, haloduracin and plantaricin lantibiotics, however, other residues escape from the activity of the dehydratase-cyclase activity: Ser30, Ser26 and Ser27, in LicA1, HalA1 and PlwA1 respectively; or Thr10-Ser21-Ser30, Ser22 and Thr21 in LicA2, HalA2 and PlwA2 respectively. On the view of these modifications in the reference peptides, we speculate no dehydration of Ser26 in LanA1, or Ser10 and Thr28 in LanA2 in CECT 8237.

Three thioeter bonds are typically found in HalA1 and PlwA1 peptides (Figure 23A), and an additional bond is formed between Ser5 and Cys7 in LicA1, while four thioeter bridges are in HalA2, PlwA2 and LicA2. Thus, we might assume three cyclizations in lanA1, due to the highly conserved amino acid sequence in comparison with the reference peptides, and at least two in LanA2 (Fig. 21B, 2 and 3) with less probable cyclizations 1 and 4 due to the absence of the residues involved in the reference peptides (Figure 23B). In addition it might be speculated a third thioether bridge between Ser11 and the internal Cys19 of LanA2, which is absent in the reference peptides.

In order to study the production of the lichenicidin lantibiotic in CECT 8237, we used the isopropanol extraction method of lichenicidin from bacterial cells as reported previously (Dischinger et al., 2009). The MALDI-TOF analysis of isopropanol extracts from *B. licheniformis* DSM13, used as control, revealed the two major peaks at 3251Da and

3020Da corresponding to LicA1 and LicA2 respectively (Figure 24). A similar pattern was observed in the CECT 8237 mass spectra, obtaining the mass peaks around 3042Da and 2500Da, that we associated to LanA1 and LanA2 respectively, based on the predicted amino acid sequence and the hypothetical processing events. Finally and as expected from the genomic comparative analysis no signal in this range of molecular weight was observed in the non-producer but closely related *B. amyloliquefaciens* CECT 8238.

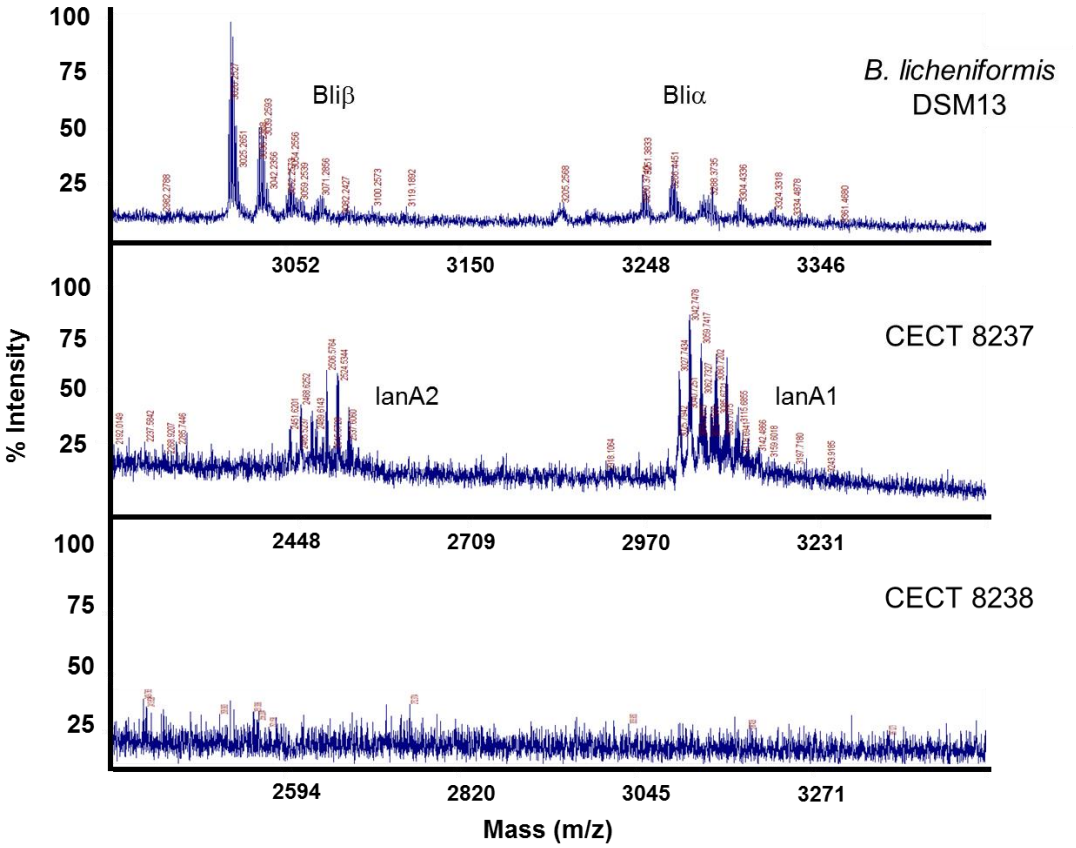


Figure 24. The two peptides related to lichenicidin-like lantibiotic are detected in cell washes of *B. amyloliquefaciens* CECT 8237. MALDI-TOF mass spectra of isopropanol extracts revealed the presence of: the two major peaks at 3251 Da and 3020 Da for Bliα and Bliβ peptides respectively in *B. licheniformis* DSM13, a producer of the lichenicidin lantibiotic (top chart); two groups of peaks of 3042 Da and 2500 Da, hypothetically associated to LanA1 and LanA2 peptides in *B. amyloliquefaciens* CECT 8237 (middle chart) and no traces in this range of molecular weight in *B. amyloliquefaciens* CECT 8238, lacking the lichenicidin related gene cluster (bottom chart).

In a separate study it was reported the detection of lichenicidin in bacterial cell washes and also supernatants from *B. licheniformis* I89 cultures grown in medium M (Caetano et al., 2011). Therefore, we decided to follow the accumulation of the lantibiotic in both fractions in cultures of CECT 8237, using in situ Mass Spectrometry, with no need of chemical extractions. First, we selected the matrix made of α-cyanohydroxycinnamic acid solution-2,5-dihydroxy benzoic acid (CHCA-DHB) among others, (CHCA), sinapinic acid (SPA) (DHB) or DHB-SPA, as the most appropriate for the detection of the molecular weight ranged between 2000 to 4000Da (Annex Figure 4) where the two peptides are located.

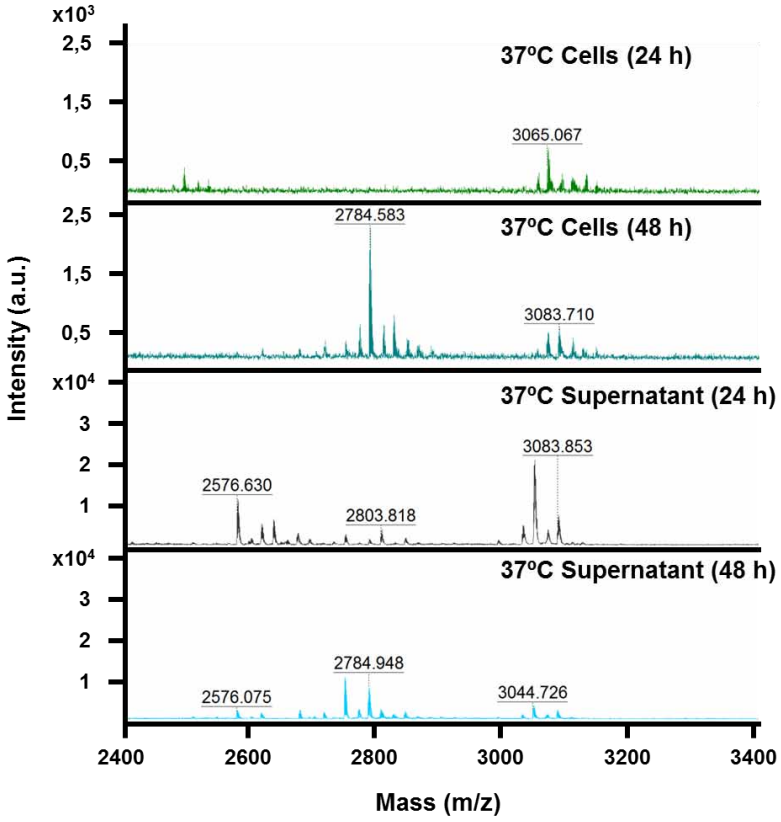


Figure 25. The two peptides related to lichenicidin-like lantibiotic are detected in cell washes and supernatants of *B. amyloliquefaciens* CECT 8237 grown in medium M. MALDI-TOF mass spectra contained the two groups of fragments above 3042 Da and 2500 Da hypothetically associated to LanA1 and LanA2 peptides in bacterial cells (top charts) and cell-free supernatants (bottom charts).

The two patterns of peaks associated to the peptides LanA1 and LanA2 were detected in both fractions, although the major presence of both peptides appeared associated to the supernatants in comparison to bacterial cells. It was also noticeable that LanA2 (2500 Da) was not detected in cells at 24 h while LanA1 (3042 Da) was present in both fractions at 24 h and 48 h.

Post-translational modifications of the two peptides that constitute the lantibiotic in *B. amyloliquefaciens* CECT 8237

Once determined the presence of the two peptides associated to a lichenicidin-like lantibiotic, we asked of their post-translational modifications responsible for the final mature and active molecules. To answer this, we combined the predictions made in comparison to related lantibiotics (Figure 23) and the MALDI-TOF fragmentation patterns of the two major peaks 3042Da and 2500Da corresponding to LanA1 and LanA2 respectively. Between the two modes of fragmentation used (CID OFF and CID ON) the CID OFF mode let us correlate the mass values and the corresponding modifications of the amino acid sequences, and propose the final structure of each mature peptide (Figure 26 and Figure 27).

The final amino acid sequence of LanA1 was identical to the anticipated in the in silico predictions (Figure 23A), however, the study of the fragmentation pattern was not resolutive enough to discriminate among four possible final structures (Figure 26B). In the four proposals, we might accommodate the three thioeter rings predicted in comparison to the orthologous LicA1, but with slight modifications (Figure 26A): the residues Cys1 and Ser7, instead of the Ser7 and Cys17, appeared to be involved in the first thioester ring. In addition to this, in the second model, an additional disulfide bond was accommodated between the residues Cys8 and Cys17.

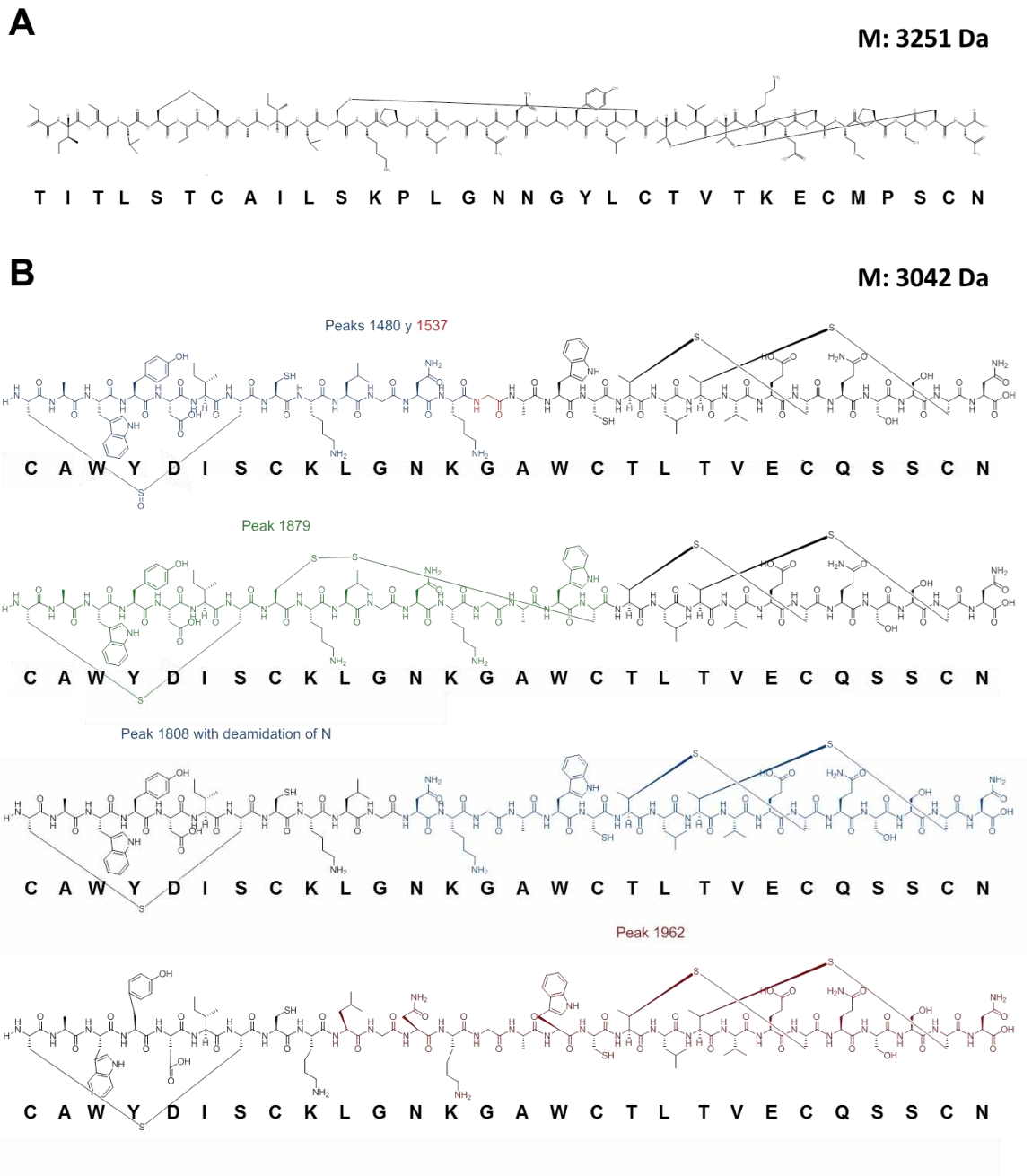


Figure 26. Structural representation of the mature peptide of lantibiotics as derived from mass spectrometry fragmentation analysis. A. Final structure of the peptide Bli1 of *B. licheniformis* DSM13 including the thioeter rings, and the corresponding molecular mass. B. Structural proposals of the mature LanA1 peptide of CECT 8237. The colour-code represents the correlation between the peak obtained in the CID OFF fragmentation and the associated sequence.

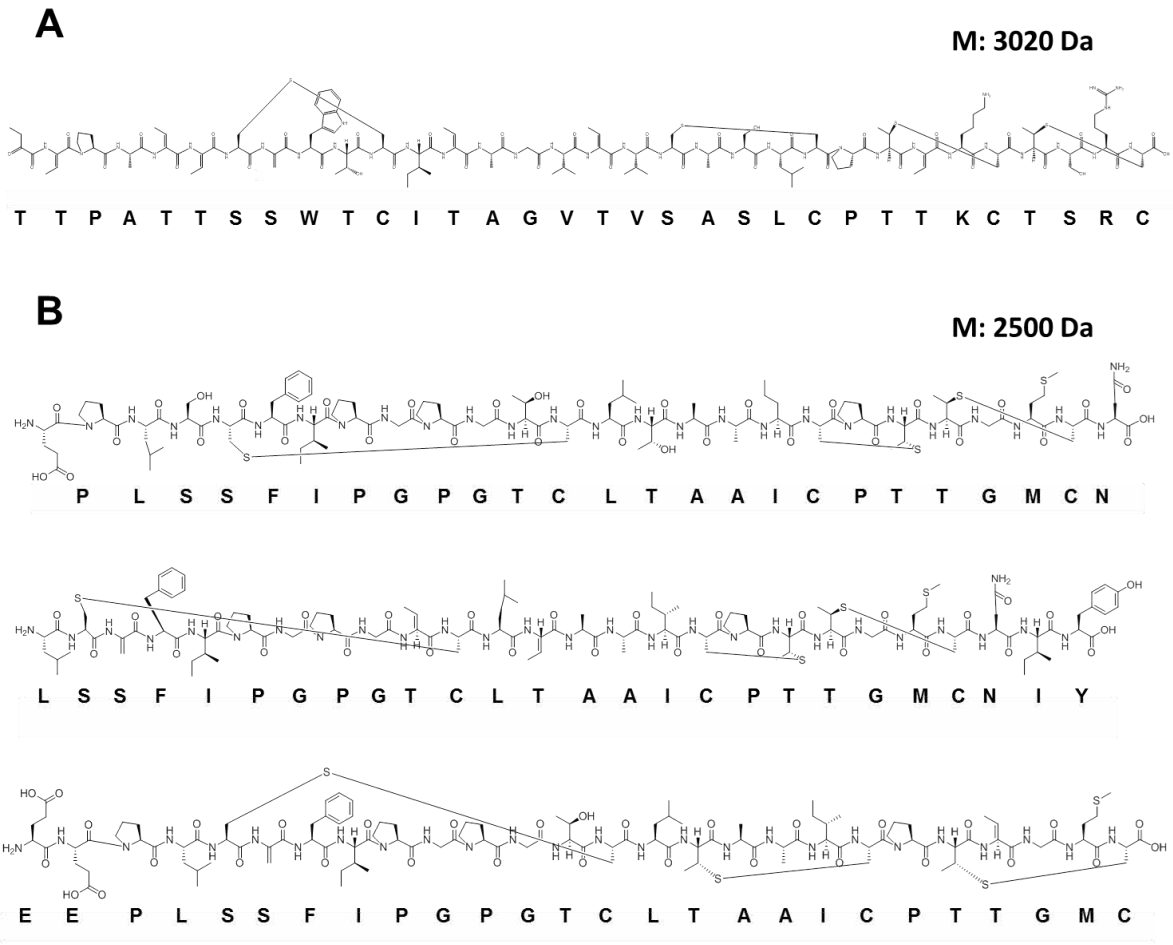


Figure 27. Structural representation of the mature peptide of lantibiotics as derived from mass spectrometry fragmentation analysis. A. Final structure of the peptide Bli2 of *B. licheniformis* DSM13 including the thioeter rings, and the corresponding molecular mass. B. Structural proposals of the mature LanA2 peptide of CECT 8237.

Two main reasons limited our capacity to analyse the peptide Lan2: one, the low intensity of the mass traces associated to the major peak (2500Da) used for the fragmentation analysis, and two, the CID ON fragmentation mode, which generated smaller fragments compared to the OFF mode (LanA1). In addition, the low conservation at amino acid level with the orthologous LicA2 of *B. licheniformis* (Figure 27A) introduced a new level of difficulty to properly interpret the results and make structural predictions. Besides all, we

could identify the mass peaks corresponding to single amino acids as well as some tandem amino acids sequences that permitted us to assemble the final sequence and propose some models of the mature peptide LanA2 (Figure 27B). The three proposals have in common some of the thioeter bonds, Cys19-Thr21 and Thr22-Cys25 in first and second models, Ser4-Cys13 in second and third proposals and other variable thioether bonds implicating Ser5-Cys13 in the first structure while in the third the rest of amino acid residues involved are Thr15-Cys19 and Thr21-Cys25 (positions considering the initial nomenclature derived from propeptides, in which the putative additional proteolysis is taken into account). These models are similar to those anticipated in our in silico predictions, and compatible with the mass of the fragments obtained after the fragmentation. The compatibility of these assumptions is correlated with the amino acid sequences indicated under the corresponding structures. Other variation of the three models is the final amino acid sequence of the mature peptide. The fact that models one and three missed some amino acids on the C-terminal of the peptide, a modification never reported for lantibiotics, led us to propose model two (that contain all residues) as the most probable, however, this is an end that needs further investigation.

The expression of the lantibiotic structural genes is higher at the stationary phase of growth

Besides the final destination of the peptides, we wanted to get insights on the kinetics of their expression. As detailed in material and methods, *B. amyloliquefaciens* CECT 8237 was grown in medium M, optimized for lichenicidin production in *B. licheniformis*, at 30ºC or 37ºC and agitation, and samples were taken at time points representative of each phase of growth for RNA extractions and qRT-PCR analysis using the synthetic genes as reporters (Figure 28).

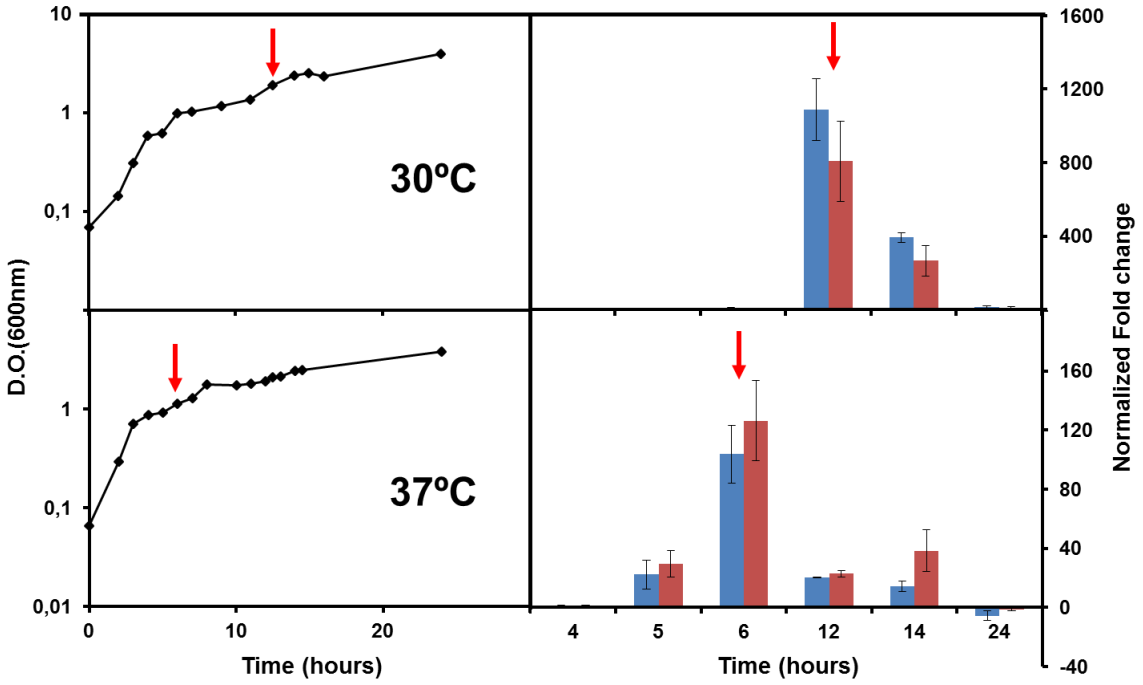


Figure 28. The peptides associated to the lichenicidin-like lantibiotic are largely expressed at the entry of the stationary phase of growth. Quantitative real-time PCR was done to study the kinetics of expression of the synthetic genes of the two peptides *lanA1* and *lanA2* in *B. amyloliquefaciens* CECT 8237.. Relative expression levels obtained in qRT-PCR assays corresponding to RNA isolated from cultures at 30ºC (right chart, top) or 37ºC (right chart, bottom) respectively. For comparison, every value of expression was divided into the data obtained at 5 h or 4 h for experiments at 30ºC or 37ºC respectively. The corresponding kinetics of growth at each temperature is represented in left charts. *lanA1*, blue boxes, and *lanA2*, red boxes. Red arrows indicate the time point of maximum level of expression.

The relative expression levels of both genes were maximum when the cultures reached the stationary phase of growth, 12 h or 6 h at 30ºC or 37ºC respectively, and decreased progressively to minimum levels at 24 h (Figure 28). Apparently, it could be said that the levels of expression of both genes were higher at 30ºC than at 37ºC. However, the increase of relative expression values relies on the reference data, 4 h of incubation at

37ºC or 5 h of incubation at 30ºC, which may lead to variations in the final value, thus, we could not conclude a clear effect of the temperature.

The spatial distribution of the lantibiotic in biofilms of *B. amyloliquefaciens* CECT 8237

Considering the ability of these strains to form biofilms, we asked of the expression profile of the lantibiotic in static growth in medium M broth or solidified (used in previous assays of expression) and 30ºC. First, we observed that these growth conditions promoted the formation of biofilms in the form of pellicle in liquid broth or colonies with distinctive morphological features, as observed in the biofilm inducible medium MSgg, especially at 37ºC (Figure 29, top). In the pellicle formation assays (Figure 29, bottom), M medium induced the formation of pellicles visually less thick than MSgg.

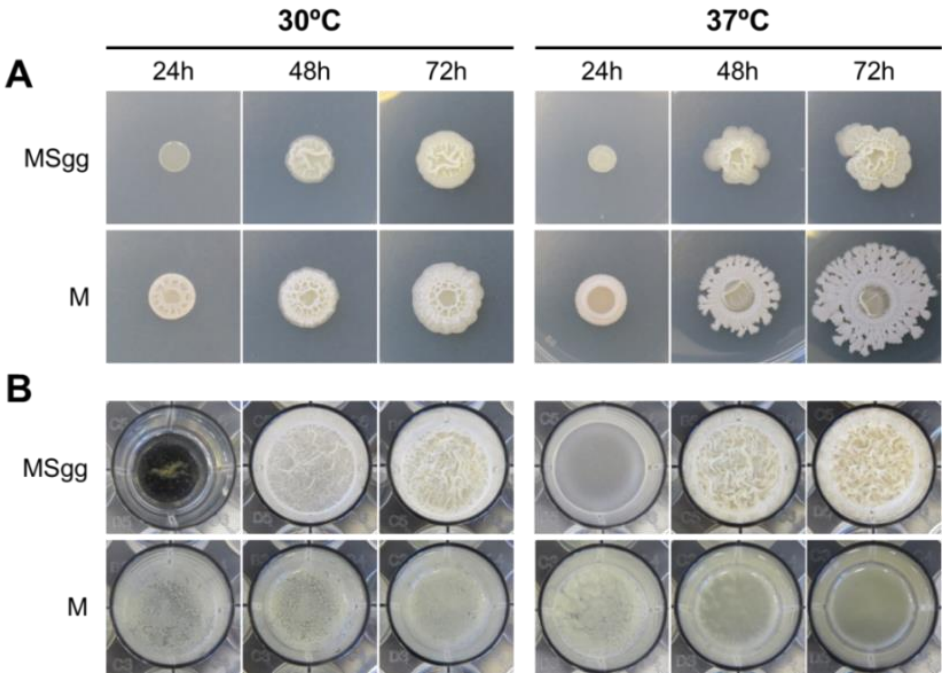


Figure 29. *B. amyloliquefaciens* CECT 8237 forms biofilms with distinguishable features in the lichenicidin inducer medium M. Biofilm formation was evaluated as: A. Colony morphology in medium M or MSgg agar or B. Pellicle formation in M or MSgg broth and incubation at 30ºC or 37ºC.

Demonstrated the ability of CECT 8237 to develop biofilms in the lantibiotic inducer medium M, we studied the production of the lantibiotic in situ using the MALDI-TOF analysis previously optimized (Annex Figure 4). The two differentiated fractions, spent medium or pellicles, of static cultures of CECT 8237 in M medium broth were considered in this analysis. Given our previous finding on agitated cultures (Figure 25), only supernatants of these fractions were analysed (Figure 30). Even that present in both fractions, larger amounts of lantibiotic were detected in spent medium compared to pellicles, a difference that became more evident at 30ºC (Figure 30A).

On the basis of these findings, we asked of differences in the gene expression between cells of the spent medium and those embedded in the pellicles doing qRT-PCR analysis using the primers that targeted the synthetic genes (Figure 31). First, we estimated the bacterial density and percentages of spores in each fraction and each temperature. The total bacterial population was of $2,5*10^7$ CFU/ml in both fractions, and the percentage of sporulation of 16% in pellicles and 1% in the spent medium, at 30ºC and 24 h. After 48 h, the bacterial counts stayed at $1,2*10^7$ CFU/ml (32% sporulated) in the spent medium, and raised to $3,7*10^7$ CFU/ml (97% sporulated) in pellicle. In biofilms at 37ºC and 24 h, the bacterial counts were $1,6*10^8$ CFU/ml in spent medium and $2,05*10^8$ CFU/ml in pellicle, and the percentages of spores of 24% and 48% respectively.

In the qRT-PCR analysis we also included: *srfAA*, involved in the synthesis of surfactin and *ituA* involved in the synthesis of iturin A. The level of expression of the structural genes for lantibiotic production was higher in pellicle associated cells at 24 h or 48 h at 30ºC (Figure 31). This was contrary to the expression of the other secondary metabolites surfactin and iturin that was more evident in cells of the spent medium. The structural genes involved in the synthesis of the two-peptide lantibiotic and the gene related to surfactin antibiotic remained more active in the pellicle fraction at 37ºC, however, the gene

that participates in iturin biosynthesis was highly expressed in planktonic cells, as previously observed at 30ºC (Annex Figure 7).

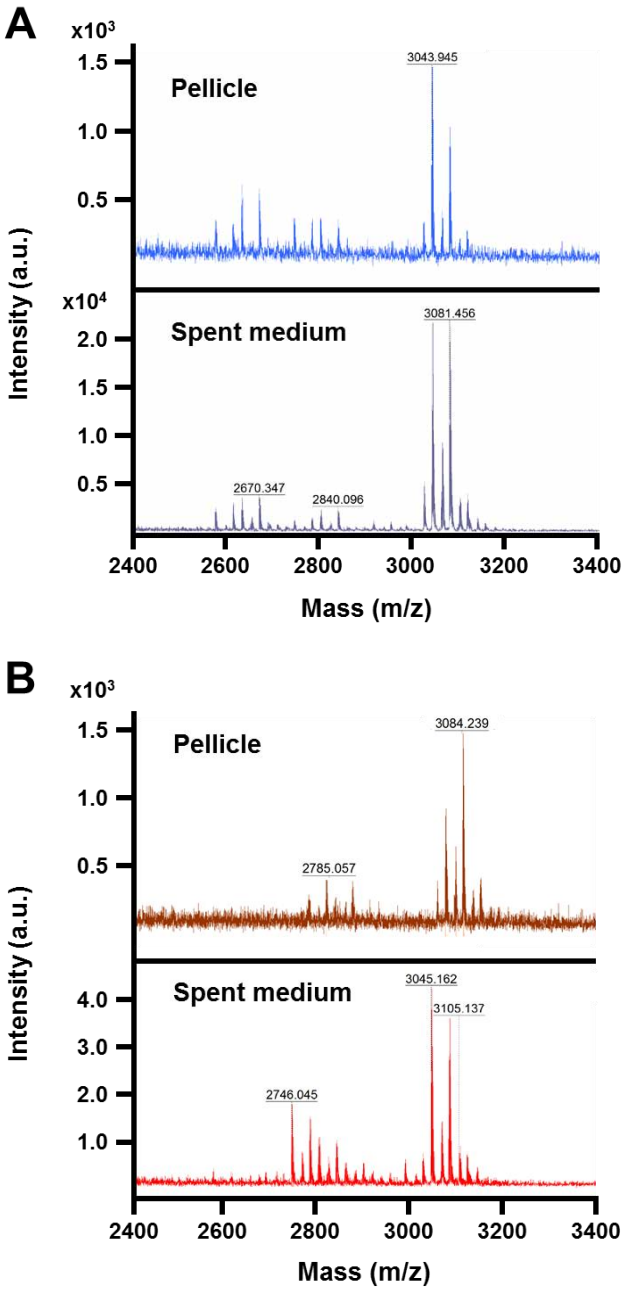


Figure 30. The lantibiotic is detected in larger amounts in the spent medium of biofilms of CECT 8237. Pellicle and spent medium of biofilms at 30ºC (A) or 37ºC (B) were separated and analysed in MALDI-TOF mass spectrometry for the presence of the lantibiotic. MALDI-TOF mass spectra profile ranging from 2400 to 3400Da.

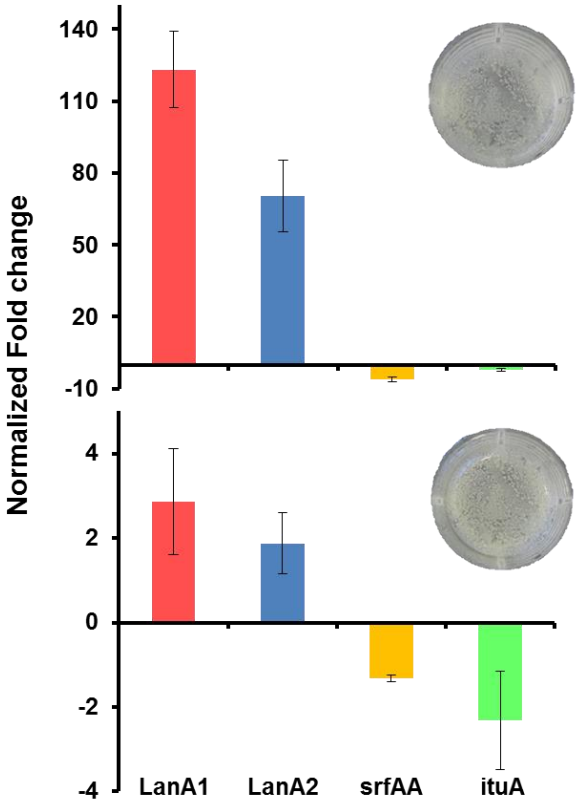
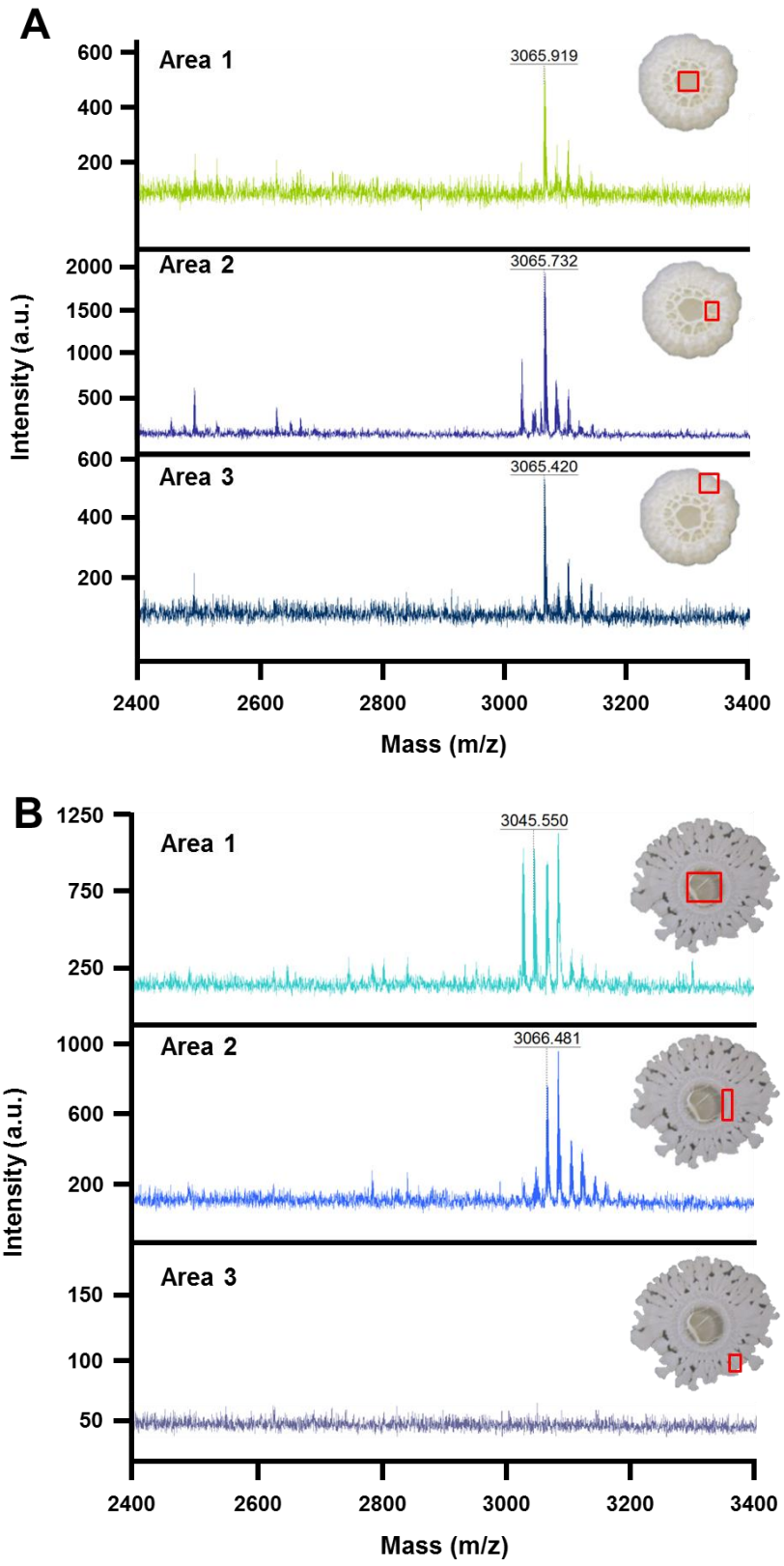


Figure 31. The synthetic genes of the lichenicidin-like peptides are more expressed in biofilms encased cells. The relative expression levels in planktonic and pellicle associated cells of biofilm grown in medium M at 30ºC for 24 h (Top) or 48 h (Bottom) was studied in qRT-PCR analysis. The genes tested were: *lanA1* (red) and *lanA2* (blue), the two peptides of the lantibiotic; *srfAA*, surfactin (yellow); *ituA*, iturinA (green). For comparison, every value of expression obtained in RNA isolated from pellicle associated cells was divided into the corresponding data obtained from planktonic phase in each condition of time and gene tested.

On the other hand, the same chemical analysis was done to determine the lantibiotic distribution in bacterial colonies. In particular, colonies of CECT 8237 grown in medium M show three areas morphologically differentiated (30ºC), which contained the lantibiotic, although the highest intensity of the signal appeared in area 2 (Figure 32A, middle chart). Looking at the mass spectra detected at 37ºC (Figure 32B), mass peaks associated to

LanA1 peptide were identified at similar levels in areas 1 and 2, but not in the outmost growth area, which is enriched in motile cells (Wang et al., 2016). Interestingly and contrary to the colony grown at 30ºC, no traces in the range of 2500Da (LanA2) were identified in any of the areas at 37ºC.

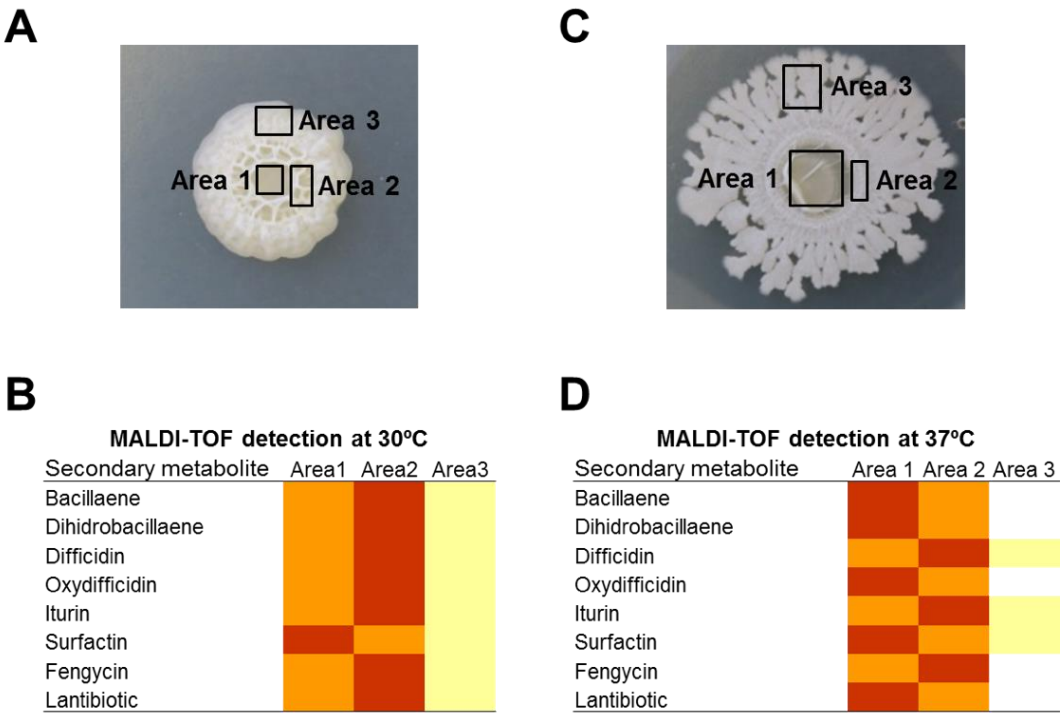
Figure 32. The spatial topography of lichenicidin-like lantibiotic production in biofilms of *B. amyloliquefaciens* CECT 8237 grown at different temperatures. A. MALDI-TOF mass spectra profile of the biofilm grown in medium M at 30ºC. MALDI-TOF mass spectra ranging from 2400 to 3400 Da, and associated to lichenicidin peptides, was detected in the core of the colony (area 1, top chart), in the middle of the colony, (area 2, middle chart) and in the outmost area of the colony (area 3, bottom chart). B. MALDI-TOF mass spectra profile of the biofilm grown in medium M at 37ºC. MALDI-TOF mass spectra ranging from 2400 to 3400 Da were detected in the core of the colony, (area 1, top chart), in the middle of the colony, (area 2, middle chart) but not in the outmost growth area (area 3, bottom chart). The location of these areas in the colony is included (insets).



Considering the ability of this strain to produce a battery of other known secondary metabolites, we decided to investigate their spatial relation with the lantibiotic in colonies at 30ºC and 37ºC (Figure 33A-C). Similar profiles of distribution were observed for the secondary metabolites at 30ºC, except for surfactin that was decreasing from the core to the outmost area of the colony (Figure 33B). The highest concentration of secondary metabolites, including the lipopeptides iturin and fengycin as well as the polyketides bacillaene and difficidin, was associated to area two (one order of magnitude compared to areas 1 and 3, Annex Figure 5) enriched in wrinkles as the morphological feature related to the presence of extracellular matrix (Figure 33A).

The accumulation of other secondary metabolites was also higher at the core and middle areas of the colony at 37ºC, although more variation in the specific profiles of each molecule could be appreciated (Figure 33D). In particular, the polyketides bacillaene and dihydrobacillaene as well as oxydifficidin and surfactin behave like the lantibiotic, more abundant in the core (area 1) and dropping towards the external areas. Difficidin, iturin and fengycin accumulated mostly in area 2, and it is noteworthy that only difficidin, surfactin and iturin are detected in the outmost area of the colony.

Figure 33. The spatial distribution of secondary metabolites varies in colonies of *B. amyloliquefaciens* grown at 30 or 37ºC. A and C, morphology of colonies of *B. amyloliquefaciens* biofilms at 30 or 37ºC respectively, and the three distinguishable areas of analysis: core (area 1), middle (area 2) and outmost (area 3). B and D, schematic representation of the quantitative composition of secondary metabolites between areas at 30ºC or 37ºC respectively. Higher concentration (red), intermediate concentration (orange), lower concentration (pale yellow) and absent (empty).



The two-peptide lantibiotic is expressed on melon leaves

On the basis of the previous results, it was demonstrated the synthesis of the two peptides that constitute the lantibiotic in CECT 8237 at transcriptional level and the final mature peptides as a result of the modification and processing steps. Then, to verify if this lantibiotic could be an additional feature that participates in the biocontrol processes developed by *B. amyloliquefaciens* CECT 8237, we studied their expression in qRT-PCR analysis in plants.

Melon plants were inoculated with 1 ml of bacterial suspensions (10^8 CFU/ml) of *Bacillus subtilis* subsp. *subtilis* 3610 or *Bacillus amyloliquefaciens* CECT 8237. The bacterial populations were $2{,}14*10^7$ CFU/g (3610) and $1{,}82*10^8$ CFU/g (8237) at 7 days post-inoculation (dpi), and $4{,}95*10^7$ CFU/g (3610) and $1{,}163*10^8$ CFU/g (8237) at 14 dpi. The percentage of spores remained at 40% (3610) and 78% (8237) in each time. The RNA was isolated from three independent samples for every treatment at each time post-

inoculation. *B. subtilis* subsp. *subtilis* 3610 strain produces surfactin, fengycin, bacilysin but not the lipopeptide iturin or the two-peptide lantibiotic, thus, the relative expression levels of the target genes in CECT 8237 strain were compared to the expression in 3610.

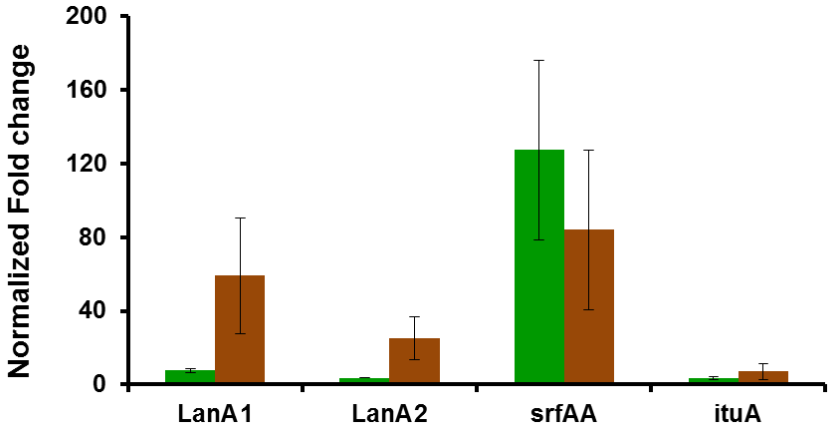


Figure 34. The genes dedicated to the synthesis of the two peptides associated to lichenicidin-like lantibiotic are expressed on melon leaves. qRT-PCR analysis was done with the total RNA isolated from melon plants after seven (green) and fourteen (brown) days post inoculation (dpi) with *B. amyloliquefaciens* CECT 8237 or *Bacillus subtilis* subsp. *subtilis* 3610. The genes studied were: *lanA1* and *lanA2*, two peptides of the lantibiotic; *srfAA*, surfactin; *ituA*, iturinA. For comparison, every value of expression obtained in RNA isolated from CECT 8237 was divided into the corresponding data obtained from 3610 in each condition of time post inoculation and gene tested.

Curiously, surfactin biosynthetic gene was more active 7dpi, but not those implicated in the synthesis of the other secondary metabolites (Figure 34), a finding that could reflect the necessity of this lipopeptide to trigger the formation of biofilm (Zeriouh et al.,2014). After 14 days however, the relative expression of the other tested genes, including those implicated in the synthesis of the lantibiotic, was comparatively higher. These data also suggested that besides the high percentage of cells sporulated (80%), the fraction of metabolic active cells are efficiently producing secondary metabolites with functionality on the defence against competitors or offense against plant pathogenic microbes.

DISCUSSION

DISCUSSION

The biological control is becoming a widespread practical in agriculture because the effectiveness of the biocontrol agents with versatile mode of actions. To select a potential biocontrol agent is essential to consider the relation of the natural enemy and the pathogen, the host plant and the environmental conditions where this interaction will occur. Keeping these principles in mind, several members of the *Bacillus* genus have emerged for their implementation as biopesticides, because of the production of secondary metabolites, the efficient colonization of diverse habitats, the ability to trigger the plant defences and the growth promotion of the host plant (Pérez-García et al., 2011). In particular, *Bacillus* strains belonging to *B. subtilis* and *B. amyloliquefaciens* have been intensely studied to reveal the specific machinery involved in all the above mentioned processes. In previous research conducted in our laboratory, two *Bacillus* strains, CECT 8237 and CECT 8238 (formerly UMAF6639 and UMAF6614 respectively), were selected in the basis of their robust and reliable performance in plants, providing successful control against the powdery mildew or bacterial soft rot diseases (Romero et al., 2004; Zeriouh et al., 2011). As other *Bacillus* species, the production of lipopeptides antibiotics, and the induction of the defence mechanisms in plants are two principal contributions to the plant health. Their reliable and robust biocontrol efficiency motivated this thesis designed to: i) bring to the light putative additional genetic factors that might contribute to the understanding of their outstanding biocontrol skills, and ii) seed future research dedicated to the improvement of their efficiency and reliability in the control of diverse microbial diseases of cucurbits and ideally other crops.

CECT 8237 and CECT 8238 are differentially allocated into the plant-associated *B. amyloliquefaciens* strains

CECT 8237 and CECT 8238 were initially identified as *Bacillus subtilis* based on sequence homology of the 16S rRNA and metabolic profiles (Romero et al., 2004). However, the recurrent microbe misidentification associated with these typical analyses and the recent acquisition of both complete genome sequences, led us to use the most reliable and accurate multi-locus phylogeny analysis (Liu et al., 2013), leading to their reclassification into the close species *B. amyloliquefaciens*. The *B. amyloliquefaciens* species has been subdivided into two distinct subspecies according to Borriss and collaborators, a finding that our results support: CECT 8237 and CECT 8238 grouped with strains of *Bacillus amyloliquefaciens* subsp. *plantarum*, a clade composed of plant-associated and typically isolated from soil or rhizosphere (Borriss et al., 2011). The increase number of complete genome sequences available nowadays are doing more accurate the understanding of the *Bacillus* taxonomy, indeed it was recently suggested that the subspecies *B. amyloliquefaciens* subsp. *plantarum* should be reclassified as a heterotypic synonym of *B. methylotrophicus* (Dunlap et al., 2015). In their study, Dunlap and collaborators found that the type strains of *B. amyloliquefaciens* subsp. *plantarum* and *B. methylotrophicus* shared around the 95% of the genes, while *B. amyloliquefaciens* subsp. *plantarum* and other members of the *B. amyloliquefaciens* taxon did not even cluster together. They also proposed that *Bacillus velezensis* should assign the name to this group given the proven link to those species through phylogenomics studies. Thus, *B. methylotrophicus*, *B. amyloliquefaciens* subsp. *plantarum* (including CECT 8237 and CECT 8238) and *B. oryzicola* should be reclassified as later heterotypic synonyms of *B. velezensis* (Dunlap et al., 2016). With this information in our hands we generated an additional phylogenetic tree, based only on the concatenated gene sequences *nusA-dnaA-spoVG-sigW-sigB,* including

the *B. methylotrophicus* KACC 13105, *B. velezensis* NRRL B-41580T and all those plant-associated *B. amyloliquefaciens* strains (Figure 35). As anticipated, both strains clustered perfectly with plant-associated strains belonging to *B. amyloliquefaciens*, reinforcing the idea that those species should be considered the same. In this work, however, we have maintained the identification of CECT 8237 and CECT 8238 as *B. amyloliquefaciens*.

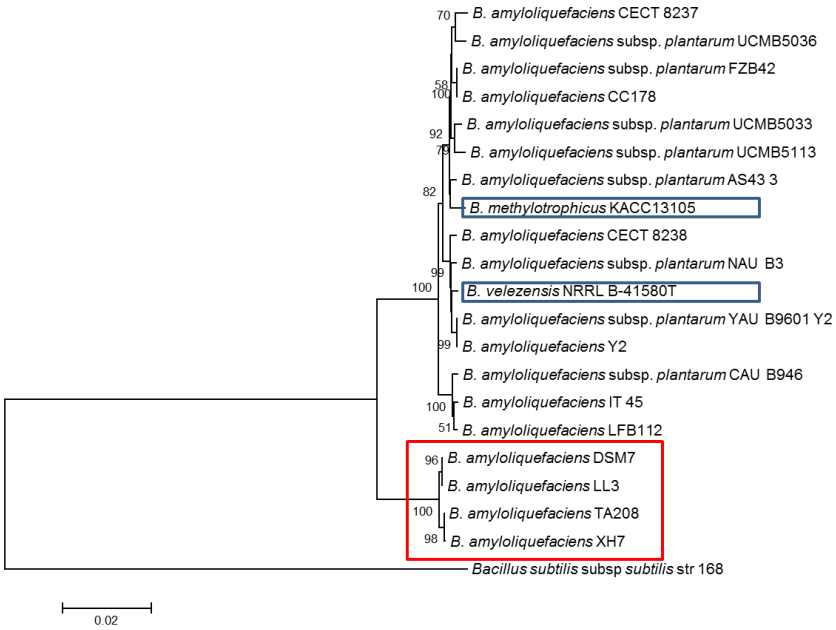


Figure 35. The strains CECT 8237 and CECT 8238 cluster with the group of plant-associated *Bacillus amyloliquefaciens* subsp *plantarum- B. methylotrophicus- B. velezensis*. Phylogenetic analysis of *B. methylotrophicus* KACC 13105 and *B. velezensis* NRRL B-41580T in relation to the plant-associated *B. amyloliquefaciens* strains, including CECT 8237 and CECT 8238. The *B. subtilis* subsp. *subtilis* 168 was used as an out-group. In this analysis, five concatenated genes (*nusA*, *dnaA*, *spoVG*, *sigW* and *sigB*) were handled to build a Neighbour-joining tree using MEGA 5 bootstrap values (10,000 repetitions), which are shown on the branches. The topology was identical for the trees produced by the minimum evolution and maximum parsimony methods. The sequences from all of the strains were extracted from published genome sequences. Squares indicate the location of the KACC 13105 and NRRL B-41580T strains and the most distantly group of *B. amyloliquefaciens* industrially relevant strains.

CECT 8237 and CECT 8238 are reservoirs of unpredicted genetic features with potential impact in bacterial ecology and interaction with plants

The existence of such a clear separation of *B. amyloliquefaciens* into two distinct clades, related to plant and non-plant-associated, led us to explore for distinguishable genes contained in each clade. In the case of plant-associated strains, their exclusive CDS were mainly classified into the categories of metabolism of amino acids, carbohydrates or lipids, synthesis of secondary metabolites, general function prediction or defence mechanisms (Figure 10A). The enhancement of their genomes in genes coding these functionalities might reflect the better adaptation of these strains to the changeable environment of plant surfaces (Carvalhais et al., 2013). One might assume that the presence of endoglucanase, N-acetylglucosamine-6-phosphate deacetylase or maltose phosphorylase enzymes would allow these strains to utilise the diversity of alternative carbon sources available in the exudates of plant surfaces (Table S3). It is also ecologically advantageous to have the genes involved in the mobilisation and bioavailability of scarce micronutrients, such as sulphur and iron, not only as a self-benefit but also for the good of plant health (Miethke et al., 2006). The overrepresentation of genes involved in the synthesis of secondary metabolites would reflect the versatility of these strains to compete with a wide range of microbes that they might encounter in the plant environment.

In our genomic analysis we also identified poorly conserved regions within the *Bacillus* genus, where similar functionalities were detected. In fact, some of these regions overlapped with the previously mentioned, being poorly conserved within the *Bacillus* genus but highly or partially conserved in *B. amyloliquefaciens*, indicating their relevance in the metabolism of this species and, in particular, in plant-associated strains. Between the atypical regions identified in the studied strains and absent in the rest of the *B. amyloliquefaciens* species (Figure 11, outer line, white boxes), it is remarkable the

presence of genes related to fatty acid metabolism, which might be involved in secondary metabolites production, and also the complete machinery that participate in the type IV secretion system in the case of CECT 8238 (Alvarez-Martinez and Christie, 2009). It is not surprising the identification of genes possibly involved in production of new secondary metabolites, as a result of the specific adaptation to the habitat they must colonize and the pathogens they must target. Those atypical regions whose genes have been classifiable, and the genes encoding hypothetical proteins with unknown domains represent potential sources of new functionalities, therefore their future analyses would be desirable as an attempt to clarify the potential biological activity of these strains.

Subtle differences in the cell-to cell communication system might contribute to variations in the developmental programme that ends in the hyper wrinkle biofilms of CECT 8237 and CECT 8238

CECT 8237 and CECT 8238 form biofilms in vitro and on melon leaves, which is relevant for the biocontrol activity (Zeriouh et al., 2014), and our genomic analysis revealed the presence of the core genes dedicated to biofilm formation identified in the type strain *B. subtilis* subsp. *subtilis* 168 (Vlamakis et al., 2013). However, it was compelling the different colony morphology of these strains compared to the type strain 168 or *B. amyloliquefaciens* subsp. *plantarum* FZB42 (Magno-Pérez-Bryan et al., 2015). The formation of biofilm is the coordination of bacterial cell-to cell signalling and expression of structural elements, thus, we first looked at the signalling network as the origin of the different outcomes. In particular, it was analysed the modifications within the operon *comQXPA*, the QS system that control the expression of several genes dedicated to competence or surfactin production among others, in several *Bacillus* strains. As a result, it was found that the strains CECT 8237 and CECT 8238 possessed different pherotypes

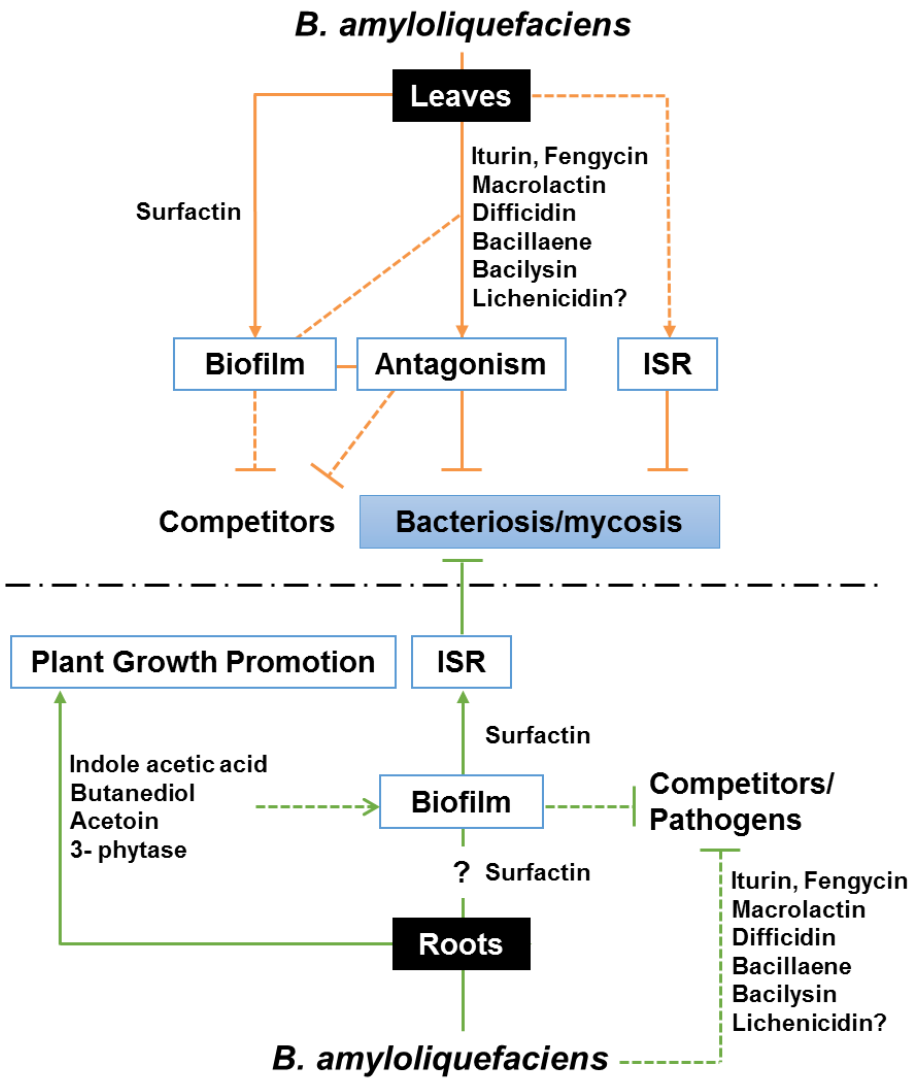
and also the type strain *B. amyloliquefaciens* subsp. *plantarum* FZB42, sharing the last the same pherotype with the *B. subtilis* group. This functional diversification into different pherotypes permits strains to communicate when they belong to the same pherotype but not across different pherotypes; thus, we might assume that this polymorphism emerges as an adaptability response (Stefanic et al., 2012; Dogsa et al., 2014). In the case of CECT 8237 and CECT 8238, these differences might be justified by the fact that they inhabit the same niche (Oslizlo et al., 2015). Therefore, harbouring different pherotypes could be beneficial to coexist and avoid interfering gene expression controlled by QS signals (Tortosa et al., 2001). Moreover, we also found differences in the *rap-phr* operons, which also participate modulating the formation of biofilm, of CECT 8237 and CECT 8238 in comparison with the type strains *B. amyloliquefaciens* subsp. *plantarum* FZB42 and *B. subtilis* subsp. *subtilis* 168 (Figure 14 and Annex Table 5). The fact that an additional aspartate phosphatase (Rap) protein was identified in *B. amyloliquefaciens* strains, but not in *B. subtilis*, proves that the regulation system *rap-phr* characterized in *B. subtilis* is not as similar as anticipated for *B. amyloliquefaciens* species. Other examples of *rap-phr* operons characterized in 168 are the *rapK-phrK* identified in both studied strains but absent in FZB42, the two copies of the gene *rapH* detected in CECT 8237 and the tandem *rapI-phrI* only contained in CECT 8238. These differences are quite striking and open a wide range of possibilities to examine in future studies to reveal the real implication and effect of these Rap proteins in the developmental programme concluding in the formation of biofilms. The structural elements dedicated to the assembly of the extracellular matrix are an EPS, the amyloid-like protein TasA, or the hydrophobic protein BslA, all of them present in both strains. We have not identified bacterial features other than those; however, whether or not additional structural components of the extracellular matrix are present in these strains need to be explored in more detail.

CECT 8237 and CECT 8238 produce a variety of secondary metabolites involved in the ISR and PGP activity

An additional contribution to the plant health is the promotion of the plant growth and induction of the systemic resistance of the plants (ISR), and both strains are promoters of ISR in melon leaves, which seems to be mediated by the activation of the jasmonic and salicylic acid pathways (García-Gutiérrez et al., 2012; García-Gutiérrez et al., 2013). Our genomic analysis revealed the presence of all the genes reported to be involved in those processes (Table 7). In relation to the gene content and identity percentage, there were no differentiable features in comparison with other biocontrol agents belonging to *Bacillus amyloliquefaciens/subtilis* species. Moreover, there were no significant differences in plant-growth promotion between the species included in this study, showing in all cases heavier and more abundant roots of melon seeds than the untreated seeds (Figure 18B-C). Besides the phytohormone indole-3-acetic acid (García-Gutiérrez et al., 2012), these strains are also producers of acetoin and 2,3-butanediol (Farag et al., 2013), as other *Bacillus* species, however, it was interesting the kinetics of their production. First, both volatiles reached their maximum peak at 24 h of growth, with the exceptions of FZB42, which accumulates more acetoin at 36 h, and CECT 8237, which accumulates more acetoin at 12 h and butanediol at 36 h. Second, the amount of butanediol was surprisingly higher (one order of magnitude) in CECT 8237 and CECT 8238 than in the rest of the strains. Nevertheless, the impact of these divergences in the accumulation of both secondary metabolites on their ecology or beneficial contribution to the plant health needs to be confirmed (Figure 36). Future analyses might be directed in two complementary directions: i) to clarify the effect of the purified molecules acetoin and 2,3-butanediol when apply directly to the melon seeds, ii) to use *Bacillus* inoculants of diverse ages attending to the kinetics of the expression of both volatiles. In addition to volatiles, beneficial bacteria

also contribute to plant health as biofertilizers, mobilizing iron and phosphate. The identification of genes related to the metabolism of both micronutrients in the genomes of both strains supports this biological activity and also suggests an ecological advantage against competitors lacking this genetic factors (Miethke et al., 2006). Among these genes we identified the gen *phy* encoding a 3-phytase, an enzyme that in phosphate-starvation conditions contributes to the phytate-associated phosphorus availability in soil and helps eliminating chelate-forming phytate (Fasimoye et al., 2013). Phytate is also abundant in plant seeds, and thus an enhanced processing would lead to higher rate of plantlet growth; in addition, both strains promote the growth of the root system in the post-germination stage of the seeds, two lines of evidences supporting the investigation of the involvement of the phytase enzyme.

Figure 36. Secondary metabolites produced by *B. amyloliquefaciens* CECT 8237 and their possible implication in the mechanisms of actions related to biocontrol. Aboveground, interactions occurring in the phyllosphere of the plants (orange lines). Belowground, interactions occurring in the rhizosphere of the plant (green lines). Continuous lines indicate proved implication while discontinuous lines denote unproved implication.



The operon for the synthesis of lichenicidin in *B. licheniformis* is among the arsenal of secondary metabolites produced by CECT 8237

We do know from previous analyses that antibiosis is the most relevant mode of action of these strains in the control of plant diseases. Separated studies demonstrated the relevant contribution of lipopeptides surfactin, iturins and fengycins to the antagonism against bacterial and fungal pathogens of melon crops (Romero et al., 2007b; Zeriouh et al., 2011; Zeriouh et al., 2014). Other studies indicated that around the 9% of the genome of *Bacillus*

is dedicated to the production of secondary metabolites, thus we scanned the genomes of CECT 8237 and CECT 8238 with the antiSMASH software and found those gene clusters dedicated to the production of known secondary metabolites of *Bacillus* and other still unknown (Figure 14, red boxes). Chemical analysis of bacterial cultures let us confirm the presence of bacillibactin siderophore, dipeptide bacilysin and the polyketides macrolactin, difficidin, bacillaene and dihydrobacillaene, all previously reported and well-characterized in other biological control strains of *B. amyloliquefaciens* species (He et al., 2012; Yuan et al., 2012; Um et al., 2013; Wu et al., 2015). However, the real implication of those molecules in the bacterial ecology of CECT 8237 and CECT 8238 or in their mechanisms of defense, against competitors, and/or offense, against pathogens, is a pending issue to deal with in future studies (Figure 36). Furthermore, the identification of genomic regions coincident with atypical regions, poorly conserved within *B. amyloliquefaciens* species (Figure 11, outer line, white boxes), and with gene clusters hypothetically implicated in the synthesis of new antibiotics (Figure 14, red boxes), attracted our attention. We identified the complete genetic machinery necessary to produce a novel non-ribosomally synthesized peptide in CECT 8237 (Figure 20A), and only found in the genome of *B. amyloliquefaciens* LFB112 (Cai et al., 2014). On the other hand, it was found a group of genes with high probability to produce an unknown non-ribosomally synthesized peptide in CECT 8238 strain (Figure 20B). This region had been previously reported in *B. amyloliquefaciens* subsp. *plantarum* NAU-B3, YAU B9601-Y2 and *B. amyloliquefaciens* Y2 strains, which are all associated to plants (He et al., 2012). The functionality and structures of the molecules potentially synthesized by these hypothetical gene clusters have not been demonstrated yet. In this thesis we decided to focus on the study of a gene cluster identified in CECT 8237, and only present in other recently published genome of *B. amyloliquefaciens* XK-4-1, and obviously *B. licheniformis* 9945A. It is important to note,

that the genome sequence of the strain XK-4-1 is available in scaffolds, so the identification of the genes was performed by *blastp* comparative analysis, and additional comparison analysis as gene arrangement could not be affordable.

A detailed analysis of this region permitted the identification of relatively conserved domains related to the synthesis of a new lantibiotic similar to lichenicidin in *B. licheniformis* DSM13 (Dischinger et al., 2009). Attending to diverse parameters, this genomic region coincident with AR21 in CECT 8237 was suggested to be acquired through horizontal transfer, which could justify its presence in different species. This is not unusual, given that other lanthipeptides associated to genomic islands were previously reported in the lantibiotics mersacidin and amylolysin produced by other *B. amyloliquefaciens* strains (Dias et al., 2015). Interestingly, it has also been suggested that the maintenance of these regions devoted to lantibiotic production is related to functional adaptation, due to the detection of genes involved in the transport and immunity of those lanthipeptides but not the biosynthesis machinery in other related *B. amyloliquefaciens* strains (Dias et al., 2015).

We demonstrated the expression of the genes of this cluster into three polycistronic units, which is consistent with the expression pattern of other lantibiotics (Willey and Donk, 2007). Thus, the genes related to synthesis and modification usually define one of these units, sometimes including the gene *lanT*, involved in the transport and processing of the signal peptides. The second unit is normally constituted by genes dedicated to transport and exportation of the peptides and thus conferring auto-immunity to the producer cells. The third unit is formed by *lanP* which encodes a serine protease supposedly involved in further processing of the peptides, and another transporter gene, both suggestively participating in the exportation of the mature peptide after these successive modifications. In other lantibiotics, an additional transcriptional unit can be identified when the regulation

is mediated by the action of a response regulator and a signal kinase, which constitute an operon within the lantibiotic gene cluster (Willey and Donk, 2007). Attending to the genes related to immunity and exportation of this novel lantibiotic, two sets of immunity genes, *lanF1G1E1* and *lanF2G2E2*, have been identified; one of them similar to that detected in lichenicidin and mersacidin, and an additional ABC transporter similar to that identified for gallidermin protection. However, no gene within the cluster was homologous to *lanI*, encoding a membrane associated lipoprotein that constitute another immunity system (Khosa et al., 2016). Considering that the studied lantibiotic is composed by two peptides, and that LanI proteins display very low sequence similarity, due to its lantibiotic binding specificity, it could be hypothesized that the gene annotated as hypothetical protein within this gene cluster might be associated to this second line of defence (Alkhatib et al., 2012). In fact, a similar scenario was previously proposed for subtilomycin, identifying an ORF with no significant homology to other sequences in the database and being designated as *subI*, so assuming to be involved in self-protection against this lantibiotic (Barbosa et al., 2015). Alternately, the two LanEFG ABC transporters identified in the studied lantibiotic cluster could supply that function given that its presence was previously reported in the two-peptide haloduracin lantibiotic (McClerren et al., 2006; Lawton et al., 2007).

Further chemical analysis of bacterial cultures proved the active production of the two peptides that constitute this new lantibiotic. Attending to the prepeptides sequences in comparison with other lantibiotic already characterized, we were able to distinguish the mass traces of 3042Da and 2500Da corresponding to LanA1 and LanA2 respectively. In addition, those groups of peaks were characterized by an increment of 16Da due to the oxidation of the thioeter bridges. A similar pattern was appreciated in the *B. licheniformis* DSM13, the reported producer strain of the original lichenicidin (Dischinger et al., 2009; Caetano et al., 2011). Based on our analysis, it appears that of the two peptides, the most

variable is the peptide LanA2, which could determine the effectiveness of the antimicrobial activity exerted against the target cell. For instance, in the case of haloduracin lantibiotic, it has been reported and experimentally demonstrated that this antibiotic requires the formation of a complex by binding Halα to lipid II at the cell surface, preventing peptidoglycan biosynthesis, while Halβ recognizes and binds to the Halα:lipid II complex, causing membrane permeabilization along with depolarization and release of K+ ions, to exert its full activity (Barbosa et al., 2015). Thus, no evidence has been proposed before that LanA2 is responsible of specificity action against target cells, but it might define how aggressive the activity is exerted through pore formation. Based on what has been described for the original lichenicidin, we tentatively propose a synergistic mode of action of the two peptides in CECT 8237. In order to characterize the antimicrobial activity of this lantibiotic, we tried their purification in two complementary ways: i) heterologous expression of entire gene cluster including structural, modifying, immunity and transport genes in the surrogate *B. subtilis* subsp. *subtilis* 168*,* lacking any traces in the range of molecular weight of the peptides, or, ii) purification of both peptides directly from the producer strain. However, we failed in both strategies so far, and it represents an issue that we are currently managing. Once both peptides have been purified, it will be possible to evaluate the antimicrobial activity spectrum associated to this novel lantibiotic, leading to disclose the potential target cells and/or its implication in the bacterial ecology (Figure 36). In most of the lantibiotics already characterized, the activity relied on Gram-positive bacteria such as *Staphylococcus aureus, Micrococcus luteus Enterococcus faecium, B. cereus,* or *Listeria monocytogenes,* some of those are opportunistic pathogens which can cause human diseases. So, depending on the activity spectrum shown in the case of the lantibiotic produced by CECT 8237, it will be determined its functionality in either cell

offense against pathogens or defence against external competitors, both processes with relevance in biological control (Figure 36).

Our qRT-PCR analysis showed an increase in the expression of the synthetic genes of both peptides at the entry of the stationary phase, and further decrease to minimum levels irrespective to the temperature (Figure 28). The expression of the lantibiotics nisin, subtilin or mersacidin is regulated either by: i) a two-component regulatory system, consisting of a receptor histidine kinase (LanK) and a transcriptional response regulator (LanR), or ii) growth phase dependent mechanisms (Guder et al., 2002; Lubelski et al., 2008). In the case of nisin, the extracellular lantibiotic content acts as an autoinducer of its production, triggering the phosphorylation of LanK, then, the phosphate group is transferred to the response regulator LanR which is bound to distinct promoters triggering the expression of the lantibiotic genes (Spieß et al., 2015). In a similar way, the two-component system spaRK of the subtilin lantibiotic is autoregulated by its own lantibiotic content; however, in this case, it was demonstrated that spaRK is positively controlled by the sigma factor H, SigH, a typical regulator in the late-growth phase, whose transcription is negatively controlled by the general transition state regulator AbrB (Hyungjae and Hae-Yeong, 2011). The former, is not the most probable scenario for the lantibiotic of CECT 8237 given that a similar two-component regulatory system is missed in the gene cluster dedicated to lantibiotic production. However, we have found an orphan regulator, *BAMY6639_14535* locus (Figure 21), which appears to be spread in the gene clusters of other lantibiotics: MutR in mutacin II, EpiQ in epidermin, LtnR in lacticin 3147 or RamR in SapB, all of them regulating positively the production of the corresponding lantibiotics. For instance, the EpiQ regulator activates the transcription of genes involved in the synthesis, post-translational modification, immunity and secretion of epidermin in *Staphylococcus*. In addition, the expression of *epiP* is irrespective to the presence of epidermin, and a

coordinated control is assisted by the Agr quorum sensing system (Kies et al., 2003). Cerecidin is a lantibiotic produced by *B. cereus*, and the gene cluster contains an orphan regulator and also the quorum sensing components comQXPA which suggest their collaborative activity. All these data let us anticipate that the orphan regulator located within the CECT 8237 gene cluster might be implicated in the transcriptional activation of this two-peptide lantibiotic genes related to production and self-protection, and additional cell-density-dependent mechanisms might participate in the fine tune of their expression.

The strength of Mass Spectrometry techniques permits to define the topological distribution of the lantibiotic and other secondary metabolites in biofilms of *B. amyloliquefaciens*

Besides this analysis we studied the expression and presence of the lantibiotic in biofilms, due to its relevance in plant colonization and persistence (Zeriouh et al., 2014). Attending to pellicle formation at the air-liquid interface, two fractions were evaluated, planktonic and pellicle-associated cells. The chemical analysis proved the major accumulation of the peptides in the spent medium compared to pellicle, a finding contrary to the qRT-PCR results that showed a major expression of the lantibiotic synthetic genes in pellicles and regardless of the temperature. Two tentative interpretations to this divergence are: i) the release of the peptides from the pellicle fraction to the spent medium or ii) the expression of the synthetic genes is not representative of the final product, the modified and mature peptides, that require the involvement of additional genes expressed in different transcriptional units. In this sense, the pellicles formed in medium M seem to be less consistent and more relaxed, compared to the commonly used MSgg, which would permit the leak of molecules. To confirm this observation, the analysis in MSgg medium, where more robust pellicles are assembled, might contribute to bring some light to this issue. In

the way around, the fact that the chemical analysis of the other molecules correlate to their genetic analysis between fractions is supportive of the second hypothesis. In addition to this, we observed how surfactin and iturin, two lipopeptides essential in the biocontrol activity of this strain, behaved differentially depending on the temperature. Surfactin was expressed and accumulated mostly in pellicles at 37ºC, while iturin was mostly expressed and accumulated in cells of the spent medium in all conditions. It was previously reported the influence of medium in the robustness of pellicles and surfactin production (Chollet-Imbert et al., 2009), which might be influenced in this case by temperature conditions in the medium M; the characterization of pellicles structure at both situations is an aspect that should be considered in future analyses. The production of the lipopeptides should be also correlated to physiological stage of cells in the pellicle, being the sporulation rate a determinant of metabolites production. Surfactin synthesis and sporulation are controlled by pheromones of the quorum sensing system, which are probably promoted in pellicles (Chollet-Imbert et al., 2009). It seems that the expression of the lantibiotic structural genes might be influenced in such a way by regulatory pathways connected to pellicle but not the final mature peptides, which might follow a different regulatory network than other relevant secondary metabolites. Thus, the more research in this direction will help elucidating the interconnection of these secondary metabolites and how they coordinate the efficient performance of this strain.

In relation to spatial distribution of the lantibiotic within the colony architecture, at 30ºC, the highest lantibiotic concentration was associated to the medium area of the colony, prone to form wrinkles; however at 37ºC, only one of the peptides was detected, in a similar concentration, in the core and middle areas of the colony but not in the outmost growth area. In general, motile cells are principally migrating to the surrounding areas of the colony exploring for novel source of nutrients, which are starved in older areas of the

colony (centre and middle) (Wang et al., 2016), a condition that might trigger the secondary metabolism. Accordingly, the analysis of other metabolites provided similar results to the lantibiotic: at 30ºC, the highest concentration of the lipopeptides iturin and fengycin as well as the polyketides bacillaene and difficidin, was detected in area 2 (middle), however surfactin is highly accumulated in the area 1 (core of the colony) where cells are supposed to be more static (Figure 33B). This is not a revealing finding considering that surfactin has been demonstrated to be essential in biofilm formation (Zeriouh et al., 2014). At 37ºC, the spatial distribution of metabolites was more variable, but still the highest accumulation occurs in the core (bacillaene, dihydrobacillaene, oxydifficidin and surfactin) and middle area (difficidin, iturin and fengycin) of the colony (Figure 33D). In this case, most of the metabolites, except difficidin, surfactin and iturin, were absence in area 3, likewise the lantibiotic production. It is interesting to notice that the concentration of the lantibiotic was comparatively lower than the rest of molecules in all scenarios. The fact that we count with a spatial pattern distribution of metabolites will help understanding their specific functionality and also provide a fingerprint that will permit to explore variations in response to a myriad of situations.

That a potential biocontrol agent efficiently inhibits pathogens growth *in vitro* not always correlates with the performance in natural settings, becoming a source of frustrating attempts. The strain CECT 8237 is well known to control fungal and bacterial diseases, by producing at least the lipopeptides in plants: i) surfactin that contributes to trigger the assembly of biofilms and ii) iturins and fengycin which target the pathogens. In this thesis, we wanted to evaluate the expression of the lantibiotic in melon leaves, using two approaches, Mass Spectrometry and genetic analyses. The qRT-PCR analysis confirmed our previous findings, surfactin appears to be highly expressed at 7 dpi, when *Bacillus* cells start to produce extracellular matrix, and iturin is highly expressed at 14 dpi (Zeriouh

et al., 2014). The expression of the lantibiotic was detected at both times, although the relative expression was higher at later stages, 14 dpi. In the absence of chemical data, it could be speculated that if any, the lantibiotic might exhibit its functionality on melon leaves.

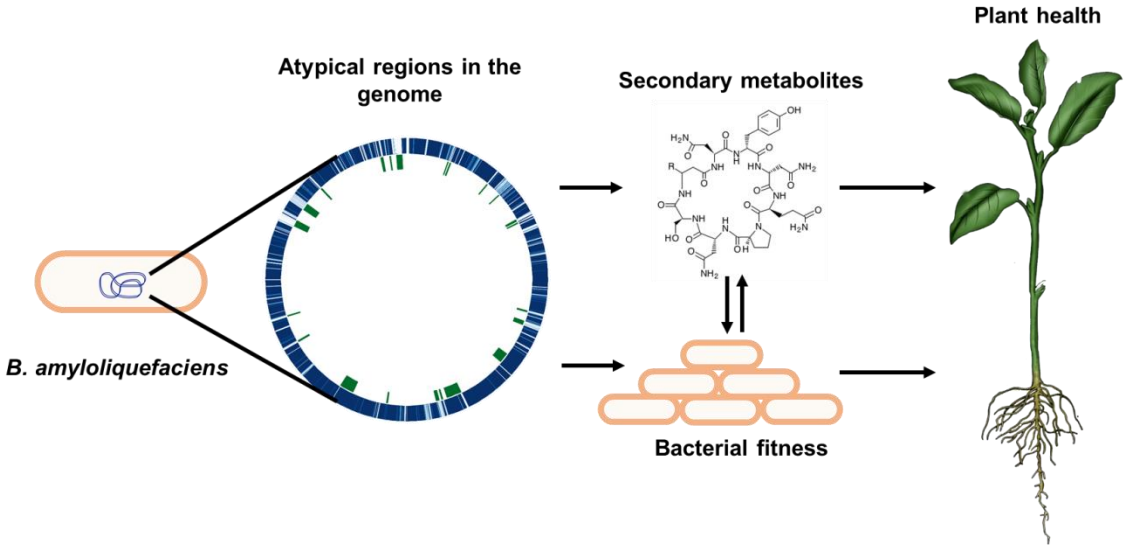


Figure 37. Atypical regions within the genomes of CECT 8237 and CECT 8238 could be sources of potential genes involved in the production of new secondary metabolites or novel bacterial features related to bacterial fitness, as biofilm formation, both contributions to the interaction and efficient protection to the host plant.

In summary, it can be stated that the analysis of the genome sequences of the biocontrol strains *B. amyloliquefaciens* CECT 8237 and CECT 8238, together with the *ad hoc* scripts specifically designed were the clue to subtract the information related to their biocontrol skills. Our analysis confirmed that these strains, similar to other *B. amyloliquefaciens* strains associated with plants, have specific genetic features that are absent in industrially relevant strains and would explain their improved performance in these habitats. The consumption of diverse carbon sources and the production of secondary metabolites

prevail as the most outstanding differences that ensure the ecological advantage over competitors. In addition, it has been reinforced the versatility in the modes of action that participate in the biocontrol capability of *B. amyloliquefaciens* CECT 8237 and CECT 8238, based on: i) multicellularity (biofilm formation and motility) that provides support to colonisation and persistence on plant surfaces, ii) production of phytohormones and MAMPs that trigger the plant growth and ISR and iii) production of secondary metabolites, which might have double functionality in direct antagonism of plant pathogens and possible competitors, or in bacterial fitness and adaptability. In particular, the wide variety of secondary metabolites whose production have been experimentally demonstrated in this thesis, together with the previous results, exhibit the importance of antagonism based on the antibiotics production in the biocontrol activity of *B. amyloliquefaciens* CECT 8237 and CECT 8238, either by already well-characterized molecules or novel compounds. In this sense, the identification of a new lantibiotic in CECT 8237 was possible due to the bioinformatics analyses carried out in this thesis that, in combination with other techniques, such as mass spectrometry, has led to its characterization and permitted partially decode its behaviour in different experimental conditions. Thus, the performance in standing conditions in vitro and in planta was related to other relevant secondary metabolites like surfactin and iturin lipopeptides whose direct implication in biocontrol processes have been extensively demonstrated. Finally, there are atypical regions in the genome of both strains that represent new and exciting areas to investigate additional genetic features with potential implications in bacterial fitness and contributions to plant health (Figure 37).

CONCLUSIONS

CONCLUSIONS

1. The multigene phylogenetic analysis is a more reliable and accurate analysis and has permitted the identification of the studied strains CECT 8237 and CECT 8238, formerly classified as *Bacillus subtilis*, as *B. amyloliquefaciens* species related to plant-associated strains.

2. The genomic analysis has contributed to the differentiation of *B. amyloliquefaciens* species into two clearly separate clades, which is correlated with a specification in gene functionality, acquiring the plant-associated strains skills in competition and environmental adaptability in comparison with industrially relevant strains.

3. The identification of morphological differences in biofilms of *B. amyloliquefaciens* CECT 8237 and CECT 8238 and other *Bacillus* species are indicative of possible divergences either in the genetic composition or their fine regulatory pathways.

4. *B. amyloliquefaciens* CECT 8237 and CECT 8238 are triggers of the immune response and growth of plants, which could be justified, apart from lipopeptides production, by the volatiles compounds 2,3-butanediol and acetoin.

5. *B. amyloliquefaciens* CECT 8237 and CECT 8238 are biological factories of secondary metabolites, that have been previously characterized in other biological control agents, and could be involved in the biocontrol abilities of both strains

6. *B. amyloliquefaciens* CECT 8237 contains a gene cluster dedicated to the production of a new two-peptide lantibiotic, originally described in *B. licheniformis*, suggestive of horizontal transference of genetic material in plant-associated strains as adaptive responses to the specific environmental conditions.

7. The lantibiotic and other well-characterized secondary metabolites production by *B. amyloliquefaciens* CECT 8237 in standing conditions seems to be associated with the organization level of the biofilms, which could indicate, apart from different regulation pathways, a specific functionality of these molecules in either cell defence or offense against external competitors.

8. There are many atypical regions in both strains, poorly conserved along the *Bacillus* genus, whose functionality was no possible to reveal. These new and exciting areas represent a challenging work to investigate in future studies, leading to discover possibly additional genetic features with potential implication in bacterial fitness and contribution to plant health.

CONCLUSIONES

CONCLUSIONES

1. Los análisis filogenéticos multigénicos han permitido una identificación y clasificación taxonómica bacteriana, más fiable y precisa, permitiendo la identificación de las cepas de estudio CECT 8237 y CECT 8238, clasificadas anteriormente como *Bacillus subtilis*, como pertenecientes a la especie *B. amyloliquefaciens*, y en especial, más relacionadas filogenéticamente con las cepas asociadas a plantas.

2. La separación en dos grupos claramente diferenciados filogenéticamente dentro de la especie de *B. amyloliquefaciens* se correlaciona con un enriquecimiento en determinadas categorías funcionales, aumentando el contenido génico asociado a competencia y adaptabilidad al medio en las cepas asociadas a plantas, en comparación con aquellas cepas relevantes a nivel industrial.

3. La identificación de diferencias morfológicas en los biofilms de *B. amyloliquefaciens* CECT 8237 y CECT 8238 y otras especies de *Bacillus* es indicativo de posibles divergencias tanto a nivel genético como en sus rutas de regulación.

4. *B. amyloliquefaciens* CECT 8237 y CECT 8238 son inductores de la respuesta de defensa y la promoción de crecimiento de la planta hospedadora, lo cual podría estar mediado, además de por la producción de lipopéptidos, por la liberación de los compuestos volátiles 2,3-butanodiol y acetoína.

5. *B. amyloliquefaciens* CECT 8237 y CECT 8238 son productores de una gran variedad de metabolitos secundarios, caracterizados previamente en otras cepas empleadas como agentes de biocontrol, y que pueden estar implicados en la actividad antimicrobiana de estas dos cepas.

6. *B. amyloliquefaciens* CECT 8237 contiene un clúster génico dedicado a la producción de un nuevo lantibiótico, originalmente descrito en *B. licheniformis*, lo cuál podría sugerir que la transferencia horizontal de material genético por parte de las cepas asociadas a plantas surge como una respuesta adaptativa al entorno medioambiental.

7. La producción del lantibiótico y otros metabolitos secundarios de función conocida por parte de *B. amyloliquefaciens* CECT 8237 en condiciones estáticas parece depender del nivel de organización de los biofilms, lo cuál aparte de sugerir rutas de regulación diferentes, podría ser indicativo de su funcionalidad en los mecanismos de defensa u ofensa de la colonia bacteriana.

8. Se han identificado numerosas regiones atípicas poco conservadas dentro del género *Bacillus* a las que no se le ha podido atribuir ninguna funcionalidad potencial. El descubrimiento de estas nuevas regiones y la asignación de una posible función que pueda estar relacionada con la adaptabilidad al entorno y/o contribución a la mejora de la salud de la planta por parte de estas bacterias, supone un trabajo desafiante para abordar en futuros estudios, que además nos ayudarán a comprender mejor el comportamiento de ambas cepas en el entorno de la planta, así como su fisiología.

REFERENCES

REFERENCES

Alkhatib, Z., Abts, A., Mavaro, A., Schmitt, L., and Smits, S.H.J. 2012. Lantibiotics: How do producers become self-protected? Journal of Biotechnology 159:145-154.

Alvarez-Martinez, C.E., and Christie, P.J. 2009. Biological diversity of prokaryotic type IV secretion systems. Microbiology and Molecular Biology Reviews 73:775-808.

Auchtung, J.M., Lee, C.A., Monson, R.E., Lehman, A.P., and Grossman, A.D. 2005. Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. Proceedings of the National Academy of Sciences of the United States of America 102:12554-12559.

Barbosa, J., Caetano, T., and Mendo, S. 2015. Class I and class II lanthipeptides produced by *Bacillus* spp. Journal of Natural Products 78:2850-2866.

Boller, T., and Felix, G. 2009. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Annual Review of Plant Biology 60:379-406.

Borriss, R., Chen, X.-H., Rueckert, C., Blom, J., Becker, A., Baumgarth, B., Fan, B., Pukall, R., Schumann, P., Spröer, C., Junge, H., Vater, J., Pühler, A., and Klenk, H.-P. 2011. Relationship of *Bacillus amyloliquefaciens* clades associated with strains DSM 7T and FZB42T: a proposal for *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* subsp. nov. and *Bacillus amyloliquefaciens* subsp. *plantarum* subsp. nov. based on complete genome sequence comparisons. International Journal of Systematic and Evolutionary Microbiology 61:1786-1801.

Butcher, R.A., Schroeder, F.C., Fischbach, M.A., Straight, P.D., Kolter, R., Walsh, C.T., and Clardy, J. 2007. The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. Proceedings of the National Academy of Sciences of the United States of America 104:1506-1509.

Caetano, T., Krawczyk, J.M., Mösker, E., Süssmuth, R.D., and Mendo, S. 2011. Heterologous expression, biosynthesis, and mutagenesis of type II lantibiotics from *Bacillus licheniformis* in *Escherichia coli*. Chemistry & biology 18:90-100.

Cai, J., Liu, F., Liao, X., and Zhang, R. 2014. Complete genome sequence of *Bacillus amyloliquefaciens* LFB112 isolated from Chinese herbs, a strain of a broad inhibitory spectrum against domestic animal pathogens. Journal of Biotechnology 175:63-64.

Carvalhais, L.C., Dennis, P.G., Fan, B., Fedoseyenko, D., Kierul, K., Becker, A., von Wiren, N., and Borriss, R. 2013. Linking plant nutritional status to plant-microbe interactions. PLoS ONE 8:e68555.

Cawoy, H., Debois, D., Franzil, L., De Pauw, E., Thonart, P., and Ongena, M. 2015. Lipopeptides as main ingredients for inhibition of fungal phytopathogens by *Bacillus subtilis/amyloliquefaciens*. Microbial Biotechnology 8:281-295.

Chen, X.H., Scholz, R., Borriss, M., Junge, H., Mögel, G., Kunz, S., and Borriss, R. 2009. Difficidin and bacilysin produced by plant-associated *Bacillus amyloliquefaciens* are efficient in controlling fire blight disease. Journal of Biotechnology 140:38-44.

Chollet-Imbert, M., Gancel, F., Slomianny, C., and Jacques, P. 2009. Differentiated pellicle organization and lipopeptide production in standing culture of *Bacillus subtilis* strains. Archives of Microbiology 191:63-71.

Dias, L., Caetano, T., Pinheiro, M., and Mendo, S. 2015. The lanthipeptides of *Bacillus methylotrophicus* and their association with genomic islands. Systematic and Applied Microbiology 38:525-533.

Dintner, S., Heermann, R., Fang, C., Jung, K., and Gebhard, S. 2014. A sensory complex consisting of an ATP-binding cassette transporter and a two-component regulatory

system controls bacitracin resistance in *Bacillus subtilis*. Journal of Biological Chemistry 289:27899-27910.

Dischinger, J., Josten, M., Szekat, C., Sahl, H.-G., and Bierbaum, G. 2009. Production of the Novel Two-Peptide Lantibiotic Lichenicidin by *Bacillus licheniformis* DSM 13. PLoS ONE 4:e6788.

Dogsa, I., Choudhary, K.S., Marsetic, Z., Hudaiberdiev, S., Vera, R., Pongor, S., and Mandic-Mulec, I. 2014. ComQXPA quorum sensing systems may not be unique to *Bacillus subtilis*: a census in prokaryotic genomes. PLoS ONE 9:e96122.

Dunlap, C.A., Kim, S.-J., Kwon, S.-W., and Rooney, A.P. 2015. Phylogenomic analysis shows that *Bacillus amyloliquefaciens* subsp. *plantarum* is a later heterotypic synonym of *Bacillus methylotrophicus*. International Journal of Systematic and Evolutionary Microbiology 65:2104-2109.

Dunlap, C.A., Kim, S.-J., Kwon, S.-W., and Rooney, A.P. 2016. *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyloliquefaciens*; *Bacillus methylotrophicus*, *Bacillus amyloliquefaciens* subsp. *plantarum* and '*Bacillus oryzicola*' are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics. International Journal of Systematic and Evolutionary Microbiology 66:1212-1217.

Farag, M., Zhang, H., and Ryu, C.-M. 2013. Dynamic chemical communication between plants and bacteria through airborne signals: induced resistance by bacterial volatiles. J Chem Ecol 39:1007-1018.

Fasimoye, F.O., Olajuyigbe, F.M., and Sanni, M.D. 2013. Purification and characterization of a thermostable extracellular phytase from *Bacillus licheniformis* PFBL-03. Preparative Biochemistry and Biotechnology 44:193-205.

Flemming, H.-C., and Wingender, J. 2010. The biofilm matrix. Nat Rev Micro 8:623-633.

Gallego del Sol, F., and Marina, A. 2013. Structural basis of Rap phosphatase inhibition by Phr peptides. PLoS Biol 11:e1001511.

García-Gutiérrez, L., Zeriouh, H., Romero, D., Cubero, J., de Vicente, A., and Pérez-García, A. 2013. The antagonistic strain *Bacillus subtilis* UMAF6639 also confers protection to melon plants against cucurbit powdery mildew by activation of jasmonate- and salicylic acid-dependent defence responses. Microbial Biotechnology 6:264-274.

García-Gutiérrez, L., Romero, D., Zeriouh, H., Cazorla, F., Torés, J., de Vicente, A., and Pérez-García, A. 2012. Isolation and selection of plant growth-promoting rhizobacteria as inducers of systemic resistance in melon. Plant and Soil.

Guder, A., Schmitter, T., Wiedemann, I., Sahl, H.-G., and Bierbaum, G. 2002. Role of the single regulator MrsR1 and the two-component system MrsR2/K2 in the regulation of mersacidin production and immunity. Applied and Environmental Microbiology 68:106-113.

He, P., Hao, K., Blom, J., Rückert, C., Vater, J., Mao, Z., Wu, Y., Hou, M., He, P., He, Y., and Borriss, R. 2012. Genome sequence of the plant growth promoting strain *Bacillus amyloliquefaciens* subsp. *plantarum* B9601-Y2 and expression of mersacidin and other secondary metabolites. Journal of Biotechnology 164:281-291.

Hobley, L., Ostrowski, A., Rao, F.V., Bromley, K.M., Porter, M., Prescott, A.R., MacPhee, C.E., van Aalten, D.M.F., and Stanley-Wall, N.R. 2013. BslA is a self-assembling bacterial hydrophobin that coats the *Bacillus subtilis* biofilm. Proceedings of the National Academy of Sciences 110:13600-13605.

Hyungjae, L., and Hae-Yeong, K. 2011. Lantibiotics, class I bacteriocins from the genus *Bacillus*. Journal of Microbiology and Biotechnology 21:229-235.

Kearns, D.B., Chu, F., Rudner, R., and Losick, R. 2004. Genes governing swarming in *Bacillus subtilis* and evidence for a phase variation mechanism controlling surface motility. Molecular Microbiology 52:357-369.

Kearns, D.B., Chu, F., Branda, S.S., Kolter, R., and Losick, R. 2005. A master regulator for biofilm formation by *Bacillus subtilis*. Molecular Microbiology 55:739-749.

Khosa, S., Lagedroste, M., and Smits, S.H. 2016. Protein defence systems against the lantibiotic nisin: Function of the immunity protein NisI and the resistance protein NSR. Frontiers in Microbiology 7.

Kierul, K., Voigt, B., Albrecht, D., Chen, X.-H., Carvalhais, L.C., and Borriss, R. 2015. Influence of root exudates on the extracellular proteome of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. Microbiology 161:131-147.

Kies, S., Vuong, C., Hille, M., Peschel, A., Meyer, C., Götz, F., and Otto, M. 2003. Control of antimicrobial peptide synthesis by the agr quorum sensing system in *Staphylococcus epidermidis*: activity of the lantibiotic epidermin is regulated at the level of precursor peptide processing. Peptides 24:329-338.

Lawton, E.M., Cotter, P.D., Hill, C., and Ross, R.P. 2007. Identification of a novel two-peptide lantibiotic, haloduracin, produced by the alkaliphile *Bacillus halodurans* C-125. FEMS Microbiology Letters 267:64-71.

Liu, Y., Lai, Q., Dong, C., Sun, F., Wang, L., Li, G., and Shao, Z. 2013. Phylogenetic diversity of the *Bacillus pumilus* group and the marine ecotype revealed by multilocus sequence analysis. PLoS ONE 8:e80097.

Lubelski, J., Rink, R., Khusainov, R., Moll, G.N., and Kuipers, O.P. 2008. Biosynthesis, immunity, regulation, mode of action and engineering of the model lantibiotic nisin. Cellular and Molecular Life Sciences 65:455-476.

Magno-Pérez-Bryan, M.C., Martínez-García, P.M., Hierrezuelo, J., Rodríguez-Palenzuela, P., Arrebola, E., Ramos, C., de Vicente, A., Pérez-García, A., and Romero, D. 2015. Comparative genomics within the *Bacillus* genus reveal the singularities of two robust *Bacillus amyloliquefaciens* biocontrol strains. Molecular Plant-Microbe Interactions 28:1102-1116.

Marahiel, M.A. 2009. Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. Journal of Peptide Science 15:799-807.

McClerren, A.L., Cooper, L.E., Quan, C., Thomas, P.M., Kelleher, N.L., and van der Donk, W.A. 2006. Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. Proceedings of the National Academy of Sciences 103:17243-17248.

Miethke, M., Klotz, O., Linne, U., May, J.J., Beckering, C.L., and Marahiel, M.A. 2006. Ferri-bacillibactin uptake and hydrolysis in *Bacillus subtilis*. Molecular Microbiology 61:1413-1427.

Nicholson, W.L. 2008. The *Bacillus subtilis ydjL* (*bdhA*) gene encodes acetoin reductase/2,3-butanediol dehydrogenase. Applied and Environmental Microbiology 74:6832-6838.

Ongena, M., Jourdan, E., Adam, A., Paquot, M., Brans, A., Joris, B., Arpigny, J.-L., and Thonart, P. 2007. Surfactin and fengycin lipopeptides of *Bacillus subtilis* as elicitors of induced systemic resistance in plants. Environmental Microbiology 9:1084-1090.

Oslizlo, A., Stefanic, P., Vatovec, S., Beigot Glaser, S., Rupnik, M., and Mandic-Mulec, I. 2015. Exploring ComQXPA quorum-sensing diversity and biocontrol potential of *Bacillus* spp. isolates from tomato rhizoplane. Microbial Biotechnology 8:527-540.

Pérez-García, A., Romero, D., and de Vicente, A. 2011. Plant protection and growth stimulation by microorganisms: biotechnological applications of Bacilli in agriculture. Current Opinion in Biotechnology 22:187-193.

Romero, D., Aguilar, C., Losick, R., and Kolter, R. 2010. Amyloid fibers provide structural integrity to *Bacillus subtilis* biofilms. Proceedings of the National Academy of Sciences.

Romero, D., Pérez-García, A., Rivera, M.E., Cazorla, F.M., and Vicente, A. 2004. Isolation and evaluation of antagonistic bacteria towards the cucurbit powdery mildew fungus *Podosphaera fusca*. Applied Microbiology and Biotechnology 64:263-269.

Romero, D., de Vicente, A., Olmos, J.L., Dávila, J.C., and Pérez-García, A. 2007a. Effect of lipopeptides of antagonistic strains of *Bacillus subtilis* on the morphology and ultrastructure of the cucurbit fungal pathogen *Podosphaera fusca*. Journal of Applied Microbiology 103:969-976.

Romero, D., de Vicente, A., Rakotoaly, R.H., Dufour, S.E., Veening, J.-W., Arrebola, E., Cazorla, F.M., Kuipers, O.P., Paquot, M., and Pérez-García, A. 2007b. The iturin and fengycin families of lipopeptides are key factors in antagonism of *Bacillus subtilis* toward *Podosphaera fusca*. Molecular Plant-Microbe Interactions 20:430-440.

Schmidt, T., Scott, E., and Dyer, D. 2011. Whole-genome phylogenies of the family Bacillaceae and expansion of the sigma factor gene family in the *Bacillus cereus* species-group. BMC Genomics 12:430.

Spieß, T., Korn, S.M., Kötter, P., and Entian, K.-D. 2015. Autoinduction specificities of the lantibiotics subtilin and nisin. Applied and Environmental Microbiology 81:7914-7923.

Stefanic, P., Decorosi, F., Viti, C., Petito, J., Cohan, F.M., and Mandic-Mulec, I. 2012. The quorum sensing diversity within and between ecotypes of *Bacillus subtilis*. Environmental Microbiology 14:1378-1389.

Steinborn, G., Hajirezaei, M.-R., and Hofemeister, J. 2005. *Bac* genes for recombinant bacilysin and anticapsin production in *Bacillus* host strains. Archives of Microbiology 183:71-79.

Tortosa, P., Logsdon, L., Kraigher, B., Itoh, Y., Mandic-Mulec, I., and Dubnau, D. 2001. Specificity and genetic polymorphism of the *Bacillus* competence *quorum-sensing* system. Journal of Bacteriology 183:451-460.

Um, S., Fraimout, A., Sapountzis, P., Oh, D.-C., and Poulsen, M. 2013. The fungus-growing termite *Macrotermes natalensis* harbors bacillaene-producing *Bacillus* sp. that inhibit potentially antagonistic fungi. Scientific Reports 3:3250.

Vlamakis, H., Chai, Y., Beauregard, P., Losick, R., and Kolter, R. 2013. Sticking together: building a biofilm the *Bacillus subtilis* way. Nature reviews. Microbiology 11:157-168.

Wang, X., Koehler, S.A., Wilking, J.N., Sinha, N.N., Cabeen, M.T., Srinivasan, S., Seminara, A., Rubinstein, S., Sun, Q., Brenner, M.P., and Weitz, D.A. 2016. Probing phenotypic growth in expanding *Bacillus subtilis* biofilms. Applied Microbiology and Biotechnology 100:4607-4615.

Willey, J.M., and Donk, W.A.v.d. 2007. Lantibiotics: peptides of diverse structure and function. Annual Review of Microbiology 61:477-501.

Wu, L., Wu, H., Chen, L., Yu, X., Borriss, R., and Gao, X. 2015. Difficidin and bacilysin from *Bacillus amyloliquefaciens* FZB42 have antibacterial activity against *Xanthomonas oryzae* rice pathogens. Scientific Reports 5:12975.

Yang, Y., Wu, H.-J., Lin, L., Zhu, Q.-q., Borriss, R., and Gao, X.-W. 2015. A plasmid-born Rap-Phr system regulates surfactin production, sporulation and genetic competence in the heterologous host, *Bacillus subtilis* OKB105. Applied Microbiology and Biotechnology:1-12.

Yuan, J., Li, B., Zhang, N., Waseem, R., Shen, Q., and Huang, Q. 2012. Production of bacillomycin- and macrolactin-type antibiotics by *Bacillus amyloliquefaciens* NJN-6 for suppressing soilborne plant pathogens. Journal of Agricultural and Food Chemistry 60:2976-2981.

Zeriouh, H., de Vicente, A., Pérez-García, A., and Romero, D. 2014. Surfactin triggers biofilm formation of *Bacillus subtilis* in melon phylloplane and contributes to the biocontrol activity. Environmental Microbiology 16:2196-2211.

Zeriouh, H., Romero, D., García-Gutiérrez, L., Cazorla, F.M., de Vicente, A., and Pérez-García, A. 2011. The iturin-like lipopeptides are essential components in the biological control arsenal of *Bacillus subtilis* against bacterial diseases of cucurbits. Molecular Plant-Microbe Interactions 24:1540-1552.

SUPPLEMENTARY MATERIALS

ANNEX I

Figures

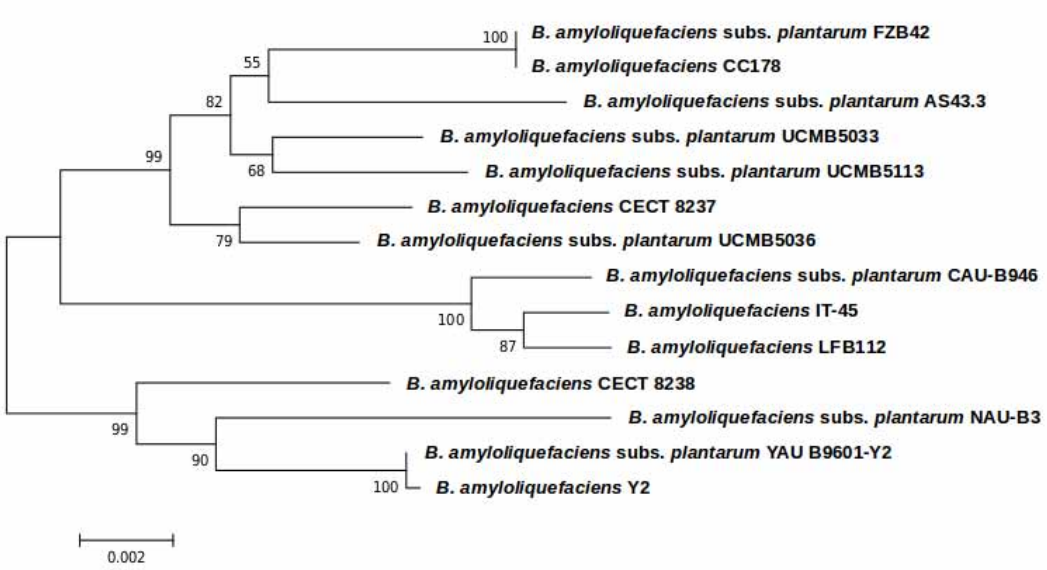


Figure 1. Phylogenetic analysis based on the gene sequences distinctive of *Bacillus amyloliquefaciens* strains associated with plants. A total of sixteen genes (genes highlighted in blue in Annex Table 1) were used to build a Neighbour-joining tree using MEGA 5 bootstrap values (10,000 repetitions), which are shown on the branches. The sequences were extracted from published genome sequences.

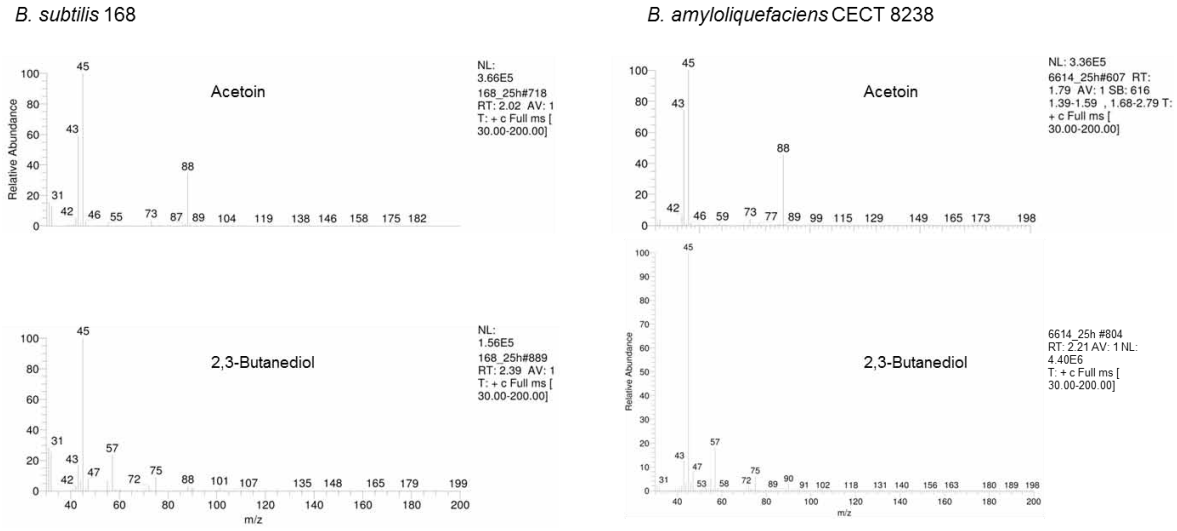


Figure 2. Detection and quantification of acetoin and 2,3-butanediol in bacterial supernatants using GC-MS analysis. A representative spectrum of each molecule detected in cultures of 24 h of *Bacillus subtilis* 168 (a known producer) and *Bacillus amyloliquefaciens* CECT 8238. The same profiles were observed in the rest of the strains included in this analysis.

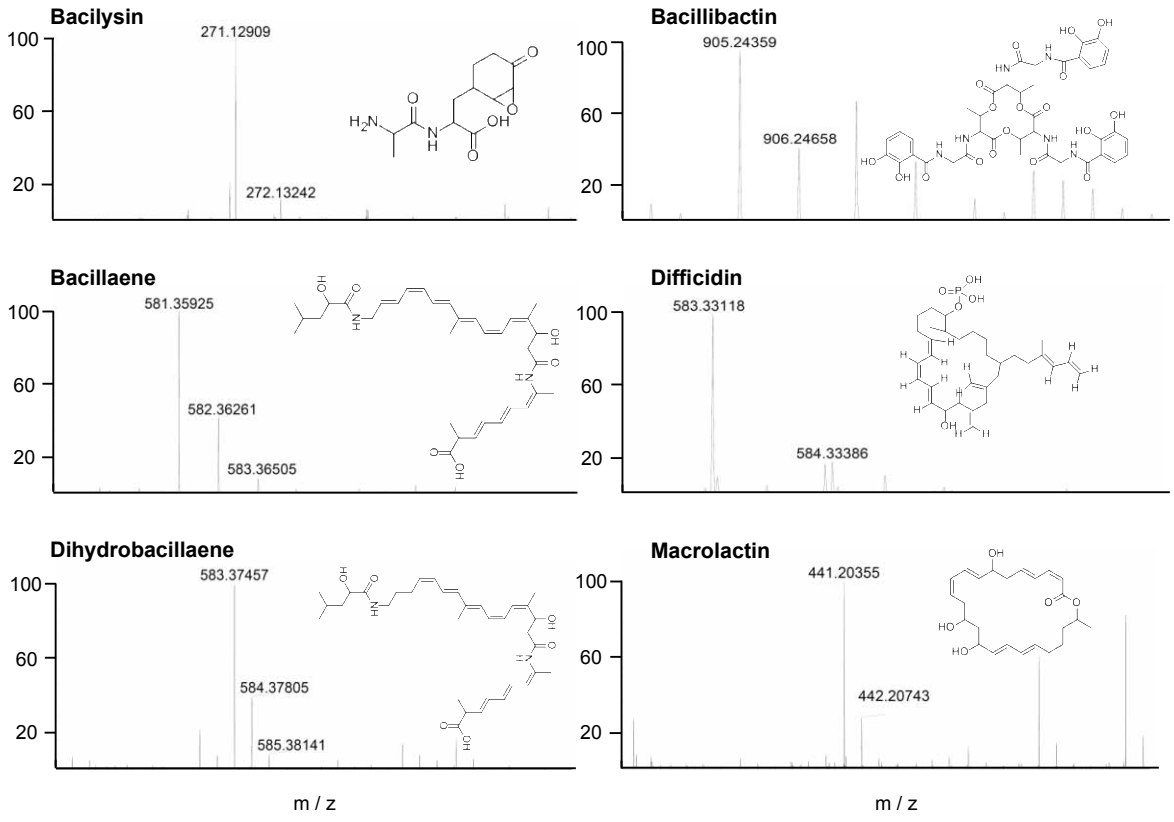
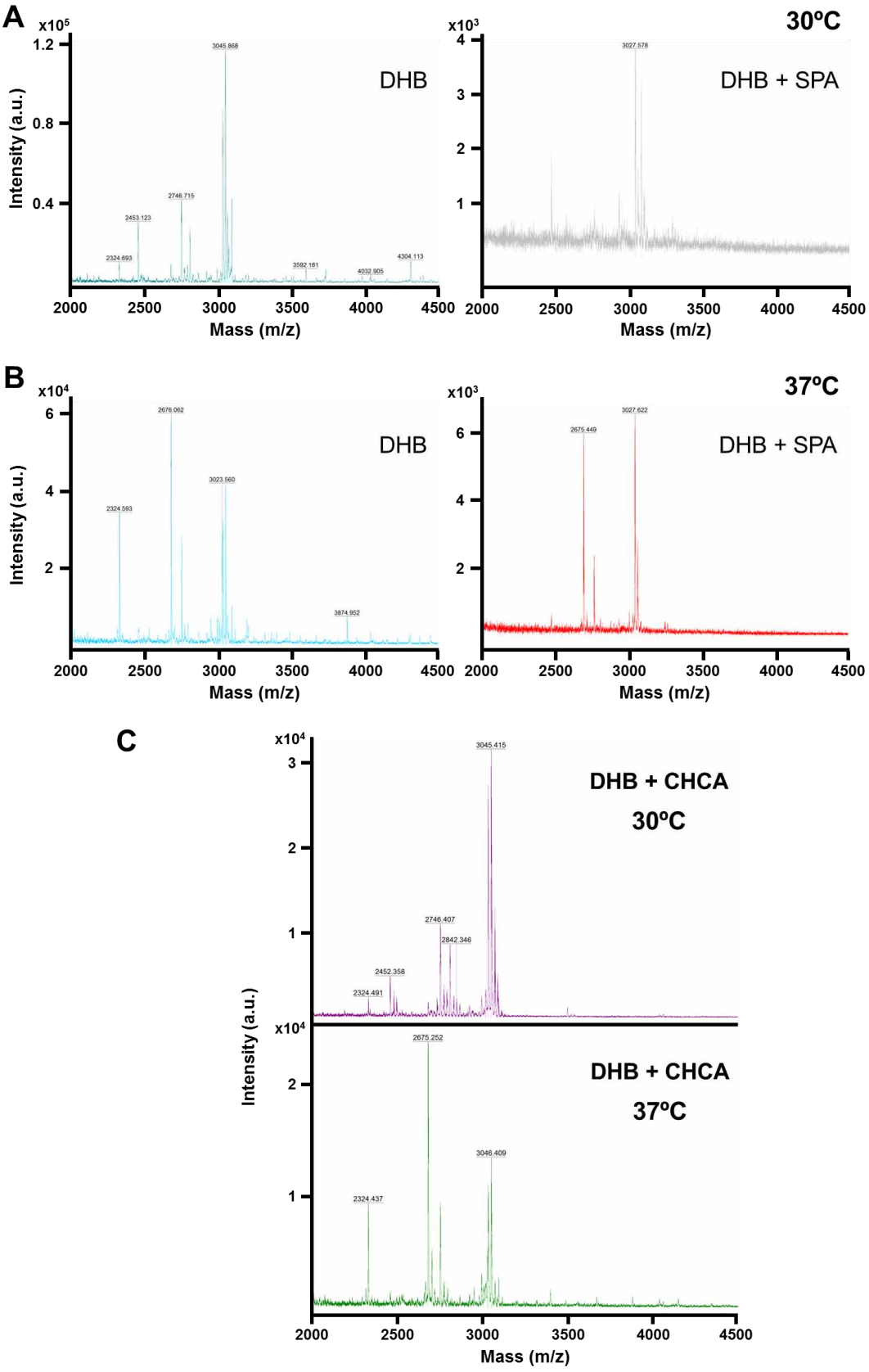


Figure 3. *Bacillus amyloliquefaciens* CECT 8237 and CECT 8238 produce a variety of known secondary metabolites. The ESI-MS spectrum in positive mode of the bacterial secondary metabolites eluted at the retention times indicated in Figure 19. The chemical structures of the molecules are included (insets). Similar profiles were observed in supernatants of *B. amyloliquefaciens* CECT 8238.

Figure 4. Optimization of the matrix used in MALDI-TOF experiments for the in situ detection of lichenicidin-like lantibiotic in *B. amyloliquefaciens* colonies. A. MALDI-TOF mass spectra at 30ºC showing values of high molecular weight in 2,5-dihydroxy benzoic acid (DHB) and DHB supplemented with sinapinic acid (SPA). B. MALDI-TOF mass spectra at 37ºC showing values of high molecular weight in DHB and DHB supplemented with SPA. C. MALDI-TOF mass spectra showing the high molecular weight values of the selected matrix, for the specific detection of the lantibiotic, composed by DHB and α-cyanohydroxycinnamic acid (CHCA) at 30ºC (up) and 37ºC (down).



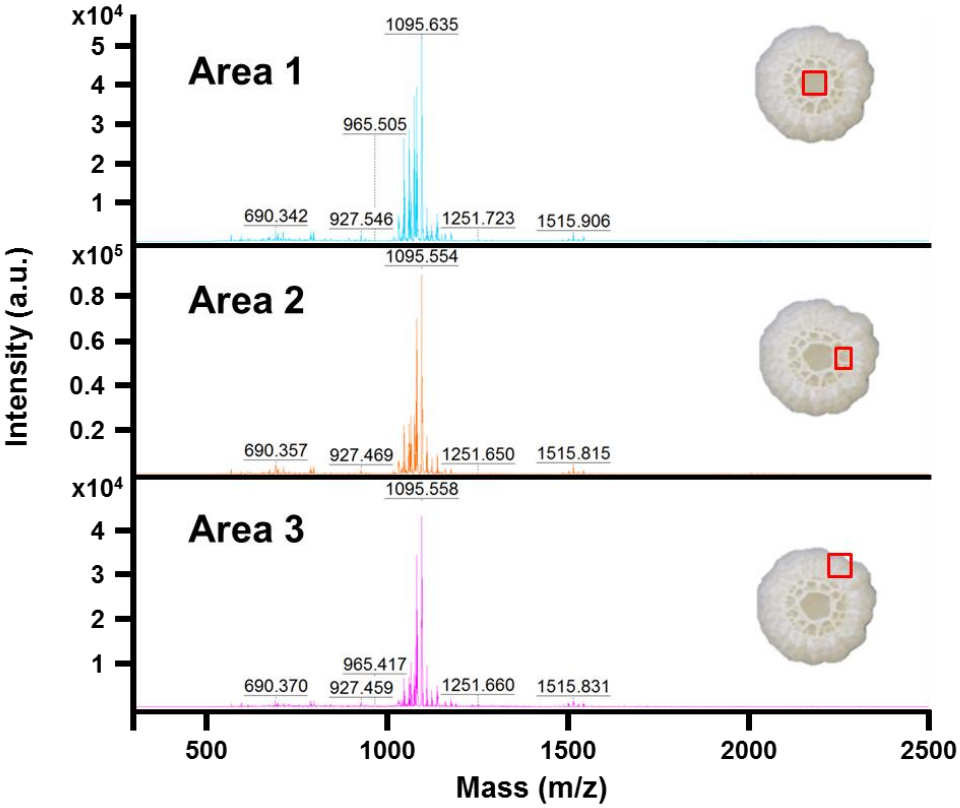


Figure 5. Spatial distribution of secondary metabolites in colonies of *B. amyloliquefaciens* CECT 8237 grown in medium M at 30ºC. MALDI-TOF mass spectra ranging from 250 to 2500 Da detected in the centre of the colony formation, area 1 (top), in the middle of the colony, area 2 (middle) and in the most external growth line, area 3 (bottom). The location in the colony is included (insets).

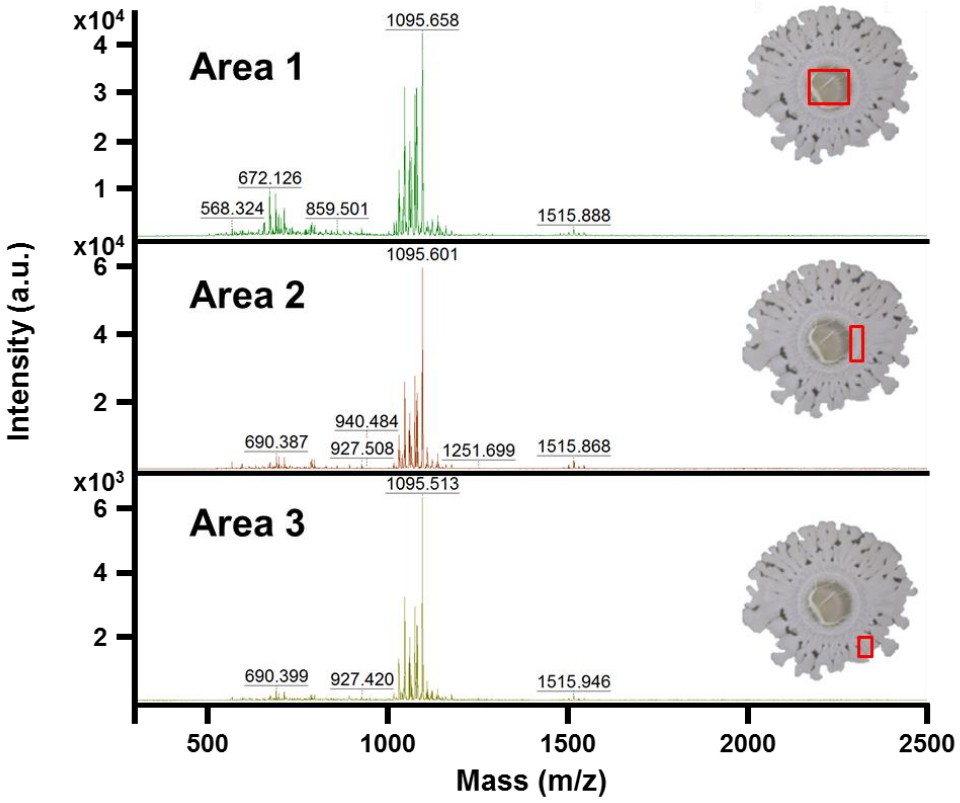


Figure 6. Spatial distribution of secondary metabolites in colonies of *B. amyloliquefaciens* CECT 8237 grown in medium M at 37ºC. MALDI-TOF mass spectra ranging from 250 to 2500 Da detected in the centre of the colony formation, area 1 (top), in the middle of the colony, area 2 (middle) and in the most external growth line, area 3 (bottom). The location in the colony is included (insets).

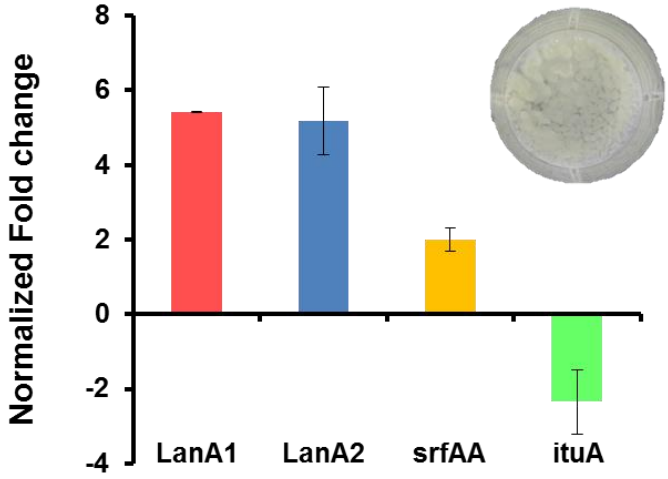


Figure 7. The synthetic genes of the lichenicidin-like peptides are highly expressed in biofilms encased cells. The relative expression levels in planktonic and pellicle associated cells of biofilm grown in medium M at 37ºC for 24 h was studied in qRT-PCR analysis. The genes tested were: *lanA1* (red) and *lanA2* (blue), the two peptides of the lantibiotic; *srfAA*, surfactin (yellow); *ituA*, iturinA (green). For comparison, every value of expression obtained in RNA isolated from pellicle associated cells was divided into the corresponding data obtained from planktonic phase in each gene tested.

ANNEX II

Tables

Table 1. Coding sequences ubiquitously present in plant-associated *Bacillus amyloliquefaciens* strains. Each row contains data relative to a given CDS in a given strain. Columns provide information on strain's GenBank code, gene identifier and protein length. CDSs included and excluded from the phylogenetic analysis of Figure S1 are highlighted in blue and red, respectively.

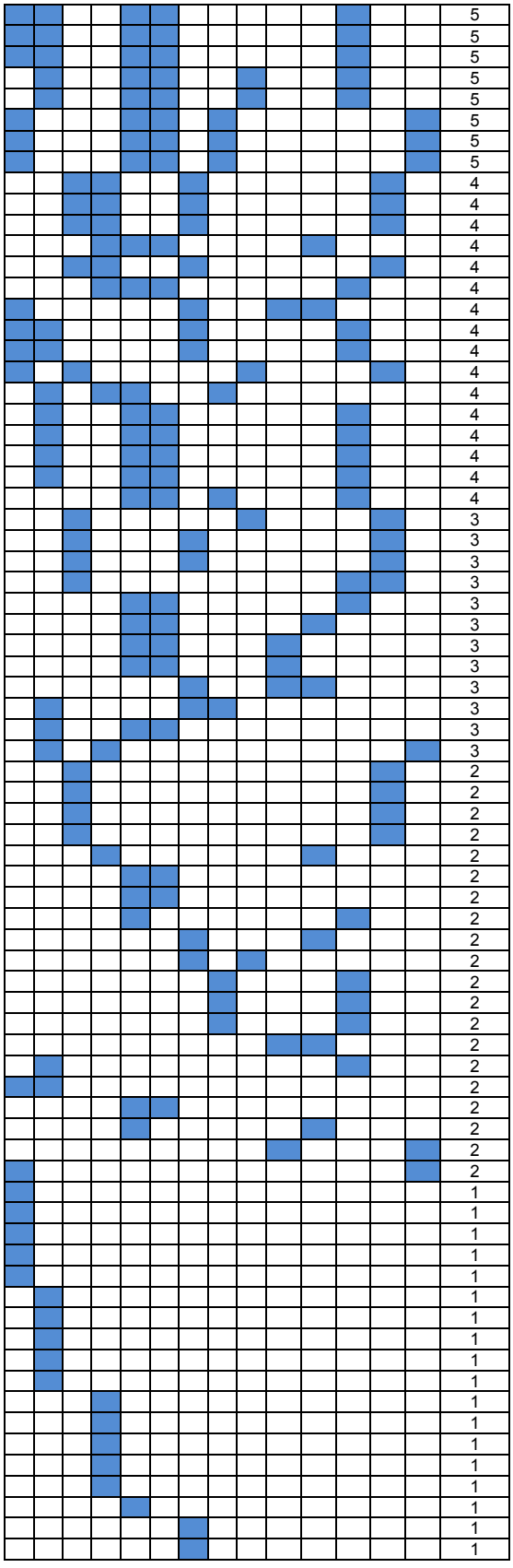
CDS	Genome	Gene_ID	Protein length	CDS	Genome	Gene_ID	Protein length
1	CP006058	BAMY6639_00510	211	4	CP006952	YP_008950396.1	46
1	CP006960	BAMY6614_13305	210	5	CP006058	BAMY6639_02585	130
1	CP006845	YP_001421391.1	211	5	CP006960	BAMY6614_15220	130
1	HE617159	YP_005130485.1	211	5	CP006845	YP_001421780.1	130
1	HE774679	YP_005421260.1	211	5	HE617159	YP_005130928.1	124
1	CP003332	YP_006328832.1	196	5	HE774679	YP_005421677.1	130
1	CP003838	YP_007186550.1	211	5	CP003332	YP_006329280.1	130
1	NC_020272	YP_007445446.1	211	5	CP003838	YP_007186946.1	130
1	HF563562	YP_007497438.1	211	5	NC_020272	YP_007445069.1	130
1	HG328253	YP_008412768.1	211	5	HF563562	YP_007497812.1	130
1	HG328254	YP_008421178.1	211	5	HG328253	YP_008413305.1	130
1	HG514499	YP_008626587.1	211	5	HG328254	YP_008421537.1	130
1	CP006845	YP_008727130.1	211	5	HG514499	YP_008626040.1	130
1	CP006952	YP_008950229.1	210	5	CP006845	YP_008727537.1	130
2	CP006058	BAMY6639_00620	499	5	CP006952	YP_008950691.1	130
2	CP006960	BAMY6614_13370	499	6	CP006058	BAMY6639_02665	90
2	CP006845	YP_001421404.1	499	6	CP006960	BAMY6614_15315	76
2	HE617159	YP_005130495.1	499	6	CP006845	YP_001421798.1	90
2	HE774679	YP_005421272.1	501	6	HE617159	YP_005130948.1	90
2	CP003332	YP_006328845.1	501	6	HE774679	YP_005421695.1	90
2	CP003838	YP_007186563.1	499	6	CP003332	YP_006329301.1	90
2	NC_020272	YP_007445439.1	499	6	CP003838	YP_007186964.1	90
2	HF563562	YP_007497448.1	499	6	NC_020272	YP_007445051.1	90
2	HG328253	YP_008412778.1	499	6	HF563562	YP_007497830.1	90
2	HG328254	YP_008421189.1	499	6	HG328253	YP_008413323.1	90
2	HG514499	YP_008626576.1	499	6	HG328254	YP_008421555.1	90
2	CP006845	YP_008727143.1	499	6	HG514499	YP_008626022.1	90
2	CP006952	YP_008950241.1	499	6	CP006845	YP_008727555.1	90
3	CP006058	BAMY6639_00685	478	6	CP006952	YP_008950711.1	76
3	CP006960	BAMY6614_13435	481	7	CP006058	BAMY6639_02670	326
3	CP006845	YP_001421417.1	478	7	CP006960	BAMY6614_15320	326
3	HE617159	YP_005130510.1	358	7	CP006845	YP_001421799.1	326
3	HE774679	YP_005421285.1	481	7	HE617159	YP_005130949.1	326
3	CP003332	YP_006328859.1	481	7	HE774679	YP_005421696.1	326
3	CP003838	YP_007186576.1	478	7	CP003332	YP_006329302.1	326
3	CP004065	YP_007445426.1	481	7	CP003838	YP_007186965.1	326
3	HF563562	YP_007497461.1	478	7	NC_020272	YP_007445050.1	326
3	HG328253	YP_008412792.1	478	7	HF563562	YP_007497831.1	326
3	HG328254	YP_008421203.1	478	7	HG328253	YP_008413324.1	326
3	HG514499	YP_008626563.1	481	7	HG328254	YP_008421556.1	326
3	CP006845	YP_008727158.1	478	7	HG514499	YP_008626021.1	326
3	CP006952	YP_008950254.1	481	7	CP006845	YP_008727556.1	326
4	CP006058	BAMY6639_01095	46	7	CP006952	YP_008950712.1	326
4	CP006960	BAMY6614_13845	46	8	CP006058	BAMY6639_02680	176
4	CP006845	YP_001421497.1	46	8	CP006960	BAMY6614_15330	176
4	HE617159	YP_005130633.1	46	8	CP006845	YP_001421801.1	176
4	HE774679	YP_005421364.1	46	8	HE617159	YP_005130951.1	176
4	CP003332	YP_006328949.1	46	8	HE774679	YP_005421698.1	176
4	CP003838	YP_007186657.1	46	8	CP003332	YP_006329305.1	176
4	NC_020272	YP_007445345.1	46	8	CP003838	YP_007186967.1	176
4	HF563562	YP_007497536.1	46	8	NC_020272	YP_007445048.1	176
4	HG328253	YP_008412867.1	46	8	HF563562	YP_007497833.1	176
4	HG328254	YP_008421277.1	46	8	HG328253	YP_008413326.1	176
4	HG514499	YP_008626484.1	46	8	HG328254	YP_008421558.1	176
4	CP006845	YP_008727243.1	46	8	HG514499	YP_008626019.1	176

8	CP006845	YP_008727558.1	176
8	CP006952	YP_008950714.1	176
9	CP006058	BAMY6639_04925	136
9	CP006960	BAMY6614_17550	136
9	CP006845	YP_001422208.1	136
9	HE617159	YP_005131365.1	136
9	HE774679	YP_005422173.1	136
9	CP003332	YP_006329803.1	55
9	CP003838	YP_007187391.1	136
9	NC_020272	YP_007444565.1	136
9	HF563562	YP_007498247.1	136
9	HG328253	YP_008413737.1	136
9	HG328254	YP_008421983.1	136
9	HG514499	YP_008625574.1	136
9	CP006845	YP_008727999.1	136
9	CP006952	YP_008951142.1	136
10	CP006058	BAMY6639_06150	214
10	CP006960	BAMY6614_18725	214
10	CP006845	YP_001422407.1	214
10	HE617159	YP_005131601.1	214
10	HE774679	YP_005422371.1	214
10	CP003332	YP_006330022.1	214
10	CP003838	YP_007187589.1	214
10	NC_020272	YP_007444366.1	214
10	HF563562	YP_007498433.1	214
10	HG328253	YP_008413949.1	214
10	HG328254	YP_008422176.1	214
10	HG514499	YP_008627663.1	214
10	CP006845	YP_008728220.1	214
10	CP006952	YP_008951341.1	214
11	CP006058	BAMY6639_06985	248
11	CP006960	BAMY6614_19540	248
11	CP006845	YP_001422565.1	248
11	HE617159	YP_005131755.1	248
11	HE774679	YP_005422528.1	248
11	CP003332	YP_006330190.1	248
11	CP003838	YP_007187752.1	248
11	NC_020272	YP_007444207.1	248
11	HF563562	YP_007498587.1	248
11	HG328253	YP_008414110.1	248
11	HG328254	YP_008422330.1	248
11	HG514499	YP_008627823.1	248
11	CP006845	YP_008728390.1	248
11	CP006952	YP_008951500.1	248
12	CP006058	BAMY6639_08725	238
12	CP006960	BAMY6614_01180	238
12	CP006845	YP_001422892.1	238
12	HE617159	YP_005132092.1	238
12	HE774679	YP_005422858.1	238
12	CP003332	YP_006330551.1	246
12	CP003838	YP_007188095.1	238
12	NC_020272	YP_007443860.1	238
12	HF563562	YP_007498919.1	238
12	HG328253	YP_008414445.1	175
12	HG328254	YP_008422654.1	238
12	HG514499	YP_008628171.1	238
12	CP006845	YP_008728741.1	246
12	CP006952	YP_008951841.1	238
13	CP006058	BAMY6639_09275	209
13	CP006960	BAMY6614_01750	209
13	CP006845	YP_001422999.1	209
13	HE617159	YP_005132202.1	209
13	HE774679	YP_005422965.1	209
13	CP003332	YP_006330668.1	209
13	CP003838	YP_007188201.1	209
13	NC_020272	YP_007443750.1	209
13	HF563562	YP_007499030.1	209
13	HG328253	YP_008414549.1	209
13	HG328254	YP_008422756.1	209
13	HG514499	YP_008628281.1	209
13	CP006845	YP_008728860.1	209
13	CP006952	YP_008951953.1	209
14	CP006058	BAMY6639_10810	239
14	CP006960	BAMY6614_03180	239
14	CP006845	YP_001423287.1	239
14	HE617159	YP_005132465.1	239
14	HE774679	YP_005423242.1	239
14	CP003332	YP_006330974.1	241
14	CP003838	YP_007188492.1	239
14	NC_020272	YP_007443476.1	239
14	HF563562	YP_007499301.1	241
14	HG328253	YP_008414822.1	241
14	HG328254	YP_008423038.1	241
14	HG514499	YP_008628557.1	239
14	CP006845	YP_008729160.1	239
14	CP006952	YP_008952222.1	239
15	CP006058	BAMY6639_11100	288
15	CP006960	BAMY6614_03480	288
15	CP006845	YP_001423337.1	288
15	HE617159	YP_005132513.1	288
15	HE774679	YP_005423290.1	288
15	CP003332	YP_006331023.1	247
15	CP003838	YP_007188542.1	288
15	NC_020272	YP_007443428.1	288
15	HF563562	YP_007499347.1	288
15	HG328253	YP_008414873.1	288
15	HG328254	YP_008423083.1	288
15	HG514499	YP_008628604.1	288
15	CP006845	YP_008729211.1	288
15	CP006952	YP_008952268.1	288
16	CP006058	BAMY6639_12160	157
16	CP006960	BAMY6614_04730	157
16	CP006845	YP_001419831.1	157
16	HE617159	YP_005128881.1	157
16	HE774679	YP_005419489.1	157
16	CP003332	YP_006326966.1	158
16	CP003838	YP_007184870.1	157
16	NC_020272	YP_007447089.1	157
16	HF563562	YP_007495895.1	157
16	HG328253	YP_008411180.1	157
16	HG328254	YP_008419608.1	157
16	HG514499	YP_008624793.1	157
16	CP006845	YP_008725445.1	157
16	CP006952	YP_008948599.1	157
17	CP006058	BAMY6639_12575	147
17	CP006960	BAMY6614_05135	147
17	CP006845	YP_001419896.1	147
17	HE617159	YP_005128937.1	147
17	HE774679	YP_005419550.1	147
17	CP003332	YP_006327031.1	147
17	CP003838	YP_007184928.1	147
17	NC_020272	YP_007447028.1	147
17	HF563562	YP_007495957.1	147
17	HG328253	YP_008411249.1	147
17	HG328254	YP_008419671.1	147
17	HG514499	YP_008624854.1	147
17	CP006845	YP_008725513.1	147
17	CP006952	YP_008948666.1	147
18	CP006058	BAMY6639_12900	659
18	CP006960	BAMY6614_05420	659
18	CP006845	YP_001419958.1	659
18	HE617159	YP_005129000.1	659
18	HE774679	YP_005419608.1	659
18	CP003332	YP_006327093.1	662
18	CP003838	YP_007184988.1	659
18	NC_020272	YP_007446968.1	659
18	HF563562	YP_007496022.1	659
18	HG328253	YP_008411312.1	659

18	HG328254	YP_008419732.1	659	21	CP006845	YP_008725784.1	141
18	HG514499	YP_008624911.1	659	21	CP006952	YP_008948628.1	448
18	CP006845	YP_008725581.1	659	22	CP006058	BAMY6639_14025	345
18	CP006952	YP_008948729.1	659	22	CP006960	BAMY6614_06575	345
19	CP006058	BAMY6639_12955	94	22	CP006845	YP_001420184.1	345
19	CP006960	BAMY6614_05465	92	22	HE617159	YP_005129217.1	345
19	CP006845	YP_001419967.1	94	22	HE774679	YP_005419819.1	221
19	HE617159	YP_005129009.1	94	22	CP003332	YP_006327321.1	221
19	HE774679	YP_005419617.1	92	22	CP003838	YP_007185220.1	345
19	CP003332	YP_006327102.1	92	22	NC_020272	YP_007446746.1	345
19	CP003838	YP_007184997.1	94	22	HF563562	YP_007496222.1	345
19	NC_020272	YP_007446959.1	94	22	HG328253	YP_008411531.1	345
19	HF563562	YP_007496031.1	94	22	HG328254	YP_008419955.1	345
19	HG328253	YP_008411321.1	94	22	HG514499	YP_008625134.1	345
19	HG328254	YP_008419741.1	94	22	CP006845	YP_008725818.1	345
19	HG514499	YP_008624920.1	92	22	CP006952	YP_008948941.1	345
19	CP006845	YP_008725591.1	94	23	CP006058	BAMY6639_14270	84
19	CP006952	YP_008948738.1	94	23	CP006960	BAMY6614_06805	84
20	CP006058	BAMY6639_13610	205	23	CP006845	YP_001420235.1	84
20	CP006960	BAMY6614_06120	205	23	HE617159	YP_005129259.1	84
20	CP006845	YP_001420094.1	192	23	HE774679	YP_005419870.1	84
20	HE617159	YP_005129133.1	192	23	CP003332	YP_006327379.1	84
20	HE774679	YP_005419733.1	205	23	CP003838	YP_007185267.1	84
20	CP003332	YP_006327233.1	94	23	NC_020272	YP_007446703.1	84
20	CP003838	YP_007185122.1	94	23	HF563562	YP_007496265.1	84
20	NC_020272	YP_007446840.1	263	23	HG328253	YP_008411577.1	84
20	HF563562	YP_007496152.1	192	23	HG328254	YP_008420000.1	84
20	HG328253	YP_008411448.1	192	23	HG514499	YP_008625180.1	84
20	HG328254	YP_008419868.1	192	23	CP006845	YP_008725876.1	84
20	HG514499	YP_008625049.1	205	23	CP006952	YP_008948982.1	84
20	CP006845	YP_008725727.1	94	24	CP006058	BAMY6639_16365	379
20	CP006952	YP_008948856.1	94	24	CP006960	BAMY6614_08900	357
21	CP006058	BAMY6639_13935	141	24	CP006845	YP_001420599.1	357
21	CP006960	BAMY6614_06425	141	24	HE617159	YP_005129692.1	357
21	CP006845	YP_001420152.1	141	24	HE774679	YP_005420246.1	357
21	HE617159	YP_005129187.1	141	24	CP003332	YP_006327783.1	379
21	HE774679	YP_005419784.1	141	24	CP003838	YP_007185660.1	379
21	CP003332	YP_006327284.1	141	24	NC_020272	YP_007446339.1	379
21	CP003838	YP_007185192.1	141	24	HF563562	YP_007496628.1	379
21	NC_020272	YP_007446779.1	141	24	HG328253	YP_008411961.1	379
21	HF563562	YP_007496207.1	141	24	HG328254	YP_008420373.1	379
21	HG328253	YP_008411503.1	141	24	HG514499	YP_008627439.1	357
21	HG328254	YP_008419928.1	141	24	CP006845	YP_008726275.1	379
21	HG514499	YP_008625100.1	141	24	CP006952	YP_008949352.1	357

Table 2. Genes exclusively detected in at least one plant-associated *Bacillus amyloliquefaciens* strain and absent in all industrial strains that were classifiable according to the annotation of COG categories.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Count	Product	COG letter
BAMY6639_00510	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	Ketohydroxyglutarate aldolase	**G**
BAMY6639_00620	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	Endoglucanase	**G**
BAMY6639_00685	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	Transporter	**G/ E**
BAMY6639_02585	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	Hypothetical protein	**E/ R**
BAMY6639_02680	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	Transcription antiterminator	**K**
BAMY6639_06150	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	Hypothetical protein	**Q**
BAMY6639_06985	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	Acetoacetate decarboxylase	**Q**
BAMY6639_08725	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	Endonuclease V	**L**
BAMY6639_09275	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	Membrane protein	**S**
BAMY6639_12160	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	DNA damage-inducible protein DinB	**S**
BAMY6639_12900	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	Alpha-amylase	**G**
BAMY6639_14025	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	Oxidoreductase	**C/ O/ P/ Q**
BAMY6639_16365	■	■	■	■	■	■	■	■	■	■	■	■	■	■	14	Hypothetical protein	**L**
BAMY6639_13935	■	■	■	■	■	■	■	■	■	■	■	■	■		13	Hypothetical protein	**I**
BAMY6639_00515	■	■	■	■	■	■	■		■	■	■	■	■	■	13	Mannonate dehydratase	**G**
BAMY6639_10825	■	■	■	■	■	■		■	■	■	■	■	■	■	13	ABC transporter substrate-binding protein	**E**
BAMY6639_13960	■	■	■	■	■	■	■		■	■	■	■	■	■	13	Hypothetical protein	**K**
BAMY6639_19655	■	■	■	■	■	■	■	■		■	■	■	■	■	13	Amino acid transporter	**J/ R**
BAMY6639_10655	■	■	■		■	■	■		■	■	■	■	■		11	Fic protein family	**S**
BAMY6639_04490	■	■	■	■	■	■				■	■	■	■		10	Type I restriction endonuclease	**V**
BAMY6639_15100	■			■	■	■	■	■	■	■	■	■			10	N-acetyltransferase	**M**
BAMY6614_02720		■	■	■	■	■		■		■	■	■	■		10	Type III restriction endonuclease subunit R	**K/ L/ V**
BAMY6639_07370	■		■	■			■	■	■	■	■		■	■	10	Carboxymuconolactone decarboxylase	**S**
BAMY6639_11105	■		■	■			■	■	■	■	■		■	■	10	4-carboxymuconolactone decarboxylase	**S**
BAMY6639_16685	■	■	■		■	■	■		■			■	■		9	MepB family protein	**S**
BAMY6614_07520		■	■	■	■	■	■		■			■	■		9	Transcriptional regulator	**K**
YP_001420166.1			■	■	■	■	■	■				■	■	■	9	Hypothetical protein	**R/ S**
YP_001420170.1			■	■	■	■	■	■				■	■	■	9	Hypothetical protein	**O**
YP_001420171.1			■	■	■	■	■	■				■	■	■	9	Hypothetical protein	**R**
YP_001421572.1			■	■	■	■	■	■				■	■	■	9	Hypothetical protein	**R**
BAMY6639_07375	■		■	■			■	■	■	■			■	■	9	Amidohydrolase	**R**
BAMY6639_01860	■		■				■	■	■	■	■		■		8	Hypothetical protein	**L**
BAMY6639_03525	■		■				■	■	■	■	■		■		8	Hypothetical protein	**L**
BAMY6639_05670	■		■				■	■	■	■	■		■		8	Hypothetical protein	**L**
BAMY6639_07705	■		■	■			■	■	■	■			■		8	N-acetylglucosamine-6-phosphate deacetylase	**G**
BAMY6639_10305	■			■	■	■	■		■	■		■			8	Wall-associated protein	**M**
BAMY6639_10485	■	■	■		■	■			■			■	■		8	MrsG	**S**
BAMY6639_10490	■	■	■		■	■			■			■	■		8	MrsE	**S**
BAMY6639_11020	■		■				■	■	■	■	■		■		8	Hypothetical protein	**L**
BAMY6639_11065	■	■	■				■		■	■	■		■		8	LysE	**E**
BAMY6639_12945	■		■				■	■	■	■	■		■		8	Hypothetical protein	**L**
BAMY6639_13505	■		■				■	■	■	■	■		■		8	Hypothetical protein	**L**
BAMY6639_13940	■	■	■			■	■	■	■				■		8	Methylated-DNA--protein-cysteine methyltransferase	**S**
BAMY6639_14525	■	■	■		■	■			■			■	■		8	Hypothetical protein	**S**
BAMY6639_14530	■	■	■		■	■			■			■	■		8	Hypothetical protein	**S**
BAMY6639_16090	■		■				■	■	■	■	■		■		8	Hypothetical protein	**L**
BAMY6639_17050	■		■				■	■	■	■	■		■		8	Hypothetical protein	**L**
BAMY6639_17115	■		■				■	■	■	■	■		■		8	Hypothetical protein	**L**
BAMY6639_17770	■		■				■	■	■	■	■		■		8	Transposase	**L**
BAMY6639_18070	■		■				■	■	■	■	■		■		8	Hypothetical protein	**L**
BAMY6639_20240	■		■				■	■	■	■	■		■		8	Tail length tape measure protein	**L**
BAMY6614_06450		■	■		■	■				■	■	■	■		8	TetR family transcriptional regulator	**K**
YP_001422899.1			■	■	■	■			■		■		■	■	8	Hypothetical protein	**T**
YP_005129181.1			■	■			■	■	■	■			■	■	8	Alpha/beta fold family hydrolase	**I/ R**
BAMY6639_03715	■		■		■	■			■			■	■		7	Phosphate-starvation-inducible protein PsiE	**S**
BAMY6639_03735	■		■		■	■			■			■	■		7	Hypothetical protein	**S**
BAMY6639_13125	■		■				■		■	■	■		■		7	Transporter	**E/ G/ R**
BAMY6614_01170		■		■	■	■		■				■		■	7	NAD(P)H-dependent oxidoreductase	**R**
BAMY6614_06505		■		■	■	■		■				■		■	7	Diaminopropionate ammonia-lyase	**E**
BAMY6614_06520		■		■	■	■		■				■		■	7	Septum formation initiator	**E**
YP_005422704.1		■			■	■	■		■			■			6	Maltose/maltodextrin-binding protein	**E/ G**
YP_008411699.1		■			■	■				■	■	■			6	Conserved protein of unknown function	**V**
BAMY6639_09385	■	■					■	■			■	■			6	LuxR family transcriptional regulator	**K/ T**
BAMY6639_05595	■		■	■						■			■	■	6	ArsR family transcriptional regulator	**K**
BAMY6614_00275		■		■	■	■		■						■	6	Hypothetical protein	**L**
YP_001420332.1			■	■						■	■		■		5	Hypothetical protein	**K/ L/ V**



BAMY6639_04485	5	Hypothetical protein **V**
BAMY6639_04495	5	Type I restriction endonuclease EcoKI subunit R **J/ K/ L/ R/ V**
BAMY6639_17535	5	HIT family hydrolase **F/ G/ R**
BAMY6614_00640	5	Saccharopine dehydrogenase **E**
BAMY6614_16370	5	Amino acid permease **E**
BAMY6639_14490	5	Hypothetical protein **V**
BAMY6639_14505	5	Hypothetical protein **V**
BAMY6639_16965	5	Hypothetical protein **R/ V**
YP_001420363.1	4	Hypothetical protein **S**
YP_001420364.1	4	Hypothetical protein **C**
YP_001420366.1	4	Hypothetical protein **H/ R**
YP_005129350.1	4	Type I restriction enzyme, S subunit **V**
YP_005129532.1	4	Alcohol dehydrogenase **C/ R**
YP_008626223.1	4	TP901 family phage tail tape measure protein **D/ M**
BAMY6639_00235	4	Hypothetical protein **R**
BAMY6639_10635	4	Hypothetical protein **O/ R**
BAMY6639_10640	4	Hypothetical protein **V**
BAMY6639_13915	4	Hypothetical protein **R**
BAMY6614_00295	4	Regulatory protein **K**
BAMY6614_00360	4	Maltose phosphorylase **G**
BAMY6614_00365	4	Maltodextrin utilization protein yvdJ **S**
BAMY6614_03500	4	FAD dependent oxidoreductase **C/ R**
BAMY6614_18355	4	Iron ABC transporter **P**
YP_008627612.1	4	Phage major capsid protein **R**
YP_001420144.1	3	Hypothetical protein **R**
YP_001420202.1	3	Hypothetical protein **I/ J**
YP_001420205.1	3	Hypothetical protein **S**
YP_001421569.1	3	Oxidoreductase **C/ R**
YP_005419846.1	3	Major Facilitator Superfamily YbjV **G/ P**
YP_005422249.1	3	Hypothetical protein BANAU_2912 **R**
YP_005423157.1	3	Superfamily I DNA/RNA helicase **L**
YP_005423227.1	3	DNA helicase-like protein **D/ L**
YP_007187174.1	3	Isocitrate/isopropylmalate dehydrogenase **C/ E**
BAMY6614_00215	3	DNA methyltransferase **L**
BAMY6614_07355	3	ABC transporter **M/ Q/ V**
BAMY6614_00290	3	Hypothetical protein **S**
YP_001420331.1	2	Hypothetical protein RBAM_007150 **V**
YP_001420334.1	2	Hypothetical protein RBAM_007180 **V**
YP_001422895.1	2	Hypothetical protein RBAM_033340 **S**
YP_001423272.1	2	Hypothetical protein RBAM_037120 **D/ L**
YP_005129348.1	2	S subunit methyltransferase **V**
YP_005419849.1	2	Fragment maltose O-acetyltransferase, partial **E/ M/ R**
YP_005420440.1	2	ImmF control region 10 kDa protein **K**
YP_005422088.1	2	Transposon Tn552 resolvase **L**
YP_007185366.1	2	Acyl carrier protein **I/ Q**
YP_007188473.1	2	Hypothetical protein B938_19025 **S**
YP_007443494.1	2	DEAD-like helicase **K/ L/ R**
YP_007447243.1	2	PGA-synthesis regulator PgsR **K**
YP_007447244.1	2	Mob **D**
YP_008411505.1	2	Recombination factor protein RarA **L**
YP_008627867.1	2	Protein AMAC1 **E/ G/ R**
BAMY6639_00595	2	Hypothetical protein **S**
YP_005421169.1	2	Hypothetical protein **S**
YP_005422248.1	2	HPP family protein **T**
YP_008411220.1	2	Conserved protein of unknown function **Q**
BAMY6639_12495	2	Hypothetical protein **H/ Q**
BAMY6639_14475	1	Hypothetical protein **K/ O/ T**
BAMY6639_14545	1	Hypothetical protein **S**
BAMY6639_14550	1	Hypothetical protein **S**
BAMY6639_14685	1	Hypothetical protein **S**
BAMY6639_16990	1	Transposase **L**
BAMY6614_03520	1	Antibiotic resistance protein VanZ **V**
BAMY6614_03685	1	Hypothetical protein **E**
BAMY6614_07400	1	Hypothetical protein **S**
BAMY6614_07460	1	Sulfate transporter subunit **P**
BAMY6614_10905	1	Transcriptional regulator **K**
YP_005129125.1	1	Hypothetical protein **S**
YP_005129345.1	1	Putative phage-associated protein **S**
YP_005129525.1	1	Beta-lactamase **V**
YP_005129531.1	1	Putative ABC transporter permease **C/ P**
YP_005132525.1	1	Hypothetical protein **R**
YP_005420746.1	1	HTH-type transcriptional regulator yozG **K**
YP_007185185.1	1	Hypothetical protein **P/ R**
YP_007187322.1	1	Hypothetical protein **E**

YP_007187938.1							■								1	Pentapeptide repeat-containing protein	**S**
YP_007443497.1								■							1	Hypothetical protein	**R**
YP_007443960.1								■							1	Transposase	**L**
YP_007446113.1								■							1	Cysteine protease	**O**
YP_007447058.1								■							1	Chromosome segregation ATPase-like protein	**D/ L**
YP_007496003.1									■						1	Predicted dTDP-4-dehydrorhamnose 3,5-epimerase and related enzymes	**M**
YP_008411613.1										■					1	Putative Phage repressor	**K**
YP_008411629.1										■					1	Conserved protein of unknown function	**S**
YP_008414809.1										■					1	Deoxyguanosinetriphosphate triphosphohydrolase-like protein	**F**
YP_008419921.1											■				1	NADP oxidoreductase, coenzyme f420-dependent	**C/ E/ R**
YP_008420117.1											■				1	Conserved membrane protein of unknown function	**S**
YP_008625285.1												■			1	Protein of unknown function DUF955	**R**
YP_008626258.1												■			1	Hypothetical protein	**S**
YP_008627872.1												■			1	OsmC family protein	**O**
YP_008627873.1												■			1	Alkyl hydroperoxide reductase ahpD	**S**
YP_008627880.1												■			1	OsmC family protein	**O**
YP_008628541.1												■			1	Pentapeptide repeat protein	**S**

*****The numbers at the top of the table are. 1: *B. amyloliquefaciens* CECT 8237; 2: *B. amyloliquefaciens* CECT 8238; 3: *B. amyloliquefaciens* subsp. *plantarum* FZB42; 4: *B. amyloliquefaciens* subsp. *plantarum* CAU-B946; 5: *B. amyloliquefaciens* subsp. *plantarum* YAU B9601-Y2; 6: *B. amyloliquefaciens* Y2; 7: *B. amyloliquefaciens* subsp. *plantarum* AS43.3; 8: *B. amyloliquefaciens* IT-45; 9: *B. amyloliquefaciens* subsp. *plantarum* UCMB5036; 10: *B. amyloliquefaciens* subsp. *plantarum* UCMB5033; 11: *B. amyloliquefaciens* subsp. *plantarum* UCMB5113; 12: *B. amyloliquefaciens* subsp. *plantarum* NAU-B3; 13: *B. amyloliquefaciens* CC178; 14: *B. amyloliquefaciens* LFB112.

******Blue or white colours represent the presence or the absence, respectively, of a particular gene in a specific strain.

Table 3. Genes contained in Atypical Regions of *B. amyloliquefaciens* CECT 8237 strain.

Genes (GenBank code)	Start	End	Product
Atypical Region 1			
BAMY6639_00145	25085	25498	Hypothetical protein
BAMY6639_00150	25701	25955	Hypothetical protein
BAMY6639_00155	25996	26562	Hypothetical protein
BAMY6639_00160	26661	28460	Hypothetical protein
BAMY6639_00165	28475	28933	Hypothetical protein
BAMY6639_00170	29021	29431	Hypothetical protein
BAMY6639_00175	31168	31353	Hypothetical protein
BAMY6639_00180	31376	32767	MFS transporter
BAMY6639_00185	32810	34411	Beta-xylosidase
Atypical Region 2			
BAMY6639_00235	45798	45959	Hypothetical protein
BAMY6639_00240	46393	46515	Hypothetical protein
BAMY6639_00245	46531	46734	Hypothetical protein
BAMY6639_00250	46725	46970	Hypothetical protein
BAMY6639_00255	47123	47551	Hypothetical protein
BAMY6639_00260	47604	48365	N-acetyltransferase
BAMY6639_00265	49755	50375	Hypothetical protein
Atypical Region 3			
BAMY6639_00550	94877	95095	Hypothetical protein
BAMY6639_00555	96045	96647	Hypothetical protein
BAMY6639_00560	96769	97104	Hypothetical protein
BAMY6639_00565	97370	97513	Hypothetical protein
BAMY6639_00575	98897	98986	Hypothetical protein
BAMY6639_00580	99143	99421	Hypothetical protein
BAMY6639_00585	99585	100250	Hypothetical protein
BAMY6639_00590	100563	100922	Hypothetical protein
BAMY6639_00595	101008	103173	Hypothetical protein
BAMY6639_00600	103170	103790	Bacitracin ABC transporter ATP-binding protein
BAMY6639_00605	103804	104013	Hypothetical protein
BAMY6639_00610	104314	104583	Hypothetical protein
BAMY6639_00615	105905	106009	Hypothetical protein
BAMY6639_00620	106265	107764	Endoglucanase
Atypical Region 4			
BAMY6639_00780	174161	177964	Peptide synthetase
BAMY6639_00785	177983	184198	Peptide synthetase
BAMY6639_00795	188784	196433	Peptide synthetase
BAMY6639_00800	196449	204146	Peptide synthetase
BAMY6639_00805	204172	211830	Peptide synthetase

BAMY6639_00810	212310	213785	D-alanyl-D-alanine carboxypeptidase
BAMY6639_00815	213804	214778	Aldose 1-epimerase
Atypical Region 5			
BAMY6639_01380	312464	312976	Transcriptional regulator
BAMY6639_01385	313048	313188	Hypothetical protein
BAMY6639_01390	313267	314046	N-acetyltransferase
BAMY6639_01395	314085	315356	*CgeD*
BAMY6639_01400	315421	315720	*CgeC*
BAMY6639_01405	315931	316362	*CgeA*
BAMY6639_01410	316369	317325	*CgeB*
BAMY6639_01415	317375	318526	3-phytase
BAMY6639_01420	318842	319267	Polysaccharide biosynthesis protein *CapD*
BAMY6639_01425	319301	319477	Hypothetical protein
BAMY6639_01430	319750	320007	Hypothetical protein
BAMY6639_01435	320554	320787	Hypothetical protein
BAMY6639_01440	320811	320930	Hypothetical protein
BAMY6639_01445	321631	322068	Hypothetical protein
BAMY6639_01450	322132	322653	Hypothetical protein
BAMY6639_01455	323297	323761	Hypothetical protein
BAMY6639_01460	323954	324259	Hypothetical protein
BAMY6639_01465	324286	324471	Hypothetical protein
BAMY6639_01470	324813	325187	Hypothetical protein
BAMY6639_01475	325399	325851	Hypothetical protein
BAMY6639_01480	326221	327336	Aspartate phosphatase
BAMY6639_01485	327333	327455	Phosphatase
BAMY6639_01490	327923	328519	Hypothetical protein
BAMY6639_01495	328521	330272	Hypothetical protein
BAMY6639_01500	330309	330842	Hypothetical protein
BAMY6639_01505	331008	331973	Endonuclease
BAMY6639_01510	332189	333856	Resolvase
BAMY6639_01515	333897	334469	Hypothetical protein
Atypical Region 6			
BAMY6639_02615	518398	519144	Polyketide biosynthesis enoyl-CoA hydratase
BAMY6639_02620	519204	520451	3-hydroxy-3-methylglutaryl-ACP synthase
BAMY6639_02625	520509	521663	Cytochrome P450
BAMY6639_02630	534132	541850	Polyketide synthase
BAMY6639_02635	541855	557469	Polyketide synthase
BAMY6639_02640	557521	563247	Polyketide synthase
BAMY6639_02645	563287	569583	Polyketide synthase

BAMY6639_02650	569602	582240	Polyketide synthase
Atypical Region 7			
BAMY6639_03730	767740	767853	Hypothetical protein
BAMY6639_03735	767920	769137	Hypothetical protein
BAMY6639_03740	769503	769598	Hypothetical protein
BAMY6639_03745	769604	770209	Hypothetical protein
BAMY6639_03750	770538	771071	Hypothetical protein
BAMY6639_03755	771553	772929	Tetracycline resistance protein
BAMY6639_03760	773449	774132	Hypothetical protein
BAMY6639_03765	774334	774456	Hypothetical protein
BAMY6639_03770	774682	775155	Hypothetical protein
BAMY6639_03775	775177	777105	Hypothetical protein
BAMY6639_03780	777306	777575	Hypothetical protein
BAMY6639_03785	777572	778156	Hypothetical protein
BAMY6639_03790	778336	779373	Hypothetical protein
BAMY6639_03795	779360	779893	Hypothetical protein
BAMY6639_03800	779914	781131	Hypothetical protein
BAMY6639_03805	781818	782294	Hypothetical protein
Atypical Region 8			
BAMY6639_04460	913084	913188	Hypothetical protein
BAMY6639_04465	913472	913612	Hypothetical protein
BAMY6639_04470	913611	913940	DNA helicase ino80
BAMY6639_04475	913980	914165	Hypothetical protein
BAMY6639_04480	914769	915011	Hypothetical protein
BAMY6639_04485	915356	916798	Hypothetical protein
BAMY6639_04490	916795	918225	Type I restriction endonuclease
BAMY6639_04495	918358	921564	Type I restriction endonuclease EcoKI subunit R
BAMY6639_04500	921775	922047	Hypothetical protein
Atypical Region 9			
BAMY6639_05150	1049855	1050412	Hypothetical protein
BAMY6639_05155	1051436	1051969	Hypothetical protein
BAMY6639_05160	1051983	1052789	Hypothetical protein
BAMY6639_05165	1052897	1053235	Hypothetical protein
BAMY6639_05170	1053283	1053489	Hypothetical protein
BAMY6639_05175	1053673	1053825	Hypothetical protein
BAMY6639_05180	1053869	1054285	Pilus biosynthesis protein *HicB*
BAMY6639_05185	1054916	1055005	Hypothetical protein
BAMY6639_05190	1055187	1055960	Hypothetical protein
BAMY6639_05195	1056008	1056115	Hypothetical protein
BAMY6639_05200	1056430	1056561	Hypothetical protein
Atypical Region 10			
BAMY6639_07360	1446464	1447315	DNA-binding protein

BAMY6639_07365	1447419	1448636	Major facilitator transporter
BAMY6639_07370	1448636	1449052	Carboxymuconolactone decarboxylase
BAMY6639_07375	1449056	1449841	Amidohydrolase
BAMY6639_07380	1449944	1450042	Hypothetical protein
BAMY6639_07385	1450043	1451371	Monooxygenase
BAMY6639_07390	1451368	1451655	Glutaredoxin
Atypical Region 11			
BAMY6639_07585	1483127	1484362	MFS transporter
BAMY6639_07590	1484337	1485257	Endonuclease
BAMY6639_07595	1485265	1485942	Acetylglucosaminylphosphatidylinositol deacetylase
BAMY6639_07600	1485905	1486663	Hypothetical protein
BAMY6639_07605	1486636	1487517	FemAB family protein
BAMY6639_07610	1487529	1488785	Perosamine synthetase
BAMY6639_07615	1488751	1489680	NAD-dependent dehydratase
BAMY6639_07620	1489677	1490807	Spore coat protein
BAMY6639_07625	1490856	1492190	UDP-glucose 6-dehydrogenase
BAMY6639_07630	1492457	1493737	LytTR family transcriptional regulator
Atypical Region 12			
BAMY6639_07885	1547441	1547599	Hypothetical protein
BAMY6639_07890	1547854	1548195	Hypothetical protein
BAMY6639_07895	1548273	1548587	Hypothetical protein
BAMY6639_07900	1548644	1549084	Hypothetical protein
BAMY6639_07905	1549239	1549526	Hypothetical protein
BAMY6639_07910	1549988	1550332	Hypothetical protein
BAMY6639_07915	1550329	1550664	Hypothetical protein
BAMY6639_07920	1550697	1550954	Hypothetical protein
BAMY6639_07925	1551063	1552490	Lipase
BAMY6639_07930	1552507	1552818	Hypothetical protein
BAMY6639_07940	1553050	1553160	Hypothetical protein
Atypical Region 13			
BAMY6639_08725	1713316	1714032	Endonuclease V
BAMY6639_08730	1714073	1714573	Hypothetical protein
BAMY6639_08735	1714765	1715205	Hypothetical protein
BAMY6639_08740	1715384	1715815	Hypothetical protein
BAMY6639_08745	1715825	1717969	Transposase
BAMY6639_08750	1717989	1718249	Hypothetical protein
BAMY6639_08755	1718249	1718683	Hypothetical protein
Atypical Region 14			
BAMY6639_10235	1987679	1988053	Hypothetical protein
BAMY6639_10240	1988385	1988822	Hypothetical protein
BAMY6639_10245	1988752	1989597	DNA-binding protein

BAMY6639_10250	1989693	1990520	Hypothetical protein
BAMY6639_10255	1990564	1990824	Hypothetical protein
BAMY6639_10260	1990821	1991042	Hypothetical protein
BAMY6639_10265	1991102	1991509	Hypothetical protein
BAMY6639_10270	1991513	1991923	Hypothetical protein
BAMY6639_10275	1992133	1992579	Hypothetical protein
BAMY6639_10280	1992563	1993120	Hypothetical protein
BAMY6639_10285	1993195	1993443	Hypothetical protein
BAMY6639_10290	1993452	1993952	Hypothetical protein
BAMY6639_10295	1994055	1994729	Hypothetical protein
BAMY6639_10300	1994833	1995528	Hypothetical protein
BAMY6639_10305	1995531	2002568	Wall-associated protein
BAMY6639_10310	2002731	2003795	Pectate lyase
Atypical Region 15			
BAMY6639_10670	2079026	2079484	(2Fe-2S)-binding protein
BAMY6639_10675	2079477	2080319	Xanthine dehydrogenase
BAMY6639_10680	2080306	2082636	Aldehyde oxidase
BAMY6639_10685	2082642	2083262	Xanthine dehydrogenase
BAMY6639_10690	2083259	2084275	Xanthine dehydrogenase
BAMY6639_10695	2084344	2085360	Hypothetical protein
BAMY6639_10700	2085675	2086229	Hypothetical protein
BAMY6639_10705	2086288	2086449	Membrane protein
BAMY6639_10710	2086552	2086710	Hypothetical protein
BAMY6639_10715	2086885	2087058	Hypothetical protein
BAMY6639_10720	2087068	2089011	Hypothetical protein
BAMY6639_10725	2089071	2089232	Hypothetical protein
BAMY6639_10730	2089814	2090119	Hypothetical protein
BAMY6639_10735	2090109	2091248	Hypothetical protein
Atypical Region 16			
BAMY6639_10820	2106366	2107556	Major facilitator transporter
BAMY6639_10825	2107575	2108897	ABC transporter substrate-binding protein
BAMY6639_10830	2108894	2109901	Molybdenum cofactor synthesis protein 3
BAMY6639_10835	2110133	2111551	Sensor histidine kinase
BAMY6639_10840	2111548	2111757	Hypothetical protein
BAMY6639_10845	2111992	2112096	Hypothetical protein
BAMY6639_10850	2112101	2112184	Hypothetical protein
Atypical Region 17			
BAMY6639_10940	2124405	2124680	Hypothetical protein
BAMY6639_10945	2124760	2126052	Glycosyl transferase family 1
BAMY6639_10950	2126075	2127130	Hypothetical protein
BAMY6639_10955	2127123	2128310	Carbamoyl-phosphate synthase

BAMY6639_10960	2128300	2129634	Glycogen synthase
BAMY6639_10965	2129631	2130287	GlcNAc-PI de-N-acetylase
BAMY6639_10975	2130974	2131072	Hypothetical protein
Atypical Region 18			
BAMY6639_12460	2396621	2397829	ABC transporter permease
BAMY6639_12465	2397832	2398998	Multidrug ABC transporter
BAMY6639_12470	2399054	2399227	Membrane protein
BAMY6639_12475	2399369	2399995	Hypothetical protein
BAMY6639_12480	2418792	2420045	Malonyl-CoA transacylase
BAMY6639_12485	2420089	2424861	Hypothetical protein
BAMY6639_12490	2425046	2429836	Hypothetical protein
BAMY6639_12495	2429858	2431381	Hypothetical protein
BAMY6639_12500	2431479	2438105	Hypothetical protein
BAMY6639_12505	2438126	2439169	Hypothetical protein
BAMY6639_12510	2439326	2439802	Hypothetical protein
BAMY6639_12515	2439880	2440131	Sporulation protein
BAMY6639_12520	2440241	2440939	Hypothetical protein
Atypical Region 19			
BAMY6639_13570	2675116	2675553	Hypothetical protein
BAMY6639_13575	2675589	2676146	Hypothetical protein
BAMY6639_13580	2676210	2676659	Hypothetical protein
BAMY6639_13585	2677318	2678448	Aspartate phosphatase
BAMY6639_13590	2678448	2678582	Hypothetical protein
BAMY6639_13595	2678941	2679171	Hypothetical protein
BAMY6639_13600	2679277	2679417	Hypothetical protein
BAMY6639_13605	2679551	2680108	Hypothetical protein
BAMY6639_13610	2680204	2680821	Hypothetical protein
Atypical Region 20			
BAMY6639_14030	2743518	2744153	Hypothetical protein
BAMY6639_14035	2744459	2745253	Dihydrodipicolinate synthase
BAMY6639_14040	2745397	2745600	Hypothetical protein
BAMY6639_14045	2745752	2746801	Radical SAM protein
BAMY6639_14050	2746911	2747291	Membrane protein
BAMY6639_14055	2747560	2748024	Hypothetical protein
BAMY6639_14060	2748021	2749031	Thymidylate synthase
BAMY6639_14065	2749028	2749786	SAM-dependent methyltransferase
Atypical Region 21			
BAMY6639_14465	2819520	2820236	Hypothetical protein
BAMY6639_14470	2820233	2820364	Hypothetical protein
BAMY6639_14475	2820390	2820995	Hypothetical protein
BAMY6639_14480	2820988	2822286	Hypothetical protein
BAMY6639_14485	2822511	2822717	Hypothetical protein
BAMY6639_14490	2822793	2825936	Hypothetical protein

BAMY6639_14495	2825941	2826162	Hypothetical protein
BAMY6639_14500	2826272	2826430	Hypothetical protein
BAMY6639_14505	2826577	2829648	Hypothetical protein
BAMY6639_14510	2829645	2831792	Hypothetical protein
BAMY6639_14515	2831903	2832205	Hypothetical protein
Atypical Region 22			
BAMY6639_15130	2962596	2963336	Hypothetical protein
BAMY6639_15135	2963789	2966209	Hypothetical protein
BAMY6639_15140	2966411	2966506	Hypothetical protein
BAMY6639_15145	2970033	2970182	Hypothetical protein
BAMY6639_15150	2970836	2971618	Hypothetical protein
BAMY6639_15155	2971773	2974862	Histidine kinase
BAMY6639_15160	2975043	2976119	Transcriptional regulator
BAMY6639_15165	2976300	2982209	Cell wall anchor
BAMY6639_15170	2982317	2982970	Sortase
Atypical Region 23			
BAMY6639_17430	3393065	3393421	tRNA-Val4
BAMY6639_17435	3393937	3395289	Transposase
BAMY6639_17440	3395423	3395725	Hypothetical protein
BAMY6639_17445	3396096	3396488	Hypothetical protein
BAMY6639_17450	3396670	3398289	Hypothetical protein
BAMY6639_17455	3399070	3399522	Hypothetical protein
BAMY6639_17460	3399541	3399942	Hypothetical protein
BAMY6639_17465	3400373	3400564	Hypothetical protein
BAMY6639_17470	3400565	3400882	Hypothetical protein
BAMY6639_17475	3400993	3401160	Hypothetical protein
BAMY6639_17480	3401294	3402412	Hypothetical protein
BAMY6639_17485	3402498	3402716	Hypothetical protein
BAMY6639_17490	3402716	3403402	Hypothetical protein
BAMY6639_17495	3404392	3404727	tRNA-Val4
BAMY6639_17500	3404749	3404976	Hypothetical protein
BAMY6639_17505	3404999	3405211	Hypothetical protein
BAMY6639_17510	3405902	3406450	Hypothetical protein
BAMY6639_17515	3406650	3406823	Hypothetical protein
BAMY6639_17520	3406854	3407552	Hypothetical protein
BAMY6639_17525	3407600	3407974	Hypothetical protein
BAMY6639_17530	3408225	3408602	Hypothetical protein
BAMY6639_17535	3408835	3409164	HIT family hydrolase
BAMY6639_17540	3409180	3409746	Hypothetical protein
BAMY6639_17545	3410458	3410820	Hypothetical protein
BAMY6639_17550	3410887	3411099	Hypothetical protein
Atypical Region 24			
BAMY6639_17740	3445170	3446615	Altronate oxidoreductase

BAMY6639_17745	3446612	3448105	Altronate hydrolase
BAMY6639_17750	3448290	3448961	Transcriptional regulator
BAMY6639_17755	3448958	3450316	Sensor histidine kinase
BAMY6639_17760	3450436	3450636	Hypothetical protein
BAMY6639_17765	3450778	3450969	Hypothetical protein
BAMY6639_17770	3451071	3451871	Transposase
BAMY6639_17775	3451889	3452221	Transposase
Atypical Region 25			
BAMY6639_17865	3462828	3463625	Terminase
BAMY6639_17870	3463622	3464920	Terminase
BAMY6639_17875	3464924	3466360	Phage portal protein
BAMY6639_17880	3466380	3467225	Phage portal protein
BAMY6639_17885	3467252	3468187	Phage portal protein
BAMY6639_17890	3468204	3468587	Hypothetical protein
BAMY6639_17895	3468584	3468940	Phage portal protein
BAMY6639_17900	3468937	3469440	Phage portal protein
BAMY6639_17905	3469437	3469883	Phage portal protein
BAMY6639_17910	3469880	3470089	Hypothetical protein
BAMY6639_17915	3470089	3471486	Phage portal protein
Atypical Region 26			
BAMY6639_17925	3472008	3472454	Phage portal protein
BAMY6639_17930	3472478	3472648	Hypothetical protein
BAMY6639_17935	3472636	3477597	Phage portal protein
BAMY6639_17940	3477590	3478249	Phage portal protein
BAMY6639_17945	3478263	3479240	Phage portal protein
BAMY6639_17950	3479240	3479506	Phage portal protein
BAMY6639_17955	3479487	3479585	Hypothetical protein
BAMY6639_17960	3479610	3480035	Phage portal protein
BAMY6639_17965	3480028	3481074	Phage portal protein
BAMY6639_17970	3481058	3481636	Phage portal protein
BAMY6639_17975	3481633	3481905	Hypothetical protein
BAMY6639_17980	3481908	3483497	Hypothetical protein
BAMY6639_17985	3483510	3483935	Hypothetical protein
BAMY6639_17990	3483940	3484137	Phage portal protein
BAMY6639_17995	3484194	3484955	Phage portal protein
Atypical Region 27			
BAMY6639_18915	3654260	3654454	Hypothetical protein
BAMY6639_18920	3654455	3656761	Poly(3-hydroxyalkanoate) synthetase
BAMY6639_18925	3656783	3669037	Polyketide synthase
BAMY6639_18930	3669037	3673809	Polyketide synthase
BAMY6639_18935	3673857	3682565	Polyketide synthase
BAMY6639_18940	3682558	3689562	Polyketide synthase

BAMY6639_18945	3689586	3695297	Polyketide synthase
BAMY6639_18950	3695297	3702676	Polyketide synthase
BAMY6639_18955	3702727	3706578	*MlnH*
BAMY6639_18960	3706611	3707702	*MlnI*
Atypical Region 28			
BAMY6639_20220	3961540	3961905	Integrase
BAMY6639_20225	3962366	3963004	Hypothetical protein
BAMY6639_20230	3963080	3963445	HNH endonuclease
BAMY6639_20235	3963731	3964063	Transposase
BAMY6639_20240	3964081	3964878	Tail length tape measure protein
BAMY6639_20245	3964949	3965764	Hypothetical protein
BAMY6639_20250	3965897	3966007	Hypothetical protein
BAMY6639_20255	3966057	3966266	Hypothetical protein
BAMY6639_20260	3967006	3967683	Polyketide biosynthesis protein

Table 4. Genes contained in Atypical Regions of *B. amyloliquefaciens* CECT 8238 strain.

Genes (GenBank code)	Start	End	Product
Atypical Region 1			
BAMY6614_00030	2469	2681	*bhlA*
BAMY6614_00035	2733	2918	Hypothetical protein
BAMY6614_00040	2918	3280	Hypothetical protein
BAMY6614_00045	3277	4413	Hypothetical protein
BAMY6614_00050	4426	6990	Peptidase G2
BAMY6614_00055	7023	8741	Alkaline phosphatase
BAMY6614_00060	8754	9590	Hypothetical protein
BAMY6614_00065	9605	13351	Hypothetical protein
BAMY6614_00070	13412	13594	Hypothetical protein
BAMY6614_00075	13606	13968	Hypothetical protein
BAMY6614_00080	14026	14604	UDP-N-acetylmuramoylalanine--D-glutamate ligase
BAMY6614_00085	14624	15013	Hypothetical protein
BAMY6614_00090	15010	15399	Hypothetical protein
BAMY6614_00095	15399	15725	Hypothetical protein
BAMY6614_00100	15715	16008	Hypothetical protein
BAMY6614_00105	16060	16464	Hypothetical protein
BAMY6614_00110	16492	17691	Capsid protein
BAMY6614_00115	17740	18336	Peptidase U35
BAMY6614_00120	18329	19555	Phage portal protein
BAMY6614_00125	19560	19766	Hypothetical protein
BAMY6614_00130	19783	21516	Terminase
BAMY6614_00135	21506	21991	Terminase
BAMY6614_00140	22238	22633	Hypothetical protein
BAMY6614_00145	22614	22883	Hypothetical protein
BAMY6614_00150	23020	23133	Hypothetical protein
BAMY6614_00155	23179	23391	Cell division protein *FtsK*
Atypical Region 2			
BAMY6614_00165	24405	24857	ArpU family transcriptional regulator
BAMY6614_00170	24912	25088	Hypothetical protein
BAMY6614_00175	25219	25494	Hypothetical protein
BAMY6614_00180	25523	25705	Hypothetical protein
BAMY6614_00185	25716	26150	Hypothetical protein
BAMY6614_00190	26153	26770	Hypothetical protein
BAMY6614_00195	26767	27012	Hypothetical protein
BAMY6614_00200	27150	27710	dUTPase
BAMY6614_00205	27724	27861	Hypothetical protein
BAMY6614_00210	27858	28196	Hypothetical protein
BAMY6614_00215	28305	29129	DNA methyltransferase

BAMY6614_00220	29133	29393	Hypothetical protein
BAMY6614_00225	29406	29807	Hypothetical protein
BAMY6614_00230	29804	30184	Hypothetical protein
BAMY6614_00235	30181	30291	Hypothetical protein
BAMY6614_00240	30325	30534	Phage portal protein
BAMY6614_00245	30608	30763	Hypothetical protein
BAMY6614_00250	30776	30916	Hypothetical protein
BAMY6614_00255	30889	31056	Hypothetical protein
BAMY6614_00260	31025	31573	Hypothetical protein
BAMY6614_00265	31570	31728	Hypothetical protein
BAMY6614_00270	31725	31874	Hypothetical protein
BAMY6614_00275	31889	32776	Hypothetical protein
BAMY6614_00280	32703	33581	Replication protein
BAMY6614_00285	33568	33792	Hypothetical protein
BAMY6614_00290	33810	34133	Hypothetical protein
BAMY6614_00295	34176	34382	Regulatory protein
BAMY6614_00300	34545	34937	Transcriptional regulator
BAMY6614_00310	35226	35354	Hypothetical protein
BAMY6614_00315	35370	36458	Hypothetical protein
BAMY6614_00320	36680	37789	Integrase
BAMY6614_00330	38237	39706	Sucrose-6-phosphate hydrolase
BAMY6614_00335	39721	40953	Galactoside permease
Atypical Region 3			
BAMY6614_00630	104743	105480	Thioesterase
BAMY6614_00635	105531	106361	Hypothetical protein
BAMY6614_00640	106685	107749	Saccharopine dehydrogenase
BAMY6614_00645	107840	109564	ABC transporter permease
BAMY6614_00650	109762	121068	Hypothetical protein
BAMY6614_00655	121080	138182	Peptide synthetase
BAMY6614_00665	139670	140032	Hypothetical protein
BAMY6614_00670	140165	140263	Hypothetical protein
Atypical Region 4			
BAMY6614_01180	242307	243023	Endonuclease V
BAMY6614_01195	244418	244780	Hypothetical protein
BAMY6614_01205	245444	245953	Hypothetical protein
BAMY6614_01210	245974	246294	Hypothetical protein
BAMY6614_01220	247169	247474	Hypothetical protein
BAMY6614_01225	247491	249236	Transposase
BAMY6614_01230	249256	249516	Hypothetical protein
BAMY6614_01235	249516	249950	Hypothetical protein
Atypical Region 5			
BAMY6614_03050	592420	592884	(2Fe-2S)-binding protein
BAMY6614_03055	592877	593719	Xanthine dehydrogenase

BAMY6614_03060	593706	596036	Aldehyde oxidase
BAMY6614_03065	596042	596662	Xanthine dehydrogenase
BAMY6614_03070	596659	597675	Xanthine dehydrogenase
BAMY6614_03075	597745	598761	Hypothetical protein
BAMY6614_03080	599078	599632	Hypothetical protein
BAMY6614_03085	599691	599852	Membrane protein
BAMY6614_03090	600023	600739	Hypothetical protein
BAMY6614_03095	600886	601878	Cobyrinic acid a,c-diamide synthase
BAMY6614_03100	602535	604478	Hypothetical protein
Atypical Region 6			
BAMY6614_03315	638301	638618	Hypothetical protein
BAMY6614_03320	638698	639990	Glycosyl transferase family 1
BAMY6614_03325	640013	641068	Hypothetical protein
BAMY6614_03330	641061	642248	Carbamoyl-phosphate synthase
BAMY6614_03335	642238	643572	Glycogen synthase
BAMY6614_03340	643569	644225	GlcNAc-PI de-N-acetylase
BAMY6614_03350	644912	645010	Hypothetical protein
Atypical Region 7			
BAMY6614_03515	674141	674449	Hypothetical protein
BAMY6614_03520	674602	675153	Antibiotic resistance protein *vanZ*
BAMY6614_03525	675244	675651	Hypothetical protein
BAMY6614_03540	676939	677229	Hypothetical protein
BAMY6614_03545	677669	679177	Hypothetical protein
BAMY6614_03550	679287	680480	Aspartate phosphatase
BAMY6614_03555	680885	681322	Terpinolene synthase
BAMY6614_03560	681343	681813	Hypothetical protein
BAMY6614_03565	682266	682652	Hypothetical protein
BAMY6614_03570	682715	683227	Hypothetical protein
BAMY6614_03575	683240	684250	Peptidase P60
BAMY6614_03580	684247	686607	Hypothetical protein
BAMY6614_03585	686618	686944	Hypothetical protein
BAMY6614_03590	686958	689453	Hypothetical protein
BAMY6614_03595	689341	689865	Hypothetical protein
BAMY6614_03600	689877	690125	Hypothetical protein
BAMY6614_03605	690141	691205	Hypothetical protein
BAMY6614_03610	691218	691346	Hypothetical protein
BAMY6614_03615	691364	691660	Hypothetical protein
BAMY6614_03620	691666	691938	Hypothetical protein
BAMY6614_03625	691963	692214	Hypothetical protein
BAMY6614_03630	692239	692511	Hypothetical protein
BAMY6614_03635	692525	692800	Hypothetical protein
BAMY6614_03640	692935	693237	Hypothetical protein
BAMY6614_03645	693500	694558	DNA polymerase

BAMY6614_03650	694551	695972	Cell division protein *FtsK*
BAMY6614_03655	696356	696925	Hypothetical protein
BAMY6614_03660	696964	697332	Hypothetical protein
BAMY6614_03665	697693	697953	Hypothetical protein
BAMY6614_03670	698004	698174	Hypothetical protein
BAMY6614_03675	698261	698449	ICEBs1 excisionase
BAMY6614_03680	698726	699106	XRE family transcriptional regulator
BAMY6614_03685	699130	699591	Hypothetical protein
BAMY6614_03690	699614	700429	Hypothetical protein
BAMY6614_03695	700447	701730	Hypothetical protein
BAMY6614_03700	701801	701962	Thymidylate kinase
BAMY6614_03705	701992	703881	Hypothetical protein
BAMY6614_03710	703878	704774	Hypothetical protein
Atypical Region 8			
BAMY6614_05040	940506	941693	ABC transporter permease
BAMY6614_05045	941690	942832	Multidrug ABC transporter
BAMY6614_05050	942846	943025	Membrane protein
BAMY6614_05055	943129	943605	Hypothetical protein
BAMY6614_05060	943686	943934	Sporulation protein
BAMY6614_05065	944043	944741	Hypothetical protein
BAMY6614_05070	944787	945047	Hypothetical protein
Atypical Region 9			
BAMY6614_05115	952705	954150	4-hydroxyphenylacetate 3-monooxygenase
BAMY6614_05120	954354	954779	Hypothetical protein
BAMY6614_05125	954853	957021	Hypothetical protein
BAMY6614_05130	957081	957209	Hypothetical protein
BAMY6614_05135	957401	957844	Hypothetical protein
BAMY6614_05140	958108	958578	XRE family transcriptional regulator
BAMY6614_05145	958651	959721	Alcohol dehydrogenase
Atypical Region 10			
BAMY6614_06070	1171632	1171781	Hypothetical protein
BAMY6614_06075	1171846	1172277	N-acetylmuramoyl-L-alanine amidase
BAMY6614_06080	1172597	1173382	Hypothetical protein
BAMY6614_06085	1173396	1173953	Hypothetical protein
BAMY6614_06090	1174048	1174326	Hypothetical protein
BAMY6614_06095	1174626	1175759	Aspartate phosphatase
BAMY6614_06100	1175756	1175890	Hypothetical protein
BAMY6614_06105	1176274	1176504	Hypothetical protein
BAMY6614_06110	1176688	1176867	Hypothetical protein
BAMY6614_06115	1176881	1177438	Hypothetical protein
BAMY6614_06120	1177535	1178152	Hypothetical protein
Atypical Region 11			
BAMY6614_06580	1253820	1254161	Hypothetical protein

BAMY6614_06585	1254438	1254563	Hypothetical protein
BAMY6614_06590	1254676	1255203	Hypothetical protein
BAMY6614_06595	1255448	1256242	Dihydrodipicolinate synthase
BAMY6614_06600	1256428	1256589	Hypothetical protein
BAMY6614_06605	1256740	1257789	Radical SAM protein
BAMY6614_06610	1257898	1258281	Membrane protein
BAMY6614_06615	1258396	1259358	Arsenic resistance protein
Atypical Region 12			
BAMY6614_07355	1408852	1411347	ABC transporter
BAMY6614_07360	1411609	1412562	ABC transporter
BAMY6614_07365	1412563	1413438	Transporter
BAMY6614_07370	1413524	1413760	Hypothetical protein
BAMY6614_07375	1413974	1414444	Hypothetical protein
BAMY6614_07380	1414448	1416454	Hypothetical protein
BAMY6614_07385	1416546	1416845	Hypothetical protein
BAMY6614_07390	1416851	1418731	Hypothetical protein
Atypical Region 13			
BAMY6614_07455	1425206	1426072	Sulfate transporter
BAMY6614_07460	1426345	1427382	Sulfate transporter subunit
BAMY6614_07465	1427398	1428234	Sulfate/thiosulfate transporter subunit
BAMY6614_07470	1428245	1429108	Sulfate ABC transporter permease
BAMY6614_07475	1429122	1430192	Sulfate ABC transporter ATP-binding protein
BAMY6614_07480	1430311	1430484	Hypothetical protein
BAMY6614_07485	1430514	1430816	Regulatory protein
Atypical Region 14			
BAMY6614_09915	1887462	1887803	tRNA-Val4
BAMY6614_09920	1888085	1888240	Hypothetical protein
BAMY6614_09925	1888237	1888539	Hypothetical protein
BAMY6614_09930	1888592	1888900	Hypothetical protein
BAMY6614_09935	1888934	1889326	Hypothetical protein
BAMY6614_09940	1889338	1891116	Hypothetical protein
BAMY6614_09945	1891604	1891777	Hypothetical protein
BAMY6614_09950	1892064	1892807	2-hydroxy-6-oxononadienedioate/2-hydroxy-6- oxononatrienedioate hydrolase 2
BAMY6614_09955	1892988	1893092	Hypothetical protein
BAMY6614_09960	1893164	1893493	HIT family hydrolase
BAMY6614_09965	1893571	1894041	Hypothetical protein
BAMY6614_09970	1894563	1894733	Hypothetical protein
BAMY6614_09975	1894730	1895092	Hypothetical protein
BAMY6614_09980	1895472	1895756	Hypothetical protein
Atypical Region 15			
BAMY6614_10270	1942870	1943667	Terminase

BAMY6614_10275	1943664	1944962	Terminase
BAMY6614_10285	1946422	1947267	Phage portal protein
BAMY6614_10290	1947294	1948229	Phage portal protein
BAMY6614_10295	1948246	1948629	Hypothetical protein
BAMY6614_10300	1948626	1948982	Phage portal protein
BAMY6614_10305	1948979	1949482	Phage portal protein
BAMY6614_10310	1949479	1949925	Phage portal protein
BAMY6614_10315	1949922	1950131	Hypothetical protein
BAMY6614_10320	1950131	1951528	Phage portal protein
BAMY6614_10325	1951530	1951973	Phage portal protein
BAMY6614_10330	1952048	1952494	Phage portal protein
BAMY6614_10335	1952518	1952688	Hypothetical protein
BAMY6614_10340	1952676	1957721	Phage portal protein
BAMY6614_10345	1957714	1958373	Phage portal protein
BAMY6614_10350	1958387	1959364	Phage portal protein
BAMY6614_10355	1959364	1959630	Phage portal protein
BAMY6614_10360	1959611	1959709	Hypothetical protein
BAMY6614_10365	1959734	1960159	Phage portal protein
BAMY6614_10370	1960152	1961198	Phage portal protein
BAMY6614_10375	1961182	1961760	Phage portal protein
BAMY6614_10380	1961757	1962029	Hypothetical protein
BAMY6614_10385	1962032	1963654	Hypothetical protein
BAMY6614_10390	1963667	1964038	Hypothetical protein
BAMY6614_10395	1964043	1964240	*XkdX*
BAMY6614_10400	1964297	1965058	Phage portal protein
Atypical Region 16			
BAMY6614_11325	2132661	2132870	Hypothetical protein
BAMY6614_11330	2132871	2135177	Malonyl CoA-ACP transacylase
BAMY6614_11335	2135199	2147456	Polyketide synthase
BAMY6614_11340	2147456	2152228	Polyketide synthase
BAMY6614_11345	2152276	2160987	Polyketide synthase
BAMY6614_11350	2160980	2167978	Polyketide synthase
BAMY6614_11355	2169631	2173710	Polyketide synthase
BAMY6614_11360	2173710	2181092	Polyketide synthase
BAMY6614_11365	2181143	2184991	*MlnH*
BAMY6614_11370	2185024	2186115	*MlnI*
Atypical Region 17			
BAMY6614_12645	2440945	2441709	Hypothetical protein
BAMY6614_12650	2441851	2442033	Hypothetical protein
BAMY6614_12655	2442105	2442206	Hypothetical protein
BAMY6614_12660	2442203	2442547	Hypothetical protein
BAMY6614_12665	2442860	2443810	Hypothetical protein
BAMY6614_12670	2443829	2444491	Hypothetical protein

BAMY6614_12675	2445363	2446178	Hypothetical protein
BAMY6614_12680	2446310	2446420	Hypothetical protein
BAMY6614_12685	2446470	2446679	Hypothetical protein
BAMY6614_12690	2447372	2448049	Polyketide biosynthesis protein
Atypical Region 18			
BAMY6614_12700	2449370	2450344	Acyltransferase
BAMY6614_12705	2450346	2452586	Malonyl CoA-ACP transacylase
BAMY6614_12710	2452652	2452900	Poly(3-hydroxyalkanoate) depolymerase
BAMY6614_12715	2452952	2454214	3-hydroxy-3-methylglutaryl-ACP synthase
BAMY6614_12720	2454211	2454984	Polyketide biosynthesis enoyl-CoA hydratase
BAMY6614_12725	2454994	2455743	Polyketide biosynthesis enoyl-CoA hydratase
BAMY6614_12735	2470729	2484141	Polyketide synthase
BAMY6614_12740	2484159	2494694	Polyketide synthase
BAMY6614_12745	2494684	2510988	Polyketide synthase
BAMY6614_12750	2511002	2518459	Polyketide synthase
Atypical Region 19			
BAMY6614_12895	2540052	2540351	Hypothetical protein
BAMY6614_12905	2540855	2540962	Hypothetical protein
BAMY6614_12910	2541107	2541424	Hypothetical protein
BAMY6614_12920	2542066	2542224	Hypothetical protein
BAMY6614_12925	2542864	2542986	Hypothetical protein
BAMY6614_12930	2542983	2544071	Hypothetical protein
BAMY6614_12935	2544429	2544614	Hypothetical protein
BAMY6614_12940	2544638	2546029	MFS transporter
BAMY6614_12945	2546072	2547673	Beta-xylosidase
Atypical Region 20			
BAMY6614_13000	2560816	2560938	Hypothetical protein
BAMY6614_13005	2560954	2561157	Hypothetical protein
BAMY6614_13010	2561148	2561393	Hypothetical protein
BAMY6614_13015	2561575	2561991	Hypothetical protein
BAMY6614_13025	2563176	2564975	Hypothetical protein
BAMY6614_13030	2564990	2565448	Hypothetical protein
BAMY6614_13035	2565504	2566100	Hypothetical protein
BAMY6614_13040	2566196	2566657	N-acetyltransferase GCN5
BAMY6614_13045	2566871	2567209	Hypothetical protein
BAMY6614_13050	2567830	2568024	Hypothetical protein
BAMY6614_13060	2568982	2569461	Hypothetical protein
BAMY6614_13065	2569507	2569734	Hypothetical protein
BAMY6614_13070	2570020	2570640	Hypothetical protein
Atypical Region 21			
BAMY6614_13080	2572153	2572773	Chitin-binding protein

BAMY6614_13085	2572938	2573036	Membrane protein
BAMY6614_13090	2573515	2573628	Hypothetical protein
BAMY6614_13095	2573792	2574328	Stress response protein *yvgO*
BAMY6614_13100	2574614	2574760	Hypothetical protein
BAMY6614_13105	2574930	2577347	Peptidase G2
BAMY6614_13110	2577548	2577730	Hypothetical protein
Atypical Region 22			
BAMY6614_14135	2827452	2827898	DNA gyrase inhibitory protein *Gyrl*
BAMY6614_14140	2828036	2828182	Hypothetical protein
BAMY6614_14145	2828261	2829040	N-acetyltransferase
BAMY6614_14150	2829079	2830350	*CgeD*
BAMY6614_14155	2830415	2830714	*CgeC*
BAMY6614_14160	2830925	2831356	*CgeA*
BAMY6614_14165	2831363	2832319	*CgeB*
BAMY6614_14170	2832371	2833522	3-phytase
BAMY6614_14175	2833838	2834863	Capsular polysaccharide biosynthesis protein *CapD*
Atypical Region 23			
BAMY6614_15250	3014638	3015384	Polyketide biosynthesis enoyl-CoA hydratase
BAMY6614_15255	3015444	3016691	3-hydroxy-3-methylglutaryl-ACP synthase
BAMY6614_15260	3016749	3017903	Cytochrome P450
BAMY6614_15265	3017985	3024200	*DfnJ*
BAMY6614_15270	3024197	3030349	Polyketide synthase
BAMY6614_15275	3030372	3038090	Polyketide synthase
BAMY6614_15280	3038095	3043911	Polyketide synthase
BAMY6614_15290	3053755	3059481	Polyketide synthase
BAMY6614_15295	3059521	3065817	Polyketide synthase
BAMY6614_15300	3065836	3078426	Polyketide synthase
Atypical Region 24			
BAMY6614_15310	3079218	3080558	AMP-dependent synthetase
BAMY6614_15315	3080579	3080809	Acyl carrier protein
BAMY6614_15320	3080876	3081856	D-fructose-6-phosphate amidotransferase
BAMY6614_15325	3081897	3084155	Malonyl CoA-ACP transacylase
BAMY6614_15330	3084890	3085420	Transcription antiterminator
BAMY6614_15335	3085447	3085692	DNA-binding protein
BAMY6614_15340	3085768	3086646	*LysR* family transcriptional regulator
BAMY6614_15345	3086785	3087108	Membrane protein
Atypical Region 25			
BAMY6614_17045	3397892	3398044	Hypothetical protein
BAMY6614_17050	3398275	3398415	Hypothetical protein
BAMY6614_17055	3398393	3398743	DNA helicase ino80
BAMY6614_17060	3398783	3398968	Hypothetical protein

BAMY6614_17075	3400699	3402132	Type I restriction endonuclease subunit S
BAMY6614_17080	3402129	3403559	Type I restriction endonuclease
BAMY6614_17085	3403673	3406879	Type I restriction endonuclease EcoKI subunit R
BAMY6614_17090	3407123	3407344	Hypothetical protein
BAMY6614_17095	3407505	3407897	Hypothetical protein
Atypical Region 26			
BAMY6614_17785	3539478	3540035	Hypothetical protein
BAMY6614_17790	3540054	3540461	Hypothetical protein
BAMY6614_17795	3541053	3542408	Hypothetical protein
BAMY6614_17800	3542516	3542854	Hypothetical protein
BAMY6614_17805	3542902	3543108	Hypothetical protein
BAMY6614_17815	3543490	3543906	Pilus biosynthesis protein *HicB*
BAMY6614_17820	3543925	3544287	Positive control sigma factor
BAMY6614_17825	3544713	3545213	Hypothetical protein

Table 5. Genes of the developmental programme leading to the formation of biofilm present in the analysed genome sequences of *Bacillus amyloliquefaciens*.

Genes	Coded protein	Functionality	CECT 8237 % ID	CECT 8238 % ID
sfp	Phosphopantetheinyl transferase	Lipopeptides production, surface motility and biofilm	99	98
yczE	Integral membrane protein	Surface motility and biofilm	99	98
swrC	Multidrug efflux pump	Immunity to surfactin; swarming motility	99	99
efp	Similar to elongation factor P	Swarming motility	99	99
swrB	Swarming motility protein	Swarming motility	99	99
swrA	Swarming motility protein	Swarming motility	99	95
spo0A	Master regulator of initiation of sporulation	Sporulation, biofilm	99	99
sigH	Sigma factor H	Initial stage of biofilm	100	99
abrB	Transition state regulator	Transcription control of biofilm formation	100	99
sigW	ECF sigma factor W	Transcription control of biofilm formation	100	100
resE	Sensor histidine kinase	Control of aerobic and anaerobic respiration	99	99
sinR	Master regulator	Transcription control of biofilm formation	99	99
sinI	Antagonist of sinR	Transcription control of biofilm formation	99	98
ydbK	Hypothetical protein	Transcription control of biofilm formation	99	99
ycbA	Sensor histidine kinase	Transcription control of biofilm formation	99	99

lytS	Sensor histidine kinase controlling autolysis	Transcription control of biofilm formation	99	99
ymcA	Conserved hypothetical protein	Transcription control of biofilm formation	99	99
ylbF	Positive regulator of ComK	Required for biofilm formation	100	99
epsA-O	Operon for exopolysaccharide biosynthesis	Assembly of extracellular matrix	99	98
yhxB	Putative phosphomannomutase	Likely involved in exopolysaccharide synthesis	99	99
tapA	Auxiliary protein	Assembly and anchoring of *tasA*-fibers	99	99
sipW	Type I signal peptidase	Involved in processing of *tasA* and *tapA*	99	99
tasA	Amyloid-like protein	Amyloid-like fibers in extracellular matrix	99	99
ecsABC	ABC-type transporter	Control of protein secretion	99	99
yqeK	Putative HD phosphatase	Predicted role in NAD metabolism	99 / 98[b]	99
bslA	hydrophobic protein	Biofilm assembly factor	99	99
pgsBCA	Operon for poly-gamma-glutamic acid biosynthesis	Adherence of charged molecules during biofilm formation	99	99
sacB	Levansucrase	Synthesis of levan	99	99
xynA	Endo-1,4-beta-xylanase	Degradation of xylan, C source	99	99
xynC	Glucuronoxylanase xynC	Degradation of xylan, C source	99	99
pel	Pectate lyase	Degradation of pectin, C source	98	98
aprE	Subtilisin E	Degradation of proteins, N source	99	99

nprE	Neutral protease	Degradation of proteins, N source	99 / 96[b]	99 / 96[b]
bpf	Bacillopeptidase F	Degradation of proteins, minor protease, N source	99	99
mpr	Extracellular metalloprotease	Degradation of proteins, minor protease, N source	99	99
epr	Extracellular protease	Degradation of proteins, minor protease, N source	99	98
vpr	Extracellular protease	Degradation of proteins, minor protease, N source	99	99
rapA	Response regulator aspartate phosphatase A	Negative regulation of Spo0F	100 / 99[b]	99 / 99[b]
phrA	Phosphatase *rapA* inhibitor	Secreted regulator of the activity of phosphatase RapA	98	94 / 75[b]
rapB	Response regulator aspartate phosphatase B	Negative regulation of Spo0F	99	99
rapC	Response regulator aspartate phosphatase C	Negative regulation of ComA	99	99
phrC	Phosphatase *rapC* inhibitor	Secreted regulator of the activity of phosphatase RapC	97 / 94[b]	97 / 94[b]
rapD	Response regulator aspartate phosphatase D	Regulation of ComA regulon	99	98
rapF	Response regulator aspartate phosphatase F	Negative regulation of ComA	99	98
phrF	Phosphatase *rapF* inhibitor	Secreted regulator of the activity of phosphatase RapF	97	95
rapH1	Response regulator aspartate phosphatase H	Negative regulation of ComA	53[a]	-
rapH2	Response regulator aspartate phosphatase H	Negative regulation of ComA	99	98

rapI	Response regulator aspartate phosphatase I	Activates ICEBs1 gene expression, excision, and transfer	-	48 / 97[a,b]
phrI	Phosphatase *rapI* inhibitor	Secreted regulator of the activity of phosphatase RapI	-	86 / 69[a,b]
rapJ	Response regulator aspartate phosphatase J	Control of expression of genes regulated by Spo0A	99	99
rapK	Response regulator aspartate phosphatase K	Inhibition of ComA activity	93[a]	49[a]
phrK	Phosphatase *rapK* inhibitor	Secreted regulator of the activity of phosphatase RapK	95[a]	41 / 77[a]
rapX	Response regulator aspartate phosphatase X	Uncharacterized role in regulation pathways	98	98
comQ	Isoprenyl transferase	Posttranslational modification of the pre-pheromone ComX	39 / 88[b]	36 / 86[b]
comX	Competence pheromone	Quorum-sensing pheromone for development of genetic competence	33 / 92[b]	41 / 50[b]
comP	Sensor histidine kinase	Regulator of surfactin production	60 / 95[b]	62 / 97[b]
comA	Two-component response regulator	Positive regulation of late competence genes and surfactin production	100	100

All of the genes have been compared to the reference strain *B. amyloliquefaciens* subsp. *plantarum* FZB42.
a. Genes absent in FZB42. The comparison was performed with the type strain *B. subtilis* subsp. *subtilis* 168.
b. When the percentage of the coverage was below 100 %, the percentage of identity (first number) and the percentage of coverage (second number) are indicated.