










Caracterización del complejo genético tracrARN-_{DR}ARN asociado al sistema CRISPR-Cas9 del fitosimbionte *Acholeplasma palmae* descriptor: interés biotecnológico

Characterization of the tracrARN-_{DR}ARN genetic complex associated with the CRISPR-Cas9 system of the phytosymbiont *Acholeplasma palmae*: biotechnological interest

Caracterização do complexo genético tracrARN-_{DR}ARN associado ao sistema CRISPR-Cas9 do descritor de *Acholeplasma palmae* fitossimbionte: interesse biotecnológico

Luis Moncayo¹, Alex Castro², Diego Arcos², Paulo Centanaro², Diego Vaca³, Cristina Maldonado⁴, Aleivi Perez⁶, Carla Lossada⁷, Lenin González-Paz^{5,8*}

¹Universidad Católica de Cuenca, Ecuador. Correo electrónico: luismoncayo834@hotmail.com, . ²Universidad Agraria de Ecuador, Facultad de Ciencias Agrarias, Ecuador. Correo electrónico: (AC): alcastro2010@hotmail.es, ; (DA): arcosda@hotmail.com,  (PC): pcentanaro@uagraria.edu.ec, . ³Universidad Laica Eloy Alfaro de Manabí, Ecuador. Correo electrónico: diegoal13@gmail.com; . ⁴Laboratorio de Biotecnología, Facultad de Ciencias Agropecuarias, Universidad Técnica de Babahoyo, Ecuador. Correo electrónico: cmaldonadoc@yahoo.com . ⁵Laboratorio de Genética y Biología Molecular, FEC-LUZ. Zulia, Maracaibo, Venezuela. Correo electrónico: [lgonzalezpaz@gmail.com](mailto:lgonzalezpaz@lgonzalezpaz@gmail.com), . ⁶Laboratorio de Microbiología General, FEC-LUZ. Zulia, Maracaibo, Venezuela. Correo electrónico: aleviciencias@gmail.com, . ⁷Laboratorio de Caracterización Molecular y Biomolecular, Centro de Investigación y Tecnología de Materiales, Instituto Venezolano de Investigaciones Científicas (IVIC), Zulia, Venezuela. 526. Correo electrónico: lossadacarla@gmail.com, . ⁸Doctorado en Ciencias Agrarias, Facultad de Agronomía, Universidad del Zulia, Zulia, Venezuela.

Recibido: 14-03-2021 • Aceptado: 28-06-2021.

*Autor de correspondencia. Correo electrónico: lgonzalezpaz@gmail.com

Resumen

La tecnología CRISPR-Cas9 usada en biotecnología vegetal se fundamenta en el empleo de nucleasas Cas9 para generar cortes precisos en el genoma y un dúplex conformado por un trans-activador CRISPR ARN (tracrARN) y un CRISPR ARN (_{DR}ARN) los cuales son precursores del ARN guía (sgARN) rediseñado comercialmente (sgARN-Cas9) para guiar la escisión de genes. La mayoría de estas herramientas provienen de bacterias clínicas. Sin embargo, existen diversos sistemas CRISPR-Cas9 en microorganismos ambientales como los fitoendosimbiontes del género *Acholeplasma*. Pero la explotación de estos sistemas más compatibles con plantas requiere usar herramientas bioinformáticas para su predicción y estudio. Fueron identificados y caracterizados los elementos asociados al dúplex tracrARN-_{DR}ARN en *A. palmae*. Para ello la información proteica se obtuvo del Protein Data Bank y la genómica del GenBank/NCBI. El sistema CRISPR fue estudiado con el software CRISPRfinder. Se utilizaron algoritmos de alineamiento y el software NUPACK para identificar los módulos tracrARN y _{DR}ARN, y también diversos softwares computacionales para la caracterización genética, estructural y biofísica. Se encontró un sistema CRISPR-Cas en *A. palmae* con características de tipo II-C, así como un dúplex con regiones flexibles, termodinámicamente muy estable, presentando una energía de acoplamiento con Cas9 termodinámicamente favorecida. Estos resultados son deseables en sistemas de edición programable de genes y evidencian la posibilidad de explorar herramientas moleculares nativas en microorganismos ambientales aplicables a la manipulación genética de plantas, conforme se realicen más investigaciones. Este estudio representa el primer reporte sobre la estabilidad termodinámica y el acoplamiento molecular de elementos asociados al dúplex tracrARN-_{DR}ARN en el fitosimbionte *A. palmae*.

Palabras clave: acoplamiento molecular, bioinformática, mycoplasma, transactivador.

Abstract

The CRISPR-Cas9 technology used in plant biotechnology is based on the use of Cas9 endonucleases to generate precise cuts in the genome, and a duplex consisting of a trans-activating CRISPR RNA (tracrRNA) and a CRISPR RNA (_{DR}RNA) which are precursors of guide RNA (sgRNA) commercially redesigned (sgRNA-Cas9) to guide gene cleavage. Most of these tools come from clinical bacteria. However, there are several CRISPR-Cas9 systems in environmental microorganisms such as phytoendosymbionts of plants of the genus *Acholeplasma*. But the exploitation of these systems more compatible with plants requires using bioinformatics tools for prediction and study. We identified and characterized the

elements associated with the duplex in the genome of *A. palmae*. For this, the protein information was obtained from the Protein Data Bank and the genomics from GenBank/NCBI. The CRISPR system was studied with the CRISPRfinder software. Alignment algorithms and NUPACK software were used to identify the tracrRNA and $_{DR}$ RNA modules, together with various computational software for genetic, structural and biophysical characterization. A CRISPR-Cas system was found in *A. palmae* with type II-C characteristics, as well as a thermodynamically very stable duplex, with flexible regions, exhibiting a docking power with Cas9 thermodynamically favored. These results are desirable in programmable gene editing systems and show the possibility of exploring native molecular tools in environmental microorganisms applicable to the genetic manipulation of plants, as more research is carried out. This study represents the first report on the thermodynamic stability and molecular docking of elements associated with the tracrRNA- $_{DR}$ RNA duplex in the phytosymbiont *A. palmae*.

Keywords: molecular docking, bioinformatics, mycoplasma, transactivator.

Resumo

A tecnologia CRISPR-Cas9 usada em biotecnologia vegetal é baseada no uso de nucleases Cas9 para gerar cortes precisos no genoma e um duplex composto de um RNA CRISPR transativador (tracrRNA) e um RNA CRISPR ($_{DR}$ ARN) que são precursores de o RNA guia comercialmente reprojetoado (sgRNA) (sgRNA-Cas9) para guiar a clivagem do gene. A maioria dessas ferramentas vem de bactérias clínicas. No entanto, existem vários sistemas CRISPR-Cas9 em microrganismos ambientais, como fitoendossimbiontes do gênero *Acholeplasma*. Mas a exploração desses sistemas mais compatíveis com as plantas requer o uso de ferramentas de bioinformática para sua previsão e estudo. Os elementos associados ao duplex tracrARN- $_{DR}$ ARN em *A. palmae* foram identificados e caracterizados. Para isso, as informações sobre proteínas foram obtidas no Protein Data Bank e a genômica no GenBank / NCBI. O sistema CRISPR foi estudado com o software CRISPRfinder. Algoritmos de alinhamento e o software NUPACK foram utilizados para identificar os módulos tracrRNA e $_{DR}$ ARN, além de diversos softwares computacionais para caracterização genética, estrutural e biofísica. Um sistema CRISPR-Cas foi encontrado em *A. palmae* com características do tipo II-C, bem como um duplex com regiões flexíveis, termodinamicamente muito estável, apresentando uma energia de acoplamento termodinamicamente favorecida com Cas9. Esses resultados são desejáveis em sistemas de edição de genes programáveis e mostram a possibilidade de explorar ferramentas moleculares nativas em microrganismos ambientais aplicáveis à manipulação genética de plantas, à medida que mais pesquisas são realizadas. Este estudo representa o primeiro relatório sobre a estabilidade termodinâmica e acoplamento molecular de elementos associados ao duplex tracrRNA- $_{DR}$ ARN no fitossimbionte *A. palmae*.

Palavras chave: acoplamento molecular, bioinformática, micoplasma, transativador

Introducción

Acholeplasma es un género de bacterias fitosimbiontes sin pared celular, clasificadas en la división Firmicutes, clase Mollicutes, orden Acholeplasmatales y familia Acholeplasmataceae. Hay 15 especies reconocidas en este género. Los genomas de las especies varían de 0,5 a 2,2 Mbp con un contenido de G+C es bajo, oscilando entre 20% y 40%. La temperatura óptima de crecimiento de estos organismos es de 30° a 37° C, son saprófitas y se encuentran en el suelo, el compost, las aguas residuales o como comensales de vertebrados, insectos o plantas, siendo consideradas como endosimbiontes de plantas (Ipoutcha *et al.*, 2019).

Hasta ahora no hay información sobre si alguna de las especies de *Acholeplasma* sea patógeno primario de plantas, pero se sabe que pueden causar efectos citopáticos en cultivos de tejidos (Windsor *et al.*, 2010; Kube *et al.*, 2014). Estas bacterias pueden aislarse de la superficie de plantas como el brócoli (*Brassica oleracea* var. *italica* descriptor) o de tejidos de la corona de la palma de coco (*Cocos nucifera* descriptor) (Tully *et al.*, 1994).

A. aidlawii ha sido la especie más estudiada y la única del género aprovechada en estudios biotecnológicos (Windsor *et al.*, 2010; Wu *et al.*, 2018; Mirabelli *et al.*, 2019). Sin embargo, existen otras especies del género *Acholeplasma* cuyos genomas han sido completamente secuenciados como *A. palmae* J233 (cuyo genoma es de 1.554.229 bp), y esto ha permitido el

Introduction

Plant's health and productivity have been linked to their microbiota dominated mainly by bacteria and fungi (Purahong *et al.*, 2018). Interactions between plants and microorganism have been present for a long time in evolution, since microorganisms have helped plants to deal with lack of nutrients, stressful conditions, and pathogens (Smith *et al.*, 2015). The study of plant-associated bacteria, their antagonistic potential in control of pathogens and their effect as growth promoters comprise the action of several genera and species of bacteria that can be isolated from the surface of plant tissues or obtained from within the plant (Daungfu *et al.*, 2019). In Mexico, and many other parts of the world, basil (*Ocimum basilicum* L., Fam. Lamiaceae) is an economically important aromatic herb. In Baja California Sur, Mexico, production of sweet basil was approximately 1,685 tons at the end of 2019 cycle (SIAP 2019). Nuffar variety was the main sweet basil grown in the state. This plant is used in the food industry and in traditional medicine (Gilardi *et al.*, 2013). Basil plants shelter a distinctive microbiota responsible for growth, quality and health via microbial metabolism and host interactions (Adamović *et al.*, 2015). However, successive cultivation in the same soil for a long time is known to be associated with changes in physicochemical properties of the ground, which in turn, provides conditions that cause diseases by pathogens, leading to

estudio comparativo entre las especies del género y las especies relacionadas (Tully *et al.*, 1994; Kube *et al.*, 2014). Es importante destacar que los análisis comparativos han revelado que *A. palmae* J233 es la única especie descrita de estos fitoendosimbiontes con un sistema genético tipo CRISPR-Cas (*Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated systems or cas genes*) confirmado, pero aun no caracterizado (Bumgardner *et al.*, 2015). Estos sistemas genéticos son importantes porque pueden estar asociados a factores de patogenicidad y virulencia en microorganismos (Westra *et al.*, 2014). También son precursores de herramientas de manipulación genética usadas en biotecnología vegetal como estrategia hacia una agricultura sostenible (Belhaj *et al.*, 2015).

Específicamente, los sistemas CRISPR son un nuevo grupo de herramientas empleadas en biotecnología vegetal basada en el uso de endonucleasas de restricción de la familia *caspasas*, como la Cas9, que provienen del sistema de inmunidad natural procarionota CRISPR tipo II, y que incluye la nucleasa, la secuencia trans-activadora CRISPR RNA (*trans-activating CRISPR RNA* o *tracrRNA*), la secuencia CRISPR ARN (*CRISPR RNA* o *crRNA*) y el motivo adyacente al protoespaciador (*protospacer adjacent motifs* o *PAM*). El crARN, tracrARN y PAM conforman los elementos genéticos del complejo tracrARN-crARN, el cual es precursor del ARN de guía única (sgARN) (*simple guide RNAs*

devastating yield losses. In this study, the objective was to isolate and identify bacteria associated to healthy basil plants from crops around La Paz, Baja California Sur, Mexico, with the ability to inhibit pathogenic fungi development without affecting seedling development.

Material and methods

Plant material and experimental site

Samples of basil plants were taken from crops with similar agroecological conditions within the area of La Paz, B.C.S. México. Three adult basil plants, ready for harvest, apparently healthy, with intact aerial parts (normal leaves without spots or color alterations), stems without deformations, separated at least 50 meters one from each other, were randomly selected per crop. The crop locations were: Pescadero (Pradera Escolar 23°21'56"N, 110°10'27"W and Las Maravillas 23°21'43"N, 110°11'18"W), Todos Santos (Sueño Tropical 23°31'29"N, 110°17'15"W) and Los Planes (Cesar Cárdenas, 23°57'11"N, 109°55'50"W). Root, stem and leaves of each plant were separated and placed in sterile plastic bags. Later, the samples were transported to the Laboratory where the microbiological analysis was carried out.

Isolation of bacteria

To isolate predominant bacteria, two culture media were used. Standard methods agar or plate count agar (SMA, DIFCO) was used to isolate bacteria and de Man Rogosa

o *sgRNA*), una molécula rediseñada comercialmente para establecer un complejo ribonucleoproteico (RNP) con el *sgARN* y la nucleasa (formado en complejo *sgARN-Cas9*) y dirigir de forma específica la escisión (o corte) programable de genes y genomas (Chyou y Brown, 2019). Por lo tanto, caracterizar los elementos *tracrARN* y *crARN* tienen un alto interés biotecnológico (Kaushik *et al.*, 2019). En este sentido esta investigación se centró en dicha caracterización. Especialmente porque se sabe que el uso de sistemas compatibles con el objeto de estudio como las “agrobacterias” para plantas, puede otorgar mejoras en los procesos de edición genética y fitomejoramiento de las células vegetales receptoras (Demirer *et al.*, 2019; Danilo *et al.*, 2019; Zhang *et al.*, 2020; Rodríguez *et al.*, 2021). El interés también radica en que existe una gran diversidad de herramientas para la manipulación genética de plantas, pero todas provienen de sistemas CRISPR tipo-II de bacterias de interés clínico como *Staphylococcus aureus* (SaCas9), *Streptococcus thermophilus* (StCas9), *Neisseria meningitidis* (NmCas9) y *Streptococcus pyogenes* (SpCas9) (Zhang *et al.*, 2019).

Sin embargo, en la naturaleza existe una gran diversidad de sistemas CRISPR-Cas tipo II y de elementos claves para la edición genética como los complejos *tracrARN-crARN* en microorganismos ambientales menos estudiados como *A. palmae*, la única especie del género *Acholeplasma* con un sistema CRISPR conocido (Bumgardner *et al.*, 2015). Pero la

and Sharp agar (MRS, DIFCO) was used to isolate lactic acid bacteria. Briefly, 10 g of root, stem or leaves were placed into 90 mL of phosphate buffer saline (PBS) (145 mM NaCl, 2.87 mM KH_2PO_4 and 6.95 mM K_2HPO_4 , pH 7.2) and triturated with a tissue homogenizer (Ultra-Turrax T25, Janke & Kunkel) until a suspension was obtained. Serial dilutions were prepared and then 100 μL of each dilution was spread on duplicated Petri dishes containing: SMA and MRS, the plates were incubated at 35 °C for 24 h. Bacteria from predominant and morphologically different colonies were cross streaked and purified onto plates containing the same media as the plate they grew first. Bacteria were characterized by Gram staining and cell morphology under a phase contrast microscope (Nikon Eclipse E-600), then inoculated in Tryptic Soy Broth (TSB, DIFCO) and incubated overnight at the above temperature and preserved with 50 % glycerol at -80 °C (Goldman and Green 2009). The isolated bacteria were labeled as follows: first two letters correspond to the investigated plant, in this case AL for basil; third letter corresponds to culture media they were isolated from, A for standard methods agar and M for de Man Rogosa and Sharpe medium. Fourth letter corresponds to the part of the plant sampled: R for root, T for stem and H for leaves. First number corresponds to the sampling area: 1-3 corresponds to Pescadero, 4-6 corresponds to Los Planes and 7-9 corresponds to Todos Santos. Finally, the last number corresponds to the isolated colony.

explotación de estos sistemas capaces de establecer interacciones con plantas requiere usar herramientas bioinformáticas para su predicción y estudio.

Para evitar confusiones el término tracrARN-crARN el cual se usa para referirse típicamente al complejo entre las secuencias llamadas repeticiones directas (*Direct Repeat o DR*) tracrARN y crARN, fue sustituido en este trabajo por tracrARN-_{DR}ARN, debido a que en un sentido estricto el crARN maduro se compone de la DR y de un elemento de ARN denominado “*espaciador*”, el cual no fue considerado en los análisis realizados, por ser el elemento genético altamente variable del dúplex tracrARN-crARN (Chyou y Brown, 2019). Por lo que se realizó la caracterización de los elementos genéticos que conforman el dúplex designado como tracrARN-_{DR}ARN asociados al sistema CRISPR-Cas tipo II en el genoma de *A. palmae* con el interés de ofrecer bases teóricas alternativas que contribuyan con los procesos de manipulación genética aplicables a la biotecnología vegetal.

Materiales y métodos

Secuencia genómica, identificación y caracterización del sistema CRISPR-Cas

La información genómica de la especie *A. palmae* J233 (*GenBank accession* GCA_000968055.1; *RefSeq accession* GCF_000968055.1; *RefSeq sequence* NC_022538.1) fue obtenida del NCBI (<http://www.ncbi.nlm.nih.gov/>). El CRISPR loci así como los genes Cas asociados se identificaron

Model Fungi used in antagonistic assay

The fungi used in this study were *Aspergillus* spp., *Alternaria* spp. and *Hendersonula* spp. currently identified as *Neoscytalidium dimidiatum*. *Aspergillus* spp. comes from a monosporic culture isolated from soil and characterized by its colonial (macroscopic) and cellular (microscopic) morphology according to general taxonomic keys in the Laboratorio de Ciencia y Tecnología de Alimentos at this University. This fungus was used in this research as a model to carry out the first screening of bacteria with antagonistic capacity. *Alternaria* spp. and *N. dimidiatum* also come from a monosporic culture and were previously isolated from *Ocimum basilicum* L. and *Ficus benjamina* respectively (Hernández-Montiel *et al.*, 2018). These fungi were used as a model for important plant pathogens in this region.

Qualitative antifungal assay

The antifungal activity was detected using a methodology developed by our research group. The confrontation assays against *Aspergillus* spp. were carried out in a method previously described (Joo *et al.*, 2015), modified with dual culture plates containing a mixture of TSA and PDA at 1:1 ratio. The isolated and purified bacteria were inoculated at 1 % in TSB and incubated overnight at 30 °C, then 10 µL drop of the suspension was placed on each Petri dish at 15 mm away from the edge. The dishes were incubated for 24 h at 30 °C. Subsequently, a 4 mm diameter filter paper disk, moistened with PBS

y caracterizaron con el servidor web CRISPRCasFinder (Couvin *et al.*, 2018) (<https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index>). Todos los espaciadores fueron comparados con la base de datos GenBank/NCBI (Benson *et al.*, 2018) para la búsqueda de secuencias homólogas con emparejamiento $\geq 95\%$ usando BLAST (Shen *et al.*, 2017). Adicionalmente se predijo el número de posibles estructuras secundarias a partir de las secuencias de repeticiones directas (DR) y del tracrARN previas a la energía libre mínima de Gibbs para la formación (ΔG) usando barriers web server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/barriers.cgi>).

Predicción del tracrARN y crARN

Para buscar la porción anti-repetición del tracrARN, se realizó una alineación local de nucleótidos usando BLAST entre la secuencia DR consenso y las secuencias de ADN dentro de una de las siguientes cuatro regiones no codificantes: dentro de 500 nt aguas arriba de cas9 y entre cas9 y cas1, y en la matriz de espaciadores más DR, o dentro de 500 nt aguas abajo de la matriz espaciadores más DR. Se utilizaron los siguientes parámetros para el BLAST: algoritmo, secuencias algo similares (blastn); tamaño de palabra, 7; puntajes de coincidencia/desajuste, 1, -2; costos de brecha, existencia 1 y extensión 2. A partir de los resultados de la alineación, se identificaron las secuencias del “potencial de tracrARN”. Una vez localizada la región anti-repetición correspondiente al tracrARN, se

was impregnated with spores of an *Aspergillus* spp. colony grown for 72 h at 30 °C in PDA medium, and then placed 15 mm away from the bacteria colony. The dual cultures, performed in duplicate, were incubated for 72 h at 30 °C and then inhibition zones were observed. A Petri dish inoculated only with the disc containing the fungus was used as positive control.

Confirmatory antifungal assay

The growth of bacteria that clearly showed an inhibition zone of *Aspergillus* spp. were tested against *Aspergillus* spp., *N. dimidiatum* and *Alternaria* spp. using methodology developed by our research group. Bacterial strains reactivated in TSB were inoculated at 1 % in tubes with the same medium and incubated for 24 h at 30 °C. Subsequently, they were centrifuged at 3,500 rpm, the supernatant was decanted, and the pellets were resuspended in the same volume of PBS. Parallely, the fungi were cultivated in PDA for 72 h, its spores were collected, and its concentration was adjusted to 10^6 spores mL^{-1} by direct count, under the microscope using a Neubauer plate. Then, 0.1 mL of the suspensions were spread onto TSA-PDA 1:1 Petri dishes. Later three 10 μL drops of each bacterial suspension were evenly distributed on the plate previously spread with a fungus. The plates were then incubated at 30 °C for 72 h. After the incubation time, the zone of inhibition was measured in millimeters. Two replicates were made per treatment and the experiment was repeated twice. Inhibition zones were measured and bacteria with

extendió la búsqueda en la dirección 3' hasta que se encontró una secuencia con semejanza a un terminador de transcripción independiente de Rho. Si no hubiese un terminador transcripcional claro, se extendería la búsqueda de tracrARN en la dirección 5', buscando una estructura similar flanqueada por una cadena rica en A. Los promotores transcripcionales y los terminadores independientes de Rho también se predijo utilizando el programa BDGP Neural Network Promoter Prediction (www.fruitfly.org/seq_tools/promoter.html) y el software TransTermHP (<http://transterm.ccb.jhu.edu/query.php>), respectivamente (Chylinski *et al.*, 2013). Si el tracrARN está codificado en la dirección de 3' a 5', el complemento inverso de la secuencia se usó durante los análisis. Empleando el software de predicción de plegamiento de RNA NUPACK (<http://www.nupack.org/>) se predijo las estructuras secundarias del dúplex tracrARN y ^{DR}ARN, así como la ΔG . Se utilizó un algoritmo de plegado que permite oscilaciones de base G-U. Para NUPACK, las secuencias de tracrARN y ^{DR}ARN se introdujeron en la dirección 5' a 3'. Usando las siguientes opciones: tipo de ácido nucleico, ARN; número de especies de hebras, 2; tamaño complejo máximo, 2; concentración de cadena1, 1 μM y concentración de cadena2, 1 μM . Se identificaron los elementos asociados a las estructuras secundarias (tallo superior, bulto, tallo inferior, nexo y horquillas) (Chylinski *et al.*, 2013; Briner *et al.*, 2016; Chyou y Brown, 2019).

large inhibition zones (more than 10 mm) were selected for molecular identification.

Molecular identification

Bacteria were identified by 16S rDNA gene sequencing. Total DNA was extracted following manufacturers recommendation (Wizard® Genomic DNA Purification Kit, Promega Corporation). PCR was carried out using universal primers pA (forward): 5'-AGAGTTTGATCCTGGCTCAG-3' and pH (reverse): 5'-AAGGAGGTGATCCAGCCGCA-3' following Broda *et al.* (1999) method. The amplicons were purified and then sequenced. The obtained data was processed using BLAST searches in the NCBI and uploaded to GenBank database.

Germination test

Microtubes were filled with one hundred twenty basil seeds. The seeds were disinfected with a 4.25 % aqueous sodium hypochlorite solution for 2 minutes, rinsed with sterile distilled water three times. Then seeds were inoculated with the suspension of bacteria that presented antifungal activity against *Aspergillus* spp. containing 10^6 CFU.mL⁻¹ and let stand for two minutes. The supernatant was decanted, and the seeds were set in a completely random design which consisted of four replicates of 30 seeds per treatment and controls. The seeds were placed in Petri dishes with a filter paper covering the bottom of the plate. Water was added to moisten the seeds and then let stand for 5 days in the dark in a germination chamber at 25 °C. Testa rupture was considered as indication of germination. Five

Análisis complementarios

Se determinó la longitud de las regiones del dúplex, así como el porcentaje GC, la temperatura de fusión (T_m), masa molecular y regiones de máxima flexibilidad usando el software Unipro UGENE-v.1.32.0. Se calculó el tamaño (en términos del radio) del dúplex tracrARN-_{DR}ARN con el método del cálculo del área de superficie accesible al solvente (SASA) usando el algoritmo Shrake-Rupley (Dupuis *et al.*, 2014). Se predijo la estabilidad termodinámica del dúplex en términos de ΔG usando NUPACK, así como las estructuras secundarias a diversos intervalos de temperatura (12°, 15°, 23°, 30°, 37°, 40° y 42° C) (Høyland-Krogsho *et al.*, 2018). En todos los casos fue estimado el tamaño de las estructuras generadas. Adicionalmente, se calculó la energía de interacción entre las regiones del dúplex y la nucleasa asociada mediante acoplamiento molecular. Para ello la información proteica se obtuvo del Protein Data Bank (<https://www.rcsb.org/>). La estructura tridimensional de la nucleasa fue modelada con SWISS-MODEL (<https://swissmodel.expasy.org/>). El complejo tracrARN-_{DR}ARN-Cas9 se preparó con el software HDOCK (<http://hdock.phys.hust.edu.cn/>) y fue posteriormente analizado usando Molegro Molecular Virtual (MMV)-v.7.0.0. Se calcularon las funciones de puntuación HDOCK (Yan *et al.*, 2017) y MolDock Score (Thomsen y Christensen, 2006). Los acoplamientos moleculares se llevaron a cabo entre la Cas9 y cada una de las regiones que componen al dúplex tracrARN-_{DR}ARN porque se sabe que

days old seedlings were separated out by hypocotyl and radicle, length of all seedlings was measured by the software RootNav (Pound *et al.*, 2013), then dried and weighted to obtain dry biomass. The significance of the obtained results was analyzed by one-way univariate analysis of variance (ANOVA), seed inoculation with different bacteria was used as a fixed factor. The difference between the means was determined by Tukey's test HSD multiple range test at $p=0.05$. Mean values were considered significantly different when $p \leq 0.05$. Statistical analysis was carried out using SAS software (Statistical Analysis System (SAS) software, SAS University Edition).

Results and discussion

The microbiological status of healthy basil plants should be determined by the growing conditions, the environment of the region, the method of cultivation and the associated beneficial microbiota at any stage of their production process. In this study a total of 165 isolates were obtained from the predominant microbiota of different parts of the sampled apparently healthy basil plants from different sites. Ninety-four isolates were from SMA and 71 from MRS media used in the microbiological analysis, all of them were Gram positive, 152 with bacillary and 13 with coccoid morphology. Among the most common plant colonizing bacteria, we can find those of the *Bacillus* genus (Moffat, 2001). It is known

en la naturaleza la Cas9 interacciona tanto con el complejo (tracrARN-DR-ARN) como con cada uno de sus elementos de manera independiente (Jinek *et al.*, 2012).

Resultados y discusión

Organización del sistema CRISPR-Cas en *A. palmae*

El genoma de *A. palmae* J233 es de 1 554 229 bp. En él se encontró un sistema CRISPR completo compuesto por una matriz de DR más espaciadores entre la región 311 136 y 314 270 (Figura 1), con 48 DR de una longitud de 36 pb y una secuencia consenso de las DR conservada dentro de la matriz en un CGTAATTTTTGCTATCTAACAAC-3'. El sistema CRISPR encontrado es del tipo II-C al estar conformado por los genes *cas9*, *cas1* y *cas2* ubicados entre la región de 305 599 y 310 195 pb, y orientados en dirección aguas arribas. Es este sentido, la matriz CRISPR exhibe las características típicas de los arreglos tipo II descritos para estos sistemas de inmunidad procariotas. Específicamente presenta la arquitectura que agrupa organismos de importancia clínico-ambiental como *Neisseria meningitidis* y *Campylobacter jejuni* (Chylinski *et al.*, 2013; Chyou y Brown, 2019). Estos presentan sistemas CRISPR que se han asociado con funciones no convencionales como factores de virulencia, regulación genética, entre otras, incluyendo la patogenicidad hacia plantas (Westra *et al.*, 2014).

La matriz presentó 47 secuencias espaciadoras muy diversas, que fueron

that induced systemic resistance is triggered by beneficial members of the root microbiome in a wide range of plant hosts making them resistant against various pathogenic threats. This systemic enhancement has also been described for many plant growth-promoting bacteria of the *Bacillus* genus (Pieterse *et al.*, 2014). It is well known that bacteria of the *Bacillus* genus have the capability of producing a series of bioactive compounds with antimicrobial activity (Raaijmakers, 2010).

The qualitative confrontation assay of the 165 isolates against *Aspergillus* spp. resulted in 15 bacilli strains that inhibited the growth of the fungus at different degrees (figure 1). Antagonistic strains were isolated from aerial part of plants collected in the towns of Los Planes and Todos Santos. It is known that conditions on the plant leaves and stem (aerial part) are harsh for bacteria resulting in local microsites at which conditions are favorable for growth and/or survival (Jacobs *et al.*, 2005).

The confirmative confrontation assay against *Aspergillus* spp., *N. dimidiatum* and *Alternaria* spp., performed with the fifteen bacterial strains that showed antagonistic activity, confirmed the growth inhibition of fungi by some of them. Bacterial strains presented non regular inhibition zones on fungi growth; therefore, length and width measurements of the inhibition area were documented (table 1). This behavior was observed because some bacteria are highly mobile.

comparadas frente a una base de datos de 4200 secuencias plasmídicas y 2718 genomas completos de bacteriófagos según la información contenida en el GenBank/NCBI database. Se encontró que el 91,5% (43/47) de los espaciadores presentaron homología con secuencias asociadas a bacteriófagos, mientras que el 8,5% (4/47) presentó homología con plásmidos. El elevado número de espaciadores dirigidos a bacteriófagos, encontrados dentro de la matriz CRISPR, en más del 90% en *A. palmarum*, imposibilita recomendar estrategias de control biológico para evitar los efectos citopáticos en cultivos de tejidos basadas en el uso de virus líticos infectivos (Fagoterapia) (São, 2018).

La predicción de la estructura secundaria de la DR consenso mostró una estructura termodinámicamente estable con un ΔG de $-2,90 \text{ kcal. mol}^{-1}$. Además se predijo un total de otras 6439 posibles estructuras en un rango de energía de $40 \text{ kcal. mol}^{-1}$ por encima de la ΔG . Curiosamente, el ΔG calculado a partir de las DR de la matriz CRISPR de *A. palmarum* es termodinámicamente menos favorable que la gran mayoría de las DR consenso descritas al presentar un ΔG con signo negativo de menor magnitud (Negahdaripour *et al.*, 2017; Shen, 2017; Zhao *et al.*, 2018; Baliga, 2018), pero sin dejar de ser una secuencia que media la transcripción de una estructura de ARN secundaria estable y de formación espontánea. Sin embargo, si bien los sistemas CRISPR tipo II han sido los más estudiados y aprovechados a nivel biotecnológico

Results of the molecular identification of bacteria with confirmed antifungal activity, showed that all of them belong to the *Bacillus* genus according to BLAST similarity analysis (table 2). *Bacillus* isolates like *B. subtilis*, *B. methylotrophicus* and *B. amyloliquefaciens* have been shown to have antagonistic activity against common plant pathogenic fungi. These bacteria species have been considered efficient colonizers that are widely spread due to their capacity to form spores, to grow in a wide range of temperatures and to produce antibiotic compounds that inhibit phytopathogens growth, besides being plant growth promoters (Rios-Velasco *et al.*, 2016).

In this investigation eight strains with antifungal activity against *Aspergillus* spp., *N. dimidiatum* and/or *Alternaria* spp., found in the samples obtained from basil plants, were identified and six of them belong to *B. amyloliquefaciens*. This is a microorganism that grows mainly in the rhizosphere, it begins to grow close to the formation point of lateral roots and it spreads on the root surface (Lugtenberg and Kamilova, 2005). Nowadays, it has been demonstrated that inoculation of crops with *B. amyloliquefaciens* strains could increase production under stressful conditions like nutrient limitations or high salinity levels (Kim *et al.*, 2017). It also has been observed that it could help to control other pathogens either by competing for nutrients or by producing antibiotics and lytic enzymes; particularly, the molecular basis of its plant growth-

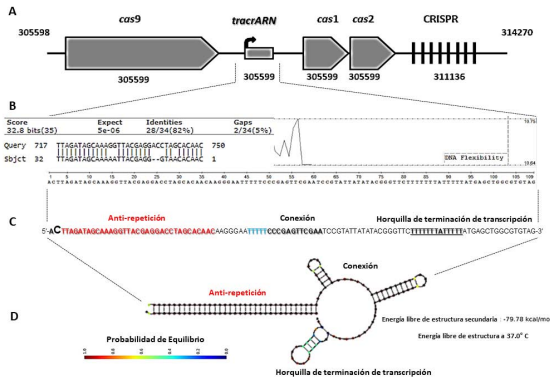


Figura 1. Descripción de los elementos del sistema CRISPR-Cas9 Tipo II-C y del *tracrARN*-*DR* *ARN* asociado en *Acholeplasma palmae*. (A) Representación esquemática de los genes *cas* adyacentes a la matriz CRISPR. (B) Identificación de la porción anti-repetición del *tracrARN* que forma el tallo superior. La secuencia “query” (consulta, en español) es la región genómica aguas arriba entre el gen *cas9* y *cas1*. La secuencia “sbjct” (Sujeto) es la secuencia de repetición CRISPR consenso. La anti-repetición detectada por BLAST tenía una identidad del 82%. Se muestra además la región de máxima flexibilidad de la secuencia entre el nt 50 y 60 que coincide con el segmento “Conexión”. (C) Secuencia de bases nitrogenadas del *tracrARN* en el locus CRISPR. Anti-repetición (rojo) del dúplex *tracrARN*-*DR* *ARN*, nucleótidos no apareados que forman la protuberancia (negrita), módulo del tallo inferior (azul). Adyacente al tallo inferior se encuentra el módulo nexa que forma una horquilla en estructura secundaria de *ARN*. La secuencia “aC” previa al segmento anti-repetición corresponde con el promotor para la secuencia del dúplex. (D) La búsqueda prevista de *tracrARN* se extiende a través de un terminador transcripcional independiente de Rho, identificado en la figura como los nucleótidos en negrita subrayados, localizados al final de la secuencia de *tracrARN*. Se muestra también la ΔG a 37°C.

Figure 1. Description of the elements of the CRISPR-Cas9 Type II-C system and the associated *tracrRNA*-*DR* *RNA* in *Acholeplasma palmae*. (A) Schematic representation of the *cas* genes adjacent to the CRISPR matrix. (B) Identification of the anti-repeat portion of the *tracrRNA* that forms the upper stem. The “query” sequence is the upstream genomic region between the *cas9* and *cas1* gene. The sequence “sbjct” (Subject) is the consensus CRISPR repeat sequence. The anti-repeat detected by BLAST had 82% identity. The region of maximum flexibility of the sequence between nt 50 and 60 that coincides with the “Connection” segment is also shown. (C) Sequence of nitrogenous bases of *tracrRNA* at the CRISPR locus. Anti-repeat (red) of the *tracrRNA*-*DR* *RNA* duplex, unpaired nucleotides forming the bulge (bold), lower stem module (blue). Adjacent to the lower stem is the nexus module that forms a hairpin in *RNA* secondary structure. The “aC” sequence prior to the anti-repeat segment corresponds to the promoter for the duplex sequence. (D) The predicted search for *tracrRNA* extends through a Rho-independent transcriptional terminator, identified in the figure as the underlined bold nucleotides, located at the end of the *tracrRNA* sequence. ΔG at 37°C is also shown.

gracias a sus enzimas asociadas de tipo nucleasas (Campenhout *et al.*, 2019; Kaushik *et al.*, 2019), no hay reportes sobre análisis comparativos *in silico* de la ΔG de las DR en sus matrices CRISPR. Los análisis termodinámicos existentes se han centrado en sistemas tipo I-B (Negahdaripour *et al.*, 2017), I-C (Zhao *et al.*, 2018), I-E (Shen, 2017), I-F (Baliga, 2018) y III-B (Negahdaripour *et al.*, 2017), y aunque en la mayoría se predicen valores de $\Delta G \geq -3,60 \text{ kcal.mol}^{-1}$ hay reportes de ΔG entre $-0,1$ y $-2,9 \text{ kcal.mol}^{-1}$ en estos sistemas (Negahdaripour *et al.*, 2017; Zhao *et al.*, 2018;), al igual que lo predicho en este estudio. Se encontró que pueden existir cerca del doble de estructuras de mínima energía a partir de la secuencia tracrARN que las calculadas a partir de las DR, esto posiblemente se deba a la mayor complejidad estructural y mayor número de interacciones nucleotídicas de los tracrARN. Es importante señalar que el número de posibles estructuras, así como su rol no ha sido aún estudiado a profundidad en los sistemas CRISPR.

Predicción y caracterización de las secuencias tracrARN y DRARN

La porción de la región denominada anti-repetidos del tracrARN, fue localizada dentro de 500 nt aguas arriba de Cas9, entre *cas9* y *cas1* (Figura 1). Se pudo identificar una región de terminación Rho-independiente en sentido 3' a 5' rica en GC seguida de una región rica en T. En relación con la ubicación del tracrARN en *A. palmae* a partir de la

promoting activity is mainly based on the production of secondary metabolites suppressing competitive microbial pathogens occurring in the plant rhizosphere, secretion of the plant growth hormone auxin and the synthesis of volatiles stimulating plant growth and induced systemic resistance (Chen *et al.*, 2009). In addition, some *B. amyloliquifaciens* strains could decrease the incidence of nematodes in some cultivars (Liu *et al.*, 2013). Other species found with antagonistic activity against fungi were *B. velezensis* and *B. subtilis*, members of the *Bacillus* genus as well and closely related to *B. amyloliquifaciens* (Rabbee *et al.*, 2019). Also, it has been shown that microbial colonization has an important impact on sweet basil quality, especially on accumulation of important secondary metabolites like essential oil (Banchio *et al.*, 2009). This outstanding ability is of great importance when production of high-quality products is of interest.

Strains with observed antagonistic effect against *Aspergillus* spp. were used in a germination experiment to determine the effect of the bacteria on total germination, length, and biomass production of basil seedlings. The total seed germination reached between 80 and 90 %. Figure 2 shows that all tested bacteria had no significant difference on total germination of basil seeds as compared to the control. Other reports regarding *Capsicum annuum* seed treatment with *B. amyloliquifaciens* resulted in maximum enhancement of seed germination (84.75 %), seedling vigor (1,423.8) along with an increase in vegetative growth parameters

DR consenso para localizar la región anti-repeticiones y usarla como guía según lo recomendado (Briner *et al.*, 2016) se encontró que la región anti-repeticiones presenta una alta complementariedad con la repetición pre-_{DR}ARN y que está ubicada entre el gen *cas9* (módulo efector) y los genes *cas1-2* (módulo de adaptación) un arreglo de genes que solo se ha descrito en aproximadamente el 1% de los sistemas CRISPR tipo II-C examinados (Chylinski *et al.*, 2013; Chyou y Brown, 2019), entre los que se encuentra la bacteria Gram negativa degradadora de la mucina intestinal humana *Akkermansia muciniphila* ATCC BAA-835 (De Vos, 2017).

Sin embargo, se observó discrepancia en esta clasificación debido a que la ubicación del tracrARN observada en *A. palmae* ha sido definida también en los sistemas tipo II-A (Faure *et al.*, 2019; Ipoutcha *et al.*, 2019). Resultados que corresponden con otros autores donde se señala que la familia de tracrARN es una familia de ARN atípica sin una conservación obvia de la estructura, secuencia o localización dentro de los loci CRISPR-Cas tipo II (Chylinski *et al.*, 2013; Chyou y Brown, 2019; Faure *et al.*, 2019). Mostrando incluso genes Cas9 alterados (Ipoutcha *et al.*, 2019). Esta divergencia en la clasificación del subtipo de CRISPR en el genoma de *A. palmae* considerado quizás se deba a la propia diversidad de estos sistemas biológicos aún en estudio, debido a que la subcategorización de los CRISPR está dada por los arreglos de genes, y no por otros elementos asociados como los transactivadores (Makarova *et al.*, 2011; Van der Oost, 2014). De hecho,

(Gowtham *et al.*, 2018). Recently, a culture-independent next-generation sequencing-based approach was applied to investigate the beneficial impact of *B. amyloliquefaciens* L-S60 on the composition and dynamics of rhizosphere microbiota, and growth conditions of cucumbers during plug seedling. Application of L-S60 significantly altered the structure of the bacterial community associated with the cucumber seedling. Presence of beneficial rhizosphere species such as *Bacillus*, *Rhodanobacter*, *Paenibacillus*, *Pseudomonas*, *Nonomuraea*, and *Agrobacterium* was higher upon L-S60 treatment than in the control group (Qin *et al.*, 2017).

Likewise, hypocotyl (figure 3A) and radicle (figure 3B), dry biomass accumulation did not show any statistically significant difference among groups upon inoculation with isolated antagonistic strains. It is important to highlight that seedlings were analyzed at early stages of growth; in future experiments, the impact of these strains on morphometric characteristic of later stages of development will be carried out. Although, there were no statistically significant differences between treatments, a tendency to increase the dry weight can be observed in the hypocotyl with ALAT81 and ALMT74 and in radicle with ALMH71 and ALMT74.

In terms of radicle length, ALAH51, ALMR73, ALMT52, ALAT64, ALMH71, ALMH42, ALAH91 and ALMH92 were statistically equivalent as compared to the control. These observations indicate that

en el primer sistema CRISPR/Cas reportado en *A. palmarum* como de tipo II-A, se identificó un gen *csn2* (Ipoutcha *et al.*, 2019), un gen que, aunque se considera una firma específica de subtipo (Ka *et al.*, 2018), curiosamente no fue identificado en este estudio bajo las condiciones planteadas y con las herramientas computacionales propuestas.

Además, se ha reportado una mayor relación filogenética entre el sistema II-A descrito en *A. palmarum* con los II-C, que con los II-B (Ipoutcha *et al.*, 2019). Por lo tanto, es probable que el sistema predicho en *A. palmarum* represente una posible variante del sistema CRISPR subtipo II-C con la misma ubicación del *tracrARN* mostrada por los subtipos II-A. Al mismo tiempo que se pudo identificar la secuencia terminadora Rho-independiente rica en Ts como se ha reportado en estudios anteriores (Briner *et al.*, 2014).

Debido a las limitaciones de los algoritmos de predicción *in silico* de los promotores y los terminadores transcripcionales, se sugiere considerar las predicciones del promotor y del terminador transcripcional solo como un paso de apoyo y como parte de la caracterización de dúplex, pero no esencial, para la identificación de los *tracrARN*, como se ha recomendado (Chylinski *et al.*, 2013).

Luego del análisis de las regiones *tracrARN* y *DR* *ARN* usando BLAST, se encontró que no existe homología con ninguna región genómica de las especies presentes en el GenBank/NCBI database. La longitud de la región *tracrARN* es de 107 nt, su contenido de GC es de 39,25%.

the interaction of the mentioned bacteria strains above did not affect radicle growth during early seedling development. However, the rest of the isolated strains tested had negative effects, some of them at higher extend like ALMT73 and ALAT81. On the other hand, hypocotyl length did not show statistically significant difference of any of the treatments as compared to the control (table 4).

Further experimentations will be carried out to find out if these *Bacillus* species strains show antagonistic activities against other important phytopathogens if they have an impact on basil growth and quality in terms of essential oil and other desirable traits. Moreover, these strains could be used for inoculation of other plants to investigate their biotechnological potential.

Conclusion

To our knowledge this is the first time that *B. amyloliquefaciens* strains have been isolated from basil cultivars. The biotechnological use of these bacteria to protect basil plants against fungi is a sustainable and reliable option to reduce crop loss due to fungi incidence. Furthermore, these bacteria are remarkable allies for basil plant development due to their ability to colonize them without damage. The investigation carried out and presented here in establishes the basis for further studies on agronomic useful these *Bacillus* strains that are naturally occurring in basil plants around La Paz, Baja California Sur, Mexico, where we can find very

Presenta una masa molar de 33 250,59 Da y un Tm de 74,71 °C, con una región de máxima flexibilidad entre los nucleótidos 50 y 60.

La región ^{DR}ARN presenta una longitud de 34 nt con un contenido de GC de 44,12%. Su masa molecular es de 10 530,77 Da y su Tm es de 63,21 °C, sin regiones flexibles predecibles. El alto contenido de pirimidinas en la región crARN y de purinas en la tracrARN calculado en *A. palmae*, y que impacta claramente en sus temperaturas de fusión, corresponden con lo reportado en la literatura (Briner *et al.*, 2016; Chyou y Brown, 2019). El ^{DR}ARN estudiado cuenta con una estabilidad medianamente conservada entre la región anti-repeticiones con la repetición pre-^{DR}ARN como se ha descrito (Briner *et al.*, 2016; Chyou y Brown, 2019). Observaciones que muestran que la versatilidad y ventaja del uso de estas moléculas está estrechamente relacionada con la flexibilidad y estabilidad termodinámica de los complejos RNP establecidos, lo cual puede ser clave en la especificidad de corte como se ha señalado en la edición programable de genes (Kaushik *et al.*, 2019).

Los análisis de la secuencia genética mostraron que la región comprendida entre el segmento denominado nexus y la secuencia de terminación es la región con máxima flexibilidad estructural, lo que coincide con la secuencia localizada entre las conformaciones tallo-bucle típicas del módulo de terminación del dúplex (Briner *et al.*, 2014; Briner

distinctive environmental conditions like extreme temperatures, high soil salinity and drought.

Acknowledgments

This work was supported by the Department of Agronomy at the Universidad Autónoma de Baja California Sur. Authors are thankful to Yozitlali Villavicencio-Velázquez and José Enrique Amador-Sánchez for technical support.

End of English Version

et al., 2016). La elasticidad en dicha región del tracrARN hace que el dúplex sea susceptible a cambios conformacionales, sin embargo, gracias a la carencia de regiones de menos rígidas detectables en el módulo ^{DR}ARN la estructura final resultante es termodinámicamente estable, sin cambios importantes en términos del tamaño expresado en radio de los dúplex. Aspecto importante para el aprovechamiento de este tipo de moléculas como precursores de sgARNs, gracias a que dicha plasticidad y estabilidad ha permitido, alternativamente al uso de los sistemas naturales Cas9-tracrARN-^{DR}ARN, generar un ARN quimérico que imita el dúplex tracrARN-^{DR}ARN para un sistema de dos componentes que combina la flexibilidad del ARN de interferencia (iARN) con la eficiencia de las enzimas de restricción (Briner *et al.*, 2014).

La predicción de la estructura secundaria del dúplex tracrARN-

DR-ARN mostró una estructura termodinámicamente muy estable en términos de ΔG de $-79.78 \text{ kcal. mol}^{-1}$ a 37° C (Figura 2), con un radio $\approx 0,334 \text{ nm}$. Adicionalmente, al examinar teóricamente tanto la ΔG como el tamaño del dúplex a distintos intervalos de temperatura se encontró que el dúplex presenta una conformación termodinámicamente estable en todas las temperaturas probadas, sin cambios importantes en un rango aproximado de 29° C , específicamente entre 13° C y 41° C . Solo se predijo cambios conformacionales en relación a su radio a temperaturas $\leq 12^\circ \text{ C}$ y $a \geq 42^\circ \text{ C}$, con radios de $0,380 \text{ nm}$ y $0,323 \text{ nm}$, respectivamente. Aunque se observó un aumento en el tamaño teórico del dúplex a bajas temperaturas, y una disminución a altas temperaturas, la diferencia del tamaño en el valor absoluto entre las conformaciones sometidas a los valores extremos de temperatura probadas fue de apenas $\approx 0,057 \text{ nm}$, este aspecto promueve la estabilidad conformacional. Curiosamente, aunque se ha descrito que otros sistemas CRISPR se ven afectados y modulados de manera importante por variaciones en la temperatura (Høyland-Krogsho *et al.*, 2018), en esta investigación el complejo genético estudiado muestra una estabilidad, conformación y termodinámica poco variable en función de la temperatura. Posiblemente esto se deba a que estos sistemas son dependientes del correcto funcionamiento de ambos módulos tracrARN-DR-ARN (Makarova *et al.*, 2011), lo que justificaría que el dúplex forme parte de un mecanismo

medianamente conservado, como se ha descrito para otras estructuras CRISPR (Kunin *et al.*, 2007).

En la Figura 2 se muestra la estabilidad conformacional en relación al ΔG del dúplex, el cual se muestra con un tamaño aproximado en términos de radio de $0,334 \text{ nm}$. Adicionalmente, se evidencia una conformación termodinámicamente estable en la mayoría de las temperaturas probadas, sin cambios importantes en un rango entre 13° C y 41° C . Solo se predijo cambios conformacionales en relación a su volumen a temperaturas $\leq 12^\circ \text{ C}$ y $a \geq 42^\circ \text{ C}$, con tamaños calculados de $0,380 \text{ nm}$ y $0,323 \text{ nm}$, respectivamente. Todas las estructuras predichas resultaron conformaciones termodinámicamente favorables.

En relación con la ΔG , la media termodinámica entre 37° y 23° C fue de $\approx 9 \text{ kcal. mol}^{-1}$. Mientras que a las temperaturas extremas de este estudio fue $\leq 4 \text{ kcal. mol}^{-1}$. Es importante destacar que la diferencia termodinámica en términos de valor absoluto entre los valores extremos de temperatura probados fue de $\approx 38 \text{ kcal. mol}^{-1}$. Se predijo además un total de otras 10.193.838 posibles estructuras de mínima energía y por lo tanto factibles en su formación a partir de la secuencia tracrARN en un rango de energía de $15,3 \text{ kcal. mol}^{-1}$ por encima de la ΔG . Por otro lado, las energías de acoplamiento entre la Cas9 y las regiones del dúplex fueron muy cercanas y termodinámicamente favorecidas según la función de puntuación HDOCK, siendo estas de $-217,14 \text{ kcal. mol}^{-1}$ para la región DR-ARN , y $-200,66 \text{ kcal. mol}^{-1}$ para la región tracrARN .

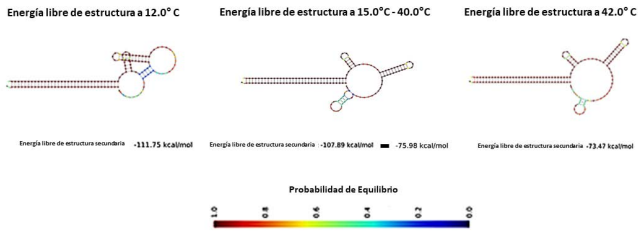


Figura 2. Cambios conformacionales y termodinámicos del dúplex tracrARN_{DR}-ARN en *A. palmarum* a los intervalos extremos de temperatura probados.

Figure 2. Conformational and thermodynamic changes of the tracrRNA_{DR}-RNA duplex in *A. palmarum* at the extreme temperature ranges tested.

Curiosamente, la función de puntuación MolDock Score solo predijo un acoplamiento termodinámicamente favorecido entre la nucleasa y la región _{DR}ARN (-409,367 kcal.mol⁻¹) contrario a lo calculado entre la región tracrARN y Cas9 (25 449,706 kcal.mol⁻¹) con una interacción no favorable (Figura 3). La función de puntuación MolDock Score mostró discrepancias y solo predijo un acoplamiento favorecido entre la nucleasa y la región _{DR}ARN, calculando una interacción no favorable con la región tracrARN. Esto puede deberse quizás a limitaciones en el algoritmo MolDock como el requerimiento de la preparación de los complejos en el software Molegro Virtual Docker (MVD-v.7.0.0) antes del análisis con el MMV (Thomsen y Christensen, 2006; Azeez *et al.*, 2012) y en este

estudio el complejo fue preparado y ensamblado en el software HDock por la naturaleza del ligando (Yan *et al.*, 2017), o posiblemente por el gran tamaño de la secuencia tracrARN utilizada. Sin embargo, en todos los casos las interacciones observadas están representadas por múltiples puentes de hidrogeno, interacciones electrostáticas y estéricas, las cuales facilitan un acoplamiento termodinámicamente favorecido entre los módulos del trans-activador y la nucleasa, especialmente con la región _{DR}ARN, y que en conjunto conforman una estructura que permite un estable ensamblaje del complejo RNP, promoviendo un eficiente reconocimiento de las secuencias dianas y el consecuente clivaje (Lau y Davie, 2017).

En la Figura 3 se muestran las energías de acoplamiento entre la Cas9 y las regiones del dúplex las cuales fueron muy cercanas y termodinámicamente favorecidas según la función de puntuación HDOCK. Las energías fueron $-217,14 \text{ kcal.mol}^{-1}$ para la región $_{DR}ARN$, y $-200,66 \text{ kcal.mol}^{-1}$ para la región $tracrARN$. La función de puntuación MolDock Score solo predijo un acoplamiento termodinámicamente favorecido entre la nucleasa y la región $_{DR}ARN$ ($-409,367 \text{ kcal.mol}^{-1}$) contrario a lo calculado entre la región $tracrARN$ y Cas9 ($25449,706 \text{ kcal.mol}^{-1}$) con una interacción no favorable, sin embargo, en todos los casos el software MMV predijo interacciones representadas por puentes de hidrógeno, interacciones electrostáticas y estéricas. Es importante señalar que si bien, este trabajo se centró en la caracterización del dúplex $tracrARN$ - $_{DR}ARN$, incluyendo su energía de interacción con la nucleasa Cas9. Esto fue predicho a partir de la secuencia reportada en el GenBank/NCBI y según los modelos tridimensionales presentes en el *Protein Data Bank*. Estos modelos fueron utilizados como plantilla por SWISS-MODEL y luego refinados con HDOCK, al no existir para el momento de realizado este estudio una cristalografía de la endonucleasa Cas9 asociada a CRISPR en *A. palmae*. El modelo utilizado estuvo representado por una estructura con una cobertura del 90 %, pero con una identidad de solo el 40 % con la estructura cristalina de la Cas9 de *Staphylococcus aureus*. Un resultado

que muestra una gran diversidad proteica y la poca similitud a nivel de esta nucleasa entre *A. palmae* y el resto de los grupos microbianos relacionados en términos de sistemas CRISPR. Lo cual puede contribuir con la predicción de afinidades diferentes y, por ende, con resultados contradictorios. Por lo tanto, se recomienda realizar un número mayor de predicciones de energía relativa de unión considerando otros tipos de Cas9, especialmente porque existe una gran diversidad de estas moléculas (Bae *et al.*, 2014), así como reportes de variantes alteradas estrechamente relacionadas (Ipoutcha *et al.*, 2019), con la finalidad de validar la metodología empleada.

Adicionalmente, se sugiere modelar diversas conformaciones de la secuencia $tracrARN$ - $_{DR}ARN$, así como realizar acoplamientos macromoleculares del dúplex completo con las nucleasas consideradas porque nuestro enfoque se centró en estudiar la afinidad de la nucleasa sobre los elementos individuales del dúplex porque se sabe que la Cas9 interactúa fuertemente con la región $tracrARN$ al mismo tiempo que participa en la maduración de la región $_{DR}ARN$ (Jinek *et al.*, 2012).

Conclusión

El uso de herramientas bioinformáticas permiten la predicción y estudio de sistemas CRISPR-Cas tipo II y de sus elementos claves para la edición genética en el fitoendosimbionte *A.*

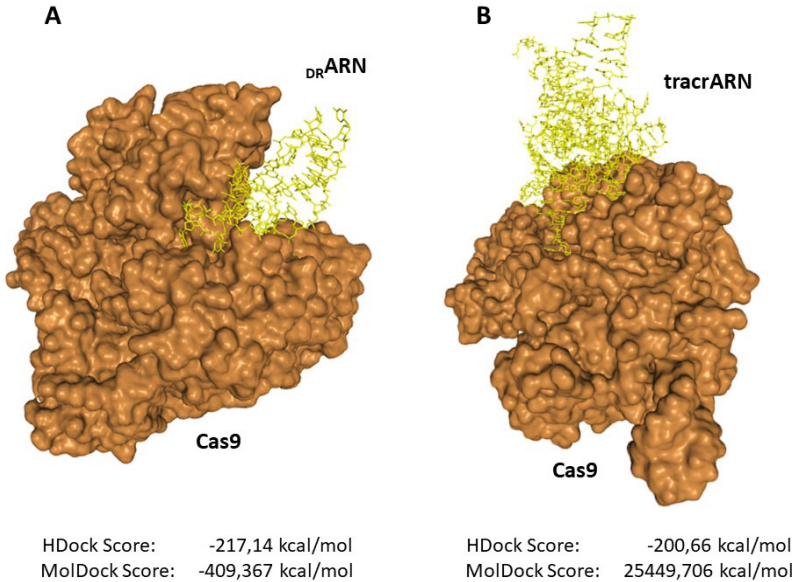


Figura 3. Acoplamiento macromolecular de las regiones del dúplex tracrARN-_{DR}ARN con la Cas9 en *A. palmae*.

Figure 3. Macromolecular docking of the regions of the tracrRNA-_{DR}RNA duplex with Cas9 in *A. palmae*.

palmae. Estas estrategias ofrecen alternativas teóricas diversas para los sistemas CRISPR comerciales aplicables a la biotecnología vegetal. Se requieren más estudios, así como demostraciones experimentales en modelos celulares. Este estudio representa el primer reporte sobre la estabilidad termodinámica y el acoplamiento molecular de elementos asociados al dúplex tracrARN-_{DR}ARN en la especie fitosimbionte *A. palmae*.

potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, 30(10),1473-1475.

Baliga, P., M. Shekar and M. Nagarajappa. 2018. Investigation of direct repeats, spacers and proteins associated with clustered regularly interspaced short palindromic repeat (CRISPR) system of *Vibrio parahaemolyticus*. *Mol Genet Genom*, 294(1),253–262.

Belhaj, K., A. Chaparro-Garcia, S. Kamoun, N. Patron and V. Nekrasov. 2015. Editing plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.*, 32(1),76–84.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., and Sayers, E. W. 2018. GenBank. *Nucleic Acids Res.*, 46(D1),D41-D47.

Cited literature

Bae, S., Park, J., and Kim, J. S. 2014. Cas-OFFinder: a fast and versatile algorithm that searches for

- Bumgardner, E., W. Kittichotirat, R. Bumgarner and P. Lawrence. 2015. Comparative genomic analysis of seven *Mycoplasma hyosynoviae* strains. *Microbiologyopen*, 4(2),343-359.
- Briner, A, P. Donohoue, A. Gomaa, K. Selle, E. Slorach, C. Nye and R. Barrangou. 2014. Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol. Cell*, 56(2),333-339.
- Briner, A., E. Henriksen and R. Barrangou. 2016. Prediction and validation of native and engineered Cas9 guide sequences. *Cold Spring Harb. Protoc.*, 2016(7),pdb-prot086785.
- Chylinski, K., A. Le Rhun and E. Charpentier. 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biology*, 10(5),726-737.
- Chyou, T. and C. Brown. 2019. Prediction and diversity of tracrRNAs from type II CRISPR-Cas systems. *RNA Biology*, 16(4),423-434.
- Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E., Vergnaud, G., Gautheret, D. and Pourcel, C. 2018. CRISPRCasFinder, an update of CRISPRfinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.*, 46(W1),W246-W251.
- De Vos, W. 2017. Microbe Profile: *Akkermansia muciniphila*: a conserved intestinal symbiont that acts as the gatekeeper of our mucosa. *Microbiology*, 165(5),646-648.
- Dupuis, N., E. Holmstrom and D. Nesbitt, D. 2014. Molecular-crowding effects on single-molecule RNA folding/unfolding thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U. S. A.*, 111(23),8464-8469.
- Høyland-Krogsho, N., K. Muñoz and B. Bassler, B. 2018. Temperature, by controlling growth rate, regulates CRISPR-Cas activity in *Pseudomonas aeruginosa*. *mBio*, 9(6),e02184-18.
- Ipoutcha, T., Tsarmopoulos, I., Talenton, V., Gaspin, C., Moisan, A., Walker, C. A., Brownlie, J., Blanchard, A., Thebault, P. and Sirand-Pugnet, P. 2019. Multiple origins and specific evolution of CRISPR/Cas9 systems in minimal bacteria (*Mollicutes*). *Front. Microbiol.*, 10(2701),1-14.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096),816-821.
- Ka, D., Jang, D. M., Han, B. W., Bae, E. 2018. Molecular organization of the type II-A CRISPR adaptation module and its interaction with Cas9 via Csn2. *Nucleic Acids Res.*, 46(18),9805-9815.
- Kaushik, I., S. Ramachandran and S. Srivastava. 2019. CRISPR-Cas9: A multifaceted therapeutic strategy for cancer treatment. *Semin. Cell Dev. Biol.*, 96(1),4-12.
- Kunin, V., R. Sorek and P. Hugenholtz. 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.*, 8(4),R61.
- Kube, M., C. Siewert, A. Migdoll, B. Duduk, S. Holz, R. Rabus and R. Reinhardt. 2014. Analysis of the Complete Genomes of *Acholeplasma brassicae*, *A. palmae* and *A. laidlawii* and their comparison to the obligate parasites from '*Candidatus Phytoplasma*'. *J. Mol. Microbiol. Biotechnol.*, 24(1),19-36.
- Lau, V. and J. Davie. 2017. The discovery and development of the CRISPR system in applications in genome manipulation. *Biochem. Cell Biol.*, 95(2),203-210.
- Makarova, K., D. Haft, R. Barrangou, S. Brouns, E. Charpentier, P. Horvath and J. Van Der Oost. 2011. Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.*, 9(6),467-477.
- Mirabelli, P., L. Coppola and M. Salvatore. 2019. Cancer Cell Lines Are Useful Model Systems for Medical Research. *Cancer*, 11(8),1098-1116.

- Negahdaripour, M., N. Nezafat, N. Hajjighahramani, S. Rahmatabadi and Y. Ghasemi. 2017. Investigating CRISPR-Cas systems in *Clostridium botulinum* via bioinformatics tools. *Infection, Genet. Evol.*, 54(2),355-373.
- Rodrigues, S. D., Karimi, M., Impens, L., Van Lerberge, E., Coussens, G., Aesaert, S., Rombaut, D., Holtappels, D., Ibrahim, H., Van Montagu, M., Wagemans, J., Jacobs, T., De Coninck, B. and Pauwels, L. 2021. Efficient CRISPR-mediated base editing in *Agrobacterium* spp. *Proc. Natl. Acad. Sci. U. S. A.*, 118(2),e2013338118.
- São, C. 2018. Engineering of phage-derived lytic enzymes: Improving their potential as antimicrobials. *Antibiotics*, 7(2),29-32.
- Shen, J., D. Zhao, R. Sasik, and J. Luebeck. 2017. Combinatorial CRISPR-Cas9 screens for *de novo* mapping of genetic interactions. *Nat. Methods*, 14(6),573-578.
- Tully, J., R. Whitcomb, D. Rose, J. Bove, P. Carle, N. Somerson and S. Eden-Green. 1994. *Acholeplasma brassicae* sp. nov. and *Acholeplasma palmae* sp. nov., two non-sterol-requiring mollicutes from plant surfaces. *Int. J. Syst. Evol. Microbiol.*, 44(4),680-684.
- Thomsen, R. and M. Christensen. 2006. MolDock: a new technique for high-accuracy molecular docking. *J. Med. Chem.*, 49(11),3315-3321.
- Westra, E., A. Buckling and P. Fineran. 2014. CRISPR-Cas systems: beyond adaptive immunity. *Nat. Rev. Microbiol.*, 12(5),317-326.
- Yan, Y., D. Zhang, P. Zhou, B. Li, and S. Huang. 2017. HDock: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.*, 45(1),365-373.
- Zhao, X., Z. Yu and Z. Xu. 2018. Study the features of 57 confirmed CRISPR loci in 38 strains of *Staphylococcus aureus*. *Front. Microbiol.*, 9(1591),1-20.
- Zhang, Y., A. Malzahn, S. Sretenovic and Y. Qi. 2019. The emerging and uncultivated potential of CRISPR technology in plant science. *Nat. Plants*, 5(1),778-794.