Universitat de Girona

Facultat de Ciències

Memòria del Treball Final de Grau

Títol del treball:

The search for possible endophytes in the growth zone of maize leaves

---

Estudiant: Anna Vidal Amill

Grau en Biologia

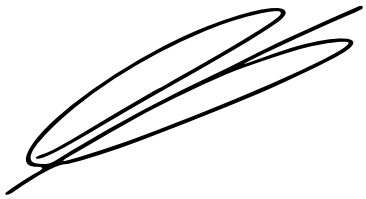
Correu electrònic: anna.vidal@hotmail.es

Tutor: Dr. Lluis Bañeras

Cotutor*: Prof. Gerrit Beemster

Empresa / institució: IMPRES – University of Antwerp

Vistiplau tutor (i cotutor*):



Nom del tutor: Dr. Lluis Bañeras

Nom del cotutor*: Prof. Gerrit Beemster

Empresa / institució: IMPRES – University of Antwerp

Correu(s) electrònic(s):

lluis.banyeras@udg.edu

gerrit.beemster@uantwerpen.be

*si hi ha un cotutor assignat

Data de dipòsit de la memòria a secretaria de coordinació: 20/07/2017

1. Resum [CAT]

L'espècie *Zea mays*, tot i ser una de les plantes més importants per la dieta humana i en el sector agroalimentari, és desconeguda respecte als endòfits que presenta associats. En els últims anys, els endòfits han guanyat rellevància i interès, ja que aquests organismes poden alterar el nostre coneixement sobre la relació que hi ha entre bacteris i plantes. Han sorgit molts estudis, en els últims anys, sobre aquest tema, però encara hi ha un gran desconeixement respecte al paper que desenvolupen aquests microorganismes al colonitzar un cultiu de moresc i la seva interacció amb el genoma de l'hoste. El principal objectiu d'aquest projecte és trobar una evidència de la presència d'endòfits bacterians en la zona de creixement de les fulles de blat de moro i, a la vegada, desenvolupar una tècnica de PCR per aïllar-los. La zona d'estudi de la planta ha sigut seleccionada perquè no s'havia estudiat prèviament en termes d'endòfits. Aquest projecte tracta de localitzar seqüències en les fulles de la planta de blat de moro que no estiguin relacionades amb el genoma de *Zea mays*, ja que poden correspondre a un bacteri i, per tant, ser una evidència de la presència d'endòfits. Per tal de trobar una evidencia robusta, primer, seqüències obtingues per Next Generation Sequencing (NGS) van ser alineades amb el genoma de referència del panís [Taxon ID: 4577]. Totes aquelles seqüències que no van alinear pel genoma de referència van ser utilitzades al llarg de l'estudi. Es va fer un BLAST (Boratyn*, et al.* 2012) amb aquestes seqüències, utilitzant la base de dades de la NCBI, per tal d'obtenir un informe taxonòmic. El millor alineament va ser per l'espècie bacteriana *Desulfitobacterium hafniense*. Per confirmar aquest resultat, es varen dissenyar varies PCRs utilitzant dos tipus diferents de primers: Els primers generalistes (A) i els primers específics (B) que eviten l'amplificació de gens del cloroplast. Els resultats d'aquestes PCRs van ser seqüenciats seguint la tècnica Illumina. Després de realitzar l'informe taxonòmic, el millor alineament obtingut per les mostres en les quals es varen utilitzar els primers generalistes (A) va ser per gens 16S rRNA de cloroplast i per bacteris no cultivables, els quals no ens aporten informació útil d'evidència d'endòfits. Tot i això, quan s'utilitzaven els primers específics (B), els millors alineaments eren pel bacteri no cultivable FFCH13347. Ja que no es van obtenir els mateixos resultats que en la primera anàlisi realitzada utilitzant la informació de NGS, no podem confirmar la presència de *Desulfitobacterium hafniense*, ni de cap altra bacteri, en la zona de creixement de les fulles de blat de moro. Per tant, són necessaris més estudis sobre la presència d'endòfits en aquesta zona de la planta.

Resumen [CAST]

La especie *Zea mays,* todo i ser una de las plantas más importantes para la dieta humana i en el sector agroalimentario, es desconocida respecto los endófitos que presenta asociados. En los últimos años, los endófitos han ganado relevancia y interés, ya que estos organismos pueden alterar nuestro conocimiento sobre la relación que hay entre bacterias y plantas. Han surgido muchos estudios, en los últimos años, sobre este tema, pero aún hay un gran desconocimiento sobre el papel que desarrollan estos microorganismos al colonizar un cultivo de maíz y su interacción con su genoma. El principal objetivo de este estudio es encontrar una evidencia de la presencia de endófitos bacterianos en la zona de crecimiento de las hojas del maíz, y a la vez, desarrollar una técnica de PCR para aislarlos. La zona de estudio de la planta ha sido seleccionada porque no se había estudiado previamente en términos de endófitos. Con tal de encontrar una evidencia robusta, primero, secuencias obtenidas por Next Generation Sequencing (NGS) fueron alineadas con el genoma de referencia del maíz [Taxon ID: 4577]. Todas las secuencias que no fueron alineadas con el genoma de referencia fueron las utilizadas al largo de este estudio. Hicimos un BLAST (Boratyn*, et al.* 2012) con estas secuencias, utilizando la base de datos de la NCBI, con el fin de obtener un informe taxonómico. El mejor alineamiento fue para la especie bacteriana *Desulfitobacterium hafniense*. Para confirmar este resultado, se diseñaron varias PCRs utilizando dos tipos de primers diferentes: primers generalistas (A) y primers específicos (B), los cuales evitan la amplificación de genes del cloroplasto. Los resultados de estas PCRs fueron secuenciados siguiendo la técnica Illumina. Después de realizar un informe taxonómico, el mejor alineamiento obtenido por las muestras en las que se usaron los primers generalistas (A) fueron por genes 16S rRNA de cloroplasto y bacterias no cultivables, los cuales no nos aporta información útil de evidencia de endófitos. Sin embargo, cuando se utilizaban los primers específicos (B), los mejores alineamientos eran por la bacteria no cultivable FFCH13347. Ya que no obtuvimos los mismos resultados en el primer análisis con la información de NGS, no pudimos confirmar la presencia de *Desulfitobacterium hafniense*, ni de ninguna otra bacteria, en la zona de crecimiento de las hojas del maíz. Por lo tanto, son necesarios más estudios sobre la presencia de endófitos en esta zona de la planta.

Abstract [ENG]

Zea mays is one of the most important agricultural crop-species for the human diet, but little is known about the endophytes associated with maize. Within the last few years, endophytes are becoming more interesting as they can change our understanding about the relationship between bacteria and plants. There are a lot of studies with regards to this topic from the last couple of years, but not much is known on the role that these microorganisms develop colonising a particular cultivar, or their interaction with the host's genome. The main objective of this project was to find evidence of bacterial endophyte organisms, and developing a PCR based protocol for isolating bacteria within the leaf growth zone of maize. This area has been chosen for this project since it has not been studied before in terms of endophytes. This project describes non-maize related sequences, extracted from maize leaf material, which could provide evidence for endophytes. In order to find robust evidence, firstly, next generation sequencing (NGS) data were mapped against the maize reference genome [Taxon ID: 4577]. The sequences that did not align against the maize reference genome, were the ones used during the study. These sequences were BLASTed (Boratyn *, et al.* 2012) using the NCBI database to get a taxonomic report. The most significant alignment was for *Desulfitobacterium hafniense.* To support this, various PCR protocols were designed by using two different kind of primers: general bacterial primers (A) and specific bacterial primers (B), which avoid chloroplast gene amplification. These were then sequenced with Illumina. The best alignments for general bacterial (A) primers were for 16S rRNA chloroplast genes and uncultured bacteria, which did not give useful information. However, when using the specific bacterial primers (B), the best hits were for uncultured bacteria clone FFCH13347. Since we were not able to get the same result as in the NGS data taxonomic study, we are not able to confirm the presence of *Desulfitobacterium hafniense* in the leaf growth zone or the presence of other bacteria organisms. Further analysis about possible endophytes within the leaf growth zone are thus needed.

2. Introduction

Endophytes are microorganisms that live inside multicellular organisms without causing any damage to its host (Wilson, 1995). All known plant endophytes are either fungi or bacteria, as plants constitute a vast and diverse niche for these organisms (Kusari, et al. 2012). Mycorrhiza and rhizobium, are the two most studied plant endophytes, typically being associated with the soil/root interface (Rosenblueth, *et al*. 2006).

No known plant species lack endophytes. The few examples of apparent absence of endophytes within plants mainly exist due to some microorganisms which are difficult to culture or isolate (Rosenblueth, *et al.* 2006). Endophytic bacteria have been reported in the roots, leaves, stem, fruits, seeds, tubers, ovules and inside some nodules (Hallmann*, et al.* 1997; Benhizia*, et al.* 2004; Roesch*, et al.* 2008). However, roots have the highest amount of endophytes, when compared against the above-ground plant tissues (Benhizia*, et al.* 2004).

The population density of endophytes is highly diverse. Mainly, it depends on the endophyte species and the host genotype, but also in the developmental host stage and environmental conditions. Therefore, biotic and abiotic factors condition endophytic populations (Rosenblueth*, et al.*2006).

If we consider the fact that endophytes reside within plants, it is evident to think that these microorganisms are continuously interacting with the host plant. There are different hypotheses on this plant-endophyte crosstalk, which results in a mutualistic relationship. The most popular and accepted hypothesis, is the *Balanced antagonism* hypothesis: the colonisation between the endophyte and the host is a balance of two antagonists. This theory can be explained with the following situations *(Figure 1)*. If bacteria or fungal virulence and plant defences are balanced, the relationship is asymptomatic and avirulent *(Figure 1A)*. The bacteria endophytes, as well as fungal endophytes, may obtain some nutrients from the plant, and at the same time protection from abiotic stress conditions, such as desiccation. But this phase is transitory, as some endophytes may become pathogenic, depending on the host defence responses and the developmental stage of host and the endophyte itself (Rosenblueth*, et al.* 2006). Also, environmental factors play a major role to modify this balanced antagonist interaction. If the delicate antagonism is destabilized, two responses are possible. Plant defence mechanisms are completely activated, and endophyte virulence factors will be counteracted causing the microorganism to die. The second response occurs where endophyte virulence is

stronger than the plant defence mechanisms causing that the relationship between these organisms will lead to a plant disease *(Figure 1B)*.

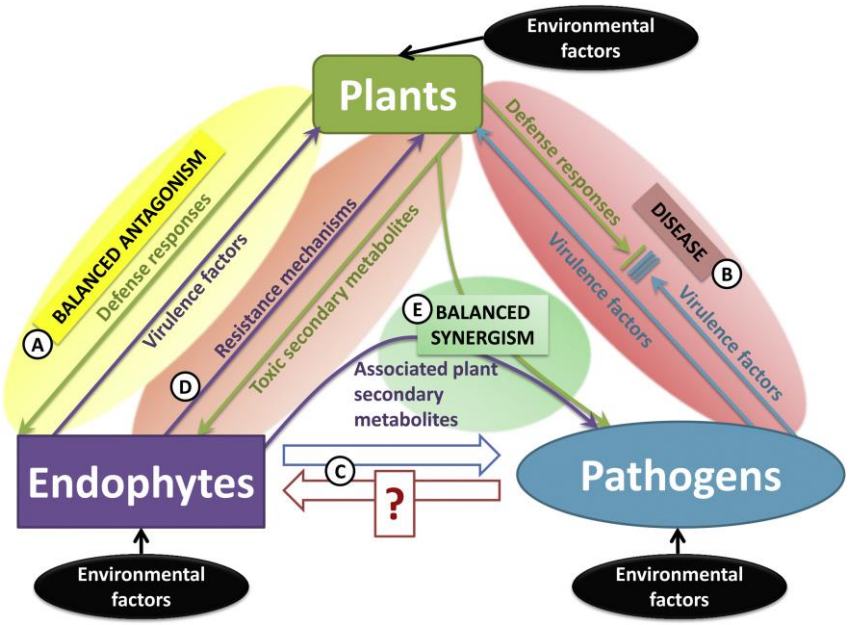


Figure 1: *Schematic interpretation of endophyte-pathogen-plant interaction. A: Balanced antagonism hypothesis representation. B: Disease caused by pathogens is shown. C: Relationship between the endophytes and pathogens. D: Survival strategy for endophytes. E: Balanced synergism is represented. "?" requires further studies to determine its significance* (Kusari, *et al*. 2012).

As many endophytes can be latent pathogens, there might be certain intrinsic or environmental factors which influence them to activate factors, which could lead to pathogenicity *(Figure 1C)* (Carson*, et al.* 2008)

In addition, it is uncommon that a plant is only colonised by a single kind of microorganism. In fact, the presence of diverse endophytes in plant tissues, support that endophytes interact, directly or indirectly, with other associated endophytes (fungus-fungus, bacteria-fungus and/or bacteria-bacteria) (Kusari*, et al.* 2012).

Hallman, *et al.* (1997), reviewed published endophytic bacteria associated with different plants. The bacterial endophytes (class and species) which have been found in association with maize are Beta-proteobacteria (*Burkholderia pickettii*, *Herbaspirillum seropedicae*), Gamma-proteobacteria (*Enterobacter spp., Klebisella variicola*), Firmicutes (*Bacillus megaterium*) and Actinobacteria (*Arthobacter globiformis, Micobacterium testaceum*) (Hallmann*, et al.* 1997).

A previous study reported that the most common bacterial phylum in the endophyte community is the *Proteobacteria* (Santoyo, *et al.* 2016). Due to the amount of interest this area has gained recently, new endophytic organisms are frequently identified.

Until now, endophytic microorganisms have been studied after culturing media. Due to this, it is difficult to understand the mutual beneficial interaction of a plant, as well as the endophytes colonising it (Battu, *et al.* 2016). However, some endophyte contributions to host plant have been described like growth promotion caused by endophyte phytohormone production (Long, *et al.* 2008), native antagonism relationship with pathogens, protection from herbivores, synthesis of novel and secondary products (Rosenblueth, *et al*. 2006), the capacity to increase the quantity of available nutrients for the plant (Bloemberg, *et al.* 2001) and regulate the biotic and abiotic stress tolerance of the plant by using ethylene and reducing the oxidative stress (Glick, *et al.* 2012; Hamilton*, et al.* 2012).

A new high-throughput genomic approach is emerging, which could aid in the identification of endophytes. Nowadays, we have a big amount of genomic plant data, due to the low cost, simplicity, and robust nature of next generation sequencing (NGS) methods (Carson, *et al.* 2008).

This new approach focuses on the analysis of sequences from bacterial and fungal genes obtained from DNA extractions residing in plant tissues (Rosenblueth, *et al.* 2006). This provide better techniques for analysing and improving our understanding of the relationships between the plant and the endophyte (Benhizia, *et al.* 2004). The 16S rRNA gene is usually used for phylogenetic studies because it is highly conserved between different species of bacteria and archaea (Coenye, *et al.* 2003). However, 16S rRNA sequences also contain hypervariable regions, which could provide species-specific signatures useful for the identification of a specific specie (Hanshew, *et al.* 2013).

Chloroplasts are evolutionary descendants from Bacteria, resulting in the 16S rRNA gene being nearly homologous between them (Hanshew*, et al.* 2013). In order to isolate bacterial 16S rRNA gene amplification, while potentially avoiding the chloroplast 16S rRNA gene, it is necessary to find a region appropriate for primer design *(*Hanshew, *et al.* 2013).

Chelius and Triplett (2001) stated that this region exists and it can be found between positions 783-799 of the chloroplast 16S gene (*E.coli* numbering system). This region presents a two base pairs mismatch between positions 798-799. They took advantage of this discovery to design the 799F primer and they also modified the sequence to add another two extra base pairs mismatch at position 783 and 784 (Chelius, *et al.* 2001). It is commonly thought that mismatches between the 3' end of the primer and the targeted sequences, blocks amplification (Kumar, *et al*. 2011; Klindworth, *et al.* 2013).

Even though fungal microorganisms are the most common endophytes in plants, this project focuses on the search of endophytic bacteria in the maize leaf growth zone, because it is a region of the plant which have not yet been studied in terms of endophytes.

3. Objectius [CAT]

Com s'ha comentat anteriorment, la presència de microorganismes en l'interior de la planta és el resultat d'una relació simbiòtica entre els dos organismes. Aquesta àrea d'estudi està guanyant més rellevància, ja que nous descobriments sobre el tema poden canviar la nostra percepció sobre la relació que estableixen plantes i microorganismes. Per tal de contribuir i millorar el coneixement actual sobre els endòfits, aquest projecte consisteix en la cerca d'evidències d'endòfits en la zona de creixement de les fulles de la planta de blat de moro. Aquesta regió de la planta ha sigut seleccionada perquè no s'ha estudiat prèviament en termes d'endòfits. A més, també s'ha intentat dissenyar una PCR que ens permeti la identificació d'endòfits bacterians. Desenvolupar aquesta tècnica farà que futures anàlisis siguin més fàcils i barates a causa de la simplicitat de la tècnica de PCR.

Objectives [ENG]

As previously stated, the presence of microorganisms within plants result in a symbiotic relationship between the organisms. This area of study is becoming more relevant since findings can make a significant difference in the way we understand the relationship between plants and microorganisms. In order to contribute and improve our understanding about endophytes, this project consisted in finding endophytic evidence in the maize leaf growth zone because it is a region of the plant that has not been studied before in terms of endophytes. In addition, we tried to design a PCR, which allows the identification of bacterial endophytes. Developing such a technique will make further analysis cheaper and easier because of the simplicity of PCR.

4. Material and methods

To find evidence of bacterial endophytes within the maize genome from the leaf growth zone, this project has been divided in two sections: NGS sequencing-based endophyte prediction and PCR based approach. Following the bioinformatics analysis and as a confirmation of the results obtained, it was decided to extract DNA from a new control *Zea mays* B73 plant and perform a PCR using different bacterial primers. Then, after getting amplification, samples were sequenced in order to see if the results were the same as the ones obtained in the previous NGS sequencing-based prediction.

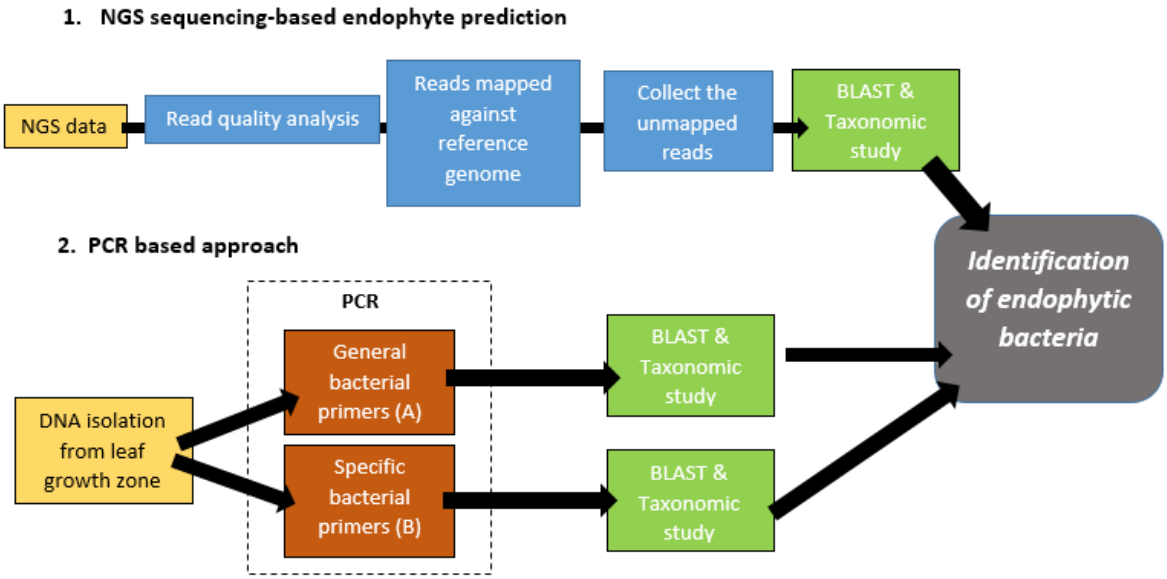


Figure 2: *Schematic workflow to search the endophytic bacterial communities in maize.*

4.1 NGS sequencing-based endophyte prediction

From leaf growth zone of control maize plants B73, next generation sequencing data were obtained. With these data, the quality was checked to remove all the sequences which did not have good quality status. In order to determine if a sequence was competent enough to perform a detailed alignment, parameters such as sequence length, GC contents, number of ambiguous positions, and nucleotide contribution were examined, using the CLC Genomic Workbench's software v.7 (https://www.qiagenbioinformatics.com/).

For these parameters, it was expected to see similar results in all samples as all of them come from the same plant and, therefore, they have the same genomic information. If a sample has a parameter result which stood out from the average values, it was removed. Doing this we ensured to only work with similar quality sequences and therefore, will

know that the differences in the alignments are because of the content of the sequence itself, and not for their quality.

Moreover, PHRED scores were also tested because it indicates the quality of the identification of the nucleobases generated by automated DNA sequencing. The higher the score is, the chances that some nucleobase was mistaken or wrongly identified are lower (Ewing, *et al.* 1998). All sequences with lower PHRED scores than 30 were removed because it was decided to work with reads with low chances of wrongly identified nucleobase.

In order to identify which sequences could be related to an endophyte, they were mapped against *Zea mays* reference genome [Taxonomy ID: 4577]. With this procedure, it was possible to determine which specific regions did not align against reference genome, which therefore, could be related to an endophyte organism rather than a sequence previously thought to has been a part of the maize genome.

Only the unmapped sequences were kept for furthers studies. Also, forward and reverse sequences of each read were joined and contigs were generated to work with longer sequences and ensure the chances of getting better alignments.

4.2 Taxonomy study

The aim of doing all these bioinformatic analyses were to get bacterial endophyte evidence in the maize genome. The evidence could be supported by further analyses explained in the following sections of this project. For example, if in the bioinformatics analysis, evidence of endophytes within the maize genome were detected, such as a high score alignment with some bacteria species, we expect to find the same bacteria after the PCR based approach.

A BLAST (Boratyn*, et al.* 2012) analysis was performed on these sequences, using the NCBI database (Bethesda & National Center for Biotechnology Information, 1998), against *Zea mays* [Taxonomy ID: 4577], in order to identify regions which are not aligned because this allows to describe them as possible endophyte sequences. All the maize-aligned sequences were removed, because this project focussed on describing sequences which did not match with maize, since there are some evidences, *as explained in the Introduction*, that these sequences could be from an endophyte. The following analyses were performed using all the non-aligned sequences.

With the new set of unmapped sequences, an additional BLAST analysis (Boratyn, *et al.* 2012) was performed selecting the parameter “Highly similar sequences (megablast)”, in order to identify if these sequences are from a possible endophyte organism. A taxonomic report generated was then used to determine these highly similar BLAST (Boratyn*, et al.* 2012) hits.

4.3 Sample collection and DNA extraction

Leaf growth zone samples were obtained from *Zea mays*, cultivar B73, plants. Three samples (C1, C2, and C3), were selected from the growth zone of maize, which were then used for further analyses. The C1 sample contained the first three centimetres of the growth zone, the C2 the following three centimetres and consecutively (*Figure 3*). It is thought that different plants have the same bacterial endophytes, so during the experimental work, C1, C2 and C3 samples from various plants were used.

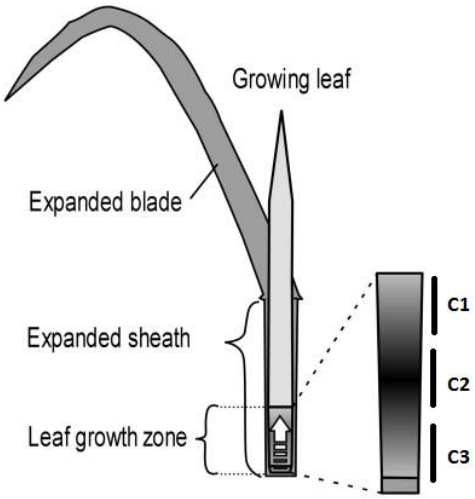


Figure 3: Schematic draw representing where C1, C2 and C3 samples are taken (modified from Kavanova*, et al*. 2006).

For DNA extraction, samples of approximately 1 cm were magnalyzed using the MagNA Lyser (Roche Molecular Systems Inc, Pleasanton, CA) at 6000 rpm for 10 seconds until getting fine powder. After, 150 µL of isolation buffer (100 mM NaCl, 100 mM Tris-HCl buffer, 1 mM EDTA, Sarkosyl and SDS) were added. The samples were magnalyzed again at 6000 rpm for 10 seconds twice. Then, 100 µL of phenol:chloroform:isamyl alcohol (25:24:1) were added and samples were mixed and centrifuged at 14000 rpm for 5 min. 100 µL of the supernatant containing DNA were transferred to a new tube and 200 µL of 100% ethanol were added. After vortexing and centrifuging the samples again, but this time at 1,200 rpm, the fluid was discarded. To make sure there is only DNA in the pellet, 500 µL of 70% ethanol was slowly added and the fluid was discarded. The samples were kept at room temperature or in the oven until the ethanol was completely evaporated. The DNA pellet was then resuspended in 20 µL of nuclease free water. To check the quality and the concentration of our extractions, the NanoDrop® ND-2000 spectrophotometer (Thermo Fisher Scientific Inc. Wilmington, DE) was used.

4.4 PCR based approach

To design a PCR technique, which allows us to detect bacterial endophytes, two different kind of bacterial primers were used: General bacterial primers (A) and specific bacterial primers (B). The second ones are able to avoid the amplification of organelles evolutionarily close to bacteria. Both were specific for the 16S rRNA gene, as this is commonly used for taxonomic and phylogenetic studies. General bacterial primers (A) forward (5' CCTACGGGNGGCWGCAG 3') and reverse (5' GACTACHVGGGTATCTAATCC 3') were specific for the V3 and V4 region, respectively, of the 16S rRNA gene. The annealing temperature suggested by the authors was 50ºC (Bartram*, et al.* 2011).

Each reaction mixture for general bacterial primers (A) contained, for a total volume of 25 µL, 12.5 µL GoTaq G2 Green Master Mix (2x), 2.5 µL of upstream and downstream bacterial primers (10 µM each), 2.5 µL genomic DNA template and 5 µL Nuclease-Free water. The reactions were placed in a Mastercycler personal (Eppendorf. Hamburg, GE) with an initial denaturation procedure at 95ºC for 2 min and 15 seconds followed by 40 cycles of 95°C for 30 seconds, 50ºC as annealing temperature for 1 minute and 72°C for 30 seconds. The extension reaction was performed at 72°C for 5 minutes (Bartram*, et al.* 2011). The tubes were held at 4°C for as much time as we needed. The PCR output quality was confirmed by 2% agarose gel electrophoresis.

Due to inaccurate findings from General bacterial primers *(all the alignments were only with chloroplast and uncultured bacteria, see Results and Discussion for full reason),* new primers were needed in order to exclude chloroplast regions to ensure the chances of getting only bacterial 16S rRNA genes amplified and exclude chloroplast 16S rRNA genes. These new primers were designed and used following the authors instructions.

Specific bacterial primers (B) are a group of five different sets of primers. These five sets of primers are composed of a combination of different forward and reverse primers, all of them specific to avoid chloroplast genes amplification.

Each reaction mixture for PCR had the same composition as the one explained before for general bacterial primers (A). For the set of primers' one, two, three, and four, the authors suggested three different annealing temperatures. To identify the best one, a PCR for each temperature was performed. The PCR products were confirmed by 2% agarose gel electrophoresis allowing us to determine the best temperature. When the temperature for each set of primer was selected, a second PCR was performed with the same conditions as before to have enough PCR product to send for sequencing. An agarose gel was not necessary for this second PCR to confirm the products as it was a repetition of the previous one in order to get enough PCR product for Illumina sequencing.

Table 1: *List of the five different sets of primers used with their sequence, the annealing temperatures suggested by the authors and the reference where the primers were obtained from.*

	Primers (Forward & Reverse)	**Sequences (5' → 3')**	**Annealing temperature (ºC)**	**Reference**
SET 1	**799F**	AACMGGATTAGATACCCKG	48,49,50	Chelsius, *et al. 2001*
	1391R	GACGGGCGGTGWGTRCA		Walker, et al. 2007
SET 2	**799F-mod7**	GGATTAGATACCCKGGT	48,49,50	Hanshew, *et al.* 2013
	1392R	ACGGGCGGTGTGTRC		Hanshew, *et al.* 2013
SET 3	**799F-mod7**	GGATTAGATACCCKGGT	48,49,50	Hanshew, *et al.* 2013
	1391R	GACGGGCGGTGWGTRCA		Walker, *et al. 2007*
SET 4	**799F**	AACMGGATTAGATACCCKG	48,49,50	Chelsius, *et al.* 2001
	1392R	ACGGGCGGTGTGTRC		Hanshew, *et al.* 2013
SET 5	**534F**	CCAGCAGCCGCGGTAAT	53	Muyzer, *et al.* 1993
	783R	ACCMGGGTATCTAATCCKG		Chelsius, *et al.* 2001

4.5 Gel purification

As the objective of this study is to find some sequences that can be identified as endophyte bacteria, we did not know which band size expect. Keeping this in mind, it was decided to work with the different bands independently to identify them as better as possible, even though the band could be a result of unspecific amplification. Gel purification was only used when the electrophoresis gel showed more than one band for the same sample, which meant that sequences with different length were amplified.

The following process was performed using the QIAquick Gel Extraction Kit (Qiagen Group, Hilden, GE), which includes a Buffer QG with pH indicator (Solubilization buffer), Buffer EB (Elution buffer) and a Buffer PE mixed with 40 mL ethanol (Wash buffer) to use during the purification procedure.

With a sterilised and sharp scalpel, the DNA fragment from the agarose gel was excised. Then, the fragment was weighted and three volumes of Buffer QG were added to one volume gel. After, the samples were incubated at 50ºC for 10 minutes in Thermomixer (Eppendorf. Hamburg, GE) until the gel was completely dissolved. A single volume of isopropanol was added to the sample and it was transferred to a QIAquick column. After, the DNA sample was centrifuged for 1 min at 13,000 rpm and the flow-through was discarded. Since our DNA samples are for sequencing, 500 µl of Buffer QG were added and centrifuged again for 1 min. The flow-through was discarded again. In order to wash out samples, 750 µl of Buffer PE were added and the QIAquick column was let to stand for two to five minutes (this last step was only performed because the DNA product will be used for salt-sensitive application as sequencing). Then, it was centrifuged again 1 min at 13,000 rpm to remove residual wash buffer. The QIAquick column was transferred into a clean 1.5 mL microcentrifuge tube. To elute DNA, 25 µl of Buffer EB (10mM Tris·Cl, pH 8.5) were added to the centre of the QIAquick membrane. After letting the column stand between 1 and 4 minutes, the tubes were centrifuged in the same condition as before. To check the quality and the concentration of our extractions, we use the NanoDrop® ND-2000 spectrophotometer (Thermo Fisher Scientific Inc. Wilmington, DE).

4.6 Sequencing and bioinformatics analysis

The sequencing of the samples were carried out by VIB Genomic Service Facility (University of Antwerpen) following an Illumina sequencing (Bennett, 2004). Using the software Geneious version 10.1.3 (http://www.geneious.com) (Kearse*, et al.* 2012) low quality regions of the reads were removed and forward and reverse paired reads were merged. Moreover, a *de novo* assemble was executed to obtain a consensus sequence for each C1, C2 and C3 samples in every set of primers. In addition, with all the consensus sequences obtained, a second *de novo* assemble was done to get a consensus general sequence.

The next stage was to repeat step *4.2 Taxonomy study* with some variations: keeping all the sequences even though they are related with maize. This process was repeated in order to analyse and understand the results of the sequencing process.

5. Results

5.1 NGS sequencing-based endophyte prediction and taxonomic study

For good quality sequences, we expected to get the same length number for all the sequences, a low number of ambiguous positions and PHRED score above 30. Also, a GC content value around 50% and a nucleotide contribution for each nitrogenous base around 25%. Both of these values mean that the sequences contain the four possible nitrogenous bases in the same proportion. All sequences that do not have these characteristics were removed. With the other ones, which fulfil these conditions, the forward and reverse reads were joined and a contig was generated to increase the chances of getting better alignments when doing BLAST (Boratyn*, et al.* 2012) analysis with the NCBI database.

After removing all the sequences that had aligned with the reference maize genome *(as explained in 4.1 NGS sequencing-based endophyte prediction),* a BLAST (Boratyn*, et al.* 2012) was performed using the NCBI database. Most of the sequences present high score alignments (>=200 bits) with sequences related with *Zea mays (Figure 4)*. Therefore, the non-aligned sequences for *Zea mays* were selected and another BLAST (Boratyn*, et al.* 2012) was performed, but this time excluding *Zea mays* at the same time bacteria was selected to identify some bacteria *(Figure 5)*.

Description	Max score	Total score	Query cover	E value	Ident	Accession
Zea mays ATPase 4 plasma membrane-type (LOC100502231), mRNA	6416	6416	96%	0.0	100%	NM_001323633.1
PREDICTED: Zea mays ATPase 4 plasma membrane-type (si687062f01(361)), transcript variant X1, mRNA	2966	2966	79%	0.0	85%	XM_008646672.2
Zea mays full-length cDNA clone ZM_BFb0089L06 mRNA, complete cds	2425	2425	36%	0.0	100%	BT087482.1
PREDICTED: Zea mays uncharacterized LOC100216710 (pco106203), transcript variant X1, mRNA	2206	2206	79%	0.0	81%	XM_008652748.2
Zea mays PCO069619 mRNA sequence	2015	2015	30%	0.0	100%	AY107171.1
Zea mays full-length cDNA clone ZM_BFb0221N13 mRNA, complete cds	1797	1797	65%	0.0	81%	BT067170.1
Zea mays uncharacterized LOC100216710 (pco106203), mRNA	1074	1074	35%	0.0	82%	NM_001143115.1
Zea mays ATPase 4 plasma membrane-type (si687062f01(361)), mRNA	998	998	26%	0.0	85%	NM_001176064.1
Zea mays full-length cDNA clone ZM_BFb0154J22 mRNA, complete cds	985	985	33%	0.0	82%	BT087675.1
Zea mays clone 1703384 mRNA sequence	832	832	12%	0.0	100%	EU943999.1

Figure 4: *Graphic summary and description of the alignments obtained after doing a BLAST* (Boratyn, et al. 2012) *with the contig number 21 (3594 base pairs) from our data set. This blast was optimized choosing the "Highly similar sequences (megablast)" parameter. This only represents the top 10 Blast hits.*

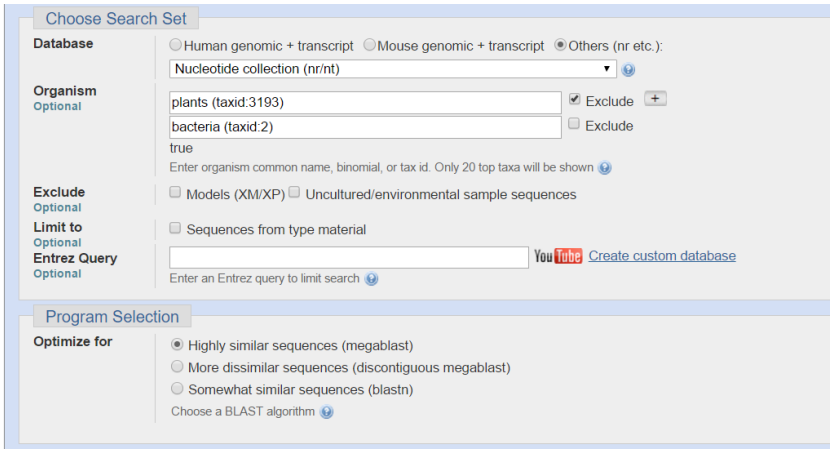


Figure 5: *Selected parameters for the BLAST* (Boratyn, *et al.* 2012) *performed with the sequences that are not maize related. This blast was optimized choosing the "Highly similar sequences (megablast)" parameter, omitting plants and selecting bacteria to find some bacteria organisms related with our sequences.*

As a result, most of the sequences did not have an alignment at all. When the similarity score required between our sequences and the ones from NCBI database was lowered, the number of sequences which present an alignment increased. This was done by choosing the option "Somewhat similar sequences (blastn)" instead of the "Highly similar sequences (megablast)" option. However, the alignments got really poor results so it was decided not to work with them. On average, the alignment score, for the sequences that have it, was between 50-80 bits. Moreover, the regions that hybridise between our sequences and the ones from NCBI database, were short. At the same time an exception

was found. Only contig number 43 (5386 base pairs) had a score of 9947 bits with an identity of 100% with *Desulfitobacterium hafniense* bacteria *(Figure 6)*.

Even though this concrete contig presents high score alignments with different organisms, the best one is with *Desulfitobacterium hafniense* strain because is the organism with higher max score (9947 bits). This means there are 9947 bits or 5386 base pairs that are homologous between contig 43 and the *D. hafniense*'s genome.

Description	Max score	Total score	Query cover	E value	Ident	Accession
Desulfitobacterium hafniense strain PCS-S genome assembly, scaffold: scaffold9	9947	10390	100%	0.0	100%	LK996026.1
Amycolatopsis lurida NRRL 2430, complete genome	9708	9953	100%	0.0	100%	CP007219.1
Actinomyces succiniciruminis strain AM4 genome assembly, scaffold: scaffold2	7260	10029	100%	0.0	100%	LK995457.1
Escherichia coli genome assembly FHI92, scaffold scaffold-6_contig-18.0_1_5295_[organism:Escherichia	6165	9774	98%	0.0	100%	LM997153.1
Escherichia coli genome assembly FHI7, scaffold scaffold-5_contig-23.0_1_5172_[organism:Escherichia	5814	9547	96%	0.0	99%	LM997376.1
Escherichia coli genome assembly FHI79, scaffold scaffold-4_contig-23.0_1_3071_[organism:Escherichia	3832	5656	57%	0.0	99%	LM996658.1
Escherichia coli genome assembly FHI83, scaffold scaffold-1_contig-26.0_1_2008_[organism:Escherichia	3709	3709	37%	0.0	100%	LM996884.1
Escherichia coli genome assembly FHI25, scaffold scaffold-3_contig-24.0_1_2121_[organism:Escherichia	3613	3918	39%	0.0	100%	LM996286.1
Escherichia coli genome assembly FHI71, scaffold scaffold-9_contig-18.1_1129_2927_[organism:Escherichia	3317	3317	33%	0.0	99%	LM996813.1
Escherichia coli genome assembly FHI43, scaffold scaffold-3_contig-14.0_1_2537_[organism:Escherichia	3144	4670	47%	0.0	99%	LM996054.1

Figure 6: *Graphic summary and description of the alignments obtained after doing a BLAST* (Boratyn*, et al.* 2012) *with contig number 43 (5386 base pairs) from our data set. The exact number of the alignment can be seen in the column max score from the description section. There are represented only the top 10 blast hits.*

5.2 PCR based approach

To confirm the presence of Desulfitobacterium hafniense in the leave growth zone of maize, various PCR were designed using new C1, C2 and C3 samples from a different B73 maize plant.

In the first place, when a PCR was performed using 50ºC as annealing temperature for general bacterial primers (A), a band of approximately 300 base pairs can be seen in the agarose gel *(Figure 7)*. However, there are non-specific amplifications that can interfere with the PCR product purity of our samples. Because we need to sequence the PCR products, these non-specific amplifications had to be removed.

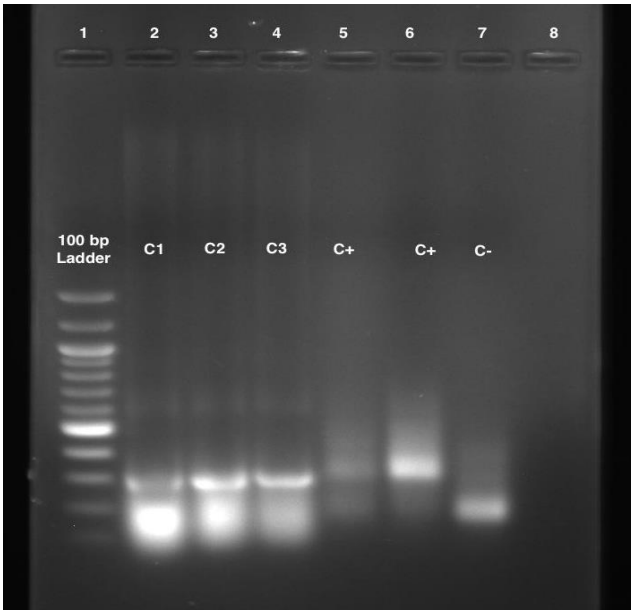


Figure 7: *2% agarose gel electrophoresis for C1, C2 and C3 PCR products using 50ºC as annealing temperature. A 100 base pairs ladder was used in the first column. Samples C1, C2 and C3 can be seen in the second, third and fourth columns respectively. Column five and six are bacterial samples used as positive controls. The 7th column has a water negative control.*

To get rid of the non-specific amplification, the strategy followed was increasing the annealing temperature. By increasing it, we can be sure that our primers will avoid aligning to non-specific regions of the maize genome. Also, one of the positive controls had amplified which means that the amplification we see can correspond to bacteria.

As a result of increasing the annealing temperature to 51ºC, the PCR product purity have improved, as the non-specific amplification could not be seen anymore (*Figure 8*). At the same time, there is no amplification in the negative control, which implies that the samples do not have contamination. Also, the positive control has amplified, which means the bacteria in our samples could have been amplified. After seeing these results, the PCR products were sequenced.

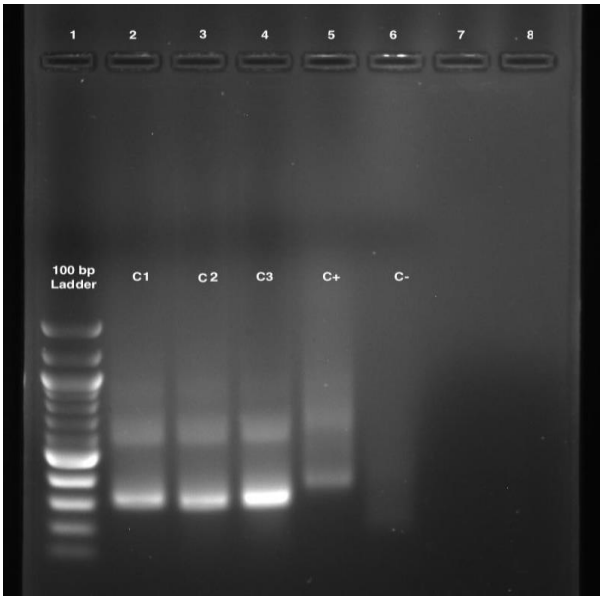


***Figure 8**: 2% agarose gel electrophoresis for C1, C2 and C3 PCR products using 51ºC as annealing temperature. A 100 base pairs ladder was used in the first column. Samples C1, C2 and C3 can be seen in the second, third and fourth columns respectively. Column five and six correspond to bacterial positive control and to negative control, respectively.*

In the second place, for the five sets of specific bacterial primers (B), a PCR reaction was necessary to determine the best annealing temperature for every set of primers *(Figure 9 & 10)*. Then, another PCR reaction was done to get only the PCR products of the ideal annealing temperature for sequencing. These two PCR reactions were performed for every set of primers except for set five, because the authors only suggested one annealing temperature *(Table 1)*.

As seen in *Figure 9*, we can see that all the columns have two bands. The one on the top, approximately 1000 bp, is the interesting one and the one on the bottom, 150 base pairs approximately, could be a result of non-specific amplification of the primers. Even though this incorrect amplification can be seen in all the temperatures, all samples and in both set of primers (1 and 3), when using the set of primer 3, the abundance of this non-specific band was lower. If the three different annealing temperatures are compared, it is clear that the best one for our samples is 49ºC because it is the one with clearer 1000 base pairs bands for both sets. Due to the fact that in all different temperatures there is non-specific amplification and it could easily affect the sequencing fiability, a gel purification was performed for the top (T) and bottom (B) bands from the 49ºC PCR product (following the protocol explained in *4.5 Gel purification*). For set 5, there is only one band in the gel (approximately 300 base pairs), so for the samples that use this combination of primers; gel purification is not necessary.

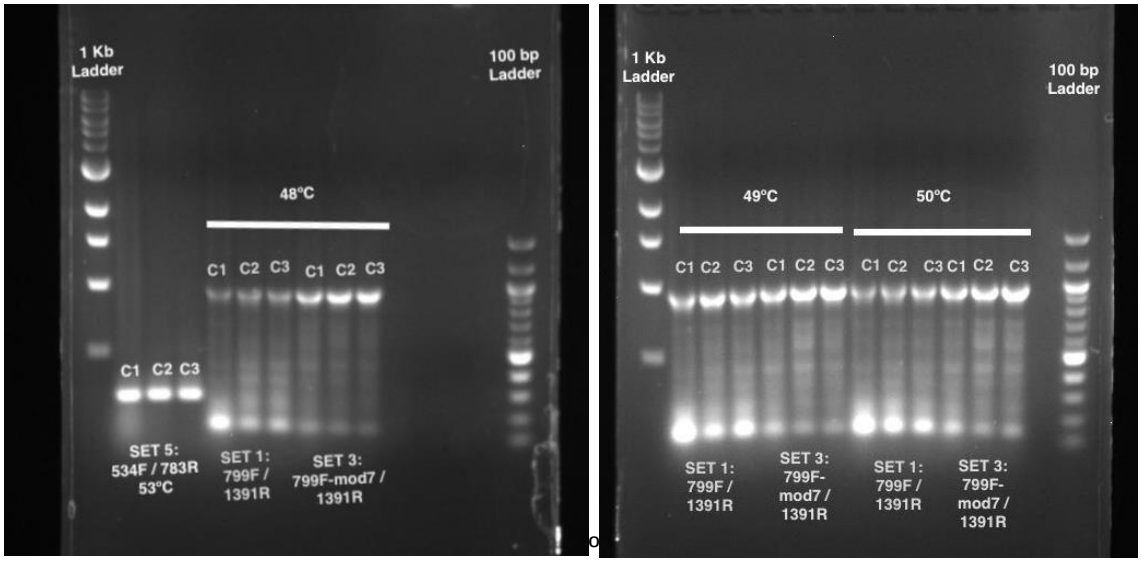


Figure 9: *2% agarose gels electrophoresis for primers set one and three using three different annealing temperatures (48ºC, 49ºC, 50ºC). Also, there is set five but only using one annealing temperature (53ºC). In the first column of the two gels there is a 1 Kbp ladder and in the last column a 100 base pairs ladder. Every set of primers were tested in the three different samples (C1, C2 and C3).*

When using primers set 2 and 4, there is only one band amplification *(Figure 10)*. This band is approximately 1000 base pairs, like the top one seen with primers sets' 1 and 3. As only one band is showed, gel purification was not necessary for these sets of primers.

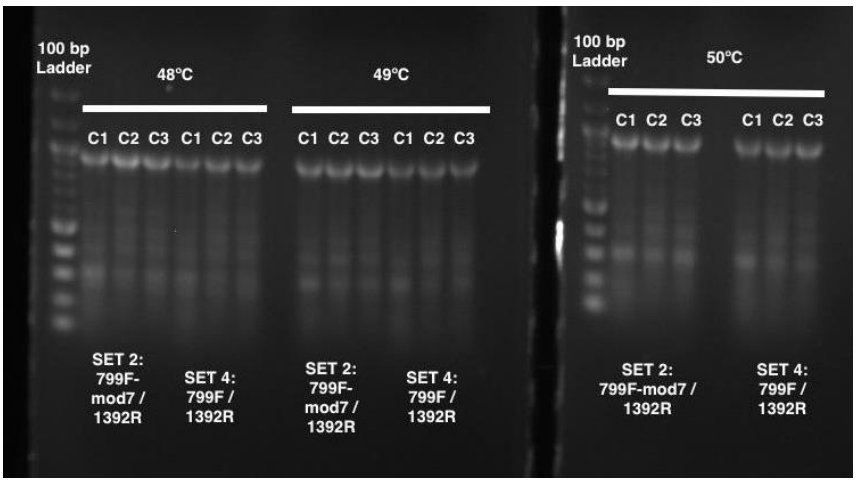


Figure 10: *2% agarose gels electrophoresis for primers set 2 and 4 using three different annealing temperatures (48ºC, 49ºC, 50ºC). In the first column of the two gels there is a 100 base pairs ladder. Every set of primers were tested in the three different samples (C1, C2 and C3).*

5.3 Taxonomy study of the PCR based approach

After working with the results of the sequencing process using the software Geneious version 10.1.3 (http://www.geneious.com) (Kearse*, et al.* 2012), for general bacterial primers, all three sequences were homologous despite being from different samples of distinct parts of the growth zone area (C1, C2 & C3). A consensus sequence was generated and the following taxonomic studies were done using it *(Table 2)*.

Table 2: *High-quality consensus sequence generated using bacterial general primers for samples C1, C2 and C3 and its length.*

Sample	Sequence (5' → 3')	Length
General primer consensus	NGGTATCTAATCCCATTTGCTCCCCTAGCTTTCGTCTCTCAGTGTCAGTGTCGGCCCAGCAGAGTG CTTTCGCCGTTGGTGTTCTTTCCGATCTCAATGCATTTCACCGCTCCACCGGAAATTCCCTCTGCC CCTACCGTACTCCAGCTTGGTAGTTTCCACCGCCTGTCCAGGGTTGAGCCCTGGGATTTGACGGC GGACTTGAAAAGCCACCTACAGACGCTTTACGCCCAATCATTCCGGATAACGCTTGCATCCTCTGT CTTACCGCGGCTGCTGGCACAGAGTTAGCCGATGCTTATTCCTCAGATACCGTCATTGTTTCTTCT CCGAGAAAAGAAGTTGACGACCCGTAGGCCTTCCACCTCCACGCGGCATTGCTCCGTCAGGCTTT CGCCCATTGCGGAAAATTCCCCACTGCN	422

For specific bacterial primers, as set 2 and 5 samples displayed identical sequences, a consensus sequence was generated and it was used for the taxonomic studies. For primers set 3 and 4, the three samples had similar sequences *(Table 3)* but they were not completely homologous. It was decided to work with each one of them to see if the differences between these sequences are reflected in the taxonomic alignments.

Notice that sequences from the sample C1, C2 and C3 of the primer set 3 bottom amplification (B-150 bp), are not included. The same is seen with the sequences from primer set 1. The reason is that the sequence obtained did not accomplish our quality standards, meaning that the nucleotide content in most of the positions was ambiguous. Not working with these sequences was decided to avoid getting uncertain results. However, this low-quality results, for samples from primers set 3, were expected because in the gel it can be seen how the band is less bright compared with the top one (1000 bp), which means there is less DNA amplification (*Table 3*).

A consensus of all the previous sequences was generated and was also used for the taxonomic studies (*Table 4*)

Table 3: *List of the different sets of primers used and the high-quality sequences obtained: consensus or individual for every sample. In the last column there is the length, the number of nucleotides, of each sequence.*

	Sample	Sequences (5' → 3')	Length
SET 2	Consensus	ACCACCGGGCGGCTGGCCTTCATGCCRAGTCTTCCTCCGCCGCCAACTCGACGTCGTCGTAACCAAGCCTAACCAAACCTCCATG CTCACTGGTACTTTGACRWCACTATAGGTAGGTCTCCGTGGGTCTTGAGTTCGGCAAGACGACTTCGGTTCACACGTGAAAGTGCT TCGAGGCTCCCAATGGTGAAATTCTTGCGGAAAGCACAGCTACGTGCTGGCACTCAATCAAGTAGTAGCGCTGGCACGTCACTCG GCTCCTCGGCTATTTCTCCTTAGGCGCATGTCTCAGCAACACAAAACGAGGGTTTCGCTCGTTATAGGACTTGACCAAACATCTCAC GACACGAGCTGACGACAGCCATGCAGCACCTGTATGAAAGTAAGTACCATCCCGTTAAAGACAGGTTTTGTTGTTCATATGTCAAG GGCTGGTAAGGTTTTGCGCGTTGTATCGAATTAAACCACATGCTCCACCGCTTGTGCAGGCCCCCGTCAATTCCTTTGAGTTTCGG TCTTGCGACCGTACTCCCCAGCGGAGTGTTTCACGCGTTAGCTGGGCCCCTGATCCGCGTAGCGTAG	582
SET 3	C1 - T	ACTAGCGATTTCTACTTCATGTTCCCGAGTTGCAGAGAACAATCCGAACTGAGGCAATCTTTCCGGATTCGCTCCGCCTTACAGCCT TGCTTCCCATTGTCATTGCCATTGTAGCACGTGTGTGGCCCAGCCCATAAGGGCCATGCGGACTTGACGTCATCCCCACCTTCCTC CAGTATCTCACTGGCAGTCCCTCGTGAGTGCGGCACGCACCTTTTTCTTTGTTTCGGAGCGGGGCGCGTACTATTACCACTACGTA CCACACCACCGGGCGGCTGGCCTTCATGCCGAGTCTTCCTCCGCCGCCAACTCGACGTCGTCGTAACCAAGCCTAACCAAACCTC CATGCTCACTGGTACTTTGACGTCACTATAGGTAGGTCTCCGTGGGTCTTGAGTTCGGCAAGACGACTTCGGTTCACACGTGAAAG TGCTTCGAGGCTCCCAATGGTGAAATTCTTGCGGAAAGCACAGCTACGTGCTGGCACTCAATCAAGTAGTAGCGCTGGCACGTCAC TCGGCTCCTCGGCTATTTCTCCTTAGGCGCATGTCTCAGCAACACAAAACGAGGGTTTCGCTCGTTATAGGACTTGACCAAACATCT CACGACACGAGCTGACGACAGCCATGCAGCACCTGTATGAAAGTAAGTACCATCCCCGTTAAAGACAGGTTTTGTTGTTCATATGTC	690
	C2 - T	CGCGATTACTAGCGATTTCTACTTCATGTTCCCGAGTTGCAGAGAACAATCCGAACTGAGGCAATCTTTCCGGATTCGCTCCGCCTT ACAGCCTTGCTTCCCATTGTCATTGCCATTGTAGCACGTGTGTGGCCCAGCCCATAAGGGCCATGCGGACTTGACGTCATCCCCAC CTTCCTCCAGTATCTCACTGGCAGTCCCTCGTGAGTGCGGCACGCACCTTTTTCTTTGTTTCGGAGCGGGGCGCGTACTATTACCA CTACGTACCACACCACCGGGCGGCTGGCCTTCATGCCGAGTCTTCCTCCGCCGCCAACTCGACGTCGTCGTAACCAAGCCTAACC AAACCTCCATGCTCACTGGTACTTTGACGTCACTATAGGTAGGTCTCCGTGGGTCTTGAGTTCGGCAAGACGACTTCGGTTCACAC GTGAAAGTGCTTCGAGGCTCCCAATGGTGAAATTCTTGCGGAAAGCACAGCTACGTGCTGGCACTCAATCAAGTAGTAGCGCTGGC ACGTCACTCGGCTCCTCGGCTATTTCTCCTTAGGCGCATGTCTCAGCAACACAAAACGAGGGTTTCGCTCGTTATAGGACTTGACC AAACATCTCACGACACGAGCTGACGACAGCCATGCAGCACCTGTATGAAAGTAAGTACCATCCCGT	668
	C3 - T	CGCGATTACTAGCGATTTCTACTTCATGTTCCCGAGTTGCAGAGAACAATCCGAACTGAGGCAATCTTTCCGGATTCGCTCCGCCTT ACAGCCTTGCTTCCCATTGTCATTGCCATTGTAGCACGTGTGTGGCCCAGCCCATAAGGGCCATGCGGACTTGACGTCATCCCCAC CTTCCTCCAGTATCTCACTGGCAGTCCCTCGTGAGTGCGGCACGCACCTTTTTCTTTGTTTCGGAGCGGGGCGCGTACTATTACCA CTACGTACCACACCACCGGGCGGCTGGCCTTCATGCCGAGTCTTCCTCCGCCGCCAACTCGACGTCGTCGTAACCAAGCCTAACC AAACCTCCATGCTCACTGGTACTTTGACGTCACTATAGGTAGGTCTCCGTGGGTCTTGAGTTCGGCAAGACGACTTCGGTTCACAC GTGAAAGTGCTTCGAGGCTCCCAATGGTGAAATTCTTGCGGAAAGCACAGCTACGTGCTGGCACTCAATCAAGTAGTAGCGCTGGC ACGTCACTCGGCTCCTCGGCTATTTCTCCTTAGGCGCATGTCTCAGCAACACAAAACGAGGGTTTCGCTCGTTATAGGACTTGACC AAACATCTCACGACACGAGCTGACGACAGCCATGCAGCACCTGTATGAAAGTW	655
SET 4	C1	ACCTTTTTCTTTGTTTCGGAGSGGGGCGCGTACTATTACCACTACGTACCACACCACCGGGCGGCTGGCCTTCATGCCGAGTCTTC CTCCGCCGCCAACTCGACGTCGTCGTAACCAAGCCTAACCAAACCTCCATGCTCACTGGTACTTTGACRWCACTATAGGTAGGTCT CCGTGGGTCTTGAGTTCGGCAAGACGACTTCGGTTCACACGTGAAAGTGCTTCGAGGCTCCCAATGGTGAAATTCTTGCGGAAAGC ACAGCTACGTGCTGGCACTCAATCAAGTAGTAGCGCTGGCACGTCACTCGGCTCCTCGGCTATTTCTCCTTAGGCGCATGTCTCAG CAACACAAAACGAGGGTTTCGCTCGTTATAGGACTTGACCAAACATCTCACGACACGAGCTGACGACAGCCATGCAGCACCTGTAT GAAAGTAAGTACCATCCCGTTAAAGACAGGTTTTGTTGTTCATATGTCAAGGGCTGGTAAGGTTTTGCGCGTTGTATCGAATTAAAC CACATGCTCCACCGCTTGTGCAGGCCCCCGTCAATTCCTTTGAGTTTCGGTCTTGCGACCGTACTCCCCAGCGGAGTGTTTCACGC GTTAGCTGGGCCCCTGATCCGCGTAGCGTAGCGTAGACAGACCAAGGCGAACACTCATCGTTTA	667
	C2	CGCGGATCAGGGGCCCAGCTAACGCGTGAAACACTCCGCCTGGGGAGTACGGTCGCAAGACCGAAACTCAAAGGAATTGACGGG GGCCTGCACAAGCGGTGGAGCATGTGGTTTAATTCGATACAACGCGCAAAACCTTACCAGCCCTTGACATATGAACAACAAAACCT GTCTTTAACGGGATGGTACTTACTTTCATACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTTGGTCAAGTCCTATA ACGAGCGAAACCCTCGTTTTGTGTTGCTGAGACATGCGCCTAAGGAGAAATAGCCGAGGAGCCGAGTGACGTGCCAGCGCTACTA CTTGATTGAGTGCCAGCACGTAGCTGTGCTTTCCGCAAGAATTTCACCATTGGGAGCCTCGAAGCACTTTCACGTGTGAACCGAAG TCGTCTTGCCGAACTCAAGACCCACGGAGACCTACCTATAGTGACGTCAAAGTACCAGTGAGCATGGAGGTTTGGTTAGGCTTGGT TACGACGACGTCGAGTTGGCGGCGGAGGAAGACTCGGCATGAAGGCCAGCCGCCCGGTGGTGTGGTACGTAGTGGTAATAGTAC GCGCCCCGCTCCGAAACAAAGAAAAAGGTG	628
	C3	CACCTTTTTCTTTGTTTCGGAGCGGGGCGCGTACTATTACCACTACGTACCACACCACCGGGCGGCTGGCCTTCATGCCRAGTCTT CCTCCGCCGCCAACTCGACGTCGTCGTAACCAAGCCTAACCAAACCTCCATGCTCACTGGTACTTTGACRWCACTATAGGTAGGTC TCCGTGGGTCTTGAGTTCGGCAAGACGACTTCGGTTCACACGTGAAAGTGCTTCGAGGCTCCCAATGGTGAAATTCTTGCGGAAAG CACAGCTACGTGCTGGCACTCAATCAAGTAGTAGCGCTGGCACGTCACTCGGCTCCTCGGCTATTTCTCCTTAGGCGCATGTCTCA GMAACACAAAACGAGGGTTTCGCTCGTTATAGGACTTGACCAAACATCTCACGACACGAGCTGACGACAGCCATGCAGCACCTGTA TGAAAGTAAGTACCATCCCGTTAAAGACAGGTTTTGTTGTTCATATGTCAAGGGCTGGTAAGGTTTTGCGCGTTGTATCGAATTAAA CCACATGCTCCACCGCTTGTGCAGGCCCCCGTCAATTCCTTTGAGTTTCGGTCTTGCGACCGTACTCCCCAGCGGAGTGTTTCACG CGTTAGCTGGGCCCCTGATCCGCGTAGCGTAGCGTAGACAGACCAA	649
SET 5	Consensus	CCCCTACACACGAAATTTCCACTCTCCTCTGTCTCACTCAAGTGAATYGGTTTCGAGAGCATTCCGCCACTTTTTGGCGACTTTCAC TTTCAACCCGATTCACCGCCTACGTGCCCTTTACGCCCAGTCATTCCGAAR	138

Table 4: *Consensus sequences created using the previous ones (Table 3). In the last column there is the length, the number of nucleotides, of each sequence.*

Sample	Sequences (5' → 3')	Length
Consensus of specific primers sequences	CGCGATTACTAGCGATTTCTACTTCATGTTCCCGAGTTGCAGAGAACAATCCGAACTGAGGCAATCTTTCCGGATTCGCTCCGCCTT ACAGCCTTGCTTCCCATTGTCATTGCCATTGTAGCACGTGTGTGGCCCAGCCCATAAGGGCCATGCGGACTTGACGTCATCCCCAC CTTCCTCCAGTATCTCACTGGCAGTCCCTCGTGAGTGCGGCACGCACCTTTTTCTTTGTTTCGGAGCGGGGCGCGTACTATTACCA CTACGTACCACACCACCGGGCGGCTGGCCTTCATGCCGAGTCTTCCTCCGCCGCCAACTCGACGTCGTCGTAACCAAGCCTAACC AAACCTCCATGCTCACTGGTACTTTGACGTCACTATAGGTAGGTCTCCGTGGGTCTTGAGTTCGGCAAGACGACTTCGGTTCACAC GTGAAAGTGCTTCGAGGCTCCCAATGGTGAAATTCTTGCGGAAAGCACAGCTACGTGCTGGCACTCAATCAAGTAGTAGCGCTGGC ACGTCACTCGGCTCCTCGGCTATTTCTCCTTAGGCGCATGTCTCAGCAACACAAAACGAGGGTTTCGCTCGTTATAGGACTTGACC AAACATCTCACGACACGAGCTGACGACAGCCATGCAGCACCTGTATGAAAGTAAGTACCATCCCGTTAAAGACAGGTTTTGTTGTTC ATATGTCAAGGGCTGGTAAGGTTTTGCGCGTTGTATCGAATTAAACCACATGCTCCACCGCTTGTGCAGGCCCCCGTCAATTCCTTT GAGTTTCGGTCTTGCGACCGTACTCCCCAGCGGAGTGTTTCACGCGTTAGCTGGGCCCCTGATCCGCGTAGCGTAGCGTAGACAG ACCAAGGCGAACACTCATCGTTTA	885

Firstly, the sequences were blasted (Boratyn*, et al.* 2012) with "Highly similar sequences (megablast)" selected. The best hit alignments were always for *Zea mays* (different subspecies) mitochondria (*Table 5*).

The best alignment for set 5's sequence was *Diplostephium hartwegii*, an unexpected result. Although the set 5 sequence presented this alignment, it also had plenty blast hits with different kinds of uncultured bacteria *(Table 5)*. At the same time, the consensus of specific bacterial primers (B) was also blasted (Boratyn*, et al.* 2012) in the same conditions *(Table 6)*.

If a deep look is taken in the results, it can be observed that set of primers 2, 4 (*Table 5*) and the consensus sequence (*Table 6*) present their top alignments with the same organisms. The only difference observed is the value of the max score, total score, and query cover. This is because of the different length of the sequences and their dissimilarities. In addition, for every set of primers, the three kinds of samples used had their top alignments with the same organisms. For example, sequences from C1, C2 and C3 of set 4, displayed the same organism in their best alignment. The only difference was in the value of the max score, total score, and query cover due to the fact they are not completely the same and had some differences both in the content of the sequence and in the length of it.

Table 5: *The top 5 hits results of blasting* (Boratyn, *et al.* 2012) *all sequences using NCBI's database with their corresponding description, maximum score, total score, query cover, E-value, Identity (Ident) and Accession number. These results were obtained after performing a general BLAST* (Boratyn, *et al.* 2012) *with the option "Highly similar sequences (megablast)" selected.C1, C2 and C3 sequences from set of primers 3 showed the same result so they are displayed together. This was also observed for the three different sequences of set 4. For both sets of primers the values in the table correspond to C1 results.*

Set	Sample	Description	Max Score	Total Score	Query value	E-value	Ident	Accession number
2	Cons.	Zea mays subsp. mays genotype,CMS-S mitochondrion, complete genome	1059	1126	100%	0.0	99%	DQ490951.2
		Zea mays subsp. parviglumis mitochondrion, complete genome	1059	1194	100%	0.0	99%	DQ645539.1
		Zea mays subsp. mays genotype CMS-T mitochondrion, complete genome	1059	1126	100%	0.0	99%	DQ490953.1
		Zea mays subsp. mays genotype male-fertile NA mitochondrion, complete genome	1059	2253	100%	0.0	99%	DQ490952.1
		Zea mays subsp. mays mitochondrion, complete genome	1059	1301	100%	0.0	99%	NC007982.1
3	C1 C2 C3	Zea mays clone 1615969 mRNA sequence	1267	1267	100%	0.0	99%	EU957780.1
		Zea mays subsp. mays genotype CMS-S mitochondrion, complete genome	1267	1335	100%	0.0	99%	DQ490951.2
		Zea mays subsp. parviglumis mitochondrion, complete genome	1267	1403	100%	0.0	99%	DQ645539.1
		Zea mays subsp. mays genotype CMS-T mitochondrion, complete genome	1267	1335	100%	0.0	99%	DQ490953.1
		Zea mays subsp. mays genotype male-fertile NA mitochondrion, complete genome	1267	2671	100%	0.0	99%	DQ490952.1
4	C1 C2 C3	Zea mays subsp. mays genotype,CMS-S mitochondrion, complete genome	1212	1280	100%	0.0	99%	DQ490951.2
		Zea mays subsp. parviglumis mitochondrion, complete genome	1212	1347	100%	0.0	99%	DQ645539.1
		Zea mays subsp. mays genotype CMS-T mitochondrion, complete genome	1212	1280	100%	0.0	99%	DQ490953.1
		Zea mays subsp. mays genotype male-fertile NA mitochondrion, complete genome	1212	2560	100%	0.0	99%	DQ490952.1
		Zea mays subsp. mays mitochondrion, complete genome	1212	1454	100%	0.0	99%	NC007982.1
5	Cons.	Diplostephium hartwegii voucher TEX:Vargas 456 mitochondrion, complete genome	243	243	99%	5e-60	99%	KX063855.1
		Zea mays isolate SM10 mitochondrion sequence	243	243	99%	5e-60	99%	KY018916.1
		Uncultured bacterium clone OTU_532 16S ribosomal RNA gene, partial sequence	243	243	99%	5e-60	99%	KY671959.1
		Uncultured bacterium clone OTU374 16S ribosomal RNA gene, partial sequence	243	243	99%	5e-60	99%	KY466320.1
		Uncultured bacterium clone OTU_4569 16S ribosomal RNA gene, partial sequence	243	243	99%	5e-60	99%	KX972094.1

Table 6: *The top 5 hits results of blasting* (Boratyn*, et al.* 2012) *the consensus sequence from specific bacterial primers (B), using NCBI's database with their corresponding description, maximum score, total score, query cover, E-value, Identity (Ident) and Accession number. These results were obtained after performing a general BLAST* (Boratyn*, et al.* 2012) *with the option "Highly similar sequences (megablast)" selected.*

Sample	Description	Max Score	Total Score	Query value	E-value	Ident	Accesion number
Specific primer consensus	Zea mays subsp. mays genotype,CMS-S mitochondrion, complete genome	1626	1693	100%	0.0	99%	DQ490951.2
	Zea mays subsp. parviglumis mitochondrion, complete genome	1626	1761	100%	0.0	99%	DQ645539.1
	Zea mays subsp. mays genotype CMS-T mitochondrion, complete genome	1626	1693	100%	0.0	99%	DQ490953.1
	Zea mays subsp. mays genotype male-fertile NA mitochondrion, complete genome	1626	3387	100%	0.0	99%	DQ490952.1
	Zea mays subsp. mays mitochondrion, complete genome	1626	1868	100%	0.0	99%	NC_007982.1

When the taxonomic study was done in the same conditions but for the consensus sequences from general bacterial primers *(Table 7)*, the best 5 alignments were for different clones of uncultured bacteria. It was observed that the first and the second best alignment described the same organisms but with different accession numbers. The same happened with the third and fourth best alignments (both accession numbers are included in *Table 7*). This can easily be explained as NCBI is a big database where every scientist can summit their organisms sequence and each of them get an individual and unique accession number. If a sequence was previously summited, it is frequently common to see the same organism description but with different accession numbers. Following the uncultured bacteria top alignments, there are hits with chloroplast of two different plant species.

Table 7: *Top 7 alignments after doing a blast* (Boratyn*, et al.* 2012) *with the NCBI database using the sequences obtained from general bacterial kind of primers (A). There is also the corresponding description of the organisms and the following parameters: maximum score, total score, query cover, E-value, Identity and Accession number. To get these results the only parameter selected was "Highly similar sequences (megablast)".*

Sample	Description	Max Score	Total Score	Query value	E-value	Ident	Accesion number
General primer Consensus	Uncultured bacterium partial 16S rRNA gene, clone C60.16_2149289	776	776	99%	0.0	100%	LT720106.1 LT721231.1
	Uncultured bacterium clone MD01e11_13424 16S ribosomal RNA gene, partial sequence	776	776	99%	0.0	100%	JQ383750.2 JQ383731.2
	Uncultured bacterium clone Otu00501 16S ribosomal RNA gene, partial sequence	774	774	99%	0.0	100%	KX990793.1
	Saccharum spontaneum chloroplast, complete genome	771	1542	99%	0.0	99%	KX139746.1
	Arundo donax voucher RSA: D. Bell 4664 chloroplast, complete genome	771	1542	99%	0.0	99%	KX229727.1

In order to get rid of all the mitochondria and chloroplast's alignments, a second BLAST (Boratyn*, et al.* 2012) was done. But this time, plants were excluded and the parameter highly similar was maintained as selected (as the one performed during the taxonomic study of the *NGS sequencing-based endophyte prediction – Figure 5*). As it can be seen in *Table 8* and *Table 9*, there is no sequence related with 16S rRNA genes from chloroplast. This means the primers were used in the correct conditions and the samples were not contaminated at any moment of the procedure by chloroplast. Also, and as expected, this is means the primers had the perfect design and accomplish our expectations. In addition, sequences from set of primers 2, 3 and 4 and the consensus show their best alignment with the same uncultured bacteria clone: Uncultured bacterium clone FFCH13347 16S ribosomal RNA gene, partial sequence (Accession number: EU133393.1).

For general bacterial primers (A), 16S rRNA genes from chloroplast could not be seen (*Table 10*). The first three hits were for the same bacterial organisms observed previously (*Table 7*).

For both kinds of primers, general (A) and specific (B), the best alignments displayed were with different uncultured bacterial organisms.

Table 8: The top 5 hits results of blasting (Boratyn, *et al.* 2012) all sequences using NCBI's database with their corresponding description, maximum score, total score, query cover, E-value, Identity (Ident) and Accession number. These results were obtained after performing a general BLAST (Boratyn, *et al.* 2012) with the option "Highly similar sequences (megablast)" selected and excluding plant kingdom.

Set	Sample	Description	Max Score	Total Score	Query value	E-value	Ident	Accession number
2	Cons.	Uncultured bacterium clone FFCH13347 16S ribosomal RNA gene, partial sequence	990	990	99%	0.0	97%	EU133393.1
		Uncultured bacterium clone Ph.pi-F-DM-HN-5-132 16S ribosomal RNA gene, partial sequence	564	564	66%	3e-157	93%	HQ639583.1
		Uncultured bacterium clone Ce.so-M-DM-HN-2-90 16S ribosomal RNA gene, partial sequence	558	558	66%	1e-155	93%	HQ639494.1
		Uncultured bacterium clone FPURT2-G08 16S ribosomal RNA gene, partial sequence	556	556	63%	5e-155	93%	GU166638.1
		Cordiimonas spp. clone DA162 16S ribosomal RNA gene, partial sequence	556	556	63%	5e-155	93%	FJ388347.1
3	C1 C2 C3	Uncultured bacterium clone FFCH13347 16S ribosomal RNA gene, partial sequence	1190	1190	98%	0.0	98%	EU133393.1
		Uncultured bacterium clone BIGO434 16S ribosomal RNA gene, partial sequence	494	755	81%	4e-136	94%	HM558599
		Uncultured bacterium clone BIGH1473 16S ribosomal RNA gene, partial sequence	494	772	79%	4e-136	94%	HM557934.1
		Uncultured bacterium clone BIGH1321 16S ribosomal RNA gene, partial sequence	494	777	79%	4e-136	94%	HM557868.1
		Uncultured bacterium clone BICH750 16S ribosomal RNA gene, partial sequence	488	783	81%	2e-134	93%	HM556942.1
4	C1 C2 C3	Uncultured bacterium clone FFCH13347 16S ribosomal RNA gene, partial sequence	1086	1086	94%	0.0	97%	EU133393.1
		Uncultured bacterium clone Ph.pi-F-DM-HN-5-132 16S ribosomal RNA gene, partial sequence	564	685	75%	3e-157	93%	HQ639583.1
		Uncultured bacterium clone Ce.so-M-DM-HN-2-90 16S ribosomal RNA gene, partial sequence	558	679	75%	1e-155	93%	HQ639494.1
		Uncultured bacterium clone FPURT2-G08 16S ribosomal RNA gene, partial sequence	556	678	69%	5e-155	93%	GU166638.1
		Cordiimonas spp. clone DA162 16S ribosomal RNA gene, partial sequence	556	678	69%	5e-155	93%	FJ388347.1
5	Cons.	Uncultured bacterium clone OTU_532 16S ribosomal RNA gene, partial sequence	243	243	99%	2e-60	99%	KY671959.1
		Uncultured bacterium clone OTU374 16S ribosomal RNA gene, partial sequence	243	243	99%	2e-60	99%	KY466320.1
		Uncultured bacterium clone OTU_4569 16S ribosomal RNA gene, partial sequence	243	243	99%	2e-60	99%	KX972094.1
		Uncultured bacterium clone Otu02783 16S ribosomal RNA gene, partial sequence	243	243	99%	2e-60	99%	KX997563.1
		Uncultured bacterium clone Otu01867 16S ribosomal RNA gene, partial sequence	243	243	99%	2e-60	99%	KX996840.1

Table 9*: The top 5 hits results of blasting* (Boratyn*, et al.* 2012) *the consensus sequence from specific bacterial primers (B), using NCBI's database with their corresponding description, maximum score, total score, query cover, E-value, Identity (Ident) and Accession number. These results were obtained after performing a general BLAST* (Boratyn*, et al.* 2012) *with the option "Highly similar sequences (megablast)" selected and excluding plant kingdom.*

Sample	Description	Max Score	Total Score	Query value	E-value	Ident	Accesion number
Specific primers consensus	Uncultured bacterium clone FFCH13347 16S ribosomal RNA gene, partial sequence	1469	1469	94%	0.0	98%	EU133393.1
	Uncultured bacterium clone Ph.pi-F-DM-HN-5-132 16S ribosomal RNA gene, partial sequence	564	1051	81%	4e-157	93%	HQ639583.1
	Uncultured bacterium clone Ce.so-M-DM-HN-2-90 16S ribosomal RNA gene, partial sequence	558	1045	81%	2e-155	93%	HQ639494.1
	Uncultured bacterium clone FPURT2-G08 16S ribosomal RNA gene, partial sequence	556	838	66%	7e-155	93%	GU166638.1
	Cordiimonas spp. clone DA162 16S ribosomal RNA gene, partial sequence	556	1047	79%	7e-155	93%	FJ388347.1

Table 10: *Top 7 alignments after doing a blast* (Boratyn*, et al.* 2012) *with the NCBI database using the sequences obtained from general bacterial kind of primers (A).There is also the corresponding description of the organisms and the following parameters: maximum score, total score, query cover, E-value, Identity and Accession number. To get these results the parameter selected was "Highly similar sequences (megablast)" and the plant kingdom was completely excluded.*

Sample	Description	Max Score	Total Score	Query value	E-value	Ident	Accesion number
General primers Consensus	Uncultured bacterium partial 16S rRNA gene, clone C60.16_2149289	776	776	99%	0.0	100%	LT720106.1 LT721231.1
	Uncultured bacterium clone MD01e11_13424 16S ribosomal RNA gene, partial sequence	776	776	99%	0.0	100%	JQ383750.2 JQ383731.2
	Uncultured bacterium clone Otu00501 16S ribosomal RNA gene, partial sequence	774	774	99%	0.0	100%	KX990793.1
	Uncultured bacterium clone T-RFLP_clone_K44 16S ribosomal RNA gene, partial sequence	771	771	99%	0.0	99%	KP780113.1
	Uncultured bacterium gene for 16S rRNA, partial sequence, clone: OYMC-Endo-CLN14	771	771	99%	0.0	99%	LC031355.1

6. Discussion

The aim of this study is to find evidence of possible endophytes in the leaf growth zone of maize. In order to accomplish this, a NGS sequencing-based prediction was done. Then a PCR based approach with samples from different B73 genotype plants was performed. New sequences were obtained and analysed again. The purpose of doing this work was to get the same bacterial evidence for both approaches and then, strongly affirm the presence of bacterial endophytes in the leaf growth zone of maize plants. Also, these results indicated that our methodology worked and could be used for future studies. All samples taken and residual components generated during this project were ethically treated.

In the first NGS sequencing-based prediction, *Desulfitobacterium hafniense* was the only bacterial species observed in the alignments. In order to support these results, a PCR based approach was designed. As stated before, two kinds of primers were used: general bacterial primers (A) and specific bacterial primers (B). The second one could be divided in five distinct set of primers identified as set 1, 2, 3, 4, and 5. After optimising the conditions for the PCR and sending the products of it for sequencing, the second bioinformatics/taxonomic study was performed.

First, one fact noticed when analysing the sequences, is that it does not matter if they are from region C1, C2 or C3 of the leaf growth zone as the sequences were completely homologous or the dissimilarities between them were negligible. We can affirm the dissimilarities are insignificant because the variation between sequences of the same primers do not affect the alignment results. In other words, the best alignments were with the same organisms.

Secondly, as expected, the general bacterial primers (A) had alignments with genes from plant chloroplast regions. This result was expected because bacteria and chloroplasts are close, in evolutionarily terms, and their 16S rRNA genes are almost homologous. However, the best hits in the BLAST (Boratyn*, et al.* 2012) analyses were with different kinds of uncultured bacteria. When the BLAST (Boratyn*, et al.* 2012) was repeated but excluding plants, the three best alignments were with the same uncultured bacteria. With these results, it cannot be confirmed if the presence of uncultured bacteria in the leaf growth zone of maize are true because the other alignments obtained had really high scores too.

Specific bacterial primers (B) were distinguishing from the previous ones because they are modified specifically to avoid chloroplast 16S rRNA gene amplification. As seen in the previous results *(Tables 5, 6, 8 and 9)*, no alignment with chloroplast was detected. This means the specific bacterial primers were correctly used and they work perfectly. Still, evidence of bacterial endophytes could not be seen, as the best alignments were with uncultured bacterium clone FFCH13347 16S ribosomal RNA gene, partial sequence (Accession number: EU133393.1).

Although high score hits were revealed, the wide diversity of organisms in the best alignments does not allow us to affirm this uncultured bacterium can be found in the plant as an endophyte. Also, as we did not get the same results from the taxonomic reports of the NGS prediction and from the PCR based approach, we cannot affirm the presence of either *Desulfitobacterium hafniense* or uncultured bacterium clone FFCH13347. Even it seems after all this work there is some evidence of the presence of bacterial endophytes in the maize genome from leaf growth zone, our results are not significant enough to make a confirmation of these microorganisms in the maize plant (*Figure 11).*

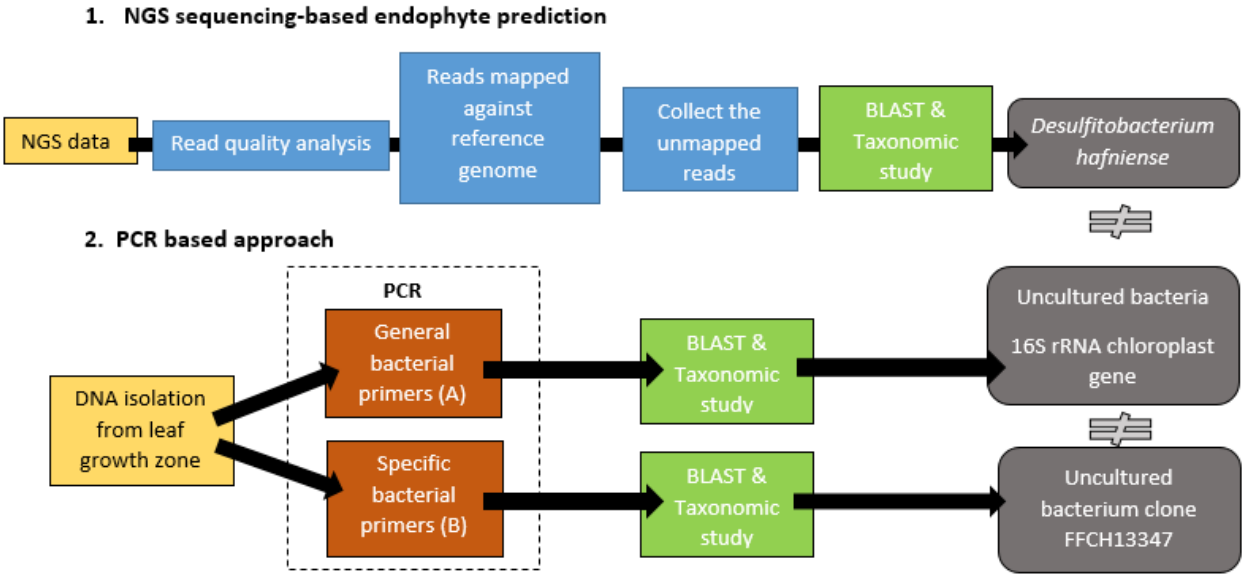


Figure 11: *Schematic workflow to search the endophytic bacterial communities and the results of this study.*

Nowadays, there are still a lot of bacteria, which have to be discovered and described. Thanks to sequencing techniques we can have a vision of how wide and unknown the bacterial world is, but further studies and discovers are needed to have a deeper look in this topic in order to be able to cultivate them. Then, a better description of these bacteria can be done to see if they can live inside a plant as an endophyte.

7. Conclusió [CAT]

El principal objectiu d'aquest treball era trobar alguna evidencia de endòfits bacterians en la zona de creixement de les fulles del blat de moro, i a la vegada, desenvolupar un protocol basat en la tècnica de PCR per tal d'identificar-los. Quan es revisen publicacions prèvies sobre el tema, es pot observar com hi ha varies evidències de la presència d'endòfits bacterians en plantes. Tot i això, després de tot el treball realitzat en aquest estudi, no podem afirmar que aquests endòfits bacterians puguin trobar-se en la zona de creixement de les fulles de les plantes de blat de moro.

A la vegada, suggerim que en futures anàlisis sobre el tema, utilitzin tècniques no relacionades amb la PCR, ja que no podem assegurar que funcionin. Si, tot i això, es necessita fer una PCR, aconsellem utilitzar primers que evitin l'amplificació tant dels gens 16S rRNA de cloroplast com de mitocòndria.

En resum, el desenvolupament d'una nova forma de trobar endòfits bacterians no ha tingut èxit i, com a conseqüència, no s'ha pogut trobar cap evidència d'endòfits bacterians en la zona de creixement de les fulles de blat de moro.

Conclusion [ENG]

The main objective of this work was to find evidence of bacterial endophytes in maize leaf growth zone and, at the same time, develop a protocol based in PCR techniques to identify them. When looking at previous publications about this topic, it can be seen that there are several evidences of bacterial endophytes in the plant. However, after all the work done in this study, we cannot affirm these bacterial endophytes can also be found in leaf growth zone of maize plants.

At the same time, we can suggest to further studies in this topic the use of other techniques not related with PCR like in this work because we cannot ensure it will work. If a PCR technique is required, using specific bacterial primers which avoids the amplification of both chloroplast and mitochondrial 16S rRNA genes are suggested.

To sum up, the development of a new way to find bacterial endophytes has not been successful and in consequence, no evidence of bacterial endophytes in leaf grown zone of maize plants has been found.

8. Bibliography

Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G., & Neufeld, J. D. (2011). Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Applied and Environmental Microbiology*, *77*(11), 3846–3852. https://doi.org/10.1128/AEM.02772-10

Battu, L., Reddy, M. M., Goud, B. S., Ulaganathan, K., & Kandasamy, U. (2016). Genome inside genome: NGS based identification and assembly of endophytic Sphingopyxis granuli and Pseudomonas aeruginosa genomes from rice genomic reads. *Genomics*. https://doi.org/10.1016/j.ygeno.2017.02.002

Benhizia, Y., Benhizia, H., Benguedouar, A., Muresu, R., Giacomini, A., & Squartini, A. (2004). Gamma Proteobacteria Can Nodulate Legumes of the Genus Hedysarum. *Systematic and Applied Microbiology*, *27*(4), 462–468. https://doi.org/10.1078/0723202041438527

Bennett, S. (2004). Solexa Ltd. *Pharmacogenomics*, *5*(4), 433–438. https://doi.org/10.1517/14622416.5.4.433

Bloemberg, G. V, & Lugtenberg, B. J. (2001). Molecular basis of plant growth promotion and biocontrol by rhizobacteria. *Current Opinion in Plant Biology*, *4*(4), 343–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11418345

Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., & Madden, T. L. (2012). Domain enhanced lookup time accelerated BLAST. *Biology Direct*, *7*(1), 12. https://doi.org/10.1186/1745-6150-7-12

Carson, W. P., Schnitzer, S. A., & Smithsonian Tropical Research Institute. (2008). *Tropical forest community ecology*. Wiley-Blackwell Pub.

Chelius, M. K., & Triplett, E. W. (2001). The Diversity of Archaea and Bacteria in Association with the Roots of Zea mays L. *Microbial Ecology*, *41*(3), 252–263. https://doi.org/10.1007/s002480000087

Coenye, T., & Vandamme, P. (2003). Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiology Letters*, *228*(1), 45–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14612235

Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, *8*(3), 175–85. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9521921

Glick, B. R., & R., B. (2012). Plant Growth-Promoting Bacteria: Mechanisms and Applications. *Scientifica*, *2012*, 1–15. https://doi.org/10.6064/2012/963401

Hallmann, J., Quadt-Hallmann, A., Mahaffee, W. F., & Kloepper, J. W. (1997). Bacterial endophytes in agricultural crops. *Canadian Journal of Microbiology*, *43*(10), 895–914. https://doi.org/10.1139/m97-131

Hamilton, C. E., Gundel, P. E., Helander, M., & Saikkonen, K. (2012). Endophytic mediation of reactive oxygen species and antioxidant activity in plants: a review. *Fungal Diversity*, *54*(1), 1–10.https://doi.org/10.1007/s13225-012-01589

Hanshew, A. S., Mason, C. J., Raffa, K. F., & Currie, C. R. (2013). Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. *Journal of Microbiological Methods*, *95*(2), 149–155. https://doi.org/10.1016/j.mimet.2013.08.007

Kavanová M, Grimoldi AA, Lattanzi FA, et al. (2006) Phosphorus nutrition and mycorrhiza effects on grass leaf growth. Pstatus- and size-mediated effects on growth zone kinematics. Plant Cell Environ29:511–20.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glockner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, *41*(1), e1–e1. https://doi.org/10.1093/nar/gks808

Kumar, P. S., Brooker, M. R., Dowd, S. E., & Camerlengo, T. (2011). Target Region Selection Is a Critical Determinant of Community Fingerprints Generated by 16S Pyrosequencing. *PLoS ONE*, *6*(6), e20956. https://doi.org/10.1371/journal.pone.0020956

Kusari, S., Hertweck, C., & Spiteller, M. (2012). Chemical ecology of endophytic fungi: Origins of secondary metabolites. *Chemistry and Biology*, *19*(7), 792–798. https://doi.org/10.1016/j.chembiol.2012.06.004

Long, H. H., Schmidt, D. D., & Baldwin, I. T. (2008). Native bacterial endophytes promote host growth in a species-specific manner; phytohormone manipulations do not result in common growth responses. *PloS One*, *3*(7), e2702. https://doi.org/10.1371/journal.pone.0002702

Muyzer, G., de Waal, E. C. and Uitterlinden, A. G. (1993) 'Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 59(3), pp. 695–700. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7683183 (Accessed: 10 July 2017).

Roesch, L. F. W., Camargo, F. A. O., Bento, F. M., & Triplett, E. W. (2008). Biodiversity of diazotrophic bacteria within the soil, root and stem of field-grown maize. *Plant and Soil*, *302*(1–2), 91–104. https://doi.org/10.1007/s11104-007-9458-3

Rosenblueth, M., & Martínez-Romero, E. (2006). Bacterial Endophytes and Their Interactions with Hosts. *Molecular Plant-Microbe Interactions*, *19*(8), 827–837. https://doi.org/10.1094/MPMI-19-0827

Santoyo, G., Moreno-Hagelsieb, G., del Carmen Orozco-Mosqueda, M., & Glick, B. R. (2016). Plant growth-promoting bacterial endophytes. *Microbiological Research*, *183*, 92–99. https://doi.org/10.1016/j.micres.2015.11.008

Walker, J. J. and Pace, N. R. (2007) 'Endolithic Microbial Ecosystems', *Annual Review of Microbiology*, 61(1), pp. 331–347. doi: 10.1146/annurev.micro.61.080706.093302.

Wilson, D. (1995). Endophyte: The Evolution of a Term, and Clarification of Its Use and Definition. *Oikos*, *73*(2), 274. https://doi.org/10.2307/3545919